The null distributions of popular distance functions for rank vectors

John I. Marden

February 8, 2024

Contents

Chapter 1

The distances

1.1 Introduction

The main goal of this paper is to gather results in the literature, and make some enhancements, on the null distributions of popular distance measures for ranking data. Diaconis (1988) has a thorough discussion of rank distances. See also Marden (1996).

We present methods for calculating the exact distributions for small values of m, the number of objects ranked, and asymptotic or other approximations for larger m. Implementations are found in the R package rankeeze. The null distributions are used for testing uniformity of the rank vectors, as well as the basis for the Mallows' parametric models.

As noted, we have m objects to rank. A rank vector is an m-dimensional vector $\boldsymbol{y} = (y_1, \ldots, y_m)$, where y_i denotes the rank of object i. We do not allow ties or missing values, so that each $\boldsymbol{y} \in \mathcal{P}_m$, where $\mathcal{P}_m$ the set of permutations of the integers from 1 to m. By the null distribution on the random vector $\boldsymbol{Y}$, we mean that $\boldsymbol{Y}$ is uniformly distributed over the set of permutations,

$$\boldsymbol{Y} \sim \text{Uniform}(\mathcal{P}_m). \tag{1.1}$$

Given a distance $d(\boldsymbol{y}, \boldsymbol{x})$ for $\boldsymbol{y}, \boldsymbol{x} \in \mathcal{P}_m$, we wish to find the distribution of $d(\boldsymbol{Y}, \boldsymbol{x})$ for $\boldsymbol{Y}$ as in (1.1). All the distances we consider are **label-invariant**, meaning the order in which the objects are numbered is irrelevant. Formally, if $\boldsymbol{Q}$ is any $m \times m$ permutation matrix (there is exactly one 1 in each row and each column), then d is label-invariant if $d(\boldsymbol{yQ}, \boldsymbol{xQ}) = d(\boldsymbol{y}, \boldsymbol{x})$ for all $\boldsymbol{y}, \boldsymbol{x} \in \mathcal{P}_m$. This property can be shown to imply that the distribution of $d(\boldsymbol{Y}, \boldsymbol{x})$ under (1.1) does not depend on $\boldsymbol{x} \in \mathcal{P}_m$. In fact, the distribution is the same even if $\boldsymbol{X}$ has a distribution, as long as it is independent of $\boldsymbol{Y}$. Thus it is enough to consider the distribution of

$$D \equiv d(\boldsymbol{Y}, \boldsymbol{\omega}), \quad \text{where} \quad \boldsymbol{\omega} = (1, 2, \ldots, m). \tag{1.2}$$

By the uniformity, we have

$$P[D = x] = \frac{\#\{\boldsymbol{y} \in \mathcal{P}_m \mid d(\boldsymbol{y}, \boldsymbol{\omega}) = x\}}{m!}. \tag{1.3}$$

A couple of the distances are also **rank-invariant**, hence are **bi-invariant**. See Section 1.7.

1.2 The distances

We have results on the seven popular distances $d(\boldsymbol{y}, \boldsymbol{x})$ listed in (1.4). There are many others, but these seem to cover the gamut quite well.

Name	Definition
Spearman	$\sum (y_i - x_i)^2$
Footrule	$\sum \lvert y_i - x_i \rvert$
Kendall	$\#\{1 \leqslant i < j \leqslant m \,\vert\, (y_i - y_j)(x_i - x_j) < 0\}$
Hamming	$\#\{i \,\vert\, y_i \neq x_i\}$
Cayley	$m - \#\text{Cycles}$
Ulam	$m -$ Length of longest increasing subsequence
Maximum	$\max\{\lvert x_i - y_i \rvert\}$

(1.4)

The Spearman distance is also known as Spearman's ρ distance, since it is related to that correlation coefficient (see Section 1.5), and the footrule is known as Spearman's footrule. See Spearman (1904). Likewise, Kendall's distance is related to Kendall's τ coefficient (M. G. Kendall, 1938). The maximum distance could also be called Chebyshev's distance.

The Cayley and Ulam distances need a little more explanation. First, suppose $\boldsymbol{x} = \boldsymbol{\omega}$. A **cycle** is a set of distinct integers $c_1, \ldots, c_K$ arranged in a cycle

$$c_1 \to c_2 \to c_3 \to \cdots \to c_K \to c_1, \tag{1.5}$$

where the beginning is arbitrary, i.e., $(c_2 \to c_3 \to \cdots \to c_K \to c_1 \to c_2)$ is the same cycle. The permutation has the cycle (1.5) if for some i,

$$y_{c_1} = c_2, y_{c_2} = c_3, \ldots, y_{c_{K-1}} = c_K \text{ and } y_{c_K} = c_1. \tag{1.6}$$

So if $\boldsymbol{y} = (3,5,2,7,1,4,6)$, the cycles are $(3 \to 2 \to 5 \to 1 \to 3)$ and $(7 \to 6 \to 4 \to 7)$. Each $\boldsymbol{y}$ has a unique decomposition into cycles, and each y_i appears in exactly one of the cycles. Cayley's distance subtracts the number of cycles from m, so that in this case $d_{\text{Cayley}}(\boldsymbol{y}, \boldsymbol{\omega}) = 7 - 2 = 5$. Note that $\boldsymbol{\omega} = (1,2,\ldots,m)$ has m cycles of one, $(1 \to 1)$, $(2 \to 2)$, etc, so that $d_{\text{Cayley}}(\boldsymbol{\omega}, \boldsymbol{\omega}) = 0$. This distance can also be defined as the minimum number of interchanges to bring $\boldsymbol{y}$ to $\boldsymbol{\omega}$. E.g.,

$$\boldsymbol{y} = (3,5,2,7,1,4,6) \to (1,5,2,7,3,4,6) \to (1,2,5,7,3,4,6) \to \tag{1.7}$$

$$(1,2,3,7,5,4,6) \to (1,2,3,4,5,7,6) \to (1,2,3,4,5,6,7), \tag{1.8}$$

which is indeed 5 interchanges. Kendall's distance is the minimum number of adjacent interchanges needed to bring $\boldsymbol{y}$ to $\boldsymbol{\omega}$.

An **increasing subsequence** in $\boldsymbol{y}$ is a subsequence $i_1 < i_2 < \cdots < i_K$ such that $y_{i_1} < y_{i_2} < \cdots < y_{i_K}$. Ulam's distance is $d_{\text{Ulam}}(\boldsymbol{y}, \boldsymbol{\omega}) = m - L_m$, where L_m is defined to be the longest such subsequence. For $\boldsymbol{y} = (3,5,2,7,1,4,6)$, the longest is $L_m = 3$. In fact, there are four subsequences of length three: (3,5,7), (3,4,6), (2,4,6), and (1,4,6). Thus $d_{\text{Ulam}}(\boldsymbol{y}, \boldsymbol{\omega}) = 7 - 3 = 4$.

To define the Ulam and Cayley distances for arbitrary $\boldsymbol{x}$, we first reorder $\boldsymbol{y}$ and $\boldsymbol{x}$ in concert until $\boldsymbol{x} \to \boldsymbol{\omega}$ and $\boldsymbol{y} \to \boldsymbol{y}^*$. Then $d(\boldsymbol{y}, \boldsymbol{x}) = d(\boldsymbol{y}^*, \boldsymbol{\omega})$. Formally, we find the permutation matrix $\boldsymbol{Q}$ for which $\boldsymbol{x}\boldsymbol{Q} = \boldsymbol{\omega}$, and define $\boldsymbol{y}^* = \boldsymbol{y}\boldsymbol{Q}$.

As mentioned above, all these distances are label-invariant. The Hamming and Cayley distances are also rank-invariant, hence bi-invariant.

1.3 Summary of results

For any distance, it is possible to find the density (1.3) by directly summing over all $\boldsymbol{y} \in \mathcal{P}_n$. Since there are $n!$ elements in $\mathcal{P}_n$, this approach is practical only for fairly small n, say $n \leqslant 10$ (depending on the speed of your computer and available time). Fortunately, there are short-cuts that allow us to find the exact distributions for values of $n \leqslant n^*$ for $n^* > 10$. When $n > n^*$, approximate and asymptotic results are available. The n^* for which the rankeeze package has exact values is given in the table in (1.9), along with approximate and/or asymptotic distributions used for larger n. ("Edgeworth" means using an Edgeworth expansion of the normal.)

Name	n^*	Approximating distribution	Asymptotic distribution
Spearman	24	Edgeworth	Normal
Footrule	350	Edgeworth	Normal
Kendall	1,000	Edgeworth	Normal
Hamming	25	Poisson	Poisson
Cayley	10,000	Edgeworth	Normal
Ulam	150	Gamma	Tracy-Widom
Maximum	24	See Section 10.4	$\sqrt{\text{Exponential}}$

(1.9)

For the Spearman, footrule, Ulam, and maximum distances, the smallish n^* are due to the limits of computational power. For the others, we could push up n^*, but the approximations work well even for $n \approx n^*$. In each case, the approximate or asymptotic distribution are based on scaled and shifted versions of the distances. Table 1.10 gathers together the first two moments of the distances, though we have only approximations for the Ulam and maximum distances, which are not especially accurate for small n. We also include the maximum value of the distances. For references and more details, see the specific chapters devoted to each distance. Note that for any distance, the variance is zero if $n = 1$.

	Maximum	Mean	Variance (if $n > 1$)
Spearman	$n(n^2-1)/3$	$n(n^2-1)/6$	$n^2(n-1)(n+1)^2/36$
Footrule	$\lfloor n^2/2 \rfloor$	$(n^2-1)/3$	$(n+1)(2n^2+7)/45$
Kendall	$n(n-1)/2$	$n(n-1)/4$	$n(n-1)(2n+5)/72$
Hamming	n	$n-1$	1
Cayley	$n-1$	$n - \sum_{i=1}^{n} 1/i$	$\sum_{i=1}^{n}(i-1)/i^2$
Ulam	$n-1$	$n - 2\sqrt{n} - n^{1/6}\mathrm{E}[W]^*$	$n^{1/3}\mathrm{Var}[W]^*$
Maximum	$n-1$	$n - \sqrt{n\pi/2}^*$	$n(1-\pi/4)^*$

$* =$ Asymptotic approximation

(1.10)

Here W is a Tracy-Widom random variable (Tracy & Widom, 1994), which we discuss further in Section 8.4. The first two moments are approximately $\mathrm{E}[W] \approx -1.7711$ and $\mathrm{Var}[W] \approx 0.8132$. Higher moments, and cumulants, which are needed in the Edgeworth expansions, are given in the individual chapters.

1.4 Practical support

The decision of which distance(s) to use can depend on a number of factors. If the scientific situation dictates a particular distance, then that is the one to use. In hypothesis testing, one would choose a distance for which the test based on it has good power, which would also depend on the particular situation. One feature that may be important based on just the null distribution is the size of the support. The more values a distance can take on, the better one can distinguish between rank vectors. Thus, e.g., we can find confidence percentages closer to 95%, or type I errors closer to 5% or 1%, or finer distinctions between p-values.

Table (1.10) indicates the maximum value each distance can take on. In each case, the minimum is 0, and all distances are integers, but the two Spearman distances have support on just the even integers, while the others have support on all integers from the minimum to the maximum. From the table we can see that Spearman's ρ distance has by far the largest support, of order m^3, at least for medium to large m. The footrule and Kendall's distances have fairly large support, too, but of order m^2. The others' support is just m or $m-1$.

Perhaps more important is the practical support, i.e., the number of values the distance is likely to take on. In table (1.11), we find the minimum number of support values for which the total probability is at least 99.9%.

$m \to$	10	25	50	100	Asymptotically
Spearman	147	1633	9485	54276	$0.55 \times m^{5/2}$
Footrule	20	85	243	690	$0.69 \times m^{3/2}$
Kendall	34	137	388	1098	$1.10 \times m^{3/2}$
Hamming	6	6	6	6	6
Cayley	7	9	11	12	$6.58 \times \sqrt{\log(m)}$
Ulam	6	8	9	12	$5.96 \times m^{1/6}$
Maximum	7	12	17	24	$2.63 \times \sqrt{m}$

(1.11)

Here we see that the Spearman,footrule and Kendall distances have practical support that grows reasonably fast, especially Spearman's. The Hamming distance's practical support does not grow at all, being stuck at 6. The others' grow quite slowly, the maximum distance's growing fastest of those three at a $\sqrt{m}$ rate.

1.5 Correlation-coefficient-like rescaling

Our main objective in studying distances is to aid in analyzing and modeling rank data. The original impetus for many of the distances was to find non-parametric alternatives to the usual product-moment correlation coefficient used for continuous data. See Spearman (1904) and M. G. Kendall & Gibbons (1990) for historical context.

For $1 \times N$ vectors $\boldsymbol{w}$ and $\boldsymbol{z}$, their product moment correlation can be defined by

$$r(\boldsymbol{w}, \boldsymbol{z}) = \frac{\boldsymbol{w}\boldsymbol{z}'}{\|\boldsymbol{w}\| \, \|\boldsymbol{z}\|}. \tag{1.12}$$

For continuous data vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, the sample correlation coefficient is r with $w_i = x_i - \overline{x}$ and $z_i = y_i - \overline{y}$, the centered versions of the vectors. It ranges from -1 to $+1$, indicating the degree to which the two variables are linearly related: $+1$ indicates they are perfectly positively linearly related ($y_i = a + bx_i$ for some $b > 1$), -1 means they are perfectly negatively linearly related (so $b < 0$), and zero indicates no linear relationship.

Spearman's ρ and Kendall's τ coefficients are analogs, where the ρ coefficient is the same as the regular correlation coefficient but with the raw data replaced by ranks. It then measures the degree of linear relationship in the ranks, which translates to monotone relationship of the original data. Kendall's τ uses pairs of indices, so that

$$w_{ij} = \text{Sign}(x_i - x_j) \text{ and } z_{ij} = \text{Sign}(y_i - y_j), \quad 1 \leqslant j < i \leqslant m, \tag{1.13}$$

where $\text{Sign}(a) = -1, 0$, or $+1$ as $a < 0, = 0$, or > 0. Now $\boldsymbol{w}$ and $\boldsymbol{z}$ are length $N = \binom{m}{2}$. The corresponding distances for rank vectors were named from these coefficients, since

$$\begin{aligned}
\rho(\boldsymbol{y}, \boldsymbol{x}) &= 1 - \frac{d_{\text{Spear}}(\boldsymbol{y}, \boldsymbol{x})}{E[d_{\text{Spear}}(\boldsymbol{Z}, \boldsymbol{\omega})]} = 1 - 2\frac{d_{\text{Spear}}(\boldsymbol{y}, \boldsymbol{x})}{\text{Max}\{d_{\text{Spear}}(\boldsymbol{y}^*, \boldsymbol{\omega})\}} \text{ and} \\
\tau(\boldsymbol{y}, \boldsymbol{x}) &= 1 - \frac{d_{\text{Ken}}(\boldsymbol{y}, \boldsymbol{x})}{E[d_{\text{Ken}}(\boldsymbol{Z}, \boldsymbol{\omega})]} = 1 - 2\frac{d_{\text{Ken}}(\boldsymbol{y}, \boldsymbol{x})}{\text{Max}\{d_{\text{Spear}}(\boldsymbol{y}^*, \boldsymbol{\omega})\}},
\end{aligned} \tag{1.14}$$

where $\boldsymbol{Z} \sim \text{Uniform}(\mathcal{P}_m)$ the maxima are taken over $\boldsymbol{y}^* \in \mathcal{P}_m$. Both of these distances are symmetric, so that their means are half their maxima. Note that the coefficients' means are zero.

Because of the nice interpretation of these correlation-like coefficients, it might help to rescale the other distances similarly. One problem is that none of the others in (1.4) are symmetric, so that the two rescalings as in (1.14) are not the same. Then either the minimum of the coefficient will not be -1, or the mean will not be zero. For example, using the footrule we have that from (1.10),

$$\min\left\{1 - \frac{d_{\text{Foot}}(\boldsymbol{y}, \boldsymbol{x})}{E[d_{\text{Foot}}(\boldsymbol{Z}, \boldsymbol{\omega})]}\right\} \approx -0.5 \text{ and } E\left[1 - 2\frac{d_{\text{Foot}}(\boldsymbol{y}, \boldsymbol{x})}{\text{Max}\{d_{\text{Foot}}(\boldsymbol{y}^*, \boldsymbol{\omega})\}}\right] \approx -0.33. \tag{1.15}$$

I prefer the first choice, presented by Spearman (1904), since having uniformity associated with "0" seems more important than being able to achieve a "-1" coefficient.

The calculations work out even worse for the other distances, so I don't believe these rescalings are appropriate. In fact, the distances have nice interpretations on their own, possibly after subtracting from m. That is, $m - d_{\text{Ham}}(\boldsymbol{y}, \boldsymbol{x})$ is the number of matches between the x_i's and y_i's; $m - d_{\text{Ulam}}(\boldsymbol{y}, \boldsymbol{x})$ is the length of the longest increasing subsequence; Cayley's distance is the number of interchanges needed to bring $\boldsymbol{y}$ to $\boldsymbol{x}$; and $d_{\text{Max}}(\boldsymbol{y}, \boldsymbol{x})$ is the maximum difference between x_i and y_i. One may wish to divide these alternatives by m, to scale between 0 and 1.

1.6 Correlations among the distances

We also consider the correlations of the distances under the uniform distribution. Chapter 11 calculates the exact covariances among five of the distances (all but Ulam's and the maximum distances), and simulates the other correlations. Below are the main take-aways:

- For small m, the correlations are fairly high, but as m increases, most correlations decrease.
- The triad of the Spearman, footrule, and Kendall distances have their inter-correlations start high, and generally increase, with that between Spearman and Kendall tending to 1, and those between the footrule and the other two tending to $3/\sqrt{10} \approx 0.95$.
- The correlations of the above triad with Ulam's distance decrease slowly; it is not clear whether they approach zero, though they appear to decline at a $\log(m)$ rate for m up to 10,000. The correlations of the triad with the remaining (Hamming, Cayley, and maximum) go to zero, at least seemingly for the maximum.
- The correlation of Hamming's and Cayley's distances approaches zero, but very slowly, at a rate of $1/\sqrt{\log(m)}$.
- All other correlations approach zero, at least seemingly for those involving Ulam or the maximum.

The exact covariances among five of the distances are given in (11.1). The asymptotic correlations as $m \to \infty$ of these distances are given next:

	Footrule	Kendall	Hamming	Cayley
Spearman	$\frac{3}{\sqrt{10}}(1-1/(4m^2))$	$1-1/(4m)$	$1/\sqrt{m-1}$	$1/\sqrt{m\log(m)}$
Footrule		$\frac{3}{\sqrt{10}}(1-1/(4m))$	$\sqrt{5/(2m)}$	$\sqrt{5/(2m\log(m))}$
Kendall			$1/\sqrt{m}$	$1/\sqrt{m\log(m)}$
Hamming				$1/\sqrt{\log(m)}$

(1.16)

See Section 11.2 for more calculations.

1.7 Bi-invariance

The distance d is label-invariant as defined in Section 1.1 if $d(\boldsymbol{y}, \boldsymbol{x})$ is not affected by permutation of the indices of $\boldsymbol{x}$ and $\boldsymbol{y}$, where the indices in the two vectors are permuted in concert. E.g., using the permutation $1 \to 2,\ 2 \to 4,\ 3 \to 3,\ 4 \to 1$ on the indices, we have

$$d((y_1, y_2, y_3, y_4), (x_1, x_2, x_3, x_4)) = d((y_2, y_4, y_3, y_1), (x_2, x_4, x_3, x_1)). \tag{1.17}$$

The distance is rank-invariant if permuting the actual rank values does not change the distance. Thus using the same permutation, but on the ranks, d is rank-invariant implies that

$$d((4,2,1,3), (1,2,3,4)) = d((1,4,2,3), (2,4,3,1)). \tag{1.18}$$

Formally, let $\mathcal{Q}_m$ be the set of $m \times m$ permutation matrices, and for $\boldsymbol{x} \in \mathcal{P}_m$, define the associated $\boldsymbol{Q}_{\boldsymbol{x}} \in \mathcal{Q}_m$ by

$$\boldsymbol{x} = \boldsymbol{\omega}\boldsymbol{Q}_{\boldsymbol{x}}, \text{ so that } x_i = j \text{ if } \boldsymbol{Q}_{\boldsymbol{x},ji} = 1. \tag{1.19}$$

Writing $d(\boldsymbol{Q}_{\boldsymbol{y}}, \boldsymbol{Q}_{\boldsymbol{x}}) = d(\boldsymbol{y}, \boldsymbol{x})$, we have that

$$\begin{array}{lcl} \text{Label-invariance} & \Rightarrow & d(\boldsymbol{Q}_{\boldsymbol{y}}, \boldsymbol{Q}_{\boldsymbol{x}}) = d(\boldsymbol{Q}_{\boldsymbol{y}}\boldsymbol{Q}^*, \boldsymbol{Q}_{\boldsymbol{x}}\boldsymbol{Q}^*) \text{ for all } \boldsymbol{Q}^* \in \mathcal{Q}_m; \\ \text{Rank-invariance} & \Rightarrow & d(\boldsymbol{Q}_{\boldsymbol{y}}, \boldsymbol{Q}_{\boldsymbol{x}}) = d(\boldsymbol{Q}\boldsymbol{Q}_{\boldsymbol{y}}, \boldsymbol{Q}\boldsymbol{Q}_{\boldsymbol{x}}) \text{ for all } \boldsymbol{Q} \in \mathcal{Q}_m; \\ \text{Bi-invariance} & \Rightarrow & d(\boldsymbol{Q}_{\boldsymbol{y}}, \boldsymbol{Q}_{\boldsymbol{x}}) = d(\boldsymbol{Q}\boldsymbol{Q}_{\boldsymbol{y}}\boldsymbol{Q}^*, \boldsymbol{Q}\boldsymbol{Q}_{\boldsymbol{x}}\boldsymbol{Q}^*) \text{ for all } \boldsymbol{Q}, \boldsymbol{Q}^* \in \mathcal{Q}_m. \end{array} \tag{1.20}$$

The **cycle distribution** of $\boldsymbol{y} \in \mathcal{P}_m$ is the $1 \times m$ vector $\boldsymbol{\gamma}(\boldsymbol{y})$, where

$$\gamma_k(\boldsymbol{y}) = \#\{\text{Cycles of length } k \text{ in } \boldsymbol{y}\}. \tag{1.21}$$

A distance d is bi-invariant if and only it is label-invariant, and for some function g,

$$d(\boldsymbol{y}, \boldsymbol{x}) = d(\boldsymbol{y}^*, \boldsymbol{\omega}) = g(\boldsymbol{\gamma}(\boldsymbol{y}^*)), \text{ where } \boldsymbol{Q}_{\boldsymbol{y}^*} = \boldsymbol{Q}_{\boldsymbol{y}}\boldsymbol{Q}'_{\boldsymbol{x}}. \tag{1.22}$$

Label-invariance yields the first equality. For the second, note first that (1.20) for bi-invariance with $\boldsymbol{Q}^* = \boldsymbol{Q}'_{\boldsymbol{x}}\boldsymbol{Q}'$ yields

$$d(\boldsymbol{y}, \boldsymbol{x}) = d(\boldsymbol{Q}\boldsymbol{Q}_{\boldsymbol{y}^*}\boldsymbol{Q}', \boldsymbol{I}) \text{ for all } \boldsymbol{Q} \in \mathcal{Q}_m. \tag{1.23}$$

The operation $\boldsymbol{Q}_{\boldsymbol{y}^*} \to \boldsymbol{Q}\boldsymbol{Q}_{\boldsymbol{y}^*}\boldsymbol{Q}'$ is called a **conjugation** in algebra. Then (1.22) follows by a result on conjugation of permutations. See Herstein (1964, pages 75 and 76). Intuitively, the idea is that a cycle as in (1.5) and (1.6) will still be a cycle upon replacing the c_i's with another set of distinct integers.

From the characterization (1.22), we see that the Hamming and Cayley distances are bi-invariant. Hamming's distance is m minus the number of matches, $y_k = k$, each of which is a cycle of length one, and Cayley's distance is just m minus the total number of cycles:

$$d_{\text{Hamming}}(\boldsymbol{y}, \boldsymbol{\omega}) = m - \gamma_1(\boldsymbol{y}) \text{ and } d_{\text{Cayley}}(\boldsymbol{y}, \boldsymbol{\omega}) = m - \sum_{k=1}^{m} \gamma_k(\boldsymbol{y}). \tag{1.24}$$

Chapter 2

Moments, cumulants, and Edgeworth expansions

For some distances, or random variables in general, it is easier to find the moments, and for some the cumulants, hence it is convenient to be able to calculate one type from the another. In Section 2.2, we present some general formulas for such conversions, along with Mathematica and R code.

The Edgeworth expansion is a series of modifications to the normal distribution that, under suitable conditions, provide successively better approximations to the distribution of a normalized sum of iid random variables. We use it for some of the distances. Even though they are not sums of iid variables, the expansions do provide very good approximations. In Section 2.3 we develop the expansion for the iid case when the variables have a density. Our applications are require an extra adjustment because they are concentrated on a lattice, i.e, the integers. See Section 2.3.1. Section 2.4 provides some Mathematica code that we use.

2.1 Moment and cumulant generating functions

For random variable X, we consider the following types of moments, where n is a positive integer:

$$\begin{aligned} \text{Raw:} \quad & \mu'_n = \mathsf{E}[X^n], \\ \text{Central:} \quad & \mathsf{E}[(X-\mu)^n], \\ \text{Regular:} \quad & \mu_n = \begin{cases} \mu'_1 \equiv \mu & \text{if } n = 1 \\ \mathsf{E}[(X-\mu)^n] & \text{if } n > 1 \end{cases}, \text{ and} \\ \text{Factorial:} \quad & \gamma_n = \mathsf{E}[(X)_n], \end{aligned} \tag{2.1}$$

where $\mu = \mathsf{E}[X] = \mu'_1$ and $(x)_n = x(x-1)\cdots(x-n+1)$, which we see again in (4.18). (Set $\mu'_0 = \mu_0 = \gamma_0 = 1$.) We usually call the "regular moments" just "moments," since they are what the term typically evokes. The first regular moment is a raw moment, the mean, the second is a central moment, the variance, and the rest are also central moments (skewness, kurtosis, etc.).

The raw moments can be found using the moment generating function $M_X(t)$ or the characteristic function $C_X(t)$ given by

$$M_X(t) = \mathsf{E}[e^{tX}] \text{ and } C_X(t) = C[e^{itX}], \tag{2.2}$$

where i is the imaginary unit. The moment generating function is said to exist if it is finite for t in a neighborhood of zero, in which case we have the expansion

$$M_X(t) = \sum_{k=0}^{\infty} \mu'_k \frac{t^k}{k!}. \tag{2.3}$$

The characteristic function always exists, and if the n^{th} moment is finite, then we have

$$C_X(t) = \sum_{k=0}^{n} \mu'_k \frac{(it)^k}{k!} + o(|t|^n) \tag{2.4}$$

Thus under the appropriate conditions ($M_X(t)$ exists, or μ'_k is finite), we can obtain the k^{th} moment via differentiation:

$$\mu'_k = M_X^{(k)}(0) = -i^k C_X^{(k)}(0). \tag{2.5}$$

We can generate the central moments by using either $M_{(X-\mu)}(t)$ or $C_{(X-\mu)}(t)$.

Cumulants are defied via the expansions of the logs of these two functions. That is, if the moment generating function exists, or μ'_n is finite, then the cumulant generating function is either

$$K_X(t) = \log(M_X(t)) = \sum_{k=1}^{\infty} \kappa_k \frac{t^k}{k!} \quad \text{or}$$

$$H_X(t) = \log(C_X(t)) = \sum_{k=1}^{n} \kappa_k \frac{(it)^k}{k!} + o(|t|^n). \tag{2.6}$$

In either case, the k^{th} cumulant is the κ_k in the expansion, and is given by $K_X^{(k)}(0)$ and $-i^k H^{(k)}(0)$. In particular, κ_1 is the mean and κ_2 is the variance. The **normalized cumulants** are the cumulants of the normalized variable $(X-\mu)/\sqrt{\mu_2}$, if the first two moments are finite. Thus

$$n^{th} \text{ normalized cumulant} = \begin{cases} 0 & \text{if } n = 1 \\ 1 & \text{if } n = 2 \\ \kappa_n/\kappa_2^{n/2} & \text{if } n \geqslant 3, \ \& \ \kappa_n \text{ is finite} \end{cases}. \tag{2.7}$$

Similarly, the factorial generating function has a real and complex version, but we will just deal with the former. If the moment generating function exists, so does the factorial generating function, and equals

$$Fa_X(t) = E[(1+t)^X] = \sum_{k=0}^{\infty} \gamma_k \frac{t^k}{k!}. \tag{2.8}$$

2.2 Moment conversions

Any one of the sets of the first n (regular) moments, raw moments, cumulants, or factorial moments can be calculated from any of the other sets. We could also include the set of central moments if we add μ to that set, which is then the set of regular moments.

The easiest conversion is to go between raw moments and regular moments. We start by relating the moments of W and $W + c$ for constant c. Using the binomial expansion, we have

$$\begin{aligned} \mathrm{E}[(W+c)^n] &= \mathrm{E}\left[\sum_{k=0}^{n} \binom{n}{k} W^k c^{n-k}\right] \\ &= \sum_{k=0}^{n} \binom{n}{k} \mathrm{E}[W^k] c^{n-k}. \end{aligned} \tag{2.9}$$

If we know $\mu'_1, \ldots, \mu'_n$ (and they are finite), then we automatically have $\mu_1 = \mu'_1$, and for $n > 1$, by setting $W = X$ and $c = -\mu$ in (2.9), we have

$$\begin{aligned} \mu_n = \mathrm{E}[(X-\mu)^n] &= \sum_{k=0}^{n} \binom{n}{k} (-1)^{n-k} \mu'_k \mu^{n-k} \\ &= (-1)^n (n-1)\mu^n + \sum_{k=2}^{n} \binom{n}{k} (-1)^{n-k} \mu'_k \mu^{n-k}. \end{aligned} \tag{2.10}$$

Reversing, we set $W = X - \mu$ and $c = \mu$, so that for $n > 1$,

$$\mu'_n = \sum_{k=0}^{n} \binom{n}{k} \mu_k \mu^{n-k} = (n+1)\mu^n + \sum_{k=2}^{n} \binom{n}{k} \mu_k \mu^{n-k}. \tag{2.11}$$

To go between raw moments and cumulants or raw moments and factorial moments, we appeal to a formula of Faà di Bruno (di Bruno, 1855) that deals with Mclaurin expansions of composite functions. Finding one generating function from another involves expanding a composite function $g(h(t))$ in terms of the individual expansions of g and h. Our version is from Appendix A of Blinnikov & Moessner (1998), proven in Section 2.2.1.

Lemma 2.1. *Suppose* $g(u)$ *can be written as a Maclaurin series. Then*

$$g\left(\sum_{l=1}^{L} \alpha_l \epsilon^l\right) = g(0) + \sum_{n=1}^{\infty} \lambda_n \epsilon^n, \tag{2.12}$$

where

$$\begin{aligned} \lambda_n &= \sum_{\boldsymbol{k} \in \mathcal{A}_n} g^{(\boldsymbol{k}^*)}(0) \prod_{\substack{l=1 \\ k_l > 0}}^{L} \frac{1}{k_l!} \alpha_l^{k_l} \\ &\equiv \sum_{\boldsymbol{k} \in \mathcal{A}_n} g^{(\boldsymbol{k}^*)}(0) \prod_{l=1}^{L} \frac{1}{k_l!} \alpha_l^{k_l}, \quad (\textit{if we take } 0^0 = 1), \end{aligned} \tag{2.13}$$

$$\mathcal{A}_n = \{\boldsymbol{k} = (k_1, \ldots, k_L) \mid \sum_{l=1}^{L} l k_l = n, \;\; k_l \textit{ are nonnegative integers}\}, \tag{2.14}$$

and

$$\boldsymbol{k}^* = \sum_{l=1}^{L} k_l. \tag{2.15}$$

In the lemma, we can take $L = \infty$. Each vector $\boldsymbol{k}$ has only a finite number of positive entries, so the product in (2.13) has effectively a finite number of components.

To simplify a little, we assume the moment generating function exists, so that we can use the real versions of the generators. Even if it doesn't exist, the formulas will be valid for moments which exist. Start with the raw moments, and consider finding the cumulants. We write the cumulant generating function as in (2.12) with $g(x) = \log(1+x)$ and the argument of g being $M_X(t) - 1$ with $\epsilon = t$, and $L = \infty$. Then (2.12) yields

$$\begin{aligned} K_X(t) = \log\left(1 + \sum_{l=1}^{\infty} \mu'_k \frac{t^k}{k!}\right) &= g\left(\sum_{l=1}^{\infty} \mu'_k \frac{t^k}{k!}\right) \\ &= 0 + \sum_{n=1}^{\infty} \lambda_n t^n. \end{aligned} \tag{2.16}$$

Thus by (2.6), the n^{th} cumulant is $n!\lambda_n$. To use (2.13), we note that $\alpha_l = \mu'_l/l!$, and $g^{(i)}(0) = (-1)^{i-1}(i-1)!$. Thus

$$\kappa_n = n! \sum_{\boldsymbol{k} \in \mathcal{A}_n} (-1)^{\boldsymbol{k}^*-1} (\boldsymbol{k}^* - 1)! \prod_{l=1}^{n} \frac{1}{k_l!} \left(\frac{\mu'_l}{l!}\right)^{k_l}. \tag{2.17}$$

(The product goes only to n, rather than $L = \infty$, since $k_l = 0$ for $l > n$ when $\boldsymbol{k} \in \mathcal{A}_n$.) This formula is given in equation (30) of Blinnikov & Moessner (1998).

We can reverse this conversion using $M_X(t) = \exp(K_X(t))$. Now $g(x) = x$, so the derivatives at $x = 0$ are all equal to one, and $\alpha_l = \kappa_l/l!$. Thus the lemma yields $\mu'_n = n!\lambda_n$, or

$$\mu'_n = n! \sum_{\boldsymbol{k} \in \mathcal{A}_n} \prod_{l=1}^{n} \frac{1}{k_l!} \left(\frac{\kappa_l}{l!}\right)^{k_l}. \tag{2.18}$$

Turn to finding the raw moments from the factorial moments. The moment generating function can be written as a function of the factorial generating function as

$$M_X(t) = E[e^{tX}] = E[(1 + e^t - 1)^X] = Fa_X(e^t - 1) = Fa_X\left(\sum_{l=1}^{\infty} \frac{t^l}{l!}\right). \tag{2.19}$$

In this case, $g(x) = Fa_X(x)$, so $g^{(j)}(0) = \gamma_j$, and $\alpha_l = 1/l!$, hence

$$\mu'_n = n! \sum_{\boldsymbol{k} \in \mathcal{A}_n} \gamma_{\boldsymbol{k}^*} \prod_{l=1}^{n} \frac{1}{k_l!} \left(\frac{1}{l!}\right)^{k_l}. \tag{2.20}$$

Perhaps more familiar to combinatorists, an equivalent formula is

$$\mu_n = \sum_{k=1}^{n} S(n,k)\gamma_k, \tag{2.21}$$

where $S(n,k)$'s are **Stirling numbers of the second kind**. See Weisstein (2018a, equation 11). The other way, we have

$$Fa_X(t) = E[(1+t)^X] = E[e^{\log(1+t)X}] = M_X(\log(1+t)) = M_X\left(\sum_{l=1}^{\infty}(-1)^{l-1}\frac{t^l}{l}\right). \tag{2.22}$$

Now $g^{(i)}(0) = \mu'_i$ and $\alpha_l = (-1)^{l-1}/l$, so that

$$\gamma_n = n! \sum_{\boldsymbol{k}\in\mathcal{A}_n} \mu'_{\boldsymbol{k}^*} \prod_{l=1}^{n} \frac{1}{k_l!}\left(\frac{(-1)^{l-1}}{l}\right)^{k_l}. \tag{2.23}$$

Converting between the regular moments and factorial moments does not appear to have a simple direct answer, so that it may be best to go through the raw moments (e.g., factorial moments $\to$ raw moments $\to$ regular moments). Converting between regular moments and cumulants is almost the same as with raw moments. The n^{th} cumulants for $n \geqslant 2$ are shift-invariant, i.e., the n^{th} cumulant of X is the same as that of $X + c$ for any c. Thus to find the cumulants from the regular moments, we set $\kappa_1 = \mu_1$, and for $n \geqslant 2$, use the set of central moments (i.e., the first is 0, the rest are μ_n) in place of the μ'_n in (2.17). To find the regular moments from the cumulants, set $\mu_1 = \kappa_1$, and for $n \geqslant 2$ use (2.18) but setting $\kappa_1 = 0$.

2.2.1 Proof of Lemma 2.1

Start by writing out the Maclaurin series for g, then use the multinomial theorem to expand out the powers of the sum:

$$\begin{aligned} g\left(\sum_{l=1}^{L}\alpha_l\epsilon^l\right) &= g(0) + \sum_{r=1}^{\infty}\frac{g^{(r)}(0)}{r!}\left(\sum_{l=1}^{L}\alpha_l\epsilon^l\right)^r \\ &= g(0) + \sum_{r=1}^{\infty}\frac{g^{(r)}(0)}{r!}\sum_{\boldsymbol{k}\in\mathcal{B}_r}\binom{r}{k_1,\ldots,k_L}\prod_{l=1}^{L}(\alpha_l\epsilon^l)^{k_l}, \end{aligned} \tag{2.24}$$

where

$$\mathcal{B}_r = \{\boldsymbol{k} = (k_1,\ldots,k_L)\,|\,\boldsymbol{k}^* = r,\ \ k_j \text{ are nonnegative integers}\}. \tag{2.25}$$

(Recall from (2.15) that $\boldsymbol{k}^* = \sum_{j=1}^{L} k_j$.) As in (2.13), in the final product, $0^0 = 1$. Rearranging a bit yields

$$g\left(\sum_{l=1}^{L}\alpha_l\epsilon^l\right) = g(0) + \sum_{r=1}^{\infty}\sum_{\boldsymbol{k}\in\mathcal{B}_r} g^{(\boldsymbol{k}^*)}(0)\,\epsilon^{\sum_{l=1}^{L} lk_l}\prod_{l=1}^{L}\frac{1}{k_l!}\,\alpha_l^{k_l}. \tag{2.26}$$

Note that the double summation is over all $1 \times L$ vectors $\boldsymbol{k}$ whose elements are nonnegative integers, and at least one element is positive. We regroup the $\boldsymbol{k}$ depending on their value $n = \sum_{l=1}^{L} lk_l$, the power of the ϵ. Then using the $\mathcal{A}_n$ from (2.14), we have

$$g\left(\sum_{l=1}^{L}\alpha_l\epsilon^l\right) = g(0) + \sum_{n=1}^{\infty}\sum_{\boldsymbol{k}\in\mathcal{A}_n} g^{(\boldsymbol{k}^*)}(0)\,\epsilon^s\prod_{l=1}^{L}\frac{1}{k_l!}\,\alpha_l^{k_l}, \tag{2.27}$$

which verifies (2.12) and (2.13).

2.3 The Edgeworth expansion

Suppose $X_1, X_2, \ldots$ are iid with distribution function $F_X(x)$, and have mean 0 and variance 1. Let Z be the normalized sum

$$Z = \frac{\sum_{i=1}^{N} X_i}{\sqrt{N}}. \tag{2.28}$$

The central limit theorem shows that $Z \to N(0,1)$ as $N \to \infty$. The Edgeworth expansion of the distribution function $F_Z(z)$ of Z to L terms has the form

$$\widehat{F}_Z(z) = \Phi(z) + \frac{1}{\sqrt{N}}\Psi_1(z) + \frac{1}{N}\Psi_2(z) + \cdots + \frac{1}{N^{L/2}}\Psi_L(z), \tag{2.29}$$

where Φ is the standard normal distribution function, and the Ψ_l are functions depending on F_X. If X has a density $f_X(x)$ with respect to Lebesgue measure, and the first $L+2$ moments of X are finite, then

$$\widehat{F}_Z(z) - F_Z(z) = o\left(\frac{1}{N^{L/2}}\right). \tag{2.30}$$

See Cramér (1946) and Esseen (1945) for details and proofs for this result and those below. The error in the expansion is not correct when F_X is a lattice distribution, which is the case for the distances we consider. Section 2.3.1 presents an adjustment for such cases.

Our presentation follows that in Blinnikov & Moessner (1998). The Edgeworth expansion relies on the complex version of the cumulant generating function, $H_X(t)$ from (2.6). Assume that the $(L+2)^{nd}$ moment of X_i is finite, and apply (2.6) to obtain

$$K_X(t) = \sum_{l=1}^{L+2} \kappa_l \frac{(it)^l}{l!} + o(|t|^{L+2}). \tag{2.31}$$

Since the X_i are independent, the cumulant generating function of $\sum X_i$ is $NK_X(t)$, hence that for Z in (2.28) is

$$K_Z(t) = NK_X\left(\frac{t}{\sqrt{N}}\right) = \sum_{l=1}^{L+2} \frac{\kappa_l}{N^{l/2-1}} \frac{(it)^l}{l!} + o(|t|^{L+2}). \tag{2.32}$$

Thus the l^{th} cumulant of Z is $\kappa_l / N^{l/2-1}$.

The X_i's, hence Z, have first two cumulants being 0 and 1. We remove the little o term to obtain our initial estimate of K_Z:

$$\widehat{K}^*_Z(t) = -\frac{t^2}{2} + \sum_{l=3}^{L+2} \frac{\kappa_l}{N^{l/2-1}} \frac{(it)^l}{l!} \tag{2.33}$$

$$= -\frac{t^2}{2} + \sum_{l=1}^{L} \epsilon^l \frac{\kappa_{l+2}(it)^{l+2}}{(l+2)!}, \quad \text{where} \quad \epsilon = \frac{1}{\sqrt{N}}. \tag{2.34}$$

The corresponding initial estimate of the characteristic function of Z is

$$\widehat{H}^*_Z(t) = e^{\widehat{K}^*_Z(t)} = \overline{H}(t)\, e^{\sum_{l=1}^{L} \epsilon^l \kappa_{l+2}(it)^{l+2}/(l+2)!}, \tag{2.35}$$

where $\overline{H}(t) = exp(-t^2/2)$ is the characteristic function of the standard normal distribution. We next find the expansion of the exponential term in (2.35) about ϵ.

Apply Lemma 2.1 to (2.35) with $g(u) = e^u$, so that $g^{(i)}(0) = 1$ for all i, and $\alpha_j = \kappa_{j+2}/(j+2)!$. Then our final estimate of H_Z is found from (2.12) and (2.13) by truncating at $n = L$:

$$\widehat{H}_Z(t) = \overline{H}(t)\left(1 + \sum_{l=1}^{L} \epsilon^l \sum_{\boldsymbol{k}\in\mathcal{A}_l} (it)^{l+2\boldsymbol{k}^*} \prod_{j=1}^{L} \frac{1}{k_j!}\left(\frac{\kappa_{j+2}}{(j+2)!}\right)^{k_j}\right). \tag{2.36}$$

See (2.14) and (2.15) for $\mathcal{A}_l$ and $\boldsymbol{k}^*$.

The characteristic function can be inverted to find the density, if the density exists. Thus if X has characteristic function $H_X(t)$, then

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx} H_X(t)\,dt = f(x), \tag{2.37}$$

where f is the density. Also, by differentiating, we obtain

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} (-itx)^l e^{-itx} H_X(t)\,dt = f^{(l)}(x) \tag{2.38}$$

if the l^{th} derivative of f exists. The estimate of the density of Z is then the inverse of the estimated characteristic function:

$$\widehat{f}_Z(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itz}\widehat{H}_Z(t)\,dt. \tag{2.39}$$

This function may not be a valid density, e.g., it may not integrate to 1, or always be non-negative, though still will be useful. We now distribute the integral over the summation on right-hand side of (2.36). Since $\overline{H}$ is the standard normal characteristic function, we can use (2.37) and (2.38) to show that

$$\widehat{f}_Z(z) = \phi(z) + \sum_{l=1}^{L} \epsilon^l \sum_{\boldsymbol{k}\in\mathcal{A}_l} (-1)^{l+2\boldsymbol{k}^*} \phi^{(l+2\boldsymbol{k}^*)}(z) \prod_{j=1}^{L} \frac{1}{k_j!}\left(\frac{\kappa_{j+2}}{(j+2)!}\right)^{k_j}, \tag{2.40}$$

where ϕ is the standard normal density, $\phi(z) = e^{-z^2/2}/\sqrt{2\pi}$. The corresponding estimate of the distribution function is found by integrating:

$$\widehat{F}_Z(z) = \Phi(z) + \sum_{l=1}^{L} \epsilon^l \sum_{\boldsymbol{k}\in\mathcal{A}_l} (-1)^{l+2\boldsymbol{k}^*} \phi^{(l+2\boldsymbol{k}^*-1)}(z) \prod_{j=1}^{L} \frac{1}{k_j!}\left(\frac{\kappa_{j+2}}{(j+2)!}\right)^{k_j}. \tag{2.41}$$

Since $\epsilon = 1/\sqrt{N}$, we have the form (2.29).

The derivatives of the standard normal density are related to **Hermite polynomials** via

$$\phi^{(k)}(z) = (-1)^k \phi(z) He_k(z), \tag{2.42}$$

where He_k is the k^{th} Hermite polynomial. (There are two sets of Hermite polynomials. The ones denoted H_k are related to the He_k's by $H_k(z) = 2^{k/2}He_k(\sqrt{2}z)$.) Differentiation shows that

$$He_0(z) = 1, \;\; He_1(z) = z, \;\; He_2(z) = z^2 - 1, \;\; He_3(z) = z^3 - 3z. \tag{2.43}$$

In fact,

$$He_{k+1}(z) = zHe_k(z) - He_k'(z). \tag{2.44}$$

Now we can write (2.40) and (2.41), respectively, as

$$\widehat{f}_Z(z) = \phi(z)\left(1 + \sum_{l=1}^{L} \epsilon^l \sum_{\boldsymbol{k}\in\mathcal{A}_l} He_{l+2\boldsymbol{k}^*}(z) \prod_{j=1}^{L} \frac{1}{k_j!}\left(\frac{\kappa_{j+2}}{(j+2)!}\right)^{k_j}\right), \text{ and} \tag{2.45}$$

$$\widehat{F}_Z(z) = \Phi(z) - \phi(z)\left(\sum_{l=1}^{L} \epsilon^l \sum_{\boldsymbol{k}\in\mathcal{A}_l} He_{l+2\boldsymbol{k}^*-1}(z) \prod_{j=1}^{L} \frac{1}{k_j!}\left(\frac{\kappa_{j+2}}{(j+2)!}\right)^{k_j}\right). \tag{2.46}$$

To illustrate, take $L = 3$. For $l = 1$, $\mathcal{A}_l$ contains just one vector, $\boldsymbol{k} = \{(1,0,0)\}$. Thus $\boldsymbol{k}^* = 1$. The coefficient of ϵ is then $He_3(z)\kappa_3/3!$. For $l = 2$, $\mathcal{A}_s = \{(2,0,0),(0,1,0)\}$, so there are two summands multiplying ϵ^2:

$$He_6(z)\,\frac{1}{2!}\left(\frac{\kappa_3}{3!}\right)^2 + He_4(z)\,\frac{\kappa_4}{4!} = He_6(z)\,\frac{\kappa_3^2}{72} + He_4(z)\,\frac{\kappa_4}{24}. \tag{2.47}$$

Finally, $l = 3$ yields $\mathcal{A}_s = \{(3,0,0),(1,1,0),(0,0,1)\}$, so the coefficient of ϵ^3 is

$$He_9(z)\,\frac{1}{3!}\left(\frac{\kappa_3}{3!}\right)^3 + He_7(z)\,\frac{\kappa_3}{3!}\frac{\kappa_4}{4!} + He_5(z)\,\frac{\kappa_5}{5!} = He_9(z)\,\frac{\kappa_3^3}{1296} + He_7(z)\,\frac{\kappa_3\kappa_4}{144} + He_5(z)\,\frac{\kappa_5}{120}. \tag{2.48}$$

2.3.1 The lattice case

If the X_i are concentrated on a lattice, the error term (2.30) is not correct. Esseen (1945) presents an adjustment, consisting of extra terms in the summation. Suppose the X_i are concentrated on the points $a + di, i \in \{\text{Integers}\}$ for some a and d. Kolassa & McCullagh (1990) prove that this adjustment is asymptotically equivalent to using the Sheppard adjustment to the cumulants in the original expansion. In place of the original cumulant κ_n, they use $\kappa_n^* = \kappa_n - d^l\kappa_n^U/N$, where κ_n^U is the n^{th} cumulant of the Uniform$(-\frac{1}{2},\frac{1}{2})$. Note that the odd cumulants of this uniform are zero. For $n \geqslant 2$, it is equivalent to find the cumulants of the Uniform$(0,1)$, for whom the moments are $\mu_j = 1/(j+1)$. Thus by (2.17),

$$\kappa_n^U = n! \sum_{\boldsymbol{k}\in\mathcal{A}_n} (-1)^{\boldsymbol{k}^*-1}(\boldsymbol{k}^*-1)! \prod_{j=1}^{L} \frac{1}{k_j!}\left(\frac{1}{(j+1)!}\right)^{k_j}. \tag{2.49}$$

These cumulants are also given by $\kappa_n^U = B_n/n$, where B_n is the n^{th} **Bernoulli number**. See Wichura (2001, cumulants chapter). The even cumulants up through $n = 20$ are given in (2.50).

n	2	4	6	8	10	12	14	16	18	20
κ_n^U	$\frac{1}{12}$	$-\frac{1}{120}$	$\frac{1}{252}$	$-\frac{1}{240}$	$\frac{1}{132}$	$-\frac{691}{32760}$	$\frac{1}{12}$	$-\frac{3617}{8160}$	$\frac{43867}{14364}$	$-\frac{174611}{6600}$

(2.50)

For the Spearman and footrule distances, the adjustments do not change the approximations appreciably, since for even moderate m, their cumulants are much larger than those of the uniform. For these distances, the gap is $d = 2$, hence for $N = 1$, the Sheppard adjustment is $2^l \kappa_l^U$. Table (2.49) displays the cumulants for $l = 4, 6$, and 8, $N = 1$, and $m = 10$, for Spearman and the footrule, as well as the corresponding Sheppard corrections. The cumulants increase as m and l increase, while the adjustments are independent of m. Thus the corrections are basically negligible.

Cumulants	$l = 4$	$l = 6$	$l = 8$
Spearman	-4.25×10^{6}	3.15×10^{10}	-5.49×10^{14}
Footrule	-683	5.05×10^{4}	-8.21×10^{6}
Sheppard correction	-0.133	0.254	-1.07

(2.51)

2.4 Mathematica code

2.4.1 Moment/cumulant conversions

Here we present some Mathematica functions for obtaining some moment conversions. Mathematica has a built in function, MomentConvert, that will symbolically convert one type of moment to another, e.g., central moments to cumulants, but I haven't figured out how to use it for what we need. The inputs for our conversion functions are

- n: the degree of the moment/cumulant;
- mom or cum: a function mom[n,m] or cum[n,m] that calculates the n^{th} moment or cumulant to be converted from;
- m: the number of objects ranked.

The functions are not specifically tied to ranking distances, so the m can represent any parameter(s), not just the number of objects.

The first helper function is moment2momentc, which implements the function in (2.9). The function mom finds the raw moments of W, and c is the constant c. Then the function raw2moment converts raw moments to regular moments, as in (2.10), and moment2raw does the reverse, as in (2.10).

```
moment2momentc[n_,mom_,c_,m_] := If[NumericQ[c]&&c==0,mom[n,m],
                                   Factor[Sum[Binomial[n,k]*mom[k,m]*c^(n-k),{k,0,n}]]];
raw2moment[n_,mom_,m_] := If[n==1,mom[1,m],moment2momentc[n,mom,-mom[1,m],m]];
moment2raw[n_,mom_,m_] := If[n==1,mom[1,m],moment2momentc[n,mom,mom[1,m],m]];
```

For the other conversions, we use the Faà di Bruno formula in Lemma 2.1. The key helper functions are fs[n], which finds the set $\mathcal{A}_n$ in (2.14) for given n, and faaLambda[g,alpha,n], which calculates the λ_n in (2.13). Here, g[i] is the $g^{(i)}(0)$ and alpha[j] is the α_j. The km in the function is the vector $\boldsymbol{k}$.

```
fs[n_]:=If[n==1,{{1}},FrobeniusSolve[Range[n],n]]
faaLambda[g_,alpha_,n_] := (
    If[n==0,Return[g[0]]];
    Factor[Sum[g[Total[km]]*Apply[Times,
                Table[If[km[[j]]==0,1,(alpha[j])^km[[j]]/km[[j]]!], {j, 1, Length[km]}]],{km,fs[n]}]])
```

Each conversion function is of the form typea2typeb, which takes as input the typea function of moments or cumulants, and outputs the n^{th} typeb value. In all cases, n should be a positive integer (not just a symbol), while m can be a number or a symbol. By "moment" we mean regular moment. Many of the conversions use faaLambda, the main task being to correctly identify g and alpha.

```
raw2cumulant[n_,mom_,m_] := Module[{g,alpha},
    If[n==0,Return[0]];
    g[i_] := (i−1)!∗(−1)^(i−1);
    alpha[j_] := mom[j,m]/j!;
    faaLambda[g,alpha,n]∗n!
]
cumulant2raw[n_,cum_,m_] := Module[{g,alpha},
    g[i_] := 1;
    alpha[j_] := cum[j,m]/j!;
    faaLambda[g,alpha,n]∗n!]
moment2cumulant[n_,mom_,m_] := Module[{},
    If[n==1,Return[mom[1,m]]];
    raw2cumulant[n,Function[{n0,m0},If[n0==1,0,mom[n0,m0]]],m]]
cumulant2moment[n_,cum_,m_] := Module[{},
    If[n==1,Return[cum[1,m]]];
    cumulant2raw[n,Function[{n0,m0},If[n0==1,0,cum[n0,m0]]],m]]
factorial2raw[n_,fmom_,m_]:= Module[{g,alpha},
    g[i_] := fmom[i,m];
    alpha[j_] := 1/j!;
    faaLambda[g,alpha,n]∗n!]
raw2factorial[n_,mom_,m_] := Module[{g,alpha},
    g[i_] := mom[i,m];
    alpha[j_] := (−1)^(j−1)/j;
    faaLambda[g,alpha,n]∗n!]
factorial2moment[n_,fmom_,m_] := Module[{raw},
    raw[n0_,m0_] := factorial2raw[n0,fmom,m0];
    raw2moment[n,raw,m]]
moment2factorial[n_,mom_,m_] := Module[{raw},
    raw[n0_,m0_] := moment2raw[n0,mom,m0];
    raw2factorial[n,raw,m]]
factorial2cumulant[n_,fmom_,m_] := Module[{raw},
    raw[n0_,m0_] := factorial2raw[n0,fmom,m0];
    raw2cumulant[n,raw,m]]
cumulant2factorial[n_,mom_,m_] := Module[{raw},
    raw[n0_,m0_] := cumulant2raw[n0,mom,m0];
    raw2factorial[n,raw,m]]
```


2.4.2 Edgeworth expansions

Again, we assume the X_i are iid with mean zero and variance one. The key input is cum, which is the function cum[l,m] that yields the l^{th} cumulant, where m is the number of objects ranked (although in non-ranking situations it could be any relevant parameter needed to find the cumulant). The main functions are edgeworthf and edgeworthF, which calculate the correction expansions for the density and distribution function, respectively. The inputs are

- L, the number of desired terms in the expansion;

- cum, m, the cumulant function and parameter m;
- z, the variable representing the normalized sum in (2.28), which is the main argument in the function;
- N, the sample size.

We again use the function faaLambda, but here the g is the Hermite polynomial in (2.45), which is a function of s and z as well as $\boldsymbol{k}^*$:

$$g^{(\boldsymbol{k}^*)}(0) = He_{s+2\boldsymbol{k}^*}(z). \tag{2.52}$$

The helper function edgeTerm implements this modification, calculating the summation over $\boldsymbol{k}^*$ for the density as in (2.45). The argument offset=0 in this case. For the distribution function in (2.46), the offset=1, since the index of the Hermite polynomial is decreased by one. Mathematica has the Hermite function HermiteH, which is the scaling favored by physicists. We need the function He_l in (2.42), which according to MathWorld (Weisstein, 2018b), "is sometimes (but rarely)" defined as we do. Our function he[r,x] performs the required rescaling.

The function edgeworthf sums the edgeTerm's, multiplying by the $\epsilon^s = (1/N)^{s/2}$, and then adding 1. The result is the correction term to the normal density in (2.45). The function edgeworthF is analgous, but finds the term multiplying $\phi(z)$ for the distribution function in (2.46).

```
edgeTerm[s_,offset_,cum_,m_,z_] := Module[{g,alpha},
    g[i_] := he[s+2*i-offset,z];
    alpha[j_] := cum[j+2,m]/(j+2)!;
    faaLambda[g,alpha,s]
]

edgeworthf[L_,cum_,m_,z_,N_] := 1+Total[Table[edgeTerm[s,0,cum,m,z]/N^(s/2),{s,1,L}]];
edgeworthF[L_,cum_,m_,z_,N_] := 1+Total[Table[edgeTerm[s,1,cum,m,z]/N^(s/2),{s,1,L}]];
```

2.5 R code

Parallel to the Mathematica code in the previous section, here we present some R code for the Edgeworth expansions. The main functions require a set of cumulants as input. We also provide a function to calculate cumulants from moments. For our purposes, we use Mathematica to obtain formulas for the cumulants (at least for Spearman's ρ and footrule), then copy and paste those functions into R to use in the Edgeworth function.

As above, the variable z represents the normalized sum (2.28) of the X_i's (which again are iid with mean zero and variance one). The key functions are edgef(z,L,ncum,n) and edgeF(z,L,ncum,n), which calculate the L-term Edgeworth expansion correction terms for the density and distribution function, respectively, at the values z (a vector). Here, ncum is a vector of length at least $L+2$, where ncum[l] is the l^{th} cumulant of X_i. That is, edgef is the multiplier of $\phi(z)$ in (2.45), and edgeF is the multiplier of $\Phi(z)$ in (2.46).

The function hermite(x,k) takes a vector x and finds the Hermite polynomials of degree $0,\ldots,k$ evaluated at the x. The output is a list with the i^{th} component being the vector $He_i(\boldsymbol{x})$. The function findA(L) finds the sets $\mathcal{A}_s$ from (2.14) for $s = 1,\ldots,L$. The output is a list with the s^{th} component being a matrix with each row a $1 \times s$ vector $\boldsymbol{k}$ with $\sum_{i=1}^{s} ik_i = s$. Finally,

pd(ku,mom,offset) finds the product as in (2.13). Here, ku is a $1 \times L$ $\boldsymbol{k}$, mom is a vector of moments or cumulants, and offset shifts the index of mom:

$$\text{pd(ku,mom,offset)} = \prod_{j=1}^{L} \frac{1}{\text{ku}_j!} \left(\frac{\text{mom}_{j+\text{offset}}}{j!} \right)^{\text{ku}_j} \tag{2.53}$$

For (2.45), mom is the set of cumulants, and offset=2.

At the end of the listing we have the function moments_to_cumulants(mom), which takes a vector of the first L moments, and finds the first L cumulants.

```
hermite <- function(x,k) {
    hh <- vector('list',k+1)
    names(hh) <- as.character(0:k)
    hh[[1]] <- x-x+1
    if(k>0) hh[[2]] <- x
    if(k<2) return(hh)
    for(n in 1:(k-1)) hh[[n+2]]<-x*hh[[n+1]]-n*hh[[n]]
    hh
}

findA <- function(L) {
    A <- vector("list",L)
    A[[1]] <- matrix(1,1,1)
    if(L==1) return(A)
    K <- rbind(c(rep(0,L-1),1),0)
    r <- c(0,L)
    for(l in (L-1):1) {
        for(i in 2:nrow(K)) {
            mx <- floor(r[i]/l)
            if(mx==0) {next}
            for(k in 1:mx) {
                newk <- K[i,]
                newk[l] <- k
                K <- rbind(K,newk)
                r <- c(r,r[i]-k*l)
            }
        }
    }

    s <- K%*%(1:L)
    for(l in 2:L) {
        A[[l]] <- K[s==l,1:l]
        rownames(A[[l]]) <- as.character(1:nrow(A[[l]]))
    }
    A
}

pd <- function(ku,mom,offset=0) {
    ii <- (1:length(ku))[ku>0]
    ku <- ku[ku>0]
    prod((mom[ii+offset]/factorial(ii+offset))^ku/factorial(ku))
}
```

```
edgef <- function(z,L,ncum,n=1) {
    if(L==0) return(1)
    A <- findA(L)
    hh <- hermite(z,3*L)
    ef <- 0
    for(s in 1:L) {
        ef0 <- 0
        for(i in 1:nrow(A[[s]])) {
            ku <- A[[s]][i,]
            ks <- sum(ku)
            ef0 <- ef0 + hh[[s+2*ks+1]]*pd(ku,ncum,2)
        }
        ef <- ef + ef0/n^(s/2)
    }
    1+ef
}

edgeF <- function(x,L,ncum,n=1) {
    if(L==0) return(0)
    A <- findA(L)
    hh <- hermite(x,3*L-1)
    ef <- 0
    for(s in 1:L) {
        ef0 <- 0
        for(i in 1:nrow(A[[s]])) {
            ku <- A[[s]][i,]
            ks <- sum(ku)
            ef0 <- ef0 + hh[[s+2*ks]]*pd(ku,ncum,2)
        }
        ef <- ef + ef0/n^(s/2)
    }
    ef
}

moments_to_cumulants <- function(mom) {
    L <- length(mom)
    ncum <- mom[1]
    A <- findA(L)
    for(s in 2:L) {
        krt <- 0
        for(i in 1:nrow(A[[s]])) {
            ku <- A[[s]][i,]
            ks <- sum(ku)
            krt <- krt + (-1)^(ks-1)*factorial(ks-1)*pd(ku,mom)
        }
        ncum <- c(ncum,krt*factorial(s))
    }
    ncum
}
```

Chapter 3

Hoeffding distances

A number of popular distances, including the two Spearman distances and Hamming distance, can be written as

$$d(\boldsymbol{y}, \boldsymbol{x}) = \sum_{i=1}^{m} \delta(y_i, x_i) \tag{3.1}$$

for some function δ with $\delta(i,j) \geqslant 0, \delta(i,i) = 0$ for $i, j \in \{1, \ldots, m\}$. Such distances are called **Hoeffding** distances, because their form is used in Hoeffding's combinatorial central limit theorem. See Section 3.2. (All the functions we consider are also symmetric, $\delta(i,j) = \delta(j,i)$, though that restriction is not necessary.) The Spearman distances can be generalized to the L_p distances, where $\delta(i,j) = |i-j|^p$ for some $p > 0$. The Hamming distance is the limit of the L_p distance as $p \to 0$. The maximum distance is in a sense the limit as $p \to \infty$, but in that case we take the limit of $(\sum |y_i - i|^p)^{1/p}$.

In Section 3.1, we present convenient formulas for the first two moments of a Hoeffding distance. Section 3.2 gives Hoeffding's condition for the central limit theorem to apply. Section 3.3 gives a general approach to finding the exact distribution of the distance which is practical for m up to 25 or so, versus 10 to 15 for using exhaustive enumeration.

3.1 First two moments

For Hoeffding distances, some moment formulas are easier to derive using matrix representations. Let d be a Hoeffding distance based on δ as in (3.1). Let $\boldsymbol{Q_y}$ be the permutation matrix corresponding to $\boldsymbol{y}$, i.e.,

$$\boldsymbol{y} = \boldsymbol{\omega} \boldsymbol{Q_y}', \tag{3.2}$$

and define $\boldsymbol{Q_x}$ via $\boldsymbol{x} = \boldsymbol{\omega} \boldsymbol{Q_x}'$. Also, let $\boldsymbol{\Delta}$ be the $m \times m$ matrix with $\boldsymbol{\Delta}_{ij} = \delta(i,j)$. We then can write

$$d(\boldsymbol{y}, \boldsymbol{x}) = \text{trace}(\boldsymbol{Q_y} \boldsymbol{\Delta} \boldsymbol{Q_x}'). \tag{3.3}$$

Since $\boldsymbol{Q_\omega} = \boldsymbol{I}_m$, the $m \times m$ identity matrix, $d(\boldsymbol{y}, \boldsymbol{\omega}) = \text{trace}(\boldsymbol{Q_y} \boldsymbol{\Delta})$. Also, if $\boldsymbol{Y} \sim \text{Uniform}(\mathcal{P}_m)$, then $\boldsymbol{Q_Y}$ is uniformly distributed over $\mathcal{Q}_m$, the set of $m \times m$ permutation matrices. Thus we are interested in the distribution of D, where

$$D = \text{trace}(\boldsymbol{Q} \boldsymbol{\Delta}), \quad \boldsymbol{Q} \sim \text{Uniform}(\mathcal{Q}_m). \tag{3.4}$$

Each element of $\boldsymbol{Q}$ is Bernoulli($\frac{1}{m}$), hence

$$\mathrm{E}[\boldsymbol{Q}] = \frac{1}{m}\mathbf{1}_m'\mathbf{1}_m, \text{ where } \mathbf{1}_m = (1,\ldots,1), \tag{3.5}$$

the $1 \times m$ vector of 1's. For the covariance matrix, first note that $\mathrm{Var}[Q_{ij}] = (m-1)/m^2$ since it is Bernoulli. For elements within the same row, we have $\mathrm{E}[Q_{ij}Q_{ij'}] = 0$ if $j \neq j'$, since at most one element can equal 1. Similarly for two elements in the same column. For two elements in different rows and columns, the chance both equal 1 is $1/(m(m-1))$. For example, $\mathrm{P}[Q_{11} = 1 \;\&\; Q_{22} = 1] = \mathrm{P}[Q_{11} = 1]\mathrm{P}[Q_{22} = 1 \,|\, Q_{11} = 1]$. The first probability is $1/m$, but given $Q_{11} = 1$, the second row must have a 1 in one of the slots $2,\ldots,m$, hence the conditional probability is $1/(m-1)$. Thus the covariances are

$$\mathrm{Cov}[Q_{ij}, Q_{i'j'}] = \begin{cases} (m-1)/m^2 & \text{if } i = i', j = j' \\ -1/m^2 & \text{if } i = i', j \neq j' \\ -1/m^2 & \text{if } i \neq i', j = j' \\ 1/(m^2(m-1)) & \text{if } i \neq i', j \neq j' \end{cases}. \tag{3.6}$$

The next lemma helps to neaten up the formula. Note that if $m = 1$, $\boldsymbol{Q} \equiv 1$, so that $\mathrm{Cov}[\boldsymbol{Q}] = 0$.

Lemma 3.1. *If* $\boldsymbol{Q} \sim$ *Uniform*$(\mathcal{Q}_m)$ *and* $m > 1$,

$$\mathrm{Cov}[\boldsymbol{Q}] = \frac{1}{m-1}\,\boldsymbol{H} \otimes \boldsymbol{H}. \tag{3.7}$$

We have to explain the notation. First, $\boldsymbol{H}$ is the $m \times m$ **centering** matrix,

$$\boldsymbol{H} = \boldsymbol{I} - \frac{1}{m}\,\mathbf{1}'\mathbf{1}. \tag{3.8}$$

If $\boldsymbol{z} = (z_1,\ldots,z_m)$, then $\boldsymbol{H}\boldsymbol{z}' = (z_1 - \bar{z},\ldots,z_m - \bar{z})'$. Also, $\boldsymbol{H}$ is symmetric and idempotent. Thus in particular

$$\boldsymbol{H}\mathbf{1}' = (0,0,\ldots,0)' \text{ and } \boldsymbol{H}\boldsymbol{H} = \boldsymbol{H}. \tag{3.9}$$

We use the convention that the covariance matrix of a matrix is the covariance matrix of the vector formed by stringing out its rows. Thus for an $a \times b$ matrix $\boldsymbol{W}$,

$$\mathrm{Cov}[\boldsymbol{W}] \equiv \mathrm{Cov}[\mathrm{row}(\boldsymbol{W})], \text{ where } \mathrm{row}(\boldsymbol{W}) = (W_{11}, W_{12},\ldots,W_{1b}, W_{21},\ldots,W_{2b}, \ldots, W_{a1},\ldots,W_{ab}). \tag{3.10}$$

The "$\otimes$" in (3.7) indicates a Kronecker product, where if $\boldsymbol{A}$ is $k \times k'$ and $\boldsymbol{B}$ is $l \times l'$, then $\boldsymbol{A} \otimes \boldsymbol{B}$ is the $kl \times k'l'$ matrix

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{pmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \cdots & a_{1k'}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & \cdots & a_{2k'}\boldsymbol{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1}\boldsymbol{B} & a_{k2}\boldsymbol{B} & \cdots & a_{kk'}\boldsymbol{B} \end{pmatrix}. \tag{3.11}$$

Kronecker products have a number of useful properties. A couple we need here follow. Suppose $\mathrm{Cov}[\boldsymbol{W}] = \boldsymbol{\Sigma} \otimes \boldsymbol{\Lambda}$. The individual variances and covariances can be found using

$$\mathrm{Cov}[W_{ij}, W_{i'j'}] = \sigma_{ii'}\lambda_{jj'}. \tag{3.12}$$

If $\boldsymbol{A}$ and $\boldsymbol{B}$ are matrices for which the multiplications make sense,

$$\mathrm{Cov}[\boldsymbol{AWB}'] = \boldsymbol{A\Sigma A}' \otimes \boldsymbol{B\Lambda B}'. \tag{3.13}$$

Proof of Lemma 3.1. To show (3.7), we note that the diagonals of $\boldsymbol{H}$ are all $(m-1)/m$, and the off-diagonals are all $-1/m$. Thus using (3.12) on the right-hand side of (3.7),

$$\begin{aligned}
\mathrm{Var}[Q_{ii}] &= \frac{1}{m-1}\left(\frac{m-1}{m}\right)^2 = \frac{m-1}{m^2};\\
\mathrm{Cov}[Q_{ij}, Q_{i'j}] = \mathrm{Cov}[Q_{ij}, Q_{ij'}] &= \frac{1}{m-1}\frac{m-1}{m}\left(-\frac{1}{m}\right) = -\frac{1}{m^2}, \quad i \neq i', j \neq j';\\
\mathrm{Cov}[Q_{ij}, Q_{i'j'}] &= \frac{1}{m-1}\left(-\frac{1}{m}\right)^2 = \frac{1}{(m-1)m^2}, \quad i \neq i', j \neq j'.
\end{aligned} \tag{3.14}$$

These equations conform to (3.6), proving (3.7). □

The main results of this section are expressions for the mean and variance of a Hoeffding distance D. We use the analysis-of-variance-like notation on δ:

$$\delta(\cdot, j) = \frac{1}{m}\sum_{i=1}^{m}\delta(i,j), \quad \delta(i,\cdot) = \frac{1}{m}\sum_{j=1}^{m}\delta(i,j), \quad \delta(\cdot,\cdot) = \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\delta(i,j), \tag{3.15}$$

and

$$\delta^*(i,j) = \delta(i,j) - \delta(\cdot,j) - \delta(i,\cdot) + \delta(\cdot,\cdot). \tag{3.16}$$

Note that the δ^*'s are the interaction terms if we consider $\boldsymbol{\Delta}$ a two-way layout.

Proposition 3.2. *If* $\boldsymbol{Y} \sim$ *Uniform*$(\mathcal{P}_m)$, $D \equiv d(\boldsymbol{Y}, \boldsymbol{\omega})$ *for* d *in (3.1) satisfies*

$$\begin{aligned}
E[D] &= \frac{1}{m}\mathbf{1}\boldsymbol{\Delta}\mathbf{1}' = m\delta(\cdot,\cdot)\\
&= -\mathit{trace}(\boldsymbol{H\Delta}) = -\mathit{trace}(\boldsymbol{H\Delta H}) = -\sum_{i=1}^{m}\delta^*(i,i), \quad \textit{and if } m > 1,\\
\mathrm{Var}[D] &= \frac{1}{m-1}\mathit{trace}(\boldsymbol{H\Delta' H\Delta}) = \frac{1}{m-1}\sum_{i=1}^{m}\sum_{j=1}^{m}\delta^*(i,j)^2.
\end{aligned} \tag{3.17}$$

If $m = 1$, $\mathrm{Var}[D] = 0$.

Proof. We use the representation (3.4), so that $D = \mathrm{trace}(\boldsymbol{Q\Delta})$ where $\boldsymbol{Q} \sim \mathrm{Uniform}(\mathcal{Q}_m)$. For the mean, (3.5) and the linearity of trace show that

$$E[D] = \mathrm{trace}(E[\boldsymbol{Q}]\boldsymbol{\Delta}) = \frac{1}{m}\mathrm{trace}(\mathbf{1}'\mathbf{1}\boldsymbol{\Delta}) = \frac{1}{m}\mathbf{1}\boldsymbol{\Delta}\mathbf{1}', \tag{3.18}$$

since trace($\boldsymbol{AB}$) = trace($\boldsymbol{BA}$). Also, $\mathbf{1}\boldsymbol{\Delta}\mathbf{1}'$ is the sum of all the $\delta(i,j)$'s, hence the second expression for the mean follows from (3.15). Next, since the diagonals of $\boldsymbol{\Delta}$ are zero, trace($\boldsymbol{I\Delta}$) = 0, hence by (3.8) and (3.9) we can also write the mean as in the second line. Then the final equality in that line holds since the elements of $\boldsymbol{H\Delta H}$ are the $\delta^*(i,j)$'s.

For the variance, from (3.7) and (3.13),

$$\mathrm{Cov}[\boldsymbol{Q\Delta}] = \frac{1}{m-1}\boldsymbol{H} \otimes \boldsymbol{\Delta}'\boldsymbol{H\Delta}. \tag{3.19}$$

Now if $\boldsymbol{W}$ is $m \times m$ with covariance matrix $\boldsymbol{\Sigma} \otimes \boldsymbol{\Omega}$, then

$$\begin{aligned}\mathrm{Var}[\mathrm{trace}(\boldsymbol{W})] &= \sum_{i=1}^{m}\sum_{j=1}^{m}\mathrm{Cov}[W_{ii}, W_{jj}] \\ &= \sum_{i=1}^{m}\sum_{j=1}^{m}\sigma_{ij}\omega_{ij} \quad \text{by (3.12)} \\ &= \mathrm{trace}(\boldsymbol{\Sigma\Lambda}).\end{aligned} \tag{3.20}$$

Thus the first expression for Var[D] in (3.17) follows from (3.19), and we can write

$$\mathrm{trace}(\boldsymbol{H\Delta}'\boldsymbol{H\Delta}) = \mathrm{trace}(\boldsymbol{H\Delta}'\boldsymbol{HH\Delta H}) = \mathrm{trace}((\boldsymbol{H\Delta H})'(\boldsymbol{H\Delta H})), \tag{3.21}$$

from which the last equality follows. □

3.2 Hoeffding's CLT

For the Hoeffding distances, we have the famous theorem by Hoeffding (1951). We can get away with a weaker version of his Theorem 4, the condition given by his Equation (12).

Theorem 3.3 (Hoeffding)**.** *Suppose that for Hoeffding distance (3.4),*

$$\frac{\max\{\delta^*(i,j)^2 \,|\, 1 \leqslant i,j \leqslant m\}}{\mathrm{Var}(D)} \longrightarrow 0 \tag{3.22}$$

as $m \to \infty$, where δ^* is defined in (3.16). Then

$$\frac{D - E[D]}{\sqrt{\mathrm{Var}(D)}} \longrightarrow N(0,1). \tag{3.23}$$

We can replace the δ^* by δ in the numerator of (3.22): By (3.16) and the triangle inequality,

$$|\delta^*(i,j)| \leqslant |\delta(i,j)| + |\delta(\cdot,j)| + |\delta(i,\cdot)| + |\delta(\cdot,\cdot)| \leqslant 4\max\{|\delta(i,j)| \,|\, 1 \leqslant i,j \leqslant m\}, \tag{3.24}$$

since the absolute value of an average is no larger than the maximum absolute value. Thus the maximum of the δ^{*2}'s is less than or equal to 16 times the maximum of the δ^2's.

See Sections 4.1 and 5.1 for applications in Spearman's ρ and footrule cases, respectively.

3.3 Exact distribution: The splitting algorithm

In what follows, we will suppose that the range of δ is in the nonnegative integers, and that $\delta(i,i) = 0$. If not, some minor modifications would be needed.

To find the exact distribution of

$$D \equiv d(\boldsymbol{Y}, \boldsymbol{\omega}) = \sum_{i=1}^{m} \delta(Y_i, i), \quad \boldsymbol{Y} \sim \text{Uniform}(\mathcal{P}_m), \tag{3.25}$$

(as in (1.1), $\mathcal{P}_m$ is the set of permutations of the integers $\{1,\ldots,m\}$), we can use a splitting algorithm as introduced in Franklin (1988) for Spearman's ρ distance. Start with two subvectors. Choose $m_1 \approx m/2$, and for $\boldsymbol{y} \in \mathcal{P}_m$, let $\boldsymbol{y}^{(1)} = (y_1,\ldots,y_{m_1})$ and $\boldsymbol{y}^{(2)} = (y_{m_1+1},\ldots,y_m)$, and similarly split up $\boldsymbol{\omega}$: $\boldsymbol{\omega}^{(1)} = (1,\ldots,m_1)$, $\boldsymbol{\omega}^{(2)} = (m_1+1,\ldots,m)$. Then

$$d(\boldsymbol{y}, \boldsymbol{\omega}) = d(\boldsymbol{y}^{(1)}, \boldsymbol{\omega}^{(1)}) + d(\boldsymbol{y}^{(2)}, \boldsymbol{\omega}^{(2)}). \tag{3.26}$$

Now $\boldsymbol{Y}^{(1)}$ and $\boldsymbol{Y}^{(2)}$ are not independent, but conditioning on the set of values each vector chooses from, they are. That is, $\mathcal{S} = (\mathcal{R}_1, \mathcal{R}_2)$ is a splitting, where $\mathcal{R}_1$ is a subset of m_1 distinct elements from $1,\ldots,m$, and $\mathcal{R}_2$ is its complement, $\{1,\ldots,m\} - \mathcal{R}_1$. Also, let

$$\mathcal{P}(\mathcal{R}_i) = \{\text{permutations of elements in } \mathcal{R}_i\}. \tag{3.27}$$

Then under the assumption that $\boldsymbol{Y} \sim \text{Uniform}(\mathcal{P}_m)$,

$$\boldsymbol{Y}^{(1)} \text{ and } \boldsymbol{Y}^{(2)} \text{ are independent } \textbf{given} \text{ that } \boldsymbol{Y}^{(1)} \in \mathcal{P}(\mathcal{R}_1) \ (\Leftrightarrow \ \boldsymbol{Y}^{(2)} \in \mathcal{P}(\mathcal{R}_2)), \tag{3.28}$$

$$\begin{aligned} \boldsymbol{Y}^{(1)} \mid \boldsymbol{Y}^{(1)} \in \mathcal{P}(\mathcal{R}_1) &\sim \text{Uniform}(\mathcal{P}(\mathcal{R}_1)), \\ \boldsymbol{Y}^{(2)} \mid \boldsymbol{Y}^{(2)} \in \mathcal{P}(\mathcal{R}_2) &\sim \text{Uniform}(\mathcal{P}(\mathcal{R}_2)), \end{aligned} \tag{3.29}$$

and

$$\mathcal{R}_1 \sim \text{Uniform}(\{\text{All possible subsets of } m_1 \text{ distinct elements from } 1,\ldots,m\}). \tag{3.30}$$

For splitting $\mathcal{S}$, let $f_i(x \mid \mathcal{S})$ be the following conditional density of d:

$$f_i(x \mid \mathcal{S}) = P[d(\boldsymbol{Y}^{(i)}, \boldsymbol{\omega}^{(i)}) = x \mid \boldsymbol{Y}^{(i)} \in \mathcal{P}(\mathcal{R}_i)], \quad i = 1,2. \tag{3.31}$$

Then by (3.26) and the conditional independence in (3.28), we can use convolutions to find the conditional density of $d(\boldsymbol{Y}, \boldsymbol{\omega})$:

$$\begin{aligned} f(x \mid \mathcal{S}) &= P[d(\boldsymbol{Y}, \boldsymbol{\omega}) = x \mid \boldsymbol{Y}^{(i)} \in \mathcal{P}(\mathcal{R}_i), i = 1,2] \\ &= P[d(\boldsymbol{Y}^{(1)}, \boldsymbol{\omega}^{(1)}) + d(\boldsymbol{Y}^{(2)}, \boldsymbol{\omega}^{(2)}) = x \mid \boldsymbol{Y}^{(i)} \in \mathcal{P}(\mathcal{R}_i), i = 1,2] \\ &= \sum_{u=0}^{x} P[d(\boldsymbol{Y}^{(1)}, \boldsymbol{\omega}^{(1)}) = u \ \& \ d(\boldsymbol{Y}^{(2)}, \boldsymbol{\omega}^{(2)}) = x-u \mid \boldsymbol{Y}^{(i)} \in \mathcal{P}(\mathcal{R}_i), i = 1,2] \\ &= \sum_{u=0}^{x} P[d(\boldsymbol{Y}^{(1)}, \boldsymbol{\omega}^{(1)}) = u \mid \boldsymbol{Y}^{(1)} \in \mathcal{P}(\mathcal{R}_1)] P[d(\boldsymbol{Y}^{(2)}, \boldsymbol{\omega}^{(2)}) = x-u \mid \boldsymbol{Y}^{(2)} \in \mathcal{P}(\mathcal{R}_2)] \\ &= \sum_{u=0}^{x} f_1(u \mid \mathcal{S}) f_2(x-u \mid \mathcal{S}). \end{aligned} \tag{3.32}$$

Then the unconditional density of $d(\boldsymbol{Y},\boldsymbol{\omega})$ is found by taking the expectation of the conditional density function over splittings $\mathcal{S}$ in (3.32):

$$f(x) = P[d(\boldsymbol{Y},\boldsymbol{\omega}) = x] = E[f(x \mid \mathcal{S})]. \tag{3.33}$$

The expression (3.33) leads us to our algorithm. For each splitting, we enumerate all the values of $d(\boldsymbol{y}^{(i)},\boldsymbol{\omega}^{(i)})$ for $\boldsymbol{y}^{(i)} \in \mathcal{P}(\mathcal{R}_i)$, $i = 1,2$, and find their densities:

$$f_i(x \mid \mathcal{S}) = \frac{1}{m_i!}\#\{\boldsymbol{y}^{(i)} \in \mathcal{P}(\mathcal{R}_i) \mid d(\boldsymbol{y}^{(i)},\boldsymbol{\omega}^{(i)}) = x\}, \tag{3.34}$$

where $m_2 = m - m_1$. The conditional density of $d(\boldsymbol{Y},\boldsymbol{\omega})$ is then the convolution

$$f(x \mid \mathcal{S}) = \sum_{u=0}^{x} f_1(u \mid \mathcal{S}) f_2(x-u \mid \mathcal{S}). \tag{3.35}$$

Let $\boldsymbol{S}$ be the set of all splittings. The final answer takes the average over the splittings:

$$f(x) = \binom{m}{m_1}^{-1} \sum_{\mathcal{S}\in\boldsymbol{S}} f(x \mid \mathcal{S}). \tag{3.36}$$

The actual algorithm accumulates the probabilities by first setting $s(i) = 0$ for i in the support of D. It then iterates over all splittings $\mathcal{S}$, and over all u and v in the supports of $d(\boldsymbol{Y}^{(1)},\boldsymbol{\omega}^{(1)})$ and $d(\boldsymbol{Y}^{(2)},\boldsymbol{\omega}^{(2)})$, respectively, for $(\boldsymbol{Y}^{(1)},\boldsymbol{Y}^{(2)}) \in (\mathcal{R}_1,\mathcal{R}_2)$, the following code:

$$s(u+v) \leftarrow s(u+v) + f_1(u \mid \mathcal{S}) f_2(v \mid \mathcal{S}). \tag{3.37}$$

Then $f(x) = s(x)/\binom{m}{m_1}$.

The basic algorithm as in (1.3) enumerates all $m!$ rank vectors. The above algorithm is somewhat better since for each of the $\binom{m}{m_1}$ splittings, it enumerates over $m_1! + m_2!$ subvectors. Thus the ratio

$$\frac{\#\{\text{enumerations with splitting algorithm}\}}{\#\{\text{enumerations with basic algorithm}\}} = \frac{\binom{m}{m_1}(m_1! + m_2!)}{m!} = \frac{m_1! + m_2!}{m_1!m_2!} \approx \frac{2}{m_1!} \tag{3.38}$$

if $m_1 \approx m/2$. For $m = 10$ we reduce the number of enumerations by a factor of 60, and for $m = 20$, by a factor of about 1.8 million.

A further splitting helps even more. Given a splitting $(\mathcal{R}_1,\mathcal{R}_2)$, we split each $\mathcal{R}_i$ into two approximately equal-sized sets $(\mathcal{R}_{i1},\mathcal{R}_{i2})$, and similarly for the $\boldsymbol{y}^{(i)}$'s and $\boldsymbol{\omega}^{(i)}$'s. For each (i,j) we enumerate over $\mathcal{P}(\mathcal{R}_{ij})$ as in (3.34):

$$\begin{aligned} f_{ij}(x \mid \mathcal{R}_{ij}) &= P[P[d(\boldsymbol{Y}^{(ij)},\boldsymbol{\omega}^{(ij)}) = x \mid \boldsymbol{Y}^{(ij)} \in \mathcal{P}(\mathcal{R}_{ij})] \\ &= \frac{1}{m_{ij}!}\#\{\boldsymbol{y}^{(ij)} \in \mathcal{P}(\mathcal{R}_{ij}) \mid d(\boldsymbol{y}^{(ij)},\boldsymbol{\omega}^{(ij)}) = x\}, \end{aligned} \tag{3.39}$$

where $m_{ij} = \#\mathcal{R}_{ij}$. Similar to (3.36), for each i, we find

$$f_i(x \mid \mathcal{R}_i) = \binom{m_i}{m_{ij}}^{-1} \sum_{\text{splittings } (\mathcal{R}_{i1},\mathcal{R}_{i2})} \sum_{u=0}^{x} f_{i1}(u \mid \mathcal{R}_{i1}) f_{i2}(x-u \mid \mathcal{R}_{i2}). \tag{3.40}$$

Then use (3.36) again to find $f(x)$. The ratio corresponding to that in (3.38) is

$$\frac{1}{m!}\binom{m}{m_1}\left(\binom{m_1}{m_{11}}(m_{11}!+m_{12}!)+\binom{m_2}{m_{21}}(m_{21}!+m_{22}!)\right)\approx\frac{4}{m_1!m_{11}!}, \tag{3.41}$$

if the m_{ij}'s are approximately equal. E.g., with $m=20$, we reduce the number of enumerations by a factor of 60 over the one-split algorithm, and by about 10^8 over the basic algorithm.

Even with this more efficient algorithm, for $m=20$, we need $4\cdot 20!/(10!5!)\approx 2.23\times 10^{10}$ enumerations. Further splitting would not help much either. E.g., splitting each 5 into 2 and 3 only reduces the enumerations by a factor of $2!3!/(2!+3!)=2.25$, and the more complicated algorithm has more overhead, which we have not factored in.

Chapter 4

Spearman's distance

4.1 First two moments

Spearman's distance has a symmetric distribution about its mean, so that the odd moments other than the first are zero. Pearson (1907) appears to be the first to find the variance, whose proof he attributes to Student. Hotelling & Pabst (1936) find the fourth moment. David, Kendall, & Stuart (1951) calculate the first eight moments. We find the first two moments in this section, and the fourth through eighth moments in Section 4.2.

Spearman's distance uses $\delta_{\text{Spear}}(i,j) = (i-j)^2 = i^2 + j^2 - 2ij$, hence we can write

$$\boldsymbol{\Delta}_{\text{Spear}} = \boldsymbol{s}'\mathbf{1} + \mathbf{1}'\boldsymbol{s} - 2\boldsymbol{\omega}'\boldsymbol{\omega}, \quad \text{where} \quad \boldsymbol{s} = (1, 2^2, \ldots, m^2), \tag{4.1}$$

and $\boldsymbol{\omega} = (1, \ldots, m)$ as before. Using (3.9),

$$\boldsymbol{H}\boldsymbol{\Delta}_{\text{Spear}}\boldsymbol{H} = -2\boldsymbol{H}\boldsymbol{\omega}'\boldsymbol{\omega}\boldsymbol{H}, \tag{4.2}$$

hence

$$\delta^*(i,j) = -2(i-\nu)(j-\nu), \quad \text{where} \quad \nu = \frac{m+1}{2}. \tag{4.3}$$

Then using (3.17) for the mean we have

$$E[d_{\text{Spear}}(\boldsymbol{Y}, \boldsymbol{\omega})] = 2\sum_{i=1}^{m}(i-\nu)^2 = \frac{m(m^2-1)}{6}. \tag{4.4}$$

The final expression follows by noting that the variance of a discrete uniform random variable on $\{1, \ldots, m\}$ is $(m^2-1)/12$.

For the variance, assuming $m > 1$, (3.17) yields

$$\begin{aligned} Var[d_{\text{Spear}}(\boldsymbol{Y}, \boldsymbol{\omega})] &= \frac{4}{m-1}\sum_{i=1}^{m}\sum_{j=1}^{m}(i-\nu)^2(j-\nu)^2 \\ &= \frac{4}{m-1}\left(\frac{m(m^2-1)}{12}\right)^2 \\ &= \frac{m^2(m-1)(m+1)^2}{36}. \end{aligned} \tag{4.5}$$

4.2 The fourth (and higher) moments

Turn to higher moments. The distribution of Spearman's ρ is symmetric about its mean, so the odd central moments (other than the mean) are all 0. David et al. (1951) find the moments up to the eighth by hand. We are lucky to have technology to do the heavy lifting. Here we describe their method, and implement it in Mathematica®.

Note that similar to above, we can write $\delta(i,j) = (i-\nu)^2 + (j-\nu)^2 - 2(i-\nu)(j-\nu)$, so that

$$D_{\text{Spear}} \equiv d_{\text{Spear}}(\boldsymbol{Y}, \boldsymbol{\omega}) = \frac{m(m^2-1)}{6} - 2W, \quad \text{where } W = \sum_{i=1}^{m} W_i, \quad W_i \equiv (Y_i - \nu)(i-\nu). \tag{4.6}$$

We will work with W, then for $n \geqslant 2$ use

$$E[(D_{\text{Spear}} - E[D_{\text{Spear}}])^n] = (-2)^n \, E[(W - E[W])^n]. \tag{4.7}$$

We start by expanding the sum of W_i's into an n-fold summation of monomials of them, then separate the overall sum into sums depending on equalities among the indices. Write

$$W^n = \left(\sum_{i=1}^{m} W_i \right)^n = \sum_{i_1}^{m} \cdots \sum_{i_n}^{m} W_{i_1} \cdots W_{i_n}. \tag{4.8}$$

For a set of indices $\boldsymbol{i} = (i_1, \ldots, i_n)$, we consider all possible sets of equalities among them. For given $\boldsymbol{i}$, let $\boldsymbol{k}(\boldsymbol{i})$ be the set partition of $\{1, \ldots, n\}$ that describes the equalities, i.e.,

$$\boldsymbol{k}(\boldsymbol{i}) = (\mathcal{K}_1, \ldots, \mathcal{K}_r), \quad \text{where } i_a = i_b \iff a, b \in \mathcal{K}_k \text{ for some } k. \tag{4.9}$$

Here, r is the number of distinct values in $\boldsymbol{i}$. Note that $r \leqslant m$, since there are only m W_i's. In order to insure uniqueness of the set partitions, we require that

$$\min(\mathcal{K}_1) < \min(\mathcal{K}_2) < \cdots < \min(\mathcal{K}_r). \tag{4.10}$$

Then

$$W^n = \sum_{\{\mathcal{K} \in \mathcal{SP}_{n,m}\}} \sum_{\{\boldsymbol{i} \,|\, \boldsymbol{k}(\boldsymbol{i}) = \mathcal{K}\}} W_{i_1} \cdots W_{i_n}, \tag{4.11}$$

where $\mathcal{SP}_{n,m}$ is the set of set partitions of $\{1, \ldots, n\}$ with at most m components. The summands above are products of powers of the W_i's. For set partition $\mathcal{K}$, let $\boldsymbol{n}$ be the vector of cardinalities of the component sets:

$$\boldsymbol{n} = \boldsymbol{n}(\mathcal{K}) = (\#\mathcal{K}_1, \ldots, \#\mathcal{K}_r). \tag{4.12}$$

Then $\boldsymbol{n}$ is a **composition** of the integer n, i.e., a set of positive integers that sum to n. Now we have for given set partition $\mathcal{K}$ with $r \leqslant m$,

$$\sum_{\boldsymbol{i} \,|\, \boldsymbol{k}(\boldsymbol{i}) = \mathcal{K}} W_{i_1} \cdots W_{i_n} = \sum_{\substack{1 \leqslant j_1, \ldots, j_r \leqslant m \\ \text{distinct}}} W_{j_1}^{n_1} \cdots W_{j_r}^{n_r} \equiv V_{\boldsymbol{n}}, \quad \boldsymbol{n} = \boldsymbol{n}(\mathcal{K}). \tag{4.13}$$

Using the definition of W_i from (4.6), the expected value of $V_{\boldsymbol{n}}$ is

$$\mathsf{E}[V_{\boldsymbol{n}}] = \sum_{\substack{1\leqslant j_1,\dots,j_r\leqslant m\\ \text{distinct}}} \mathsf{E}[(Y_{j_1}-\nu)^{n_1}\cdots(Y_{j_r}-\nu)^{n_r}](j_1-\nu)^{n_1}\cdots(j_r-\nu)^{n_r}. \tag{4.14}$$

Since the distribution of $\boldsymbol{Y}$ is permutation invariant, the expected value in the summand does not depend on the $(j_1,\dots,j_r)$, as long as they are distinct. Thus

$$\mathsf{E}[V_{\boldsymbol{n}}] = \mathsf{E}[(Y_1-\nu)^{n_1}\cdots(Y_r-\nu)^{n_r}]\,\tau(\boldsymbol{n}) \;\text{ where }\; \tau(\boldsymbol{n}) = \sum_{\substack{1\leqslant j_1,\dots,j_r\leqslant m\\ \text{distinct}}}\cdots\sum (j_1-\nu)^{n_1}\cdots(j_r-\nu)^{n_r}. \tag{4.15}$$

The expected value can be found by averaging over all sets $(y_1,\dots,y_r)$, where the entries are distinct integers between 1 and m, just like the indices in τ. That is,

$$\mathsf{E}[(Y_1-\nu)^{n_1}\cdots(Y_r-\nu)^{n_r}] = \frac{\tau(\boldsymbol{n})}{(m)_r} \tag{4.16}$$

and

$$\mathsf{E}[V_{\boldsymbol{n}}] = \frac{\tau(\boldsymbol{n})^2}{(m)_r}. \tag{4.17}$$

The **Pochhammer symbol** $(m)_r$ is the falling factorial defined for nonnegative integers r by

$$(m)_0 = 1, (m)_r = m(m-1)\cdots(m-r+1) \text{ for } r>0. \text{ If } \; 0\leqslant r\leqslant m,\; (m)_r = \frac{m!}{(m-r)!}. \tag{4.18}$$

We can now find $\mathsf{E}[W^n]$ from (4.11) by summing the $\mathsf{E}[V_{\boldsymbol{n}}]$'s over the set partitions $\mathcal{K}$, but first note that $\tau(\boldsymbol{n})$ does not depend on the order of the n_j's in $\boldsymbol{n}$. Thus we can restrict to integer **partitions** of n, which are combinations with the elements in nonincreasing order, $n_1\geqslant n_2\geqslant\cdots\geqslant n_u$. Then

$$\mathsf{E}[W^n] = \sum_{\boldsymbol{n}\in\mathcal{IP}_{n,m}} \zeta_{\boldsymbol{n}}\,\frac{\tau(\boldsymbol{n})^2}{(m)_r}, \tag{4.19}$$

where $\mathcal{IP}_{n,m}$ is the set of integer partitions of n with at most m components, and

$$\begin{aligned}\zeta_{\boldsymbol{n}} &= \#\{\text{Set partitions } \mathcal{K} \text{ corresponding to } \boldsymbol{n}\}\\ &= \binom{n}{n_1,\dots,n_r}\frac{1}{u_1!\cdots u_r!}, \;\text{ where }\; u_j = \#\{n_i = j \mid i=1,\dots,r\}.\end{aligned} \tag{4.20}$$

We illustrate with $n=4$. The table below contains the integer partitions of n and the corresponding set partitions $\mathcal{K}$, where, e.g., $(13,2,4)$ represents $\mathcal{K}=(\{1,3\},\{2\},\{4\})$.

Integer partition $\boldsymbol{n}$	Set partitions $\mathcal{K}$	$\zeta_{\boldsymbol{n}}$
(4)	(1234)	1
$(3,1)$	$(123,4),(124,3),(134,2),(1,234)$	4
$(2,2)$	$(12,34),(13,24),(14,23)$	3
$(2,1,1)$	$(12,3,4),(13,2,4),(14,2,3),(1,23,4),(1,24,3),(1,2,34)$	6
$(1,1,1,1)$	$(1,2,3,4)$	1

(4.21)

The next task is to calculate the τ. The complication is the requirement that the indices be distinct. To simplify the calculations, we write each sum without that requirement, then subtract the summands that violate it. Those summands are in fact also τ's for some partition with smaller number of elements. Analogous to (4.11), for given $\boldsymbol{n} = (n_1, \ldots, n_r)$, we write

$$\sum_{i_1=1}^{m} \cdots \sum_{i_r=1}^{m} (i_1 - \nu)^{n_1} \cdots (i_r - \nu)^{n_r} = \sum_{\mathcal{K} \in \mathcal{SP}_{r,m}} \sum_{\{\boldsymbol{i} \,|\, \boldsymbol{k}(\boldsymbol{i}) = \mathcal{K}\}} (i_1 - \nu)^{n_1} \cdots (i_r - \nu)^{n_r}. \tag{4.22}$$

The individual summations on the left-hand side of (4.22) can be distributed to obtain

$$\sum_{i_1=1}^{m} (i_1 - \nu)^{n_1} \cdots \sum_{i_r=1}^{m} (i_r - \nu)^{n_r} \equiv \eta_{n_1} \cdots \eta_{n_r}, \tag{4.23}$$

where

$$\eta_k = \sum_{i=1}^{m} (i - \nu)^k = m \, E[(U - E[U])^k], \quad U \sim \text{Uniform}\{1, \ldots, m\}, \tag{4.24}$$

the k^{th} central moment of the discrete uniform, times m. By symmetry, $\eta_k = 0$ if k is odd. On the right-hand side of (4.22), $\tau(\boldsymbol{n})$ is the summation for all the indices being distinct, so that $\mathcal{K} = (\{1\}, \{2\}, \ldots, \{m\})$. For the other summations, some of the indices are equal, hence some of the powers are added together. That is, for $\boldsymbol{n}$ and $\mathcal{K}$, let

$$\boldsymbol{n}^*(\boldsymbol{n}, \mathcal{K}) = \text{sort}(n_1^*, \ldots, n_s^*), \quad \text{where} \quad n_j^* = \sum_{i \in \mathcal{K}_j} n_i, \tag{4.25}$$

where we put the elements of n^* into nonincreasing order. Then

$$\sum_{\boldsymbol{i} \,|\, \boldsymbol{k}(\boldsymbol{i}) = \mathcal{K}} (i_1 - \nu)^{n_1} \cdots (i_r - \nu)^{n_r} = \tau(n^*(\boldsymbol{n}, \mathcal{K})). \tag{4.26}$$

Rearranging (4.22), we obtain

$$\tau(\boldsymbol{n}) = \eta_{n_1} \cdots \eta_{n_r} - \sum_{\substack{\mathcal{K} \in \mathcal{SP}_{r,m} \\ \mathcal{K} \neq \{\{1\},\{2\},\ldots,\{r\}\}}} \tau(\boldsymbol{n}^*(\boldsymbol{n}, \mathcal{K})). \tag{4.27}$$

Each of the $\boldsymbol{n}^*$'s on the right-hand side has fewer than r components. Also, for the case all indices are equal ($r = 1$),

$$\tau(n) = \sum_{i=1}^{m} (i - \nu)^n = \eta_n. \tag{4.28}$$

Thus we can find the τ's sequentially, since for $r > 1$, the summation term in (4.27) is a function of the τ's for $\boldsymbol{n}$ of length less than r.

Turn to $E[W^4]$, taking $m \geqslant 4$. From (4.19) and (4.21), we have

$$E[W^4] = \frac{\tau(4)^2}{m} + 4\,\frac{\tau(3,1)^2}{m(m-1)} + 3\,\frac{\tau(2,2)^2}{m(m-1)} + 6\,\frac{\tau(2,1,1)^2}{m(m-1)(m-2)} + \frac{\tau(1,1,1,1)^2}{m(m-1)(m-2)(m-3)}. \tag{4.29}$$

By (4.28), we have $\tau(4) = \eta_4$. For partition $(3,1)$, there is only one possible equality among the two indices, hence

$$\tau(3,1) = \eta_3\eta_1 - \tau(4) = -\eta_4, \tag{4.30}$$

since $\eta_1 = \eta_3 = 0$. Similarly, for $(2,2)$,

$$\tau(2,2) = \eta_2^2 - \eta_4. \tag{4.31}$$

The set partitions $\mathcal{K}$ for $(2,1,1)$ are $(12,3), (13,2), (1,23)$, and (123), so that the respective $\boldsymbol{n}^*((2,1,1),\mathcal{K})$'s are $(3,1), (3,1), (2,2)$, and (4). Then by (4.27),

$$\tau(2,1,1) = -2\tau(3,1) - \tau(2,2) - \eta_4 = -\eta_2^2 + 2\eta_4. \tag{4.32}$$

For $\tau(1,1,1,1)$, we need the set partitions of $\{1,2,3,4\}$. These we have already exhibited in (4.21). Thus we can write

$$\begin{aligned}\tau(1,1,1,1) &= \eta_1^4 - 6\,\tau(2,1,1) - 3\,\tau(2,2) - 4\,\tau(3,1) - \tau(4)\\ &= 3\,\eta_2^2 - 6\,\eta_4.\end{aligned} \tag{4.33}$$

By (4.24), we need the variance and fourth central moment of the discrete uniform, which result in

$$\eta_2 = \frac{m(m^2-1)}{12} \quad \text{and} \quad \eta_4 = \frac{m(m^2-1)(3m^2-7)}{240}. \tag{4.34}$$

We have the ingredients in (4.29) through (4.34) to find $\mathsf{E}[W^4]$. Mathematica® will easily calculate the following:

$$\mathsf{E}[W^4] = \frac{(m-1)m^3(m+1)^3\left(25m^3-38m^2-35m+72\right)}{172800}. \tag{4.35}$$

Finally, by (4.7), we have

$$\mathsf{E}[(D_{\text{Spear}} - \mathsf{E}[D_{\text{Spear}}])^4] = 16\,\mathsf{E}[W^4] = \frac{(m-1)m^3(m+1)^3\left(25m^3-38m^2-35m+72\right)}{10800}. \tag{4.36}$$

Using the code in Section 4.2.1, we obtain the sixth central moment (if $m \geqslant 6$) as

$$\begin{aligned}\mathsf{E}[(D_{\text{Spear}} - \mathsf{E}[D_{\text{Spear}}])^6] = \frac{(m-1)m^3(m+1)^3}{3810240}&(1225m^8 - 4361m^7 - 178m^6 + 23818m^5\\ &- 22783m^4 - 50081m^3 + 54280m^2 + 44160m - 28800),\end{aligned} \tag{4.37}$$

and the eighth central moment (if $m \geqslant 8$) as

$$\begin{aligned}\mathsf{E}[(D_{\text{Spear}} - \mathsf{E}[D_{\text{Spear}}])^8] = \frac{(m-1)m^3(m+1)^3}{489888000}&(30625m^{13} - 218050m^{12} + 451718m^{11}\\ &+ 1090534m^{10} - 6275976m^9 + 2142858m^8 + 30402746m^7 - 27330110m^6 - 79689881m^5\\ &+ 71871632m^4 + 110888256m^3 - 74721024m^2 - 51867648m + 40642560).\end{aligned} \tag{4.38}$$

These conform to the formulas in David et al. (1951), though they look at the correlation rather than distance.

4.2.1 Mathematica code

The key functions find the various n^{th} moments (see Section 2.1) when there are m objects ranked: spearmanRawMoment[n,m], spearmanMoment[n,m], spearmanCumulant[n,m] and spearmanNormalizedCumulant[n,m]. In each case, the n must be a nonnegative integer (not just a symbol), while m can be either a positive integer, or a symbol (i.e., "m"). If it is a symbol, then the output is a function of m that is correct if $m \geqslant n$. For $m < n$, the routines need to have both n and m input as integers.

The code below shows the above functions plus the helper functions. Note that we need to load the Combinatorica package. The functions setPartitions and integerPartitions find the set and integer partitions of $\{1,\ldots,n\}$ and n, respectively, but limit the output to just those with at most m components. The eta is the η_k from (4.24) and etaprod is the product in (4.23); nstar is the function n^* in (4.25); tauk (and taukN) and tau (and tauN) denote the summands and sum, respectively, of τ in (4.20); zeta is ζ_n in (4.20); and denom is the product $m(m-1)\cdots(m-r+1)$ as in the denominator of the summands in (4.19). The moments of D_{Spear} are based on the moments of W in (4.6), which are calculated in spearmanWMoment, which uses the two other functions depending on whether m is an integer or symbol. We also use the functions moment2momentc and raw2cumulant from Section 2.4.

```
Needs["Combinatorica`"]

setPartitions[n_,m_] := Select[Combinatorica`SetPartitions[n],Function[Length[#1]<=m]];
integerPartitions[n_,m_] := Select[IntegerPartitions[n],Function[Length[#1]<=m]];
eta[k_,m_] := If[EvenQ[k],Sum[(i - (m + 1)/2)^k, {i, 1, m}],0];
etaprod[nn_,m_] := Apply[Times, Table[eta[ni,m], {ni,nn}]];
nstar[nn_, s_] := Sort[Map[Total, Table[Part[nn, si], {si, s}]], Greater];
zeta[nn_] := Apply[Multinomial, nn]/Apply[Times, Map[Factorial, Values[Counts[nn]]]];
denom[l_, m_] := Pochhammer[m - l + 1, l];
spearmanWMomentS[n_,m_] := Module[{tau,tauk},
    tau[nn_,m0_] := tau[nn,m0] = Total[Table[tauk[nn, s,m0],
                {s, Combinatorica`SetPartitions[Length[nn]]}]];
    tauk[nn_, s_,m0_] := If[Length[s] == Length[nn], etaprod[nn,m0], -tau[nstar[nn, s],m0]];
    Factor[Sum[zeta[nn]*tau[nn,m]^2/denom[Length[nn], m], {nn,IntegerPartitions[n]}]]];
spearmanWMomentN[n_,m_] := Module[{tauN,taukN},
    tauN[nn_,m0_] := tauN[nn,m0] = Total[Table[taukN[nn, s,m0],
                {s, setPartitions[Length[nn],m0]}]];
    taukN[nn_, s_,m0_] := If[Length[s] == Length[nn], etaprod[nn,m0], -tauN[nstar[nn, s],m0]];
    Factor[Sum[zeta[nn]*tauN[nn,m]^2/denom[Length[nn], m], {nn,integerPartitions[n,m]}]]];
spearmanWMoment[n_,m_] := (
    If[!IntegerQ[n],return[(Print["n must be a nonnegative integer"];Abort[])]];
    If[OddQ[n],Return[0]];
    If[IntegerQ[m],spearmanWMomentN[n,m],spearmanWMomentS[n,m]]);
spearmanRawMoment[n_,m_] := (-2)^n*moment2momentc[n,spearmanWMoment,-m*(m^2-1)/12,m];
spearmanMoment[n_,m_] := If[n==1,m*(m^2-1)/6,2^n*spearmanWMoment[n,m]];
spearmanCumulant[n_,m_] := If[n==1,m*(m^2-1)/6,(-2)^n*raw2cumulant[n,spearmanWMoment,m]];
spearmanNormalizedCumulant[n_,m_] := (
    If[n==1,Return[0]];
    If[n==2,Return[1]];
    spearmanCumulant[n,m]/spearmanCumulant[2,m]^(n/2)
)
```

4.3 Exact distribution

Maciak (2009) improves the Franklin (1988) splitting algorithm described in Section 3.3 for the Spearman distance by taking advantage of some redundancies. Also, van de Wiel & Bucchianico (2001) develop a different, but seemingly as effective, algorithm based on permanents. Here we describe some of Maciak's improvements.

It is easier to deal with the cross-product of the rankings, where

$$\begin{aligned} d_{\text{Spear}}(\boldsymbol{y}, \boldsymbol{x}) = \sum_{i=1}^{m} (y_i - x_i)^2 &= 2\sum_{i=1}^{m} i^2 - 2\sum_{i=1}^{m} y_i x_i \\ &= \frac{m(m+1)(2m+1)}{3} - 2\sum_{i=1}^{m} y_i x_i. \end{aligned} \tag{4.39}$$

The algorithm finds the distribution of

$$C(\boldsymbol{Y}, \boldsymbol{\omega}) = \sum_{i=1}^{m} Y_i \omega_i = \sum_{i=1}^{m} i Y_i. \tag{4.40}$$

The procedure runs along the same lines as that for D in (3.25), but with some shortcuts.

As in Section 3.3, we take $m_1 \approx m/2$, $m_2 = m - m_1$, and split C into $C^{(1)} + C^{(2)}$, where

$$C_1 = C(\boldsymbol{Y}^{(1)}, \boldsymbol{\omega}^{(1)}) = \sum_{i=1}^{m_1} i Y_i \quad \text{and} \quad C_2 = C(\boldsymbol{Y}^{(2)}, \boldsymbol{\omega}^{(2)}) = \sum_{i=m_1+1}^{m} i Y_i. \tag{4.41}$$

Condition on the splitting $\mathcal{S} = (\mathcal{R}_1, \mathcal{R}_2)$, so that the vectors $\boldsymbol{Y}^{(1)}$ and $\boldsymbol{Y}^{(2)}$ are independent, $\boldsymbol{Y}^{(i)}$ being uniformly distributed over permutations of $\mathcal{R}_i$, $i = 1, 2$, as in (3.28) and (3.29). Let $\boldsymbol{y}^{(1)}$ and $\boldsymbol{y}^{(2)}$ be fixed members of $\mathcal{P}(\mathcal{R}_1)$ and $\mathcal{P}(\mathcal{R}_2)$, respectively. (E.g., if $\mathcal{R}_1 = \{r_1, \ldots, r_{m_1}\}$, then $\boldsymbol{y}^{(1)} = (r_1, \ldots, r_{m_1})$.) Then we can equivalently represent the conditional distributions of the C_i's by

$$C_1 = C(\boldsymbol{y}^{(1)}, \boldsymbol{W}^{(1)}) \quad \text{and} \quad C_2 = m_1 \sigma_2 + C(\boldsymbol{y}^{(2)}, \boldsymbol{W}^{(2)}), \tag{4.42}$$

where

$$\boldsymbol{W}^{(1)} \text{ and } \boldsymbol{W}^{(2)} \text{ are independent, } \boldsymbol{W}^{(1)} \sim \text{Uniform}(\mathcal{P}_{m_1}), \text{ and } \boldsymbol{W}^{(2)} \sim \text{Uniform}(\mathcal{P}_{m_2}), \tag{4.43}$$

and we set $\sigma_i = \sum_{j=1}^{m_i} y_j^{(i)}$. (Note that σ_i is the same for every $\boldsymbol{y}^{(i)} \in \mathcal{R}_i$.)

To calculate the distribution of $C(\boldsymbol{y}^{(i)}, \boldsymbol{W}^{(i)})$, we enumerate the permutations of $1, \ldots, m_i$:

$$f_i(c \mid \mathcal{S}) = \frac{1}{m_i!} \#\{\boldsymbol{w}^{(i)} \in \mathcal{P}_{m_i} \mid C(\boldsymbol{y}^{(i)}, \boldsymbol{w}^{(i)}) = c\}. \tag{4.44}$$

But note that for $\boldsymbol{w}^{(i)} \in \mathcal{P}_{m_i}$, $(m_i + 1) - \boldsymbol{w}^{(i)} \in \mathcal{P}_{m_i}$ as well, and

$$C(\boldsymbol{y}^{(i)}, (m_i + 1) - \boldsymbol{w}^{(i)}) = (m_i + 1)\sigma_i - C(\boldsymbol{y}^{(i)}, \boldsymbol{w}^{(i)}). \tag{4.45}$$

If we let $\mathcal{P}^*_{m_i}$ be the set of permutations that has exactly one of each pair $\{\boldsymbol{w}, m_i + 1 - \boldsymbol{w}\}$, $\boldsymbol{w} \in \mathcal{P}_{m_i}$, then we need enumerate over only half the permutations:

$$f_i(c \mid \mathcal{S}) = \frac{g_i(c) + g_i((m_i+1)\sigma_i - c)}{m_i!}, \tag{4.46}$$

where

$$g_i(c) = \#\{\boldsymbol{w}^{(i)} \in \mathcal{P}^*_{m_i} \mid C(\boldsymbol{y}^{(i)}, \boldsymbol{w}^{(i)}) = c\}. \tag{4.47}$$

To find the conditional distribution of $C(\boldsymbol{Y}, \boldsymbol{\omega})$, i.e., of $C_1 + C_2$ in (4.42), we convolve the f_1 and f_2 as in (3.35), and also shift by a constant:

$$f(c + m_1\sigma_2 \mid \mathcal{S}) = h(c \mid \mathcal{S}), \quad \text{where} \quad h(c \mid \mathcal{S}) = \sum_{u=0}^{c} f_1(u \mid \mathcal{S}) f_2(c - u \mid \mathcal{S}). \tag{4.48}$$

To find the unconditional density of $C(\boldsymbol{Y}, \boldsymbol{\omega})$, we average over the splittings as in (3.36):

$$f(x) = \binom{m}{m_1}^{-1} \sum_{\mathcal{S} \in \boldsymbol{S}} f(x \mid \mathcal{S}) = \binom{m}{m_1}^{-1} \sum_{\mathcal{S} \in \boldsymbol{S}} h(x - m_1\sigma_2 \mid \mathcal{S}). \tag{4.49}$$

If m is even and we take $m_1 = m_2 \equiv \overline{m}$, another shortcut arises from noting that $\boldsymbol{W}^{(1)}$ and $\boldsymbol{W}^{(2)}$ have the same distribution. Thus the distribution of $(C(\boldsymbol{y}^{(1)}, \boldsymbol{W}^{(1)}), C(\boldsymbol{y}^{(2)}, \boldsymbol{W}^{(2)}))$ conditioning on the splitting $(\mathcal{R}_1, \mathcal{R}_2)$ is the same as that of $(C(\boldsymbol{y}^{(2)}, \boldsymbol{W}^{(1)}), C(\boldsymbol{y}^{(1)}, \boldsymbol{W}^{(2)}))$ conditioning on the splitting $(\mathcal{R}_2, \mathcal{R}_1)$. This symmetry implies that we have to find the convolution over only half of the splittings. Let $\boldsymbol{S}^*$ be splittings $(\mathcal{R}_1, \mathcal{R}_2)$ for which $1 \in \mathcal{R}_1$. (Thus each $\mathcal{S} \in \boldsymbol{S}$ is either $(\mathcal{R}_1, \mathcal{R}_2)$ or $(\mathcal{R}_2, \mathcal{R}_1)$ for some $(\mathcal{R}_1, \mathcal{R}_2) \in \mathcal{S}^*$.) The unconditional density of C is found by using the same h in (4.48) for $\mathcal{S} = (\mathcal{R}_1, \mathcal{R}_2)$ and $\mathcal{S} = (\mathcal{R}_2, \mathcal{R}_1)$, but with different shifts for the f:

$$f(x) = \binom{m}{m_1}^{-1} \sum_{\mathcal{S} \in \boldsymbol{S}^*} (h(x - \overline{m}\sigma_2 \mid \mathcal{S}) + h(x - \overline{m}\sigma_1 \mid \mathcal{S})). \tag{4.50}$$

As in (3.39) through (3.41), we can further split $\mathcal{R}_1$ and $\mathcal{R}_2$, and use the extra symmetry as in (4.50) if either or both of m_1 and m_2 are even. Maciak (2009) has even more symmetries that can speed up the algorithm, but we stopped at the above.

Once we obtain the density for $C(\boldsymbol{Y}, \boldsymbol{\omega})$ in (4.40), we can obtain the density of $d_{\text{Spear}}(\boldsymbol{Y}, \boldsymbol{\omega})$ in (4.39):

$$f_{\text{Spear}}(z) = P[d_{\text{Spear}}(\boldsymbol{Y}, \boldsymbol{\omega}) = z] = f((z - \theta)/2), \quad \theta = \frac{m(m+1)(2m+1)}{3}. \tag{4.51}$$

The support is $\{0, 2, 4, \ldots, m(m^2-1)/3\}$.

4.4 Normal and Edgeworth approximations

Hotelling & Pabst (1936) prove the asymptotic normality of Spearman's distance. We will use Hoeffding's theorem, our Theorem 3.3. From (4.3), we have

$$\max\{\delta^*(i,j)^2 \mid 1 \leqslant i, j \leqslant m\} = 4(m - \nu)^4 = \frac{(m-1)^4}{4}. \tag{4.52}$$

By (4.5), $Var[D_{Spear}] = m^2(m-1)(m+1)^2/36$. Thus the ratio in (3.22), the maximum over the variance, is of order $1/m$, which goes to zero. The theorem then shows that

$$\frac{D_{Spear} - E[D_{Spear}]}{\sqrt{Var[D_{Spear}]}} \longrightarrow N(0,1). \tag{4.53}$$

David, Kendall, & Stuart (1951) present the 6-term Edgeworth expansion for the density of Spearman's ρ, calculated the first eight moments and cumulants by hand. Best & Roberts (1975) implement the approximation, simplified somewhat, in FORTRAN. Maciak (2009), who found the exact distribution for m up to 25, compared a number of approximations to the exact distribution for these m. See also Best & Roberts and Franklin (1988) for comparisons for smaller m. Maciak recommends the Edgeworth expansion for common p-value calculations.

We compared L-term Edgeworth expansions of Spearman's distance for $L = 0, 2, 4, 6, 8$, and 10 (an odd one is the same as the previous even one), using both the density and distribution function versions of the expansions, for m between 5 and 24. (The sample size is $N = 1$. Larger sample size yield better approximations.) Summary statistics included the maxima of the errors for the density and the distribution function, as well as the maximum error for the relative p-values for p-values above 0.00001. That is, with $\widehat{f}$ and $\widehat{F}$ being the estimates of the true density f and true distribution function F, respectively, we calculate

$$ME_{dens} = \max_{x \in \mathcal{X}} |\widehat{f}(x) - f(x)|, \quad ME_{DF} = \max_{x \in \mathcal{X}} |\widehat{F}(x) - F(x)|, \tag{4.54}$$

and with $pv(x) = \min\{F(x), 1 - F(x) + f(x)\}$ and $\widehat{pv}(x) = \min\{\widehat{F}(x), 1 - \widehat{F}(x) + \widehat{f}(x)\}$ being the p-value and its estimate,

$$MRE_{pv} = \max_{x \in \mathcal{X} \,|\, pv(x) \geqslant .00001} \frac{|\widehat{pv}(x) - pv(x)|}{pv(x)}. \tag{4.55}$$

The estimates based on the expansion for the distribution function were overall somewhat better than those for the density, so we focus on the former. As expected, the estimates generally improve as m and L increase. Figure 4.1 compares the maximum errors in the density estimation. We see that there is substantial improvement going from the normal approximation to the 2-term expansion to the 4-term expansion. These is less separation among the 6-, 8-, and 10-term expansions. The main driver of the error appears to be the center of the distribution, where the true density is very spiky. Figure 4.4 shows the true density for x near the mean of 2300. We see the heights jump up and down for consecutive values of x. The smooth curve is the 10-term Edgeworth estimate, though the 6- and 8-term estimates are virtually indistinguishable from the 10-term estimate. Figure 4.5 plots the same data but as the error in the estimate, over the entire range of x. Note that the largest errors are indeed in the center, and they errors jump between positive and negative for nearby values of x. The conclusion is that no smooth density is going to be able follow closely the ups-and-downs, hence higher L will not improve by much the maximum error in the density.

Figure 4.2 shows the maximum errors in estimating the distribution function. Here the picture is a bit clearer, with substantial gains as we increase L. The approximations are quite good, with maximum errors between 10^{-3} and 10^{-4} even for $m = 10$ and $L = 4$. For $m = 24$

and $L = 10$, the error is about 10^{-6}. The approximations are not very good for the p-values according to the maximum relative errors (for p-values over 0.00001). See 4.3. For $m \leqslant 15$ and any of the L, the relative error is over 1, and sometimes beween 10 and 100. For $M \geqslant 20$, $L = 10$ has relative error less than 10%, and for $m = 24$, about 1%. Equation (4.56) exhibits the errors in the $L = 10$ expansion for select values of m.

Considering all three types of error comparisons, it appears that the extra calculation in the 10-term expansion is worth it, and reasonably good for larger m. If the relative errors in the tails is not as important, even $L = 4$ is reasonable for $m \geqslant 15$, say.

	5	10	15	20	22	24
Density	0.0231	0.000344	3.54×10^{-5}	3.71×10^{-6}	9.94×10^{-8}	5.92×10^{-7}
Distribution function	0.0175	0.000349	2.59×10^{-5}	3.37×10^{-6}	9.96×10^{-7}	8.20×10^{-7}
Relative, p-value	2.11	32.9	1.65	0.0852	0.0226	0.00741

(4.56)

4.5 R code

Below are functions for finding moments, cumulants, and Edgeworth expansions for the distribution of Spearman's distance. The first twelve moments, cumulants, and normalized cumulants based on m are produced by the functions spearman_moments(m), spearman_cumulants(m), and spearman_normalized_cumulants(m), respectively. The function spearman_edgeworthf(x,m,L,n) calculates the values of the Edgeworth approximation to the density at the values x, where L is the number of terms in the expansion, and n is the sample size. If $n > 1$, then the values of x represent the sum of the distances, $\sum_{i=1}^{n} d_{Spear}(\boldsymbol{y}_i, \boldsymbol{\omega})$, the $\boldsymbol{y}_i$ being the m-length rank vectors. The corresponding distribution-function-based Edgeworth estimates are found by spearman_edgeworthF(x,m,L,n). The Edgeworth functions use the functions edgef and edgeF from Section 2.5.

The functions for the moments and cumulants have many unnecessarily long integers. Although I could have shortened them substantially, I decided to leave them as is since then the formulas are exact (as calculated by Mathematica in Section 4.2.1), hence may be of interest to some.

```
spearman_moments <- function(m) {
    if(m==1) return(rep(0,12))
    if(m==2) return(c(1,1,0,1,0,1,0,1,0,1,0,1))
    mfactor <- (m-1)*m^3*(1 + m)^3
    mu <- (m-1)*m*(1 + m)/6
    s2 <- (m-1)*m^2*(1 + m)^2/36
    if(m==3) return(c(mu,s2,0,96, 0, 1408, 0, 22016, 0, 350208, 0, 5595136))
    m4 <- mfactor*(72 + m*(-35 + m*(-38 + 25*m)))/10800
    if(m==4) return(c(mu,s2,0,m4, 0, 462400/3, 0, 38068480/3, 0, 3321472000/3, 0, 302100582400/3))
```

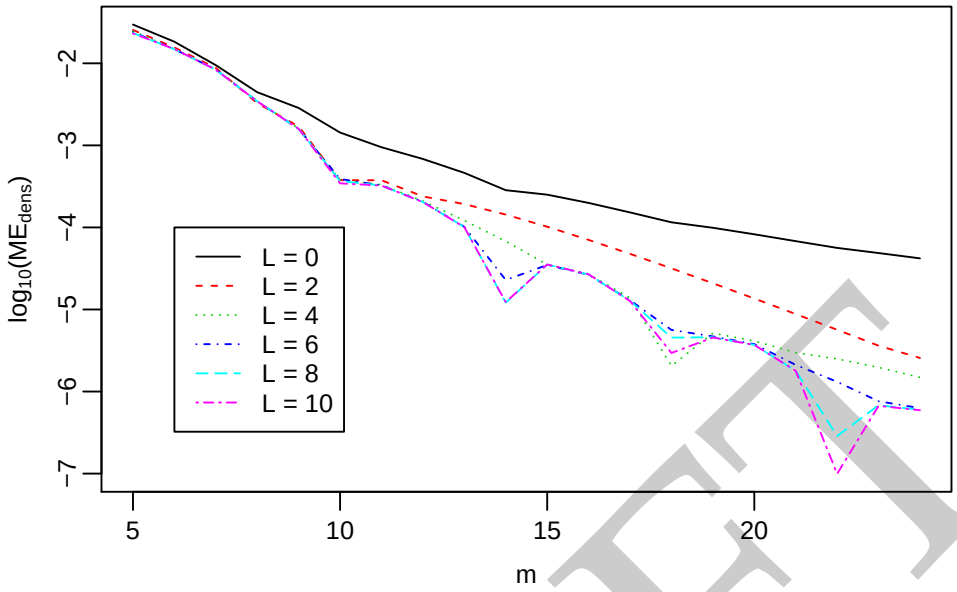

Figure 4.1: The maximum error in estimating the density for Spearman's ρ, as a function of m. The values are $\log_{10}$ of the ME_{dens}; the lines depend on L, the number of terms in the Edgeworth expansion.

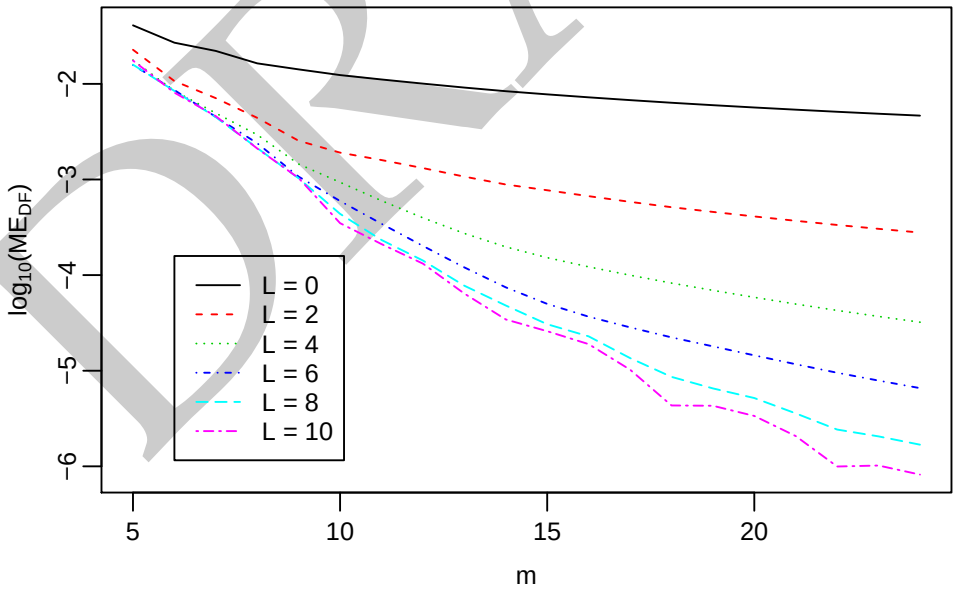

Figure 4.2: The maximum error in estimating the distribution function for Spearman's ρ, as a function of m. The values are $\log_{10}$ of the ME_{DF}; the lines depend on L, the number of terms in the Edgeworth expansion.

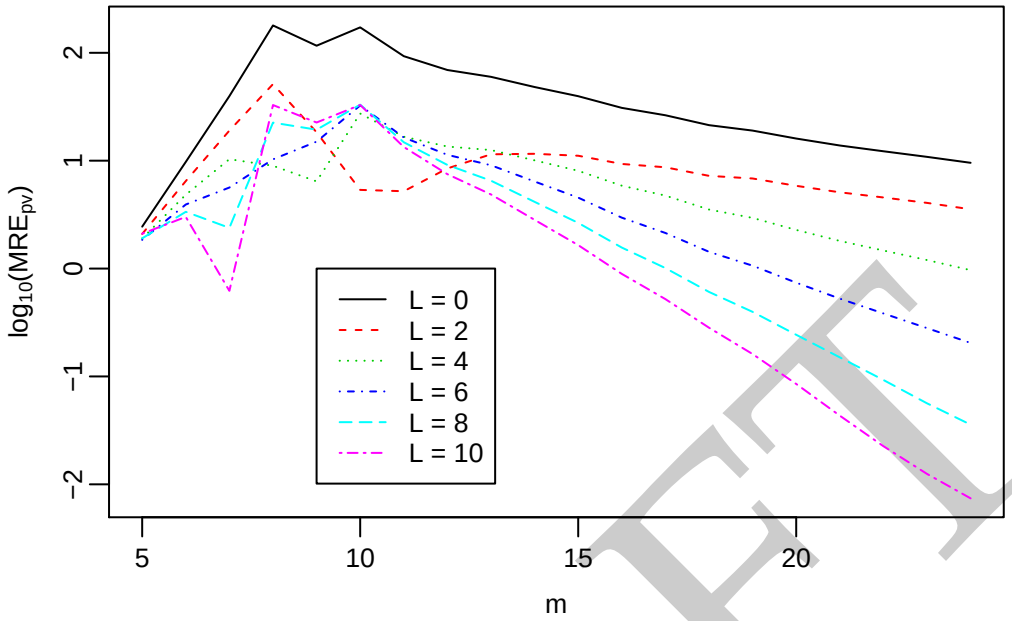

Figure 4.3: The maximum relative error in estimating the p-value (for p-values > 0.00001) for Spearman's ρ, as a function of m. The values are $\log_{10}$ of the MRE_{pv}; the lines depend on L, the number of terms in the Edgeworth expansion.

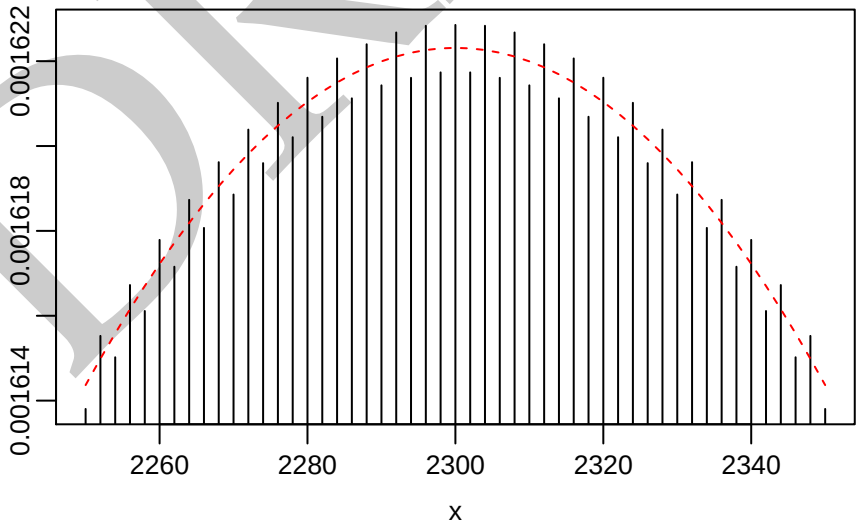

Figure 4.4: The true density $f(x)$ of Spearman's ρ for $m = 24$ and x near the center of the distribution. The smooth line is the estimate from the 10-term Edgeworth expansion.

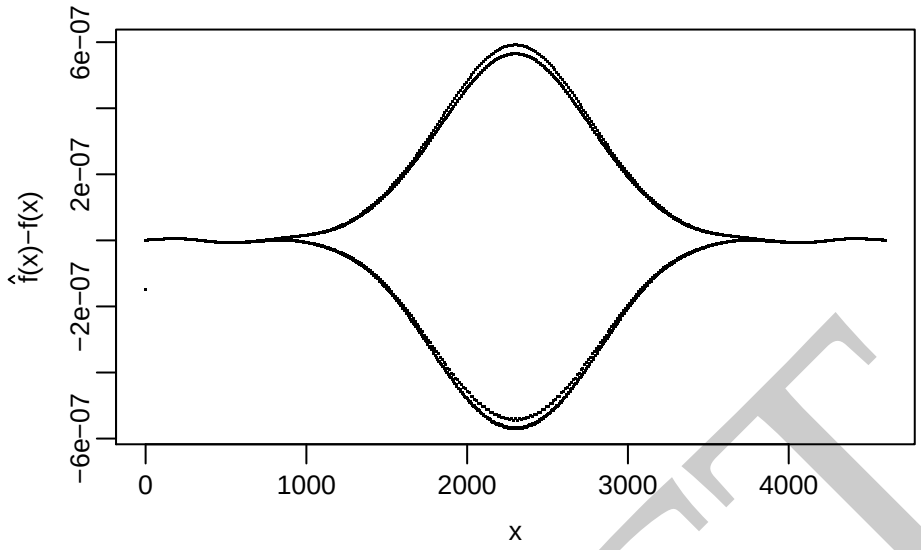

Figure 4.5: The error $\widehat{f}(x) - f(x)$ of Spearman's ρ for $m = 24$ as a function of x, where the estimate is based on the 10-term Edgeworth expansion.

```
  if(m==5) return(c(mu,s2,0,m4, 0, 5400000, 0, 1585227520, 0, 500904320000, 0, 166277056327680))
  m6 <- mfactor*(-28800 + m*(44160 + m*(54280 + m*(-50081 + m*(-22783 +
      m*(23818 + m*(-178 + 49*m*(-89 + 25*m)))))))/3810240
  if(m==6) return(c(mu,s2,0,m4, 0,m6, 0, 386558298581/5, 0, 68673221950725, 0, 323462714247836541/5))
  if(m==7) return(c(mu,s2,0,m4, 0,m6, 0, 30010131730432/15, 0, 12656746257956864/3, 0,
      142711169198502510592/15))
  m8 <- mfactor*(40642560 + m*(-51867648 + m*(-74721024 + m*(110888256 + m*(71871632 +
      m*(-79689881 + m*(-27330110 + m*(30402746 + m*(2142858 + m*(-6275976 +
      m*(1090534 + m*(451718 + 1225*m*(-178 + 25*m))))))))))))/489888000
  if(m==8) return(c(mu,s2,0,m4, 0,m6, 0, m8, 0, 145340586776641536, 0, 24283028012629343404032/35))
  if(m==9) return(c(mu,s2,0,m4, 0,m6, 0, m8, 0, 3235097260221696000, 0, 20849270770389434019840 0/7))
  m10 <- mfactor*(-188116992000 + m*(198866534400 + m*(414838609920 + m*(-433387155456 +
      m*(-404189047296 + m*(442652192448 + m*(251159442832 + m*(-237499131899 +
      m*(-85964555005 + m*(78306584239 + m*(12607819481 + m*(-16568926574 +
      m*(675454398 + m*(1946453438 + m*(-475483358 + 121*m*(-426751 +
      m*(368743 + 1225*m*(-61 + 5*m))))))))))))))))/47421158400
  if(m==10) return(c(mu,s2,0,m4, 0,m6, 0, m8, 0, m10, 0, 5914520439997407834095335/7))
  if(m==11) return(c(mu,s2,0,m4, 0,m6, 0, m8, 0, m10, 0, 60146680213794456424126 0544/35))
  m12 <- mfactor*(19321772487671808000 +m*(-16809105623907041280 + m*(-47949509758034903040 +
      m*(36488300660503068672 + m*(54419976845675882496 + m*(-36488654366312540160 +
      m*(-36118404438893610240 + m*(23517322130866345792 + m*(16323753091762066456 +
      m*(-9960665856120429375 + m*(-4862551541725722430 + m*(2911340308818055007 +
      m*(854178582308599116 + m*(-607856933067242975 + m*(-59540744251778210 +
      m*(85428878340287847 + m*(-7210452777013984 +
      m*(-6425184317282485 + m*(1832998897635870 + 169*m*(77461248253 + 5929*m*(-100938884 +
      25*m*(980379 + 1225*m*(-94 + 5*m))))))))))))))))))))/32129559221760000
  c(mu,s2,0,m4,0,m6,0,m8,0,m10,0,m12)
}
```

```
spearman_cumulants <- function(m) {
    if(m==1) return(rep(0,12))
    if(m==2) return(c(1,1,0,-2, 0, 16, 0, -272, 0, 7936, 0, -353792))
    mfactor <- (m-1)*m^3*(1 + m)^3
    mu <- (m-1)*m*(1 + m)/6
    s2 <- (m-1)*m^2*(1 + m)^2/36
    if(m==3) return(c(mu,s2, 0, -96, 0, 5248, 0, -615936, 0, 124028928, 0, -38150168576))
    k4 <- -mfactor*(-36 + m*(5 + 19*m))/5400
    if(m==4) return(c(mu,s2,0,k4, 0, 2147200/9, 0, -884610560/9, 0, 210287872000/3, 0,
        -229566520729600/3))
    if(m==5) return(c(mu,s2,0,k4, 0, 4320000, 0, -4536916480, 0, 8338196480000, 0, -23573413833400320))
    k6 <- mfactor*(-1800 + 2760*m + 4054*m^2 - 2637*m^3 - 2603*m^4 + 723*m^5 + 583*m^6)/238140
    if(m==6) return(c(mu,s2,0,k4, 0,k6, 0, -501636997456/5, 0, 393383544180480, 0,
        -11925767764861042176/5))
    if(m==7) return(c(mu,s2,0,k4, 0,k6, 0, -6781406220288/5, 0, 30034763619352576/3, 0,
        -5165787362075227783168/45))
    k8 <- -mfactor*(-846720 + m*(1080576 + m*(1616688 + m*(-2358048 + m*(-1800776 +
        m*(1690125 + m*(1012323 + m*(-578442 + m*(-304254 + m*(83709 + 41939*m))))))))))/10206000
    if(m==8) return(c(mu,s2,0,k4, 0,k6, 0, k8, 0, 163576921879461888, 0, -113479202433204782825472/35))
    if(m==9) return(c(mu,s2,0,k4, 0,k6, 0, k8, 0, 1912729128097536000, 0, -6126621752049800417280 0))
    k10 <- mfactor*(-244944000 + m*(258940800 + m*(546557760 + m*(-566728128 + m*(-553076496 +
        m*(587593488 + m*(380118062 + m*(-321580899 + m*(-166918373 + m*(105303339 +
        m*(46553241 + m*(-20933373 + m*(-8319131 + m*(2008773 + 784937*m))))))))))))/61746300
    if(m==10) return(c(mu,s2,0,k4, 0,k6, 0, k8, 0, k10, 0, -5917889802167831516695040/7))
    if(m==11) return(c(mu,s2,0,k4, 0,k6, 0, k8, 0, k10, 0,-45297198502217473065852928/5))
    k12 <- -mfactor*(-12579278963328000 + m*(10943428140564480 + m*(3136925734352064 0 +
        m*(-2377037605784371 2 + m*(-360828483577447 68 + m*(2381161308195648 0 +
        m*(247130910210826 48 + m*(-154338409246524 80 + m*(-119376164206330 52 +
        m*(659434748936128 9 + m*(409715058650945 5 + m*(-192408359073064 4 +
        m*(-97878435062693 2 + m*(39631821049902 2 + m*(16433032610474 6 + m*(-5470229696736 4 +
        m*(-1934786765144 8 + m*(388330607852 9 + 131683559231 1*m))))))))))))))))/20917681785000
    c(mu,s2,0,k4,0,k6,0,k8,0,k10,0,k12)
}

spearman_normalized_cumulants <- function(m) {
    if(m<2) return(rep(0,12))
    fc <- spearman_cumulants(m)
    sigma <- sqrt(fc[2])
    c(0,1,fc[3:12]/sigma^(3:12))
}


spearman_edgeworthf <- function(x,m,L,n=1) {
    mu <- n*(m-1)*m*(1 + m)/6
    sigma <- mu/sqrt(n*(m-1))
    kum <- spearman_normalized_cumulants(m)
    z <- (x-mu)/sigma
    dnorm(z)*edgef(z,L,kum,n)*2/sigma
}

spearman_edgeworthF <- function(x,m,L,n=1) {
```

```
    mu <- n*(m-1)*m*(1 + m)/6
    sigma <- mu/sqrt(n*(m-1))
    kum <- spearman_normalized_cumulants(m)
    z <- (x-mu+1)/sigma
    pnorm(z) - dnorm(z)*edgeF(z,L,kum,n)
}
```

Chapter 5

The footrule

In this section we consider the distribution of the footrule distance:

$$D_{\text{Foot}} \equiv d_{\text{Foot}}(\boldsymbol{Y}, \boldsymbol{\omega}) = \sum_{i=1}^{m} |Y_i - \omega_i|. \tag{5.1}$$

As before, we take $\boldsymbol{\omega} = (1, 2, \ldots, m)$, and suppose $\boldsymbol{Y} \sim \text{Uniform}(\mathcal{P}_m)$. To find the exact null distribution, we could use the splitting method as in Section 10.2 (see Franklin, 1988), but in Section 5.3 we present a clever algorithm developed in Sen and Salama (1983) and Salama and Quade (1990) that allows us to calculate the distribution quickly for m up to 350. Section 5.6 exhibits the Edgeworth expansion approximation to the distribution for larger m.

5.1 First two moments

The first two moments appear to be first presented in Spearman (1904), who attributed the calculation of the variance to Professor Hausdorff. Ury & Kleinecke (1979) note that these formulas were proved in Kleinecke, Ury, & Wagner (1962). Salama & Quade (1990) present the third moment.

For the footrule, we note that $\delta_{\text{Footrule}}(i, j) = |i - j| = 2\max\{i, j\} - i - j$, hence we can write the distance as in (3.3) with

$$\boldsymbol{\Delta}_{\text{Footrule}} = 2\boldsymbol{M} - \mathbf{1}'\boldsymbol{\omega} - \boldsymbol{\omega}'\mathbf{1}, \quad \boldsymbol{M}_{ij} = \max\{i, j\}. \tag{5.2}$$

Next, let $\boldsymbol{K}$ be the $m \times m$ matrix with 1's on the diagonal and above, and 0's elsewhere:

$$\boldsymbol{K} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & 1 & 1 & \cdots & 1 \\ 0 & 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}. \tag{5.3}$$

Then $\boldsymbol{K}\boldsymbol{K}'$ has ij^{th} element $m+1-\max\{i,j\}$, i.e.,

$$\boldsymbol{K}\boldsymbol{K}' = \begin{pmatrix} m & m-1 & m-2 & \cdots & 1 \\ m-1 & m-1 & m-2 & \cdots & 1 \\ m-2 & m-2 & m-2 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix} = (m+1)\mathbf{1}'\mathbf{1} - \boldsymbol{M}. \tag{5.4}$$

We obtain

$$\boldsymbol{\Delta}_{\text{Footrule}} = 2(m+1)\mathbf{1}'\mathbf{1} - \mathbf{1}'\boldsymbol{\omega} - \boldsymbol{\omega}'\mathbf{1} - 2\boldsymbol{K}\boldsymbol{K}'. \tag{5.5}$$

The distance can then be written

$$\begin{aligned} d_{\text{Foot}}(\boldsymbol{y},\boldsymbol{\omega}) &= \operatorname{trace}(\boldsymbol{Q}_{\boldsymbol{y}}\boldsymbol{\Delta}_{\text{Foot}}) \\ &= 2m(m+1) - 2\boldsymbol{\omega}\mathbf{1}' - 2\operatorname{trace}(\boldsymbol{Q}_{\boldsymbol{y}}\boldsymbol{K}\boldsymbol{K}') \\ &= m(m+1) - 2\operatorname{trace}(\boldsymbol{Q}_{\boldsymbol{y}}\boldsymbol{K}\boldsymbol{K}'), \end{aligned} \tag{5.6}$$

since $\boldsymbol{\omega}\mathbf{1}' = \sum_{i=1}^{m} i = m(m+1)/2$.

Before launching into the calculations, we present for reference the sums of the first four powers of $1,\ldots,m$:

n	$\sigma_m^{(n)} \equiv \sum_{i=1}^{m} i^n$
1	$m(m+1)/2$
2	$m(m+1)(2m+1)/6$
3	$m^2(m+1)^2/4$
4	$m(m+1)(2m+1)(3m^2+3m-1)/30$

(5.7)

From (5.5),

$$\boldsymbol{H}\boldsymbol{\Delta}_{\text{Foot}}\boldsymbol{H} = -2\boldsymbol{H}\boldsymbol{K}\boldsymbol{K}'\boldsymbol{H}. \tag{5.8}$$

To find the mean, using the second line of (3.17), we have

$$\begin{aligned} \mathrm{E}[d_{\text{Foot}}(\boldsymbol{Y},\boldsymbol{\omega})] &= -\operatorname{trace}(\boldsymbol{H}\boldsymbol{\Delta}_{\text{Foot}}\boldsymbol{H}) \\ &= 2\operatorname{trace}(\boldsymbol{H}\boldsymbol{K}\boldsymbol{K}'\boldsymbol{H}) = 2\operatorname{trace}(\boldsymbol{H}\boldsymbol{K}\boldsymbol{K}'). \end{aligned} \tag{5.9}$$

Writing out the $\boldsymbol{H}$ yields

$$\boldsymbol{H}\boldsymbol{K}\boldsymbol{K}' = \boldsymbol{K}\boldsymbol{K}' - \frac{1}{m}\mathbf{1}'\mathbf{1}\boldsymbol{K}\boldsymbol{K}'. \tag{5.10}$$

From (5.4) we see that the diagonals of $\boldsymbol{K}\boldsymbol{K}'$ are $m, m-1, \ldots, 1$, and from (5.3) that $\mathbf{1}\boldsymbol{K} = \boldsymbol{\omega}$. Thus

$$\begin{aligned} \mathrm{E}[d_{\text{Foot}}(\boldsymbol{Y},\boldsymbol{\omega})] = 2\operatorname{trace}(\boldsymbol{H}\boldsymbol{K}\boldsymbol{K}') &= 2\sigma_m^{(1)} - 2\frac{1}{m}\sigma_m^{(2)} \\ &= m(m+1) - \frac{1}{3}(m+1)(2m+1) \\ &= \frac{m^2-1}{3}. \end{aligned} \tag{5.11}$$

For the variance, we can similarly use (3.17) to show that

$$\begin{aligned}\mathrm{Var}[d_{\mathrm{Foot}}(\boldsymbol{Y},\boldsymbol{\omega})] &= \frac{1}{m-1}\,\mathrm{trace}((\boldsymbol{H}\boldsymbol{\Delta}_{\mathrm{Foot}}\boldsymbol{H})^2)\\ &= \frac{4}{m-1}\,\mathrm{trace}(\boldsymbol{H}\boldsymbol{K}\boldsymbol{K}'\boldsymbol{H}\boldsymbol{K}\boldsymbol{K}').\end{aligned} \tag{5.12}$$

From (5.10),

$$\begin{aligned}\boldsymbol{H}\boldsymbol{K}\boldsymbol{K}'\boldsymbol{H}\boldsymbol{K}\boldsymbol{K}' = \boldsymbol{K}\boldsymbol{K}'\boldsymbol{K}\boldsymbol{K}' - \frac{1}{m}1'1\boldsymbol{K}\boldsymbol{K}'\boldsymbol{K}\boldsymbol{K}'\\ - \frac{1}{m}\boldsymbol{K}\boldsymbol{K}'1'1\boldsymbol{K}\boldsymbol{K}' + \frac{1}{m^2}1'1\boldsymbol{K}\boldsymbol{K}'1'1\boldsymbol{K}\boldsymbol{K}',\end{aligned} \tag{5.13}$$

hence using $1\boldsymbol{K} = \boldsymbol{\omega}$ again, and with some more manipulations,

$$\mathrm{trace}(\boldsymbol{H}\boldsymbol{K}\boldsymbol{K}'\boldsymbol{H}\boldsymbol{K}\boldsymbol{K}') = \mathrm{trace}(\boldsymbol{K}\boldsymbol{K}'\boldsymbol{K}\boldsymbol{K}') - \frac{2}{m}\boldsymbol{\omega}\boldsymbol{K}'\boldsymbol{K}\boldsymbol{\omega}' + \frac{1}{m^2}(\boldsymbol{\omega}\boldsymbol{\omega}')^2. \tag{5.14}$$

We need to find the terms in the final expression of (5.14). The first term is the sum of squares of the elements in $\boldsymbol{K}\boldsymbol{K}'$ in (5.4). We can see that for each i, row i has i elements equal to $m-i+1$, as does column i. Since we are counting element ii twice, there are $2i-1$ elements equal to $m-i+1$ in the matrix. Thus using (5.7) and some algebra

$$\begin{aligned}\mathrm{trace}(\boldsymbol{K}\boldsymbol{K}'\boldsymbol{K}\boldsymbol{K}') &= \sum_{i=1}^{m}(2i-1)(m-i+1)^2\\ &= \sum_{i=1}^{m}(2m+1-2i)i^2\\ &= (2m+1)\sigma_m^{(2)} - 2\sigma_m^{(3)}\\ &= \frac{1}{6}m(m+1)((2m+1)^2 - 3m(m+1))\\ &= \frac{1}{6}m(m+1)(m^2+m+1).\end{aligned} \tag{5.15}$$

For the second term, note that

$$(\boldsymbol{\omega}\boldsymbol{K}')_i = \sum_{j=i}^{m} j = (m-i+1)(i-1) + \frac{1}{2}(m-i+1)(m-i+2). \tag{5.16}$$

Thus

$$\begin{aligned}
\boldsymbol{\omega K' K \omega'} &= \sum_{i=1}^{m} ((m-i+1)(i-1) + (m-i+1)(m-i+2)/2)^2 \\
&= \sum_{i=1}^{m} (i(m-i) + i(i+1)/2)^2 \\
&= \frac{1}{4} \sum_{i=1}^{m} ((2m+1)i - i^2)^2 \\
&= \frac{1}{4} \sum_{i=1}^{m} ((2m+1)^2 i^2 - 2(2m+1)i^3 + i^4) \\
&= \frac{1}{4} ((2m+1)^2 \sigma_m^{(2)} - 2(2m+1)\sigma_m^{(3)} + \sigma_m^{(4)}) \\
&= \frac{1}{30} m(m+1)(2m+1)(2m^2+2m+1). \qquad (5.17)
\end{aligned}$$

For the third term, we have

$$(\boldsymbol{\omega\omega'})^2 = (\sigma_m^{(2)})^2 = \frac{1}{36} m^2 (m+1)^2 (2m+1)^2. \qquad (5.18)$$

Then from (5.12) and (5.14),

$$\begin{aligned}
\mathrm{Var}[d_{\mathrm{Foot}}(\boldsymbol{Y}, \boldsymbol{\omega})] &= \frac{4}{m-1}(m+1) \left(\frac{1}{6} m(m^2+m+1) - \frac{1}{15}(2m+1)(2m^2+2m+1) \right. \\
&\qquad \left. + \frac{1}{36}(m+1)(2m+1)^2 \right) \\
&= \frac{4}{m-1}(m+1) \frac{1}{180}(2m^3 - 2m^2 + 7m - 7) \\
&= \frac{1}{45}(m+1)(2m^2+7). \qquad (5.19)
\end{aligned}$$

5.2 The Sen-Salama decomposition

Sen & Salama (1983) developed a clever decomposition of the footrule that leads to a fast algorithm for calculating the exact null distribution (see Section 5.3), and aids in finding higher moments.

For $i = 1, \ldots, m$, define

$$T_i = \#\{y_j \leqslant i \,|\, j = 1, \ldots, i\} = \sum_{j=1}^{i} I[y_j \leqslant i], \qquad (5.20)$$

and let S be their sum

$$S = \sum_{i=1}^{m} T_i. \qquad (5.21)$$

(Note that $T_m \equiv m$.) This statistic is equivalent to the footrule statistic D_{Foot} as shown in Theorem 1 of Sen and Salama, which we show now:

Lemma 5.1. *For D_{Foot} in (5.1) and S in (5.21),*

$$D_{Foot} = m(m+1) - 2S. \tag{5.22}$$

Proof. Consider the matrix $\boldsymbol{K}'\boldsymbol{Q}_y'$ for $\boldsymbol{K}$ in (5.3). Its ij^{th} element is $I[y_j \leqslant i]$. Multiplying on the right by $\boldsymbol{K}$ finds the cumulative sums of the rows:

$$(\boldsymbol{K}'\boldsymbol{Q}_y'\boldsymbol{K})_{ij} = \sum_{a=1}^{j} I[y_a \leqslant i]. \tag{5.23}$$

In particular, the diagonals of $\boldsymbol{K}'\boldsymbol{Q}_y'\boldsymbol{K}$ are the T_i. Thus

$$S = \text{trace}(\boldsymbol{K}'\boldsymbol{Q}_y'\boldsymbol{K}) = \text{trace}(\boldsymbol{K}\boldsymbol{K}'\boldsymbol{Q}_y') = \text{trace}(\boldsymbol{Q}_y\boldsymbol{K}\boldsymbol{K}'), \tag{5.24}$$

and (5.22) follows from (5.6). □

Sen & Salama (1983) show that marginally each T_i is hypergeometric, since we are choosing i of the rankings out of m, and counting how many are less than or equal to i. Thus

$$E[T_i] = \frac{i^2}{m} \text{ and } Var[T_i] = \frac{i^2(m-i)^2}{m^2(m-1)}. \tag{5.25}$$

They also find the pairwise joint distribution of the T_i's as well as their covariances, and third moment of S. The latter we show in Section 5.4.1. The covariance matrix of $\boldsymbol{T} = (T_1, \ldots, T_m)$ can be obtained by applying (3.7) and (3.13) to the matrix in (5.23),

$$Cov[\boldsymbol{K}'\boldsymbol{Q}_y'\boldsymbol{K}] = \frac{1}{m-1}\boldsymbol{K}'\boldsymbol{H}\boldsymbol{K} \otimes \boldsymbol{K}'\boldsymbol{H}\boldsymbol{K}. \tag{5.26}$$

Writing out the $\boldsymbol{H}$, we have

$$\begin{aligned} \boldsymbol{K}'\boldsymbol{H}\boldsymbol{K} &= \boldsymbol{K}'\boldsymbol{K} - \frac{1}{m}\boldsymbol{K}'\boldsymbol{1}'\boldsymbol{1}\boldsymbol{K} \\ &= \boldsymbol{K}'\boldsymbol{K} - \frac{1}{m}\boldsymbol{\omega}'\boldsymbol{\omega}. \end{aligned} \tag{5.27}$$

Note that $(\boldsymbol{K}'\boldsymbol{K})_{ij} = \min\{i, j\}$. Thus since $\boldsymbol{T} = \text{diag}(\boldsymbol{K}'\boldsymbol{Q}_y'\boldsymbol{K})$, (3.12) shows that

$$Cov[T_i, T_j] = (\boldsymbol{K}'\boldsymbol{H}\boldsymbol{K})_{ij}^2 = \frac{1}{m-1}\left(\min\{i, j\} - \frac{1}{m}ij\right)^2 = \frac{(m \cdot \min\{i, j\} - ij)^2}{m^2(m-1)}. \tag{5.28}$$

Note that from (5.22), (5.11) and (5.19),

$$E[S] = \frac{1}{6}(m+1)(2m+1) \text{ and } Var[S] = \frac{1}{180}(m+1)(2m^2+7). \tag{5.29}$$

5.3 Exact distribution

The task here is to find the exact distribution of S. Consider the joint distribution of $(T_1, \ldots, T_m)$. Theorem 2 and Lemma 3 of Sen and Salama (1983) show that the T_i's form a Markov chain, and present the conditional distributions of T_i given T_{i-1}, which we present in Lemma 5.2. Finally, we use convolutions to obtain the distribution of S, from which that for F is easily found.

First, some preliminaries. Clearly $0 \leqslant T_i \leqslant i$, but we also need that

$$\#\{y_j \leqslant i \,|\, j = 1, \ldots, i\} + \#\{y_j \leqslant i \,|\, j = i+1, \ldots, m\} = i, \tag{5.30}$$

which implies that $i - T_i \leqslant m - i$, i.e.,

$$\max\{0, m - 2i\} \leqslant T_i \leqslant i. \tag{5.31}$$

Also, $T_1 = I[Y_1 = 1]$, hence is Bernoulli($\frac{1}{m}$).

Lemma 5.2. *For each $i = 1, \ldots, m-1$, if the r_j's and r are such that the conditioning event has positive probability,*

$$P[T_{i+1} = r + k \,|\, T_1 = r_1, \ldots, T_{i-1} = r_{i-1}, T_i = r] = P[T_{i+1} = r + k \,|\, T_i = r]$$
$$= \begin{cases} \frac{(m-2i+r)(m-2i+r-1)}{(m-i)^2} & \text{if } k = 0 \\ \frac{(m-2i+r)(2i-2r+1)}{(m-i)^2} & \text{if } k = 1 \\ \frac{(i-r)^2}{(m-i)^2} & \text{if } k = 2 \\ 0 & \text{otherwise} \end{cases}. \tag{5.32}$$

Proof. Fix $i = 1, \ldots, m-1$, and for each $h = 1, \ldots, i$, let $Y^*_h = Y_h$ if $Y_h \leqslant i$, and $Y^*_h = i+1$ if $Y_h \geqslant i+1$. Then for $j \leqslant i+1$,

$$T_j \equiv \#\{y_h \leqslant j \,|\, h = 1, \ldots, j\} = \#\{y^*_h \leqslant j \,|\, h = 1, \ldots, j\}, \tag{5.33}$$

since in the definition the exact value of y_h is irrelevant if it is more than i. We consider the conditional distribution

$$T_{i+1} - T_i \,|\, Y^*_1 = y^*_1, \ldots, Y^*_i = y^*_i. \tag{5.34}$$

The difference can be written

$$\#\{y_j \leqslant i+1 \,|\, j = 1, \ldots, i+1\} - \#\{y_j \leqslant i \,|\, j = 1, \ldots, i\} = B + C, \tag{5.35}$$

where B and C are the two $0/1$ functions

$$B \equiv I[y_{i+1} \leqslant i+1] \text{ and } C \equiv \#\{y_j = i+1 \,|\, j = 1, \ldots, i\}. \tag{5.36}$$

(The C is in $\{0, 1\}$ since the y_j's are distinct.) Thus $T_{i+1} - T_i$ can take on only the values 0, 1, and 2. We will find the conditional distribution of C, and the conditional distribution of B given C, conditioning on the Y^*_j's. Let A be the conditioning event

$$A = \{Y^*_1 = y^*_1, \ldots, Y^*_i = y^*_i\}. \tag{5.37}$$

Start with C, and condition as in (5.34). Let $r = \#\{y_j^* \leqslant i \,|\, j = 1, \ldots, i\} (= T_i)$. Then there are $i - r$ of the y_j^*'s equal to $i+1$, which means $i-r$ of the y_j's are greater than i. Then if $C = 0$, we have to have all $i-r$ of those greater than $i+1$. Thus (since there are $m-i$ values greater than i and $m-i-1$ greater than $i+1$),

$$P[C = 0 \,|\, A] = \frac{\binom{m-i-1}{i-r}}{\binom{m-i}{i-r}} = \frac{m-2i+r}{m-i}, \tag{5.38}$$

and, subtracting from 1,

$$P[C = 1 \,|\, A] = \frac{i-r}{m-i}. \tag{5.39}$$

Next look at B conditioning on C and A. To find

$$P[B = 0 \,|\, C = 0, A] = P[Y_{i+1} > i+1 \,|\, C = 0, A], \tag{5.40}$$

note than with $C = 0$, $i-r$ of the first i y_j's are greater than $i+1$. Thus there are $(m-i-1) - (i-r)$ choices for Y_{i+1} that exceed $i+1$, out of $m-i$ remaining possibilities. Thus

$$P[B = 0 \,|\, C = 0, A] = \frac{m-2i+r-1}{m-i} \;\Rightarrow\; P[B = 1 \,|\, C = 0, A] = \frac{i-r+1}{m-i}. \tag{5.41}$$

Similarly, if $C = 1$, one of the first y_j's has taken $i+1$, hence Y_{i+1} has one more value over $i+1$ to choose from, and

$$P[B = 0 \,|\, C = 1, A] = \frac{m-2i+r}{m-i} \;\Rightarrow\; P[B = 1 \,|\, C = 1, A] = \frac{i-r}{m-i}. \tag{5.42}$$

Assembling the probabilities, we have that

$$P[T_{i+1} - T_i = 0 \,|\, A] = P[B = 0 \,|\, C = 0, A] P[C = 0 \,|\, A] = \frac{(m-2i+r-1)(m-2i+r)}{(m-i)^2}, \tag{5.43}$$

$$P[T_{i+1} - T_i = 2 \,|\, A] = P[B = 1 \,|\, C = 1, A] P[C = 1 \,|\, A] = \frac{(i-r)^2}{(m-i)^2}, \tag{5.44}$$

and, subtracting those from 1,

$$P[T_{i+1} - T_i = 1 \,|\, A] = \frac{(m-2i+r)(2i-2r+1)}{(i-r)^2}. \tag{5.45}$$

Note that the conditional probabilities in (5.43) through (5.45) are functions of r alone, hence they depend on $(Y_1^*, \ldots, Y_i^*)$ through only T_i. Thus we have the same conditional probabilities when conditioning on T_i. The same can be said about $(T_1, \ldots, T_i)$. That is,

$$\begin{aligned} P[T_{i+1} - T_i = k \,|\, A] &= P[T_{i+1} - T_i = k \,|\, T_1 = r_1, \ldots, T_{i-1} = r_{i-1}, T_i = r] \\ &= P[T_{i+1} - T_i = k \,|\, T_i = r]. \end{aligned} \tag{5.46}$$

Thus (5.43) through (5.46) prove (5.32). □

To obtain the distribution of $S = T_1 + \cdots + T_m$, we use convolutions, but need to carry along both the T_i and the partial sums $S_i = T_1 + \cdots + T_i$'s because of the dependence of the T_i's. We sequentially find the joint distributions of $(T_i, S_i), i = 1, \ldots, m$. Then $S = S_m$, and since $T_m \equiv m$, the marginal distribution of S is essentially the same as the joint of (T_m, S).

Start with (T_1, S_1), where $S_1 = T_1$, hence $P[T_1 = 0, S_1 = 0] = \frac{m-1}{m}$ and $P[T_1 = 1, S_1 = 1] = \frac{1}{m}$. Suppose we have the joint distribution $P[T_i = r, S_i = s]$. Then T_{i+1} may be $k = 0$, 1, or 2 larger than T_i, in which case S_{i+1} will be $T_i + k$ larger than S_i. Thus by the Markov property in Lemma 5.2,

$$P[T_{i+1} = r + k, S_{i+1} = s + r + k \,|\, T_i = r, S_i = s] = P[T_{i+1} = r + k \,|\, T_i = r], \tag{5.47}$$

which we know from (5.32). The i^{th} convolution is then

$$\begin{aligned} P[T_{i+1} = r^*, S_{i+1} = s^*] &= \sum_{k=0}^{2} P[T_{i+1} = r^*, S_{i+1} = s^* \,|\, T_i = r^* - k, S_i = s^* - r^* - k] \\ &\qquad \times P[T_i = r^* - k, S_i = s^* - r^* - k] \\ &= \sum_{k=0}^{2} P[T_{i+1} = r^* \,|\, T_i = r^* - k] P[T_i = r^* - k, S_i = s^* - r^* - k]. \end{aligned} \tag{5.48}$$

We find these values for $i = 2, \ldots, m$. Then at the end we have from (5.22)

$$P[D_{\text{Foot}} = x] = P\left[S_m = \frac{m(m+1) - x}{2}, T_m = m\right], x = 0, 2, \ldots, \left\lfloor \frac{m^2}{2} \right\rfloor. \tag{5.49}$$

5.4 Higher moments

We follow an approach similar to that in Section 4.2 for Spearman's ρ distance. First, note that by (5.22), the central moments of the footrule are related to those of S via

$$E[(D_{\text{Foot}} - E[D_{\text{Foot}}])^n] = (-2)^n E[(S - E[S])^n]. \tag{5.50}$$

We will first find $E[S^n]$, then calculate the central moment from that and the previous moments.

We have $S^n = (\sum T_i)^n$ as in (5.21), and expand the summation similar to that in (4.11), except that because the distribution of the T_i's is not permutation invariant, we decompose the sum into sums over ordered indices. That is, we write

$$E[S^n] = \sum_{\boldsymbol{n} \in \mathcal{IC}_{n,m}} \binom{n}{n_1, \ldots, n_r} \sigma(\boldsymbol{n}). \tag{5.51}$$

where $\mathcal{IC}_{n,m}$ is the set of integer compositions of n with at most m components, and

$$\sigma(\boldsymbol{n}) = \sum_{j_1 < j_2 < \cdots < j_r} \cdots \sum E[T_{j_1}^{n_1} T_{j_2}^{n_2} \cdots T_{j_r}^{n_r}]. \tag{5.52}$$

(Recall that a composition of n is a vector $\boldsymbol{n} = (n_1, \ldots, n_r)$ of positive integers than sum to n.)

Next, for each monomial in (5.52), write the T_i's as sums of indicator functions as in (5.20), and expand into a multiple sum of products of the indicator functions:

$$T_{j_1}^{n_1} \cdots T_{j_r}^{n_r} = \sum_{a_1=1}^{i_1} \cdots \sum_{a_n=1}^{i_n} I[Y_{a_1} \leqslant i_1] \cdots I[Y_{a_n} \leqslant i_n], \tag{5.53}$$

where on the right-hand side, there are equalities among the i_k's determined by the j_k's. Specifically, let $\boldsymbol{j} = (j_1, \ldots, j_r)$, and define the function

$$\boldsymbol{i}(\boldsymbol{n}, \boldsymbol{j}) = (j_1, \ldots, j_1, j_2, \ldots, j_2, \ldots, j_r, \ldots, j_r), \text{ where there are } n_k \text{ of the } j_k\text{'s.} \tag{5.54}$$

Then in (5.53), $(i_1, \ldots, i_n) = \boldsymbol{i}(\boldsymbol{n}, \boldsymbol{j})$.

Now split the overall sum into summations determined by the pattern of the equalities among the indices $\boldsymbol{a} = (a_1, \ldots, a_n)$. We again use the set partitions of $\{1, \ldots, n\}$ as in (4.9) and (4.10). Then

$$T_{j_1}^{n_1} \cdots T_{j_r}^{n_r} = \sum_{\mathcal{K} \in \mathcal{SP}_{n,m}} h(\boldsymbol{n}, \mathcal{K}), \tag{5.55}$$

where

$$h(\boldsymbol{n}, \mathcal{K}) = \sum_{\substack{a_1=1 \\ \boldsymbol{k}(\boldsymbol{a})=\mathcal{K}}}^{i_1} \cdots \sum_{a_n=1}^{i_n} I[Y_{a_1} \leqslant i_1] \cdots I[Y_{a_n} \leqslant i_n]. \tag{5.56}$$

Here, equalities among the i_k's are defined by the $\boldsymbol{n}$, and equalities among the a_k's are defined by the $\mathcal{K}$. With $\mathcal{K} = (\mathcal{K}_1, \ldots, \mathcal{K}_u)$, consider the component $\mathcal{K}_k$, and the set of i's in $\mathcal{K}_k$. Their associated j's are then

$$\{\boldsymbol{i}(\boldsymbol{n}, \boldsymbol{j})_i \mid i \in \mathcal{K}_k\}. \tag{5.57}$$

The upper limit of the index a_i in its summation is $\boldsymbol{i}(\boldsymbol{n}, \boldsymbol{j})_i$. Then since $\boldsymbol{i}(\boldsymbol{n}, \boldsymbol{j})_i$ is nondecreasing in i, the upper limit of all the indices with $i \in \mathcal{K}_k$ must be that of the smallest such i. We can then write

$$h(\boldsymbol{n}, \mathcal{K}, \boldsymbol{j}) = \sum_{\substack{a_{\min(\mathcal{K}_1)}=1 \\ \text{distinct}}}^{\boldsymbol{i}(\boldsymbol{n},\boldsymbol{j})_{\min(\mathcal{K}_1)}} \cdots \sum_{a_{\min(\mathcal{K}_u)}=1}^{\boldsymbol{i}(\boldsymbol{n},\boldsymbol{j})_{\min(\mathcal{K}_u)}} I[Y_{a_1} \leqslant i_1] \cdots I[Y_{a_n} \leqslant i_n]. \tag{5.58}$$

The indices are now distinct because they are from different components of $\mathcal{K}$. The same idea shows that if we group the indicator functions in each summand according to the $\mathcal{K}_k$, we find the summand is

$$\prod_{k=1}^{u} \prod_{i \in \mathcal{K}_k} I[Y_{a_i} \leqslant \boldsymbol{i}(\boldsymbol{n}, \boldsymbol{j})_i] = \prod_{k=1}^{u} I[Y_{a_{\min(\mathcal{K}_k)}} \leqslant \boldsymbol{i}(\boldsymbol{n}, \boldsymbol{j})_{\min(\mathcal{K}_k)}]. \tag{5.59}$$

To clean up the notation a bit, for given $\boldsymbol{n}$ and $\mathcal{K}$, let $j_k^* = \boldsymbol{i}(\boldsymbol{n}, \boldsymbol{j})_{\min(\mathcal{K}_k)}$. Then

$$E[h(\boldsymbol{n}, \mathcal{K}, \boldsymbol{j})] = \sum_{\substack{b_1=1 \\ \text{distinct}}}^{j_1^*} \cdots \sum_{b_u=1}^{j_u^*} E[I[Y_{b_1} \leqslant j_1^*] \cdots I[Y_{b_u} \leqslant j_u^*]]. \tag{5.60}$$

By the exchangeability of the distribution of $\boldsymbol{Y}$, the expectations of the summands are equal. Because the j_k^*'s are in nondecreasing order, the number of summands can be calculated as $j_1^*(j_2^*-1)\cdots(j_u^*-u+1)$. Since the Y_{b_k}'s are also distinct integers, that last value is also the numerator of the expected value of the summand. That is,

$$E[h(\boldsymbol{n},\mathcal{K},\boldsymbol{j})] = \frac{\big(j_1^*(j_2^*-1)\cdots(j_u^*-u+1)\big)^2}{m(m-1)\cdots(m-u+1)}. \tag{5.61}$$

Now we have

$$\begin{aligned} E[S^n] &= \sum_{\boldsymbol{n}\in\mathcal{IC}_{n,m}} \binom{n}{n_1,\ldots,n_r} \sum_{1\leqslant j_1<\cdots<j_r\leqslant m} \cdots \sum_{\mathcal{K}\in\mathcal{SP}_{n,m}} E[h(\boldsymbol{n},\mathcal{K},\boldsymbol{j})] \\ &= \sum_{\boldsymbol{n}\in\mathcal{IC}_{n,m}} \binom{n}{n_1,\ldots,n_r} \sum_{\mathcal{K}\in\mathcal{SP}_{n,m}} \lambda(\boldsymbol{n},\mathcal{K}), \end{aligned} \tag{5.62}$$

where

$$\lambda(\boldsymbol{n},\mathcal{K}) = \sum_{1\leqslant j_1<\cdots<j_r\leqslant m} \cdots \sum \frac{\big(j_1^*(j_2^*-1)\cdots(j_u^*-u+1)\big)^2}{m(m-1)\cdots(m-u+1)}. \tag{5.63}$$

The interchanging of summations in (5.62) is valid because the partitions we sum over are given by m, independently of $\boldsymbol{n}$. The λ's are the computationally most complex part of the calculations. There are a numerous redundancies among them. For example, with $n = 4$, there are 15 possible ordered set partitions $\mathcal{K}$, and 8 combinations, but only 33 different λ's.

5.4.1 Third central moment

Sen & Salama (1983) present the third central moment of the footrule, which they find using their decomposition. Here we present a proof, and similarly find the fourth moment in Section 5.4.2.

Assume that $m \geqslant n$, where $n = 3$, so that the combinations $\boldsymbol{n}$ are $(3), (2,1), (1,2)$, and $(1,1,1)$, and from (5.51),

$$E[S^3] = \sigma(3) + 3\,\sigma(2,1) + 3\,\sigma(1,2) + 6\,\sigma(1,1,1). \tag{5.64}$$

The set partitions $\mathcal{K}$ are $(123), (12,3), (13,2), (1,23)$ and $(1,2,3)$. For each $(\boldsymbol{n}, \mathcal{K})$ pair, we need to find the j^*'s in terms of the j's. In the next table, the entries are $j_1^*, \ldots, j_r^*$.

$\boldsymbol{n}$	(3)	$(2,1)$	$(1,2)$	$(1,1,1)$
$\boldsymbol{j}$	(j_1)	(j_1,j_2)	(j_1,j_2)	(j_1,j_2,j_3)
$\boldsymbol{i}(\boldsymbol{n},\boldsymbol{j})$	(j_1,j_1,j_1)	(j_1,j_1,j_2)	(j_1,j_2,j_2)	(j_1,j_2,j_3)
$\mathcal{K}$				
(123)	j_1	j_1	j_1	j_1
$(12,3)$	j_1,j_1	j_1,j_2	j_1,j_2	j_1,j_3
$(13,2)$	j_1,j_1	j_1,j_1	j_1,j_2	j_1,j_2
$(1,23)$	j_1,j_1	j_1,j_1	j_1,j_2	j_1,j_2
$(1,2,3)$	j_1,j_1,j_1	j_1,j_1,j_2	j_1,j_2,j_2	j_1,j_2,j_3

(5.65)

For example, take the $\boldsymbol{n} = (2,1)$ combination, which correspond to $T_{j_1}^2 T_{j_2} = T_{j_1} T_{j_1} T_{j_2}$, so that $\boldsymbol{i}(\boldsymbol{n},\boldsymbol{j}) = \boldsymbol{i}((2,1),(j_1,j_2)) = (j_1,j_1,j_2)$. For each ordered set partition $\mathcal{K}$, we consider just the minima for the component sets. Thus $\mathcal{K} = (123)$ gives just the index "1," which applied to $\boldsymbol{i}(\boldsymbol{n},\boldsymbol{j})$ yields "j_1." For $\mathcal{K} = (12,3)$ we find the minima $(1,3)$, which in turn picks out the first and third components of $\boldsymbol{i}(\boldsymbol{n},\boldsymbol{j})$, i.e., "$j_1, j_2$."

Now for each column, we sum each entry's (5.61) over $1 \leqslant j_1 < \cdots < j_r \leqslant m$ to find the λ in (5.63), then add up over partitions. For the first column, the middle three entries are equal, so we obtain

$$\sum_{j_1=1}^{m} E[T_{j_1}^3] = \sum_{j_1=1}^{m} \frac{j_1^2}{m} + 3 \sum_{j_1=1}^{m} \frac{(j_1(j_1-1))^2}{m(m-1)} + \sum_{j_1=1}^{m} \frac{(j_1(j_1-1)(j_1-2))^2}{m(m-1)(m-2)}. \tag{5.66}$$

The $(2,1)$ and $(1,2)$ columns both sum over $j_1 < j_2$, but have slightly different sets of summands:

$$\begin{aligned}\sum_{1\leqslant j_1<j_2\leqslant m} \sum E[T_{j_1}^2 T_{j_2}] = \sum_{1\leqslant j_1<j_2\leqslant m} \sum \frac{j_1^2}{m} + \sum_{1\leqslant j_1<j_2\leqslant m} \sum \frac{(j_1(j_2-1))^2}{m(m-1)} \\ + 2 \sum_{1\leqslant j_1<j_2\leqslant m} \sum \frac{(j_1(j_1-1))^2}{m(m-1)} + \sum_{1\leqslant j_1<j_2\leqslant m} \sum \frac{(j_1(j_1-1)(j_2-2))^2}{m(m-1)(m-2)},\end{aligned} \tag{5.67}$$

and

$$\sum_{1\leqslant j_1<j_2\leqslant m} \sum E[T_{j_1} T_{j_2}^2] = \sum_{1\leqslant j_1<j_2\leqslant m} \sum \frac{j_1^2}{m} + 3 \sum_{1\leqslant j_1<j_2\leqslant m} \sum \frac{(j_1(j_2-1))^2}{m(m-1)} + \sum_{1\leqslant j_1<j_2\leqslant m} \sum \frac{(j_1(j_2-1)(j_2-2))^2}{m(m-1)(m-2)}. \tag{5.68}$$

Finally,

$$\begin{aligned}\sum_{1\leqslant j_1<j_2<j_3\leqslant m} \cdots \sum E[T_{j_1} T_{j_2} T_{j_3}] = \sum_{1\leqslant j_1<j_2<j_3\leqslant m} \cdots \sum \frac{j_1^2}{m} + \sum_{1\leqslant j_1<j_2<j_3\leqslant m} \cdots \sum \frac{(j_1(j_3-1))^2}{m(m-1)} \\ + 2 \sum_{1\leqslant j_1<j_2<j_3\leqslant m} \cdots \sum \frac{(j_1(j_2-1))^2}{m(m-1)} + \sum_{1\leqslant j_1<j_2<j_3\leqslant m} \cdots \sum \frac{(j_1(j_2-1)(j_3-2))^2}{m(m-1)(m-2)}.\end{aligned} \tag{5.69}$$

Note that although all columns sum over j_1^2/m in the first row, the summation depends on r, the number of indices summed over.

Here we turn to the computer algebra system again. The various λ's are then easy to find, and we assemble them as in (5.66) to (5.69), then use (5.62) to find that

$$E[S^3] = \frac{(m+1)(m+2)(280m^4 + 504m^3 + 452m^2 + 315m + 153)}{7560}. \tag{5.70}$$

Then the third central moment is by (5.11) and (5.19),

$$\begin{aligned}\mathrm{E}[(S-\mathrm{E}[S])^3] &= \mathrm{E}[S^3] - 3\,\mathrm{E}[S]Var[S] - \mathrm{E}[S]^3 \\ &= \mathrm{E}[S^3] - 3\frac{(m+1)(2m+1)}{6}\frac{(m+1)(2m^2+7)}{180} - \left(\frac{(m+1)(2m+1)}{6}\right)^3 \\ &= \frac{(m+1)(m+2)\left(2m^2+31\right)}{3780}. \end{aligned} \tag{5.71}$$

Thus by (5.65),

$$\mathrm{E}[(D_{\mathrm{Foot}} - \mathrm{E}[(D_{\mathrm{Foot}}])^3] = -\frac{2(m+1)(m+2)\left(2m^2+31\right)}{945}. \tag{5.72}$$

5.4.2 Fourth central moment

We turn to the fourth moment. For (5.51), we need the compositions of $n = 4$, of which there are $2^{n-1} = 8$. Now we assume that $m \geqslant 4$. We find that

$$\begin{aligned}\mathrm{E}[S^4] = \sigma(4) + 4\,\sigma(3,1) + 4\,\sigma(1,3) + 6\,\sigma(2,2) + 12\,\sigma(2,1,1) \\ + 12\,\sigma(1,2,1) + 12\,\sigma(1,1,2) + 24\,\sigma(1,1,1,1). \end{aligned} \tag{5.73}$$

Table 5.1 has the sets of j^* for the pairs of $(\boldsymbol{n}, \mathcal{K})$. After assembling all the calculations analogous to those in (5.66) through (5.69), we find with the computer algebra system that

$$\begin{aligned}\mathrm{E}[S^4] = \frac{1}{226800}(m+1)(2800m^7 + 15680m^6 + 34684m^5 \\ + 45500m^4 + 46315m^3 + 37775m^2 + 22086m + 7920), \end{aligned} \tag{5.74}$$

and using previous moments,

$$\mathrm{E}[(S-\mathrm{E}[S])^4] = \frac{(m+1)(28m^5 + 180m^3 + 160m^2 + 887m + 1265)}{75600}. \tag{5.75}$$

Finally, multiplying by 2^4, we find

$$\begin{aligned}\mathrm{E}[(D_{\mathrm{Foot}} - \mathrm{E}[(D_{\mathrm{Foot}}])^4] &= 16\,\mathrm{E}[(S-\mathrm{E}[S])^4] \\ &= \frac{(m+1)(28m^5 + 180m^3 + 160m^2 + 887m + 1265)}{4725}. \end{aligned} \tag{5.76}$$

5.5 Mathematica code

The key functions here parallel those for Spearman's distance in Section 4.2.1, finding the n^{th} raw and central moments, and the regular and normalized cumulants. Again, n must be a nonnegative integer and m either a positive integer or symbol. Also, if m is a symbol, the formula is valid for $m \geqslant n$.

$\boldsymbol{n}$	(4)	$(3,1)$	$(1,3)$	$(2,2)$
$\boldsymbol{i}(\boldsymbol{n},\boldsymbol{j})$	(j_1,j_1,j_1,j_1)	(j_1,j_1,j_1,j_2)	(j_1,j_2,j_2,j_2)	(j_1,j_1,j_2,j_2)
(1234)	j_1	j_1	j_1	j_1
$(123,4)$	j_1,j_1	j_1,j_2	j_1,j_2	j_1,j_2
$(124,3)$	j_1,j_1	j_1,j_1	j_1,j_2	j_1,j_2
$(134,2)$	j_1,j_1	j_1,j_1	j_1,j_2	j_1,j_1
$(1,234)$	j_1,j_1	j_1,j_1	j_1,j_2	j_1,j_1
$(12,34)$	j_1,j_1	j_1,j_1	j_1,j_2	j_1,j_2
$(13,24)$	j_1,j_1	j_1,j_1	j_1,j_2	j_1,j_1
$(14,23)$	j_1,j_1	j_1,j_1	j_1,j_2	j_1,j_1
$(12,3,4)$	j_1,j_1,j_1	j_1,j_1,j_2	j_1,j_2,j_2	j_1,j_2,j_2
$(13,2,4)$	j_1,j_1,j_1	j_1,j_1,j_2	j_1,j_2,j_2	j_1,j_1,j_2
$(14,2,3)$	j_1,j_1,j_1	j_1,j_1,j_1	j_1,j_2,j_2	j_1,j_1,j_2
$(1,23,4)$	j_1,j_1,j_1	j_1,j_1,j_2	j_1,j_2,j_2	j_1,j_1,j_2
$(1,24,3)$	j_1,j_1,j_1	j_1,j_1,j_1	j_1,j_2,j_2	j_1,j_1,j_2
$(1,2,34)$	j_1,j_1,j_1	j_1,j_1,j_1	j_1,j_2,j_2	j_1,j_1,j_2
$(1,2,3,4)$	j_1,j_1,j_1,j_1	j_1,j_1,j_1,j_2	j_1,j_2,j_2,j_2	j_1,j_1,j_2,j_2

$\boldsymbol{n}$	$(2,1,1)$	$(1,2,1)$	$(1,1,2)$	$(1,2,3,4)$
$\boldsymbol{i}(\boldsymbol{n},\boldsymbol{j}) \to$	(j_1,j_1,j_2,j_3)	(j_1,j_2,j_2,j_3)	(j_1,j_2,j_3,j_3)	(j_1,j_2,j_3,j_4)
(1234)	j_1	j_1	j_1	j_1
$(123,4)$	j_1,j_3	j_1,j_3	j_1,j_3	j_1,j_4
$(124,3)$	j_1,j_2	j_1,j_2	j_1,j_3	j_1,j_3
$(134,2)$	j_1,j_1	j_1,j_2	j_1,j_2	j_1,j_2
$(1,234)$	j_1,j_1	j_1,j_2	j_1,j_2	j_1,j_2
$(12,34)$	j_1,j_2	j_1,j_2	j_1,j_3	j_1,j_3
$(13,24)$	j_1,j_1	j_1,j_2	j_1,j_2	j_1,j_2
$(14,23)$	j_1,j_1	j_1,j_2	j_1,j_2	j_1,j_2
$(12,3,4)$	j_1,j_2,j_3	j_1,j_2,j_3	j_1,j_3,j_3	j_1,j_3,j_4
$(13,2,4)$	j_1,j_1,j_3	j_1,j_2,j_3	j_1,j_2,j_3	j_1,j_2,j_4
$(14,2,3)$	j_1,j_1,j_2	j_1,j_2,j_2	j_1,j_2,j_3	j_1,j_2,j_3
$(1,23,4)$	j_1,j_1,j_3	j_1,j_2,j_3	j_1,j_2,j_3	j_1,j_2,j_4
$(1,24,3)$	j_1,j_1,j_2	j_1,j_2,j_2	j_1,j_2,j_3	j_1,j_2,j_3
$(1,2,34)$	j_1,j_1,j_2	j_1,j_2,j_2	j_1,j_2,j_3	j_1,j_2,j_3
$(1,2,3,4)$	j_1,j_1,j_2,j_3	j_1,j_2,j_2,j_3	j_1,j_2,j_3,j_3	j_1,j_2,j_3,j_4

Table 5.1: The indices for (5.63) to find the fourth moment of S.

There are a number of helper function. The function integerCompositions[n,m] finds the list of integer compositions of n with at most m components, $\mathcal{IC}_{n,m}$. Mathematica doesn't seem to have a built-in such function, as they do have for integer partitions. The function indicesO[r,m] constructs the indices for the multiple summation in (5.63). The (numerators of the) summands are found using smnd[hv], where hv is a vector that indicates which indices j appear, so that hv = 1,1,3,5 represents the product $j_1(j_1-1)(j_3-2)(j_5-3)$. Function lambda[nn,sp,m] calculates $\lambda(\boldsymbol{n},\mathcal{K})$ in (5.63), where nn is the integer combination $\boldsymbol{n}$ and sp is the set partition $\mathcal{K}$. Function footruleSRawMoment[n,m] finds the n^{th} raw moment of S, calling upon the two functions for m being an integer or symbol. The functions setPartitions and denom are the same as for Spearman's distance.

```
Needs["Combinatorica`"]

setPartitions[n_,m_] := Select[Combinatorica`SetPartitions[n],Function[Length[#1]<=m]];
integerCompositions[n_] := Table[Append[p,n]-Prepend[p,0],{p,Subsets[Range[n-1]]}];
integerCompositions[n_,m_] := Select[Table[Append[p, n] - Prepend[p, 0],
            {p, Subsets[Range[n - 1]]}],Function[Length[#1]<=m]];
indicesO[r_, m_] := Module[{inds, inds2},
    inds = Table[i[j], {j, Range[r]}];
    inds2 = Prepend[Drop[inds, -1] + 1, 1];
    Transpose[{inds, inds2, m + Range[-r + 1, 0]}]];
smnd[hv_] := Module[{inds},
    inds = Table[i[j], {j, hv}] - Range[0, Length[hv] - 1];
    Apply[Times, inds]];
lambda[nn_, sp_, m_] := Module[{ns, hv,lambda0},
    lambda0[hv_, r_, m0_] := lambda0[hv,r,m0] = Module[{si, sind},
        si = smnd[hv];
        sind = indicesO[r, m0];
        Sum[si^2, Evaluate[Sequence @@ sind]]];
    ns = Flatten[Table[ConstantArray[i, nn[[i]]], {i, 1, Length[nn]}]];
    hv = ns[[Apply[Min, sp, {1}]]];
    lambda0[hv, Length[nn], m]];
denom[l_, m_] := Pochhammer[m - l + 1, l];
footruleSRawMomentS[n_,m_] :=
    Factor[Sum[Apply[Multinomial, nn]*lambda[nn, sp, m]/denom[Length[sp], m],
        {nn, integerCompositions[n]}, {sp, Combinatorica`SetPartitions[n]}]]
footruleSRawMomentN[n_, m_] := Sum[Apply[Multinomial, nn]*lambda[nn, sp, m]/denom[Length[sp], m],
    {nn, integerCompositions[n,m]}, {sp, setPartitions[n,m]}];
footruleSRawMoment[0, m_] := 1;
footruleSRawMoment[n_, m_] := (
    If[!IntegerQ[n],return[(Print["n must be a nonnegative integer"];Abort[])]];
    If[IntegerQ[m] && m<=n,footruleSRawMomentN[n,m],footruleSRawMomentS[n,m]]);
footruleRawMoment[n_,m_] := Module[{mmp1},
    mmp1 = m*(m+1);
    Factor[Sum[Binomial[n, k]*mmp1^k*(-2)^(n-k)*footruleSRawMoment[n-k,m], {k, 0, n}]]];
footruleCentralMoment[n_, m_] := Module[{mu},
    mu = footruleSRawMoment[1, m];
    (-2)^n*Factor[Sum[Binomial[n, k]*footruleSRawMoment[k, m]*(-mu)^(n - k), {k, 0, n}]]];
footruleCumulant[n_, m_] := cumulant[n, footruleRawMoment, m];
footruleNormalizedCumulant[n_,m_] :=
            If[n==1,0,footruleCumulant[n,m]/footruleCentralMoment[2,m]^(n/2)];
```

5.6 Normal and Edgeworth approximations

For the asymptotic normality of the footrule, we use Theorem 3.3, Hoeffding's theorem with our note that we can use δ in place of δ^* in the numerator (3.22). Here

$$\max\{\delta^2(i,j) \,|\, 1 \leqslant i,j \leqslant m\} = \max\{|i-j|^2 \,|\, 1 \leqslant i,j \leqslant m\} = (m-1)^2. \tag{5.77}$$

From (5.19), the variance is of order $2m^3/45$ as $m \to \infty$. Thus the ratio $(m-1)^2/Var[d_{Footrule}]$ approaches zero, proving that

$$\frac{D_{Foot} - E[D_{Foot}]}{\sqrt{Var[D_{Foot}]}} \longrightarrow N(0,1). \tag{5.78}$$

We next tried the Edgeworth expansion approximations for $L = 1, \ldots, 6$. Generally, those based on the density expansion and distribution function expansion were similar in accuracy, except for $L = 5$ and 6, where the density-based approximations were better. From here on, we will restrict to those estimates. Figures 5.1, 5.2, and 5.3 compare the approximations using criteria in (4.54) and (4.55), i.e., the maximum error in estimating the density, in estimating the distribution function, and the maximum relative error in estimating the p-value for p-values over 0.00001, respectively.

Overall, the approximations are very good. Even for $m = 50$ and $L = 2$, the errors for the three criteria are, respectively, 3.3×10^{-7}, 4.5×10^{-6}, and 0.04. Table 5.79 shows the values for $L = 6$ and select values of m from 10 to 350. Except for the relative error in the p-value for $m < 50$, all the errors are quite small.

	10	25	50	100
Density	0.00145	1.44×10^{-6}	2.71×10^{-8}	7.28×10^{-10}
Distribution function	0.000803	5.07×10^{-6}	3.44×10^{-7}	2.64×10^{-8}
Relative, p-value	0.482	0.156	0.00354	0.000481

	150	200	275	350
Density	9×10^{-11}	2.06×10^{-11}	4.06×10^{-12}	1.19×10^{-12}
Distribution function	6.02×10^{-9}	2.12×10^{-9}	6.74×10^{-10}	2.84×10^{-10}
Relative, p-value	0.000124	4.47×10^{-5}	1.43×10^{-5}	6.05×10^{-6}

(5.79)

5.7 R code

Here we present R functions to calculate the first eight moments and cumulants of the distribution of the footrule, and the functions for the Edgeworth approximations up to $L = 6$. See the parallel section for Spearman's distance, Section 4.5, for more explanation of these routines.

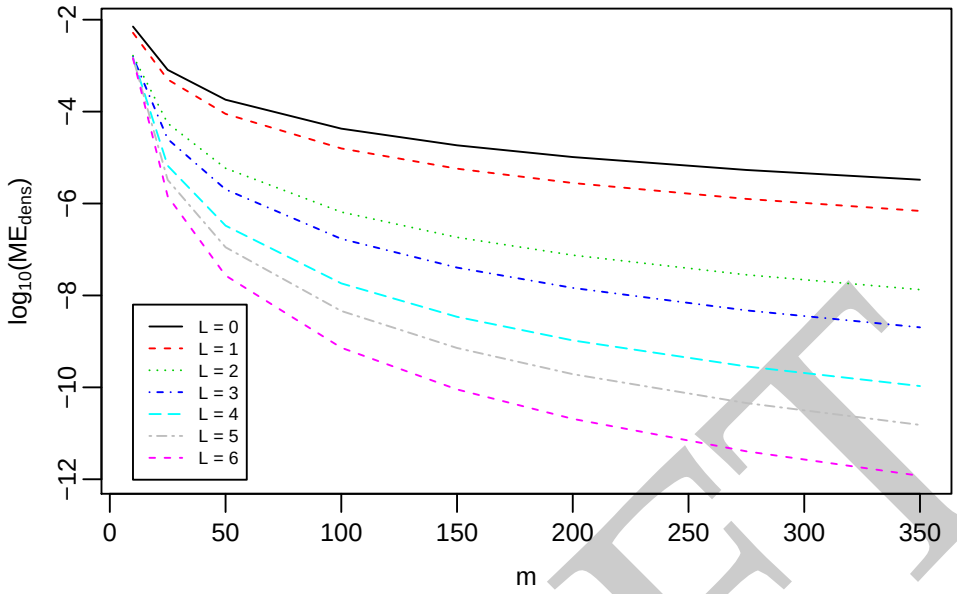

Figure 5.1: The maximum error in estimating the density for the footrule, as a function of m. The values are $\log_{10}$ of the ME_{dens}; the lines depend on L, the number of terms in the Edgeworth expansion.

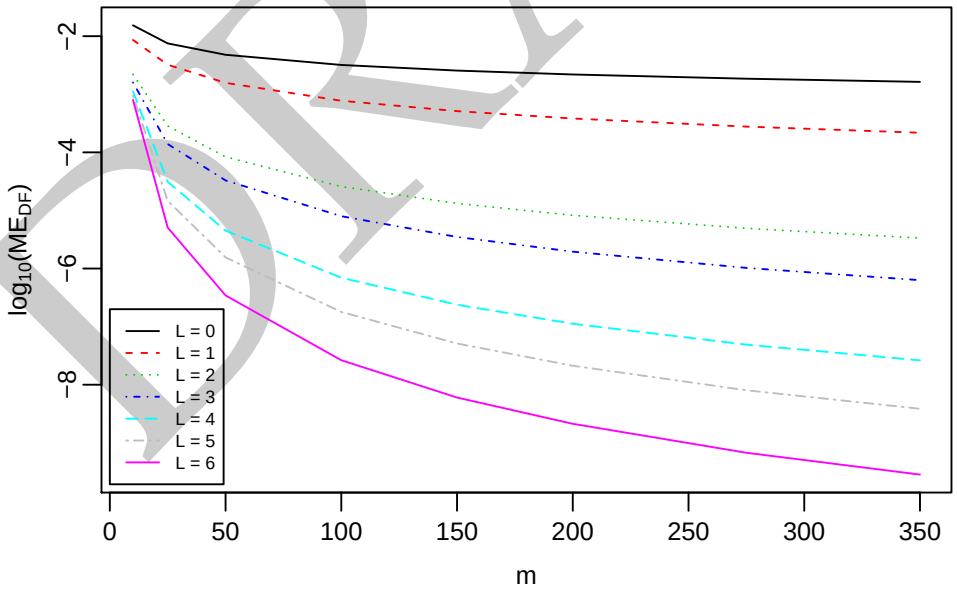

Figure 5.2: The maximum error in estimating the distribution function for the footrule, as a function of m. The values are $\log_{10}$ of the ME_{DF}; the lines depend on L, the number of terms in the Edgeworth expansion.

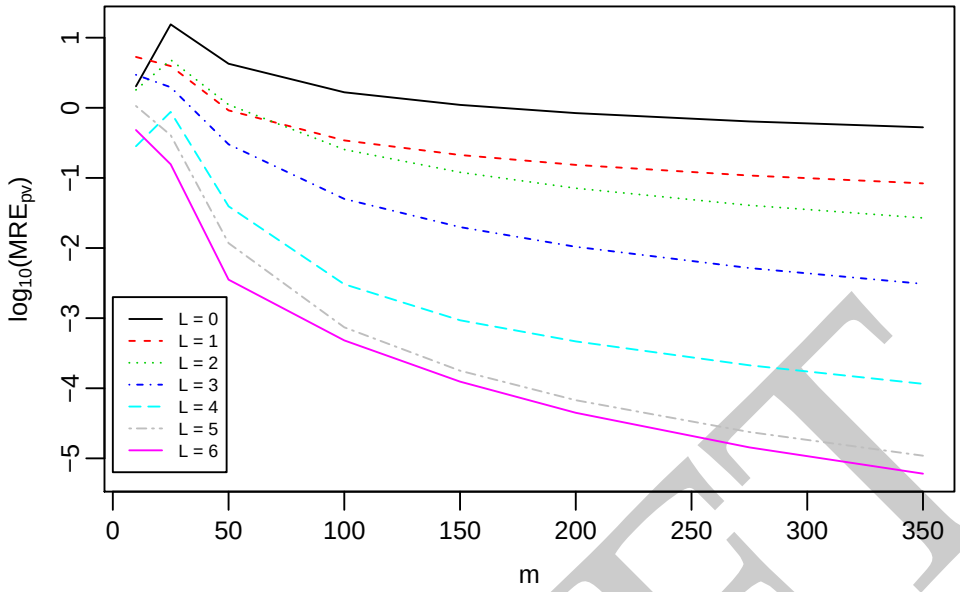

Figure 5.3: The maximum relative error in estimating the p-value (for p-values > 0.00001) for the footrule, as a function of m. The values are $\log_{10}$ of the $\mathrm{MRE}_{\mathrm{pv}}$; the lines depend on L, the number of terms in the Edgeworth expansion.


```
footrule_moments <- function(m) {
    if(m==1) return(rep(0,8))
    if(m==2) return(c(1,1,0,1,0,1,0,1))
    mu <- (m^2-1)/3
    s2 <- (m+1)*(2*m^2+7)/45
    m3 <- -2*(m+1)*(m+2)*(2*m^2+31)/945
    if(m==3) return(c(mu, s2, m3, 272/27, -(4960/243), 45760/729,-(113792/729), 2829056/6561))
    m4 <- (m+1)*(28*m^5+180*m^3+160*m^2+887*m+1265)/4725
    if(m==4) return(c(mu,s2, m3, m4, -120, 2593/3,-3164, 54571/3))
    m5 <- -((4*(1 + m)*(2 + m)*(8555 + m*(3587 + 2*m*(43 + m*(394 + m*(-5 + 22*m))))))/93555)
    if(m==5) return(c(mu,s2, m3, m4, m5, 24288/5,-(128576/5), 213376))
    m6 <- (1 + m)*(368963105 + m*(385870348 + m*(112117257 + 2*m*(16273614 +
              m*(9254091 + 2*m*(545223 + m*(450037 + 286*m*(-127 + 147*m))))))))/127702575
    if(m==6) return(c(mu,s2, m3, m4, m5, m6,-(449092336/3645), 49981557493/32805))
    m7 <- -2*(1 + m)*(2 + m)*(73541545 + m*(46078520 + m*(4890161 + 2*m*(1494444 +
                m*(822501 + 2*m*(-22140 + 7*m*(6113 + 26*m*(-16 + 11*m))))))))/18243225
    if(m==7) return(c(mu,s2, m3, m4, m5, m6,m7, 23910656/3))
    m8 <- (1 + m)*(1690532291725 + m*(2016696623115 + m*(758605059019 + m*(125630091477 +
                8*m*(6671943200 + m*(2137251630 + m*(249602164 + m*(168465567 +
                    2*m*(3955550 + 17*m*(408960 + 143*m*(-361 + 147*m))))))))))/13956067125
    c(mu,s2,m3,m4,m5,m6,m7,m8)
}


footrule_cumulants <- function(m) {
    if(m==1) return(rep(0,8))
    if(m==2) return(c(1, 1, 0, -2, 0, 16, 0, -272))
```

```
    k1 <- (m^2-1)/3
    k2 <- (1 + m)*(7 + 2*m^2)/45
    k3 <- -2*(1 + m)*(2 + m)*(31 + 2*m^2)/945
    if(m==3) return(c(k1,k2,k3, -(128/27), 2080/81, 3200/243, -(151424/243), 1083904/729))
    k4 <- -2*(1 + m)*(-461 + 2*m*(-136 + m*(9 + m*(4 + 7*m))))/4725
    if(m==4) return(c(k1,k2,k3, k4, 160/3, -(1136/9), -(2912/3), 102736/9))
    k5 <- 8*(1 + m)*(2 + m)*(-1028 + m*(-200 + m*(125 + m*(8 + 9*m))))/31185
    if(m==5) return(c(k1,k2,k3, k4,k5, 4512/25, -17248/25, -1229216/125))
    k6 <- 16*(1 + m)*(5867107 + m*(4896338 + m*(411595 + 2*m*(-147077 + m*(5687 +
        m*(4789 + 5673*m)))))/42567525
    if(m==6) return(c(k1,k2,k3,k4,k5,k6, -67053056/6075,-13725231184/91125))
    k7 <- -((1/6081075)*(16*(1 + m)*(2 + m)*(2503157 + m*(1131328 + m*(-122143 +
        m*(-33296 + m*(21407 + 2*m*(1028 + 681*m))))))))
    if(m==7) return(c(k1,k2,k3,k4,k5,k6,k7,-8488960/21))
    k8 <- -(1/1550674125)*(16*(1 + m)*(-10129376957 + 2*m*(-5353166424 + m*(-1413344332 +
        m*(109175700 + m*(25532019 + m*(-15456588 + m*(-312530 + m*(281688 + 350635*m)))))))))
    c(k1,k2,k3,k4,k5,k6,k7,k8)
}

footrule_normalized_cumulants <- function(m) {
    if(m<2) return(rep(0,8))
    fc <- footrule_cumulants(m)
    sigma <- sqrt(fc[2])
    c(0,1,fc[3:8]/sigma^(3:8))
}

footrule_edgeworthf <- function(x,m,L,n=1) {
    mu <- n*(m^2-1)/3
    sigma <- sqrt(n*(1 + m)*(7 + 2*m^2)/45)
    kum <- footrule_normalized_cumulants(m)
    z <- (x-mu)/sigma
    dnorm(z)*edgef(z,L,kum,n)*2/sigma
}

footrule_edgeworthF <- function(x,m,L,n=1) {
    mu <- n*(m^2-1)/3
    sigma <- sqrt(n*(1 + m)*(7 + 2*m^2)/45)
    kum <- footrule_normalized_cumulants(m)
    z <- (x-mu+1)/sigma
    pnorm(z) - dnorm(z)*edgeF(z,L,kum,n)
}
```

Chapter 6

Kendall's distance

A popular distance is **Kendall's distance**, which is the distance upon which Kendall's τ measure of correlation is based. See M. G. Kendall & Gibbons (1990). For $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{P}_m$, Kendall's distance counts the number of **discordant pairs** in the two vectors, a discordant pair being an (x_i, y_i) and (x_j, y_j) for which $x_i < x_j$ but $y_i > y_j$, or *vice versa*. The distance is then given by

$$d_{\text{Ken}}(\boldsymbol{y}, \boldsymbol{x}) = \sum_{1 \leqslant j < i \leqslant m} \sum I[(x_i - x_j)(y_i - y_j) < 0]. \tag{6.1}$$

This distance is label-invariant, hence for the null distribution it is enough to take $\boldsymbol{x} = \boldsymbol{\omega} = (1, \ldots, m)$ and consider the distribution of

$$D_{\text{Ken}} = d_{\text{Ken}}(\boldsymbol{Y}, \boldsymbol{\omega}), \quad \boldsymbol{Y} \sim \text{Uniform}(\mathcal{P}_m). \tag{6.2}$$

The $d_{\text{Ken}}(\boldsymbol{y}, \boldsymbol{\omega})$ can also be given as the number of adjacent interchanges in the $\boldsymbol{y}$ needed to bring $\boldsymbol{y}$ to $\boldsymbol{\omega}$. E.g., if $m = 4$ and $\boldsymbol{y} = (3, 1, 4, 2)$, we can proceed as

$$(3, 1, 4, 2) \to (1, 3, 4, 2) \to (1, 3, 2, 4) \to (1, 2, 3, 4). \tag{6.3}$$

There are three interchanges, so $d_{\text{Ken}}((3, 1, 4, 2), (1, 2, 3, 4)) = 3$.

To find the exact and asymptotic distribution of D_{Ken}, we present a convenient decomposition into independent random variables in the next section, which we then apply in later sections.

6.1 Decomposition

First, we organize the sum according to the y_i's. With $\boldsymbol{x} = \boldsymbol{\omega} = (1, 2, \ldots, m)$, the $x_i - x_j > 0$ in the summation (6.1). Then we can write

$$\begin{aligned} d_{\text{Ken}}(\boldsymbol{Y}, \boldsymbol{\omega}) &= \sum_{1 \leqslant j < i \leqslant m} \sum I[Y_i < Y_j] \\ &= \sum_{i=2}^{m} V_i, \quad \text{where} \quad V_i = \sum_{j=1}^{i-1} I[Y_i < Y_j]. \end{aligned} \tag{6.4}$$

So, for example, V_4 is the number of $\{Y_1, Y_2, Y_3\}$ that exceed Y_4. Note that V_i has space $\{0, \ldots, i-1\}$, for $i = 2, \ldots, m$. Feller (1968, page 256-257) argues that the V_i are independent and uniformly distributed over their spaces. Let

$$\mathcal{V}_m = \{0,1\} \times \{0,1,2\} \times \cdots \times \{0, \ldots, m-1\} \tag{6.5}$$

be the Cartesian product of these spaces. Then the map $\boldsymbol{y} \to \boldsymbol{u} = (v_1, \ldots, v_{m-1})$ is clearly from $\mathcal{P}_m$ into $\mathcal{V}_m$. Also, if $\boldsymbol{y}^{(1)} \neq \boldsymbol{y}^{(2)}$, then the corresponding $\boldsymbol{v}^{(i)}$'s are not equal. (Find the largest index i for which $y_i^{(1)} \neq y_i^{(2)}$. Then since the set of values $\{y_1^{(k)}, \ldots, y_i^{(k)}\}$ is the same for both vectors $k = 1$ and 2, but in different orders, the fact that they differ $y_i^{(k)}$ means they differ in $v_i^{(k)}$.) Since $\#\mathcal{P}_m = \#\mathcal{V}_m = m!$, the map must be onto. Since the distribution of $\boldsymbol{Y}$ is uniform, so must be the induced distribution on $\boldsymbol{V}$. From that fact, and the Cartesian nature of $\mathcal{V}_m$, we have

$$V_2, \ldots, V_m \text{ are independent, } \; V_i \sim \text{Discrete Uniform}(\{0, \ldots, i-1\}). \tag{6.6}$$

6.2 Moments

The moments of Kendall's distance can be found from the moments of the discrete uniforms. The mean and variance of a discrete uniform on $\{0, \ldots, k\}$ are $k/2$ and $k(k+2)/12$, respectively. We present again for reference the sums of the first four powers of $1, \ldots, k$, as in (5.7):

n	$\sigma_k^{(n)} \equiv \sum_{i=1}^k i^n$
1	$k(k+1)/2$
2	$k(k+1)(2k+1)/6$
3	$k^2(k+1)^2/4$
4	$k(k+1)(2k+1)(3k^2+3k-1)/30$

(6.7)

Then, with $i' = i-1$,

$$\begin{aligned} E[D_{Ken}] = \sum_{i=2}^{m} E[V_i] &= \sum_{i=1}^{m-1} \frac{i}{2} \\ &= \frac{\sigma_{m-1}^{(1)}}{2} = \frac{m(m-1)}{4}, \end{aligned} \tag{6.8}$$

and

$$\begin{aligned} Var[D_{Ken}] = \sum_{i=2}^{m} Var[V_i] &= \sum_{i=1}^{m-1} \frac{i(i+2)}{12} \\ &= \frac{\sigma_{m-1}^{(2)} + 2\sigma_{m-1}^{(1)}}{12} = \frac{m(m-1)(2m+5)}{72}. \end{aligned} \tag{6.9}$$

Higher moments can be found similarly, from which we can find cumulants, but Moran (1950) and Silverstone (1950) find the cumulants directly. We will show that the n^{th} cumulant of V_i is

$$\kappa_n^{(i)} = \kappa_n^U (i^n - 1), \tag{6.10}$$

where κ_n^U is the n^{th} cumulant of $U \sim \text{Uniform}(0,1)$. See (2.49) and (2.50). To verify (6.10), note that the cumulant generating function of the $\text{Uniform}(0,a)$ is

$$c_U(t\,;\,a) = \log\left(\frac{e^{at}-1}{at}\right), \tag{6.11}$$

while that of V_i is

$$\begin{aligned} c_V(t) &= \log\left(\frac{1}{k+1}\frac{1-e^{it}}{1-e^{t}}\right) \\ &= \log\left(\frac{e^{it}-1}{it}\right) - \log\left(\frac{e^{t}-1}{t}\right) \\ &= c_U(t\,;\,i) - c_U(t\,;\,1). \end{aligned} \tag{6.12}$$

Thus the n^{th} cumulant of V_i is the n^{th} cumulant of the $\text{Uniform}(0,i)$ minus that of the $\text{Uniform}(0,1)$. Since the $\text{Uniform}(0,i) = i\cdot\text{Uniform}(0,1)$, its n^{th} cumulant is $i^n\kappa_n^U$, hence (6.10) follows from (6.12).

Let κ_n be the n^{th} cumulant of Kendall's distance D_{Ken}. Since D_{Ken} is a sum of the independent V_i's as in (6.4), and the cumulants of V_i are given by (6.10), we have

$$\begin{aligned} \kappa_n = \sum_{i=2}^{m} \kappa_n^{(i)} &= \kappa_n^U \sum_{i=2}^{m} (i^n - 1) \\ &= \kappa_n^U \sum_{i=1}^{m} (i^n - 1) \\ &= \kappa_n^U (\sigma_m^{(n)} - m). \end{aligned} \tag{6.13}$$

The odd cumulants of the uniform are zero except for the first, so the same is true for Kendall's distance. Using (6.7) and (2.50) in (6.13), we have that the fourth cumulant of D_K is

$$\begin{aligned} \kappa_4 &= -\frac{1}{120}\left(\frac{1}{30}m(m+1)(2m+1)(3m^2+3m-1) - m\right) \\ &= -\frac{m(m-1)\left(6m^3+21m^2+31m+31\right)}{3600}. \end{aligned} \tag{6.14}$$

Similarly, using Mathematica, we can find that

$$\begin{aligned} \kappa_6 &= \frac{(m-1)m\left(6m^5+27m^4+48m^3+48m^2+41m+41\right)}{10584}, \quad \text{and} \\ \kappa_8 &= -\frac{(m-1)m\left(10m^7+55m^6+115m^5+115m^4+73m^3+73m^2+93m+93\right)}{21600}. \end{aligned} \tag{6.15}$$

The conversion formulas in Section 2.2 can be used to find various types of moments from the cumulants.

6.2.1 Mathematica code

Equation (6.13) is used in kendallCumulant[n,m] to find the n^{th} cumulant of Kendall's distance. We use the Mathematica function Cumulant to find the Uniform(0,1) cumulants. The regular moments are found using kendallMoment[n,m], which call the function cumulant2moment from Section 2.4.1.

```
kendallCumulant[n_,m_] := Factor[(Sum[k^n,{k,1,m}]−m)∗Cumulant[UniformDistribution[{0,1}],n]];
kendallMoment[n_,m_] := cumulant2moment[n,kendallCumulant,m];
```

6.3 Exact distribution

The exact distribution of Kendall's distance can be found using convolutions of the V_i's. The basic algorithm is to let $W_k = V_2 + \cdots + V_k$, $k = 2, \ldots, m$, so that W_{k+1} is the convolution of W_k and V_{k+1}. Note that W_k has the distribution of Kendall's distance with k objects ranked. Thus the range of W_k is $\{0, \ldots, k(k-1)/2\}$. Let $f_k(w)$ be the density of W_k, and g_i be that of V_i, so that $g_i(u) = 1/i$ for $u = 0, \ldots, i-1$. We start with

$$f_2(w) = g_2(w) = \frac{1}{2}, \quad w = 0, 1, \tag{6.16}$$

then for $k = 2, \ldots, m-1$, we have

$$f_{k+1}(w) = \sum_{u=\max\{0,w-k\}}^{\min\{w,k(k-1)/2\}} f_k(u) g_{k+1}(w-u) = \frac{1}{k+1} \sum_{u=\max\{0,w-k\}}^{\min\{w,k(k-1)/2\}} f_k(u). \tag{6.17}$$

The distribution of W_k is symmetric, $f_k(w) = f_k(k(k-1)/2 - w)$, hence we need to calculate (6.17) for only $w = 0, \ldots, k^* \equiv \text{floor}(k(k+1)/4)$, then fill in the rest. The upper limit in the sum will then be just w, so that this step is

$$\begin{aligned} f_{k+1}(w) &= \frac{1}{k+1} \sum_{u=\max\{0,w-k\}}^{w} f_k(u), \quad w = 0, \ldots, k^*, \\ f_{k+1}(w) &= f_{k+1}\left(\frac{k(k+1)}{2} - w\right), \quad w = k^*+1, \ldots, \frac{k(k+1)}{2}. \end{aligned} \tag{6.18}$$

6.4 Normal and Edgeworth approximations

To prove the asymptotic normality of Kendall's distance, we use the following corollary of the Lindeberg–Feller theorem. See, e.g., Serfling (1980, page 30).

Theorem 6.1. *Suppose* $V_i, i = 1, 2, \ldots$ *are independent, and let* $D_m = \sum_{i=1}^m V_i$. *If for some* $\nu > 2$,

$$\frac{\sum_{i=1}^m E|V_i - E[V_i]|^\nu}{Var[D_m]^{\nu/2}} \to 0 \quad \text{as } m \to \infty, \tag{6.19}$$

then $(D_m - E[D_m])/\sqrt{Var[D_m]} \to N(0,1)$.

For V_i as in (6.6) (where $V_1 \equiv 0$), $|V_i - E[V_i]| < i - 1$, hence

$$\sum_{i=1}^{m} E|V_i - E[V_i]|^\nu = O(m^{\nu+1}). \tag{6.20}$$

Now $D_m = D_{Ken}$, and by (6.9) has variance of order $m^3/36$. Thus the ratio in (6.19) is $O(m^{1-\nu/2})$, which goes to zero for any $\nu > 2$, proving by Theorem 6.1 that

$$\frac{D_{Ken} - E[D_{Ken}]}{\sqrt{Var[D_{Ken}]}} \longrightarrow N(0,1) \text{ as } m \to \infty. \tag{6.21}$$

Figures 6.1 (density), 6.2 (distribution function), and 6.3 (p-values, relative) illustrate the maximum errors of the normal approximation and Edgeworth approximations (up to $L = 10$ terms) for m from 10 to 500. See (4.54) and (4.55) for definitions of these errors. These graphs are based on the Edgeworth expansion of the density, which overall, but especially for larger m, showed smaller errors than those based on the distribution function's expansion.

For $m = 500$, which can be found quickly using the exact algorithm of Section 6.3, the simple normal approximation is very good for the density and distribution function, with maximum errors of about 1×10^{-7} and 1×10^{-4}, respectively. The relative error for the p–values has maximum error about 0.05, which may be a little high. With $L = 2$ terms in the Edgeworth expansion, these three maximum errors are about 1.4×10^{-10}, 1.4×10^{-7} and 7.5×10^{-4}, which are probably sufficient for the usual situations.

A moderate $m = 100$ may need the $L = 4$ term expansion, with maximum errors around 6.5×10^{-10}, 1×10^{-8} and 2×10^{-4}. For small $m = 10$, we need to go to the $L = 10$ term expansion for the relative error of the p-value to achieve a maximum under 0.1, in which case the density and distribution function errors are about 1×10^{-6} and 3×10^{-6}, respectively.

The bottom line is that for moderate to large m, $L = 4$ is fine. For small m the exact algorithm is easy, anyway, so might as well use that.

6.5 R code

Below are the R functions for the moments (up to order 12), cumulants, and normalized cumulants, as well as the Edgeworth expansions of the density and distribution. These expansions take up to $L = 10$ terms. Note that the formulas for the cumulants are much nicer than those for the moments.

```
kendall_moments <- function(m) {
    if(m<2) return(rep(0,12))
    (1/4)*(-1 + m)*m * c(1,
    (2*m+5)/18,
    0,
    (-372 + m*(-997 + m*(-127 + 4*m*(82 + 25*m))))/10800,
    0,
    (118080 + m*(391500 + m*(400733 + m*(-72460 + m*(-230695 +
        2*m*(-21005 + 98*m*(167 + 50*m)))))))/7620480,
```

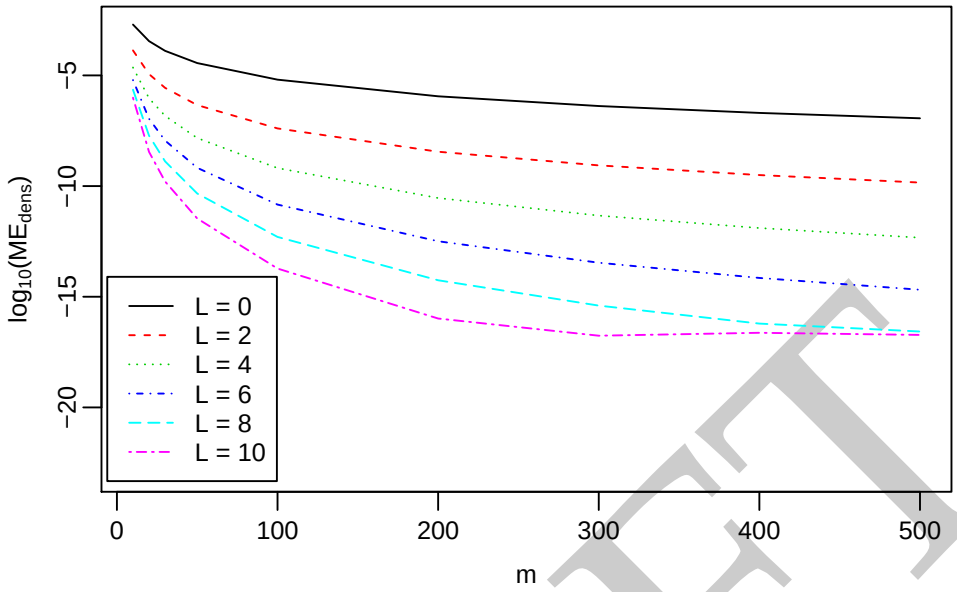

Figure 6.1: The maximum error in estimating the density for Kendall's distance, as a function of m. The values are $\log_{10}$ of the ME_{dens}; the lines depend on L, the number of terms in the Edgeworth expansion.

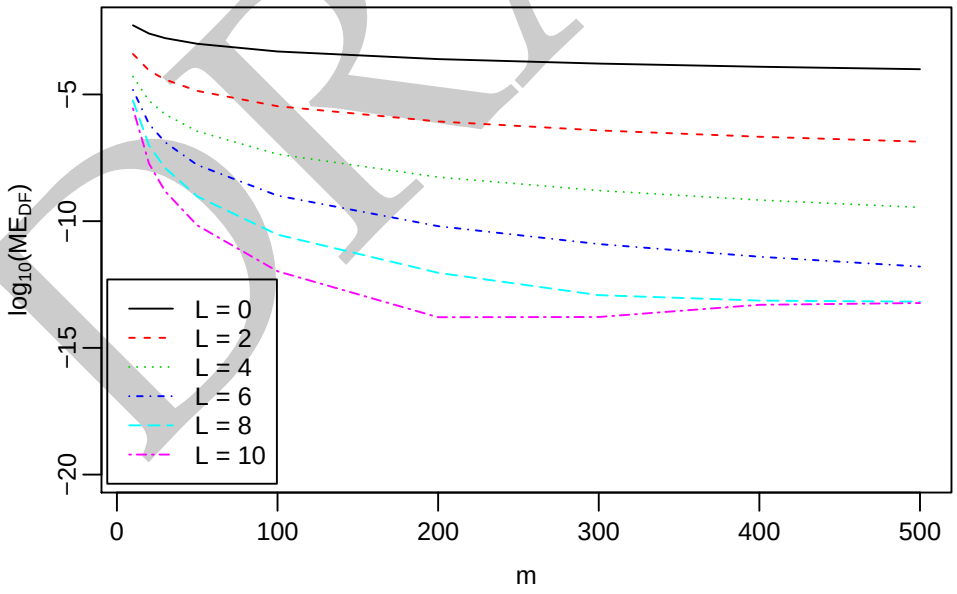

Figure 6.2: The maximum error in estimating the distribution function for Kendall's distance, as a function of m. The values are $\log_{10}$ of the ME_{DF}; the lines depend on L, the number of terms in the Edgeworth expansion.

```
    0,
    −31/1800 + (m∗(−113130288 + m∗(−138803688 + m∗(−49258153 + m∗(59298827 +
        m∗(62414373 + m∗(4246473 + 8∗m∗(−1994064 + m∗(−549519 +
            2450∗m∗(67 + 25∗m)))))))))/1959552000,
    0,
    61/2178 + (m∗(41672603520 + m∗(65336937840 + m∗(36134049576 + m∗(−18039907415 +
        m∗(−30431836250 + m∗(−9671655490 + m∗(3953929208 + m∗(4562897865 + 2∗m∗(553095465 +
                484∗m∗(−664175 + 2∗m∗(−167719 + 1225∗m∗(13 + 10∗m))))))))))))/379369267200,
    0,
    −2363911/22358700 + (m∗(−172335083530314240 + m∗(−197459676044536896 +
            m∗(−150837210953698896 + m∗(−77972420124688716 + m∗(88247642830732759 +
                m∗(135417602903986777 + m∗(11392020884057902 + m∗(−42894902437432958 +
                m∗(−11183279701895213 + m∗(3773773972549669 + 4∗m∗(418337303120486 +
                    169∗m∗(849771419099 + 23716∗m∗(−1327741 +
                                    25∗m∗(−373789 + 4900∗m∗(−4 + 5∗m))))))))))))))/514072947548160000)}

kendall_cumulants <− function(m) {
    if(m<2) return(rep(0,12))
    (1/4)∗(−1 + m)∗m ∗ c(1,
    (2∗m+5)/18,
    0,
    (1/900)∗(−31 − m∗(31 + 3∗m∗(7 + 2∗m))),
    0,
    (41 + m∗(41 + 3∗m∗(16 + m∗(16 + m∗(9 + 2∗m)))))/2646,
    0,
    (−93 − m∗(93 + m∗(73 + m∗(73 + 5∗m∗(23 + m∗(23 + m∗(11 + 2∗m)))))))/5400,
    0,
    (61 + m∗(61 + m∗(94 + m∗(94 + m∗(28 + m∗(2 + m)∗(1 + 2∗m)∗(14 + 3∗m∗(4 + m)))))))/2178,
    0,
    −((1/22358700)∗(691∗(3421 + m∗(3421 + m∗(−1129 + m∗(−1129 + 5∗m∗(1576 + m∗(1576 + 7∗m∗(−20 +
                    m∗(−20 + 3∗m∗(41 + m∗(41 + m∗(15 + 2∗m))))))))))))}

kendall_normalized_cumulants <− function(m) {
    kc <− kendall_cumulants(m)
    c(0,1,kc[−(1:2)]/kc[2]^((1:length(kc))[−(1:2)]/2))}

kendall_edgeworthf <− function(x,m,L,n=1) {
    mu <− n∗(m−1)∗m/4
    sigma <− sqrt(n∗m∗(m−1)∗(2∗m+5)/72)
    kum <− kendall_normalized_cumulants(m)
    z <− (x−mu)/sigma
    dnorm(z)∗edgef(z,L,kum,n)/sigma}

kendall_edgeworthF <− function(x,m,L,n=1) {
    mu <− n∗(m−1)∗m/4
    sigma <− sqrt(n∗m∗(m−1)∗(2∗m+5)/72)
    kum <− kendall_normalized_cumulants(m)
    z <− (x−mu+.5)/sigma
    pnorm(z)−dnorm(z)∗edgeF(z,L,kum,n)}
```

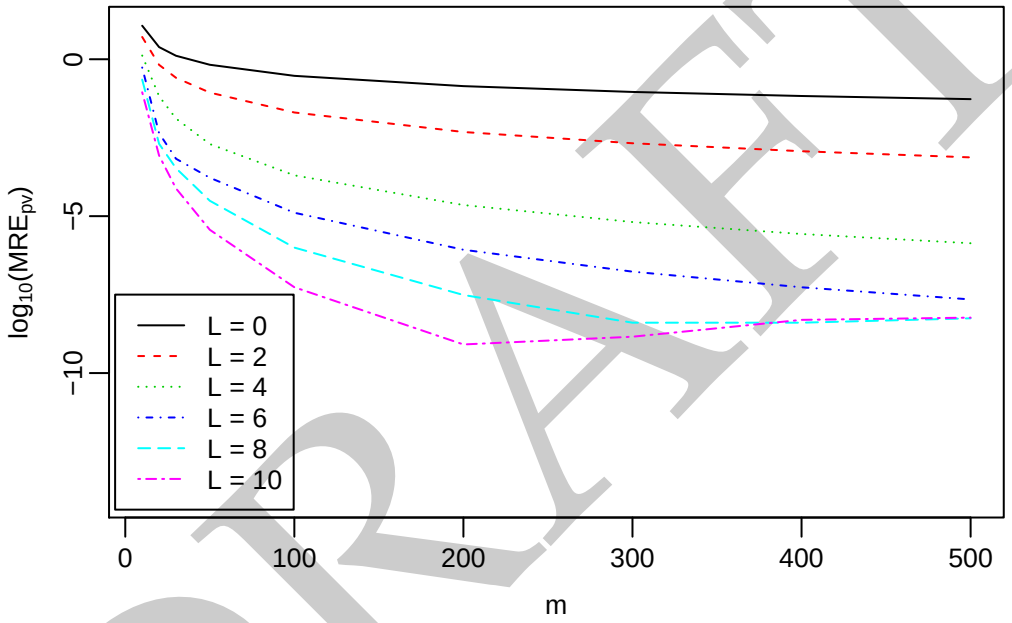

Figure 6.3: The maximum relative error in estimating the p-value (for p-values > 0.00001) for Kendall's distance, as a function of m. The values are $\log_{10}$ of the MRE_{pv}; the lines depend on L, the number of terms in the Edgeworth expansion.

Chapter 7

Hamming distance

7.1 Exact distribution

It is not too hard to find the exact distribution of the Hamming distance, and from that to derive the moments. As a Hoeffding distance, one can also use (3.17) for the first two moments, which is also computationally easy. See Section 7.2. Here we sketch the argument given by Feller (1968), pages 100–106.

For given $\boldsymbol{y} \in \mathcal{P}_m$ and each $i = 1, \ldots, m$, let $\mathcal{A}_i$ be the event that $y_i = i$, and $\mathcal{B}_i$ be its complement, $y_i \neq i$. Then for $k = 0, \ldots, m$,

$$P[d_{\text{Ham}}(\boldsymbol{Y}, \boldsymbol{\omega}) = k] = P[\#\{i \,|\, Y_i = i\} = m - k]. \tag{7.1}$$

There are $\binom{m}{k}$ ways to choose which of the $y_i = i$, and each of those has the same probability. E.g., $P[Y_1 = 1, Y_2 = 2, Y_i \neq i, i = 3, \ldots, m] = P[Y_{m-1} = m-1, Y_m = m, Y_i \neq i, i = 1, \ldots, m-2]$. Letting the last $m - k$ be the ones that match, we then have

$$\begin{aligned} P[d_{\text{Ham}}(\boldsymbol{Y}, \boldsymbol{\omega}) = k] &= \binom{m}{k} P[Y_1 \neq 1, \ldots, Y_k \neq k, Y_{k+1} = k+1, \ldots, Y_m = m] \\ &= \binom{m}{k} P[Y_1 \neq 1, \ldots, Y_k \neq k \,|\, Y_{k+1} = k+1, \ldots, Y_m = m] \\ &\qquad \times P[Y_{k+1} = k+1, \ldots, Y_m = m] \\ &= \frac{1}{(m-k)!} P[Y_1 \neq 1, \ldots, Y_k \neq k \,|\, Y_{k+1} = k+1, \ldots, Y_m = m], \end{aligned} \tag{7.2}$$

the last equation following from

$$P[Y_{k+1} = k+1, \ldots, Y_m = m] = \frac{1}{m} \frac{1}{m-1} \cdots \frac{1}{k+1} = \frac{1}{(m)_k} = \frac{k!}{m!}. \tag{7.3}$$

Consider the conditional probability given in (7.2). Conditionally, the $Y_1, \ldots, Y_k$ all must take values in the range $1, \ldots, k$, and any arrangement is equally likely. Thus $(Y_1, \ldots, Y_k)$ is conditionally Uniform$(\mathcal{P}_k)$, so that is sufficient to find

$$P[Y_1^* \neq 1, \ldots, Y_k^* \neq k] \text{ where } \boldsymbol{Y}^* \sim \text{Uniform}(\mathcal{P}_k). \tag{7.4}$$

We take the complement of the intersection in the probability:

$$P[Y_1^* \neq 1, \ldots, Y_k^* \neq k] = 1 - P[Y_1^* = 1 \text{ or } Y_2^* = 2 \text{ or } \ldots \text{ or } Y_k^* = k]. \tag{7.5}$$

Now we use the union-intersection principle on the probability of the union (Feller, 1968, theorem on page 99):

$$\begin{aligned} P[Y_1^* = 1 \text{ or } Y_2^* = 2 \text{ or } \ldots \text{ or } Y_k^* = k] = &\sum_{i=1}^{k} P[Y_i^* = i] - \sum_{1 \leqslant i < j \leqslant k} P[Y_i^* = i, Y_j^* = j] \\ &+ \sum_{1 \leqslant i < j < l \leqslant k} P[Y_i^* = i, Y_j^* = j, Y_l^* = l] \\ &\pm \cdots + (-1)^{k+1} P[Y_1^* = 1, Y_2^* = 2, \ldots, Y_k^* = k]. \end{aligned} \tag{7.6}$$

As above, each of the summations on the right-hand side is a sum of equal summands, on which can use the formula (7.3), to obtain

$$\sum_{1 \leqslant i_1 < i_2 < \cdots < i_r \leqslant k} P[Y_{i_1}^* = i_1, \ldots, Y_{i_r}^* = i_r] = \binom{k}{r} \frac{1}{(k)_r} = \frac{1}{r!}. \tag{7.7}$$

Then inserting these values into (7.6) and subtracting from 1 we have from (7.5) that

$$P[Y_1^* \neq 1, \ldots, Y_k^* \neq k] = 1 - \sum_{r=1}^{k} \frac{(-1)^{r+1}}{r!} = \sum_{r=0}^{k} \frac{(-1)^r}{r!} \equiv E_k. \tag{7.8}$$

We take $E_0 = 1$. Note that E_k is the sum of the first $k+1$ terms of the expansion of e^{-1}. Also, $E_1 = 0$, which makes sense since it is impossible to have exactly one $Y_i^* \neq i$. Thus from (7.3),

$$P[d_{\text{Ham}}(\boldsymbol{Y}, \boldsymbol{\omega}) = k] = \frac{1}{(m-k)!} E_k, \quad k = 0, \ldots, m. \tag{7.9}$$

7.2 Moments

If interest is mainly in the first two moments, the Hoeffding approach is quite simple. The Hamming distance has $\delta(i, j) = I[i \neq j]$, so that we can write,

$$\boldsymbol{\Delta}_{\text{Ham}} = 1'1 - \boldsymbol{I}, \tag{7.10}$$

and find from (3.9) that

$$\boldsymbol{H} \boldsymbol{\Delta}_{\text{Ham}} \boldsymbol{H} = -\boldsymbol{H}. \tag{7.11}$$

Then by (3.17),

$$\begin{aligned} E[d_{\text{Ham}}(\boldsymbol{Y}, \boldsymbol{\omega})] &= \text{trace}(\boldsymbol{H}) = m - 1 \text{ and} \\ Var[d_{\text{Ham}}(\boldsymbol{Y}, \boldsymbol{\omega})] &= \frac{1}{m-1} \text{trace}(\boldsymbol{H}) = 1. \end{aligned} \tag{7.12}$$

For higher moments, it is easiest to consider factorial moments of the number of matches. That is, let C be the random variable

$$C = m - d_{\text{Ham}}(\boldsymbol{Y}, \boldsymbol{\omega}). \tag{7.13}$$

The density of C is found from (7.9) by switching k and $m-k$:

$$P[C = k] = \frac{1}{k!}\, E_{m-k}, \quad k = 0, \ldots, m. \tag{7.14}$$

The s^{th} factorial moment (see (2.1)) of C for positive integer s is

$$\overline{\gamma}_s = E[(C)_s] = E[C(C-1)\cdots(C-s+1)]. \tag{7.15}$$

Note that $(c)_s = 0$ if $c < s$, hence $(c)_s = 0$ if $m < s$ and c is in the range of C. Thus $\gamma_s = 0$ for $s > m$. For $0 \leqslant s \leqslant m$, we have

$$\begin{aligned} E[(C)_s] = \sum_{c=0}^{m} (c)_s \frac{1}{c!}\, E_{m-c} = \sum_{c=s}^{m} (c)_r \frac{1}{c!}\, E_{m-c} = \sum_{c=s}^{m} \frac{1}{(c-s)!}\, E_{m-c} \quad (\text{set } k = c-s) \\ = \sum_{k=0}^{m-s} \frac{1}{k!}\, E_{(m-s)-k} = 1, \end{aligned} \tag{7.16}$$

the final equation following from the fact that the summation is over the density of C for $m-s$ objects. Thus

$$\overline{\gamma}_s = I[0 \leqslant s \leqslant m]. \tag{7.17}$$

We can use (2.20) to find the raw moments from the factorial moments:

$$\overline{\mu}'_s \equiv E[C^s] = s! \sum_{\boldsymbol{k} \in \mathcal{A}_s} I[\boldsymbol{k}^* \leqslant m] \prod_{l=1}^{s} \frac{1}{k_l!} \left(\frac{1}{l!}\right)^{k_l}. \tag{7.18}$$

See Lemma 2.1 for the definitions of $\mathcal{A}_s$ and $\boldsymbol{k}^*$. Then use (2.10) (where $E[C] = 1$) to obtain the central moments from the raw moments:

$$\overline{\mu}'_n = E[(C-1)^n] = \sum_{s=0}^{n} \binom{n}{s} (-1)^{n-s} \overline{\mu}'_s. \tag{7.19}$$

The central moments for the Hamming distance can then be easily obtained from those of C via

$$\mu_n = (-1)^n \overline{\mu}'_n. \tag{7.20}$$

We note that the Poisson(1) random variable has all factorial moments as well as cumulants equal to 1. (For the Poisson(λ), the factorial moments are $\gamma_n = \lambda^n$ and the cumulants are all λ.) Thus the n^{th} moment (cumulant) of C is the same as the n^{th} moment (cumulant) of the Poisson(1) as long as $n \leqslant m$. The central moments for $n = 2, \ldots, 8$ for the Hamming distance are given next, each assuming $m \geqslant n$:

n	2	3	4	5	6	7	8
μ_n	1	−1	4	−11	41	−162	715

(7.21)

7.3 Asymptotics

The fact that the k^{th} factorial moments of C are the same as those for the Poisson(1) for $k \leqslant m$ suggests that as $m \to \infty$, the C approaches that Poisson. Indeed, it does. From (7.8) we have that E_k is the sum of the first k terms in the expansion of e^{-1}, hence for the density of C in (7.14), for any fixed k, $E_{m-k} \to e^{-1}$ as $m \to \infty$. That is,

$$P[C = k] \longrightarrow e^{-1} \frac{1}{k!}, \tag{7.22}$$

the Poisson(1) density. Thus for the footrule,

$$D_{\text{Ham}} \longrightarrow^{\mathcal{D}} m - \text{Poisson}(1). \tag{7.23}$$

Diaconis (1988, page 117) states that the total variation distance (the maximum absolute difference between the two distributions of the probability of any set) between C and the Poisson(1) is less than $2^m/m!$. For $m = 10$ the value is 0.00028, and for $m = 25$ it is 2.1×10^{-18}. Thus even though the exact distribution is fairly easy to compute for moderate m, the Poisson approximation is essentially exact for $m \geqslant 25$.

Chapter 8

Ulam's distance

8.1 Definition

For a rank vector $\boldsymbol{y} \in \mathcal{P}_m$, an **increasing subsequence** is a subsequence $(y_{i_1}, \ldots, y_{i_k})$ where

$$1 < i_1 < \cdots < i_k \leqslant m \text{ and } y_{i_1} < y_{i_2} < \cdots < y_{i_k}. \tag{8.1}$$

Then a **longest increasing subsequence** is a subsequence with the largest k. There may be more than one subsequence with the largest k. Ulam suggested a distance between $\boldsymbol{\omega} = (1, 2, \ldots, m)$ and $\boldsymbol{y}$ be $m - k$:

$$d_{\text{Ulam}}(\boldsymbol{y}, \boldsymbol{\omega}) = m - \text{Length of longest increasing subsequence of } \boldsymbol{y}. \tag{8.2}$$

Then if $\boldsymbol{y} = \boldsymbol{\omega}$, the distance is zero. The worst is if $\boldsymbol{y} = (m, m-1, \ldots, 1)$ where there are only increasing subsequences of length 1, i.e., the distance is $m - 1$. Some examples for $m = 6$:

$\boldsymbol{y}$	Longest increasing subsequence(s)	Length	$d_{\text{Ulam}}(\boldsymbol{y}, \boldsymbol{\omega})$
123456	123456	6	0
512463	1246	4	2
241635	246, 245, 135	3	3
654321	1, 2, 3, 4, 5, 6	1	5

(8.3)

The distance between $\boldsymbol{y}$ and an arbitrary $\boldsymbol{x} \in \mathcal{P}_m$ is defined as m minus the length of the longest increasing subsequence on which $\boldsymbol{x}$ and $\boldsymbol{y}$ are increasingly related, i.e., $m - L$ for the largest L for which there are indices $i_1, \ldots, i_L$ such that

$$x_{i_1} < x_{i_2} < \cdots < x_{i_L} \text{ and } y_{i_1} < y_{i_2} < \cdots < y_{i_L}. \tag{8.4}$$

Equivalently, reorder the objects such that $\boldsymbol{x} = \boldsymbol{\omega}$, then use (8.2) on the reordered $\boldsymbol{y}$.

The analysis of this distance involves some deep and interesting mathematics, which we review in this chapter. Section 8.2 we present an efficient way to find the longest increasing subsequence. Section 8.3 uses fairly advanced combinatorics to find a way to calculate the null distribution that is much faster than basic enumeration. The asymptotic distribution was only recently discovered, being the Tracy-Widom distribution (Section 8.4), which arises in the asymptotic distribution of eigenvalues in random matrices. The approximation based on this distribution needs very large m to kick in, so in Section 8.4 uses beta approximations based on simulations preformed by Wellner (2002).

8.2 Calculating the distance

Schensted (1961) presents an algorithm that calculates the length of the longest increasing subsequence in one pass through the elements of $\boldsymbol{y}$. Aldous & Diaconis (1999) relate the algorithm to card game they term **patience sorting**, where there are m cards with the numbers 1 to m in the order given by the $\boldsymbol{y}$. ("Patience" is British for games called "solitaire" in the U.S.)

The algorithm sequentially places each y_i into one of a row of m potential cells. We start by placing y_1 in the left-most cell. Suppose after placing y_i, we have the left-most l_i of the cells occupied. Let $s_j^{(i)}$ be the y-value in cell j, $j = 1, \ldots, l_i$. If $y_{i+1} > s_{l_i}^{(i)}$ (the right-most cell), then we place y_{i+1} in the next empty cell, so that we have one more occupied cell, $l_{i+1} = l_i + 1$ and $s_{l_{i+1}}^{(i+1)} = y_{i+1}$. The other cells remain the same, $s_j^{(i+1)} = s_j^{(i)}$, $j = 1, \ldots, l_i$. If $y_{i+1} < s_{l_i}^{(i)}$, then we find the smallest value among the cells that exceeds y_{i+1}, say the value in cell h. Then we place y_{i+1} in cell h, and "bump" the previous value from the cell. The rest of the cells remain the same, i.e., $s_h^{(i+1)} = y_{i+1}$, and $s_j^{(i+1)} = s_j^{(i)}$, $j \neq h$.

When we finish, we have l_m cells filled. The claim is that the length of the longest increasing subsequence is l_m. It is easier to see an illustration, after which we justify the claim. Let $\boldsymbol{y} = (4, 1, 3, 6, 5, 7, 2)$. We proceed through the vector.

$y_1 = 4$,	which we place in the first cell.	$\Rightarrow$	$\boxed{4}$	$(l_1 = 1)$
$y_2 = 1$	is less than $s_1^{(1)} = 4$, hence we replace the 4 with the 1.	$\Rightarrow$	$\boxed{1}$	$(l_2 = 1)$
$y_3 = 3$	is larger than what is there, hence we place it in the next-right cell.	$\Rightarrow$	$\boxed{1}\boxed{3}$	$(l_3 = 2)$
$y_4 = 6$	is again larger than what is in the cells, hence we place it in the third cell.	$\Rightarrow$	$\boxed{1}\boxed{3}\boxed{6}$	$(l_4 = 3)$
$y_5 = 5$	is smaller than just the $s_3^{(4)} = 6$, hence we bump the 6 and replace it with the 5	$\Rightarrow$	$\boxed{1}\boxed{3}\boxed{5}$	$(l_5 = 3)$
$y_6 = 7$	is larger than anything, hence we place it is a new cell.	$\Rightarrow$	$\boxed{1}\boxed{3}\boxed{5}\boxed{7}$	$(l_6 = 4)$
$y_7 = 2$	is smaller than 3, 5, and 7, so the 2 bumps the smallest of those.	$\Rightarrow$	$\boxed{1}\boxed{2}\boxed{5}\boxed{7}$	$(l_7 = 4)$

(8.5)

There are $l_m = 4$ occupied cells. Inspecting $\boldsymbol{y}$, we see that the longest increasing subsequences are 1367 and 1357, which are indeed of length 4. Thus $d_{\text{Ulam}}(\boldsymbol{y}, \boldsymbol{\omega}) = m - l_m = 2$. Note that the sequence of numbers in the cells at the end of the process in (8.5), 1257, is **not** the longest increasing subsequence, being not an increasing subsequence within $\boldsymbol{y}$.

By construction, the numbers in the cells increase as we go from left to right, and over time, for each cell, the numbers do not increase, $s_j^{(i+1)} \leqslant s_j^{(i)}$ (if the cells are occupied).

If we are interested in knowing the best sequence in addition to its length, we add some notes to our above procedure. For each y_i, we note the number directly to its left when first placed in a cell, unless it is placed in the first cell. For the sequence in (8.5) we would have

$$4(\emptyset);\ 1(\emptyset);\ 3(1);\ 6(3);\ 5(3);\ 7(5);\ 2(1). \tag{8.6}$$

Starting with the number in the right-most cell in (8.5), in this case 7, we see that 5 was directly to its left, then 3 was directly to the left of 5, and 1 was directly to the left of 3. There is nothing the left of 1, so we have the sequence from right to left, 7531, or from left to right, 1357. This sequence *is* one of the longest increasing subsequences.

To show that the above algorithm in general finds the length of the longest increasing subsequence, we argue the following two facts. Here, L is the number of occupied cells resulting from the procedure.

1. *There exists an increasing subsequence of length* L. As above, for each y_i, let $y_{j(i)}$ be the value in the cell just to the left of that into which y_i is placed (at the time y_i is placed). Then $j(i) < i$ because the values are placed sequentially by index, and $y_{j(i)} < y_i$, since $y_{j(i)}$ is to the left of y_i among the cells. Let a_L be the index of the value in the L^{th} (right-most) pile, so that $s_L^{(m)} = y_{a_L}$. For $l = L-1, \ldots, 1$, set $a_l = j(a_{l+1})$ as in (8.6). Note that y_{a_L} is in cell L, hence y_{L-1} was initially placed in cell $L-1$, and y_{a_l} was initially placed in cell l, $l = L-2, \ldots, 1$. That is, this sequence is of length L. Then we have that

   $$a_1 < a_2 < \cdots < a_L \ \text{ and } \ y_{a_1} < y_{a_2} < \cdots < y_{a_L}, \tag{8.7}$$

   which is an increasing subsequence of length L.

2. *If an increasing subsequence has length* K, *then* $K \leqslant L$. Suppose $y_{i_1}, \ldots, y_{i_K}$ is an increasing subsequence as in (8.1). During the procedure, y_{i_1} will be placed in a cell, say cell k, $s_k^{(i_1)} = y_{i_1}$. At y_{i_2}'s turn, the k^{th} cell will be no more than y_{i_1} since $s_k^{(i_2)} \leqslant s_k^{(i_1)}$. Thus $y_{i_1} < y_{i_2}$ implies that y_{i_2} will be placed in a cell to the right of k. Continuing, each y_{i_j} will be placed to the right of the previous. Since there are only L occupied cells in the entire patience process, the K distinct cells the y_{i_j}'s are placed in must all be at or to the left of L, hence $K \leqslant L$.

Thus items 1 and 2 show that the maximum length of the set of increasing subsequences is L.

8.3 The exact distribution

Schensted used his algorithm from the previous section to show a correspondence between permutations (rank vectors for us) and pairs of what are known as **Young tableaux**. The numbers of such tableaux are easily counted using a remarkable formula by Frame, Robinson, and Thrall (1954), which can then be used to efficiently determine the exact distribution of the length of the longest subsequence, or Ulam's distance, for m up to about 150. Here we present the basic ideas in Schensted's development.

Schensted continues his algorithm as in (8.5) by taking the bumped values, and using them to fill further rows of cells. That is, if a value is bumped from the initial row, it is inserted into

a second row using the same rules as for the first row. Any value bumped from the second row is placed in a third row, etc. Consider again the example with $\boldsymbol{y} = (4,1,3,6,5,7,2)$. We start as before, with $y_1 = 4$ in the first cell. But when $y_2 = 1$ bumps 4, we now place it in the first cell of the second row, as in table (a) in (8.8). Then $y_3 = 3$ and $y_4 = 6$ go in the first row as before, with no bumping, yielding table (b).

$$(a)\ \begin{array}{|c|}\hline 1\\\hline 4\\\hline\end{array}\quad (b)\ \begin{array}{|c|c|c|}\hline 1&3&6\\\hline 4\\\cline{1-1}\end{array}\quad (c)\ \begin{array}{|c|c|c|}\hline 1&3&5\\\hline 4&6\\\cline{1-2}\end{array}\quad (d)\ \begin{array}{|c|c|c|c|}\hline 1&3&5&7\\\hline 4&6\\\cline{1-2}\end{array}\quad (e)\ \begin{array}{|c|c|c|c|}\hline 1&2&5&7\\\hline 3&6\\\cline{1-2}4\\\cline{1-1}\end{array} \tag{8.8}$$

Next, $y_5 = 5$ means the 6 gets bumped to the second row. Since it is larger than the 4, it is placed in the second cell, as in table (c). The $y_6 = 7$ has no bumping, yielding table (d). Finally, $y_7 = 2$, which bumps the 3 from the first row. Now the 3 bumps the 4 in the second row, which then must start a third row, hence we obtain table (e).

Consider the shapes of the tables in (8.8). In each case, the lengths of the rows are nonincreasing as we go down. Such tables, ignoring the numbers in the boxes, are called **Young diagrams**. (If the boxes are replaced by dots, they are **Ferrers diagrams**.) Such diagrams are graphical representations of partitions of integers. That is, in table (e), the row lengths are $(4,2,1)$, which is a partition of the integer $m = 7$. By construction, the values in each row are increasing as we go from left to right. Notice also that in each column, the values are increasing from top to bottom. We will call such arrangements **monotone Young tableaux** (which is not standard terminology). If as for table (e), a monotone tableau with m cells contains exactly the numbers $1,\dots,m$, then it is a **standard Young tableaux**. (A plain Young tableau is a Young diagram with distinct numbers in the boxes, but in no particular order.) We summarize the first key result due to Schensted. The example in (8.8) may be convincing enough, but we provide a fairly detailed proof.

Lemma 8.1. *For each $\boldsymbol{y} \in \mathcal{P}_m$, Schensted's algorithm yields a standard Young tableau with m cells.*

Proof. Let $\boldsymbol{P}^{(i)}$ be the table at stage i (so that, e.g., table (a) in (8.8) is $\boldsymbol{P}^{(2)}$, table (b) is $\boldsymbol{P}^{(4)}$, and table (e) is $\boldsymbol{P}^{(7)} \equiv \boldsymbol{P}$). $\boldsymbol{P}^{(1)}$ is just a single box with y_1, hence is a monotone Young tableau. Suppose $\boldsymbol{P}^{(i)}$ is a monotone Young tableau, and start the process of placing y_{i+1}.

We first verify that the Schensted algorithm guarantees that $\boldsymbol{P}^{(i+1)}$ is a Young diagram, i.e., the lengths of the rows are nonincreasing as we go down. If two consecutive rows of $\boldsymbol{P}^{(i)}$ are such that the top one of the two is longer than the other, then any number bumped from the top will increase the next row by at most one, hence will still satisfy the monotonicity. If the two have the same length, and a number is bumped from the top one, that number will be less than the number in the box directly below it by the assumption of increasingness down the columns. Thus it will be placed in that box, or one to the left, which in either case will leave the length of that row the same.

Now to make sure the monotonicity is preserved. Going from $\boldsymbol{P}^{(i)}$ to $\boldsymbol{P}^{(i+1)}$ will involve a series of placings and, possibly, bumpings. Each such placing will preserve the row monotonicity, so for the rest of the proof we will focus on showing the column monotonicity. We show that after each placing, the monotonicity holds. First, y_{i+1} is placed in row 1. If y_{i+1} is larger than anything in that row, it is added to the end of that row. Since it will be the only value in its column, the column monotonicity is fine. Everything else stays the same, and we

are finished. Otherwise y_{i+1} bumps the lowest value in the first row that is larger than y_{i+1}, say $z \equiv \boldsymbol{P}^{(i)}_{1,c}$. By the monotonicity of the i^{th} table,

$$\boldsymbol{P}^{(i)}_{1,1} < \cdots < \boldsymbol{P}^{(i)}_{r,c-1} < y_{i+1} < z = \boldsymbol{P}^{(i)}_{1,c} < \cdots < \boldsymbol{P}^{(i)}_{r+1,c_1} \tag{8.9}$$

and

$$y_{i+1} < z = \boldsymbol{P}^{(i)}_{1,c} < \boldsymbol{P}^{(i)}_{2,c} < \cdots < \boldsymbol{P}^{(i)}_{r_c,c}, \tag{8.10}$$

where c_1 is the number of boxes in the first row, and r_c is the number in the c^{th} column. Thus replacing $\boldsymbol{P}^{(i)}_{1,c}$ with y_{i+1} preserves the monotonicity of the table.

Next, z must be placed in the second row. It will go in column c' where $c' \leqslant c$. (Obvious if box c in row 2 is empty; otherwise, $z < \boldsymbol{P}^{(i)}_{2,c}$ by (8.10).) Consider the two possibilities:

1. The z is larger than anything in row 2, so that we append it to the end of row. If $c' = c$, z is in the box directly under y_{i+1}, which is less than z. If $c' < c$, z is directly under $\boldsymbol{P}^{(i)}_{1,c'}$, which is even smaller than y_{i+1} by (8.9). Since there is nothing below z in either case, the column monotonicity holds.

2. We have $\boldsymbol{P}^{(i)}_{2,c'-1} < z < \boldsymbol{P}^{(i)}_{2,c'}$, so that z replaces $w \equiv \boldsymbol{P}^{(i)}_{2,c'}$. By (8.9), $\boldsymbol{P}^{(i)}_{1,c'} < y_{i+1} < z$, and by the column monotonicity of $\boldsymbol{P}^{(i)}$, $z < \boldsymbol{P}^{(i)}_{2,c'} < \boldsymbol{P}^{(i)}_{3,c'}$. Thus column monotonicity holds.

If possibility 2 occurs, then w has been bumped, and must be place in row 3. By similar reasoning, it will either be appended to the row (possibly a previously empty row), thus finishing the process, or will bump another number. We continue until nothing is bumped. The end result is $\boldsymbol{P}^{(i+1)}$, a monotone tableau. Thus by the induction hypothesis, all the tables are monotone. The final one, $\boldsymbol{P}^{(m)}$, in addition contains all the values from $\boldsymbol{y}$, hence is a standard Young tableau. □

To summarize, each $\boldsymbol{y} \in \mathcal{P}_m$ is associated with a standard Young tableau, and the length of the first row in the tableau is equal to the length of the longest increasing subsequence in $\boldsymbol{y}$. Schensted also shows that the length of the first column is the length of the longest decreasing subsequence. (In fact, the tableau for the reversed $\boldsymbol{y}$, $(y_m, \ldots, y_1)$, is the transpose of the tableau for $\boldsymbol{y}$. See his Lemma 7.) However, there may be several $\boldsymbol{y}$'s with the same $\boldsymbol{P}$. To use these tableaux to find the exact distribution of Ulam's distance, we need to find a one-to-one representation, which is achieved by noting when each new box is added to the table during the Schensted algorithm. Specifically, we fill out a matrix $\boldsymbol{Q}$ as we go along, which is also a standard Young tableau, of the same shape as $\boldsymbol{P}$.

At each stage in the algorithm, a new box is added to the table. For the $\boldsymbol{Q}$ matrix, we add a box at the same location, but fill it with the index i of the stage. There is no bumping for

the $\boldsymbol{Q}$ matrices. We illustrate with $\boldsymbol{y} = (4,1,3,6,5,7,2)$ as in (8.5) and (8.8):

i	y_i	$\boldsymbol{P}^{(i)}$	$\boldsymbol{Q}^{(i)}$
1	4	$\begin{array}{\|c\|}\hline 4 \\ \hline \end{array}$	$\begin{array}{\|c\|}\hline 1 \\ \hline \end{array}$
2	1	$\begin{array}{\|c\|}\hline 1 \\ \hline 4 \\ \hline \end{array}$	$\begin{array}{\|c\|}\hline 1 \\ \hline 2 \\ \hline \end{array}$
3	3	$\begin{array}{\|c\|c\|}\hline 1 & 3 \\ \hline 4 \\ \cline{1-1} \end{array}$	$\begin{array}{\|c\|c\|}\hline 1 & 3 \\ \hline 2 \\ \cline{1-1} \end{array}$
4	6	$\begin{array}{\|c\|c\|c\|}\hline 1 & 3 & 6 \\ \hline 4 \\ \cline{1-1} \end{array}$	$\begin{array}{\|c\|c\|c\|}\hline 1 & 3 & 4 \\ \hline 2 \\ \cline{1-1} \end{array}$
5	5	$\begin{array}{\|c\|c\|c\|}\hline 1 & 3 & 5 \\ \hline 4 & 6 \\ \cline{1-2} \end{array}$	$\begin{array}{\|c\|c\|c\|}\hline 1 & 3 & 4 \\ \hline 2 & 5 \\ \cline{1-2} \end{array}$
6	7	$\begin{array}{\|c\|c\|c\|c\|}\hline 1 & 3 & 5 & 7 \\ \hline 4 & 6 \\ \cline{1-2} \end{array}$	$\begin{array}{\|c\|c\|c\|c\|}\hline 1 & 3 & 4 & 6 \\ \hline 2 & 5 \\ \cline{1-2} \end{array}$
7	2	$\begin{array}{\|c\|c\|c\|c\|}\hline 1 & 2 & 5 & 7 \\ \hline 3 & 6 \\ \cline{1-2} 4 \\ \cline{1-1} \end{array}$	$\begin{array}{\|c\|c\|c\|c\|}\hline 1 & 3 & 4 & 6 \\ \hline 2 & 5 \\ \cline{1-2} 7 \\ \cline{1-1} \end{array}$

(8.11)

It is not too hard to see that $\boldsymbol{Q}$ is a standard Young tableau, since each new added box contains a number higher than anything else so far in the table. The interesting fact is that the pair $(\boldsymbol{P}, \boldsymbol{Q})$ uniquely determines $\boldsymbol{y}$, which follows from the next lemma.

Lemma 8.2. *Suppose $\boldsymbol{P}$ and $\boldsymbol{Q}$ are standard Young tableaux of the same shape with m cells. Then there exists $\boldsymbol{y} \in \mathcal{P}_m$ such that the Schensted algorithm applied to $\boldsymbol{y}$ yields $(\boldsymbol{P}, \boldsymbol{Q})$.*

Proof. We reverse the steps in the Schensted algorithm, the bumpers and bumpees trading places. Start with the cell in $\boldsymbol{Q}$ that contains m, say $\boldsymbol{Q}_{r,c}$. Since m is the largest entry, it must be at the end of one of the rows and the end of its column. If it is in the first row, set y_m to be the value at the end of the first row in $\boldsymbol{P}$, $\boldsymbol{P}_{r,c}$. If it is in row $r > 1$, then we know that the number right above it is smaller, $\boldsymbol{P}_{r-1,c} < \boldsymbol{P}_{r,c}$. Find the largest value in row $r-1$ that is smaller than $\boldsymbol{P}_{r,c}$, say $\boldsymbol{P}_{r-1,c'}$, where $c' \geqslant c$. If we bump the value in $(r-1, c')$, replace it with $\boldsymbol{P}_{r,c}$, and remove the (r, c) box, then we still have a monotone Young tableau. Now if $r-1 = 1$, set $y_m = \boldsymbol{P}_{r-1,c'}$, the bumped value. If $r-1 > 1$, then we use the same procedure to bump a value in row $r-2$, replacing it with $\boldsymbol{P}_{r-1.c'}$. We continue bumping until we have bumped a value from row 1, which we assign to y_m.

Call the resulting tableaux $\boldsymbol{P}^{(m-1)}$, and let $\boldsymbol{Q}^{(m-1)}$ be $\boldsymbol{Q}$ with the m-square removed. We apply the same procedure to the $m-1$ box in $\boldsymbol{Q}^{(m-1)}$, eventually finding y_{m-1} as the value bumped from row 1. Continue until we have $(y_1, \ldots, y_m)$, which since $\boldsymbol{P}$ contained the integers 1 to m, is in $\mathcal{P}_m$. □

Thus we have that $\mathcal{P}_m$ and the set of pairs $(\boldsymbol{P}, \boldsymbol{Q})$ as in Lemma 8.2 are in one-to-one correspondence. To determine the length of the longest increasing subsequence in $\boldsymbol{y}$, we need only the length of the corresponding tableau $\boldsymbol{P}$ (or $\boldsymbol{Q}$), which is a function of just the shape of $\boldsymbol{P}$. Recall that the shapes are in one-to-one correspondence with the possible partitions of the integer m. Thus a key to finding the null distribution of the length of the longest increasing subsequence of $\boldsymbol{Y}$ is to count the number of $\boldsymbol{y}$ associated with a particular shape. Fortunately, there is an easily-applied formula, using the **hook length formula**, which counts the number of standard Young tableau of a given shape.

First we define **hooks**. For a Young diagram, the hook for a particular cell is a set consisting of the cell itself, plus the cells directly below and directly to the right of it. The **hook length** for a cell is the number of cells in its hook. The table below shows a diagram with the hook lengths in the cells. So, for example, cell (1,2) has hook = {(1,2), (1,3), (1,4), (1,5), (2,2), (3,2)}, hence its hook length is 6.

8	6	5	2	1
5	3	2		
4	2	1		
1				

(8.12)

The next lemma is due to Frame, de B. Robinson, & Thrall (1954).

Lemma 8.3. *For a given Young diagram with m cells and shape $\boldsymbol{\lambda}$, the number of standard Young tableaux with entries $1, \ldots, m$ of shape $\boldsymbol{\lambda}$ is given by*

$$F(\boldsymbol{\lambda}) = \frac{m!}{\prod_{\text{cells } (i,j)} h_{i,j}}, \quad \text{where} \quad h_{i,j} = \text{hook length of cell } (i,j). \tag{8.13}$$

The proof is in Section 8.3.1. The number of $\boldsymbol{y} \in \mathcal{P}_m$ whose $\boldsymbol{P}$ or $\boldsymbol{Q}$ is of a given shape is then simply the number of such $\boldsymbol{P}$ times the number of such $\boldsymbol{Q}$, i.e., $F(\boldsymbol{\lambda})^2$. Dividing by $m!$ yields the probability of $\boldsymbol{\lambda}$. Then the distribution of the length of the longest increasing subsequence is the marginal of λ_1.

Proposition 8.4. *If $\boldsymbol{Y} \sim$ Uniform$(\mathcal{P}_m)$, for shape $\boldsymbol{\lambda}$ with m cells,*

$$P[\boldsymbol{Y} \text{ has shape } \boldsymbol{\lambda}] = \frac{1}{m!} F(\boldsymbol{\lambda})^2, \tag{8.14}$$

and if $L =$ length of longest increasing subsequence of $\boldsymbol{Y}$, then

$$P[L = l] = \sum_{\boldsymbol{\lambda} \,|\, \lambda_1 = l} \frac{1}{m!} F(\boldsymbol{\lambda})^2. \tag{8.15}$$

Baer & Brock (1968) used this idea to generate the exact distribution up to $m = 36$, and Odlyzko & Rains (2000) pushed it to $m = 120$.

To implement the proposition, we need an efficient method for running through all the partitions of m. Stojmenović & Zoghbi (1998) give an overview of such algorithms. We will use one of theirs called **ZS1**. It is based on an idea common to many proposed algorithms, which they have determined first appeared in Stockmal (1962). The partitions are generated in anti-lexicographical order, which is the numerical analog of reverse alphabetical order. If

$\boldsymbol{\lambda}$ and $\boldsymbol{\delta}$ are both partitions of m, then $\boldsymbol{\lambda} < \boldsymbol{\delta}$ in the anti-lexicographical ordering if for some index j, $\lambda_i = \delta_i$ for $i < j$, and $\lambda_j > \delta_j$. For example, with $m = 6$, the partitions in this order are

$$\begin{aligned}(6) < (5,1) < (4,2) < (4,1,1) < (3,3) < (3,2,1)\\ < (3,1,1,1) < (2,2,2) < (2,2,1,1) < (2,1,1,1,1) < (1,1,1,1,1,1).\end{aligned} \tag{8.16}$$

The basic algorithm starts with $\boldsymbol{\lambda}^{(1)} = (m)$. Then for each $i > 1$, $\boldsymbol{\lambda}^{(i+1)}$ is obtained from $\boldsymbol{\lambda}^{(i)}$ by finding the right-most element greater than 1, decreasing it by 1, summing 1 plus the elements to its right (which are all 1's); then distributing the sum in the most anti-lexicographical way possible while respecting the monotonicity of the partition. For example, suppose $m = 14$ and $\boldsymbol{\lambda}^{(i)} = (6,4,1,1,1,1)$. For the next partition, we decrease the 4 by 1, then add that 1 to the other four 1's. This 5 is then distributed to the left of the 3, hence must be 3, 2: $\boldsymbol{\lambda}^{(i+1)} = (6,3,3,2)$. The algorithm stops when the partition is $(1,1,\ldots,1)$. Symbolically, at step i, if we haven't stopped, let k_i be the length of $\boldsymbol{\lambda}^{(i)}$, and h_i be the index for which $\lambda^{(i)}_{h_i} > 1$ and $\lambda^{(i)}_h = 1$ for $h > h_i$. Then

$$\lambda^{(i+1)}_h = \lambda^{(i)}_h \text{ if } h < h_i \text{ and } \lambda^{(i+1)}_{h_i} = \lambda^{(i)}_{h_i} - 1. \tag{8.17}$$

Let $l_i = 1 + k_i - h_i$, the part of m still to be allotted. The largest the elements to the right of the h_i^{th} can be is $\lambda^{(i+1)}_{h_i}$, hence we do integer division to find the integers d_i and $r_i \in \{0,\ldots,\lambda^{(i+1)}_{h_i} - 1\}$ such that

$$l_i = d_i \lambda^{(i+1)}_{h_i} + r_i. \tag{8.18}$$

Then we set

$$\lambda^{(i+1)}_{h_i+1} = \cdots = \lambda^{(i+1)}_{h_i+d_i} = \lambda^{(i+1)}_{h_i} \text{ and if } r_i > 0,\ \ \lambda^{(i+1)}_{h_i+d_i+1} = r_i. \tag{8.19}$$

We now have $k_{i+1} = h_i + d_i + I[r_i > 0]$.

Notice that if $\lambda^{(i)}_{h_i} = 2$, then the algorithm can skip the adding and dividing steps. After (8.17), we just add a 1 to the end, so that $k_{i+1} = k_i + 1$ and $\lambda^{(i+1)}_h = 1$ for $h_i < h \leqslant k_{i+1}$. Stojmenović & Zoghbi exploit this simplification, showing that having 2 be the target value occurs very frequently: In 66% of partitions for $m = 30$ and 78% for $m = 90$, the percentage asymptotically approaching 100%. They show empirically for $m = 75$ that this modification speeds up the algorithm by a factor of four.

We use the SZ1 algorithm to run through the partitions, then find the hook lengths and the product for each partition generated, to calculate (8.15). There may be further efficiencies by modifying the hook length product along with updating the partitions.

8.3.1 Proof of hook length formula

There have been a number of proofs of the hook length formula since first discovered by Frame, de B. Robinson, & Thrall (1954). Here we present a fun probabilistic one by Greene, Nijenhuis, & Wilf (1979).

The proof uses induction on m. It is clear that the hook length formula is valid for the $m = 1$ case. Assume it is valid for any tableau with $m - 1$ cells. Consider the diagram with m cells and shape $\boldsymbol{\lambda} = (\lambda_1,\ldots,\lambda_r)$. Any standard Young tableau with values $1,\ldots,m$

will have the largest value m in one of the corner cells, say at the end of row r^*. If we remove that cell, then we have a standard Young tableau containing the values $1, \ldots, m-1$ with shape $\boldsymbol{\lambda}_{r^*}$ the same as the original, except for row r^* being one cell shorter: $\boldsymbol{\lambda}_{r^*} = (\lambda_1, \ldots, \lambda_{r^*-1}, \lambda_{r^*} - 1, \lambda_{r^*+1}, \ldots, \lambda_r)$. (For illustrations, consider any of the tables in (8.11), and remove the cell with the largest value. What remains is a standard tableau.) Thus the number of standard tableaux of shape $\boldsymbol{\lambda}$ with m in the corner of row r^* is the same as the number of standard tableaux with shape $\boldsymbol{\lambda}_{r^*}$, which by the induction hypothesis is $F(\boldsymbol{\lambda}_{r^*})$ in (8.13). Then the number of standard tableau corresponding to the shape $\boldsymbol{\lambda}$ is the sum of such values, adding over the possible corners in which the m falls. The induction step is to show that this number is $F(\boldsymbol{\lambda})$:

$$F(\boldsymbol{\lambda}) = \sum_{r^* \,|\, \text{row } r^* \text{ in } \boldsymbol{\lambda} \text{ contains a corner cell}} F(\boldsymbol{\lambda}_{r^*}). \tag{8.20}$$

Greene, Nijenhuis, & Wilf approach the problem by defining a random walk over the cells of the diagram which always ends in one of the corners, say (R, C). They show that

$$p(r^*, c^*) \equiv P[\text{Random walk ends with } (R, C) = (r^*, c^*)] = \frac{F(\boldsymbol{\lambda}_{r^*})}{F(\boldsymbol{\lambda})} \quad \text{if } (r^*, c^*) \text{ is a corner cell.} \tag{8.21}$$

Because the probabilities sum to 1, (8.21) implies (8.20).

The random walk proceeds as follows:

- Step 1: Start in cell $(r_1, c_1) \in \boldsymbol{D}$ with probability $\frac{1}{m}$. If (r_1, c_1) is a corner cell, then stop and set $(R, C) = (r_1, c_1)$. Otherwise, go to the next step.
- Step $i+1$: Let H_{r_i, c_i} be the hook for (r_i, c_i). Since we have not stopped, (r_i, c_i) must not be a corner cell, hence there must be at least two cells in the hook. Randomly choose a cell from $H_{r_i, c_i} - \{(r_i, c_i)\}$, each with probability $1/(h_{r_i, c_i} - 1)$. Call it (r_{i+1}, c_{i+1}). If it is a corner cell, then stop and set $(R, C) = (r_{i+1}, c_{i+1})$. Otherwise, repeat this step with $i \to i+1$.

Note that at each step, we increase either the row index or the column index by at least 1. Thus there can be only a finite number of steps. We need to verify (8.21). Consider the ratio $F(\boldsymbol{\lambda}_{r^*})/F(\boldsymbol{\lambda})$ for (r^*, c^*) a corner cell. Since it is a corner, the only hook lengths that will be different in $\boldsymbol{\lambda}_{r^*}$ than in $\boldsymbol{\lambda}$ will be those in row r^* or column c^*, each of which will have their hook lengths decrease by 1. In (8.22), we see removing the corner $(r^*, c^*) = (3, 3)$ decreases the hook lengths in column 3 and row 3 by 1.

$$\boldsymbol{\lambda}: \begin{array}{|c|c|c|c|c|} \hline 8 & 6 & 5 & 2 & 1 \\ \hline 5 & 3 & 2 \\ \cline{1-3} 4 & 2 & 1 \\ \cline{1-3} 1 \\ \cline{1-1} \end{array} \longrightarrow \boldsymbol{\lambda}_{r^*}: \begin{array}{|c|c|c|c|c|} \hline 8 & 6 & 4 & 2 & 1 \\ \hline 5 & 3 & 1 \\ \cline{1-3} 3 & 1 \\ \cline{1-2} 1 \\ \cline{1-1} \end{array} \tag{8.22}$$

Note also that the hook lengths of the corners is always 1. Thus from (8.13),

$$\frac{F(\boldsymbol{\lambda}_{r^*})}{F(\boldsymbol{\lambda})} = \frac{1}{m} \prod_{r=1}^{r^*-1} \frac{h_{r,c^*}}{h_{r,c^*} - 1} \prod_{c=1}^{c^*-1} \frac{h_{r^*,c}}{h_{r^*,c} - 1}. \tag{8.23}$$

To show (8.21) for given corner (r^*, c^*), we need to sum up the probabilities of all the paths that lead to that corner. We will sort the paths into groups depending on the rows and columns they visit. A path $\boldsymbol{p}$ starts at cell (r_1, c_1), say, and proceeds through (possibly) several cells until arriving at the corner:

$$\boldsymbol{p} = (r_1, c_1) \to (r_2, c_2) \to \cdots \to (r_t, c_t) = (r^*, c^*). \tag{8.24}$$

Since each step in the random walk is either to the right or down, $r_i \leqslant r^*$ and $c_i \leqslant c^*$. Also, one of the row index or column index stays the same, and the other increases, for consecutive cells in the path:

$$\text{For each } i = 1, \ldots, t-1, \text{ either } r_{i+1} = r_i \;\&\; c_{i+1} > c_i \;\text{ or }\; r_{i+1} > r_i \;\&\; c_{i+1} = c_i. \tag{8.25}$$

For any such path $\boldsymbol{p}$, let $(\text{row}(\boldsymbol{p}), \text{col}(\boldsymbol{p}))$ be the row and column **projections**, i.e., the sets of indices of the rows and columns the path hits:

$$\text{row}(\boldsymbol{p}) = \{r_1, \ldots, r_t\} \;\&\; \text{col}(\boldsymbol{p}) = \{c_1, \ldots, c_t\}. \tag{8.26}$$

The sets do not contain multiple elements of the same value, so that if the path is $(1,3) \to (1,5) \to (4,5) \to (6,5)$, we would have $(\text{row}(\boldsymbol{p}), \text{col}(\boldsymbol{p})) = (\{1,4,6\}, \{3,5\})$.

Now let $(\mathcal{R}, \mathcal{C})$ be any pair of subsets such that

$$r^* \in \mathcal{R} \subset \{1, \ldots, r^*\} \text{ and } c^* \in \mathcal{C} \subset \{1, \ldots, c^*\}. \tag{8.27}$$

(There will be at least one path that has $(\mathcal{R}, \mathcal{C})$ as its projections for such pairs.) Denote $r_1 = \min(\mathcal{R})$ and $c_1 = \min(\mathcal{C})$, so that (r_1, c_1) is the starting point of any of its paths. Lemma 3 in Greene, Nijenhuis, & Wilf (1979) shows that if $\boldsymbol{P} = (\boldsymbol{P}_1 \to \ldots \to \boldsymbol{P}_t)$ is a random path,

$$\begin{aligned} P[(\text{row}(\boldsymbol{P}), \text{col}(\boldsymbol{P})) = (\mathcal{R}, \mathcal{C}) \,|\, \boldsymbol{P}_1 = (r_1, c_1)] &= \prod_{r \in \mathcal{R} - \{r^*\}} \frac{1}{h_{r,c^*} - 1} \prod_{c \in \mathcal{C} - \{c^*\}} \frac{1}{h_{r^*,c} - 1} \\ &\equiv q(\mathcal{R}, \mathcal{C}), \end{aligned} \tag{8.28}$$

where the product over the empty set is 1. (It is interesting that given the starting point, the rows the path hit are independent of the columns hit.) We again use induction, now on the path length t. Note that for given projections $(\mathcal{R}, \mathcal{C})$, the corresponding paths have length $t = \#\mathcal{R} + \#\mathcal{C} - 1$. If $t = 1$, then the path must be (r^*, c^*) itself, hence the probability we start on that corner is $\frac{1}{m}$, which checks.

Next, assume (8.28) holds for paths of length less than t. The second step in any such path must be to either (r_1, c_2) where $c_2 \in \min(\mathcal{C} - \{c_1\})$, or to (r_2, c_1), where $r_2 = \min(\mathcal{R} - \{r_1\})$, either of which has transition probability $1/(h_{r_1,c_1} - 1)$. Subsequent steps would then have a path with projection either $(\mathcal{R}, \mathcal{C} - \{c_1\})$ or $(\mathcal{R} - \{r_1\}, \mathcal{C})$, which has length $t - 1$. Thus by the Markov property of the random walk, and the induction hypothesis on q,

$$\begin{aligned} P[(\text{row}(\boldsymbol{P}), \text{col}(\boldsymbol{P})) = (\mathcal{R}, \mathcal{C}) \,|\, \boldsymbol{P}_1 = (r_1, c_1)] &= P[\boldsymbol{P}_2 = (r_1, c_2) \,|\, \boldsymbol{P}_1 = (r_1, c_1)] q(\mathcal{R}, \mathcal{C} - \{c_1\}) \\ &\quad + P[\boldsymbol{P}_2 = (r_2, c_1) \,|\, \boldsymbol{P}_1 = (r_1, c_1)] q(\mathcal{R} - \{r_1\}, \mathcal{C}) \\ &= \frac{1}{h_{r_1,c_1} - 1} (q(\mathcal{R}, \mathcal{C} - \{c_1\}) + q(\mathcal{R} - \{r_1\}, \mathcal{C})). \end{aligned} \tag{8.29}$$

By (8.28),

$$q(\mathcal{R}, \mathcal{C} - \{c_1\}) = (h_{r^*,c_1} - 1)q(\mathcal{R}, \mathcal{C}) \text{ and } q(\mathcal{R} - \{r_1\}, \mathcal{C})) = (h_{r_1,c^*} - 1)q(\mathcal{R}, \mathcal{C}), \tag{8.30}$$

hence

$$P[(\text{row}(\boldsymbol{P}), \text{col}(\boldsymbol{P})) = (\mathcal{R}, \mathcal{C}) \,|\, \boldsymbol{P}_1 = (r_1, c_1)] = \frac{h_{r^*,c_1} - 1 + h_{r_1,c^*} - 1}{h_{r_1,c_1} - 1} q(\mathcal{R}, \mathcal{C}). \tag{8.31}$$

It is not hard to see that $h_{r_1,c_1} = h_{r^*,c_1} + h_{r_1,c^*} - 1$ by looking at an example. In (8.32) take $(r_1, c_1) = (2, 1)$ and the corner $(r^*, c^*) = (5, 3)$.

a	a	ab	ab	ab	ab	
a		b				
a		b				
ac	c	bc				
ac						
ac						

(8.32)

The cells in the hook for $A = (r_1, c_1)$ are labelled "a," and those for cells $B = (r_1, c^*)$ and $C = (r^*, c_1)$ are labelled "b" and "c," respectively. Counting, we see that $\#a = \#b + \#c - 1$, as desired. The formal proof would proceed by noting the cells at (r_1, c^*) and to the left are in the hooks for A and B; the cells at (r^*, c_1) and below are in the hooks for A and C; the number of cells between columns $c_1 + 1$ and $c^* - 1$ are the same for A and C; and the number of cells between rows $r_1 + 1$ and $r^* - 1$ are the same for A and B. Then only A has (r_1, c_1) in its hook, and both B and C have (r^*, c^*) in their hooks. Thus we count one extra cell in the sum of the B and C hooks.

Since we now have that the factor in (8.31) is 1, (8.28) holds.

Back to (8.21). The total probability that the random walk ends up in corner (r^*, c^*) is then found by summing over the possible starting points and projections. Letting $(\mathcal{R}, \mathcal{C})$ run over all pairs of sets as in (8.27), and setting $(r_1, c_1) = (\min(\mathcal{R}), \min(\mathcal{C}))$, we have

$$\begin{aligned}
p(r^*, c^*) &= \sum_{(\mathcal{R},\mathcal{C})} P[(\text{row}(\boldsymbol{P}), \text{col}(\boldsymbol{P})) = (\mathcal{R}, \mathcal{C}) \,|\, \boldsymbol{P}_1 = (r_1, c_1)] P[\boldsymbol{P}_1 = (r_1, c_1)] \\
&= \frac{1}{m} \sum_{(\mathcal{R},\mathcal{C})} q(\mathcal{R}, \mathcal{C}) \\
&= \frac{1}{m} \sum_{(\mathcal{R},\mathcal{C})} \left(\prod_{r \in \mathcal{R} - \{r^*\}} \frac{1}{h_{r,c^*} - 1} \prod_{c \in \mathcal{C} - \{c^*\}} \frac{1}{h_{r^*,c} - 1} \right) \quad \text{(by (8.28))} \\
&= \frac{1}{m} \sum_{\mathcal{R}} \left(\prod_{r \in \mathcal{R} - \{r^*\}} \frac{1}{h_{r,c^*} - 1} \right) \times \sum_{\mathcal{C}} \left(\prod_{c \in \mathcal{C} - \{c^*\}} \frac{1}{h_{r^*,c} - 1} \right).
\end{aligned} \tag{8.33}$$

The first summation is the sum over all subsets of $\{1, \ldots, r^* - 1\}$ (including the empty set) of the product of the $1/(h_{r,c^*} - 1)$'s for r in each subset, which is the expansion of the product of

all $(1+1/(h_{r,c^*}-1))$'s. Similarly for the second summation. Thus

$$p(r^*,c^*) = \frac{1}{m}\prod_{r=1}^{r^*-1}\left(1+\frac{1}{h_{r,c^*}-1}\right)\prod_{c=1}^{c^*-1}\left(1+\frac{1}{h_{r^*,c}-1}\right) = \frac{F(\boldsymbol{\lambda}_{r^*})}{F(\boldsymbol{\lambda})}, \tag{8.34}$$

by (8.23), proving (8.21), hence the hook length formula, Lemma 8.3.

8.4 Approximations and asymptotics

The asymptotic distribution of Ulam's distance, or actually of L_m, the length of the longest increasing subsequence, has an interesting and surprisingly recent history. Baik, Deift, & Johansson (1999) show that as $m \to \infty$,

$$\frac{L_m - 2\sqrt{m}}{m^{1/6}} \longrightarrow Z, \tag{8.35}$$

where Z has the **Tracy-Widom distribution** (Tracy & Widom, 1994) with parameter $\beta = 2$. The definition of this distribution is rather involved, so we refer the reader to the Tracy-Widom article or, more briefly, to the Wikipedia article (Wikipedia contributors, 2018b). Figure 8.1 exhibits the density, calculated using the function dtw from the R package RMTstat (Johnstone, Ma, Perry, & Shahram, 2014). It looks fairly symmetric, but is actually slightly skewed. The first four summary statistics have been found by Bornemann (2010), Table 4:

Mean	Variance	Skewness	Kurtosis
-1.7710868074	0.8131947928	0.2240842036	0.0934480876

(8.36)

The skewness is the normalized third central moment, and the kurtosis is the normalized fourth central moment minus 3. The Tracy-Widom distributions arose originally from studying the asymptotic distributions of the largest eigenvalue of certain random matrices. The $\beta = 2$ indicates the matrices are random complex normal hermitian matrices.

Rather than trying to present a proof of (8.35), we will present a brief history of the result. For more discussion, insights, and references, see Aldous & Diaconis (1999), Wellner (2002), and Romik (2015), this last being a thorough and comprehensive book on the mathematical background and results surrounding the study of the distribution of L_m.

Stanislaw Ulam, the famous physicist and mathematician, among myriad other achievements, pioneered the idea of using Monte Carlo simulations to estimate intractable distributions of random variable. See Metropolis & Ulam (1949). Ulam (1961) noted a Monte Carlo experiment (by an E. Neighbor) to study the distribution of the lengths of longest monotone (increasing or decreasing) subsequences in permutations. Baer & Brock (1968) used Monte Carlo to conjecture that

$$\frac{E[L_m]}{\sqrt{m}} \longrightarrow 2. \tag{8.37}$$

Hammersley (1972) picked up on this problem, showing that as $m \to \infty$, there exists a constant c such that

$$\frac{L_m}{\sqrt{m}} \longrightarrow c \quad \text{in probability.} \tag{8.38}$$

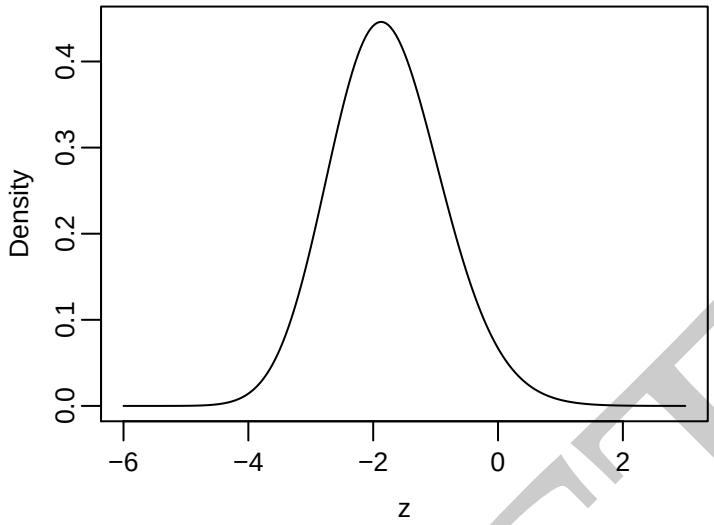

Figure 8.1: The Tracy-Widom density.

He also showed that $\pi/2 \leqslant c \leqslant e$, and, using Monte Carlo and other methods, suggested strongly that $c = 2$. Later, Logan & Shepp (1977) and Vershik & Kerov (1977) verified (8.37) conclusively.

The next challenge was to find the asymptotic variance of L_m. Using analysis, Kim (1996) conjectured that the variance would be of order $n^{1/3}$, as did Odlyzko & Rains (2000) and Wellner (2002) from extensive simulations, also proposing (8.35) for some Z. These conjectures were soon confirmed by Baik, Deift, & Johansson (1999), who showed that the Z had the Tracy -Widom distribution.

The simulations in Wellner (2002) suggest that the behavior of the empirical distribution is close to that predicted by the theoretical results only when m is quite large, say over 10^5 or 10^6. Figure 8.2 compares the actual or estimated density of L_m to the approximation using the Tracy-Widom distribution suggested by the asymptotic result (8.35). These plots show that the Tracy-Widom density is slightly to the left of the empirical density, even for $m = 10^7$.

Chiani (2014) showed that a shifted gamma distribution can approximate the Tracy-Widom distribution very well. We use the idea to approximate the distribution of the Ulam distance, where the gamma parameters depend on m. Let $Y \sim \text{Gamma}(\alpha, \theta)$, where α is the shape parameter and θ is the scale. Then we approximate the distribution of L_m by $\kappa + Y$ for the shift parameter κ. We use the method of moments. Let the estimated mean, variance, and skewness for L_m be $\widehat{\mu}$, $\widehat{\sigma}^2$, and $\widehat{\gamma}$, respectively. Then solving for (α, θ, κ):

$$\widehat{\mu} = \kappa + E[Y] = \kappa + \alpha\theta, \quad \widehat{\sigma}^2 = Var[Y] = \alpha\theta^2, \quad \widehat{\gamma} = \text{Skewness}(Y) = \frac{2}{\sqrt{\alpha}}$$
$$\Longrightarrow \alpha = \frac{4}{\widehat{\gamma}^2}, \quad \theta = \frac{\widehat{\sigma}}{\sqrt{\alpha}}, \quad \kappa = \widehat{\mu} - \alpha\theta. \quad (8.39)$$

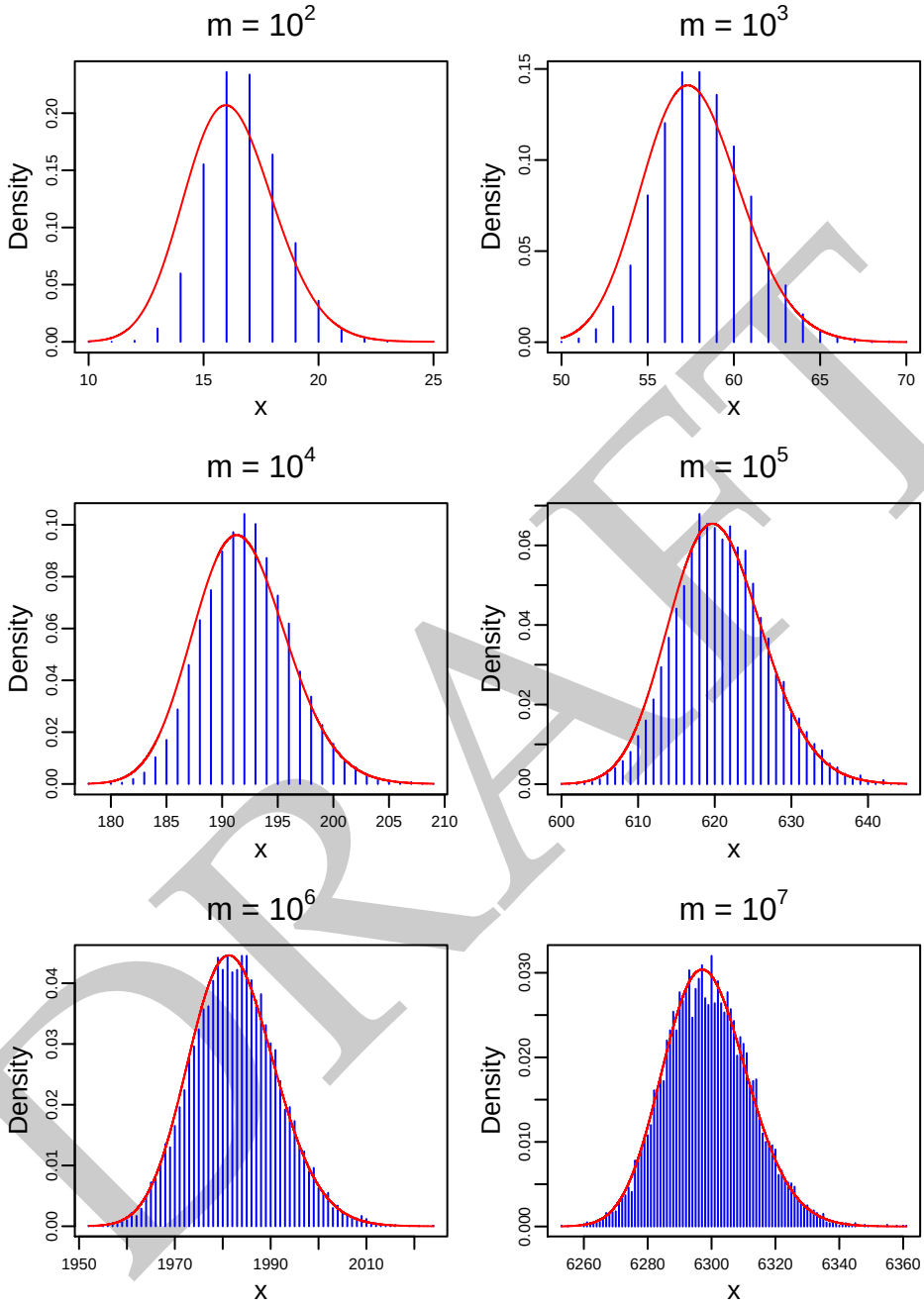

Figure 8.2: The Tracy-Widom density compared to the observed density of L_m. The spikes indicate the exact probabilities ($P[L_m = x]$) for $m = 100$, and the estimated probabilities from simulations by Wellner (2002) for $m \geqslant 1000$. The smooth curves represent the Tracy-Widom approximation to the density.

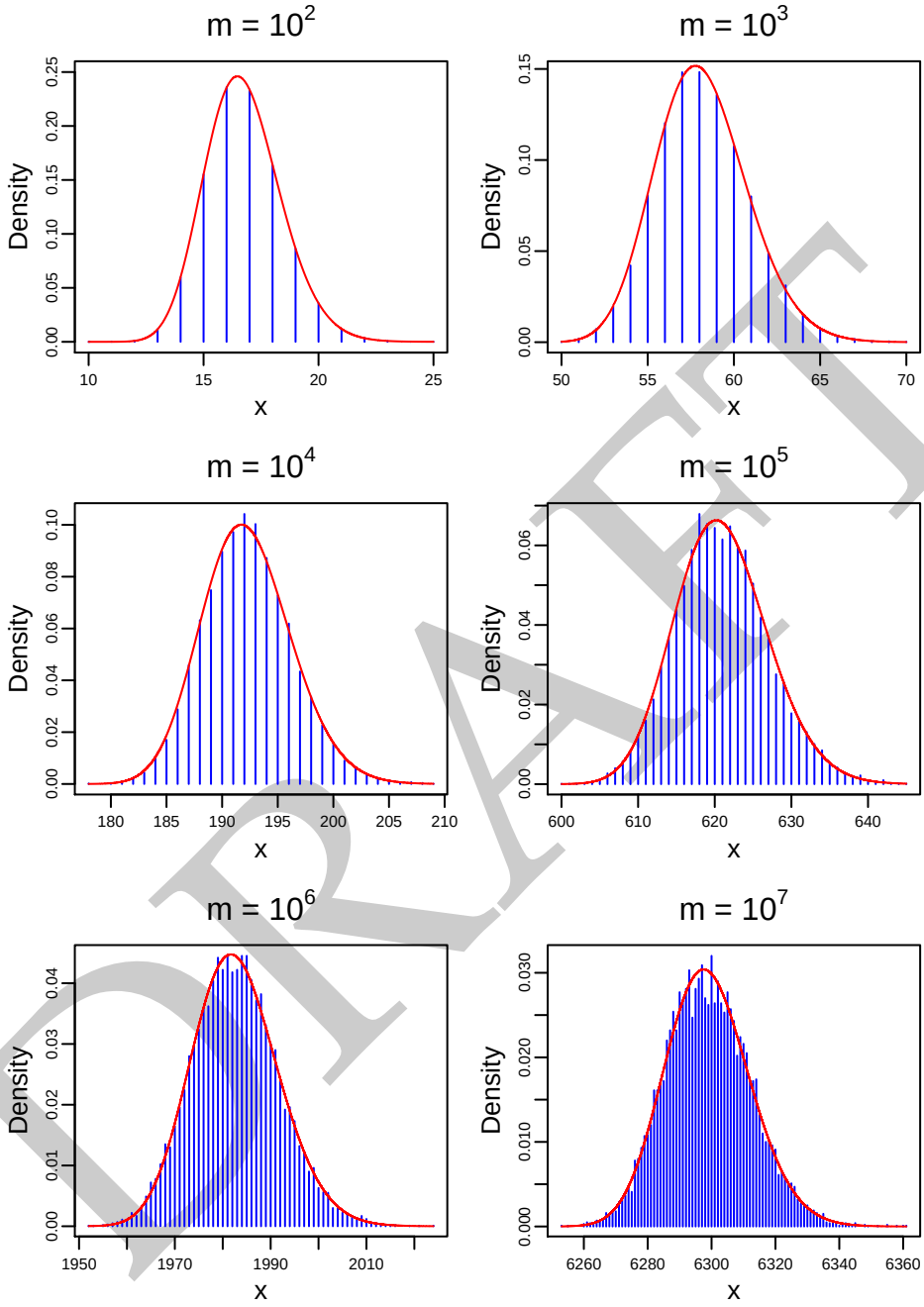

Figure 8.3: The gamma density compared to the observed density of L_m. The spikes indicate the exact probabilities ($P[L_m = x]$) for $m = 100$, and the estimated probabilities from simulations by Wellner (2002) for $m \geqslant 1000$. The smooth curves represent the gamma approximation to the density.

For $m \leqslant 150$, we can calculate the moments exactly. For larger m, we use the simulations in Wellner (2002). To smooth out the estimated moments, and to interpolate, we use linear regression to find models for the moments, so that upon finding the coefficients using least squares, we have the models for the mean and variance suggested by (8.35):

$$\widehat{\mu}_m = \widehat{\beta}_0 + \widehat{\beta}_1\sqrt{m} + \widehat{\beta}_2 m^{1/6} \text{ and } \widehat{\sigma}^2_m = \widehat{\beta}^*_0 + \widehat{\beta}^*_1 m^{1/3}. \tag{8.40}$$

We used a simple average for the skewness. We ran separate regressions for batches of m depending on the exponents in scientific notation, i.e., one batch used $m = 10^2, 2\times 10^2, \ldots, 10^3$, the next $m = 10^3, 2\times 10^3, \ldots, 10^4$, etc. Figure 8.3 has the analogous plots to those in Figure 8.2 for our gamma approximations. Visually, we see that the gamma does do better, especially for $m \leqslant 10^5$ or 10^6.

To summarize the graphs, we found the maximum and sum of absolute discrepancies of the observed probabilities and the smooth densities. Figure 8.4 compares the results using Tracy-Widom and gamma approximations for m between 10^2 and 10^7. Table (8.42) summarizes the graphs, averaging over groups of m's. We see that the gamma approximation is better for maximum discrepancy for $m \leqslant 10^5$, and for the sum of discrepancies for $m \leqslant 10^6$. For the sum of discrepancies, we also included a benchmark based on the expected sum. That is, for $\boldsymbol{X} \sim \text{Multinomial}_K(n, \boldsymbol{p})$ (multinomial with K categories and n observations), we calculate using the normal approximation

$$\sum_{k=1}^{K} E|X_k/n - p_k| \approx \sqrt{\frac{2}{\pi n}} \sum_{k=1}^{K} \sqrt{p_k(1-p_k)}. \tag{8.41}$$

We used the gamma density to approximate the $\boldsymbol{p}$ for each m, and here $n = 10,000$, the size of the simulations. Note that the gamma approximation in Figure 8.4 tracks the benchmark quite closely, meaning it is about as accurate as we can obtain using the simulations.

Maximum discrepancy

m	Tracy-Widom	Gamma	$\frac{\text{Tracy-Widom}}{\text{Gamma}}$
$10^2 - 10^3$	0.0314	0.0046	6.8013
$10^3 - 10^4$	0.0142	0.0051	2.7544
$10^4 - 10^5$	0.0078	0.0048	1.6239
$10^5 - 10^6$	0.0050	0.0042	1.1918
$10^6 - 10^7$	0.0043	0.0039	1.1135

(8.42)

Total discrepancy

m	Tracy-Widom	Gamma	Benchmark	$\frac{\text{Tracy-Widom}}{\text{Gamma}}$	$\frac{\text{Gamma}}{\text{Benchmark}}$
$10^2 - 10^3$	0.2045	0.0225	0.0253	9.0949	0.8884
$10^3 - 10^4$	0.1245	0.0326	0.0321	3.8164	1.0173
$10^4 - 10^5$	0.0893	0.0410	0.0398	2.1796	1.0292
$10^5 - 10^6$	0.0658	0.0478	0.0488	1.3774	0.9789
$10^6 - 10^7$	0.0709	0.0627	0.0595	1.1317	1.0535

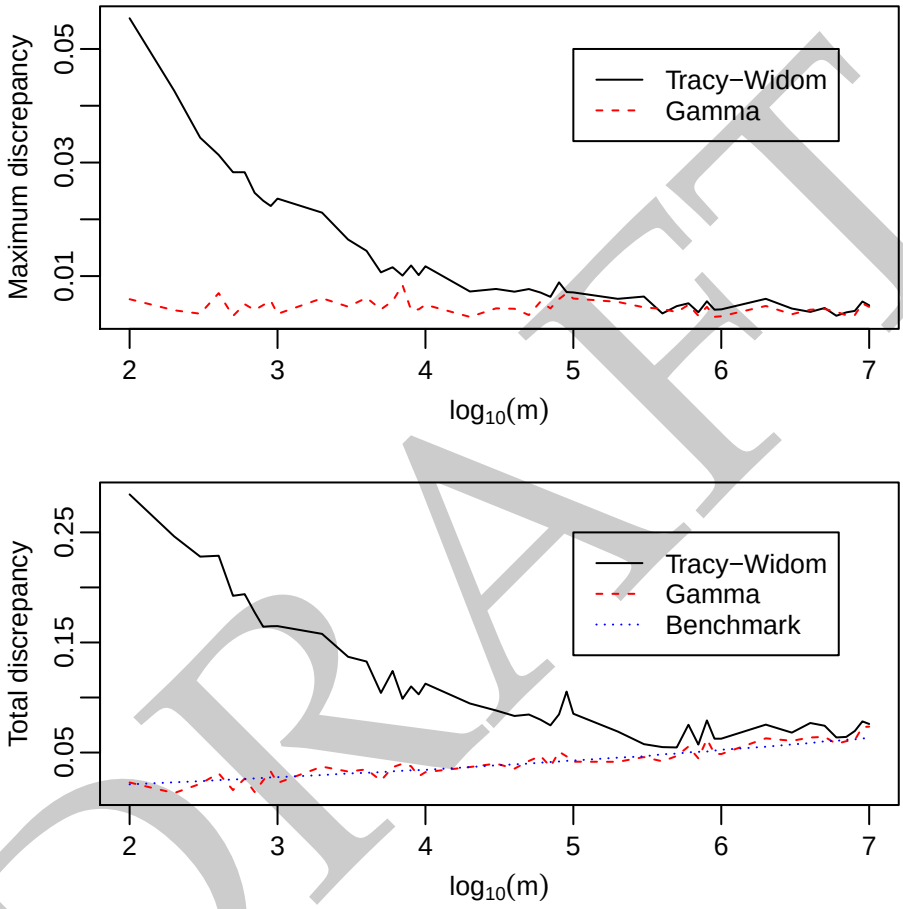

Figure 8.4: The discrepancy between the observed and approximate density of L_m. The top graph compares the Tracy-Widom approximation to that of the gamma using the maximum discrepancy. The bottom graph compares them on the sum of the discrepancies. The latter also exhibits the estimated expected sum of discrepancies of the simulations from the true probabilities.

Chapter 9

Cayley's distance

There are two distinct (though equivalent) approaches to calculating Cayley's distance: counting the number of cycles subtracted from m, or counting the number of interchanges. The latter is useful for finding exact and asymptotic distributions and moments, while the former is faster for actually calculating the distance.

For a vector $\boldsymbol{y} \in \mathcal{P}_m$, a **cycle** is a vector of indices $(i_1, \ldots, i_k)$ such that

$$y_{i_j} = i_{j+1}\ , j = 1, \ldots, k, \text{ and } \ y_{i_k} = i_1. \tag{9.1}$$

For example, if $\boldsymbol{y} = (4, 1, 6, 2, 5, 3)$, then

$$\begin{aligned} y_1 = 4,\ y_4 = 2,\ y_2 = 1 \ &\Rightarrow \ C_1 = (1, 4, 2) \text{ is a cycle;} \\ y_3 = 6,\ y_6 = 3 \ &\Rightarrow \ C_2 = (3, 6) \text{ is a cycle;} \\ y_5 = 5 \ &\Rightarrow \ C_3 = (5) \text{ is a cycle.} \end{aligned} \tag{9.2}$$

Note that we can cycle the cycles, i.e., $(4, 2, 1)$ and $(2, 1, 4)$ are also cycles, but are considered to be equivalent to $(1, 4, 2)$. (Also equivalent are $(3, 6)$ and $(6, 3)$.) Any $\boldsymbol{y}$ can be uniquely (up to equivalence) decomposed into a set of cycles. Cayley's distance is then defined as

$$d_{\text{Cay}}(\boldsymbol{y}, \boldsymbol{\omega}) = m - \#\text{Cycles}(\boldsymbol{y}). \tag{9.3}$$

For the above, $d_{\text{Cay}}(4, 1, 6, 2, 5, 3)) = 6 - 3 = 3$.

Recall as in (6.3) that Kendall's distance between $\boldsymbol{y}$ and $\boldsymbol{\omega} = (1, \ldots, m)$ is the minimum number of adjacent interchanges of elements of $\boldsymbol{y}$ needed to obtain $\boldsymbol{\omega}$. Cayley's distance is the minimum number of interchanges as well, but is not restricted to adjacent. For example, with the above $\boldsymbol{y}$:

$$\begin{aligned} \text{Kendall}: \ & (4, 1, 6, 2, 5, 3) \to (1, 4, 6, 2, 5, 3) \to (1, 4, 2, 6, 5, 3) \to (1, 2, 4, 6, 5, 3) \\ & \to (1, 2, 4, 6, 3, 5) \to (1, 2, 4, 3, 6, 5) \to (1, 2, 3, 4, 6, 5) \to (1, 2, 3, 4, 5, 6) \\ \text{Cayley}: \ & (4, 1, 6, 2, 5, 3) \to (2, 1, 6, 4, 5, 3) \to (1, 2, 6, 4, 5, 3) \to (1, 2, 3, 4, 5, 6) \end{aligned} \tag{9.4}$$

So $d_{\text{Ken}}(\boldsymbol{y}, \omega) = 7$, while $d_{\text{Cay}}(\boldsymbol{y}, \boldsymbol{\omega}) = 3$ as before.

The two methods for Cayley's distance can be seen to yield the same distance. Note that in order to arrange $\boldsymbol{y}$ in order, each of the cycles has to be arranged in order. The minimum

number of interchanges to arrange a cycle C_l in order is $\#C_l - 1$. Hence with $L = \#\text{Cycles}$, the total number of interchanges is

$$\sum_{l=1}^{L} (\#C_l - 1) = m - L = d_{\text{Cay}}(\boldsymbol{y}, \boldsymbol{\omega}), \tag{9.5}$$

since lengths of the cycles must sum to the number of objects.

In Section 9.1 we present a decomposition of Cayley's distance based on interchanges as in (9.4). This representation facilitates finding moments, the exact distribution, and the asymptotic normality.

9.1 Decomposition

Feller (1968, page 257–258) and Diaconis (1988, page 118) show that Cayley's distance can be written as a sum of independent Bernoulli's. Consider the interchange algorithm as in (9.4). We start by making an interchange, if necessary, so that $y_1 = 1$. Next, we make the interchange so that $y_2 = 2$. We continue until $y_{m-1} = m - 1$, which necessitates that $y_m = m$. Then v_i is the indicator function of having actually made a switch at stage i, and Cayley's distance is

$$D_{\text{Cay}} \equiv d_{\text{Cay}}(\boldsymbol{Y}, \boldsymbol{\omega}) = \sum_{i=1}^{m-1} V_i. \tag{9.6}$$

More formally, let $\boldsymbol{y}^{(0)} = \boldsymbol{y}$. For $i = 1, \ldots, m-1$, we obtain $\boldsymbol{y}^{(i)}$ from $\boldsymbol{y}^{(i-1)}$ by switching (if necessary) the i^{th} element with the element equalling k:

$$\begin{aligned} \boldsymbol{y}_k^{(i)} &= \boldsymbol{y}_k^{(i-1)} \quad \text{for } k \in \{1, \ldots, m\} - \{i, j(i)\} \quad \text{where} \quad \boldsymbol{y}_{j(i)}^{(i-1)} = i; \\ \boldsymbol{y}_i^{(i)} &= i; \\ \boldsymbol{y}_{j(i)}^{(i)} &= \boldsymbol{y}_i^{(i-1)}. \end{aligned} \tag{9.7}$$

A switch is made at stage i if $\boldsymbol{y}_i^{(i-1)} \neq i$ (i.e., $j(i) \neq i$). Assume $\boldsymbol{Y} \sim \text{Uniform}(\mathcal{P}_m)$. The indicator function of a switch at stage i is

$$V_i = I[\boldsymbol{Y}_i^{(i-1)} \neq i]. \tag{9.8}$$

Since $\boldsymbol{Y}_i^{(i-1)} = i$ for $i = 1, \ldots, i-1$ (if $i > 1$), the chance that $Y_i^{(i-1)} = i$ is $1/(m - i + 1)$. The V_i can also be shown to be independent, since how $Y_i^{(i)}$ came to equal i does not affect the order of $Y_{i+1}^{(i)}, \ldots, Y_m^{(i)}$, which implies that

$$V_1, \ldots, V_{m-1} \text{ are independent}, \quad V_i \sim \text{Bernoulli}\left(1 - \frac{1}{m - i + 1}\right). \tag{9.9}$$

Thus as for Kendall's distance, it is easy to find the mean, variance, a convolution formula for the exact distribution, and asymptotic approximations for D_{Cay}.

9.2 Moments and cumulants

Since the mean and variance of a Bernoulli(p) are p and $p(1-p)$, respectively, (9.9) gives us

$$\mathrm{E}[D_{\mathrm{Cay}}] = \sum_{i=1}^{m-1} \mathrm{E}[V_i] = \sum_{i=1}^{m-1} \left(1 - \frac{1}{m-i+1}\right) = m - 1 - \sum_{i=2}^{m} \frac{1}{i} = m - H_m^{(1)}, \text{ and}$$

$$\mathrm{Var}[D_{\mathrm{Cay}}] = \sum_{i=1}^{m-1} \mathrm{Var}[V_i] = \sum_{i=1}^{m-1} \frac{1}{m-i+1}\left(1 - \frac{1}{m-i+1}\right) = \sum_{i=2}^{m} \frac{i-1}{i^2} = H_m^{(1)} - H_m^{(2)}, \quad (9.10)$$

where $H_m^{(k)}$ is the generalized harmonic function

$$H_m^{(k)} = \sum_{i=1}^{m} \frac{1}{i^k}. \quad (9.11)$$

To find higher moments and cumulants, it appears easiest to start with the raw moments of the V_i, find the cumulants of the V_i, then sum over i to find the cumulants of D_{Cay}. We can then obtain the moments from the cumulants. None of these quantities seem to have a very compact expression owing to the harmonic functions present. We do not present Mathematica functions because they just express everything in terms of the $H_m^{(k)}$'s.

Start with $W \sim$ Bernoulli(p), so that $\mu_n'(p) = \mathrm{E}[W^n] = p$ for any $n \geqslant 1$. Using (2.17), we have that the n^{th} cumulant of W is

$$\begin{aligned}\kappa_n(p) &= n! \sum_{\boldsymbol{k} \in \mathcal{A}_n} (-1)^{\boldsymbol{k}^*-1}(\boldsymbol{k}^*-1)! \prod_{l=1}^{n} \frac{1}{k_l!}\left(\frac{p}{l!}\right)^{k_l} \\ &= n! \sum_{\boldsymbol{k} \in \mathcal{A}_n} p^{\boldsymbol{k}^*} (-1)^{\boldsymbol{k}^*-1}(\boldsymbol{k}^*-1)! \prod_{l=1}^{n} \frac{1}{k_l!}\left(\frac{1}{l!}\right)^{k_l}. \end{aligned} \quad (9.12)$$

As in (2.14) and (2.15), $\mathcal{A}_n$ is the set of vectors of nonnegative integers $\boldsymbol{k}$ for which $\sum l k_l = n$, and $\boldsymbol{k}^* = \sum k_l$. We will rewrite things a bit. First, add V_m to the sum in (9.6), where by (9.9), $V_m \equiv 0$. Then consider

$$D_{\mathrm{Cay}}^* \equiv m - D_{\mathrm{Cay}} = m - \sum_{i=1}^{m} V_i = \sum_{i=1}^{m} W_i, \text{ where } W_i = 1 - V_{m-i+1} \sim \text{Bernoulli}\left(\frac{1}{i}\right). \quad (9.13)$$

If κ_n and κ_n^* are the n^{th} cumulants of D_{Cay} and D_{Cay}^*, respectively, we have

$$\kappa_n = (-1)^n \kappa_n^*, \ n \geqslant 2. \quad (9.14)$$

Since the V_i are independent, so are the W_i, hence the n^{th} cumulant of D_{Cay}^* sums over the

appropriate $\kappa_n(p)$:

$$\begin{aligned}\kappa_n^* &= \sum_{i=1}^{m} \kappa_n\left(\frac{1}{i}\right) \\ &= \sum_{i=1}^{m} n! \sum_{\boldsymbol{k}\in\mathcal{A}_n} \left(\frac{1}{i}\right)^{\boldsymbol{k}^*} (-1)^{\boldsymbol{k}^*-1}(\boldsymbol{k}^*-1)! \prod_{l=1}^{n} \frac{1}{k_l!}\left(\frac{1}{l!}\right)^{k_l} \\ &= n! \sum_{\boldsymbol{k}\in\mathcal{A}_n} H_m^{(\boldsymbol{k}^*)} (-1)^{\boldsymbol{k}^*-1}(\boldsymbol{k}^*-1)! \prod_{l=1}^{n} \frac{1}{k_l!}\left(\frac{1}{l!}\right)^{k_l}. \end{aligned} \tag{9.15}$$

See (9.11).

9.3 Normal and Edgeworth approximations

To prove asymptotic normality, we again appeal to the Lindeberg-Feller theorem, Theorem 6.1. For the W_i in (9.13),

$$\sum_{i=1}^{m-1} E[|V_i - E[V_i]|]^\nu = \sum_{i=1}^{m} E[|W_i - E[W_i]|]^\nu. \tag{9.16}$$

By the triangle inequality, and noting that the W_i's are nonnegative,

$$|W_i - E[W_i]| \leqslant W_i + E[W_i] \quad\Rightarrow\quad E[|W_i - E[W_i]|]^\nu \leqslant 2^\nu E[W_i]^\nu = 2^\nu H_m^{(\nu)}. \tag{9.17}$$

Thus by (9.10),

$$\frac{\sum_{i=1}^{m-1} E[|V_i - E[V_i]|]^\nu}{Var[D_{Cay}]^{\nu/2}} \leqslant 2^\nu \frac{H_m^{(\nu)}}{(H_m^{(1)} - H_m^{(2)})^{\nu/2}}. \tag{9.18}$$

Since $H_m^{(1)}$ is asymptotically $\log(m)$ as $m \to \infty$, and $H_m^{(\nu)}$ is bounded in m for $\nu > 1$, the ratio in (9.18) goes to zero for any $\nu > 1$. Thus the theorem shows that

$$\frac{D_{Cay} - E[D_{Cay}]}{\sqrt{Var[D_{Cay}]}} \longrightarrow N(0,1) \;\text{ as }\; m \to \infty. \tag{9.19}$$

We tested the Edgeworth approximations for $L = 0, \ldots, 10$ terms, for m up to 10,000. Even for $m = 10,000$, the Edgeworth expansion is not much faster than the exact algorithm, so it is reasonable to use the latter when available. The density version of the Edgeworth expansions were better than those using the distribution-function-based version, especially for larger L, so we will restrict discussion to the density-based ones. From (9.15), we can see that the n^{th} cumulant is a linear combination of the harmonic functions $H_m^{(i)}$ for $i = 1, \ldots, n$, and the coefficient for $H_m^{(1)}$ is one (using $\boldsymbol{k}$ with "1" in the n^{th} slot and zeroes elsewhere). Thus all the cumulants are asymptotic to $\log(m)$, and the n^{th} normalized cumulant is asymptotic to $\log(m)^{1-n/2}$. This rate shows that the normalized cumulants decline very slowly in m, and

reasonably log-linearly in n, suggesting that the Edgeworth expansion errors behave similarly. Figures 9.1 for the errors in the density, 9.2 for the error in the distribution function (see (4.54)), and 9.3 for the relative errors (see (4.55)), seem to bear out this idea.

The maximal errors for $L = 10$ terms are given in (9.20) for various values of m. Even for $m = 25$, the approximation is quite good. The three types of error go from about $1\times10^{-5}, 6\times10^{-6}$, and 0.01 for $m = 50$ to about $1.5\times10^{-7}, 7.5\times10^{-8}$, and 0.0007 for $m = 10,000$.

	10	25	50	100	250
Density	0.000307	4.41×10^{-5}	1.08×10^{-5}	8.14×10^{-6}	1×10^{-6}
Distribution function	0.000255	2.72×10^{-5}	6.07×10^{-6}	4.76×10^{-6}	5.84×10^{-7}
p-value (relative)	0.229	0.012	0.0103	0.00429	0.0013
	500	1000	1250	5000	10000
Density	1.09×10^{-6}	6.97×10^{-7}	5.02×10^{-7}	1.96×10^{-7}	1.47×10^{-7}
Distribution function	6.23×10^{-7}	4.25×10^{-7}	3.47×10^{-7}	1.18×10^{-7}	7.42×10^{-8}
p-value (relative)	0.00308	0.000888	0.000813	0.000592	0.000742

(9.20)

9.4 R code

The function cayley_cumulants(m,L) produces the first L cumulants of Cayley's distance. It uses the function pd from (2.53) to calculate the product term in (9.15) (with mom==1). We use an approximation to the harmonic numbers in (9.11) that takes effect for larger values of n and m. For $n > 1$, $H_m^{(n)} \to \zeta(n)$ as $m \to \infty$, where ζ is the **Riemann zeta function** (Wikipedia contributors, 2018a). If the difference between the zeta function and the harmonic number is estimated to be less than 10^{-20}, then we use the zeta function. For the zeta function itself, we use a numerical value accurate to twenty decimal places for $2 \leqslant n \leqslant 20$, and for $n > 20$, just use $H_{10}^{(n)}$.

The Edgeworth functions are cayley_edgeworthf and cayley_edgeworthF, which find the estimated density and distribution function, respectively. The former is the preferable one. The arguments are x,m,L,N, where x is the variable in f(x) or F(x), m is the number of objects, L is the number of terms in the expansion, and N is the sample size. These functions rely on edgef and edgeF from (2.5).

```
cayley_cumulants <- function(m,L=12) {
    if(m<2) return(rep(0,L))
    ii <- 1:m
    mom <- rep(1,L)
    kurt <- m-sum(1/ii)
    A <- findA(L)
    for(s in 2:L) {
        krt <- 0
        for(i in 1:nrow(A[[s]])) {
            ku <- A[[s]][i,]
            ks <- sum(ku)
            krt <- krt + harmonic_number(m,ks)*(-1)^(ks-1)*factorial(ks-1)*pd(ku,mom)
```

```
        }
        kurt <- c(kurt,krt*factorial(s)*(-1)^s)
    }
    kurt
}

harmonic_number <- function(m,n) {
    M <- min(m,10^((20-log10(n-1))/(n-1))+.5)
    if(n>1&&m>M) return(zeta(n))
    sum(1/(1:m)^n)
}


zeta <- function(n) {
    if(n==1) {return(Inf)}
    if(n>20) {return(sum(1/(1:10)^n))}
    c(pi^2/6, 1.2020569031595942854,pi^4/90,1.0369277551433699263, pi^6/945,
        1.0083492773819228268,pi^8/9450,1.0020083928260822144, pi^10/93555, 1.0004941886041194646,
        (691*pi^12)/638512875, 1.000122713347578489 1,
        (2*pi^14)/18243225, 1.0000305882363070205,(3617*pi^16)/325641566250, 1.0000076371976378998,
        (43867*pi^18)/38979295480125, 1.0000019082127165539,(174611*pi^20)/1531329465290625)[n-1]
}

cayley_normalized_cumulants <- function(m,L=12) {
    if(m<2) return(rep(0,L))
    cc <- cayley_cumulants(m,L)
    c(0,1,cc[-(1:2)]/cc[2]^((3:L)/2))
}

cayley_edgeworthf <- function(x,m,L=10,N=1) {
    cum <- cayley_cumulants(m,L+2)
    sigma <- sqrt(cum[2])
    kum <- c(0,1,cum[-(1:2)]/sigma^(3:length(cum)))
    z <- (x-cum[1])/sigma
    dnorm(z)*edgef(z,L,kum,N)/sigma
}

cayley_edgeworthF <- function(x,m,L=10,N=1) {
    cum <- cayley_cumulants(m,L+2)
    sigma <- sqrt(cum[2])
    kum <- c(0,1,cum[-(1:2)]/sigma^(3:length(cum)))
    z <- (x-cum[1]+.5)/sigma
    pnorm(z) - dnorm(z)*edgeF(z,L,kum,N)
}
```

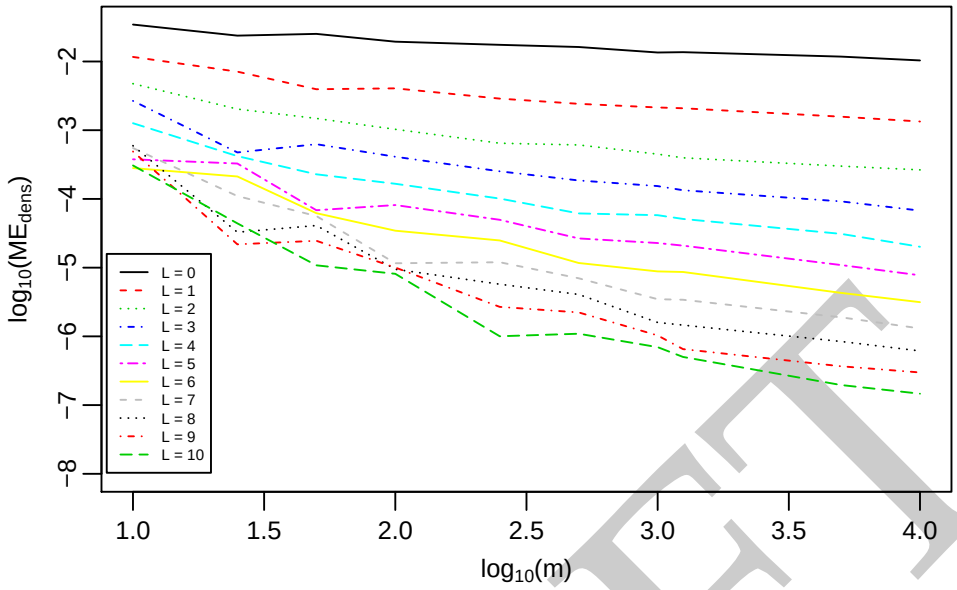

Figure 9.1: The maximum error in estimating the density for Kendall's distance, as a function of $\log_{10}(m)$. The values are $\log_{10}$ of the ME_{dens}; the lines depend on L, the number of terms in the Edgeworth expansion.

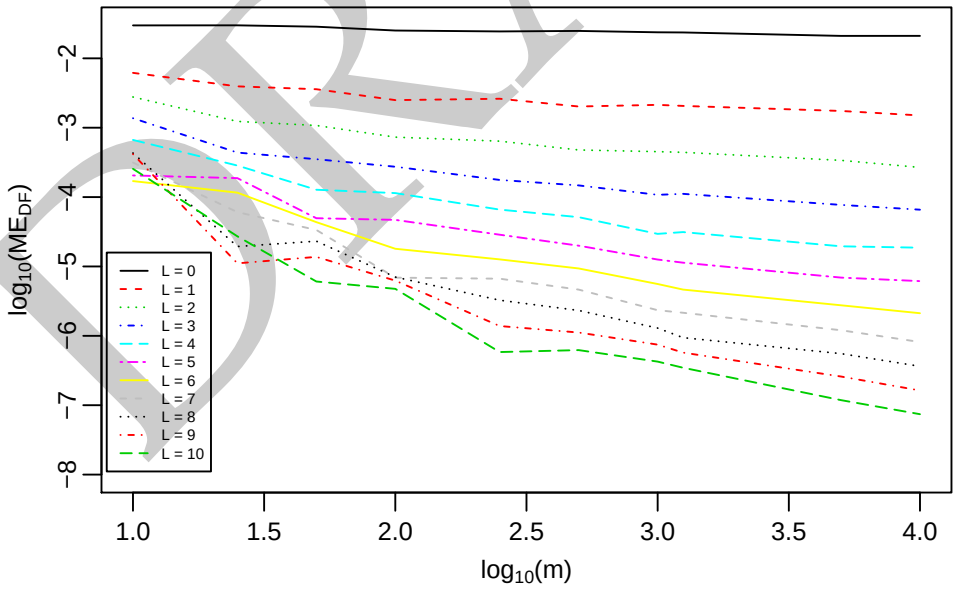

Figure 9.2: The maximum error in estimating the distribution function for Kendall's distance, as a function of $\log_{10}(m)$. The values are $\log_{10}$ of the ME_{DF}; the lines depend on L, the number of terms in the Edgeworth expansion.

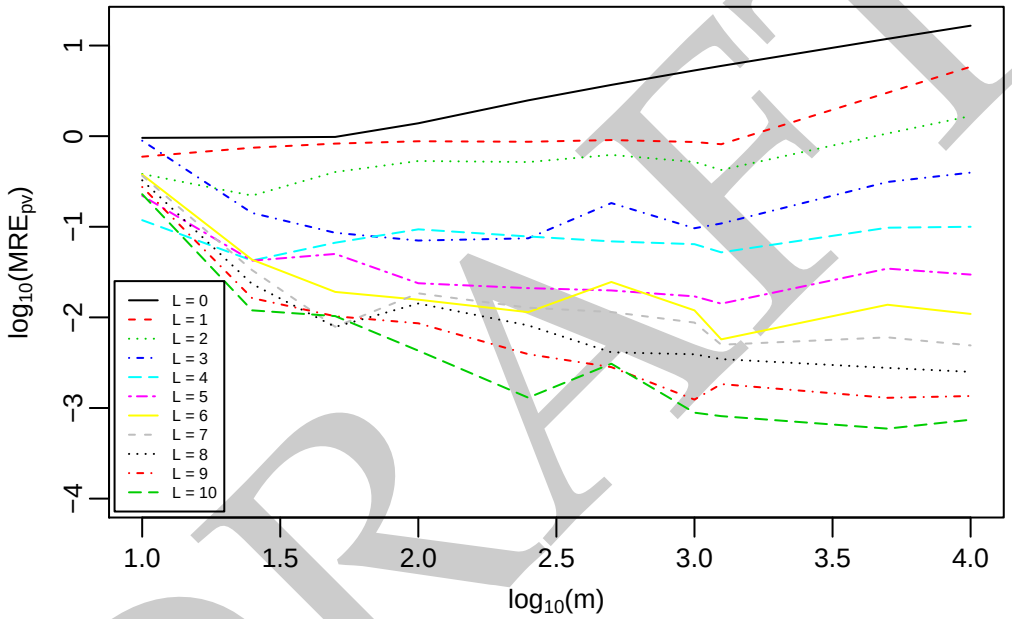

Figure 9.3: The maximum relative error in estimating the p-value (for p-values > 0.00001) for Kendall's distance, as a function of $\log_{10}(m)$. The values are $\log_{10}$ of the MRE_{pv}; the lines depend on L, the number of terms in the Edgeworth expansion.

Chapter 10

Maximum distance

10.1 Introduction

We again take $\boldsymbol{Y} \sim \text{Uniform}(\mathcal{P}_m)$, but now consider the exact and asymptotic distribution of $d_{\text{Max}}(\boldsymbol{Y}, \boldsymbol{\omega})$, where $\boldsymbol{\omega}$ is a fixed vector in $\mathcal{P}_m$, and

$$d_{\text{Max}}(\boldsymbol{y}, \boldsymbol{\omega}) = \text{Max}_{\{i=1,\ldots,m\}}\{|y_i - \omega_i|\}, \tag{10.1}$$

the maximum elementwise difference of the two vectors. The distribution is the same for any $\boldsymbol{\omega} \in \mathcal{P}_m$, and in fact the same of as $d_{\text{Max}}(\boldsymbol{Y}, \boldsymbol{W})$ for $\boldsymbol{W}$ any random vector over $\mathcal{P}_m$ independent of $\boldsymbol{Y}$. For convenience we can from now on take

$$\boldsymbol{\omega} = (1, 2, \ldots, m). \tag{10.2}$$

If m is small ($m \leqslant 10$, say), the exact distribution of $D_{\text{Max}} \equiv d_{\text{Max}}(\boldsymbol{Y}, \boldsymbol{\omega})$ can be found reasonably quickly by enumerating over $\mathcal{P}_m$. For m up to about 24, in Section 10.2 we present a method based on one in the literature for finding the distribution of Spearman's sum-of-squares distance.

Section 10.3 has an expression for the distribution for values of d_{Max} over $m/2$. If m is large, the probability $M > m/2$ is very close to 1. This expression can be used to find an asymptotic expansion of the distribution that works well for moderate m. It also leads the way to the interesting asymptotic distribution (Section 10.7)

$$P\left[\frac{m - D_{\text{Max}}}{\sqrt{m}} \leqslant x\right] \longrightarrow 1 - e^{-x^2}, x > 0. \tag{10.3}$$

Equivalently, $(D_{\text{Max}} - m)^2/m \to \text{Exponential}(1)$.

10.2 The exact distribution

For small m ($m \leqslant 10$, say), the exact distribution can be found by calculating $d_{\text{Max}}(\boldsymbol{y}, \boldsymbol{\omega})$ explicitly for each $\boldsymbol{y} \in \mathcal{P}_m$. For slightly large values ($11 \leqslant m \leqslant 24$, for us), we can use the method of treating subvectors of $\boldsymbol{y}$ separately, then combining the results, similar to the process in Section 3.3 for Hoeffding distances. (There we convolved densities; here we multiply

distribution functions.) Start with two subvectors. Choose $m_1 \approx m/2$, and for $\boldsymbol{y} \in \mathcal{P}_m$, let $\boldsymbol{y}^{(1)} = (y_1, \ldots, y_{m_1})$ and $\boldsymbol{y}^{(2)} = (y_{m_1+1}, \ldots, y_m)$, and similarly split up $\boldsymbol{\omega}$: $\boldsymbol{\omega}^{(1)} = (1, \ldots, m_1)$, $\boldsymbol{\omega}^{(2)} = (m_1 + 1, \ldots, m)$. Then

$$d_{\text{Max}}(\boldsymbol{y}, \boldsymbol{\omega}) = \max\{d_{\text{Max}}(\boldsymbol{y}^{(1)}, \boldsymbol{\omega}^{(1)}), d_{\text{Max}}(\boldsymbol{y}^{(2)}, \boldsymbol{\omega}^{(2)})\}. \tag{10.4}$$

Recall the discussion around (3.27) to (3.30). We let $\mathcal{R}^{(1)}$ be a subset of m_1 distinct elements from $1, \ldots, m$, $\mathcal{R}^{(2)}$ be its complement, $\{1, \ldots, m\} - \mathcal{R}^{(1)}$, and

$$\mathcal{P}(\mathcal{R}^{(i)}) = \{\text{permutations of elements in } \mathcal{R}^{(i)}\}. \tag{10.5}$$

With $\boldsymbol{Y} \sim \text{Uniform}(\mathcal{P}_m)$ we again have that

$$\boldsymbol{Y}^{(1)} \text{ and } \boldsymbol{Y}^{(2)} \text{ are independent } \textbf{given} \text{ that } \boldsymbol{Y}^{(1)} \in \mathcal{P}(\mathcal{R}^{(1)}) \ (\Leftrightarrow\ \boldsymbol{Y}^{(2)} \in \mathcal{P}(\mathcal{R}^{(2)})), \tag{10.6}$$

$$\begin{aligned} \boldsymbol{Y}^{(1)} \mid \boldsymbol{Y}^{(1)} \in \mathcal{P}(\mathcal{R}^{(1)}) &\sim \text{Uniform}(\mathcal{P}(\mathcal{R}^{(1)})), \\ \boldsymbol{Y}^{(2)} \mid \boldsymbol{Y}^{(2)} \in \mathcal{P}(\mathcal{R}^{(2)}) &\sim \text{Uniform}(\mathcal{P}(\mathcal{R}^{(2)})), \end{aligned} \tag{10.7}$$

and

$$\mathcal{R}^{(1)} \sim \text{Uniform}(\{\text{All possible subsets of } m_1 \text{ distinct elements from } 1, \ldots, m\}). \tag{10.8}$$

Now let $F^{(i)}(x \mid \mathcal{R}^{(i)})$ be the following conditional distribution function of d_{Max}:

$$F^{(i)}(x \mid \mathcal{R}^{(i)}) = P[d_{\text{Max}}(\boldsymbol{Y}^{(i)}, \boldsymbol{\omega}^{(i)}) \leqslant x \mid \boldsymbol{Y}^{(1)} \in \mathcal{P}(\mathcal{R}^{(1)})], \quad i = 1, 2. \tag{10.9}$$

Then by (10.4) and the conditional independence in (10.6),

$$\begin{aligned} F(x \mid \mathcal{R}^{(i)}) &= P[d_{\text{Max}}(\boldsymbol{Y}, \boldsymbol{\omega}) \leqslant x \mid \boldsymbol{Y}^{(1)} \in \mathcal{P}(\mathcal{R}^{(1)})] \\ &= P[d_{\text{Max}}(\boldsymbol{Y}^{(1)}, \boldsymbol{\omega}^{(1)}) \leqslant x \ \&\ d_{\text{Max}}(\boldsymbol{Y}^{(2)}, \boldsymbol{\omega}^{(2)}) \leqslant x \mid \boldsymbol{Y}^{(1)} \in \mathcal{P}(\mathcal{R}^{(1)})] \\ &= P[d_{\text{Max}}(\boldsymbol{Y}^{(1)}, \boldsymbol{\omega}^{(1)}) \leqslant x \mid \boldsymbol{Y}^{(1)} \in \mathcal{P}(\mathcal{R}^{(1)})] P[d_{\text{Max}}(\boldsymbol{Y}^{(2)}, \boldsymbol{\omega}^{(2)}) \leqslant x \mid \boldsymbol{Y}^{(1)} \in \mathcal{P}(\mathcal{R}^{(1)})] \\ &= F^{(1)}(x \mid \mathcal{R}^{(1)}) F^{(2)}(x \mid \mathcal{R}^{(2)}). \end{aligned} \tag{10.10}$$

Then the unconditional distribution function of $d_{\text{Max}}(\boldsymbol{Y}, \boldsymbol{\omega})$ is found by taking the expectation of the condition distribution function over $\mathcal{R}^{(i)}$ in (10.8):

$$\begin{aligned} F(x) &= P[d_{\text{Max}}(\boldsymbol{Y}, \boldsymbol{\omega}) \leqslant x] \\ &= E[F^{(1)}(x \mid \mathcal{R}^{(1)}) F^{(2)}(x \mid \mathcal{R}^{(2)})]. \end{aligned} \tag{10.11}$$

Now for each splitting $(\mathcal{R}^{(1)}, \mathcal{R}^{(2)})$, we find the distributions of $d_{\text{Max}}(\boldsymbol{y}^{(i)}, \boldsymbol{\omega}^{(i)})$ for $\boldsymbol{y}^{(i)} \in \mathcal{P}(\mathcal{R}^{(i)})$, $i = 1, 2$, by enumeration:

$$P[d_{\text{Max}}(\boldsymbol{Y}^{(i)}, \boldsymbol{\omega}^{(i)}) = x \mid \boldsymbol{Y}^{(i)} \in \mathcal{P}(\mathcal{R}^{(i)})] = \frac{1}{m_i!} \#\{\boldsymbol{y}^{(i)} \in \mathcal{P}(\mathcal{R}^{(i)}) \mid d_{\text{Max}}(\boldsymbol{y}^{(i)}, \boldsymbol{\omega}^{(i)}) = x\}, \tag{10.12}$$

where $m_2 = m - m_1$. We take the relevant cumulative sums to find the conditional $F^{(i)}$'s, then multiply the them as in (10.11). Then the final answer takes the average over the $\mathcal{R}^{(i)}$'s:

$$F(x) = \binom{m}{m_1}^{-1} \sum_{\text{splittings } (\mathcal{R}^{(1)}, \mathcal{R}^{(2)})} F^{(1)}(x \mid \mathcal{R}^{(1)}) F^{(2)}(x \mid \mathcal{R}^{(2)}). \tag{10.13}$$

As in Section 3.3, we can proceed with further splitting each $\mathcal{R}^{(i)}$.

10.3 An expression for the distribution

This section presents a useful expression for the distribution function of D_{Max} for values larger than $m/2$. The range of D_{Max} is $\{0, \ldots, m-1\}$. We look at the distribution function for values of D_{Max} near m, using integer k:

$$F(m-k) = P[D_{Max} \leqslant m-k], \;\; 1 \leqslant k < \left\lfloor \frac{m}{2} \right\rfloor + 1. \tag{10.14}$$

Note that

$$D_{Max} \leqslant m-k \;\text{ if and only if }\; |y_i - i| \leqslant m-k, \;\; i = 1, \ldots, m. \tag{10.15}$$

Then for each i, we can write out the values of y_i that satisfy the inequality:

$$\begin{aligned}
y_1 &\in \{1, \ldots, m-k+1\} \\
y_2 &\in \{1, \ldots, m-k+2\} \\
&\vdots \\
y_{k-1} &\in \{1, \ldots, m-1\} \\
y_k &: \quad \text{all of them} \\
&\vdots \\
y_{m-k+1} &: \quad \text{all of them} \\
y_{m-k+2} &\in \{2, \ldots, m\} \\
&\vdots \\
y_{m-1} &\in \{k-1, \ldots, m\} \\
y_m &\in \{k, \ldots, m\}.
\end{aligned} \tag{10.16}$$

Since the y_i's for $k \leqslant i \leqslant m-k+1$ always satisfy the inequality, we can ignore them in our calculations. Thus we have

$$P[D_{Max} \leqslant m-k] = P[Y_i \leqslant m-k+i, i = 1, \ldots, k-1, \;\text{ and }\; Y_i \geqslant i-m+k, \; i = m-k+1, \ldots, m]. \tag{10.17}$$

Let $\mathcal{Y}_1$ be the first set in the last probability, and $\mathcal{Y}_2$ be the second:

$$\begin{aligned}
\mathcal{Y}_1 &= \{\boldsymbol{y} \,|\, y_i \leqslant m-k+i, i = 1, \ldots, k-1\} \;\; \& \\
\mathcal{Y}_2 &= \{\boldsymbol{y} \,|\, y_i \geqslant i-m+k, \; i = m-k+1, \ldots, m\}.
\end{aligned} \tag{10.18}$$

For the rest of this section, we take k and m fixed. The marginal probabilities of these two sets are fairly easy to find, and are equal. But they are not independent. So we will first find $P[\mathcal{Y}_1]$ explicitly, then an expression for $P[\mathcal{Y}_2 \,|\, \mathcal{Y}_1]$.

Lemma 10.1. *In the above setup,*

$$P[\mathcal{Y}_1] = \frac{(m-k+1)^{k-1}}{(m)_{k-1}}. \tag{10.19}$$

Here, $(m)_l = m(m-1)\cdots(m-l+1) = m!/l!$.

Proof. We proceed iteratively. $P[Y_1 \leqslant m-k+1] = (m-k+1)/m$, since Y_1 is Uniform $(\{1,\ldots,m\})$. Given $y_1 \leqslant m-k+1$, Y_2 must be drawn from one of the $m-1$ values left that are less than or equal to $m-k+2$. Since whatever y_1 is, it is less than $m-k+2$, there are only $m-k+1$ left. Thus

$$P[Y_2 \leqslant m-k+2 \,|\, Y_1 = y_1, \text{ where } y_1 \leqslant m-k+1] = \frac{m-k+1}{m-1}. \tag{10.20}$$

Now for Y_3, we choose from $m-2$ values, and two of the values in $\{1,\ldots,m-k+3\}$ have been removed. Thus again we are left with $m-k+1$ values that satisfy the inequality:

$$P[Y_3 \leqslant m-k+3 \,|\, Y_1 = y_1, Y_2 = y_2 \text{ where } y_1 \leqslant m-k+1 \;\&\; y_2 \leqslant m-k+2] = \frac{m-k+1}{m-2}. \tag{10.21}$$

Continue if necessary for Y_i, $i = 4,\ldots,k-1$, each time obtaining a conditional probability of $(m-k+1)/(m-i+1)$. Multiplying those probabilities yields (10.19). □

We can use the same approach on $Y_m, Y_{m-1}, \ldots, Y_{m-k+2}$ (in that order) to show that $P[\mathcal{Y}_2] = P[\mathcal{Y}_1]$. But $P[\mathcal{Y}_2 \,|\, \mathcal{Y}_1]$ is more problematic, since the conditional probability

$$P[\mathcal{Y}_2 \,|\, Y_1 = y_1, \ldots, Y_{k-1} = y_{k-1}, \;\&\; \mathcal{Y}_1] \tag{10.22}$$

depends on the specific values of $y_1,\ldots,y_{k-1}$. To whit, the next lemma:

Lemma 10.2. *Fix $y_1,\ldots,y_{k-1}$, distinct values from $\{1,\ldots,m\}$. Then*

$$P[\mathcal{Y}_2 \,|\, Y_1 = y_1, \ldots, Y_{k-1} = y_{k-1}] = \frac{\prod_{a=2}^{k}(m-k+1-c_a)}{(m-k+1)_{k-1}}, \tag{10.23}$$

where

$$c_a \equiv c_a(y_1,\ldots,y_{k-1}) = \#\{y_j \geqslant a,\ j = 1,\ldots,k-1\}. \tag{10.24}$$

Proof. Start with $P[Y_m \geqslant k \,|\, Y_1 = y_1, \ldots, Y_{k-1} = y_{k-1}]$. There are conditionally $m-k+1$ values left for Y_m to choose from. Unconditionally, $m-k+1$ values satisfy the inequality $Y_m \geqslant k$, but conditionally c_k have been taken from the set $\{k,\ldots,m\}$. Thus

$$P[Y_m \geqslant k \,|\, Y_1 = y_1, \ldots, Y_{k-1} = y_{k-1}] = \frac{m-k+1-c_k}{m-k+1}. \tag{10.25}$$

Further condition on $Y_m = y_m$ for $y_m \geqslant k$. We want $Y_{m-1} \geqslant k-1$. There is one fewer value to choose from, i.e., $m-k+1-1 = m-k$. Initially there are $m-k+2$ values satisfying the inequality, but y_m has been taken by Y_m, so there are $m-k+1$ left, then $y_1,\ldots,y_{k-1}$ have taken c_{k-1} of them, yielding

$$P[Y_{m-1} \geqslant k-1 \,|\, Y_1 = y_1, \ldots, Y_{k-1} = y_{k-1}, Y_m = y_m, y_m \geqslant k] = \frac{m-k+1-c_{k-1}}{m-k}. \tag{10.26}$$

We continue, noting that for $Y_i \geqslant i-m+k$, the $Y_{i+1},\ldots,Y_m$ have removed $m-i$ that satisfy the inequality, and $y_1,\ldots,y_{k-1}$ have removed c_{i-m+k}, hence there are $m-(i-m+k)+1-(m-i)-c_{i-m+k} = m-k+1-c_{i-m+k}$ left. Thus

$$P[\mathcal{Y}_2 \,|\, Y_1 = y_1, \ldots, Y_{k-1} = y_{k-1}] = \frac{\prod_{i=m}^{m-k+2}(m-k+1-c_{i-m+k})}{(m-k+1)_{k-1}}, \tag{10.27}$$

which by setting $a = m-i+2$ is the same as (10.23). □

Now let $C_2, \ldots, C_k$ be random variables whose distribution is that induced by $\boldsymbol{Y}$ via (10.24):

$$C_a = c_a(Y_1, \ldots, Y_{k-1}), \quad a = 2, \ldots, k. \tag{10.28}$$

The main result of this section follows.

Proposition 10.3. *For* $k \leqslant \lfloor m/2 \rfloor + 1$,

$$P[D_{Max} \leqslant m-k] = \frac{E[\prod_{a=2}^{k}(m-k+1-C_a) \,|\, \mathcal{Y}_1] \times (m-k+1)^{k-1}}{(m)_{2k-2}}. \tag{10.29}$$

Proof. Let $\boldsymbol{Y}^{(1)} = (Y_1, \ldots, Y_{k-1})$, and define $\boldsymbol{y}^{(1)}$ similarly. Then

$$\begin{aligned} P[D_{Max} \leqslant m-k] &= P[\mathcal{Y}_1 \cap \mathcal{Y}_2] \\ &= \sum_{\boldsymbol{y}^{(1)} \in \mathcal{Y}_1} P[\boldsymbol{Y}^{(1)} = \boldsymbol{y}^{(1)} \;\&\; \mathcal{Y}_2] \\ &= \sum_{\boldsymbol{y}^{(1)} \in \mathcal{Y}_1} P[\mathcal{Y}_2 \,|\, \boldsymbol{Y}^{(1)} = \boldsymbol{y}^{(1)}] P[\boldsymbol{Y}^{(1)} = \boldsymbol{y}^{(1)}] \\ &= \sum_{\boldsymbol{y}^{(1)} \in \mathcal{Y}_1} \frac{\prod_{a=2}^{k}(m-k+1-c_a(\boldsymbol{y}^{(1)}))}{(m-k+1)_{k-1}} \times \frac{1}{(m)_{k-1}}, \end{aligned} \tag{10.30}$$

where the last equation uses (10.23) and the fact that $P[\boldsymbol{Y}^{(1)} = \boldsymbol{y}^{(1)}] = 1/(m)_{k-1}$. The conditional distribution of $\boldsymbol{Y}^{(1)}$ given that it is in $\mathcal{Y}_1$ is uniform over $\mathcal{Y}_1$. Since $\#\mathcal{Y}_1 = (m-k+1)^{k-1}$, we can rewrite (10.29) as

$$P[D_{Max} \leqslant m-k] = \frac{\sum_{\boldsymbol{y}^{(1)} \in \mathcal{Y}_1} \prod_{a=2}^{k}(m-k+1-c_a(\boldsymbol{y}^{(1)}))}{\#\mathcal{Y}_1} \times \frac{(m-k+1)^{k-1}}{(m-k+1)_{k-1}(m)_{k-1}}. \tag{10.31}$$

The first term in that last expression is $E[\prod_{a=2}^{k}(m-k+1-C_a) \,|\, \mathcal{Y}_1]$, and $(m)_{k-1}(m-k+1)_{k-1} = (m)_{2k-2}$, which proves (10.29). □

If we had the exact conditional distribution of $(C_2, \ldots, C_k)$, given $\mathcal{Y}_1$, then we could find the exact probability in (10.29). In the next section we use the proposition to find an approximation to the probabilities for moderate m. Section 10.7 uses the proposition to prove the asymptotic result in (10.3).

10.4 Approximation

Successive approximations to the probability in (10.14) can be obtained by expanding in a Taylor series in the C_a's around their $E[C_a \,|\, \mathcal{Y}_1]$'s, then taking the conditional expected value. To that end, let

$$\mu_a = E[C_a \,|\, \mathcal{Y}_1], \quad W_a = C_a - \mu_a, \quad \text{and set } \lambda = m-k+1. \tag{10.32}$$

Expanding, or just multiplying out, we have

$$\prod_{a=2}^{k}(\lambda-\mu_a-w_a)=\prod_{a=2}^{k}(\lambda-\mu_a)-\sum_{a=2}^{k}w_a P[-\{a\}]+\sum_{2\leqslant a<b\leqslant k}w_a w_b P[-\{a,b\}]$$
$$-\sum_{2\leqslant a<b<c\leqslant k}w_a w_b w_c P[-\{a,b,c\}]\pm\cdots+(-1)^{k-1}\prod_{i=2}^{k}w_i, \quad (10.33)$$

where we use the shorthand for a set of integers $\mathcal{I}$,

$$P[-\mathcal{I}]=\prod_{a\notin\mathcal{I}}(\lambda-\mu_a)=\frac{\prod_{a=2}^{k}(\lambda-\mu_a)}{\prod_{a\in\mathcal{I}}(\lambda-\mu_a)}. \quad (10.34)$$

Take the conditional expected value over the W_a's on both sides of (10.33). Note that $E[W_a \mid \mathcal{Y}_1]=0$. Then

$$E\left[\prod_{a=2}^{k}(\lambda-\mu_a-w_a)\mid\mathcal{Y}_1\right]=(1+s_2-s_3\pm\cdots+(-1)^{k-1}s_{k-1})\prod_{a=2}^{k}(\lambda-\mu_a), \quad (10.35)$$

where

$$s_l=\sum_{2\leqslant a_1<a_2<\cdots<a_l\leqslant k}E\left[\frac{W_{a_1}}{\lambda-\mu_{a_1}}\frac{W_{a_2}}{\lambda-\mu_{a_2}}\cdots\frac{W_{a_l}}{\lambda-\mu_{a_l}}\mid\mathcal{Y}_1\right]. \quad (10.36)$$

To make use of the expansion, we need to at least be able to find some of the mixed moments. The next lemma gives explicit expressions for the first, second, and third mixed moments. The proof is in Section 10.6.

Lemma 10.4. *Suppose* $0\leqslant a<b<c\leqslant k\leqslant\lfloor\frac{m}{2}\rfloor+1$. *Then*

$$\mu_a\equiv E[C_a\mid\mathcal{Y}_1]=k-a+(a-1)\left(1-\frac{1}{\lambda}\right)^{k-1}; \quad (10.37)$$

$$\mu_{ab}\equiv E[C_aC_b\mid\mathcal{Y}_1]=(k-a)(k-b)$$
$$+(b-1+(b-2)(k-a)+a(k-b))\left(1-\frac{1}{\lambda}\right)^{k-1}+(a-1)(b-2)\left(1-\frac{2}{\lambda}\right)^{k-1}; \quad (10.38)$$

and

$$\mu_{abc}\equiv E[C_aC_bC_c\mid\mathcal{Y}_1]=(k-a)(k-b)(k-c)+(x_0+x_1(k-1)+x_2(k-1)^2)\left(1-\frac{1}{\lambda}\right)^{k-1}$$
$$+(y_0+y_1(k-1))\left(1-\frac{2}{\lambda}\right)^{k-1}+(a-1)(b-2)(c-3)\left(1-\frac{3}{\lambda}\right)^{k-1}, \quad (10.39)$$

where

$$\begin{aligned}
x_0 &= (a-1)(7-4c+b(3c-5)),\\
x_1 &= 4c-9-2a(b+c-3)-b(2c-5),\\
x_2 &= a+b+c-3,\\
y_0 &= -(a-1)(12-5c+b(3c-7)),\ \textit{and}\\
y_1 &= (b-2)(c-3)+a(b+c-4).
\end{aligned} \tag{10.40}$$

In Section 10.6.1 we present a roadmap for finding higher-order moments.

For fixed k, the W_a's ($2 \leqslant a \leqslant k$) are bounded since $0 \leqslant C_a \leqslant k-1$, hence the term s_l in the expansion (10.35) is the sum of order k^l terms, each of order $1/m^l$. Thus with k of order $\sqrt{m}$ we have that s_l is of order $1/m^{l/2}$, showing that the expansion is reasonable. In Section 10.7 we consider asymptotics where k is of order $\sqrt{m}$ as $m \to \infty$.

For very large m, the first-order approximation works well:

$$\widehat{P}[D_{Max} \leqslant m-k] = \frac{\lambda^{k-1}\prod_{a=2}^{k}(\lambda-\mu_a)}{(m)_{2k-2}}. \tag{10.41}$$

If k is very large as well, a quicker approximation is available. Stirling's approximation to the factorial is

$$n! = \sqrt{2\pi}\,n^{n+\frac{1}{2}}e^{-n}\left(1+O\left(\frac{1}{n}\right)\right). \tag{10.42}$$

Taking logs in (10.41), we have

$$\begin{aligned}
\log(\widehat{P}[D_{Max} \leqslant m-k]) &= (k-1)\log(\lambda)+\sum_{a=2}^{k}\log(\lambda-\mu_a)-\log(m!)+\log((m-2k+2)!)\\
&\approx (k-1)\log(\lambda)+\sum_{a=2}^{k}\log(\lambda-\mu_a)-(m+\tfrac{1}{2})\log(m)\\
&\quad +(m-2k+2\tfrac{1}{2})\log(m-2k+2)+2k-2.
\end{aligned} \tag{10.43}$$

The summation can be approximated using an integral. From (10.37) we see that μ_1 is a linear function of a, hence

$$\sum_{a=2}^{k}\log(\lambda-\mu_a) \approx \int_{1\frac{1}{2}}^{k+\frac{1}{2}}\log(c_0+c_1a)\,da;\quad c_0 = k-\left(1-\frac{1}{\lambda}\right)^{k-1},\ c_1 = \left(1-\frac{1}{\lambda}\right)^{k-1}-1. \tag{10.44}$$

The indefinite integral of $\log(x)$ is $x\log(x)-x$, hence

$$\begin{aligned}
\int_{1\frac{1}{2}}^{k+\frac{1}{2}}\log(c_0+c_1a)\,da &= (\lambda-c_0-(k+\tfrac{1}{2})c_1)\log(\lambda-c_0-(k+\tfrac{1}{2})c_1)\\
&\quad -(\lambda-c_0-1\tfrac{1}{2}c_1)\log(\lambda-c_0-1\tfrac{1}{2}c_1)-k+1.
\end{aligned} \tag{10.45}$$

Replacing the summation in (10.43) with (10.45), then exponentiating, yields the quick estimate of the first-order estimate. It is accurate even for small m and k.

10.5 Results of the approximations

For $m \leqslant 24$, we can find the exact null distribution of D_{Max} using the algorithm in Section 10.2. For larger m, we use the expansion in (10.35) up to the J^{th} term:

$$P[D_{Max} \leqslant m-k] \approx \frac{\lambda^{k-1}\prod_{a=2}^{k}(\lambda-\mu_a)}{(m)_{2k-2}}\left(1+\sum_{j=2}^{J}(-1)^j s_j\right), \tag{10.46}$$

for $k \leqslant \lfloor m/2 \rfloor + 1$, where $\lambda = m-k+1$. The approximation is best for large m and small k, and in fact exact if $k \leqslant J+1$. The s_j's themselves become more time-consuming to calculate as k and j increase. Our strategy is to use as large a J as is computationally quick, which we take to be as follows:

$$\begin{array}{lccccc} J: & 2 & 3 & 4 & 5 & 6 \\ k: & 13,000 & 680 & 165 & 75 & 50 \end{array} \tag{10.47}$$

That is, if $k \leqslant 50$ we take $J = 6$, if $50 < k \leqslant 75$ we take $J = 5$, etc.

The approximations are reasonable for not-too-small probabilities of $P[D_{Max} \leqslant m-k]$, so are useful for testing hypotheses at the usual levels, but not accurate for large deviation results. For moderate or larger m, the probabilities $P[D_{Max} \leqslant m-k]$ for $k \geqslant k_0 \equiv \lfloor m/2 \rfloor + 1$ are very small. Here are some selected values for m between 25 and 100 and $k = k_0$:

$$\begin{array}{lcccc} m & 25 & 30 & 35 & 40 \\ \widehat{P}[D_{Max} \leqslant m-k_0] & 3.73\times10^{-4} & 3.73\times10^{-5} & 1.41\times10^{-5} & 1.31\times10^{-6} \\ m & 45 & 50 & 75 & 100 \\ \widehat{P}[D_{Max} \leqslant m-k_0] & 5.01\times10^{-7} & 4.04\times10^{-8} & 1.29\times10^{-11} & 4.05\times10^{-17} \end{array} \tag{10.48}$$

To estimate the probabilities $P[D_{Max} \leqslant m-k]$ for $k > k_0$ (for which the expansion is not valid), we use a couple of approaches. For $25 \leqslant m \leqslant 50$, we use simulations for some of the smaller k's. We could simulate directly the distribution of D_{Max}, but since we are interested only in the values $D_{Max} \leqslant m-k_0$, relatively few simulations are of use. We could improve the efficiency hugely by simulating directly conditioning on these values of D_{Max}, but are unable to find such an algorithm. We have found it easy to simulate from $\boldsymbol{Y} \in \mathcal{Y}_1$ of (10.18) for $k = k_0$. For $x \leqslant m-k_0$, $d_{Max}(\boldsymbol{y},\boldsymbol{\omega}) = x \Rightarrow \boldsymbol{y} \in \mathcal{Y}_1$, hence

$$P[d_{Max}(\boldsymbol{Y},\boldsymbol{\omega}) = x \,|\, \boldsymbol{Y} \in \mathcal{Y}_1] = \frac{P[d_{Max}(\boldsymbol{Y},\boldsymbol{\omega}) = x]}{P[\mathcal{Y}_1]}. \tag{10.49}$$

Then with $\widehat{P}_{Sim}$ denoting the simulated conditional probability, we have the unconditional approximation

$$\widehat{P}[d_{Max}(\boldsymbol{Y},\boldsymbol{\omega}) = x] = \widehat{P}_{Sim}[d_{Max}(\boldsymbol{Y},\boldsymbol{\omega}) = x \,|\, \boldsymbol{Y} \in \mathcal{Y}_1]\frac{(m-k_0+1)^{k_0-1}}{(m)_{k_0-1}} \tag{10.50}$$

by Lemma 10.1. This approach is substantially more efficient than direct simulation. For example, for $m = 25$ and 10^7 simulated $\boldsymbol{y}$'s, direct simulation yields about 2,400 of them in the target area (less than or equal to $m-k_0$), while the conditional approach expects about 4×10^5.

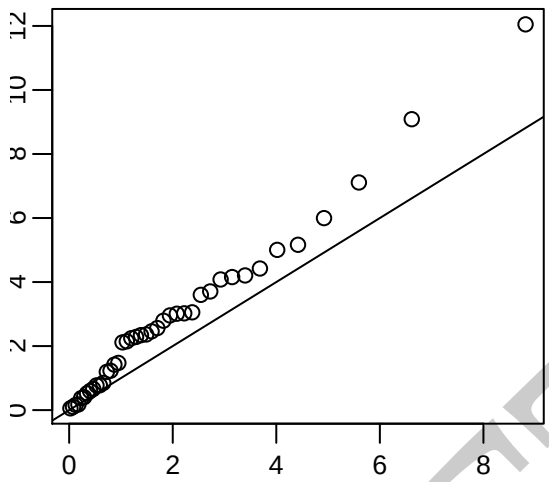

Figure 10.1: The QQ plot.

For $m = 50$, the numbers are 22,700 and 0.26. We use this approximation for $x < m - k_0$ for which the simulation yields at least 100 observations at x.

For $m > 50$, and for $m \leqslant 50$ when the above simulation is too sparse, we use the asymptotic distribution in (10.3) to estimate the ratio of the probabilities at the two values:

$$\widehat{P}[D_{Max} \leqslant m - k] = \widehat{P}[D_{Max} \leqslant m - k_0] \, \frac{e^{-(k-.5)^2/m}}{e^{-(k_0-.5)^2/m}}. \tag{10.51}$$

To test out our approximations, for each of several values of m we simulate 10^8 $\boldsymbol{y}$'s. We have two evaluations: testing the fit using χ^2's, and estimating the absolute error using Bayes analysis.

10.5.1 Testing goodness-of-fit

Here we compare the simulated distribution to the approximation using χ^2 goodness-of-fit tests. For given m, we looked at $P[D_{Max} = x]$ for $x \in \{m_0, \ldots, m_1\}$, where m_0 is the smallest values such that the expected counts is at least 5, i.e., the smallest integer such that $\widehat{P}[D_{Max} = m_0] \geqslant 5 \times 10^{-8}$, and $m_1 = m - 8$ (because the algorithm is exact for $x = m-7, \ldots, m-1$).

The bottom line is that we did not see any systematic problems with the approximation. Because we looked only at probabilities over 5×10^{-8}, we cannot claim anything about smaller probabilities. Such large deviation results would need a different approach.

The m's we use are 25, 26, ..., 50, 60, ..., 100, 150, 200, 300, 400, 500, 750, 1000, 2000, 5000, 10000. For each, we found the Pearson χ^2 statistic. Summing those, we find a total $\chi^2 = 1101.34$ on 1017 degrees of freedom, which yields a p-value of 0.0332. This value is fine, maybe a little low. Using the Sellke-Bayarri-Berger calibration, this p-value yields a posterior

probability of the null of 0.307 ($= -e(\text{p-value})\log(\text{p-value})$), not small. Inspecting the 41 individual p-values, there are six below 0.01: 0.0024, 0.0106, 0.0286, 0.0498, 0.0756, and 0.0818. The only worrisome one is the lowest, at least from a multiple-comparison's viewpoint. Just to check, we reran the simulations for the three m's with lowest p-values ($m = 28, 44, 90$). The rerun yielded p-values of 0.22, 0.82, and 0.97. So there does not seem to be anything particularly concerning with these three values.

Figure 10.1 contains a QQ plot based on the p-values, where we used $-2\log(\text{p-value})$'s, which under the null they are approximately independent χ^2_2's, and the smallest p-values are the largest in the transform. We again see that there is some lack of fit, but it is not too bad.

10.5.2 Assessing the absolute error

To estimate the absolute error in the approximations, we use a Bayes procedure. For given m and x, let $p = P[D_{Max} = x]$, $\widehat{p} = \widehat{P}[D_{Max} = x]$, our approximation to the probability, and S be the observed number of simulated $d_{Max}(\boldsymbol{y}, \boldsymbol{\omega})$'s that equal x (out of $n = 10^8$). Then $S \sim \text{Binomial}(n, p)$, hence with a $\text{Beta}(\frac{1}{2}, \frac{1}{2})$ prior on p, we have the posterior

$$p \,|\, S = s \sim \text{Beta}(a, b) \quad \text{where} \quad a = s + \tfrac{1}{2}, b = n - s + \tfrac{1}{2}. \tag{10.52}$$

Then letting $b(x\,;\,a,b)$ be the beta density and $B(x\,;\,a,b)$ the distribution function, the posterior expected error is

$$\begin{aligned} e_{\text{post}} \equiv E[|\widehat{p} - p| \,|\, S = s] &= \int |\widehat{p} - p| b(p\,;\,a,b)dp \\ &= \int_0^{\widehat{p}} (\widehat{p} - p) b(p\,;\,a,b)dp + \int_{\widehat{p}}^1 (p - \widehat{p}) b(p\,;\,a,b)dp \\ &= \widehat{p}\left(\int_0^{\widehat{p}} b(p\,;\,a,b)dp - \int_{\widehat{p}}^1 b(p\,;\,a,b)dp\right) \\ &\quad - \left(\int_0^{\widehat{p}} p\, b(p\,;\,a,b)dp - \int_{\widehat{p}}^1 p\, b(p\,;\,a,b)dp\right). \end{aligned} \tag{10.53}$$

For the beta,

$$x\, b(x\,;\,a,b) = \frac{a}{a+b}\, b(x\,;\,a+1,b), \tag{10.54}$$

hence

$$e_{\text{post}} = \widehat{p}(2B(\widehat{p}\,;\,a,b) - 1) - \frac{a}{a+b}(2B(\widehat{p}\,;\,a+1,b) - 1). \tag{10.55}$$

As $n \to \infty$, e_{Post} goes to the absolute error in our approximation, thus it combines this error with the uncertainty arising from the finite number of simulations. We expect it to be an overestimate of the actual error. (The smallest e_{Post} can be over $\widehat{p}$ is when $\widehat{p}$ equals the median of the beta, which in our examples this minimum is generally at least half the observed e_{Post}.)

We also look at the upper 95^{th} percentile of the posterior of the absolute error. This value is the point w_α such that

$$1 - B(\widehat{p} + w_\alpha\,;\,a,b) + B(\widehat{p} - w_\alpha,\,;\,a,b) = \alpha, \tag{10.56}$$

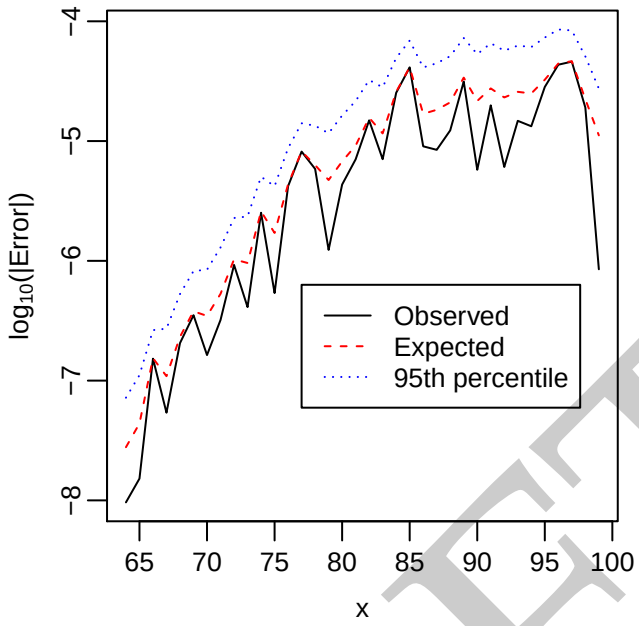

Figure 10.2: The observed, expected, and 95$^{\text{th}}$ percentile of the absolute error, by x, for $m = 100$.

for $\alpha = 0.05$. There doesn't seem to be an automatic function that will solve for w_α, so we need a root-finding procedure. Bisection works nicely, which requires initial bounds for w_α. For a lower bound, we find the two quantiles q_α and $q_{1-\alpha}$ of the Beta(a, b). Then we must have that

$$\widehat{p} + w_\alpha \geqslant q_{1-\alpha} \text{ and } \widehat{p} - w_\alpha \leqslant q_\alpha \;\Rightarrow\; w_\alpha \geqslant \max\{q_{1-\alpha} - \widehat{p}, \widehat{p} - q_\alpha\}. \tag{10.57}$$

We have two extreme cases to consider. Note that by the final inequality above,

$$q_{1-\alpha} - \widehat{p} \geqslant \widehat{p} \;\Rightarrow\; w_\alpha \geqslant \widehat{p} \;\Rightarrow\; B(\widehat{p} - w_\alpha\,;\, a, b) = 0 \;\Rightarrow\; \widehat{p} + w_\alpha = q_{1-\alpha}. \tag{10.58}$$

Similarly,

$$\widehat{p} - q_\alpha \geqslant 1 - \widehat{p} \;\Rightarrow\; w_\alpha \geqslant 1 - \widehat{p} \;\Rightarrow\; B(\widehat{p} + w_\alpha\,;\, a, b) = 1 \;\Rightarrow\; \widehat{p} - w_\alpha = q_\alpha. \tag{10.59}$$

Both the implication strings can be reversed, hence we have that

$$w_\alpha = \begin{cases} q_{1-\alpha} - \widehat{p} & \text{if } \; q_{1-\alpha} \geqslant 2\widehat{p} \\ \widehat{p} - q_\alpha & \text{if } \; q_\alpha \leqslant 2\widehat{p} - 1 \end{cases}. \tag{10.60}$$

If neither of the above cases hold, then we have $w_\alpha \leqslant \min\{\widehat{p}, 1 - \widehat{p}\}$, and we use bisection with this upper bound and the lower bound from (10.57).

For each m and x, we look at the observed absolute error between our approximation of $p \equiv P[D_{\text{Max}} = x]$, $\widehat{p}$, and the estimate from the simulation (when $n\widehat{p} \geqslant 5$). We also

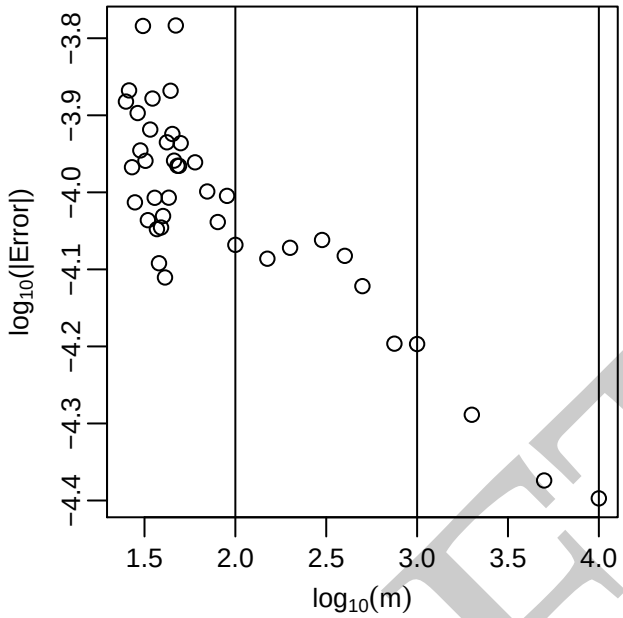

Figure 10.3: The maximum 95^{th} percentile of the absolute error, by m.

find the mean and 95^{th} percentiles of the posterior distribution of the absolute error of our approximation $|p - \widehat{p}|$. These three quantities track each other quite closely, with the observed absolute error being less than the expected absolute error, which in turn is less than the 95^{th} percentile by a factor of about 2 or 3. See Figure 10.2 for the $m = 100$ case. We will hence focus on the percentile.

For the most part, the 95^{th} percentiles of the errors are under 10^{-4}, with improvement as m increases, and as we get further into the lower tail of the distribution. Figure 10.3 plots the maximum of these percentiles for our values of m. For $25 \leqslant m < 100$, the maxima range from 7.75×10^{-5} to 1.65×10^{-4}; for $150 \leqslant m < 1000$, from 6.36×10^{-5} to 8.67×10^{-5}; and from $1000 \leqslant n \leqslant 10000$, from 4.01×10^{-5} to 6.36×10^{-5}.

Figure 10.4 plots the percentiles for each x against $F(x) = \widehat{P}[D_{\text{Max}} \leqslant x]$, the lower-tail probability (distribution function). The bulk of the values are between 10^{-4} and 10^{-5}, with the errors decreasing sharply for very small $F(x)$. The next table summarizes these percentiles.

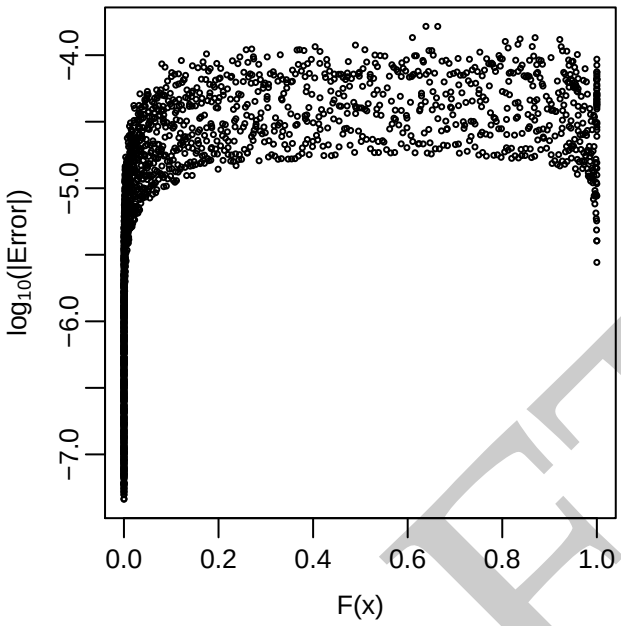

Figure 10.4: The 95^{th} percentile of the absolute error, versus the lower-tail probability.

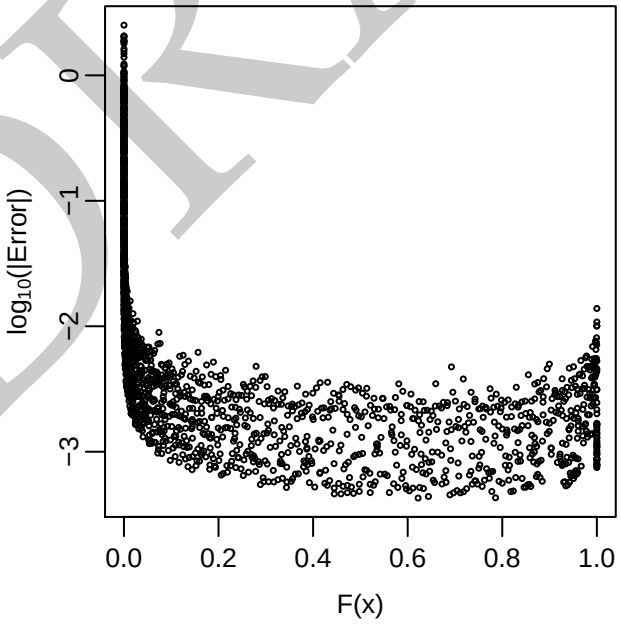

Figure 10.5: The 95^{th} percentile of the relative error, versus the lower-tail probability.

Absolute errors of the $\widehat{p}$

Range of $F(x)$	Median percentile	Maximum percentile
$(0, 10^{-6}]$	8.92×10^{-8}	2.64×10^{-7}
$(10^{-6}, 10^{-5}]$	2.30×10^{-7}	7.66×10^{-7}
$(10^{-5}, 10^{-4}]$	7.02×10^{-7}	4.02×10^{-6}
$(10^{-4}, 10^{-3}]$	1.95×10^{-6}	7.49×10^{-6}
$(10^{-3}, 10^{-2}]$	5.77×10^{-6}	2.38×10^{-5}
$(10^{-2}, 10^{-1}]$	1.70×10^{-5}	8.59×10^{-5}
$(10^{-1}, 1]$	3.85×10^{-5}	1.65×10^{-4}

(10.61)

Figure 10.5 looks at the relative errors, i.e., the posterior 95^{th} percentiles of $|\widehat{p} - p|/\widehat{p}$. Now the bulk are in the 0.001 to 0.01 range, but in the lower tail the relative errors become worse. Here is the table:

Relative absolute errors of the $\widehat{p}$

Range of $F(x)$	Median percentile	Maximum percentile
$(0, 10^{-6}]$	0.64194	2.51479
$(10^{-6}, 10^{-5}]$	0.24823	0.79992
$(10^{-5}, 10^{-4}]$	0.08920	0.38993
$(10^{-4}, 10^{-3}]$	0.03038	0.13229
$(10^{-3}, 10^{-2}]$	0.01021	0.04050
$(10^{-2}, 10^{-1}]$	0.00357	0.01589
$(10^{-1}, 1]$	0.00157	0.01386

(10.62)

The next two tables make the same comparisons as above, but on the lower-tail probabilities, the $\widehat{F}(x)$'s, which are typical p-values for testing uniformity of the rank vectors.

Absolute errors of the $\widehat{F}$

Range of $F(x)$	Median percentile	Maximum percentile
$(0, 10^{-6}]$	1.19×10^{-7}	3.56×10^{-7}
$(10^{-6}, 10^{-5}]$	4.17×10^{-7}	1.02×10^{-6}
$(10^{-5}, 10^{-4}]$	1.29×10^{-6}	4.74×10^{-6}
$(10^{-4}, 10^{-3}]$	4.27×10^{-6}	1.18×10^{-5}
$(10^{-3}, 10^{-2}]$	1.36×10^{-5}	4.03×10^{-5}
$(10^{-2}, 10^{-1}]$	4.40×10^{-5}	1.44×10^{-4}
$(10^{-1}, 1]$	9.66×10^{-5}	2.03×10^{-4}

(10.63)

Relative absolute errors of the $\widehat{F}$

Range of $F(x)$	Median percentile	Maximum percentile
$(0, 10^{-6}]$	0.49359	1.93113
$(10^{-6}, 10^{-5}]$	0.12521	0.46193
$(10^{-5}, 10^{-4}]$	0.04087	0.12025
$(10^{-4}, 10^{-3}]$	0.01296	0.03773
$(10^{-3}, 10^{-2}]$	0.00418	0.01026
$(10^{-2}, 10^{-1}]$	0.00124	0.00344
$(10^{-1}, 1]$	0.00022	0.00118

(10.64)

10.6 The mixed moments of the C_a's

For fixed m and $k \leqslant \lfloor m/2 \rfloor + 1$, we use an iterative approach to find an expression for the distribution of the C_a's in (10.28). Recall C_a is the number of the first $k-1$ y_j's that are at least a. Extend the definition to an arbitrary number of y_j's:

$$C_a^{(i)} = \#\{y_1, \ldots, y_i \geqslant a\}, \tag{10.65}$$

so that $C_a = C_a^{(k-1)}$. The joint distribution of $C_a^{(1)}, C_a^{(2)}, \ldots, C_a^{(k-1)}$ conditional on $\boldsymbol{Y} \in \mathcal{Y}_1$ (see (10.18)) is a Markov chain:

$$C_a^{(i)} \,|\, C_a^{(1)}, \ldots, C_a^{(i-1)}, y_1 \;\; =^{\mathcal{D}} \;\; C_a^{(i)} \,|\, C_a^{(i-1)}, y_1. \tag{10.66}$$

That is, if one is at stage $i-1$, for $C_a^{(i)}$ it matters only how many of the first $j-1$ are greater than or equal to a, not which ones. Next, note that if $C^{(i-1)} = c$, then $C_a^{(i)}$ can either be c, if $y_i < a$, or $c+1$, if $y_i \geqslant a$. By (10.18), $y_i \in \{1, \ldots, m-k+i\}$. At the i^{th} stage, $i-1$ possible values have been removed, with c of them greater than or equal to a, hence

$$\begin{aligned} P[C_a^{(i)} = x \,|\, C_a^{(i-1)} = c, y_1] &= \begin{cases} \frac{a-1-(i-1-c)}{m-k+1} & \text{if } x = c \\ \frac{m-k+i-(a-1)-c}{m-k+1} & \text{if } x = c+1 \end{cases} \\ &= \begin{cases} \frac{a-i+c}{m-k+1} & \text{if } x = c \\ 1 - \frac{a-i+c}{m-k+1} & \text{if } x = c+1 \end{cases}. \end{aligned} \tag{10.67}$$

This expression can be used to find the distribution of C_a, but here we are concerned with the mean. Now

$$\begin{aligned} E[C_a^{(i)} \,|\, C_a^{(i-1)} = c, y_1] &= c\frac{a-i+c}{m-k+1} + (c-1)\left(1 - \frac{a-i+c}{m-k+1}\right) \\ &= c\left(1 - \frac{1}{\lambda}\right) + 1 - \frac{a-i}{\lambda}, \end{aligned} \tag{10.68}$$

where again $\lambda = m-k+1$. Thus

$$\mu_a^{(i)} \equiv E[C_a^{(i)} \,|\, y_1] = \mu_a^{(i-1)}\left(1 - \frac{1}{\lambda}\right) + 1 - \frac{a-i}{\lambda}, \tag{10.69}$$

and $\mu_a^{(1)} = P[Y_1 \geqslant a \,|\, y_1] = 1 - (a-1)/\lambda$. We can obtain an expression for these means as a sum, but instead use induction. The claim is that

$$\mu_a^{(j)} = j - a + 1 + (a-1)\left(1 - \frac{1}{\lambda}\right)^j. \tag{10.70}$$

It checks for $j = 1$. If it is true for $j = i-1$, then by (10.69),

$$\begin{aligned}\mu_a^{(i)} &= \left(i - a + (a-1)\left(1-\frac{1}{\lambda}\right)^{i-1}\right)\left(1-\frac{1}{\lambda}\right) + 1 - \frac{a-i}{\lambda} \\ &= (i-a)\left(1-\frac{1}{\lambda}\right) + 1 - \frac{a-i}{\lambda} + (a-1)\left(1-\frac{1}{\lambda}\right)^{i} \\ &= i - a + 1 + (a-1)\left(1-\frac{1}{\lambda}\right)^{i}, \end{aligned} \tag{10.71}$$

as desired. Setting $i = k-1$ proves Lemma 10.4.

For the second moments, we fix $1 \leqslant a < b \leqslant k-1$. We again use the Markov approach, but with pairs $(C_a^{(i)}, C_b^{(i)})$. Note that since $a < b$, $C_a^{(i)} \geqslant C_b^{(i)}$. If for $0 \leqslant d \leqslant c \leqslant i-1$, $C_a^{(i-1)} = c$ and $C_b^{(i-1)} = d$, then

$$(C_a^{(i)}, C_b^{(i)}) = \begin{cases} (c,d) & \text{if } y_i \in \{1,\ldots,a-1\} \\ (c+1,d) & \text{if } y_i \in \{a,\ldots,b-1\} \\ (c+1,d+1) & \text{if } y_i \in \{b,\ldots,m-k+i\} \end{cases}. \tag{10.72}$$

At stage i, there are again $m-k+i$ values for y_1 left to choose from, and

$$\begin{aligned} \#\text{ left in } \{1,\ldots,a-1\} &= a-1-(i-1-c) = a-i+c, \\ \#\text{ left in } \{a,\ldots,b-1\} &= b-1-(a-1)-(c-d) = b-a-c+d, \text{ and} \\ \#\text{ left in } \{b,\ldots,m-k+i\} &= m-k+i-(b-1)-d = m-k+1+i-b-d. \end{aligned} \tag{10.73}$$

Thus

$$\begin{aligned} &E[C_a^{(i)}C_b^{(i)} \mid (C_a^{(i-1)}, C_b^{(i-1)}) = (c,d), y_1] \\ &\quad = cd\frac{a-i+c}{\lambda} + (c+1)d\frac{b-a-c+d}{\lambda} + (c+1)(d+1)\frac{\lambda+i-b-d}{\lambda} \\ &\quad = cd\left(1-\frac{2}{\lambda}\right) + d\left(1-\frac{a-i+1}{\lambda}\right) + (c+1)\left(1-\frac{b-i}{\lambda}\right). \end{aligned} \tag{10.74}$$

We now have a recurrence relationship as in (10.69) but including the mixed moments $\mu_{ab}^{(i)} = E[C_a^{(i)}C_b^{(i)} \mid y_1]$:

$$\begin{pmatrix} \mu_{ab}^{(i)} \\ \mu_a^{(i)} \\ \mu_b^{(i)} \end{pmatrix} = \begin{pmatrix} 1-\frac{2}{\lambda} & 1-\frac{b-i}{\lambda} & 1-\frac{a-i+1}{\lambda} \\ 0 & 1-\frac{1}{\lambda} & 0 \\ 0 & 0 & 1-\frac{1}{\lambda} \end{pmatrix} \begin{pmatrix} \mu_{ab}^{(i-1)} \\ \mu_a^{(i-1)} \\ \mu_b^{(i-1)} \end{pmatrix} + \begin{pmatrix} 1-\frac{b-i}{\lambda} \\ 1-\frac{a-i}{\lambda} \\ 1-\frac{b-i}{\lambda} \end{pmatrix}. \tag{10.75}$$

We'll use induction to show that

$$\begin{aligned} E[C_a^{(i)}C_b^{(i)} \mid y_1] &= (i-a+1)(i-b+1) + \\ &\quad ((a+b-2)i-(a-1)(2b-3))\left(1-\frac{1}{\lambda}\right)^{i} + (a-1)(b-2)\left(1-\frac{2}{\lambda}\right)^{i}, \end{aligned} \tag{10.76}$$

It works for $i = 0$. Inserting the expressions (10.76) and (10.71) with $i-1$ into (10.75), we can obtain $\mu_{ab}^{(i)} = A + B + C$, where

$$\begin{aligned}
A &= \left(1-\frac{2}{\lambda}\right)(i-a)(i-b) + \left(1-\frac{b-i}{\lambda}\right)(i-a) + \left(1-\frac{a-i+1}{\lambda}\right)(i-b) + \left(1-\frac{b-i}{\lambda}\right); \\
B &= \left(1-\frac{1}{\lambda}\right)^{i-1}\left[\left(1-\frac{2}{\lambda}\right)((a+b-2)(i-1)-(a-1)(2b-3))\right. \\
&\qquad \left. + \left(1-\frac{b-i}{\lambda}\right)(a-1) + \left(1-\frac{a-i+1}{\lambda}\right)(b-1)\right]; \text{ and} \\
C &= (a-1)(b-2)\left(1-\frac{2}{\lambda}\right)^{i}. \qquad (10.77)
\end{aligned}$$

Part C is indeed the last term in (10.76). Part A is

$$\begin{aligned}
A &= (i-a)(i-b) + (i-a) + (i-b) + 1 \\
&\quad - \frac{2(i-a)(i-b) + (b-i)(i-a) + (a-i+1)(i-b) + b - i}{\lambda} \\
&= (i-a+1)(i-b+1) + \frac{0}{\lambda}, \qquad (10.78)
\end{aligned}$$

which is the first term in (10.76). Consider the term in the square brackets of B. Call it B^*, and write it factoring out the i and λ, i.e.,

$$\begin{aligned}
B^* &= -(a+b-2) - (a-1)(2b-3) + (a-1) + (b-1) + i(a+b-2) \\
&\quad + \frac{2(a+b-2) + 2(a-1)(2b-3) - b(a-1) - (a+1)(b-1)}{\lambda} \\
&\quad + i\frac{-2(a+b-2) + a - 1 + b - 1}{\lambda} \\
&= -(a-1)(2b-3) + i(a+b-2) + \frac{(a-1)(2b-3) - i(a+b-2)}{\lambda} \\
&= ((a+b-2)i - (a-1)(2b-3))\left(1-\frac{1}{\lambda}\right). \qquad (10.79)
\end{aligned}$$

Maybe nonobvious:

$$\begin{aligned}
2(a+b-2) - b(a-1) - (a+1)(b-1) &= 2a + 2b - 4 - ab + b - ab + a - b + 1 \\
&= -2ab + 3a + 2b - 3 = -(a-1)(2b-3). \qquad (10.80)
\end{aligned}$$

Thus

$$B = ((a+b-2)i - (a-1)(2b-3))\left(1-\frac{1}{\lambda}\right)^{i}, \qquad (10.81)$$

the middle term of (10.76), completing the induction.

10.6.1 Third-degree mixed moments

The approach we take for higher moments is to guess the form, then use Mathematica® to find the coefficients. Consider the q^{th} mixed moment,

$$\mu^{(i)}_{a_1,\ldots,a_q} = E[C^{(i)}_{a_1} \cdots C^{(i)}_{a_q} \,|\, \mathcal{Y}_1], \quad \text{where} \quad a_1 < a_2 < \cdots < a_q, \tag{10.82}$$

where $\mathcal{Y}_1$ was defined in (10.16) and (10.18). The equation analog to (10.76) is arranged according to the factors $(1-j/\lambda)^i$:

$$\begin{aligned}
\mu^{(i)}_{a_1\ldots,a_q} &= (i-a_1+1)(i-a_2+1)\cdots(i-a_q+1) \\
&\quad + (x^{(1)}_0 + x^{(1)}_1 i + x^{(1)}_2 i^2 + \cdots + x^{(1)}_{q-1} i^{q-1})(1-1/\lambda)^i \\
&\quad + (x^{(2)}_0 + x^{(2)}_1 i + \cdots + x^{(2)}_{q-2} i^{q-2})(1-2/\lambda)^i \\
&\quad + \cdots + (x^{(q-1)}_0 + x^{(q-1)}_1 i)(1-(q-1)/\lambda)^i \\
&\quad + (a_1-1)(a_2-2)\cdots(a_q-q)(1-q/\lambda)^i.
\end{aligned} \tag{10.83}$$

That is, the form is

$$\mu^{(i)}_{a_1,\ldots,a_q} = \sum_{j=0}^{q} P^{(q)}_j(i;\, a_1,\ldots,a_q)(1-j/\lambda)^i, \tag{10.84}$$

where $P^{(q)}_j(i;\, a_1,\ldots,a_q)$ is a $(q-j)^{\text{th}}$-degree polynomial in i, with coefficients depending on just $a_1,\ldots, a_q$, i.e., not on λ. We happen to know the first and last terms.

The steps for finding the $x^{(j)}_l$'s are

1. Find the recurrence as in (10.75), where $\mu^{(i)}_{a_1,\ldots,a_q}$ is a linear function of all the $\mu^{(i-1)}$ for all the subsets of $\{a_1,\ldots,a_q\}$.

2. Use the representations in (10.83) to obtain an expression for $\mu^{(i)}_{a_1,\ldots,a_q}$ that collects terms according to the $(1-j/\lambda)^{i-1}$'s.

3. Equate each the term multiplying the $(1-j/\lambda)$'s in (10.83) to $(1-j/\lambda)$ times the terms multiplying $(1-j/\lambda)^{i-1}$ from step 2. From these equations, we can solve for the $x^{(j)}_l$'s.

4. Check that the resulting formula yields $\mu^{(0)}_{a_1,\ldots,a_q} = 0$. Then induction and step 1 will verify the result.

We illustrate with $q = 3$, letting $(a_1, a_2, a_3) = (a, b, c)$, with $a < b < c$. To find the recurrence, we first find the conditional distribution

$$(C^{(i)}_a, C^{(i)}_b, C^{(i)}_c) \mid (C^{(i-1)}_a, C^{(i-1)}_b, C^{(i-1)}_c) = (u, v, w), \mathcal{Y}_1. \tag{10.85}$$

(Recall the definition of $C_a^{(i)}$ in (10.65) and y_1 in (10.16) and (10.18).) Denote the conditional probabilities as

$$P\left[(C_a^{(i)}, C_b^{(i)}, C_c^{(i)}) = \left\{\begin{array}{l}(u,v,w)\\(u+1,v,w)\\(u+1,v+1,w)\\(u+1,v+1,w+1)\end{array}\right\} \mid (C_a^{(i-1)}, C_b^{(i-1)}, C_c^{(i-1)}) = (u,v,w), y_1\right] \tag{10.86}$$

$$= \left\{\begin{array}{l}p_0\\p_1\\p_2\\p_3\end{array}\right\} = P\left[Y_i \in \left\{\begin{array}{l}\{0,\ldots,a-1\}\\\{a,\ldots,b-1\}\\\{b,\ldots,c-1\}\\\{c,\ldots,m-k+i\}\end{array}\right\} \mid (C_a^{(i-1)}, C_b^{(i-1)}, C_c^{(i-1)}) = (u,v,w), y_1\right]. \tag{10.87}$$

The relevant conditional expectation is then

$$\begin{aligned}E[C_a^{(i)}C_b^{(i)}C_c^{(i)} \mid (C_a^{(i-1)}, C_b^{(i-1)}, C_c^{(i-1)}) = (u,v,w), y_1]&\\ = uvwp_0 + (u+1)vwp_1 + (u+1)(v+1)wp_2 + (u+1)(v+1)(w+1)p_3&\\ = uvw + vwQ_a + (uw+w)Q_b + (uv+u+v+1)Q_c,&\\ \text{where } Q_r = P[Y_1 \geqslant r \mid (C_a^{(i-1)}, C_b^{(i-1)}, C_c^{(i-1)}) = (u,v,w), y_1].&\end{aligned} \tag{10.88}$$

To find Q_a, note that by definition of y_1 in (10.18), Y_i must be choosing from the values $\{1,\ldots,m-k+i\}$. Since we are at the i^{th} stage, already $i-1$ have been chosen, leaving $\lambda \equiv m-k+1$ to choose from. Initially there are $m-k+i-(a-1)$ that are greater than or equal to a, but since $C^{(i-1)} = u$, u of them have been chosen, leaving $m-k+i-a+1-u$. Thus $Q_a = (m-k+i-a+1-u)/\lambda$. Similarly for the others, or taking from 1:

$$Q_a = 1 - \frac{a+u-i}{\lambda},\ Q_b = 1 - \frac{b+v-i}{\lambda},\ \text{and } Q_c = 1 - \frac{c+w-i}{\lambda}. \tag{10.89}$$

Inserting these formulas into (10.88) and collecting products of the u, v, w, we find the conditional probability to be

$$uvw\left(1-\frac{3}{\lambda}\right) + vw\left(1-\frac{a+2-i}{\lambda}\right) + (uw+w)\left(1-\frac{b+1-i}{\lambda}\right) + (uv+u+v+1)\left(1-\frac{c-i}{\lambda}\right). \tag{10.90}$$

To complete step 1, we take expectations in (10.88), obtaining the recurrence

$$\mu_{abc}^{(i)} = \mu_{abc}^{(i-1)}\left(1-\frac{3}{\lambda}\right) + \mu_{bc}^{(i-1)}\left(1-\frac{a+2-i}{\lambda}\right) + (\mu_{ac}^{(i-1)} + \mu_c^{(i-1)})\left(1-\frac{b+1-i}{\lambda}\right) + (\mu_{ab}^{(i-1)} + \mu_a^{(i-1)} + \mu_b^{(i-1)} + 1)\left(1-\frac{c-i}{\lambda}\right). \tag{10.91}$$

For step 2, we insert the expressions (10.83) into (10.91) for the various μ's, and collect terms according to $(1-j/\lambda)^{i-1}$. To start, we find the coefficients for those quantities. For $j=3$, the coefficient is 0 for all but the "abc" means, whose coefficient is

$$(a-1)(b-2)(c-3). \tag{10.92}$$

The expression for sets of two is in (10.77), and for single values is in (10.71):

	Constant	$\left(1-\frac{1}{\lambda}\right)^{i-1}$	$\left(1-\frac{2}{\lambda}\right)^{i-1}$
$\mu_{abc}^{(i-1)}$	$(i-a)(i-b)(i-c)$	$x_0+x_1(i-1)+x_2(i-1)^2$	$y_0+y_1(i-1)$
$\mu_{ab}^{(i-1)}$	$(i-a)(i-b)$	$-(a-1)(2b-3)+(a+b-2)(i-1)$	$(a-1)(b-2)$
$\mu_{ac}^{(i-1)}$	$(i-a)(i-c)$	$-(a-1)(2c-3)+(a+c-2)(i-1)$	$(a-1)(c-2)$
$\mu_{bc}^{(i-1)}$	$(i-b)(i-c)$	$-(b-1)(2c-3)+(b+c-2)(i-1)$	$(b-1)(c-2)$
$\mu_{a}^{(i-1)}$	$i-a$	$(a-1)$	0
$\mu_{b}^{(i-1)}$	$i-b$	$(b-1)$	0
$\mu_{c}^{(i-1)}$	$i-c$	$(c-1)$	0
1	1	0	0

(10.93)

We then use the above to find the coefficients of the $(1-j/\lambda)^i$'s on both sides of (10.91). Step 3 is then to equate the coefficients from the two sides. For the $j=3$ it is easy, since the abc terms both have the coefficient $(a-1)(b-2)(c-3)$, and the lower-order terms' coefficients are 0. For the constant term, we seem to have the answer already, but for completeness we need to verify the value. The right-hand side's constant is

$$\begin{aligned}&(i-a)(i-b)(i-c)\left(1-\frac{3}{\lambda}\right)+(i-b)(i-c)\left(1-\frac{a+2-i}{\lambda}\right)\\&+((i-a)(i-c)+i-c)\left(1-\frac{b+1-i}{\lambda}\right)+((i-a)(i-b)+i-a+i-b+1)\left(1-\frac{c-i}{\lambda}\right).\end{aligned} \tag{10.94}$$

Tediously writing things out, or, better, using Mathematica®, we find that the expression in (10.94) reduces to $(i-a+1)(i-b+1)(i-c+1)$, which is indeed the constant in (10.83).

Now consider the $(1-2/\lambda)^{i-1}$ terms. The coefficient on the left-hand side of (10.91) is, by (10.83),

$$(y_0+y_1 i)\left(1-\frac{2}{\lambda}\right). \tag{10.95}$$

On the right-hand side we have

$$\begin{aligned}&(y_0+y_1(i-1))\left(1-\frac{3}{\lambda}\right)+(b-1)(c-2)\left(1-\frac{a+2-i}{\lambda}\right)\\&\qquad+(a-1)(c-2)\left(1-\frac{b+1-i}{\lambda}\right)+(a-1)(b-2)\left(1-\frac{c-i}{\lambda}\right).\end{aligned} \tag{10.96}$$

Subtract (10.95) from (10.96), then find the y_0 and y_1 that sets the difference to zero. The coefficient of i in the difference is

$$\frac{1}{\lambda}((b-2)(c-3)+a(b+c-4)-y_1), \tag{10.97}$$

which means that

$$y_1 = (b-2)(c-3)+a(b+c-4). \tag{10.98}$$

Now if we substitute that y_1 into the difference and set to zero, we can solve for y_0, which turns out to be

$$y_0 = -(a-1)(12-5c+b(3c-7)). \tag{10.99}$$

We do the same approach for $(1-1/\lambda)$. The coefficient on the left-hand side of (10.91) is

$$(x_0+x_1 i+x_2 i^2)\left(1-\frac{1}{\lambda}\right). \tag{10.100}$$

The right-hand side's is

$$\begin{aligned}(x_0+x_1(i-1)+x_2(i-1)^2)\left(1-\frac{3}{\lambda}\right)&+(-(b-1)(2c-3)+(b+c-2)(i-1))\left(1-\frac{a+2-i}{\lambda}\right)\\&+(-(a-1)(2c-3)+(a+c-2)(i-1)+c-1)\left(1-\frac{b+1-i}{\lambda}\right)\\&+(-(a-1)(2b-3)+(a+b-2)(i-1)+a-1+b-1)\left(1-\frac{c-i}{\lambda}\right).\end{aligned} \tag{10.101}$$

Equating (10.100) and (10.102), and solving for the x_j's, we find

$$\begin{aligned}x_0 &= (a-1)(7-4c+b(3c-5)),\\ x_1 &= 4c-9-2a(b+c-3)-b(2c-5),\\ x_2 &= a+b+c-3.\end{aligned} \tag{10.102}$$

Thus with the y_j's and x_j's in (10.98), (10.99), and (10.102), we have

$$\begin{aligned}\mu^{(i)}_{abc} = (i-a+1)(i-b+1)(i-c+1)&+(x_0+x_1 i+x_2 i^2)\left(1-\frac{1}{\lambda}\right)^i\\&+(y_0+y_1 i)\left(1-\frac{2}{\lambda}\right)^i+(a-1)(b-2)(c-3)\left(1-\frac{3}{\lambda}\right)^i.\end{aligned} \tag{10.103}$$

Step 4 is to verify that setting $i=0$ in (10.103) yields zero, which it does, proving the equation.

10.6.2 General q

Consider finding the $x_l^{(j)}$'s in (10.83) for arbitrary q, assuming we have the results (10.82) for values less than q. We condition on

$$\boldsymbol{C}^{(i-1)} \equiv (C_{a_1}^{(i-1)},\ldots,C_{a_q}^{(i-1)}) = (u_1,\ldots,u_q) \equiv \boldsymbol{u}, \tag{10.104}$$

so that

$$\boldsymbol{C}^{(i)} \,|\, \boldsymbol{C}^{(i-1)} = \boldsymbol{u} \;\in\; \{(u_1,\ldots,u_q),(u_1+1,u_2,\ldots,u_q), \\ (u_1+1,u_2+1,u_3,\ldots,u_q),\ldots,(u_1+1,u_2+1,\cdots,u_q+1)\}. \quad (10.105)$$

Let

$$p_j = P\left[\boldsymbol{C}^{(i)} = (u_1+1,\ldots,u_j+1,u_{j+1},\ldots,u_q) \,|\, \boldsymbol{C}^{(i-1)} = \boldsymbol{u}, \mathcal{Y}_1\right], j = 0,\ldots,q. \quad (10.106)$$

After multiplying out the (u_j+1) terms, the conditional expectation can be written as a sum of all possible products of the elements $\{1,u_1,\ldots,u_q\}$, times a sum of some of the p_h's:

$$E[\boldsymbol{C}^{(i)} \,|\, \boldsymbol{C}^{(i-1)} = \boldsymbol{u}, \mathcal{Y}_1] = \sum_{j=0}^{q} \sum_{1\leqslant l_1<\cdots<l_j\leqslant q} u_{l_1}u_{l_2}\cdots u_{l_j} Q_r, \quad r = r(\{l_1,\ldots,l_j\}). \quad (10.107)$$

Here

$$Q_r = p_r + \cdots + p_q, \quad (10.108)$$

and the function r is the largest integer ($0 \leqslant r \leqslant q$) not in the subset $\{l_1,\ldots,l_j\}$:

$$r(\mathcal{I}) = \max\{r \in \{0,\ldots,q\} \,|\, r \notin \mathcal{I}\} \text{ for } \mathcal{I} \subset \{1,\ldots,q\}. \quad (10.109)$$

For example, if $q = 5$, then we have the following:

$\{l_1,\ldots,l_j\}$	r
$\{1,2,3,4,5\}$	0
$\{1,2,4,5\}$	3
$\{3,4,5\}$	2
$\{4,5\}$	3
$\{1,3\}$	5
$\emptyset$	5

(10.110)

Now

$$Q_r = P[Y_i \geqslant a_r \,|\, \boldsymbol{C}^{(i-1)} = \boldsymbol{u}] = \frac{m-k+i-(a_r-1)-u_r}{m-k+i-(i-1)} = 1 - \frac{a_r+u_r-i}{\lambda}, \quad \lambda = m-k+1. \quad (10.111)$$

By $\mathcal{Y}_1$ in (10.18), $Y_i \leqslant m-k+i$, and after the $(i-1)^{st}$ stage, there have been $i-1$ removed, hence there are $m-k+1$ left to choose from. There are $m-k+i-(a_r-1)$ values at least a_r for the numerator, but since $C_{a_r}^{(i-1)} = u_r$, there have been u_r already taken. Thus we obtain (10.111).

Place the Q_r's into (10.107). Since the u_l's appear in the Q_r's, we have to rework the formula so that the products of u_l's are all together. Studying (10.91), it appears that the pattern is

$$E[\boldsymbol{C}^{(i)} \,|\, \boldsymbol{C}^{(i-1)} = \boldsymbol{u}, \mathcal{Y}_1] = u_1\cdots u_q\left(1-\frac{q}{\lambda}\right) + \sum_{j=0}^{q-1} \sum_{1\leqslant l_1<\cdots<l_j\leqslant q} u_{l_1}\cdots u_{l_j}\left(1-\frac{a_r+q-r-i}{\lambda}\right), \quad (10.112)$$

for $r = r(\{l_1, \ldots, l_j\})$ in (10.109).

Now taking expectations in (10.112), we have

$$\mu^{(i)}_{a_1,\ldots,a_q} = \mu^{(i-1)}_{a_1,\ldots,a_q}\left(1-\frac{q}{\lambda}\right) + \sum_{j=0}^{q-1}\sum_{1\leqslant l_1<\cdots<l_j\leqslant q}\mu^{(i-1)}_{a_{l_1},\ldots,a_{l_j}}\left(1-\frac{a_r+q-r-i}{\lambda}\right). \tag{10.113}$$

To find the formulas in (10.83), we have to pick out the coefficients of each $(1-j/\lambda)^{(i-1)}$ $(j = 1, \ldots, q-1)$ for each $\mu^{(i-1)}$, multiply by the appropriate $(1-(a_r+q-r-i)/\lambda)$, and sum them. That is, using the polynomials $P^{(q)}_j$ in (10.84), we have for each j,

$$\begin{aligned} P^{(q)}_j(i\,;\,a_1,\cdots,a_q)(1-j/\lambda) &= P^{(q)}_j(i-1\,;\,a_1,\cdots,a_q)(1-q/\lambda) \\ &\quad + \sum_{h=j}^{q-1}\sum_{1\leqslant l_1<\cdots<l_h\leqslant q} P^{(j)}_h(i-1\,;\,a_{l_1},\ldots,a_{l_h})\left(1-\frac{a_r+q-r-i}{\lambda}\right). \end{aligned} \tag{10.114}$$

Both sides of (10.114) are polynomials in i of degree $q-j$ (since the $P^{(j)}_h$'s in the summation are of degree $h-j$, $h<q$). Given we know the $P^{(j)}_h$ for $j<q$, the only unknowns in (10.114) are the coefficients for $P^{(q)}_j$. Thus we can solve for them, obtaining $P^{(q)}_j$.

10.6.3 Central moments

Taking $i = k-1$, the moments we have found are the raw mixed moments,

$$\mu_{a_1,\ldots,a_q} = E[C_{a_1}\cdots C_{a_q}\,|\,y_1], \tag{10.115}$$

but in order to find the s_l's in (10.36), we need the central mixed moments. E.g., for $q=2$, $a<b$, from (10.38) we obtain

$$\begin{aligned} \mathrm{Cov}[C_a, C_b\,|\,y_1] &= E[W_aW_b\,|\,y_1] \\ &= (a-1)\left(\left(1-\frac{1}{\lambda}\right)^{k-1} + (b-2)\left(1-\frac{2}{\lambda}\right)^{k-1} - (b-1)\left(1-\frac{1}{\lambda}\right)^{2k-2}\right). \end{aligned} \tag{10.116}$$

In general they can be found from the raw moments using

$$\begin{aligned} E[W_{a_1}\cdots W_{a_q}\,|\,y_1] &= E[(C_{a_1}-\mu_{a_1})\cdots(C_{a_q}-\mu_{a_q})\,|\,y_1] \\ &= \sum_{j=0}^{q}(-1)^{q-j}\sum_{1\leqslant l_1<\cdots<l_j\leqslant q}\left(\mu_{a_{l_1},\ldots,a_{l_j}}\prod_{i\notin\{l_1,\ldots,l_j\}}\mu_{a_i}\right) \\ &= \sum_{j=2}^{q}(-1)^{q-j}\sum_{1\leqslant l_1<\cdots<l_j\leqslant q}\left(\mu_{a_{l_1},\ldots,a_{l_j}}\prod_{i\notin\{l_1,\ldots,l_j\}}\mu_{a_i}\right) + (-1)^{q-1}(q-1)\prod_{i=1}^{q}\mu_{a_i} \end{aligned} \tag{10.117}$$

10.7 Asymptotics

In this section we find the asymptotic distribution of D_{Max} as in (10.3). It turns out not to be a great approximation for even fairly large m, so we recommend the approximation given at the end of Section 10.4, although the first-order approximation from (10.33) is almost as easy to apply. The main result for this section is next.

Proposition 10.5. *For $D_{Max} = d_{Max}(\boldsymbol{Y}, \boldsymbol{\omega})$, where $\boldsymbol{Y} \sim$ Uniform$(\mathcal{P}_m)$, as $m \to \infty$,*

$$P\left[\frac{m - D_{Max}}{\sqrt{m}} \leqslant x\right] \longrightarrow 1 - e^{-x^2}, \quad x > 0. \tag{10.118}$$

Proof. Fix $x > 0$, and let $k \equiv k_m$ be a sequence of positive integers such that $k/\sqrt{m} \to x$. Now for large enough m, $k < \lfloor \frac{m}{2} \rfloor + 1$, hence we obtain from Proposition 10.3, i.e., (10.29),

$$\begin{aligned} P\left[\frac{m - D_{\text{Max}}}{\sqrt{m}} \geqslant \frac{k}{\sqrt{m}}\right] &= P[D_{\text{Max}} \leqslant m - k] \\ &= \frac{E[\prod_{a=2}^{k}(m - k + 1 - C_a) \,|\, Y_1] \times (m - k + 1)^{k-1}}{(m)_{2k-2}}. \end{aligned} \tag{10.119}$$

Recall from (10.24) that C_a is the number of $Y_1, \ldots, Y_{k-1}$ that are greater than or equal to a, so that for $2 \leqslant a \leqslant k$, $C_k \leqslant C_a \leqslant k - 1$. Letting $V_k = k - 1 - C_k$, we have

$$\begin{aligned} (m - 2k + 2)^{k-1} &\leqslant E\left[\prod_{a=2}^{k}(m - k + 1 - C_a) \,|\, Y_1\right] \\ &\leqslant E[(m - k + 1 - C_k)^{k-1} \,|\, Y_1] \\ &= E[(m - 2k + 2 + V_k)^{k-1} \,|\, Y_1] \\ &= (m - 2k + 2)^{k-1} E\left[\left(1 + \frac{V_k}{m - 2k + 2}\right)^{k-1} \,|\, Y_1\right]. \end{aligned} \tag{10.120}$$

Hence by (10.119) and (10.120), we have the bounds

$$P_{k,m} Q_{k,m} \leqslant P[D_{\text{Max}} \leqslant m - k] \leqslant P_{k,m} Q_{k,m} R_{k,m}, \tag{10.121}$$

where, since $(m)_{2k-2} = (m)_{k-1}(m - k + 1)_{k-1}$,

$$P_{k,m} = \frac{(m - 2k + 2)^{k-1}}{(m - k + 1)_{k-1}}, \quad Q_{k,m} = \frac{(m - k + 1)^{k-1}}{(m)_{k-1}}, \tag{10.122}$$

and

$$R_{k,m} = E\left[\left(1 + \frac{V_k}{m - 2k + 2}\right)^{k-1} \,|\, Y_1\right]. \tag{10.123}$$

Below we show that as $m \to \infty$,

$$P_{k,m} \longrightarrow e^{-\frac{1}{2}x^2}, \quad Q_{k,m} \longrightarrow e^{-\frac{1}{2}x^2}, \quad \text{and} \quad R_{k,m} \longrightarrow 1. \tag{10.124}$$

Thus by (10.121), $P[D_{\text{Max}} \leqslant m - k] \to \exp(-x^2)$, proving (10.118).

For the first two components, use the following lemma, proven after this proof.

Lemma 10.6. *Suppose* $l/\sqrt{n} \to x$ *as* $n \to \infty$. *Then*

$$\frac{(n-l)^l}{(n)_l} \longrightarrow e^{-\frac{1}{2}x^2}. \tag{10.125}$$

Proof. Now apply the lemma to $P_{k,m}$ and $Q_{k,m}$ in (10.122), with $(n,l) = (m-k+1, k-1)$ and $(n,l) = (m, k-1)$, respectively. Since $k/\sqrt{m} \to x$, in both cases, $l/\sqrt{n} \to x$. Thus the first two results in (10.124) hold.

Turn to $R_{k,m}$ in (10.123). We will show that for any $\epsilon > 0$, there exists a finite T such that

$$P[V_k \geqslant T \,|\, \mathcal{Y}_1] \leqslant \epsilon \text{ for sufficiently large } m. \tag{10.126}$$

(That is, the sequence V_k is bounded in probability.) Fix such an ϵ, and find the T as in (10.126). Then since $V_k \leqslant k-1$,

$$\begin{aligned} R_{k,m} &= E\left[\left(1+\frac{V_k}{m-2k+2}\right)^{k-1} I[V_k < T] \,|\, \mathcal{Y}_1\right] + E\left[\left(1+\frac{V_k}{m-2k+2}\right)^{k-1} I[V_k \geqslant T] \,|\, \mathcal{Y}_1\right] \\ &\leqslant \left(1+\frac{T}{m-2k+2}\right)^{k-1} + \left(1+\frac{k-1}{m-2k+2}\right)^{k-1} \epsilon \end{aligned} \tag{10.127}$$

for large m. Let $m \to \infty$, with $k/\sqrt{m} \to x$. The first term in the last line of (10.127) goes to 1, since T is fixed and $(k-1)/(m-2k+2) \to 0$, and the second goes to $\exp(x^2)\epsilon$ since $(k-1)^2/(m-2k+2) \to x^2$. Thus

$$\limsup_{m\to\infty} R_{k,m} \leqslant 1 + e^{x^2}\epsilon. \tag{10.128}$$

Since ϵ is arbitrary, and $R_{k,m} \geqslant 1$, we have completed (10.124).

All that is left is showing (10.126). Since $V_k \geqslant 0$, we can use Markov's inequality, $P[V_k \geqslant T \,|\, \mathcal{Y}_1] \leqslant E[V_k \,|\, \mathcal{Y}_1]/T$. Then by Lemma 10.4 with $a = k$ in (10.37), recalling $V_k = k-1-C_k$, we have

$$P[V_k \geqslant T \,|\, \mathcal{Y}_1] \leqslant \frac{1}{T}(k-1)\left(1-\left(1-\frac{1}{m-k+1}\right)^{k-1}\right). \tag{10.129}$$

Use the expansion $(1-z)^l = 1 - lz + (l(l-1)z^2/2)(1-z^*)^{l-2}$, $|z^*| \leqslant |z|$. With $l = k-1$ and $z = 1/(m-k+1)$, (10.129) yields

$$\begin{aligned} \limsup_{m\to\infty} P[V_k \geqslant T \,|\, \mathcal{Y}_1] &\leqslant \frac{1}{T} \limsup_{m\to\infty} \left(\frac{(k-1)^2}{m-k+1} - \frac{(k-1)^2(k-2)}{2(m-k+1)^2}(1-z^*)^{k-3}\right) \\ &= \frac{x^2}{T}. \end{aligned} \tag{10.130}$$

Then for given $\epsilon > 0$ in (10.126), we can take $T = (x^2+1)/\epsilon$. □

Proof of Lemma 10.6. Recall Stirling's approximation,

$$n! = s(n)\left(1+O\left(\frac{1}{n}\right)\right), \quad \text{where} \quad s(n) = \sqrt{2\pi}\, n^{n+\frac{1}{2}} e^{-n}. \tag{10.131}$$

Then

$$\frac{(n-l)^l}{(n)_l} = \frac{(n-l)^l(n-l)!}{n!} = t(n)\left(1+O\left(\frac{1}{n}\right)\right), \tag{10.132}$$

where

$$t(n) = \frac{(n-l)^l(n-l)^{n-l+\frac{1}{2}}e^{-(n-l)}}{n^{n+\frac{1}{2}}e^{-n}} = \left(1-\frac{l}{n}\right)^n e^l\sqrt{1-\frac{l}{n}}. \tag{10.133}$$

As $n \to \infty$, the term in the square root goes to 1. Consider the log of the rest, and use the Taylor expansion $\log(1-z) = -(z+z^2/2+(z^*)^3/3)$, $|z^*| \leqslant |z|$:

$$\begin{aligned} n\log\left(1-\frac{l}{n}\right)+l &= -n\left(\frac{l}{n}+\frac{1}{2}\frac{l^2}{n^2}+\frac{1}{3}(z^*)^3\right)+l \\ &= -\frac{1}{2}\frac{l^2}{n}+\frac{n}{3}(z^*)^3, \ \ |n(z^*)^3| \leqslant \frac{l^3}{n^2}. \end{aligned} \tag{10.134}$$

Since $l/\sqrt{n} \to x$ by assumption, the final equality in (10.134) goes to $-x^2/2$, proving (10.125). □

Chapter 11

Covariances of some of the distances

In Section 1.6 we summarized results about the correlations among the distances. This chapter provides the details. Section 11.1 finds the exact covariances among five of the distances. Section 11.2 presents some simulations for the covariances we could not deal with analytically.

11.1 Exact covariances of five of the distances

In this section we find the exact covariances of the Spearman, footrule, Kendall, Hamming, and Cayley distances. Ulam and Maximum have proven very difficult to deal with analytically. David, Kendall, & Stuart (1951) find the first-order (in $1/m$) correlation between Spearman's ρ and Kendall's distance assuming normal samples with correlation r. We are restricting to $\boldsymbol{Y} \sim \text{Uniform}(\mathcal{P}_m)$, which is equivalent to their setting $r = 0$. Table (11.1) contains the covariances.

	Footrule	Kendall	Hamming	Cayley
Spearman	$m(m+1)(m^2+1)/30$	$m(m-1)(m+1)^2/36$	$m(m+1)/6$	$(m^2-1)/6$
Footrule		$(m+1)(m^2+1)/30$	$(m+1)/3$	$(m^2-1)/3m$
Kendall			$(m+1)/6$	$(2m-1)/12$
Hamming				$1-1/m$

(11.1)

All but the Kendall/Cayley covariance can be handled by at least one of the following formulas, where D_A is a Hoeffding distance (Chapter 3) with $\delta_A(i,j)$ and Δ_A as in (3.1) and (3.4):

$$\text{Cov}[D_A, D_\rho] = -\frac{2}{m-1}\,\boldsymbol{\omega}\boldsymbol{H}\boldsymbol{\Delta}_A\boldsymbol{H}\boldsymbol{\omega}', \tag{11.2a}$$

$$\text{Cov}[D_A, D_{\text{Kendall}}] = \frac{1}{m}\,\text{Cov}[D_A, D_\rho], \tag{11.2b}$$

$$\text{Cov}[D_A, D_{\text{Hamming}}] = \frac{1}{m-1}\,E[D_A], \text{ and} \tag{11.2c}$$

$$\text{Cov}[D_A, D_{\text{Cayley}}] = \frac{1}{m}\,E[D_A]. \tag{11.2d}$$

Here, $\boldsymbol{H}$ is the $n \times n$ centering matrix from (3.8). The proofs are given in Section 11.3.

Start with the covariance between the Spearman and footrule distances. By (5.5), we have

$$\boldsymbol{H}\boldsymbol{\Delta}_{\text{Footrule}}\boldsymbol{H} = -2\boldsymbol{H}\boldsymbol{K}\boldsymbol{K}'\boldsymbol{H}, \tag{11.3}$$

where $\boldsymbol{K}$ is the $n \times n$ matrix with ones on and above the diagonal, and zeros below, as in (5.3). Thus (11.2a) shows that

$$\text{Cov}[D_{\text{Foot}}, D_{\text{Spear}}] = \frac{4}{n-1}\,\|\boldsymbol{\omega}\boldsymbol{H}\boldsymbol{K}\|^2. \tag{11.4}$$

Now $\boldsymbol{\omega}\boldsymbol{H} = \boldsymbol{\omega} - (n+1)\mathbf{1}/2$, hence, with help from (5.7),

$$\begin{aligned}(\boldsymbol{\omega}\boldsymbol{H}\boldsymbol{K})_i &= \sum_{j=1}^{i}\left(j - \frac{n+1}{2}\right) = \frac{i(i-n)}{2} \\ \Rightarrow \|\boldsymbol{\omega}\boldsymbol{H}\boldsymbol{K}\|^2 &= \sum_{i=1}^{n}\left(\frac{i(i-n)}{2}\right)^2 = \frac{1}{120}\,n(n^2-1)(n^2+1).\end{aligned} \tag{11.5}$$

Thus (11.4) proves the Spearman/footrule entry in (11.1).

For convenience, we repeat the relevant means and covariances from (1.10):

	Mean	Variance (if $n > 1$)
Spearman	$n(n^2-1)/6$	$n^2(n-1)(n+1)^2/36$
Footrule	$(n^2-1)/3$	$(n+1)(2n^2+7)/45$
Kendall	$n(n-1)/4$	$n(n-1)(2n+5)/72$
Hamming	$n-1$	1
Cayley	$n - \sum_{i=1}^{n} 1/i$	$\sum_{i=1}^{n}(i-1)/i^2$

(11.6)

The covariance of Spearman or footrule with Hamming or Cayley, and Hamming with Cayley, can all be found using (11.2c), (11.2d), and the means from (11.6). The covariance of Kendall's distance with the Spearman, footrule, or Hamming distances is then found using (11.2b) and the previously found covariances (or variance, in the first case) with Spearman.

Finally, consider the Kendall/Cayley covariance. Using the expression for Kendall's distance in (6.4), we have

$$\text{Cov}[D_{\text{Ken}}, D_{\text{Cay}}] = \sum_{1 \leqslant j < i \leqslant n}\sum \text{Cov}[I[Y_i < Y_j], D_{\text{Cay}}]. \tag{11.7}$$

Recall the decomposition of Cayley's distance from Section 9.1, where V_k was the indicator of whether an interchange was needed to set $y_k = k$, proceeding in the order $k = 1, 2, \ldots, n-1$. We could just as well take any other order. For fixed i, j, $j < i$, consider any order that starts with $k = j, i$, so that V_1 is the indicator of whether at the first step we need an interchange to have $y_j = j$, and V_2 is the indicator at the second step of whether an interchange is needed to have $y_i = i$. After these two steps, the further steps $(V_3, \ldots, V_{n-1})$ are independent of V_1 and V_2, and independent of Y_i and Y_j since after the first two interchanges, the remaining elements

of $\boldsymbol{Y}$ are equally likely to be in any order. Also, as in (9.9), $V_k \sim \text{Bernoulli}(1-1/(m-k+1))$ for each k. The independence yields

$$\text{Cov}[I[Y_i < Y_j], D_{\text{Cay}}] = \text{Cov}[I[Y_i < Y_j], \sum_{k=1}^{m} V_k] = \text{Cov}[I[Y_i < Y_j], V_1 + V_2]. \tag{11.8}$$

Now $V_1 = I[Y_j \neq j]$, and V_2 equals one if $Y_i \neq i$ unless the first interchange places an i into the i^{th} slot, that is, unless $Y_i{=}j$ and $Y_j{=}i$. Thus

$$\begin{aligned} V_1 + V_2 &= I[Y_j \neq j] + I[Y_i \neq i] - I[Y_i = j \ \& \ Y_j = i] \\ &= 2 - I[Y_j = j] - I[Y_i = i] - I[Y_i = j \ \& \ Y_j = i]. \end{aligned} \tag{11.9}$$

To find the covariance in (11.8), since $E[I[Y_i < Y_j]] = 1/2$ we have the cross-product

$$\begin{aligned} E[I[Y_i < Y_j](V_1 + V_2)] &= 1 - P[Y_i < Y_j \,|\, Y_i = i]P[Y_i = i] - P[Y_i < Y_j \,|\, Y_j = j]P[Y_j = j] \\ &\quad - P[Y_i = j \ \& \ Y_j = i] \\ &= 1 - \frac{m-i}{m(m-1)} - \frac{j-1}{m(m-1)} - \frac{1}{m(m-1)}. \end{aligned} \tag{11.10}$$

For the final term, note that since $j{<}i$, $Y_i{=}j$ and $Y_j{=}i$ implies that $Y_i{<}Y_j$. Now $E[V_1 + V_2] = 2 - 1/m - 1/(m-1)$, which yields

$$\text{Cov}[I[Y_i < Y_j], V_1 + V_2] = \frac{i - j - \frac{1}{2}}{m(m-1)}. \tag{11.11}$$

Using the expression in (11.7), the overall covariance simplifies to

$$\begin{aligned} \text{Cov}[D_{\text{Ken}}, D_{\text{Cay}}] &= \sum_{j=1}^{m-1} \sum_{i=j+1}^{m} \frac{i - j - \frac{1}{2}}{m(m-1)} \\ &= \frac{1}{m(m-1)} \sum_{j=1}^{m-1} \sum_{k=1}^{m-j} (k - \frac{1}{2}) \\ &= \frac{1}{2m(m-1)} \sum_{j=1}^{m-1} (m-j)^2 \\ &= \frac{2m-1}{12}. \end{aligned} \tag{11.12}$$

This equation completes the matrix in (11.1).

11.2 Simulations for Ulam and maximum vs. the others

We simulated 10,000 random vector $\boldsymbol{Y}$'s of various lengths m to estimate the correlations of the Ulam and maximum distances versus the other five, and each other. Figure 11.1 graphs the correlations for a number of values of m from 10 to 10,000 for the correlations with Ulam's

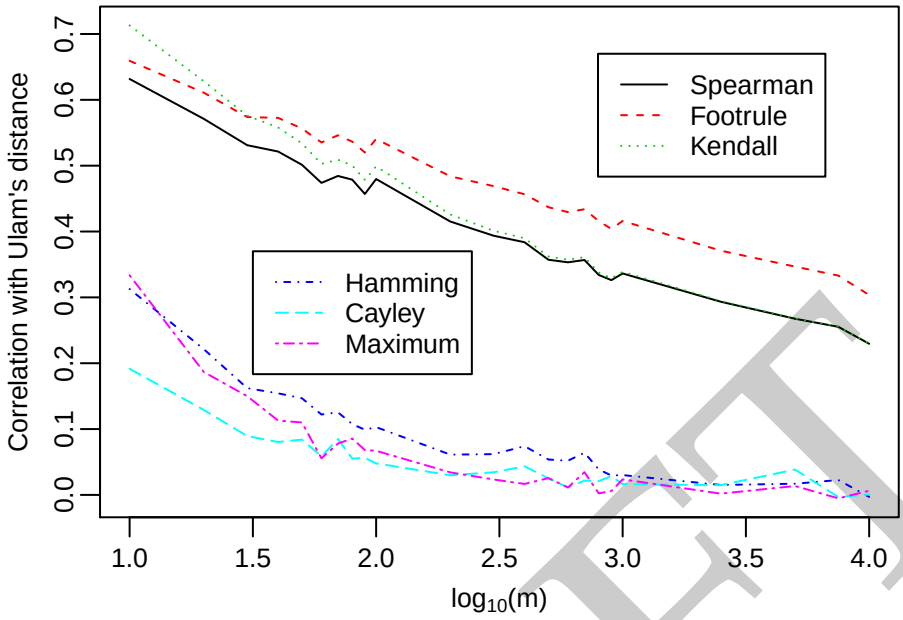

Figure 11.1: The correlation of Ulam's distance with the other distances as a function of $\log_{10}(m)$, for m from 10 to 10000.

distance, and Figure 11.2 has similar graphs for the correlations with the maximum distance. We used $\log_{10}(m)$ for the horizontal axis. Table (11.13) extracts these estimated correlations for a few values of m.

For the Ulam distance, we see that as m increases, the correlations decline. The correlations with the Spearman, footrule, and Kendall distances are much higher than those with Hamming, Cayley, and the maximum distance. The latter do appear to be close to zero, while the former seem to be declining fairly linearly with the log of m. For the maximum distance, all the correlations are approaching very close to zero.

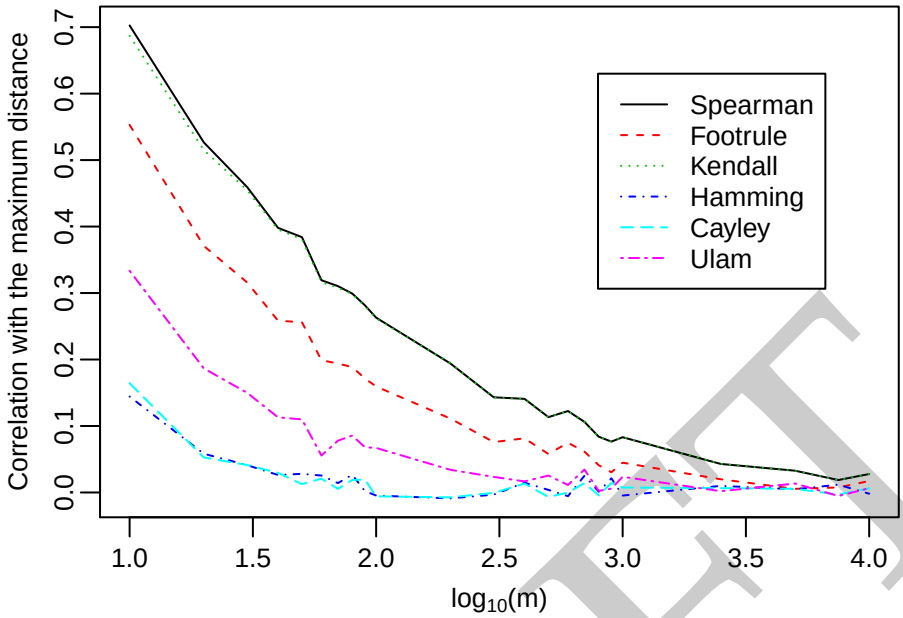

Figure 11.2: The correlation of the maximum distance with the other distancesas a function of $\log_{10}(m)$, for m from 10 to 10,000. The Spearman and Kendall distances' curves are almost on top of each other.

	Spearman	Footrule	Kendall	Hamming	Cayley	Maximum
$m = 10$						
Ulam	0.6317	0.6593	0.7130	0.3126	0.1916	0.3336
Maximum	0.7025	0.5530	0.6871	0.1443	0.1644	
$m = 100$						
Ulam	0.4797	0.5405	0.499	0.1030	0.0478	0.0670
Maximum	0.2630	0.1595	0.2617	−0.0046	−0.0058	
$m = 1,000$						
Ulam	0.3363	0.4161	0.3388	0.0301	0.0165	0.0236
Maximum	0.0831	0.0449	0.0833	−0.0046	0.0075	
$m = 10,000$						
Ulam	0.2295	0.3036	0.2299	−0.0030	0.0006	0.0064
Maximum	0.0278	0.0173	0.0278	−0.0018	0.0063	

(11.13)

We end this section with some thoughts about possible reasons why most of the correlations with Ulam and the maximum distances approach zero. Formal mathematical analysis so far has proven elusive.

Consider any bi-invariant distance. Its distribution depends on a given y_i only based on

whether $y_i = i$ or not. On the other hand, the Ulam and maximum distances are sensitive to the distance y_i is from i, thus would tend to have little relationship to the bi-invariant one.

Think about D_{Max}. From the asymptotic distribution in (10.3), we see that for a moderate value of the constant c, say $c = 10$, $P[D_{Max} \geqslant m - c\sqrt{m}] \approx 1$. Now

$$D_{Max} \geqslant m - c\sqrt{m} \implies D_{Max} = \max\{|Y_i - i| \,|\, i \in \mathcal{O}_{c,m}\}, \tag{11.14}$$

where

$$\mathcal{O}_{c,m} = \{i \,|\, 1 \leqslant i \leqslant c\sqrt{m} \text{ or } m - c\sqrt{m} \leqslant i \leqslant m\}, \tag{11.15}$$

since if $i \notin \mathcal{O}_{c,m}$, then $|y_i - i| < m - c\sqrt{m}$. Thus for large m, D_{Max} depends almost exclusively on only $2c\sqrt{m}$ of the y_i's. For Spearman, footrule, and Kendall, and possibly for Ulam, the asymptotics are not substantially affected by ignoring those indices, which suggests an asymptotically low correlation with D_{Max}

11.3 Proofs

We first deal with cases where both distances are Hoeffding distances.

Lemma 11.1. *Suppose D_A and D_B are Hoeffding distances (Chapter 3), with matrices $\boldsymbol{\Delta}_A$ and $\boldsymbol{\Delta}_B$ as in (3.4), respectively. Then*

$$Cov[D_A, D_B] = \frac{1}{m-1} \, trace\left(\boldsymbol{H}\boldsymbol{\Delta}_A'\boldsymbol{H}\boldsymbol{\Delta}_B\right), \tag{11.16}$$

where $\boldsymbol{H}$ is the $m \times m$ centering matrix from (3.8).

Proof. Similar to (3.19) through (3.21), with $\boldsymbol{W} = \boldsymbol{Q}(\boldsymbol{\Delta}_A \;\; \boldsymbol{\Delta}_B)$,

$$Cov[\boldsymbol{W}] = \frac{1}{m-1}\boldsymbol{H}_m \otimes \begin{pmatrix} \boldsymbol{\Delta}_A' \\ \boldsymbol{\Delta}_B' \end{pmatrix} \boldsymbol{H}_m(\boldsymbol{\Delta}_A \;\; \boldsymbol{\Delta}_B) \equiv \boldsymbol{\Sigma} \otimes \boldsymbol{\Omega}, \tag{11.17}$$

where here $\boldsymbol{\Omega}$ is $(2m)\times(2m)$. Now

$$\begin{aligned} Cov[D_a, D_B] &= Cov[\text{trace}(\boldsymbol{Q}\boldsymbol{\Delta}_A), \text{trace}(\boldsymbol{Q}\boldsymbol{\Delta}_B)] \\ &= \text{trace}\left(\sum_{i=1}^{m}\sum_{j=1}^{m} Cov[W_{ii}, W_{j,j+m}]\right) \\ &= \sum_{i=1}^{m}\sum_{j=1}^{m} \sigma_{ij}\omega_{i,j+m} \\ &= \frac{1}{m-1} \text{trace}\left(\boldsymbol{H}\boldsymbol{\Delta}_A'\boldsymbol{H}\boldsymbol{\Delta}_B\right). \end{aligned} \tag{11.18}$$

□

First, let $D_B = D_\rho$. From (4.1), we have

$$\boldsymbol{H\Delta}_\rho\boldsymbol{H} = -2\boldsymbol{H\omega}'\boldsymbol{\omega H}. \tag{11.19}$$

Thus

$$\begin{aligned} \mathrm{Cov}[D_A, D_\rho] &= \frac{1}{m-1}\,\mathrm{trace}(\boldsymbol{H\Delta}_A\boldsymbol{H}\Delta_\rho) \\ &= -\frac{2}{m-1}\,\mathrm{trace}(\boldsymbol{\Delta}_A\boldsymbol{H\omega}'\boldsymbol{\omega H}) \\ &= -\frac{2}{m-1}\,\boldsymbol{\omega H\Delta}_A\boldsymbol{H\omega}', \end{aligned} \tag{11.20}$$

as in (11.2a).

Now take $D_B = D_{\mathrm{Ham}}$, for which $\boldsymbol{\Delta}_{\mathrm{Ham}} = \mathbf{1}'\mathbf{1} - \boldsymbol{I}$, where $\boldsymbol{I}$ is the $m \times m$ identity matrix, and $\mathbf{1}$ is the $1 \times m$ vector of 1's. Then $\boldsymbol{H\Delta}_{\mathrm{Ham}}\boldsymbol{H} = -\boldsymbol{H}$, so that

$$\mathrm{Cov}[D_A, D_{\mathrm{Ham}}] = -\frac{1}{m-1}\,\mathrm{trace}(\boldsymbol{H\Delta}_A) = \frac{1}{m-1}\,E[D_A] \tag{11.21}$$

by (3.17), which is (11.2c)

The next result deals with one Hoeffding distance.

Lemma 11.2. *Suppose* $D_A = d_A(\boldsymbol{Y}, \boldsymbol{\omega})$ *is a Hoeffding distance with* $D_A = \sum \delta_A(Y_i, i)$ *as in (3.1), and* D *is any other distance. Then*

$$\mathrm{Cov}[D_A, D] = \sum_{k=1}^{m} \mathrm{Cov}[\delta_A(Y_k, k), E[D \,|\, Y_k]]. \tag{11.22}$$

Furthermore, if D *is bi-invariant,*

$$\mathrm{Cov}[D_A, D] = \frac{1}{m-1} E[D_A](E[D] - E[D \,|\, Y_k = k]). \tag{11.23}$$

Proof. First,

$$\mathrm{Cov}[D_A, D] = \sum_{k=1}^{m} \mathrm{Cov}[\delta_A(Y_k, k), D]. \tag{11.24}$$

For each summand, we use the conditional/unconditional formula for covariances,

$$\mathrm{Cov}[\delta_A(Y_k, k), D] = E[\mathrm{Cov}[\delta_A(Y_k, k), D \,|\, Y_k]] + \mathrm{Cov}[E[\delta_A(Y_k, k) \,|\, Y_k], E[D \,|\, Y_k]]. \tag{11.25}$$

The conditional covariance in the first term on the right-hand side is zero since conditionally $\delta_A(Y_k, k)$ is a constant. For the other term, $E[\delta_A(Y_k, k) \,|\, Y_k] = \delta_A(Y_k, k)$, hence we have (11.22).

It can be shown that for any bi-invariant distance D, the conditional distribution $D \,|\, Y_k = l$ depends only on whether $l = k$. Then for some a and b,

$$E[D \,|\, Y_k = k] = a \quad \text{and} \quad E[D \,|\, Y_k = l] = b \quad \text{for any} \quad k, l \quad \text{with} \quad k \neq l. \tag{11.26}$$

Furthermore, the unconditional expectation can be written

$$\begin{aligned} E[D] = E[D \mid Y_k = k]P[Y_k = k] + E[D \mid Y_k \neq k]P[Y_k \neq k] &= a\,\frac{1}{m} + b\,\frac{m-1}{m} \\ \Longrightarrow b &= \frac{m\,E[D] - a}{m-1}. \end{aligned} \tag{11.27}$$

Then we can write

$$E[D \mid Y_k] = (a-b)\,I[Y_k = k] + b. \tag{11.28}$$

Each summand in (11.22) here is

$$\begin{aligned} Cov[\delta_A(Y_k, k), E[D \mid Y_k]] &= (a-b)\,Cov[\delta_A(Y_k, k), I[Y_k = k]] \\ &= (a-b)\,(E[\delta_A(Y_k, k)\,I[Y_k = k]] - E[\delta_A(Y_k, k)]E[I[Y_k = k]]) \\ &= \frac{b-a}{m}\,E[\delta_A(Y_k, k)], \end{aligned} \tag{11.29}$$

since $\delta_A(Y_k, k) = 0$ if $Y_k = k$. Then by (11.22), $Cov[D_A, D]$ is found by summing (11.29) over k, which yields $(b-a)E[D_A]/m$, and (11.23) follows by writing $b-a$ as a function of $E[D]$ and $E[D \mid Y_k = k]$. □

Consider the covariance of Kendall's distance with a Hoeffding distance, so that $D = D_{Ken}$ in (11.22). Now

$$\begin{aligned} E[D_{Ken} \mid Y_k = l] &= \sum_{1 \leqslant i < j \leqslant m} \sum P[Y_i > Y_j \mid Y_k = l] \\ &= \sum_{i=1}^{k-1} P[Y_i > l \mid Y_k = l] + \sum_{j=k+1}^{m} P[l > Y_j \mid Y_k = l] + \sum_{1 \leqslant i < j \leqslant m, i, j \neq k} \sum P[Y_i > Y_j] \\ &= (k-1)\frac{m-l}{m-1} + (m-k)\frac{l-1}{m-1} + \frac{(m-1)(m-2)}{2}\frac{1}{2} \\ &= -\frac{2}{m-1}\left(k - \frac{m+1}{2}\right) l + C_{k,m}, \end{aligned} \tag{11.30}$$

where $C_{k,m}$ is a constant independent of l. The parallel calculation for Spearman's distance is

$$\begin{aligned} E[D_{Spear} \mid Y_k = l] &= \sum_{i=1}^{m} E[Y_i^2 \mid Y_k = l] + \sum_{i=1}^{m} i^2 - 2\sum_{i=1}^{m} i\,E[Y_i \mid Y_k = l] \\ &= \frac{m(m+1)(2m+1)}{3} - 2\left(k\,E[Y_k \mid Y_k = l] + \sum_{i \neq k} i\,E[Y_i \mid Y_k = l]\right). \end{aligned} \tag{11.31}$$

Note that

$$E[Y_i \mid Y_k = l] = \begin{cases} l & \text{if } i = k \\ \frac{1}{m-1}\left(\frac{m(m+1)}{2} - l\right) & \text{if } i \neq k \end{cases}, \tag{11.32}$$

since if $i \neq k$, Y_i is conditionally equally likely to be anything but l. Thus

$$\begin{aligned} E[D_{\text{Spear}} \,|\, Y_k = l] &= -2\left(kl + \frac{1}{m-1}\left(\frac{m(m+1)}{2} - k\right)\left(\frac{m(m+1)}{2} - l\right)\right) + C^*_m \\ &= -\frac{2m}{m-1}\left(k - \frac{m+1}{2}\right) l + C^{**}_{k,m}. \end{aligned} \tag{11.33}$$

In (11.22), for Spearman's or Kendall's distance, we can ignore the constants, so that by setting $l = Y_k$ in (11.30) and (11.33),

$$\begin{aligned} \text{Cov}[D_A, D_{\text{Ken}}] &= -\frac{2}{m-1}\sum_{k=1}^{m}\left(k - \frac{m+1}{2}\right) \text{Cov}[\delta_A(Y_k, k), Y_k], \text{ and} \\ \text{Cov}[D_A, D_{\text{Spear}}] &= -\frac{2m}{m-1}\sum_{k=1}^{m}\left(k - \frac{m+1}{2}\right) \text{Cov}[\delta_A(Y_k, k), Y_k]. \end{aligned} \tag{11.34}$$

They differ only by a factor of m, which gives us (11.2b)

Turn to Cayley's distance, which is bi-invariant (see Section 1.7). Given $Y_m = m$, the conditional distribution of $(Y_1, \ldots, Y_{m-1})$ is Uniform$(\mathcal{P}_{m-1})$. Thus the conditional distribution of Cayley's distance given $Y_m = m$ is the same as the unconditional distribution when there are just $m-1$ objects to rank. From (1.10) for the mean, then, we have

$$E[D_{\text{Cay}}] = m - \sum_{i=1}^{m}\frac{1}{i} \text{ and } E[D_{\text{Cay}} \,|\, Y_k = k] = m - 1 - \sum_{i=1}^{m-1}\frac{1}{i}, \tag{11.35}$$

since by bi-invariance, conditioning on $Y_k = k$ is the same for any k. Thus (11.23) gives us

$$\begin{aligned} \text{Cov}[D_A, D_{\text{Cay}}] &= \frac{1}{m-1} E[D_A](E[D_{\text{Cay}}] - E[D_{\text{Cay}} \,|\, Y_k = k]) \\ &= \frac{1}{m-1}(1 - \frac{1}{m})E[D_A] = \frac{1}{m}\, E[D_A]. \end{aligned} \tag{11.36}$$

Thus we have (11.2d).

Chapter 12

Tied Rankings

Often, rank data is incomplete, either by design or by random ties or judges leaving out information. For example, people might rank only the top three of five choices, or an experiment might be conducted where each judge is given only a subset of the objects to rank. In nonparametrics, even nominally continuous variables will often have ties (from round-off error or inherent discreteness, such as age in years).

There are a couple of general approaches to extending distances on full rankings to those on partial rankings. Both are based on the notion of **compatibility sets**, as in Alvo & Cabilio (1991) and Critchlow (1985), and earlier in M. G. Kendall & Gibbons (1990) (a recent update to M. Kendall (1948)) for Spearman's ρ. The idea is that for any incomplete ranking, there is a well-defined set of complete rankings that is compatible with it. For example, suppose $m = 5$, and a ranking with ties ranks object 1 first, has objects 2 and 3 tied in second place, and objects 4 and 5 tied in last place. We will represent this tied ranking as $\boldsymbol{x} = (1, 2, 2, 3, 3)$. Then the complete rankings which are compatible with $\boldsymbol{x}$ are those with object 1 first, objects 2 and 3 next in either order, and objects 4 and 5 taking the last two slots, in either order. If the tied ranking is $\boldsymbol{x} = (1, 2, 3, 3, 3)$, then the compatible complete rankings are all those with objects 1 and 2 ranked 1 and 2, respectively. That is,

Tied ranking	Compatible complete rankings
$(1,2,2,3,3)$	$(1,2,3,4,5),\ (1,2,3,5,4),\ (1,3,2,4,5),\ (1,3,2,5,4).$
$(1,2,3,3,3)$	$(1,2,3,4,5),\ (1,2,3,5,4),\ (1,2,4,3,5),\ (1,2,4,5,3),$ $(1,2,5,3,4),\ (1,2,5,4,3).$

(12.1)

If only the first three objects are ranked, and they are ranked 1, 2, and 3, then the compatible complete rankings are those with the first three ranked in that order, but not necessarily ranked 1, 2, 3:

Incomplete ranking	Compatible complete rankings
$(1,2,3,*,*)$	$(1,2,3,4,5),\ (1,2,3,5,4),\ (1,2,4,3,5),\ (1,2,4,5,3),$ $(1,2,5,3,4),\ (1,2,5,4,3),\ (1,3,4,2,5),\ (1,3,4,5,2),$ $(1,3,5,2,4),\ (1,3,5,4,2),\ (1,4,5,2,3),\ (1,4,5,3,2),$ $(2,3,4,1,5),\ (2,3,4,5,1)\ (2,3,5,1,4),\ (2,3,5,4,1),$ $(2,4,5,1,3),\ (2,4,5,3,1),\ (3,4,5,1,2),\ (3,4,5,2,1).$

(12.2)

Other schemes can be imagined. For example, suppose with $m = 4$ objects, a judge prefers object 1 to object 2, and object 3 to object 4. Then the compatible complete rankings would be (1,2,3,4), (1,3,2,4), (1,4,2,3), (2,3,1,4), (2,4,1,3), and (3,4,1,2).

Suppose we have two incomplete rankings, represented by their compatibility subsets $\mathcal{W}, \mathcal{Z} \subset \mathcal{P}_m$. We wish to define a distance between the two subsets based on the interpoint distances $d(\boldsymbol{x}, \boldsymbol{y})$ for $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{P}_m$. The two main approaches in the literature are the use of **Hausdorff** distances proposed by Critchlow (1985), and the averaging of Alvo & Cabilio (1991). The former looks at the Hausdorff distance between $\mathcal{W}$ and $\mathcal{Z}$ induced by d:

$$d^H(\mathcal{W}, \mathcal{Z}) = \max\{\max_{\boldsymbol{x} \in \mathcal{W}} d^m(\boldsymbol{x}, \mathcal{Z}), \max_{\boldsymbol{y} \in \mathcal{Z}} d^m(\mathcal{W}, \boldsymbol{y})\}, \tag{12.3}$$

where

$$d^m(\boldsymbol{x}, \mathcal{Z}) = \min_{\boldsymbol{y} \in \mathcal{Z}} d(\boldsymbol{x}, \boldsymbol{y}) \quad \text{and} \quad d^m(\mathcal{W}, \boldsymbol{y}) = \min_{\boldsymbol{x} \in \mathcal{W}} d(\boldsymbol{x}, \boldsymbol{y}). \tag{12.4}$$

There are various ways to interpret Hausdorff distance. E.g., $d^H(\mathcal{W}, \mathcal{Z})$ is the smallest K such that for each $\boldsymbol{x} \in \mathcal{W}$, there exists a $\boldsymbol{y} \in \mathcal{Z}$ such that $d(\boldsymbol{x}, \boldsymbol{y}) \leqslant K$, and for each $\boldsymbol{y} \in \mathcal{Z}$, there exists an $\boldsymbol{x} \in \mathcal{W}$ such that $d(\boldsymbol{x}, \boldsymbol{y}) \leqslant K$.

The averaging approach just averages the distances $d(\boldsymbol{x}, \boldsymbol{y})$ for $\boldsymbol{x} \in \mathcal{W}$ and $\boldsymbol{y} \in \mathcal{Z}$:

$$d^A(\mathcal{W}, \mathcal{Z}) = \frac{\sum_{\boldsymbol{x} \in \mathcal{W}} \sum_{\boldsymbol{y} \in \mathcal{Z}} d(\boldsymbol{x}, \boldsymbol{y})}{\#\mathcal{W} \times \#\mathcal{Z}}, \tag{12.5}$$

The idea here is the same as for hierarchical agglomerative clustering using average linkage. As in clustering, we could also use the maximum of the interpoint distances (complete linkage) or minimum of them (single linkage) in place of the average:

$$\begin{aligned} d^M(\mathcal{W}, \mathcal{Z}) &= \max_{\boldsymbol{x} \in \mathcal{W}, \boldsymbol{y} \in \mathcal{Z}} d(\boldsymbol{x}, \boldsymbol{y}); \\ d^m(\mathcal{W}, \mathcal{Z}) &= \min_{\boldsymbol{x} \in \mathcal{W}, \boldsymbol{y} \in \mathcal{Z}} d(\boldsymbol{x}, \boldsymbol{y}). \end{aligned} \tag{12.6}$$

If the scientific context leading to the ties rankings dictates which method to use, then certainly that is the one to use. For general purposes, the minimum and maximum seem too extreme, while either the averaging or Hausodrff approach has its advantages. The Hausdorff distance is an actual metric on the subsets of $\mathcal{P}_m$ (if d is metric on $\mathcal{P}_m$), whereas the average distance is typically not, i.e., it is likely that $d_A(\mathcal{W}, \mathcal{W}) > 0$, and there is no guarantee that the triangle inequality holds. Any Hausdorff distance will take values in the support of $d(\boldsymbol{x}, \boldsymbol{y})$. It is also easier than averaging to calculate for Ulam's and the maximum distances. The averaging approach is the usual one, especially for Hoeffding and Kendall distances. In particular, for Spearman's distance, it leads naturally to the popular midrank statistic for tied ranks.

For situations with not too many ties, the two approaches yield similar results. There can be interesting differences in other cases. For an extreme example, suppose $m = 10$ and $\boldsymbol{w} = (1, 1, 1, 1, 1, 2, 2, 2, 2, 2)$. We consider three $\boldsymbol{z}$'s that conform maximally to $\boldsymbol{w}$ but with increasing numbers of ties. Because of the ties (in one or both vectors), we might be reluctant to declare their distances from $\boldsymbol{w}$ to be zero. We will look at Spearman's ρ as in (1.14), but with the distance extended to ties in the numerator. E.g., for the averaging approach,

$\rho^A(\boldsymbol{w},\boldsymbol{z}) = 1 - d^A_{Spear}(\boldsymbol{w},\boldsymbol{z})/E[d_{Spear}(\boldsymbol{Z},\boldsymbol{\nu})]$, where in the denominator the $\boldsymbol{Z}$ and $\boldsymbol{\nu}$ are in $\mathcal{P}_m$. Then for the four approaches we have

$\boldsymbol{z}$	Averaging	Hausdorff	Minimum	Maximum
$(1,2,3,4,5,6,7,8,9,10)$	0.7576	0.5152	1	0.5152
$(1,1,1,2,2,3,3,3,4,4)$	0.7576	0.6364	1	0.5152
$(1,1,1,1,1,2,2,2,2,2)$	0.7576	1	1	0.5152

(12.7)

(Note that the minimum distance yields the maximum ρ, and vice versa.)

The values of Spearman's ρ for the averaging approach are high (0.76), but not too close to 1, as would seem reasonable. The maximum and minimum are going to be extreme (and in fact average to the d^A). What is interesting is that the Hausdorff ranges from being equal to the maximum when $\boldsymbol{z}$ has no ties, to being equal to the minimum when $\boldsymbol{z}$ has the most ties. At first it might seem unusual that the less-specific $\boldsymbol{z}$ achieves the smallest distance from $\boldsymbol{w}$, but in this case for the last $\boldsymbol{z}$, the compatibility sets for $\boldsymbol{w}$ and $\boldsymbol{z}$ are the same, hence their Hausdorff distance is zero. When $\boldsymbol{z} = (1,2,\ldots,10)$, we are comparing a large compatibility set $\mathcal{W}$ with a singleton set $\{\boldsymbol{z}\}$, hence the Hausdorff distance is the maximum between $\boldsymbol{z}$ and those in $\mathcal{W}$. Thus we may expect Hausdorff to tend to be larger the more discrepant the numbers of ties in the vectors.

We redid the calculations after randomly shuffling the $\boldsymbol{z}$. Now among the first five slots, there are two low values and three high, and vice versa for the last five, hence intuitively we might expect ρ to be slightly negative, as indeed it is for the averaging approach. Note that the minimum and maximum are quite extreme, and differing in sign. Hausdorff again shows the distance decreasing as the compatibility sets become closer in size, where the ρ increases quite a bit from -0.62 to $+0.03$.

$\boldsymbol{z}$	Averaging	Hausdorff	Minimum	Maximum
$(6,9,4,10,1,8,5,3,2,7)$	-0.1515	-0.6242	0.3212	-0.6242
$(3,4,2,4,1,3,2,1,1,3)$	-0.3030	-0.5879	0.3212	-0.8667
$(2,2,1,2,1,2,2,1,1,2)$	0.0000	0.0303	0.8182	-0.8182

(12.8)

The above are just a few examples, but they do point out that the approaches are quite different, and at least to me suggest that the averaging is most intuitive.

The main goal of this chapter is to calculate the averaging approach for the given distances. See Section 12.1. Critchlow (1985) gives a thorough analysis of the Hausdorff approach, which we summarize in Section 12.2. Chapters 13, 14, and 16 go into more detail on the distributional aspects for the Spearman and Kendall distances.

12.1 Averaging

As above, we will define a tied ranking to be an $m \times 1$ vector $\boldsymbol{w}$ with values from the set $\{1,\ldots,K\}$ for some K, where each value in that set is represented in $\boldsymbol{w}$. Thus $\boldsymbol{w} = (1,3,2,1,3)$ is valid, while $\boldsymbol{w} = (1,3,4,1,3)$ is not. The lower the value of w_i, the more object i is preferred. If $w_i = w_j$, then objects i and j are equally preferred. The compatibility set as in (12.1) is here

$$\mathcal{C}(1,3,2,1,3) = \{(1,4,3,2,5),\ (1,5,3,2,4),\ (2,4,3,1,5),\ (2,5,3,1,4)\} \subset \mathcal{P}_m. \tag{12.9}$$

A formal general definition of the compatibility set uses Kendall's distance:

$$\mathcal{C}(\boldsymbol{w}) = \{\boldsymbol{x} \in \mathcal{P}_m \mid \sum_{1 \leqslant j < i \leqslant m} \sum I[(w_i - w_j)(x_i - x_j) < 0] = 0\}. \tag{12.10}$$

The averaging approach seems most reasonable for the Spearman and footrule (and more generally, L_p) distances, and Kendall's distance. It also makes sense for the Ulam and maximum distances, although for ease of calculation, one might prefer to use the Hausdorff, maximum, or minimum in (12.5) rather than the average. Section 12.3 presents a modification of the Hausdorff approach that may be more appealing for these distances. In Section 12.2.1, we discuss whether this approach to tied rankings makes sense for bi-invariant distances.

We can represent the distribution of elements in the compatibility set as

$$\boldsymbol{X} \mid \boldsymbol{W} = \boldsymbol{w} \sim \text{Uniform}(\mathcal{C}(\boldsymbol{w})). \tag{12.11}$$

Then if the marginal distribution of $\boldsymbol{W}$ is uniform over the permutations of $\boldsymbol{w}$, the unconditional distribution of $\boldsymbol{X}$ is Uniform$(\mathcal{P}_m)$. If $\boldsymbol{z}$ is another tied ranking, and $\boldsymbol{Y} \mid \boldsymbol{Z} = \boldsymbol{z} \sim$ Uniform$(\mathcal{C}(\boldsymbol{z}))$, independent of $(\boldsymbol{W}, \boldsymbol{X})$, then we can represent the distance in (12.5) as

$$d^A(\boldsymbol{w}, \boldsymbol{z}) = E[d(\boldsymbol{X}, \boldsymbol{Y}) \mid \boldsymbol{W} = \boldsymbol{w}, \boldsymbol{Z} = \boldsymbol{z}]. \tag{12.12}$$

We need to find the distributions (12.11). Let $\boldsymbol{m}$ be pattern of ties for $\boldsymbol{w}$, i.e., the counts for its elements:

$$\boldsymbol{m} = (m_1, \ldots, m_K), \quad m_a = \#\{i \mid w_i = a\}. \tag{12.13}$$

As $\boldsymbol{x}$ runs over the compatible set $\mathcal{C}(\boldsymbol{w})$, the individual x_i's run over ranks determined by their values. That is, the x_i's with $w_i = 1$ have the lowest ranks, then come those with $w_i = 2$, etc. Specifically, denoting the cumulative sums of the $\boldsymbol{m}$ by

$$m_{\leqslant a} = \begin{cases} 0 & \text{if } a = 0 \\ m_1 + \cdots + m_a & \text{if } a = 1, \ldots, K \end{cases}, \tag{12.14}$$

and let $m_{<a} = m_{\leqslant a-1}$. Then we have

$$\begin{aligned} w_i = 1 &\Rightarrow x_i \in \{1, \ldots, m_1\}; \\ &\vdots \\ w_i = a &\Rightarrow x_i \in \{m_{<a} + 1, \ldots, m_{\leqslant a}\}; \\ &\vdots \\ w_i = K &\Rightarrow x_i \in \{m_{<K} + 1, m\}. \end{aligned} \tag{12.15}$$

For example, suppose $\boldsymbol{w} = (1, 3, 4, 4, 2, 2, 3, 3, 2, 3)$. Then there is one 1, three 2's, etc. so $\boldsymbol{m} = (1, 3, 4, 2)$ and we have $x_i \in \{m_{\leqslant w_1 - 1} + 1, \ldots, m_{\leqslant w_i}\}$, where

i	w_i	$m_{<w_1} + 1$	$m_{\leqslant w_i}$	m_{w_i}
1	1	1	1	1
5, 6, 9	2	2	4	3
2, 7, 8, 10	3	5	8	4
3, 4	4	9	10	2

(12.16)

E.g., w_2, w_7, w_8 and w_{10} all equal 3, so the corresponding x_i's go from $m_1 + m_2 + 1 = 5$ to $m_1 + m_2 + m_3 = 8$.

The values of X_i and X_j are conditionally independent if they are from different groups, that is, if $w_i \neq w_j$. The vector of X_i's from the same group are distributed uniformly over the permutations of their range,

$$(X_i \,|\, W_i = a) \,|\, \boldsymbol{W} = \boldsymbol{w} \sim \text{Uniform}(\text{Permutations}(m_{<a} + 1, \ldots, m_{\leqslant a})). \tag{12.17}$$

Thus for the example in (12.16),

$$\begin{aligned} X_1 \,|\, \boldsymbol{W} = \boldsymbol{w} &\sim \text{Point mass at 1,} \\ (X_5, X_6, X_9) \,|\, \boldsymbol{W} = \boldsymbol{w} &\sim \text{Uniform}(\text{Permutations}(2, 3, 4)), \\ (X_2, X_7, X_8, X_{10}) \,|\, \boldsymbol{W} = \boldsymbol{w} &\sim \text{Uniform}(\text{Permutations}(5, 6, 7, 8)), \\ (X_3, X_4) \,|\, \boldsymbol{W} = \boldsymbol{w} &\sim \text{Uniform}(\text{Permutations}(9, 10)), \end{aligned} \tag{12.18}$$

and the four vectors are conditionally independent.

Consider two tied rankings, $\boldsymbol{w}$ and $\boldsymbol{z}$, for the same set of objects, so that $\mathcal{C}(\boldsymbol{w})$ and $\mathcal{C}(\boldsymbol{z})$ are their compatibility sets. Let $\boldsymbol{n}$ be the vector of counts for $\boldsymbol{z}$, so that $n_b = \#\{i \,|\, z_i = b\}$, and define $n_{\leqslant b}$ as in (12.14) for $\boldsymbol{n}$. The remainder of this section considers specific distances. The Hoeffding distances and Kendall's distance have reasonably tractable extensions using the averaging, and is the traditional approach for Spearman and Kendall. We have not found simple formulas to apply the averaging to the other distances. We can always enumerate over the $\prod m_a! \prod n_b!$ values for $\boldsymbol{X}$ and $\boldsymbol{Y}$ in (12.5) if the counts are fairly small, or use simulations if not.

In Section 12.2.1, we note that the above method for dealing with ties may not make sense for the bi-invariant distances, and present alternatives.

12.1.1 Hoeffding distances

As in (3.1), a Hoeffding distance is of the form

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{m} \delta(x_i, y_i) \tag{12.19}$$

for given function δ. The averaging (12.11) is now

$$d^A(\boldsymbol{w}, \boldsymbol{z}) = \sum_{i=1}^{m} E[\delta(X_i, Y_i) \,|\, \boldsymbol{W} = \boldsymbol{w}, \boldsymbol{Z} = \boldsymbol{z}]. \tag{12.20}$$

Thus we need just the marginal distributions of X_i's and Y_i's. By (12.17) we see that they are conditionally independent and uniform over their conditional spaces spaces given in (12.15). Then (12.20) can be reasonably easily applied to specific δ's. In fact, we can write the distance in (12.20) as a Hoeffding distance with

$$\delta^*(a, b) = \frac{\sum_{x=m_{<a}+1}^{m_{\leqslant a}} \sum_{y=m_{<b}+1}^{m_{\leqslant b}} \delta(x, y)}{m_a n_b}. \tag{12.21}$$

Then we have

$$d^A(\boldsymbol{w}, \boldsymbol{z}) = \sum_{i=1}^{m} \delta^*(w_i, z_i). \tag{12.22}$$

Consider Spearman's distance. We have from (4.39)

$$d_{\text{Spear}}(\boldsymbol{x}, \boldsymbol{y}) = \frac{m(m+1)(2m+1)}{3} - 2\sum_{i=1}^{m} x_i y_i. \tag{12.23}$$

Then for fixed $\boldsymbol{w}$ and $\boldsymbol{z}$, we have

$$d^A_{\text{Spear}}(\boldsymbol{w}, \boldsymbol{z}) = \frac{m(m+1)(2m+1)}{3} - 2\sum_{i=1}^{m} E[X_i \,|\, \boldsymbol{W} = \boldsymbol{w}]E[Y_i \,|\, \boldsymbol{Z} = \boldsymbol{z}]. \tag{12.24}$$

The $E[X_i \,|\, \boldsymbol{W} = \boldsymbol{w}]$ and $E[Y_i \,|\, \boldsymbol{Z} = \boldsymbol{z}]$ are well-known as the **midranks** of the vectors $\boldsymbol{w}$ and $\boldsymbol{z}$, respectively. In R, the default output of the function rank(w) is the midrank vector, hence we set

$$\text{rank}(\boldsymbol{w})_i = E[X_i \,|\, \boldsymbol{W} = \boldsymbol{w}] = \frac{m_{<w_i-1} + m_{\leqslant w_i} + 1}{2}, \tag{12.25}$$

and similarly for $\boldsymbol{z}$. Then we can write

$$d^A_{\text{Spear}}(\boldsymbol{w}, \boldsymbol{z}) = \frac{m(m+1)(2m+1)}{3} - 2\sum_{i=1}^{m} \text{rank}(\boldsymbol{w})_i \text{rank}(\boldsymbol{z})_i. \tag{12.26}$$

By (1.4), for Hamming's distance, $\delta(x, y) = 1 - I[x = y]$, hence

$$\begin{aligned}
\delta^*(a, b) &= 1 - \frac{\sum_{x=m_{<a}+1}^{m_{\leqslant a}} \sum_{y=m_{<b}+1}^{m_{\leqslant b}} I[x = y]}{m_a n_b} \\
&= 1 - \frac{\#(\{m_{<a} + 1, \ldots, m_{\leqslant a}\} \cap \{n_{<b} + 1, \ldots, n_{\leqslant b}\})}{m_a n_b} \\
&= 1 - \frac{(\min\{m_{\leqslant a}, n_{\leqslant b}\} - \max\{m_{<a}, n_{<b}\})^+}{m_a n_b}.
\end{aligned} \tag{12.27}$$

Here, $z^+ = \max\{z, 0\}$.

For the footrule in (1.4), $\delta(x, y) = |x - y|$, so that

$$\delta^*(a, b) = \frac{\sum_{x=m_{<a}+1}^{m_{\leqslant a}} \sum_{y=m_{<b}+1}^{m_{\leqslant b}} |x - y|}{m_a n_b}. \tag{12.28}$$

To find a formula for the numerator of (12.28), we need to figure out which parts of the ranges of x's and y's overlap, and which are disjoint. It's useful to define the following summations:

$$\begin{aligned}
&\text{If } i \leqslant j \leqslant k \leqslant l: \quad f(i, j, k, l) = \sum_{x=i+1}^{j} \sum_{y=k+1}^{l} |x - y| = \left(\frac{k+l}{2} - \frac{i+j}{2}\right)(j - i)(l - k); \\
&\text{if } i \leqslant j: \quad g(i, j) = \sum_{x=i+1}^{j} \sum_{y=i+1}^{j} |x - y| = \frac{(j-i)^2 - 1}{3}(j - i).
\end{aligned} \tag{12.29}$$

The first case has $y \geqslant x$, so we are finding a multiple of the difference in average values of the y's and the x's. The second case is m times the average value of the footrule in (1.10) when there are no ties and $m = j - i$.

For the general case, for fixed a and b, set $i = m_{<a}$, $j = m_{\leqslant a}$, $k = m_{<b}$ and $l = m_{\leqslant b}$. We consider some cases based on the ordering of i, j, k, l, with $i \leqslant k$, where we automatically have $i < j$ and $k < l$. Then

$$\sum_{x=i+1}^{j} \sum_{y=k+1}^{l} |x-y| = \begin{cases} f(i,j,k,l) & \text{if } i < j \leqslant k < l \\ f(i,k,k,l) + g(k,j) + f(k,j,j,l) & \text{if } i \leqslant k < j \leqslant l \\ f(i,k,k,l) + g(k,l) + f(k,l,l,j) & \text{if } i \leqslant k < l < j \end{cases}. \tag{12.30}$$

If $k < i$, then we switch the roles of (i,j) and (k,l), i.e.,

$$(i,j,k,l) = \begin{cases} (m_{<a}, m_{\leqslant a}, n_{<b}, n_{\leqslant b}) & \text{if } m_{<a} \leqslant n_{<b} \\ (n_{<b}, n_{\leqslant b}, m_{<a}, m_{\leqslant a}) & \text{if } n_{<b} < m_{<a} \end{cases}. \tag{12.31}$$

12.1.2 Kendall's distance

Turn to Kendall's distance. Using (12.5) and (6.1), we have

$$d^{A}_{Ken}(\boldsymbol{w}, \boldsymbol{z}) = \sum_{1 \leqslant j < i \leqslant m} \sum P[(X_i - X_j)(Y_i - Y_j) < 0 \,|\, \boldsymbol{W} = \boldsymbol{w}, \boldsymbol{Z} = \boldsymbol{z}]. \tag{12.32}$$

If $w_i \neq w_j$, then the ranges of X_i and X_j do not overlap, and are in the same direction as w_i and w_j. That is

$$w_i > w_j \;\;\Rightarrow\;\; X_i > X_j \quad \text{and} \quad w_i < w_j \;\;\Rightarrow\;\; X_i < X_j. \tag{12.33}$$

If $w_i = w_j$, then (X_i, X_j) are equally likely to be any distinct pair of values from $l_w(i)$ to $u_w(i)$, so that $P[X_i - X_j > 0 \,|\, \boldsymbol{W} = \boldsymbol{w}] = P[X_i - X_j < 0 \,|\, \boldsymbol{W} = \boldsymbol{w}] = \frac{1}{2}$. Similarly for the $\boldsymbol{z}$ and $\boldsymbol{Y}$. Since $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent, we have that

$$P[(X_i - X_j)(Y_i - Y_j) < 0 \,|\, \boldsymbol{W} = \boldsymbol{w}, \boldsymbol{Z} = \boldsymbol{z}] = \begin{cases} 1 & \text{if } (w_i - w_j)(z_i - z_j) < 0 \\ \frac{1}{2} & \text{if } (w_i - w_j)(z_i - z_j) = 0 \\ 0 & \text{if } (w_i - w_j)(z_i - z_j) > 0 \end{cases}. \tag{12.34}$$

Thus

$$d^{A}_{Ken}(\boldsymbol{w}, \boldsymbol{z}) = \sum_{1 \leqslant j < i \leqslant m} \sum \left(I[(w_i - w_j)(z_i - z_j) < 0] + \tfrac{1}{2}\, I[(w_i - w_j)(z_i - z_j) = 0] \right). \tag{12.35}$$

This distance can be written as the sum of the usual Kendall distance, $d_{Ken}(\boldsymbol{w}, \boldsymbol{z})$, plus an adjustment for ties, which is

$$\begin{aligned}\frac{1}{2}\sum_{1\leqslant j<i\leqslant m} I[(w_i - w_j)(z_i - z_j) = 0] &= \frac{1}{2}(\#\{i < j \,|\, w_i = w_j\} + \#\{i < j \,|\, z_i = z_j\} \\ &\quad - \#\{i < j \,|\, w_i = w_j \;\&\; z_i = z_j\}) \\ &= \frac{1}{2}\left(\sum_{a=1}^{K}\binom{m_a}{2} + \sum_{b=1}^{L}\binom{n_b}{2} - \sum_{a=1}^{K}\sum_{b=1}^{L}\binom{t_{ab}}{2}\right) \\ &= \frac{1}{4}\left(\sum_{a=1}^{K} m_a^2 + \sum_{b=1}^{L} n_b^2 - \sum_{a=1}^{K}\sum_{b=1}^{L} t_{ab}^2 - m\right). \qquad (12.36)\end{aligned}$$

where

$$t_{ab} = T_{ab}(\boldsymbol{w}, \boldsymbol{z}) \equiv \#\{i \,|\, w_i = a \;\&\; z_i = b\}. \qquad (12.37)$$

12.1.3 Other distances

We have not been able to discover simpler methods for applying the averaging to ties for the Ulam, Cayley, or maximum distances. Section 12.2.1 suggests alternatives for Cayley (and Hamming) that better conform to bi-invariance, and the Hausdorff approach is convenient for Ulam and maximum as seen in Section 12.2. In any case, if the compatibility sets $\mathcal{W}$ and $\mathcal{Z}$ are not too large (say, $\#\mathcal{C}(\boldsymbol{w}) \times \#\mathcal{C}(\boldsymbol{z}) \leqslant 20$ million), then the averaging (12.5) can be calculated directly. For larger sizes, randomly sampling pairs $(\boldsymbol{x}^{(k)}, \boldsymbol{y}^{(k)}) \in \mathcal{C}(\boldsymbol{w}) \times \mathcal{C}(\boldsymbol{z})$ as in (12.17), then averaging the $d(\boldsymbol{x}^{(k)}, \boldsymbol{y}^{(k)})$'s, will at least yield an unbiased estimate of the true $d^A(\boldsymbol{w}, \boldsymbol{z})$.

12.2 Hausdorff distances

Critchlow (1985) develops expressions for the Hausdorff distance (12.3, 12.4) based on the Spearman, footrule, Kendall, Hamming, and Ulam distances. He also treats Cayley's distance for tied rankings with patterns $\boldsymbol{m} = (1, 1, \ldots, 1, m - q)$, where there are q 1's. (So people rank their top q choices, the rest being tied at the bottom.) See his theorem in section D. We present the results for all but Cayley's distance here, as well as that for the maximum distance.

We can think of finding the Hausdorff distance as a two-person game, the two people being W and Z, with tied rankings $\boldsymbol{w}$ and $\boldsymbol{z}$. Each has a set of compatible rankings to choose from, $\mathcal{W}$ and $\mathcal{Z}$, respectively. Fix a distance d on the rankings. There are two rounds. In the first round, W first chooses a ranking $\boldsymbol{x}$ from $\mathcal{W}$, then Z chooses $\boldsymbol{y}$ from $\mathcal{Z}$ knowing what W's choice is. W's score is then $d(\boldsymbol{x}, \boldsymbol{y})$. In round two, the roles are switched, so that Z chooses a $\boldsymbol{y}' \in \mathcal{Z}$ first, then W chooses an $\boldsymbol{x}' \in \mathcal{W}$, and Z's score is $d(\boldsymbol{x}', \boldsymbol{y}')$. Whoever has the highest score wins, so W's goal in the first round is to choose $\boldsymbol{x}$ to maximize, and Z's to choose $\boldsymbol{y}$ to minimize, $d(\boldsymbol{x}, \boldsymbol{y})$. In round two it's the reverse. As a game it is not very exciting, since the outcome is predetermined unless someone makes a mistake. But we do have that the Hausdorff distance is the winning score, i.e.,

$$d^H(\boldsymbol{w}, \boldsymbol{z}) = \max\{d(\boldsymbol{x}, \boldsymbol{y}), d(\boldsymbol{x}', \boldsymbol{y}')\}. \qquad (12.38)$$

The strategy for Spearman, footrule, Kendall, Ulam, and maximum is the same. We consider an illustrative example. Let $\boldsymbol{w} = (1,1,1,2,2,3,3,3)$ and $\boldsymbol{z} = (2,4,3,3,1,3,2,3)$. We start with round one. Since the distances are label-invariant, we can rearrange the objects so that $\boldsymbol{w}$ is nondecreasing (it already is), and for objects with w_j equal given value i, the rankings in $\boldsymbol{z}$ are nondecreasing. Thus we have

$$\begin{array}{rcccccccc} \boldsymbol{w} \to & 1 & 1 & 1 & 2 & 2 & 3 & 3 & 3 \\ \boldsymbol{z} \to & 2 & 3 & 4 & 1 & 3 & 2 & 3 & 3 \end{array}. \tag{12.39}$$

W wants to maximize the distance, so chooses an $\boldsymbol{x}$ that as much as possible goes in the reverse order of $\boldsymbol{z}$. That is, for $w_j = 1$, the z_j's are $2,3,4$, so W chooses the corresponding x_j's to be 3, 2, 1. For $x_j = 2$, the x_j's are 5, 4. For $x_j = 3$, the x_j's are 8, 7, 6, though 8, 6, 7 works just as well. Thus we have

$$\begin{array}{rcccccccc} \boldsymbol{x} \to & 3 & 2 & 1 & 5 & 4 & 8 & 7 & 6 \\ \boldsymbol{z} \to & 2 & 3 & 4 & 1 & 3 & 2 & 3 & 3 \end{array}', \tag{12.40}$$

i.e., within each grouping of w_j's we put the compatible ranks in reverse order. Now rearrange again so that the $\boldsymbol{z}$ is in nondecreasing order, and within each value of z_j's, the $\boldsymbol{x}$ is in increasing order

$$\begin{array}{rcccccccc} \boldsymbol{x} \to & 5 & 3 & 8 & 2 & 4 & 6 & 7 & 1 \\ \boldsymbol{z} \to & 1 & 2 & 2 & 3 & 3 & 3 & 3 & 4 \end{array}. \tag{12.41}$$

Now Z chooses $\boldsymbol{y}$ to minimize the distance, which entails for each value of z_j choosing the y_j's in the same order as the x_j's. Since the latter are increasing within each group, we take $\boldsymbol{y} = (1,2,\ldots,m)$:

$$\begin{array}{rcccccccc} \boldsymbol{x} \to & 5 & 3 & 8 & 2 & 4 & 6 & 7 & 1 \\ \boldsymbol{y} \to & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array}. \tag{12.42}$$

Now W's score is $d(\boldsymbol{x},\boldsymbol{y})$.

Round two switches the roles of W and Z, the steps beginning and ending as

$$\begin{array}{rcccccccc} \boldsymbol{z} \to & 1 & 2 & 2 & 3 & 3 & 3 & 3 & 4 \\ \boldsymbol{w} \to & 2 & 1 & 3 & 1 & 2 & 3 & 3 & 1 \end{array} \to \begin{array}{rcccccccc} \boldsymbol{y}' \to & 3 & 7 & 8 & 1 & 6 & 2 & 4 & 5 \\ \boldsymbol{x}' \to & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array}. \tag{12.43}$$

Then Z's score is $d(\boldsymbol{x}',\boldsymbol{y}')$, and the Hausdorff distance is as in (12.38). The results for the four distances are next, along with the averaging distance:

	Spearman	Footrule	Kendall	Ulam	Max
$d(\boldsymbol{x},\boldsymbol{y})$	96	20	15	4	7
$d(\boldsymbol{x}',\boldsymbol{y}')$	98	26	15	4	5
$d^H(\boldsymbol{w},\boldsymbol{z})$	98	26	15	4	7
$d^A(\boldsymbol{w},\boldsymbol{z})$	96.5	22.92	15.5	4.63	6.09

(12.44)

We formalize this process. We will say that a pair of possibly tied m rankings ($\boldsymbol{w}$,$\boldsymbol{z}$) is in **lexicographical order** if for any $i < j$, $w_i \leqslant w_j$, and if $w_i = w_j$, then $z_i \leqslant z_j$. See, e.g., (12.39). Given $\boldsymbol{w}$, $\boldsymbol{z}$, and a distance d among the distances Spearman, footrule, Kendall, Ulam, and maximum, $d^H(\boldsymbol{w},\boldsymbol{z})$ can be calculated using the following steps.

1. Relabel the objects so that $(\boldsymbol{w},\boldsymbol{z})$ is in lexicographical order.

2. Defining $l^{(w)}$ and $u^{(w)}$ as in (12.15) and (12.16) for $\boldsymbol{w}$ (with values from 1 to K), set

$$\boldsymbol{x} = (u_1^{(w)}, \ldots, l_1^{(w)}, u_2^{(w)}, \ldots, l_2^{(w)}, \ldots, u_K^{(w)}, \ldots, l_K^{(w)}). \tag{12.45}$$

   See (12.40).

3. Relabel the objects so that now $(\boldsymbol{z}, \boldsymbol{x})$ is in lexicographical order. See (12.41).

4. Relabel the objects so that $(\boldsymbol{z}, \boldsymbol{w})$ is in lexicographical order.

5. Define $l^{(z)}$ and $u^{(z)}$ for $\boldsymbol{z}$ (with values from 1 to L), and set

$$\boldsymbol{y}' = (u_1^{(z)}, \ldots, l_1^{(z)}, u_2^{(z)}, \ldots, l_2^{(z)}, \ldots, u_L^{(z)}, \ldots, l_L^{(z)}). \tag{12.46}$$

6. Relabel the objects so that $(\boldsymbol{w}, \boldsymbol{y}')$ is in lexicographical order.

7. With $\boldsymbol{\nu} = (1, \ldots, m)$, $\boldsymbol{x}$ from step 3, and $\boldsymbol{y}'$ from step 6, we have

$$d^H(\boldsymbol{w}, \boldsymbol{z}) = \max\{d(\boldsymbol{x}, \boldsymbol{\nu}), d(\boldsymbol{\nu}, \boldsymbol{y}')\}. \tag{12.47}$$

The result for the first four distances are given by the theorem in Section D of Critchlow (1985). We can show that the method also holds for the maximum distance by using Lemma 2 in that section. Consider the partial ordering on $\mathcal{P}_m$ proposed by Henery (1981) that states that $\boldsymbol{x}$ is less than $\boldsymbol{y}$ if they agree on all but two objects, and only for $\boldsymbol{x}$, they are in increasing order:

$$\boldsymbol{x} <_H \boldsymbol{y} \text{ if for some } i < j \;\; x_k = y_k \text{ for } k \neq i, j, \;\; x_i < x_j, \text{ and } y_i > y_j. \tag{12.48}$$

Critchlow's lemma shows that if $\boldsymbol{x} <_H \boldsymbol{y}$ implies that $d(\boldsymbol{\nu}, \boldsymbol{x}) \leqslant d(\boldsymbol{\nu}, \boldsymbol{y})$ for $\boldsymbol{\nu}$ as in step 7, then the Hausdorff measure can be calculated as in (12.47) using the steps 1 to 7. To see that the lemma can be applied to the maximum distance, take $\boldsymbol{x}$ and $\boldsymbol{y}$ as in (12.48). Then $(x_i, x_j) = (y_j, y_i)$, and we will show that

$$\max\{|i - y_i|, |j - y_j|\} = \max\{|i - x_j|, |j - x_i|\} \geqslant \max\{|i - x_i|, |j - x_j|\}. \tag{12.49}$$

Consider the cases defined by the relative ordering of x_i, x_j, i, j:

$$\begin{aligned}
x_i \leqslant i, x_j \leqslant j &\;\Rightarrow\; |j - x_i| > |i - x_i| \;\&\; |j - x_j|; \\
x_i \leqslant i < j \leqslant x_j &\;\Rightarrow\; |j - x_i| > |i - x_i| \;\;\&\;\; |i - x_j| > |j - x_j|; \\
i \leqslant x_i < x_j \leqslant j &\;\Rightarrow\; |j - x_i| > |j - x_j| \;\;\&\;\; |i - x_j| > |i - x_i|; \\
i \leqslant j, x_i \leqslant x_j &\;\Rightarrow\; |i - x_j| > |i - x_i| \;\&\; |j - x_j|.
\end{aligned} \tag{12.50}$$

Any of those possibilities implies (12.49), which in turn implies

$$\begin{aligned}
d_{Max}(\boldsymbol{\nu}, \boldsymbol{y}) &= \max\{\max_{k \neq i,j} |k - x_k|, |i - x_j|, |j - x_i|\} \\
&\geqslant \max\{\max_{k \neq i,j} |k - x_k|, |i - x_i|, |j - x_j|\} = d_{Max}(\boldsymbol{\nu}, \boldsymbol{x}).
\end{aligned} \tag{12.51}$$

The above does not work for Hamming's distance, nor I imagine for any bi-invariant distance. We again refer to the theorem of Critchlow (1985). We need two $K \times L$ matrices, $\boldsymbol{A}$ and $\boldsymbol{B}$, where

$$a_{ij} = \#\{k \,|\, w_k = i \;\&\; z_k = j\} \text{ and } b_{ij} = \#\{l_w(i), \ldots, \ldots, u_w(i)\} \cap \{l_z(j), \ldots, \ldots, u_z(j)\}. \quad (12.52)$$

Then with $\boldsymbol{\lambda_w}$ and $\boldsymbol{\lambda_z}$ being the patterns of ties (12.13), we have

$$d^H_{Ham}(\boldsymbol{w}, \boldsymbol{z}) = \max\{\sum_{i=1}^{K}\sum_{j=1}^{L}(a_{ij} + b_{ij} - \lambda_{\boldsymbol{w}}(i))^+, \sum_{i=1}^{K}\sum_{j=1}^{L}(a_{ij} + b_{ij} - \lambda_{\boldsymbol{z}}(j))^+\}. \quad (12.53)$$

For Cayley's distance, we can find the Hausdorff extension by explicitly enumerating the compatible rankings, which is practical only if the sizes of $\mathcal{W}$ and $\mathcal{Z}$ are not too large. We refer to Critchlow (1985) for the extension in the special case mentioned in the first paragraph.

12.2.1 Bi-invariant distances

If d is bi-invariant, any of the formulas in (12.3), (12.5), and (12.6) can certainly be used, but whether the compatibility approach to ties makes sense when the situation is inherently rank-invariant is questionable. For example, take $m = 3$, and let $\boldsymbol{w} = (2, 1, 2)$ be a tied ranking. Then the compatible set is $\mathcal{C}(\boldsymbol{w}) = \{(2, 1, 3), (3, 1, 2)\}$. Consider the permutation of the ranks where 1 and 2 are switched. What is the new tied ranking $\boldsymbol{w}^*$? The compatibility set becomes $\mathcal{C}(\boldsymbol{w}^*) = \{(1, 2, 3), (3, 2, 1)\}$. But there is no $\boldsymbol{w}^*$ that yields this compatibility set. The problem is that two objects can be tied only if their rankings are consecutive, and consecutivity is not invariant under rank permutations. There may be an alternate definition of compatibility sets that will preserve rank-invariance when there are ties. Instead, it is probably advisable to take a different tack.

Hamming's distance was originally, and still primarily, used to measure the distance of any two finite sequences, not just permutations of integers. Thus it has an immediate extension to rankings with non-distinct elements. Matching is key, so it may be easier to think of the ranks as a set of m colors, say. A ranking is a vector of m colors. Ties would be repetitions of the same color. Then for two such vectors, Hamming's distance would be as before,

$$d_{Ham}(\boldsymbol{w}, \boldsymbol{z}) = m - \#\{i \,|\, w_i = z_i\}. \quad (12.54)$$

So if $m = 4$ and the colors are {red, blue, green, yellow}, we could have $\boldsymbol{w}$ = (red, green, red, blue) and $\boldsymbol{z}$ = (blue, green, yellow, blue), so that $d_{Ham}(\boldsymbol{w}, \boldsymbol{x}) = 2$. The setup is invariant under permuting the colors. E.g., the permutation red $\to$ green, green $\to$ yellow, and yellow $\to$ red yields

$$\boldsymbol{w}^* = (\text{green, yellow, green, blue}) \text{ and } \boldsymbol{z}^* = (\text{blue, yellow, red, blue}), \quad (12.55)$$

which again has $d_{Ham}(\boldsymbol{w}^*, \boldsymbol{z}^*) = 2$. Using integers, we'd have a tied ranking be a vector with components drawn from $\{1, \ldots, m\}$, so that now, e.g., $\boldsymbol{w} = (1, 3, 4, 1, 3)$ would be a valid tied ranking.

Cayley's distance can be similarly extended, at least for tied rankings with the same patterns. That is, suppose $\boldsymbol{x}$ and $\boldsymbol{y}$ are tied ranks with the values from 1 to K, where $\#\{i \,|\, w_i = k\} = \#\{i \,|\, z_i = k\}$ for each k. Then $d_{Cay}(\boldsymbol{w}, \boldsymbol{z})$ is the minimum number of transpositions to bring $\boldsymbol{z}$ to $\boldsymbol{x}$. Diaconis (1988, page 127) has another suggested distance using adjacent transpositions.

12.3 Alternatives for Ulam and maximum

As mentioned at the start of this Section 12.1, the calculations for the averaging applied to the Ulam or maximum distance can be challenging if the m_a and n_a are not small. Tractable alternatives include the Hausdorff distance in Section 12.2, as well as the minimum and maximum approaches in (12.6).

For the maximum distance, the latter two methods are easily applied since we can calculate them on the ranges of (X_i, Y_i) for each i individually. That is,

$$\begin{aligned} d^M_{Max}(\boldsymbol{w}, \boldsymbol{z}) &= \max_{1 \leqslant i \leqslant m} \max\{|u_i^{(\boldsymbol{w})} - l_i^{(\boldsymbol{z})}|, |u_i^{(\boldsymbol{z})} - l_i^{(\boldsymbol{w})}|\} \text{ and} \\ d^m_{Max}(\boldsymbol{w}, \boldsymbol{z}) &= \max_{1 \leqslant i \leqslant m} \max\{l_i^{(\boldsymbol{z})} - u_i^{(\boldsymbol{w})}, l_i^{(\boldsymbol{w})} - u_i^{(\boldsymbol{z})}, 0\}. \end{aligned} \tag{12.56}$$

Turn to Ulam's distance. With no ties, Ulam's distance is based on the length of the longest increasing subsequence of the y_i's, once the components are rearranged so that $\boldsymbol{x} = \boldsymbol{\omega} = (1, \ldots, m)$. Then we can define Ulam's distance as in (8.4) to be $m - S$, where S is the longest set of distinct indices $i_1, \ldots, i_S$ such that $x_{i_1}, \ldots, x_{i_S}$ and $y_{i_1}, \ldots, y_{i_S}$ are both strictly increasing sequences. With tied rankings $\boldsymbol{w}$ and $\boldsymbol{z}$, we could consider either or both of these sequences to be nondecreasing. The most conservative approach, which yields $d^M_{Ulam}(\boldsymbol{w}, \boldsymbol{z})$, is to demand strict increasingness. That is, we find longest sequence such that

$$w_{i_1} < w_{i_2} < \cdots < w_{i_S} \text{ and } z_{i_1} < z_{i_2} < \cdots < z_{i_S}. \tag{12.57}$$

The most generous is to find the longest sequence of nondecreasing components, which yields $d^m_{Ulam}(\boldsymbol{w}, \boldsymbol{z})$:

$$w_{i_1} \leqslant w_{i_2} \leqslant \cdots \leqslant w_{i_S} \text{ and } z_{i_1} \leqslant z_{i_2} \leqslant \cdots \leqslant z_{i_S}. \tag{12.58}$$

We could also contemplate using strict inequality for just one vector, i.e., consider $m - S$ or $m - T$, where S is the longest sequence such that

$$w_{i_1} < w_{i_2} < \cdots < w_{i_S} \text{ and } z_{i_1} \leqslant z_{i_2} \leqslant \cdots \leqslant z_{i_S}, \tag{12.59}$$

while T is the longest such that

$$w_{i_1} \leqslant w_{i_2} \leqslant \cdots \leqslant w_{i_T} \text{ and } z_{i_1} < z_{i_2} < \cdots < z_{i_T}. \tag{12.60}$$

It may be that the particular context of the ranking situation dictates one of the above to be the most appropriate. If not, the maximum (12.57) or minimum (12.58) approaches are generally too extreme. Also, the alternative in (12.59) is generally conservative, since S is bounded by K, the number of distinct values in $\boldsymbol{w}$, and similarly for (12.60).

Chapter 13

Tied Ranks: Spearman

If $\boldsymbol{w}$ and $\boldsymbol{z}$ are rank vectors with ties, then Spearman's distance can be given by expressions analogous to (4.39) and (4.6), replacing the rank vectors by their midrank vectors. That is,

$$d^{A}_{Spear}(\boldsymbol{w},\boldsymbol{z}) = \frac{2m(m+1)(2m+1)}{3} - 2\sum_{i=1}^{m} r_i s_i, \quad \text{and}$$

$$= \mu_{Spear}(m) - 2\sum_{i=1}^{m}(r_i-\nu)(s_i-\nu), \tag{13.1}$$

where

$$\boldsymbol{r} \equiv \text{rank}(\boldsymbol{w}) \ \text{ and } \ \boldsymbol{s} \equiv \text{rank}(\boldsymbol{z}) \tag{13.2}$$

are the corresponding vectors of midranks (12.25),

$$\mu_{Spear}(m) \equiv E[d^{A}_{Spear}(\boldsymbol{W},\boldsymbol{Z})] = \frac{m(m^2-1)}{6}, \ \text{ and } \ \nu = (m+1)/2. \tag{13.3}$$

Consider the range of d^{A}_{Spear}. Let $\boldsymbol{r}^{(o)}$ and $\boldsymbol{s}^{(o)}$ be the midrank vectors with the elements in order from lowest to highest, e.g.,

$$\boldsymbol{r}^{(o)} = (t_1,\ldots,t_1,t_2,\ldots,t_2,\ldots,t_K,\ldots,t_K), \ \ t_a = \frac{m_{<a}+m_{\leqslant a}+1}{2}, \tag{13.4}$$

and there are m_a of the t_a's, $a = 1,\ldots,K$. The $\boldsymbol{s}^{(o)}$ is defined similarly in terms of n_b's. Then we have the rearrangement inequalities

$$\nu_1 \equiv \sum_{i=1}^{m} r_i^{(o)} s_{m-i+1}^{(o)} \leqslant \sum_{i=1}^{m} r_i s_i \leqslant \sum_{i=1}^{m} r_i^{(o)} s_i^{(o)} \equiv \nu_2. \tag{13.5}$$

Thus with $c_m = 2m(m+1)(2m+1)/3$,

$$u_1 \equiv c_m - 2\nu_2 \leqslant d^{A}_{Spear}(\boldsymbol{w},\boldsymbol{z}) \leqslant c_m - 2\nu_1 \equiv u_2. \tag{13.6}$$

Since the midranks are either integers or half integers, so are the values of d^{A}_{Spear}. Thus we have

$$d^{A}_{Spear}(\boldsymbol{w},\boldsymbol{z}) \in \{u_1 + \frac{i}{2} \,|\, i = 0,\ldots,2(u_2-u_1)\} \equiv \mathcal{U}(\boldsymbol{m},\boldsymbol{n}), \tag{13.7}$$

though depending on the actual values in the $\boldsymbol{r}$ and $\boldsymbol{s}$, it may be that not all values in the set are achievable.

In Chapter 4, we treat the case with no ties. A splitting algorithm is used to find the exact distribution of Spearman's distance for $m \leqslant 25$, and Edgeworth expansions are used to find approximations for larger m. Fortunately, the approximations work well for m near 25, so one of these two techniques will always be useful. In the ties case, we again have an exact splitting algorithm and Edgeworth expansions, but they do not appear to cover all possibilities adequately. That is, the splitting algorithm works quickly for $m \leqslant 13$, and generally Edgeworth expansions are good for large m and most patterns of ties, but depending on m and the patterns of ties, it may be that neither approach works well. For some patterns, especially those with few distinct values in the rankings, we can use a contingency table approach to find the exact distribution. In case none of these approaches is useful, we use randomizations. An extra difficulty is that it is not always easy to determine which approach is best.

The next section briefly covers two methods for finding the exact distribution, the splitting algorithm and one based on contingency tables. Section 13.2 presents formulas for the moments, and Section 13.3 gives conditions for asymptotic normality to hold. Section **??** compares the approaches and makes some recommendations about when to use which approach.

13.1 Exact distribution

We have two algorithms for finding the exact distribution of Spearman's distribution with ties. The first is the splitting algorithm from Section 4.3 for the case with no ties. Though when there are ties, some of the symmetries that speed up the algorithm are unavailable. This procedure is reasonably fast for $m \leqslant 13$ or so.

The other algorithm is based on writing a pair of tied rankings as a contingency table, then writing Spearman's distance as a simple function of the table. The exact distribution is found by enumerating all the possible contingency tables given the rankings, and finding each one's probability. This procedure is reasonable if there are not too many possible tables, e.g., under about three million, which generally needs $m \leqslant 50$, possibly much less. Unfortunately, it is not easy to determine the number of tables apriori, i.e., without enumerating. See the end of Section 13.1.2.

If $K = 2$ and $\boldsymbol{Z}$ has no ties (so $L = m$), we have the Mann-Whitney/Wilcoxon statistic, in which case the Spearman and Kendall distances adjusted for ties are equivalent. Thus we can use the exact algorithm in Section 14.3.

13.1.1 Splitting algorithm

We start by considering the distribution of the random part of Spearman's distance (13.1), $\boldsymbol{S}\boldsymbol{r}'$ where $\boldsymbol{S} = \text{rank}(\boldsymbol{Z})$. Take $m_1 \approx m/2$ and $m_2 = m - m_1$. As in Section 3.3, we consider all splittings $\mathcal{S} = (\mathcal{R}_1, \mathcal{R}_2)$, where $\mathcal{R}_1$ is a subset of m_1 distinct elements from $\{1, \ldots, m\}$, and $\mathcal{R}_2$ is its complement. For given splitting, let

$$\boldsymbol{s}^{(j)} = (s_{i_1}, \ldots, s_{i_{m_j}}) \text{ where } \mathcal{R}^{(j)} = \{i_1, \ldots, i_{m_j}\},\ i_1 < \cdots < i_{m_j};\ j = 1, 2. \tag{13.8}$$

Then conditional on the splitting,

$$\begin{aligned} \boldsymbol{S}^{(1)} = (S_1, \ldots, S_{m_1}) \,|\, \mathcal{S} &\sim \text{Uniform}(\text{Permutations of } \boldsymbol{s}^{(1)}\}, \\ \boldsymbol{S}^{(2)} = (S_{m_1+1}, \ldots S_m) \,|\, \mathcal{S} &\sim \text{Uniform}(\text{Permutations of } \boldsymbol{s}^{(2)}\}, \end{aligned} \tag{13.9}$$

where $\boldsymbol{S}^{(1)}$ and $\boldsymbol{S}^{(2)}$ are conditionally independent, as in (3.28) and (3.29). We write

$$\boldsymbol{S}\boldsymbol{r}' = \boldsymbol{S}^{(1)}\boldsymbol{r}^{(1)\prime} + \boldsymbol{S}^{(2)}\boldsymbol{r}^{(2)\prime}. \tag{13.10}$$

Then as in (3.31), we can enumerate over the permutations of each $\boldsymbol{s}^{(j)}$ to find the conditional probabilities

$$f_j(u \,|\, \mathcal{S}) = P[\boldsymbol{S}^{(j)}\boldsymbol{r}^{(j)\prime} = u \,|\, \mathcal{S}], \;\; j = 1, 2, u \in \mathcal{C}, \tag{13.11}$$

where $\boldsymbol{r}^{(1)} = (r_1, \ldots, r_{m_1})$, $\boldsymbol{r}^{(2)} = (r_{m_1+1}, \ldots, r_m)$. The $\mathcal{C}$ is a set large enough to accomodate all the spaces we need, which at most is

$$\mathcal{C} = \{1 + k/4 \,|\, k = 0, \ldots, 4(m-1)\}. \tag{13.12}$$

Then the distribution of $\boldsymbol{S}\boldsymbol{r}'$ given the splitting is, as in (3.32),

$$f(x \,|\, \mathcal{S}) = \sum_{u \in \mathcal{C}} f_1(u \,|\, \mathcal{S}) f_2(x - u \,|\, \mathcal{S}); \tag{13.13}$$

and the unconditional distribution is, as in (3.36),

$$f(x) = P[\boldsymbol{S}\boldsymbol{r}' = x] = \binom{m}{m_1}^{-1} \sum_{\text{Splittings } \mathcal{S}} f(x \,|\, \mathcal{S}), \;\; x \in \mathcal{C}. \tag{13.14}$$

The distribution of the distance is

$$P[d^A_{\text{Spear}}(\boldsymbol{w}, \boldsymbol{z}) = u] = f\left(\frac{c_m - u}{2}\right), \; u \in \mathcal{U}(\boldsymbol{m}, \boldsymbol{n}), \tag{13.15}$$

where $\mathcal{U}$ is given in (13.7).

13.1.2 Contingency tables

If K and L are small, and m not too large, a more efficient approach to finding the exact distribution is based on representing the tied rankings in a contingency table. Thus for tied rankings $\boldsymbol{w}$ and $\boldsymbol{z}$, set

$$T_{ab}(\boldsymbol{w}, \boldsymbol{z}) = \#\{i \,|\, w_i = a, z_i = b\}, \;\; 1 \leqslant a \leqslant K, \; 1 \leqslant b \leqslant L, \tag{13.16}$$

as in (12.37). Then the matrix $\boldsymbol{T}(\boldsymbol{w}, \boldsymbol{z})$ of the $T_{ab}(\boldsymbol{w}, \boldsymbol{z})$'s is a $K \times L$ contingency table with row totals $\boldsymbol{m} = (m_1, \ldots, m_K)$ and column totals $\boldsymbol{n} = (n_1, \ldots, n_L)$, the vectors of counts as in (12.13). Then for a general Hoeffding distance, if $\boldsymbol{T}(\boldsymbol{w}, \boldsymbol{z}) = \boldsymbol{t}$, we have

$$d^A_{\text{Hoeff}}(\boldsymbol{w}, \boldsymbol{z}) = \sum_{a=1}^{K} \sum_{b=1}^{L} t_{ab} \delta^*(a, b) \tag{13.17}$$

for δ^* in (12.21). This formula applied to Spearman is

$$d^A_{Spear}(\boldsymbol{w},\boldsymbol{z}) = \overline{d}^A_{Spear}(\boldsymbol{t}) \equiv c_m - 2\sum_{a=1}^{K}\sum_{b=1}^{L} t_{ab}\frac{m_{<a}+m_{\leqslant a}+1}{2}\frac{n_{<b}+n_{\leqslant b}+1}{2}. \tag{13.18}$$

The exact distribution can be found by iterating over all possible such contingency tables, and multiplying by the tables' probabilties, which are given by

$$g(\boldsymbol{t}) \equiv P[\boldsymbol{T}(\boldsymbol{W},\boldsymbol{Z}) = \boldsymbol{t}] = \frac{\prod_{a=1}^{K} m_a! \prod_{b=1}^{L} n_b!}{m! \prod_{a=1}^{K}\prod_{b=1}^{L} t_{ab}!} \tag{13.19}$$

Thus

$$f(u) \equiv P[d^A_{Spear}(\boldsymbol{W},\boldsymbol{Z}) = u] = \sum_{\boldsymbol{t}\,|\,\overline{d}^A_{Spear}(\boldsymbol{t})=u} g(\boldsymbol{t}), \quad u \in \mathcal{U}(\boldsymbol{m},\boldsymbol{n}). \tag{13.20}$$

Verbeek & Kroonenberg (1985) and Mehta & Patel (1983) describe some iterative procedures, and provide references to others. We next sketch a fairly simple approach. We wish to find all the $K \times L$ contingency tables $\boldsymbol{t}$ with row marginals $\boldsymbol{m} = (m_1,\ldots,m_K)$ and column marginals $\boldsymbol{n} = (n_1,\ldots,n_L)$. Since the last row and column are determined by the other cells, it is enough to consider the upper-left $(K-1)\times(L-1)$ part of the table. We next string out the rows, defining the $(K-1)(L-1)$-length vector

$$\boldsymbol{v} = (t_{11},\ldots,t_{1,L-1},t_{21},\ldots,t_{2,L-1},\ldots,t_{K-1,1},\ldots,t_{K-1,L-1}). \tag{13.21}$$

We fill the $\boldsymbol{v}$ in from left to right. For each element i, we need the minimum and maximum values it could take on, given the first $i-1$ elements (and the marginal constraints). Denote these minima and maxima, respectively, by

$$\alpha_i(v_1,\ldots,v_{i-1}) \text{ and } \omega_i(v_1,\ldots,v_{i-1}). \tag{13.22}$$

For $i = 1$, these are the minimum and maximum the first cell could possibly be.

The algorithm has two basic steps. We start with $f(u) = 0$ for all u, $i = 1$, and $v_1 = l_1$.

1. Given the values of $v_1,\ldots,v_i$, fill in the rest of the table from left to right with the minimum possible at that stage:

$$v_j \leftarrow \alpha_j(v_1,\ldots,v_{j-1}), j = i+1,\ldots,(K-1)(L-1). \tag{13.23}$$

The final result is a valid table, which we record. Letting $\boldsymbol{t}$ be its $K \times L$ version, we update the f:

$$f(u) \leftarrow f(u) + P[\boldsymbol{T} = \boldsymbol{t}], \text{ where } u = \overline{d}^A_{Spear}(\boldsymbol{t}). \tag{13.24}$$

2. Start at the end of the table from step 1, and backtrack until you find an element that can be increased by one. That is,

$$i \leftarrow \max\{j\,|\,v_j < \omega_j(v_1,\ldots,v_{j-1})\}. \tag{13.25}$$

If i does not exist (i.e., all cells are at their maxima), then we are done. Otherwise, set

$$v_i \leftarrow v_i + 1. \tag{13.26}$$

Go back to step 1, where $v_1,\ldots,v_i$ are now determined.

To find the bounds in (13.22), we use the $K \times L$ version of the table. Suppose the i^{th} element in $\boldsymbol{v}$ corresponds to cell (a, b), so that $v_i = t_{ab}$. It is enough to consider the part of the table extending from cell (a, b) to the lower right, and collapse it into a 2×2 table with t_{ab} as the upper-left cell. Extend the definition in (12.14) to the t_{ab}, so that

$$t_{a,\leqslant b} = \sum_{l=1}^{b} t_{a,l}, \quad t_{\leqslant a,\leqslant b} = \sum_{k=1}^{a} \sum_{l=1}^{b} t_{kl}, \tag{13.27}$$

and similarly for other inequalities. Then the collapsed table looks like

$$\begin{array}{cc|c} t_{ab} & * & m_a - t_{a,<b} \\ * & * & * \\ \hline n_b - t_{<a,b} & * & m - m_{<a} - n_{<b} + t_{<a,<b} \end{array} \tag{13.28}$$

Here the asterisked cells can be found by subtraction. Note that the marginals are given in terms of the known quantities $\boldsymbol{m}, \boldsymbol{n}$ and $v_1, \ldots, v_{i-1}$. Now we are in the familiar hypergeometric situation, where we have that t_{ab} is bounded from above by its row total and its column total, and bounded from below by zero and the sum of the row and column totals minus the overall total. Thus the quantities in (13.22) are

$$\alpha_i = \max\{0, m_{\leqslant a} + n_{\leqslant b} - m - t_{<a,\leqslant b} - t_{a,<b}\} \text{ and } \omega_i = \min\{m_a - t_{a,<b}, n_b - t_{<a,b}\} \tag{13.29}$$

Estimating the number of tables

To decide whether enumerating all the possible contingency tables is viable for a particular pair of tied rankings, we would like to efficiently approximate the total number of such tables. There are several approaches. See Diaconis & Gangolli (1995) for an overview. Most of these demand larger m than we can accommodate, or other conditions such as sparseness. The best simple approach we have found for our purposes is in Good (1976). For fixed $\boldsymbol{w}$ and $\boldsymbol{z}$, let $\#\mathcal{T}$ be the number of possible $\boldsymbol{t}$'s. Then Good suggests that

$$\#\mathcal{T} \approx \frac{A_L(\boldsymbol{m}) A_K(\boldsymbol{n})}{B_{KL}(m)}, \tag{13.30}$$

where

$$\begin{aligned} A_L(\boldsymbol{m}) &= \prod_{a=1}^{K} \binom{m_a + L - 1}{m_a} \text{ and,} \\ B(\boldsymbol{m}, \boldsymbol{n}) &= \binom{m + KL - 1}{m}. \end{aligned} \tag{13.31}$$

Note that $A_L(\boldsymbol{m})$ is the number of $K \times L$ tables with row marginals given by $\boldsymbol{m}$ (and no restrictions on column marginals); $A_K(\boldsymbol{n})$ is the number with column marginals $\boldsymbol{n}$; and $B_{KL}(m)$ is the number with no restrictions other than having a total of m. The intuition given is based on imagining randomly picking a $K \times L$ table with a total of m. The chance it has row marginals $\boldsymbol{m}$ is $A_L(\boldsymbol{m})/B_{KL}(m)$, and the chance that it has column marginals $\boldsymbol{n}$ is $A_K(\boldsymbol{n})/B_{KL}(m)$. If the

row and column marginals are independent (which they are generally not), then the chance of any single table having those marginals is the product of those two probabilities, and the number of such tables is $A_L(\boldsymbol{m})A_K(\boldsymbol{n})/B_{KL}(m)$. He also proposes the extension

$$\#\mathcal{T} \approx \frac{A_L(\boldsymbol{m})A_K(\boldsymbol{n})C_L(\boldsymbol{m})C_K(\boldsymbol{n})}{B_{KL}(m)}, \text{ where } C_L(\boldsymbol{m}) = \frac{1.3m^2}{L\sum_{a=1}^{K} m_a^2}. \tag{13.32}$$

The C factors in (13.32) are given as adjustments for roughness. We have found this adjustment does not improve the estimates, so we will focus on (13.30).

A slightly slower approach uses simulations to estimate $\#\mathcal{T}$. In (13.19) we have $g(\boldsymbol{t})$, the probability of any particular table $\boldsymbol{t}$. Consider the random variable $1/g(\boldsymbol{T}(\boldsymbol{W},\boldsymbol{Z}))$ for random $(\boldsymbol{X},\boldsymbol{Y})$. Then

$$E\left[\frac{1}{g(\boldsymbol{T}(\boldsymbol{X},\boldsymbol{Y}))}\right] = \sum_{\boldsymbol{t}\in\mathcal{T}} \frac{1}{g(\boldsymbol{t})} P[\boldsymbol{T}(\boldsymbol{W},\boldsymbol{Z}) = \boldsymbol{t}] = \#\mathcal{T}. \tag{13.33}$$

To estimate $\#\mathcal{T}$ we randomly sample n observations $(\boldsymbol{x},\boldsymbol{y})$, for each find their tables $\boldsymbol{t}$, and average their $1/g(\boldsymbol{t})$. Even for $n = 1000$, this estimate improves on that in (13.30). The variance of $1/g$ is

$$Var[1/g(\boldsymbol{T}(\boldsymbol{W},\boldsymbol{Z}))] = E\left[\frac{1}{g(\boldsymbol{T}(\boldsymbol{X},\boldsymbol{Y}))^2}\right] - \#\mathcal{T}^2 = \sum_{\boldsymbol{t}\in\mathcal{T}} \frac{1}{g(\boldsymbol{t})} - \#\mathcal{T}^2, \tag{13.34}$$

which can be quite high if some of the individual probabilities are very low, though in our simulations it hasn't seemed to be too bad. The estimated standard errors tend to be of the order of 10% of the mean.

13.2 Moments

The moments for Spearman's distance with ties are found much as they are when there are no ties in Sections 4.1 and 4.2. Let $\boldsymbol{w}$ and $\boldsymbol{z}$ be fixed values for the vectors with ties, so that the distributions of the vectors can be represented as $\boldsymbol{W} = \boldsymbol{wQ}$ and $\boldsymbol{Z} = \boldsymbol{zQ}$ where, as in (3.4), $\boldsymbol{Q}$ is distributed uniformly over the $m \times m$ permutations matrices. Thus since $\text{rank}(\boldsymbol{wQ}) = \text{rank}(\boldsymbol{w})\boldsymbol{Q}$, setting $\boldsymbol{r} = \text{rank}(\boldsymbol{w})$ and $\boldsymbol{s} = \text{rank}(\boldsymbol{z})$,

$$\begin{aligned} Var[\text{rank}(\boldsymbol{W})\text{rank}(\boldsymbol{Z})'] &= Var[\boldsymbol{rQs}'] \\ &= \frac{(\boldsymbol{rHr}')(\boldsymbol{sHs}')}{m-1}. \end{aligned} \tag{13.35}$$

The covariance of $\boldsymbol{Q}$ is given in (3.7), and $\boldsymbol{H}$ is the centering matrix in (3.8). Then by (13.1), because of the factor 2,

$$Var[d^A_{Spear}(\boldsymbol{W},\boldsymbol{Z})] = 4\,\frac{\sum(r_i-\nu)^2\sum(s_i-\nu)^2}{m-1}. \tag{13.36}$$

For higher moments, we can follow the development in Section 4.2, but need to adjust the τ's to take into account the midrank values. The main modification involves the expression in

(4.14), where here we have

$$\mathrm{E}[V_{\boldsymbol{n}}] = \sum_{\substack{1\leqslant j_1,\ldots,j_d\leqslant m \\ \text{distinct}}} \mathrm{E}[(\mathrm{rank}(W_{j_1})-\nu)^{n_1}\cdots(\mathrm{rank}(W_{j_d})-\nu)^{n_d}](r_{j_1}-\nu)^{n_1}\cdots(r_{j_d}-\nu)^{n_d}. \quad (13.37)$$

Then we can derive, as from (4.7) and (4.19), the formula

$$\mathrm{E}[(d^A_{\mathrm{Spear}}(\boldsymbol{W},\boldsymbol{Z}) - \mathrm{E}[d^A_{\mathrm{Spear}}(\boldsymbol{W},\boldsymbol{Z})])^n] = (-2)^n \sum_{\boldsymbol{n}\in\mathcal{IP}_{n,m}} \zeta_{\boldsymbol{n}} \frac{\tau_{\boldsymbol{r}}(\boldsymbol{n})\tau_{\boldsymbol{s}}(\boldsymbol{n})}{(m)_d}, \quad (13.38)$$

where $\mathcal{IP}_{n,m}$ is the set of integer partitions of n with at most m components, $\zeta_{\boldsymbol{n}}$ is the number of set partitions corresponding to $\boldsymbol{n}$, d is the number of elements in the vector $\boldsymbol{n}$, and

$$\tau_{\boldsymbol{r}}(\boldsymbol{n}) = \sum_{\substack{1\leqslant j_1,\ldots,j_d\leqslant m \\ \text{distinct}}} \cdots \sum (r_{j_1}-\nu)^{n_1}\cdots(r_{j_d}-\nu)^{n_d}. \quad (13.39)$$

The iterative formula in (4.27) also holds here for $\tau_{\boldsymbol{r}}$, but with η also depending on $\boldsymbol{r}$, i.e., instead of (4.24) we have

$$\eta_{\boldsymbol{r},k} = \sum_{i=1}^{m} (r_i-\nu)^k. \quad (13.40)$$

13.3 Asymptotic distributions

The asymptotics for Spearman's distance with ties depend on the most common tied value. Specifically, denote $m_x = \max\{m_a\}$ and $n_z = \max\{n_a\}$. If these two are small relative to m, then asymptotic normality holds. If $q \equiv m - m_x$ is fixed, then we do not have normality, but may be asymptotically approaching a sum of q iid variables. The main theorem is next. The proof of this and the other results in this section are found in Section 13.5.

Theorem 13.1. *If*

$$\frac{(m-m_x)(m-n_z)}{m} \to \infty, \quad (13.41)$$

then

$$\frac{d^A_{Spear}(\boldsymbol{W},\boldsymbol{Z}) - \mu_{Spear}(m)}{\sqrt{\mathrm{Var}[d^A_{Spear}(\boldsymbol{W},\boldsymbol{Z})]}} \longrightarrow^{\mathcal{D}} N(0,1). \quad (13.42)$$

Theorem 3.5 of Alvo & Yu (2014) has a more general result that allows missing data as well as ties. Note that condition (13.41) is weaker than, but implied by,

$$\limsup \frac{m_x}{m} < 1 \quad \text{and} \quad \limsup \frac{n_z}{m} < 1, \quad (13.43)$$

which means as long as we do not have that almost all elements are equal in the vectors, the asymptotic normality holds.

The condition is violated if $m - m_x$ (or $m - n_z$) is bounded, so consider with the case that $m - m_x$ is fixed. Let $\boldsymbol{R} = \text{rank}(\boldsymbol{W})$ and $\boldsymbol{S} = \text{rank}(\boldsymbol{Z})$ be the midrank vectors. In order to describe the asymptotic distribution of Spearman's distance, we need the asymptotic distribution of the S_i/m's. Let F_m be their distribution function (the S_i are identically distributed), so that

$$F_m(x) = P\left[\frac{S_i}{m} \leqslant x\right], \; x \in [0,1]. \tag{13.44}$$

Theorem 13.2. *Suppose that for large enough* m, $m_{<x} = q_1$ *and* $m_{>x} = q_2$ *are fixed, with* $q = q_1 + q_2$, *and that there exists a distribution function* F *on [0,1] such that* $F_m(x) \to F(x)$ *on points of continuity. Then*

$$\frac{1}{m^2}\left(d^A_{Spear}(\boldsymbol{W},\boldsymbol{Z}) - \mu_{Spear}(m)\right) + \frac{q}{2} \longrightarrow^{\mathcal{D}} V_1 + \ldots + V_{q_1} + (1 - V_{q_1+1}) + \cdots + (1 - V_q), \tag{13.45}$$

where $V_1, \ldots, V_q$ *are iid with distribution* F.

If there are no, or few, ties in the $\boldsymbol{Z}$, then the V_i are uniform, as in the next result.

Corollary 13.3. *Suppose that for large enough* m, $m - m_x = q$ *is fixed, and* $n_z/m \to 0$. *Then*

$$\frac{1}{m^2}\left(d^A_{Spear}(\boldsymbol{W},\boldsymbol{Z}) - \mu_{Spear}(m)\right) + \frac{q}{2} \longrightarrow^{\mathcal{D}} U_1 + \ldots + U_q, \tag{13.46}$$

where $U_1, \ldots, U_q$ *are iid Uniform(0,1).*

Thus the asymptotic distribution in (13.67) is called the **Irwin-Hall distribution** with parameter q. See Wikipedia contributors (2019a) and Marengo, Farnsworth, & Stefanic (2017) for overviews. The latter give a nice geometric proof of the density, which is

$$f_{IH}(t\,;\,q) = \frac{1}{(q-1)!}\sum_{k=0}^{\lfloor t \rfloor}(-1)^k\binom{q}{k}(t-k)^{q-1}, \;\; 0 < t < q. \tag{13.47}$$

The density is symmetric about $q/2$, and if $t > q/2$, it is more efficient and more stable numerically to calculate $f_{IH}(q - t; q)$.

If $n_z/m \not\to 0$, it is unlikely one would know what the limiting distribution function is. If m is large and q is small, then a reasonable approach is to use (13.45) but approximate the distribution of the sum by a convolution of the empirical F_m. If the $\boldsymbol{Z}$ has most values tied at one value, so that $n_z/m \to 1$, then (13.45) holds, but the V_i are identically 1/2. In fact,

$$\frac{(m - m_x)(m - n_z)}{m} \to 0 \Longrightarrow P\left[\frac{1}{m^2}\left(d^A_{Spear}(\boldsymbol{W},\boldsymbol{Z}) - \mu_{Spear}(m)\right) = c_m\right] \longrightarrow 1 \tag{13.48}$$

for some constant $c_m \to 0$. See the end of Section 13.5.2. Thus the statistic is not likely to be very useful with just one observation $(\boldsymbol{w}, \boldsymbol{z})$.

The above cover most situations likely to arise, but do not address cases where $m - m_x \to \infty$, $m - n_z \to \infty$, and $(m - m_x)(m - n_z)/m$ is bounded away from 0 and infinity. It is also open whether the condition in (13.41) is necessary as well as sufficient for the asymptotic normality in (13.42). It is in the case that $n_z/m \to 0$, or if $K = L = 2$, as a consequence of Theorem 2.2 of Kou & Ying (1996). Thus I suspect the condition is indeed necessary for normality.

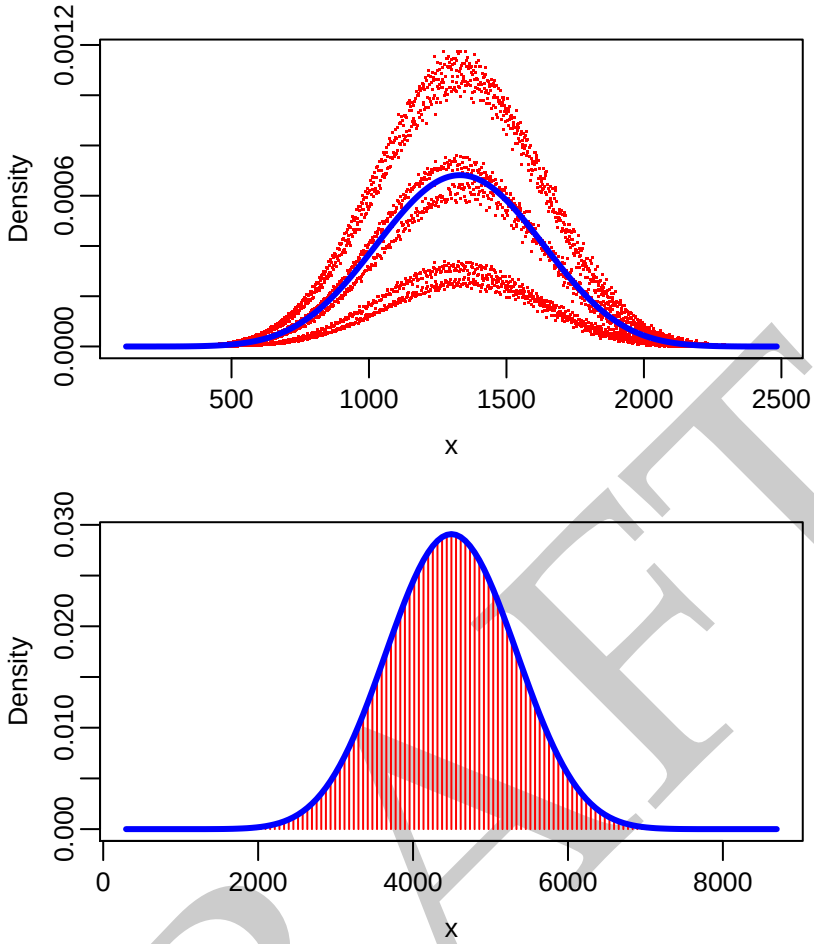

Figure 13.1: Comparing the exact density for Spearman's distance with ties to the Edgeworth approximation with $L = 4$ for two examples with $m = 20$. The top graph has $\boldsymbol{n} = (8, 3, 3, 3, 2, 1)$ and $\boldsymbol{m} = (7, 4, 3, 3, 3)$, where the dots represent the heights of the density, and the solid line is the Edgeworth approximation. The bottom graph has $\boldsymbol{n} = (5, 5, 5, 5, 5, 5)$ and $\boldsymbol{n} = (6, 6, 6, 6, 6)$, where now the exact density is represented by the vertical bars.

13.4 Edgeworth and simulation approximations

The previous section guarantees asymptotic normality for most situations, but for small to medium m, the actual fidelity of the normal (or Edgeworth modifications) to the true distributions is often not very good. It appears that the larger the number of possible tables, $\#\mathcal{T}$, the better the approximation. If that number is not too large (order of 1.5×10^6, say), the exact distribution can be found in under a second or so. Larger numbers of tables may still not lead to very good normal approximations, in which case simulation is preferable. Here we try to present some guidelines of which approach to use when.

Generally, the more unequal the entries in the patterns of ties are, the less normal the

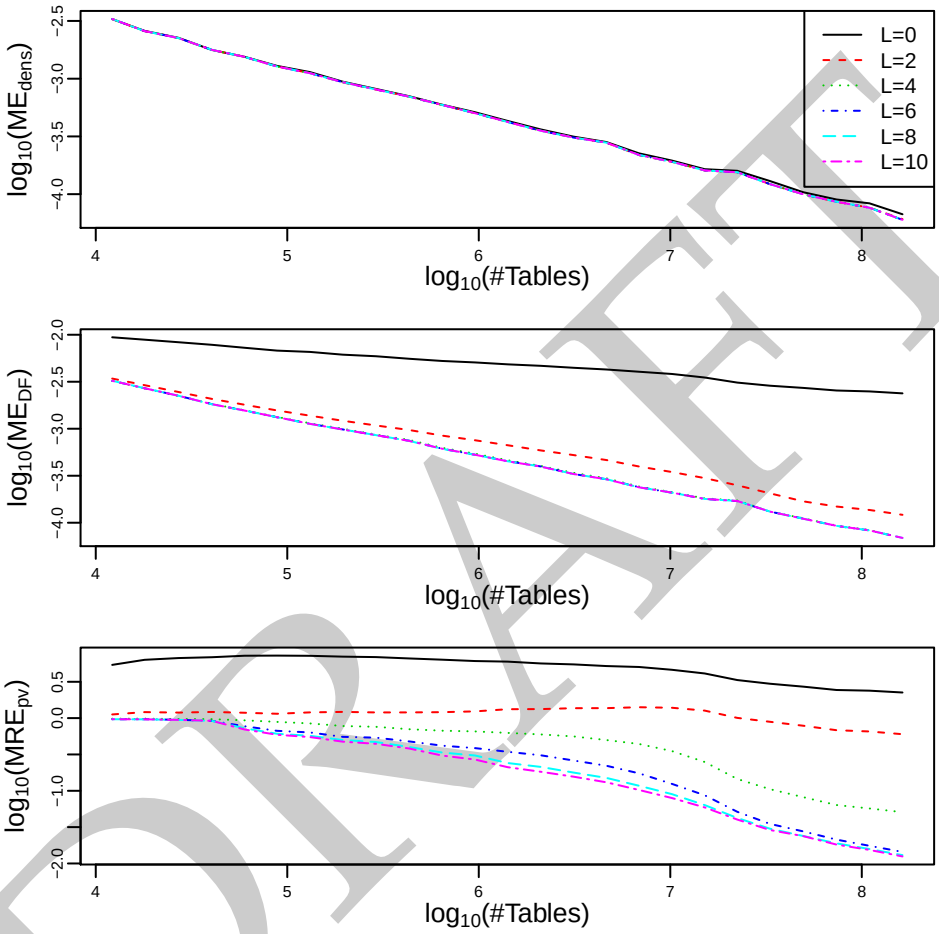

Figure 13.2: Comparing the median $\log_{10}$(errors) of the Edgeworth approximation to the distribution of the Spearman distance with ties for L=0, 2, 4, 6, 8, and 10. The horizontal axis is the $\log_{10}(\#\mathcal{T})$. The panel graphs the maximum error in the density, the middle graphs the maximum error in the distribution functions, and the bottom graphs the maximum relative error in the p-values. See (4.54) and (4.54).

density looks. Figure 13.1 shows the density and Edgeworth approximation for two examples with $m = 20$. In the first, the patterns of ties are somewhat uneven, $\boldsymbol{n} = (8,3,3,3,2,1)$ and $\boldsymbol{m} = (7,4,3,3,3)$. We see that the approximation follows the general shape well, but the density is very spiky, and in fact looks somewhat like a mixture of many normalish curves. This approximation tends to be good for the distribution function and p-value, but has trouble with the density because of that spikiness. We saw this behavior to a lesser extent in Figure 4.4 for the case with no ties. The second example has equal values in the patterns, with $\boldsymbol{n} = (5,5,5,5,5,5)$ and $\boldsymbol{n} = (6,6,6,6,6)$. Now the density is very regular, and the Edgeworth approximation matches it almost exactly.

Another consideration arises in using the Edgeworth expansion. For the case with no ties, we could pre-calculate the moments and cumulants based on polynomials m, as seen in Section 4.5, hence the calculations are very fast. Here, since each set of tied rankings has a different set of moments, we have to calculate the quantities as in (13.38) and (13.39) each time. In particular, the $\tau_r(\boldsymbol{n})$ can be onerous to find for larger n and m. Thus in some cases calculation time could be an issue.

For $m \leqslant 13$, the splitting algorithm is reasonably quick. For each m between 13 and 50, we randomly chose at about 1000 pairs $(\boldsymbol{m}, \boldsymbol{n})$, then found the exact number of corresponding tables and exact distribution, up to 2×10^8; the time the algorithm expended; the Edgeworth approximation to the distribution for $L = 0, \ldots, 10$; an estimate of the distribution using 500,000 simulations; and the errors of the two approximations.

First, we compare the Edgeworth approximations for the different values of L, using the maximum error in the density, the maximum error in the distribution functions, and the maximum relative error in the p-values. See (4.54) and (4.54). In general, the improvement going from an even L to $L+1$ is negligible, so in Figure 13.2 we show just the even L, though even in that case many of the lines are indistiguishable. For each graph, the data was grouped into 25 categories depending on $\log_{10}(\#\mathcal{T})$, and we then found the median $\log_{10}$(error)'s for each type of error by group. Table 13.1 summarizes the results by finding the median error for $7 < \log_{10}(\#\mathcal{T}) < 8$ for each L, as well as the jump (usually down) for the errors from $L-1$ to L.

If we look at the error in the density, then there is a small improvement from $L = 0$ to 2, then little improvement for larger L. For the distribution function, there is a large improvement from $L = 0$ to 2, then substantial improvement from 2 to 4. After that, there is little change. For the relative error in the p-value, there is substantial improvement from 0 to 2 to 4 to 6, then small improvements to 8 and 10. Thus $L = 4$ or 6 seem to be reasonable choices, or 8 if the relative p-value is of most importance.

Figure 13.3 compares the errors using either the Edgeworth or simulated estimates, on their 5%, 50% , and 95% quantiles. Overall, for the density and distribution function, the errors are mostly in the 10^{-3} to 10^{-4} area. In each case, the simulations are fairly constant as the number of tables increases, while Edgeworth improves. Simulations tend to give better approximations for the density, until we get to 10^8 tables, when Edgeworth is about the same. For the distribution function, Edgeworth is better, especially if the number of tables is over 10^6 (and under that we can find the exact distribution). For the relative p-value, they are similar for smaller numbers of tables, but Edgeworth becomes much better after 10^7 tables. It would

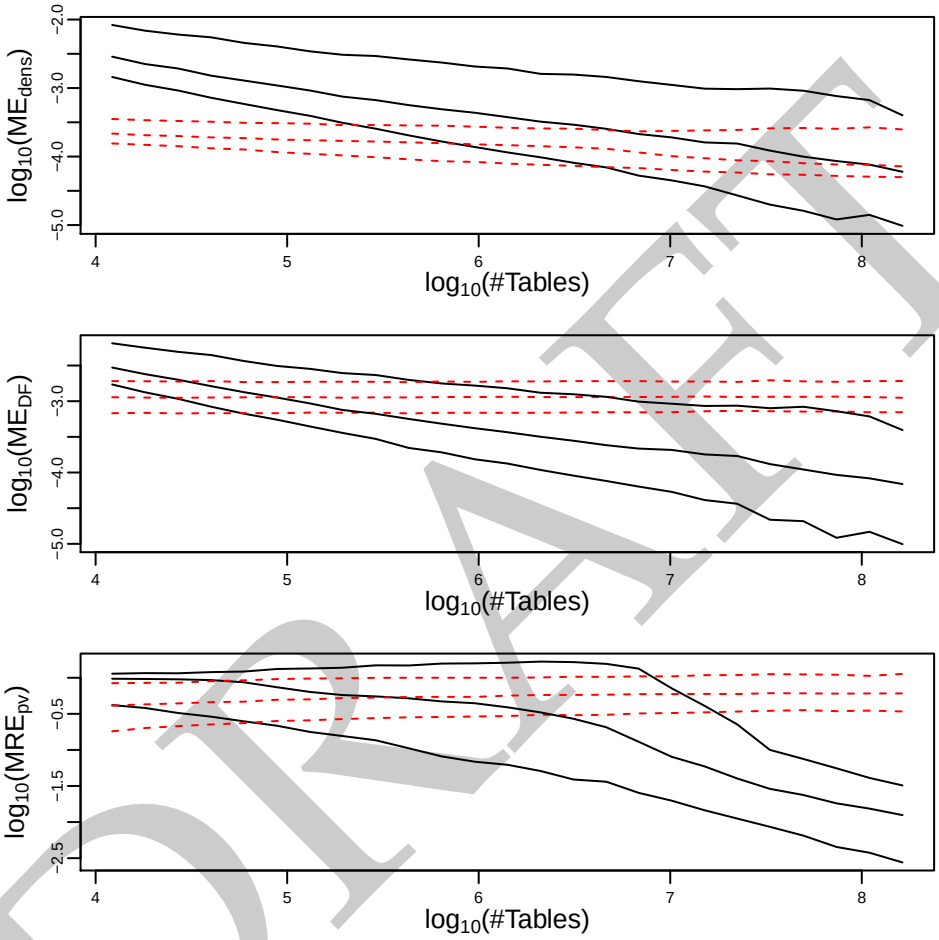

Figure 13.3: Comparing the $\log_{10}$(errors) of the Edgeworth approximation with $L = 10$ to estimation with 500,000 simulations. The horizontal axis is the $\log_{10}(\#\mathcal{T})$. The three panels graph the errors in the density, distribution function, and relative p-values, respectively. In each plot, the solid lines represent the $5^{\text{th}}, 50^{\text{th}}$, and 95^{th} percentiles for the Edgeworth approximations; the dotted lines represent the same percentiles for the simulations.

	Density		Distribution function		P-value	
L	$\log_{10}$(error)	Jump	$\log_{10}$(error)	Jump	$\log_{10}$(error)	Jump
0	−3.8351		−2.5103		0.5268	
1	−3.8354	−0.0003	−2.5414	−0.0311	0.4867	−0.0400
2	−3.8507	−0.0153	−3.5932	−1.0518	0.0058	−0.4809
3	−3.8507	0.0000	−3.6033	−0.0101	−0.0797	−0.0855
4	−3.8507	0.0000	−3.7956	−0.1923	−0.8327	−0.7530
5	−3.8507	0.0000	−3.7978	−0.0023	−0.8510	−0.0183
6	−3.8507	0.0000	−3.8018	−0.0040	−1.2980	−0.4470
7	−3.8507	0.0000	−3.8019	−0.0001	−1.2867	0.0113
8	−3.8507	0.0000	−3.8022	−0.0003	−1.3713	−0.0846
9	−3.8507	0.0000	−3.8023	−0.0001	−1.3715	−0.0002
10	−3.8507	0.0000	−3.8023	0.0000	−1.3909	−0.0193
Simulations	−4.0463		−2.9373		−0.2222	

Table 13.1: The median $\log_{10}$(errors) when $10^7 < \#\mathcal{T} < 10^8$ for the Edgeworth approximation (L=0, ..., 10), and for simulations with 500,000 replications. The "Jump" columns show the difference log errors between L and L + 1.

be reasonable to use simulations if $\#\mathcal{T} \leqslant 10^7$, and Edgeworth otherwise. See also the last line in Table 13.1. Of course, the simulations can be improved by increasing their number.

13.5 Proofs of asymptotic results

13.5.1 Asymptotic normality

To prove Theorem 13.1, we use Theorem 4 of Hoeffding (1951). Let $\boldsymbol{r} = \text{rank}(\boldsymbol{w})$ and $\boldsymbol{s} = \text{rank}(\boldsymbol{z})$ be the midrank vectors. Then Hoeffding shows that the asymptotic normality in (13.42) holds if

$$\frac{h(\boldsymbol{r})h(\boldsymbol{s})}{m} \to \infty, \tag{13.49}$$

where

$$h(\boldsymbol{r}) = \frac{\sum(r_i - \nu)^2}{\max\{(r_i - \nu)^2\}}, \quad \nu = \frac{m+1}{2}. \tag{13.50}$$

We will show that $h(\boldsymbol{r})$ and $m - m_x$ are asymptotically equivalent, so that (13.41) is equivalent to (13.49).

Consider the maximum term. Without loss of generality, we can take $\boldsymbol{r} = \boldsymbol{r}^{(o)}$ of (13.4), so that the elements are in nondecreasing order. Since ν is the mean of the elements of $\boldsymbol{r}$, the maximum of $(r_i - \nu)^2$ must occur at one of the extremes, $i = 1$ or m. Now $r_1 = (m_1 + 1)/2$

and $r_m = (m - m_K) + (m_K + 1)/2$, so that

$$\max\{(r_i - \nu)^2\} = \frac{1}{4}\max\{(m - m_1)^2, (m - m_K)^2\} = \frac{1}{4}(m - \min\{m_1, m_K\})^2. \tag{13.51}$$

The smallest m_a can be is 1, and the largest the minimum of m_1 and m_L can be is when they are both $m/2$ ($(m \pm 1)/2$ if m is odd), hence

$$\frac{m^2}{16} \leqslant \max\{(r_i - \nu)^2\} < \frac{m^2}{4}. \tag{13.52}$$

Using the conditional expectation representation in (12.25) for r_i, we have

$$\mathrm{Var}[\mathrm{E}[X_i \,|\, \boldsymbol{W}]] = \frac{\sum (r_i - \nu)^2}{m}. \tag{13.53}$$

Also, since the variance of a discrete uniform on $\{c+1, \ldots, c+n\}$ is $(n^2 - 1)/12$, from (12.17),

$$\mathrm{Var}[X_i] = \frac{m^2 - 1}{12} \quad \text{and} \quad \mathrm{Var}[X_i \,|\, \boldsymbol{W} = \boldsymbol{w}] = \frac{m_{w_i}^2 - 1}{12} \tag{13.54}$$

since $m_a = u_a^{(\boldsymbol{w})} - l_a^{(\boldsymbol{w})} + 1$. Using the decomposition of variance by conditioning, write

$$\mathrm{Var}[X_i] = \mathrm{Var}[\mathrm{E}[X_i \,|\, \boldsymbol{W}]] + \mathrm{E}[\mathrm{Var}[X_i \,|\, \boldsymbol{W}]], \tag{13.55}$$

so that

$$\frac{\sum (r_i - \nu)^2}{m} = \frac{m^2 - 1}{12} - \sum_{a=1}^{K} \frac{m_a}{m} \frac{m_a^2 - 1}{12}, \tag{13.56}$$

hence

$$\sum_{i=1}^{m} (r_i - \nu)^2 = \frac{m^3 - \sum m_a^3}{12}. \tag{13.57}$$

Now

$$m_x^3 \leqslant \sum m_a^3 \leqslant \sum m_a m_x^2 = m m_x^2, \tag{13.58}$$

so that

$$\begin{aligned} &m^3 - \sum m_a^3 \leqslant m^3 - m_x^3 = (m - m_x)(m^2 + m m_x + m_x^2) \leqslant 3m^2(m - m_x), \text{ and} \\ &m^3 - \sum m_a^3 \geqslant m^3 - m m_x^2 = m(m - m_x)(m + m_x) \geqslant m^2(m - m_x). \end{aligned} \tag{13.59}$$

Together with (13.52) and (13.57), by (13.50), we have

$$\frac{(m - m_x)}{3} \leqslant h(\boldsymbol{r}) \leqslant 4(m - m_x), \tag{13.60}$$

and a similar formula for $\boldsymbol{s}$. Thus condition (13.41) holds if (and only if) (13.49) does, which verifies the theorem.

13.5.2 Asymptotics with $m_x \approx m$

Here we have $\boldsymbol{W}$ with pattern of ties such that $q = m - m_x$ is fixed. The $\boldsymbol{Y}$ may or may not have ties, but if $\boldsymbol{Z}$ has pattern $(n_1, \ldots, n_L)$. Let $\boldsymbol{R} = \text{rank}(\boldsymbol{W})$ and $\boldsymbol{S} = \text{rank}(\boldsymbol{Z})$ be the midrank vectors. As in (13.44), let F_m be the distribution function of the S_i/m, which are identically distributed. The next lemma shows that if F_m has a limit, any finite set of S_i/m's are asymptotically independent as well.

Lemma 13.4. *Suppose there exists a distribution function F on [0,1] such that $F_m(x) \to F(x)$ on points of continuity of F. Then if q is fixed, as $m \to \infty$,*

$$\frac{1}{m}(S_1, \ldots, S_q) \longrightarrow^{\mathcal{D}} (V_1, \ldots, V_q), \tag{13.61}$$

where $V_1, \ldots, V_q$ are iid with distribution function F.

Proof. Similar to (13.4), let $u_1 < u_2 < \cdots < u_L$ be the possible values of the S_i, so that the ordered values of any realization of $\boldsymbol{S}$ is

$$\boldsymbol{s}^{(o)} = (u_1, \ldots, u_1, u_2, \ldots, u_2, \ldots, u_L, \ldots, u_L), \quad u_l = n_{<l} + \frac{n_l + 1}{2}, \tag{13.62}$$

where there are n_l of the u_l's in the $\boldsymbol{s}^{(o)}$. See (12.14). The distribution function F_m can be thought of as the empirical distribution function of the $s_i^{(o)}/m$'s. Thus we have

$$F_m(x) = \#\{i \mid s_i^{(o)} \leqslant mx\} = \frac{n(mx)}{m}, \quad \text{where} \quad n(w) = \sum_{k \mid u_k \leqslant w} n_k. \tag{13.63}$$

This distribution is a discrete one, where the pmf is $f_m(u_l/m) = n_l/m$. Take $0 \leqslant x_1 \leqslant x_2 \leqslant \cdots \leqslant x_q \leqslant 1$. Then

$$P\left[\frac{S_1}{m} \leqslant x_1, \cdots, \frac{S_q}{m} \leqslant x_q\right] = \frac{n(mx_1)}{m} \frac{n(mx_2) - 1}{m - 1} \cdots \frac{n(mx_q) - q + 1}{m - q + 1}. \tag{13.64}$$

To see (13.64), note that there are $n(mx_1)$ ways to choose one of the $s_i^{(o)}$'s such that $ms_i^{(o)} \leqslant x_1$. Since $x_1 \leqslant x_2$, there are only $n(mx_2) - 1$ of the $s_i^{(o)}$'s left such that $ms_i^{(o)} \leqslant x_2$, out of $m - 1$ total. We continue until the q^{th} choice. Thus from (13.63), since q is fixed,

$$P\left[\frac{S_1}{m} \leqslant x_1, \cdots, \frac{S_q}{m} \leqslant x_q\right] = \frac{m^q}{(m)_q} F_m(x_1) \cdots F_m(x_q) + O\left(\frac{1}{m}\right). \tag{13.65}$$

Now let $m \to \infty$. If the x_i's are continuity points of F, since we have assumed that $F_m(x_i) \to F(x_i)$, we have

$$P\left[\frac{S_1}{m} \leqslant x_1, \cdots, \frac{S_q}{m} \leqslant x_q\right] \longrightarrow F(x_1) \cdots F(x_q). \tag{13.66}$$

Finally, since the distribution of $\boldsymbol{S}$ is permutation-invariant, (13.66) holds for any order of the x_i's, hence (13.61) follows from (13.66). □

We apply the lemma to prove the main asymptotic result for q fixed.

Proof of Theorem 13.2. Fix $\boldsymbol{W} = \boldsymbol{w}$ with the w_i's nondecreasing, and set $\text{rank}(\boldsymbol{w}) = \boldsymbol{r} = (t_1, \ldots, t_1, \ldots, t_K, \ldots, t_K)$. Thus there are m_k elements equal to t_k, and $t_k = m_{<k} + (m_k + 1)/2, k = 1, \ldots, K$. Also, let $\boldsymbol{S} = \text{rank}(\boldsymbol{Z})$. From (13.1) and (13.3), we can obtain

$$d^A_{\text{Spear}}(\boldsymbol{w}, \boldsymbol{z}) - \mu_{\text{Spear}}(m) = -2 \sum_{i=1}^{m} (r_i - \nu) s_i. \tag{13.67}$$

Writing the r_i in terms of t_k, and singling out the $k = x$ term, the crossproduct term can be written

$$\begin{aligned}
\sum_{i=1}^{m} (r_i - \nu) s_i &= \sum_{k=1}^{K} (t_k - \nu) \sum_{i=m_{<k}+1}^{m_{\leqslant k}} s_i \\
&= \sum_{k \neq x} (t_k - \nu) \sum_{i=m_{<k}+1}^{m_{\leqslant k}} s_i + (t_x - \nu) \left(\frac{m(m+1)}{2} - \sum_{i=1}^{m_{<x}} s_i - \sum_{i=m_{\leqslant x}+1}^{m} s_i \right) \\
&= \sum_{k \neq x} (t_k - t_x) \sum_{i=m_{<k}+1}^{m_{\leqslant k}} s_i + (t_x - \nu) \frac{m(m+1)}{2}.
\end{aligned} \tag{13.68}$$

Now

$$t_k - t_x = m_{<k} - m_{<x} + \frac{m_k - m_x}{2}. \tag{13.69}$$

If $k < x$, all the terms in (13.69) are bounded except for m_x, which has coefficient $-1/2$. If $k > x$, then $m_{<k}$ equals m_x plus some bounded terms, hence m_x has coefficent $1/2$ in the difference. Thus as $m \to \infty$, $m_x/m \to 1$, and

$$\frac{t_k - t_x}{m} = \begin{cases} -\frac{1}{2} + O\left(\frac{1}{m}\right) & \text{if } k < x, \\ \frac{1}{2} + O\left(\frac{1}{m}\right) & \text{if } k > x. \end{cases} \tag{13.70}$$

Also, $t_x = q_1 + (m - q + 1)/2$, so that $t_x - \nu = (q_1 - q_2)/2$. (Recall that $m_{<x} = q_1$ and $m_{>x} = q_2$ are fixed.) Thus

$$\frac{1}{m^2} \left(\sum_{i=1}^{m} (r_i - \nu) S_i \right) = -\frac{1}{2} \sum_{i=1}^{q_1} \frac{S_i}{m} + \frac{1}{2} \sum_{i=m-q_2+1}^{m} \frac{S_i}{m} + \frac{q_1 - q_2}{4} + O\left(\frac{1}{m}\right) \tag{13.71}$$

Now let $m \to \infty$. Then by (13.61),

$$\frac{1}{m}(S_1, \ldots, S_{q_1}, S_{m-q_1+1}, \ldots, S_m) \longrightarrow^{\mathcal{D}} (V_1, \ldots, V_q), \tag{13.72}$$

where the $V_1, \ldots, V_q$ are iid F. Rewinding through (13.71) and(13.67), we have

$$\frac{1}{m^2} \left(d^A_{\text{Spear}}(\boldsymbol{w}, \boldsymbol{Z}) - \mu_{\text{Spear}}(m) \right) \longrightarrow^{\mathcal{D}} V_1 + \cdots + V_{q_1} - (V_{q_1+1} + \cdots + V_q) + \frac{q_2 - q_1}{2}. \tag{13.73}$$

Now add $q/2$ to both sides to obtain (13.45). □

Next we specialize to the case where there are not many ties in the $\boldsymbol{Z}$.

Proof of Corollary 13.3. Now we have that $n_z/m \to 0$ holds as well as q being fixed. For given $x \in [0,1]$, let l be the index such that $t_l \leqslant mx < t_{l+1}$, where $t_0 = 0$ and $t_{L+1} = m+1$. Then $n(mx) = n_{\leqslant l}$. From (13.63), we can write

$$n_{\leqslant l} + \frac{-n_l + 1}{2} \leqslant mx < n_{\leqslant l} + \frac{n_{l+1} + 1}{2}, \tag{13.74}$$

(where $n_{L+1} = 0$), hence since $F_m(x) = n_{\leqslant l}/m$,

$$\frac{-n_l + 1}{2m} \leqslant x - F_m(x) < \frac{n_{l+1} + 1}{2m}. \tag{13.75}$$

That is, $|F_m(x) - x| \leqslant (n_z + 1)/(2m) \to 0$, so that $F(x) = x$, the Uniform(0,1) distribution function. Thus in Theorem 13.2, the V_i are Uniform(0,1), which means so are the $1 - V_i$, and no matter what q_1 and q_2 are, as long as they sum to q, the asymptotic distribution in (13.45) is the sum of q iid Uniform(0,1)'s. Thus the corollary follows. □

Finally, suppose both $q = m - m_x$ and $p = m - n_z$ are small relative to m, in that $pq/m \to 0$. We wish to show (13.48), that the distribution of Spearman's distance approaches placing all its mass at one point. Let $\boldsymbol{r}$ and $\boldsymbol{s}$ be the midrank vectors, so that $\boldsymbol{r}$ has m_k elements equal to t_k, and $\boldsymbol{s}$ has n_l values of u_l. Thus t_x and u_z are the most common values in $\boldsymbol{r}$ and $\boldsymbol{s}$, respectively. The most common value of the crossproduct $\sum r_i s_i$ occurs if the pair of vectors is in the following set:

$$\mathcal{A}_m = \{(\boldsymbol{r}, \boldsymbol{s}) \,|\, \text{for each } i,\ r_i = t_x \text{ or } s_i = u_z \text{ (or both)}\}. \tag{13.76}$$

Then if $(\boldsymbol{r}, \boldsymbol{s}) \in \mathcal{A}_m$,

$$\begin{aligned}
(\boldsymbol{r}, \boldsymbol{s}) \in \mathcal{A}_m &\Longrightarrow \sum_{i=1}^{m} (r_i - \nu)(s_i - \nu) \\
&= (t_x - \nu) \sum_{i \,|\, s_i \neq u_z} (s_i - \nu) + (u_z - \nu) \sum_{i \,|\, r_i \neq t_x} (r_i - \nu) + (t_x - \nu)(u_z - \nu)\, \#\{i \,|\, r_i = t_x \ \&\ s_i = u_z\} \\
&= -(t_x - \nu)(u_z - v) n_z - (u_z - \nu)(t_x - \nu) m_x + (t_x - \nu)(u_z - \nu)(m_x + n_z - m) \\
&= -m(t_x - \nu)(u_z - \nu).
\end{aligned} \tag{13.77}$$

The third line follows from the second by noting that $\sum_i (r_i - \nu) = 0 = \sum_k m_k (t_k - \nu)$ (and similarly for $\boldsymbol{s}$), and the final set has $m - (m - m_x) - (m - n_z)$ values.

Now fix $\boldsymbol{r}$. Then if $p + q < m$, which happens for large enough m since $pq/m \to 0$ implies $(p+q)/m \to 0$,

$$P[(\boldsymbol{r}, \boldsymbol{Z}) \in \mathcal{A}_m] = \frac{(m_x)_{m-n_z}}{(m)_{m-n_z}} = \frac{(m-p)!(m-q)!}{(m-p-q)!m!}. \tag{13.78}$$

To see (13.78), note that we need to match each $r_i \neq u_z$ with $z_i = t_x$. Thus we have to count the number of ways to choose $m - n_z$ items, without replacement, from a pool of m_x. Using Stirling's approximation,

$$\log(k!) = (k + \tfrac{1}{2}) \log(k) - k + O\left(\frac{1}{k}\right), \tag{13.79}$$

we can show that

$$\log\left(\frac{(m-p)!(m-q)!}{(m-p-q)!m!}\right) = (m-p)\log\left(1-\frac{p}{m}\right) + (m-q)\log\left(1-\frac{q}{m}\right) - (m-p-q)\log\left(1-\frac{p+q}{m}\right) + \frac{1}{2}\log\left(\frac{(m-p)(m-q)}{m(m-p-q)}\right) + O\left(\frac{1}{m}\right). \quad (13.80)$$

The big O term holds since $(p+q)/m \to 0$ implies that $m-p, m-q$, and $m-p-q$ are asymptotically equivalent to m. Those convergences also show that the second to last term in (13.81) approaches 0. We expand $\log(1-x)$ for the other three terms, so that their sum equals

$$\begin{aligned} &-(m-p)\sum_{k=1}^{\infty}\frac{1}{k}\left(\frac{p}{m}\right)^k - (m-q)\sum_{k=1}^{\infty}\frac{1}{k}\left(\frac{q}{m}\right)^k + (m-p-q)\sum_{k=1}^{\infty}\frac{1}{k}\left(\frac{p+q}{m}\right)^k \\ &\qquad = \sum_{k=1}^{\infty}\frac{1}{k}\frac{(p+q)^k - p^k - q^k}{m^{k-1}} - \sum_{k=1}^{\infty}\frac{1}{k}\frac{(p+q)^{k+1} - p^{k+1} - q^{k+1}}{m^k} \\ &\qquad = -\sum_{k=1}^{\infty}\frac{1}{k(k+1)}\frac{(p+q)^{k+1} - p^{k+1} - q^{k+1}}{m^k}. \end{aligned} \quad (13.81)$$

Using the binomial theorem, we have

$$\begin{aligned} (p+q)^{k+1} - p^{k+1} - q^{k+1} = \sum_{l=1}^{k}\binom{k+1}{l}p^l q^{k+1-l} &= pq\sum_{l=1}^{k}\binom{k}{l}p^{l-1}q^{k-l} \\ &\leqslant pq(p+q)^{k-1}2^{k+1}, \end{aligned} \quad (13.82)$$

where in the last step, we bound p and q by $p+q$, and add the $l=0$ and $l=k+1$ terms to the sum of the binomial coefficients. Thus

$$0 \geqslant -\sum_{k=1}^{\infty}\frac{1}{k(k+1)}\frac{(p+q)^{k+1} - p^{k+1} - q^{k+1}}{m^k} \geqslant -4\frac{pq}{m}\sum_{k=1}^{\infty}\frac{1}{k(k+1)}\left(\frac{2(p+q)}{m}\right)^{k-1}. \quad (13.83)$$

For any $0 < x < 1$, eventually $2(p+q)/m \leqslant x$, in which case the final summation in (13.83) converges. Now the assumption that $pq/m \to 0$ implies that the expression in (13.81) goes to zero, hence by (13.81) and (13.78), $P[(\boldsymbol{r}, \boldsymbol{Z}) \in \mathcal{A}_m] \to 1$, hence via (13.77),

$$P\left[\frac{1}{m^2}\left(d_{\text{Spear}}(\boldsymbol{w}, \boldsymbol{Z}) - \mu_{\text{Spear}}(m)\right) = c_m\right] \longrightarrow 1, \quad (13.84)$$

where

$$c_m = 2\frac{(t_x - \nu)(u_z - \nu)}{m} = \frac{1}{m}\left(m_{<x} - \frac{p}{2}\right)\left(n_{<z} - \frac{q}{2}\right). \quad (13.85)$$

Since $0 \leqslant m_{<x} \leqslant p$ and $0 \leqslant n_{<z} \leqslant q$, $|c_m| \leqslant pq/(4m)$, which also goes to zero by our assumption. Thus (13.48) follows.

13.5.3 Proof of Irwin-Hall density

Here we use the approach given in Marengo et al. (2017) to verify the Irwin-Hall density in (14.24). Let $T = U_1 + \cdots + U_q$, where the U_i are iid Uniform(0,1). The space of T is $(0, q)$. Since $(U_1, \ldots, U_q)$ is distributed over the unit q-dimensional hypercube, the distribution function can be expressed as

$$F_{IH}(t) = \text{Volume}\{(u_1, \ldots, u_q) \,|\, 0 < u_i < 1, i = 1, \ldots, q, \text{ and } u_1 + \cdots + u_q \leqslant t\}. \tag{13.86}$$

If we remove the constraint that the $u_i < 1$ in the set in (13.86), the resulting set has volume $t^q/q!$. To find the distribution function F_{IH}, we start with the volume of this expanded set, then subtract off the volume of the parts where one or more $u_i > 1$. That is,

$$F_{IH}(t) = \text{Volume}(\mathcal{A}(t)) - \text{Volume}(\mathcal{B}_1(s) \cup \mathcal{B}_2(t) \cup \cdots \cup \mathcal{B}_q(t)), \tag{13.87}$$

where

$$\begin{aligned} \mathcal{A}(t) &= \{(u_1, \ldots, u_q) \,|\, 0 < u_i, i = 1, \ldots, q, \text{ and } u_1 + \cdots + u_q \leqslant t\} \text{ and} \\ \mathcal{B}_j(t) &= \{(u_1, \ldots, u_q) \in \mathcal{A}(t) \,|\, u_j > 1\}, \quad j = 1, \ldots, q. \end{aligned} \tag{13.88}$$

We find the volume of the union in (13.87) using the union-intersection principle. By symmetry, the volume of an intersection of j of the $\mathcal{B}_i(t)$'s depends just on the number j, and since there are $\binom{q}{j}$ possible sets of j of them, we have

$$\text{Volume}(\mathcal{B}_1(t) \cup \mathcal{B}_2(t) \cup \cdots \cup \mathcal{B}_q(t)) = \sum_{j=1}^{q} (-1)^{j+1} \binom{q}{j} \text{Volume}(\mathcal{B}_1(t) \cap \cdots \cap \mathcal{B}_j(t)). \tag{13.89}$$

Now take $t \in (0, q)$ and let k be the integer such that $k \leqslant t < k + 1$. Note that at most k of the u_i's can exceed 1 in order that the sum doesn't exceed t. For $1 \leqslant j \leqslant k$, we can write

$$\begin{aligned} \mathcal{B}_1(t) \cap \cdots \cap \mathcal{B}_j(t) = \{(u_1, \ldots, u_q) \,|\, 1 < u_i, i = 1, \ldots, j;\ 0 < u_i, i = j + 1, \ldots, q; \\ \text{and } u_1 + \cdots + u_q \leqslant t\}. \end{aligned} \tag{13.90}$$

In the final set, let $v_i = u_i - 1$ for $i = 1, \ldots, j$, and $v_i = u_i$ for $i = j + 1, \ldots, q$. Then all $v_i > 0$, and

$$u_1 + \cdots + u_q \leqslant t \iff v_1 + \cdots + v_q \leqslant t - j. \tag{13.91}$$

Thus the right-hand side of (13.90) is the same as $\mathcal{A}(t - j)$ with the u_i's replaced by v_i's. Hence

$$\text{Volume}(\mathcal{B}_1(t) \cap \cdots \cap \mathcal{B}_j(t)) = \text{Volume}(\mathcal{A}(t - j)) = \frac{(t - j)^q}{q!}, \tag{13.92}$$

and by (13.87) and (13.89), for $k = 0, \ldots, q - 1$,

$$F_{IH}(t) = \sum_{j=0}^{k} (-1)^j \binom{q}{j} \frac{(t - j)^q}{q!} \quad \text{for} \quad k \leqslant t < k + 1. \tag{13.93}$$

The density in (14.24) is found by differentiating, at least for t not equalling an integer, and using the floor function ($k = \lfloor t \rfloor$).

Chapter 14

Tied rankings in just one variable: Kendall

In this chapter we treat Kendall's distance when there are ties in just one of the variables. We assume the second variable $\boldsymbol{Y}$ has no ties, and treat the first variable $\boldsymbol{W}$ as fixed at $\boldsymbol{w}$. In this case, the Kendall distance adjusted for ties is equal, modulo an additive constant, to the Mann-Whitney/Wilcoxon (MWW) statistic if $\boldsymbol{w}$ has just two distinct values, and the Jonckheere-Terpstra (JT) statistic for arbitrary $\boldsymbol{w}$. See Wilcoxon (1945) and Mann & Whitney (1947) for the former, and Terpstra (1952) and Jonckheere (1954) for the latter. Most books on nonparametric methods cover these procedures. Because the analysis in this case is so much easier to deal with than the case with ties in both variables, and the MWW and JT statistics are of interest in their own rights, we split the analysis into two chapters. Chapter 16 handles the more general case.

A special case of the JT statistic is what Silverberg (1980) calls a q-permutation, which is a tied ranking in which the top q items are ranked from 1 to q, but all other items are tied at $q+1$. This type of ranking arises quite frequently, e.g., when people are asked to rank their top ten movies out of possibly hundreds. The exact distribution is easier to find than that for the general JT statisistic.

In Section 14.1, we present the above statistics, and their relationship to Kendall's distance with ties. In Section 14.2 we give fairly simple formulas for the cumulants. An algorithm for the exact distribution for medium m is given in Section 14.3. Section 14.4 contains results for asymptotic normality, and another approximation based on the sum of independent uniforms for special cases. Proofs are in Section **??**. We apply the asymptotic approximations in Section 14.5.

14.1 The Mann-Whitney/Wilcoxon and Jonckheere-Terpstra statistics

The Mann-Whitney statistic (Mann & Whitney, 1947) is a nonparametric statistic used to test the equality of two populations based on independent random samples of a single variable from each population. The statistic counts the number of times a value in the first sample exceeds one from the second, at least if there are no ties. Let $\boldsymbol{y}$ be the vector of ranks for the combined samples, where the first m_1 components are the ranks for the first sample, and the remaining m_2 components are the ranks of the second sample. If we set $\boldsymbol{w} = (1, 1, \ldots, 1, 2, \ldots, 2)$,

where there are m_1 1's and m_2 2's, then the Mann-Whitney statistic is exactly the Kendall distance (without adjustment for ties) given in (1.4) between $\boldsymbol{w}$ and $\boldsymbol{y}$:

$$\begin{aligned} d_{\text{Ken}}(\boldsymbol{w},\boldsymbol{y}) = \sum_{1\leqslant i<j\leqslant m} I[(w_i - w_j)(y_i - y_j) < 0] &= \sum_{i=1}^{m_1}\sum_{j=m_1+1}^{m_1+m_2} I[y_i > y_j] \\ &\equiv d_{\text{MWW}}(\mathcal{Y}_1,\mathcal{Y}_2);\ \ \mathcal{Y}_j = \{y_i \,|\, w_i = j\}. \end{aligned} \tag{14.1}$$

The Wilcoxon (1945) statistic is based on the sum of the ranks in the first sample, $y_1 + \cdots + y_{m_1}$. It is well-known to be equivalent to the Mann-Whitney statistic, since it can be written

$$\begin{aligned} d_{\text{Ken}}(\boldsymbol{w},\boldsymbol{y}) &= \sum_{i=1}^{m_1}\sum_{j=m_1+1}^{m} I[y_i > y_j] \\ &= \sum_{i=1}^{m_1}\sum_{j=1}^{m} I[y_i > y_j] - \sum_{i=1}^{m_1}\sum_{j=1}^{m_1} I[y_i > y_j] \\ &= \sum_{i=1}^{m_1}(y_i - 1) - \binom{m_1}{2} \\ &= \sum_{i=1}^{m_1} y_i - \frac{m_1(m_1+1)}{2}. \end{aligned} \tag{14.2}$$

We will refer to this statistic as the Mann-Whitney/Wilcoxon (MWW) statistic. It is also equivalent to Spearman's distance adjusted for ties in the first variable as in (13.1), so that with the r_i's being the midranks for the w_i's,

$$\begin{aligned} d^{A}_{\text{Spear}}(\boldsymbol{w},\boldsymbol{y}) &= \frac{2m(m+1)(2m+1)}{3} - 2\sum_{i=1}^{m} r_i y_i \\ &= \frac{2m(m+1)(2m+1)}{3} - 2\left(\frac{m_1+1}{2}\sum_{i-1}^{m_1} y_i + \left(m_1 + \frac{m_2+1}{2}\right)\sum_{i=m_1+1}^{m_1+m_2} y_i\right) \\ &= \frac{m(m+1)}{6}(1 + 5m - 3m_1) + m\sum_{i=1}^{m_1} y_i. \end{aligned} \tag{14.3}$$

The Jonckheere-Terpstra (JT) statistic is a similar statistic that compares $K \geqslant 2$ populations, looking to see if there is a trend in the variable over the population indices. Here we have $\boldsymbol{w} = (1,\ldots,1,2,\ldots,2,\ldots,K,\ldots,K)$, and define the JT statistic by

$$\begin{aligned} d_{\text{Ken}}(\boldsymbol{w},\boldsymbol{y}) = d_{\text{JT}}(\mathcal{Y}_1,\ldots,\mathcal{Y}_K) &= \sum_{a=1}^{K-1}\sum_{i\in\mathcal{Y}_a}\sum_{j=m_a+1}^{m} I[y_i > y_j] \\ &= \sum_{a=1}^{K-1} W_a, \ \text{ where } \ W_a = d_{\text{MWW}}(\mathcal{Y}_a, \mathcal{Y}_{a+1}\cup\cdots\cup\mathcal{Y}_K). \end{aligned} \tag{14.4}$$

In (14.1) and (14.4), we have not adjusted the distance for ties as in (12.35). Since there are no ties in the $\boldsymbol{y}$, the adjustment as in (12.36) reduces to

$$d^{A}_{Ken}(\boldsymbol{w},\boldsymbol{y}) = d_{JT}(y_1,\ldots,y_K) + \frac{\sum m_a^2 - m}{4}. \tag{14.5}$$

14.2 Moments and cumulants

The results here were originally obtained by Jonckheere (1954). We start by showing the independence of the MWW statistic and the ranks of the Y_i's within each group, which is Theorem 1 of Terpstra (1952).

Lemma 14.1. *Suppose* $\boldsymbol{Y} \sim$ *Uniform*$(\mathcal{P}_m)$, $m = m_1 + m_2$, *and let* $y_1 = \{Y_1,\ldots,Y_{m_1}\}$ *and* $y_2 = \{Y_{m_1+1},\ldots,Y_m\}$. *Then the three quantities*

$$\boldsymbol{Y}^{(1)} \equiv \textit{rank}(Y_1,\ldots,Y_{m_1}), \boldsymbol{Y}^{(2)} \equiv \textit{rank}(Y_{m_1+1},\ldots,Y_m), \textit{ and } d_{MWW}(y_1,y_2) \tag{14.6}$$

are independent, where $\boldsymbol{Y}^{(1)} \sim$ *Uniform*$(\mathcal{P}_{m_1})$ *and* $\boldsymbol{Y}^{(2)} \sim$ *Uniform*$(\mathcal{P}_{m_2})$.

The proof is based on the observation that the order of the elements within the two subsets y_1 and y_2 are independent since the subsets are disjoint, and also independent of the MWW statistic since the latter only compares the values between the two subsets.

The lemma can be easily extended to several groups, where for the situation in (14.4), we have the $2L-1$ quantities

$$W_1, W_2, \ldots, W_{K-1}, \text{ and } \text{rank}(\boldsymbol{Y}^{(1)}), \text{ rank}(\boldsymbol{Y}^{(2)}), \ldots, \text{ rank}(\boldsymbol{Y}^{(K)}) \text{ are independent.} \tag{14.7}$$

This result follows by first using the lemma to show that W_1, $\text{rank}(\boldsymbol{Y}^{(1)})$, and $\text{rank}(\boldsymbol{Y}^{(2)},\ldots,\boldsymbol{Y}^{(K)})$ are independent. Then apply the lemma to that last vector to show that W_2, $\text{rank}(\boldsymbol{Y}^{(2)})$, and $\text{rank}(\boldsymbol{Y}^{(3)},\ldots,\boldsymbol{Y}^{(K)})$) are independent, and still independent of W_1 and $\text{rank}(\boldsymbol{Y}^{(1)})$. Continuing, we have (14.7).

Equation (14.7) can be used to find the cumulants for the JT statistic. Letting $\boldsymbol{e}_k = (1,\ldots,k)$, write

$$\begin{aligned} d_{Ken}(\boldsymbol{e}_m,\boldsymbol{Y}) &= \sum_{1\leqslant i<j\leqslant m} I[Y_i > Y_j] \\ &= \sum_{a=1}^{K-1}\sum_{i=1}^{m_a}\sum_{j=m_a+1}^{m} I[y_i > y_j] + \sum_{1\leqslant i<j\leqslant m_1} I[Y_i > Y_j] \\ &+ \sum_{m_1+1\leqslant i<j\leqslant m_1+m_2} I[Y_i > Y_j] + \cdots + \sum_{m_1+\cdots+m_{K-1}+1\leqslant i<j\leqslant m} I[Y_i > Y_j] \\ &= d_{JT}(y_1,\ldots,y_K) + d_{Ken}(\boldsymbol{e}_{m_1},\boldsymbol{Y}^{(1)}) + d_{Ken}(\boldsymbol{e}_{m_2},\boldsymbol{Y}^{(2)}) + \cdots + d_{Ken}(\boldsymbol{e}_{m_K},\boldsymbol{Y}^{(K)}). \end{aligned} \tag{14.8}$$

By Lemma 14.1, the K+1 terms in the final expression are independent. Since cumulants are linear, the n^{th} cumulant of the (first) left-hand side of (14.8) equals the sum of the n^{th}

cumulants of the independent terms. Let $\kappa_n^{\text{Ken}}(k)$ be the n^{th} cumulant for Kendall's distance based on $\boldsymbol{Y}$ of length k. This cumulant is given in (6.13). Then we have that the n^{th} cumulant of the JT statistic, with group sizes $m_1, \ldots, m_K$, is

$$\kappa_n^{\text{JT}}(m_1, \ldots, m_K) = \kappa_n^{\text{Ken}}(m) - \sum_{a=1}^{K} \kappa_n^{\text{Ken}}(m_a). \tag{14.9}$$

In particular, from (1.10),

$$E[d_{\text{JT}}(y_1, \ldots, y_K)] = \frac{m(m+1)}{4} - \sum_{a=1}^{K} \frac{m_a(m_a+1)}{4} = \frac{m^2 - \sum_{a=1}^{K} m_a^2}{4}, \tag{14.10}$$

and

$$\begin{aligned} Var[d_{\text{JT}}(y_1, \ldots, y_K)] &= \frac{m(m-1)(2m+5)}{72} - \sum_{a=1}^{K} \frac{m_a(m_a-1)(2m_a+5)}{72} \\ &= \frac{m^3 - \sum_{a=1}^{K} m_a^3}{36} + \frac{m^2 - \sum_{a=1}^{K} m_a^2}{24}. \end{aligned} \tag{14.11}$$

For the MWW statistic, we simplify to

$$E[d_{\text{MWW}}(y_1, y_2)] = \frac{m_1 m_2}{2} \quad \text{and} \quad Var[d_{\text{MWW}}(y_1, y_2)] = \frac{m_1 m_2 (m+1)}{12}. \tag{14.12}$$

14.3 Exact distribution

We start with the recursive algorithm from Gibbons & Chakraborti (2010, page 265) for finding the exact distribution of the MWW statistic. It is based on the algorithm in Terpstra (1953) for the JT statistic. Fixing m_1 and m_2, the MWW statistic in (14.1) can be written

$$d_{\text{Ken}}(\boldsymbol{w}, \boldsymbol{y}) = d_{\text{MWW}}(y_1, y_2) = \#\{y_i > y_j \,|\, y_i \in y_1, y_2 \in y_2\}. \tag{14.13}$$

The possible values range from 0 to $m_1 m_2$. There are $\binom{m}{m_1}$ ways to allocate m_1 of the y_i's to y_1, and the rest to y_2, hence the distribution of the MWW statistic is

$$P[d_{\text{MWW}}(y_1, y_2) = u] = \frac{c(u; m_1, m_2)}{\binom{m}{m_1}}, \quad u = 0, \ldots, m_1 m_2, \tag{14.14}$$

where

$$c(u; m_1, m_2) = \#\{\text{Ways to allocate } m_1 \text{ of the } y_i\text{'s to } y_1 \,|\, \#\{y_i > y_j \,|\, y_i \in y_1, y_j \in y_2\} = u\}. \tag{14.15}$$

For each such allocation, the y_k that equals m is either in y_1 or y_2. Removing that element will be an allocation of $1, \ldots, m-1$, for which we can find the statistic as follows, where $d_{\text{MWW}}(y_1, y_2) = u$:

$$\begin{array}{rcl} m \in y_1 & \Rightarrow & d_{\text{MWW}}(y_1 - \{m\}, y_2) = u - m_2; \\ m \in y_2 & \Rightarrow & d_{\text{MWW}}(y_1, y_2 - \{m\}) = u \end{array} \tag{14.16}$$

The first line holds because m is larger than anything in $\mathcal{Y}_2$, hence contributes m_2 to the statistic. (Note that if $u < m_2$, then m cannot be in $\mathcal{Y}_1$.) For the second line, it contributes nothing since everything in $\mathcal{Y}_1$ is smaller than m. Now any configuration of m y_i's that has statistic value u is associated with (possibly) two configurations of $m-1$ y_i's, as in (14.16). Thus

$$c(u; m_1, m_2) = c(u - m_2; m_1 - 1, m_2) + c(u; m_1, m_2 - 1). \tag{14.17}$$

Then the function c for any m_1, m_2 can be found iteratively from those with smaller m_i's. We can start the process with one of the m_i's being one, where

$$c(u; 1, m_2) = 1, u = 0, \ldots, m_2 \quad \text{and} \quad c(u; m_1, 1) = 1, u = 0, \ldots, m_1. \tag{14.18}$$

Also, $c(u; m_1, m_2) = 0$ if $u < 0$.

For the JT statistic (14.4), we first find the exact distribution of all the component W_a's, which by Lemma 14.1 are independent. Now the convolution of those $L-1$ distributions yields the exact distribution of the JT statistic.

Turn to the q-permutations. Using (14.6), we can write

$$d_{\text{Ken}}(\boldsymbol{w}, \boldsymbol{Y}) = \sum_{j=1}^{q} V_j \text{ where } V_j = \sum_{i=j+1}^{m} I[Y_i < Y_j]. \tag{14.19}$$

This decomposition is equivalent to that in (6.4), but with the summations on i and j switched. Then the $V_1, \ldots, V_q$ are independent, with

$$V_j \sim \text{Uniform}\{0, \ldots, m-j\}. \tag{14.20}$$

Thus we can perform a straightforward convolution of the q discrete uniforms, just as we did for Kendall's distance without ties in Section 6.3, though we need fewer terms.

14.4 Asymptotic distributions

Here we find conditions for the JT statistic to have a limit, either normal or a sum of uniforms. Section 14.5 evaluates the resulting approximations.

We start with conditions for asymptotic normality. Let

$$D_{\text{JT}}^{(\boldsymbol{m})} = d_{\text{JT}}(\mathcal{Y}_1, \ldots, \mathcal{Y}_K). \tag{14.21}$$

The components m_a in $\boldsymbol{m}$, and K, depend on m, though the notation won't reflect the dependence. Next is the main result. It is a bit more general than the results in Terpstra (1952) and Jonckheere (1954), in that these authors seem to assume K is bounded and $\limsup m_\chi/m < 1$, where $m_\chi = \max\{m_a\}$. The result follows from the more general Theorem 15.1.

Theorem 14.2. *If as* $m \to \infty$ *we have* $m - m_\chi \to \infty$, *then*

$$\frac{D_{JT}^{(\boldsymbol{m})} - E[D_{JT}^{(\boldsymbol{m})}]}{\sqrt{Var[D_{JT}^{(\boldsymbol{m})}]}} \longrightarrow^{\mathcal{D}} N(0,1). \tag{14.22}$$

If the condition $m - m_x \to \infty$ fails, we have the following alternative result, which is shown to be a consequence of Theorem 15.1 at the end of Section 15.2.

Theorem 14.3. *Suppose that for some finite* q, $m - m_x \to q$ *as* $m \to \infty$. *Then*

$$\frac{D_{JT}^{(\boldsymbol{m})}}{m} \longrightarrow^{\mathcal{D}} U_1 + \cdots + U_q, \tag{14.23}$$

where $U_1, \ldots, U_q$ *are independent Uniform(0,1) random variables.*

If $m - m_x \not\to \infty$, then there must exist a finite q such that $m - m_x \to q$, at least on a subsequence. Theorem 14.3 shows (14.23) on that subsequence, hence either that limit holds for the entire sequence, or the sequence does not have a limit. Thus the condition in Theorem 14.2 is also necessary for asymptotic normality.

The distribution of a sum of q independent Uniform(0,1)'s is called the **Irwin-Hall distribution**, with parameter q. See Wikipedia contributors (2019a) and Marengo et al. (2017) for overviews. The latter give a nice geometric proof of the density, which is

$$f_{IH}(t\,;\,q) = \frac{1}{(q-1)!}\sum_{k=0}^{\lfloor t \rfloor}(-1)^k\binom{q}{k}(t-k)^{q-1}, \quad 0 < t < q. \tag{14.24}$$

We sketch their proof in Section 13.5.3. The density is symmetric about $q/2$, and if $t > q/2$, it is more efficient and more stable numerically to calculate $f_{IH}(q - t; q)$.

14.5 Edgeworth and Irwin-Hall approximations

The exact algorithm for Kendall's distance with no ties given in Section 6.3 is reasonably fast for m up to about 1325. The related algorithm for q-permutations in Section 14.3 is also fast for m up to 1200, or if $q \leqslant 950000/m + 150$. The algorithm for more general patterns of ties seems to be slower. In the MWW case, so that $K = 2$, it appears that it is fine if $m_1 m_2 \leqslant 25000$, or for the general JT case if $m^*(m - m^*) \leqslant 25000$, where $m^* = \max\{m_1, \ldots, m_K, m/2\}$.

If the m_a's are too large for the exact algorithms to be feasible, either the Edgeworth or Irwin-Hall approximations are usually very accurate. The latter are good when the $m_x = \max\{m_1, \ldots, m_K\}$ is close to m. Figure 14.1 graphs the log of the errors for the Edgeworth approximation with $L = 10$ and the Irwin-Hall approximation for the MWW statistic, where $m_1 = 2, \ldots, 10$ and $m_2 = 100, 200, \ldots, 500$. (If $m_1 = 1$, then the distribution is exactly discrete uniform from 0 to m_2.) For Edgeworth, the approximations generally improve as m_1 and m_2 increase, and are excellent if $m_1 = 10$, being about $10^{-8}, 10^{-7}$, and 10^{-3} for the maximum error in density, distribution function, and relative p-value, respectively, as given in (4.54) and (4.54). For larger m_1 and m_2, the approximation is even better. The errors in the Irwin-Hall approximation tends to be relatively flat, and are better than Edgeworth only if $m_1 \leqslant 3$ or 4, or maybe 6 if the relative p-value is of most import. Note that for the situations in these plots, the exact distribution is easy to find.

The table in (14.25) gives the cutoff points for m_1 when the better approximation for the MWW statistic switches from Irwin-Hall to Edgeworth, and $m_2 \geqslant 500$. Values of m_1 less than

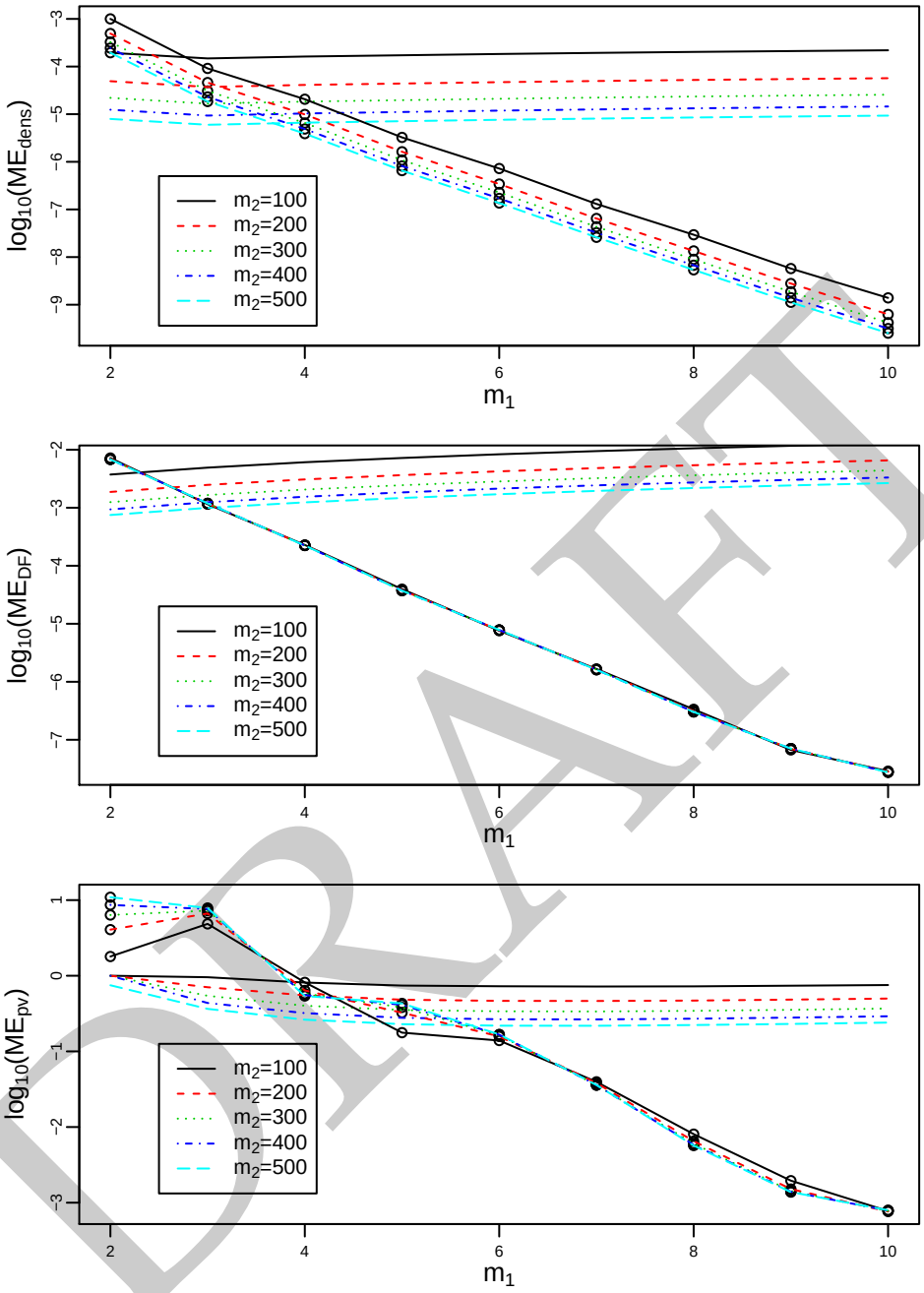

Figure 14.1: Comparing the $\log_{10}$(errors) of the Edgeworth with $L = 10$ and Irwin-Hall approximations for MWW statistics with $\boldsymbol{m} = (m_1, m_2)$, where m_1 is small. The horizontal axis is m_1, and the vertical is the log of the errors. The different lines represent different m_2. In each graph, the relatively horizontal lines denote the errors for the Irwin-Hall approximation, and the lines decreasing substantially denote those for the Edgeworth approximation. The three panels graph the errors in the density, distribution function, and relative p-values, respectively.

or equal to the entry favor the former, and those greater than favor the latter. A reasonable guideline is to use Irwin-Hall if $m_1 \leqslant \lceil (4/3) \log_{10}(m_2) \rceil$, and Edgeworth otherwise.

$m_2 \rightarrow$	500	1000	2000	3000	4000	5000	7500	10000
Density	3	4	4	4	4	4	5	5
Distribution function	3	3	3	4	4	4	4	4
Relative p-value	5	6	6	6	7	7	7	7

(14.25)

The MWW statistic is generally a worst-case, in that these approximations improve as the groups in the $\boldsymbol{m}$ are further subdivided, so that the JT statistic with $\boldsymbol{m} = (10, 100, 100, 100)$ has better approximations than the MW statistic with $\boldsymbol{m} = (10, 300)$.

Chapter 15

Tied rankings in both variables: Kendall

This chapter treats Kendall's distance when there are ties in both variables, where as in (12.35),

$$d^{A}_{Ken}(\boldsymbol{w},\boldsymbol{z}) = \sum_{1\leqslant j<i\leqslant m}\sum \left(I[(w_i - w_j)(z_i - z_j) < 0] + \tfrac{1}{2}\, I[(w_i - w_j)(z_i - z_j) = 0] \right). \tag{15.1}$$

The results proceed similar to those in Chapter 13 for Spearman's distance. Section 15.1 relates the distance to the contingency tables in Section 13.1.2, so that we can iterate over all the tables (if there are not too many) to find the exact distribution. Section 15.2 describes conditions under which the distance is asymptotically normal, which are verified in Section **??**. We apply the normal and Edgeworth approximations to the distribution in Section 15.3, and compare them to using simulations. The Edgeworth approximations need higher moments, but so far we have only been able to find analytic expressions for the moments up to order four, hence we can use Edgeworth expansions up to $L = 2$. Their (tedious) derivations are in Section 15.4.

15.1 Exact distribution: Contingency tables

Here we use contingency tables as for Spearman's distance in Section 13.1.2, as in Brown (1988). For tied rankings $\boldsymbol{w}$ and $\boldsymbol{z}$, let $\boldsymbol{T}(\boldsymbol{w},\boldsymbol{z})$ be the $K \times L$ contingency table as in (13.16), so that $t_{ab} = T_{ab}(\boldsymbol{w},\boldsymbol{z}) = \#\{i \,|\, w_i = a, z_i = b\}$. We can then find the number of concordances and discordances from the table:

$$\begin{aligned} C &\equiv \sum_{1\leqslant j<i\leqslant m}\sum I[(w_i - w_j)(z_i - z_j) > 0] = \sum_{1\leqslant a_1<a_2\leqslant K, 1\leqslant b_1<b_2\leqslant L}\cdots\sum t_{a_1b_1}t_{a_2b_2}, \\ D &\equiv \sum_{1\leqslant j<i\leqslant m}\sum I[(w_i - w_j)(z_i - z_j) < 0] = \sum_{1\leqslant a_1<a_2\leqslant K, 1\leqslant b_1<b_2\leqslant L}\cdots\sum t_{a_1b_2}t_{a_2b_1}. \end{aligned} \tag{15.2}$$

Since

$$C + D + \sum_{1\leqslant j<i\leqslant m}\sum I[(w_i - w_j)(z_i - z_j) = 0] = \binom{m}{2}, \tag{15.3}$$

from (12.35) we have

$$\begin{aligned} d^A_{Ken}(\boldsymbol{w}, \boldsymbol{z}) &= D + \tfrac{1}{2} I[(w_i - w_j)(z_i - z_j) = 0] \\ &= \frac{1}{2}\left(D - C + \binom{m}{2}\right). \end{aligned} \tag{15.4}$$

To find the exact distribution for Kendall's distance, we can run through all the possible tables and sum up their probabilities, as for Spearman in (13.20).

15.2 Asymptotic distributions

Alvo & Yu (2014) prove the asymptotic normality of Kendall's distance with ties and missing values by showing it is close enough to Spearman's distance to borrow its asymptotics. We again have $\boldsymbol{m}$ and $\boldsymbol{n}$ being the pattern of ties for $\boldsymbol{W}$ and $\boldsymbol{Z}$, respectively, and set $m_x = \max\{m_a\}$ and $n_z = \max\{n_a\}$. As in (13.44), let F_m be the distribution function for S_i/m, where $\boldsymbol{S} = \text{rank}(\boldsymbol{Z})$, Let $\mu_{Ken}(m)$ and $\sigma^2_{Ken}(\boldsymbol{m}, \boldsymbol{n})$ denote the mean and variance of $d^A_{Ken}(\boldsymbol{W}, \boldsymbol{Z})$. Then we have the following.

Theorem 15.1. *Suppose that* $m \to \infty$. *Then*

$$\frac{(m - m_x)(m - n_z)}{m} \to \infty \Longrightarrow \frac{d^A_{Ken}(\boldsymbol{W}, \boldsymbol{Z}) - \mu_{Ken}(m)}{\sigma_{Ken}(\boldsymbol{m}, \boldsymbol{n})} \longrightarrow^{\mathcal{D}} N(0, 1). \tag{15.5}$$

On the other hand, if $q_1 = m_{<x}$ *and* $q_2 = m_{>x}$ *are fixed, with* $q = q_1 + q_2$, *and* $F_m(x) \to F(x)$ *on points of continuity, then*

$$\frac{1}{m}\left(d^A_{Ken}(\boldsymbol{w}, \boldsymbol{Z}) - \mu_{Ken}(m)\right) + \frac{q}{2} \longrightarrow^{\mathcal{D}} V_1 + \ldots + V_{q_1} + (1 - V_{q_1+1}) + \cdots + (1 - V_q), \tag{15.6}$$

where the V_i *are iid with distribution given by* F.

If there are no ties in $\boldsymbol{Y}$, then we have the Jonckheere-Terpstra statistic as in (14.4). In this case, $n_z = 1$, hence the condition in (15.5) for asymptotic normality is that $m - m_x \to \infty$, proving Theorem 14.2.

Next, suppose $q_1 = m_{<x}$ and $q_2 = m_{>x}$ are fixed. Using (14.5), and (1.10) for the mean,

$$\begin{aligned} d^A_{Ken}(\boldsymbol{w}, \boldsymbol{y}) - \mu_{Ken}(m) &= d_{JT}(y_1, \ldots, y_K) + \frac{\sum m_a^2 - m}{4} - \frac{m(m-1)}{4} \\ &= d_{JT}(y_1, \ldots, y_K) - \frac{m^2 - \sum m_a^2}{4}. \end{aligned} \tag{15.7}$$

Now

$$m^2 - \sum m_a^2 = m^2 - m_x^2 - \sum_{a \neq x} m_a^2 = (m + m_x)q - \sum_{a \neq x} m_a^2. \tag{15.8}$$

With q being fixed, $m_x/m \to 1$ and $\sum_{a \neq x} m_a^2$ is bounded. Thus (15.6) and (15.8) show that

$$\frac{1}{m} d_{JT}(y_1, \ldots, y_K) \longrightarrow^{\mathcal{D}} V_1 + \ldots + V_{q_1} + (1 - V_{q_1+1}) + \cdots + (1 - V_q). \tag{15.9}$$

Since the $n_a = 1$, the second part of Lemma 13.4 shows that the V_i are Uniform(0,1), hence so are the $1 - V_i$. Note also that the asymptotic distribution depends on q_1 and q_2 only through their sum q, hence need not be fixed as long as q is. Thus we have Theorem 14.3.

To prove Theorem 15.1, we follow Alvo and Yu's Theorem 3.4, we starting with the following lemma, then apply it to show that the results for Spearman imply the results in the theorem.

Lemma 15.2. *Let* $c_m = m + 1 - \sqrt{m+1}$. *Then*

$$E\left[\left(\frac{d^A_{Spear}(\boldsymbol{w}, \boldsymbol{Z}) - \mu_{Spear}(m)}{c_m} - (d^A_{Ken}(\boldsymbol{w}, \boldsymbol{Z}) - \mu_{Ken}(m))\right)^2\right] \leqslant \frac{m(m - m_x)}{12}. \tag{15.10}$$

Proof of Lemma 15.2. We first find the covariance of the two distances in the Jonckheere-Terpstra case, i.e., $\boldsymbol{W}$ has ties given by $\boldsymbol{m}$, and $\boldsymbol{Y}$ is without ties. We fix $\boldsymbol{W} = \boldsymbol{w}$, and set

$$\text{rank}(\boldsymbol{w}) = \boldsymbol{r} = (t_1, \ldots, t_1, \ldots, t_K, \ldots, t_K), \quad \text{where } t_a = m_{<a} + \frac{m_a + 1}{2}, \tag{15.11}$$

and there are m_a elements equal to t_a in $\boldsymbol{r}$. Then from (13.67),

$$\text{Cov}[d^A_{\text{Spear}}(\boldsymbol{w}, \boldsymbol{Y}), d^A_{\text{Ken}}(\boldsymbol{w}, \boldsymbol{Y})] = -2\,\text{Cov}[\sum_{i=1}^{m} r_i Y_i, d^A_{\text{Ken}}(\boldsymbol{w}, \boldsymbol{Y})]. \tag{15.12}$$

Using the decomposition in (14.8) for Kendall, we have

$$\text{Cov}[\sum_{i=1}^{m} r_i Y_i, d^A_{\text{Ken}}(\boldsymbol{w}, \boldsymbol{Y})] = \text{Cov}[\sum_{i=1}^{m} r_i Y_i, d_{\text{Ken}}(\boldsymbol{e}_m, \boldsymbol{Y})] - \sum_{a=1}^{K} \text{Cov}[\sum_{i=1}^{m} r_i Y_i, d_{\text{Ken}}(\boldsymbol{e}_{m_a}, \boldsymbol{Y}^{(a)})], \tag{15.13}$$

where $\boldsymbol{e}_l = (1, 2, \ldots, l)$, and $\boldsymbol{Y}^{(a)} = (Y_{m_{<a}+1}, \ldots, Y_{m_{\leqslant a}})$. Note that since the r_i are constant within each group defined by the t_a's, the $\sum r_i Y_i$ depends on $\boldsymbol{Y}$ only through the sums of the component $\boldsymbol{Y}^{(a)}$'s, hence only on the comparisons of the Y_i's between groups. Also, the $d_{\text{Ken}}(\boldsymbol{e}_{m_a}, \boldsymbol{Y}^{(a)})$'s depend only on the comparisons within groups. Thus Lemma 14.1 implies that the summation is independent of the Kendall distances, hence the covariances in the final sum of (15.14) are all zero.

Turning to the first covariance on the right-hand side, consider

$$\text{Cov}[Y_i, d_{\text{Ken}}(\boldsymbol{e}_m, \boldsymbol{Y})] = \sum_{1 \leqslant h < j \leqslant m} \sum \text{Cov}[Y_i, I[Y_h > Y_j]]. \tag{15.14}$$

If $i \neq h$ and $i \neq j$, Y_i is independent of $I[Y_h > Y_j]$. With $i = h < j$, we have

$$\begin{aligned}
\text{Cov}[Y_i, I[Y_i > Y_j]] &= \text{Cov}[E[Y_i \,|\, Y_i], E[I[Y_i > Y_j] \,|\, Y_i]] + E[\text{Cov}[Y_i, I[Y_i > Y_j] \,|\, Y_i]] \\
&= \text{Cov}[Y_i, E[I[Y_i > Y_j] \,|\, Y_i]] + 0 \\
&= \text{Cov}[Y_i, (Y_i - 1)/(m-1)] \\
&= \frac{1}{m-1} \text{Var}[Y_i] = \frac{1}{m-1} \frac{m^2 - 1}{12} = \frac{m+1}{12}.
\end{aligned} \tag{15.15}$$

(In the last term on the first line, the conditional covariance is zero since conditionally the Y_i is a constant.) Then if $i = j > h$, $\mathrm{Cov}[Y_i, I[Y_h > Y_i]] = -(m+1)/12$. Thus on the right-hand side of (15.14), there are $m - i$ positive terms and $i - 1$ negative ones, and the rest are zero, hence

$$\mathrm{Cov}[Y_i, d_{\mathrm{Ken}}(\boldsymbol{e}_m, \boldsymbol{Y})] = \frac{(m-2i+1)(m+1)}{12} = -\frac{(m+1)(i-\nu)}{6}, \quad \nu = \frac{m+1}{2}. \tag{15.16}$$

Thus from (15.14) and (15.16), noting that replacing r_i with $r_i - \nu$ does not change the covariance since the sum of the Y_i's is constant, we can write

$$\begin{aligned} \mathrm{Cov}[\sum_{i=1}^{m} r_i Y_i, d_{\mathrm{Ken}}(\boldsymbol{e}_m, \boldsymbol{Y})] &= -\frac{(m+1)\sum_{i=1}^{m}(r_i-\nu)(i-\nu)}{6} \\ &= -\frac{(m+1)\sum_{i=1}^{m}(r_i-\nu)^2}{6}, \end{aligned} \tag{15.17}$$

since the average of the i for which $r_i = t_a$ is t_a. We multiply that value by -2 to find the covariance in (15.12). Using the expression for the variance Spearman's distance from (13.36) when there are no ties in $\boldsymbol{Y}$, we have that

$$\mathrm{Cov}[d^A_{\mathrm{Spear}}(\boldsymbol{w}, \boldsymbol{Y}), d^A_{\mathrm{Ken}}(\boldsymbol{w}, \boldsymbol{Y})] = \frac{(m+1)\sum_{i=1}^{m}(r_i-\nu)^2}{3} = \frac{1}{m}\,\mathrm{Var}[d^A_{\mathrm{Spear}}(\boldsymbol{w}, \boldsymbol{Y})], \tag{15.18}$$

the variance for Spearman being found in (13.36), where here the $s_i = 1$.

The variance for Kendall is found in (14.11):

$$\begin{aligned} \mathrm{Var}[d^A_{\mathrm{Ken}}(\boldsymbol{w}, \boldsymbol{Y})] &= \frac{m^3 - \sum_{a=1}^{K} m_a^3}{36} + \frac{m^2 - \sum_{a=1}^{K} m_a^2}{24} \\ &= \frac{\sum_{i=1}^{m}(r_i-\nu)^2}{3} + \frac{m^2 - \sum_{a=1}^{K} m_a^2}{24}, \end{aligned} \tag{15.19}$$

where we use (13.57) to go from the second to third equality. The two previous equations yield

$$\begin{aligned} &\mathrm{E}\left[\left(\frac{d^A_{\mathrm{Spear}}(\boldsymbol{w}, \boldsymbol{Y}) - \mu_{\mathrm{Spear}}(m)}{c_m} - (d^A_{\mathrm{Ken}}(\boldsymbol{w}, \boldsymbol{Y}) - \mu_{\mathrm{Ken}}(m))\right)^2\right] \\ &\quad = \frac{\sum_{i=1}^{m}(r_i-\nu)^2}{3}\left(\frac{m(m+1)}{c_m^2} - 2\frac{m+1}{c_m} + 1\right) + \frac{m^2 - \sum_{a=1}^{K} m_a^2}{24} \\ &\quad = \frac{m^2 - \sum_{a=1}^{K} m_a^2}{24} \leqslant \frac{m(m-m_x)}{12}, \end{aligned} \tag{15.20}$$

since the given $c_m = m + 1 - \sqrt{m+1}$ renders the first term in the second line zero, and the inequality follows from (15.8).

Now for ties in the $\boldsymbol{Y}$, i.e., in $\boldsymbol{Z}$, we have for any distance that $d^A(\boldsymbol{w}, \boldsymbol{z}) = \mathrm{E}[d^A(\boldsymbol{w}, \boldsymbol{Y}) \mid \boldsymbol{Z} = \boldsymbol{z}]$. Following Alvo and Yu, we use Jensen's inequality to show that for a given function h,

$$\begin{aligned} \mathrm{E}[h(\boldsymbol{w}, \boldsymbol{Y})^2] &= \mathrm{E}[\mathrm{E}[h(\boldsymbol{w}, \boldsymbol{Y})^2 \mid \boldsymbol{Z}]] \\ &\geqslant \mathrm{E}[\mathrm{E}[h(\boldsymbol{w}, \boldsymbol{Y}) \mid \boldsymbol{Z}]^2]. \end{aligned} \tag{15.21}$$

Letting h be the quantity that is squared in the expectation in the first expression in (15.20), (15.10) then follows. □

Proof of Theorem 15.1. Consider (15.5), and assume that

$$\frac{(m-m_x)(m-n_z)}{m} \longrightarrow \infty. \tag{15.22}$$

Using (13.36), (13.57), and (15.19), we can show that

$$\begin{aligned}\sigma^2_{\text{Spear}}(\boldsymbol{m},\boldsymbol{n}) \equiv \mathbb{V}\text{ar}[d^A_{\text{Spear}}(\boldsymbol{W},\boldsymbol{Z})] &= \frac{(m^3-\sum m_a^3)(m^3-\sum n_a^3)}{36(m-1)}, \text{ and}\\ \sigma^2_{\text{Ken}}(\boldsymbol{m},\boldsymbol{n}) \equiv \mathbb{V}\text{ar}[d^A_{\text{Ken}}(\boldsymbol{W},\boldsymbol{Z})] &= \frac{(m^3-\sum m_a^3)(m^3-\sum n_a^3)}{36(m)_3} + O(m^2)\\ &= \frac{\sigma^2_{\text{Spear}}(\boldsymbol{m},\boldsymbol{n})}{m(m-2)} + O(m^2).\end{aligned} \tag{15.23}$$

Let

$$S_m = \frac{d^A_{\text{Spear}}(\boldsymbol{W},\boldsymbol{Z}) - \mu_{\text{Spear}}(m)}{\sigma_{\text{Spear}}(\boldsymbol{m},\boldsymbol{n})} \text{ and } K_m = \frac{d^A_{\text{Ken}}(\boldsymbol{W},\boldsymbol{Z}) - \mu_{\text{Ken}}(m)}{\sigma_{\text{Ken}}(\boldsymbol{m},\boldsymbol{n})}. \tag{15.24}$$

Multiply both sides of (15.10) by $c_m^2/\sigma^2_{\text{Spear}}(\boldsymbol{m},\boldsymbol{n})$ to obtain

$$\mathbb{E}\left[\left(S_m - \frac{c_m\sigma_{\text{Ken}}(\boldsymbol{m},\boldsymbol{n})}{\sigma_{\text{Spear}}(\boldsymbol{m},\boldsymbol{n})}K_m\right)^2\right] \leqslant \frac{c_m^2 m(m-m_x)}{12\,\sigma^2_{\text{Spear}}(\boldsymbol{m},\boldsymbol{n})} \leqslant \frac{(m+1)^2m^2}{12\,\sigma^2_{\text{Spear}}(\boldsymbol{m},\boldsymbol{n})}. \tag{15.25}$$

By (13.59), we have $m^3 - \sum m_a^3 \geqslant m^2(m-m_x)$, so that

$$\frac{\sigma^2_{\text{Spear}}(\boldsymbol{m},\boldsymbol{n})}{m^4} \geqslant \frac{(m-m_x)(m-n_z)}{36(m-1)}, \tag{15.26}$$

hence

$$\frac{(m+1)^2m^2}{12\,\sigma^2_{\text{Spear}}(\boldsymbol{m},\boldsymbol{n})} \leqslant \frac{3(m+1)^2}{m^2}\frac{m-1}{(m-m_x)(m-n_z)} \longrightarrow 0, \tag{15.27}$$

by assumption (15.22). That is, the right-hand side of (15.25) goes to zero. By (15.23), the constant multiplying K_m, squared, is

$$\frac{c_m^2\sigma^2_{\text{Ken}}(\boldsymbol{m},\boldsymbol{n})}{\sigma^2_{\text{Spear}}(\boldsymbol{m},\boldsymbol{n})} = \frac{(m+1-\sqrt{m+1})^2}{m(m-2)} + \frac{O(m^4)}{\sigma^2_{\text{Spear}}(\boldsymbol{m},\boldsymbol{n})} \longrightarrow 1, \tag{15.28}$$

the last term going to zero by (15.22) and (15.26). Theorem 13.1 shows that $S_m \to N(0,1)$, thus we have that $K_m \to N(0,1)$ as well, proving (15.5).

Next turn to the case that $m_{<x} = q_1$ and $m_{>x} = q_2$ are fixed. We have from Theorem 13.2 that

$$\frac{1}{m^2}\left(d^A_{\text{Spear}}(\boldsymbol{W},\boldsymbol{Z}) - \mu_{\text{Spear}}(m)\right) + \frac{q}{2} \longrightarrow^{\mathcal{D}} V_1 + \ldots + V_{q_1} + (1-V_{q_1+1}) + \cdots + (1-V_q), \tag{15.29}$$

where $V_1, \ldots, V_q$ are iid with distribution F. Consider the inequality (15.10) again. Dividing both sides by m^2, since $q = m - m_x$, we have

$$\mathsf{E}\left[\left(\frac{m}{c_m}\frac{d^A_{\text{Spear}}(\boldsymbol{w},\boldsymbol{Z}) - \mu_{\text{Spear}}(m)}{m^2} - \frac{d^A_{\text{Ken}}(\boldsymbol{w},\boldsymbol{Z}) - \mu_{\text{Ken}}(m)}{m}\right)^2\right] \leqslant \frac{q}{12\,m}. \tag{15.30}$$

Letting $m \to \infty$, we have the bound in (15.30) goes to zero, and $m/c_m \to 1$. Thus by (15.29) we have (15.6). □

15.3 Edgeworth and simulation approximations

In Section 13.4, we compared the Edgeworth and simulation approximations of the distribution of Spearman's distance with ties. Here we do the same for Kendall's distance with ties in both variables. The results are fairly similar. Those here are based on randomly take 100 examples for each m from 13 to 50, restricting then to situations where the number of contingency tables ($\#\mathcal{T}$) corresponding to the particular $(\boldsymbol{m},\boldsymbol{n})$ pair is between 10^4 and 10^8. In general, finding the exact distribution using the contingency table approach is slower for Kendall than Spearman, owing to the extra time it takes to calculate the distance from the table. The algorithm for Kendall is reasonably quick if $\#\mathcal{T} \leqslant 5 \times 10^5$.

As for Spearman, the differences between the Edgeworth approximations with even L versus that with the corresponding $L+1$ is negligible, hence Figure 15.1 compares the even cases for $L = 0$ to 10. Here we see that comparing the maximum errors in the density, there is a slight jump from $L = 0$ to 2, but little benefit to higher L. For the distribution function, there is a large jump from $L = 0$ to 2, a slight jump to $L = 4$, but then little more improvement. For the relative p-value, there is substantial improvement from $L = 0$ to 2, then to 4, then to 6, and a little jump to 8. Thus one would probably choose $L = 4$ or 6, but that choice is a bit academic at this point, since we have analytical expressions (Section 15.4) only for cumulants up to the fourth degree, so that we can use Edgeworth for just L up to 2. The higher cumulants for the graphs were found using the exact distribution.

Figure 15.2 compares the Edgeworth approximation with $L = 2$ to the sumulations, where for each group of numbers of tables, we have the median as well as the 5^{th} and 95^{th} percentiles of the errors. It looks like Edgeworth is better than the the simulations for the error in the density $\#\mathcal{T} > 10^7$; in the distribution $\#\mathcal{T} > 10^6$; and in the relative p-value if $\#\mathcal{T} > 10^8$, though they are failry simlar for $10^7 \leqslant \#\mathcal{T} \leqslant 10^8$. Note that if could use larger L, then Edgeworth would be relative better for the p-value. Thus a reasonable rule of thumb is to use the exact algorithm if $\#\mathcal{T} \leqslant 5 \times 10^5$; use simulations if $5 \times 10^5 < \#\mathcal{T} \leqslant 10^7$; and Edgeworth otherwise.

15.4 Moments and cumulants

In the definition (12.35) of Kendall's distance in the presence of ties, we sum over pairs of indices with $j < i$. Here, it is more convenient to first subtract the mean ($\mu = m(m-1)/4$),

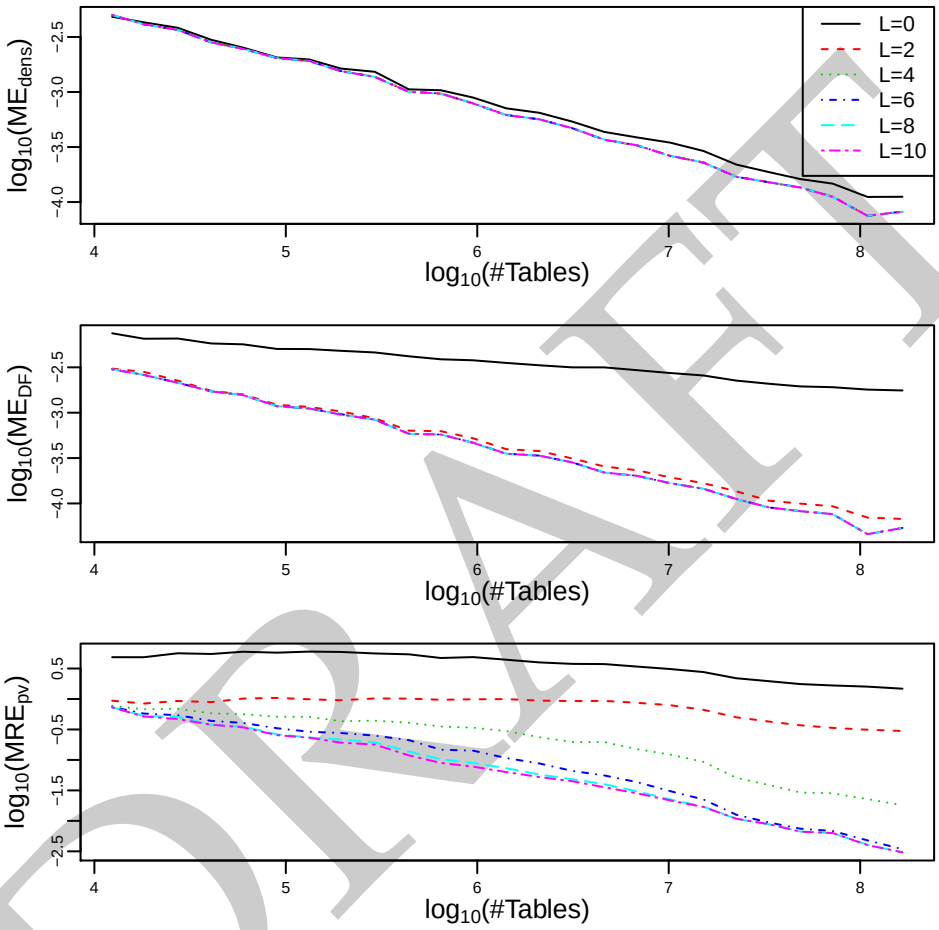

Figure 15.1: Comparing the median $\log_{10}$(errors) of the Edgeworth approximation to the distribution of the Kendall distance with ties for L=0, 2, 4, 6, 8, and 10. The horizontal axis is the $\log_{10}(\#\mathcal{T})$. The panel graphs the maximum error in the density, the middle graphs the maximum error in the distribution functions, and the bottom graphs the maximum relative error in the p-values. See (4.54) and (4.54).

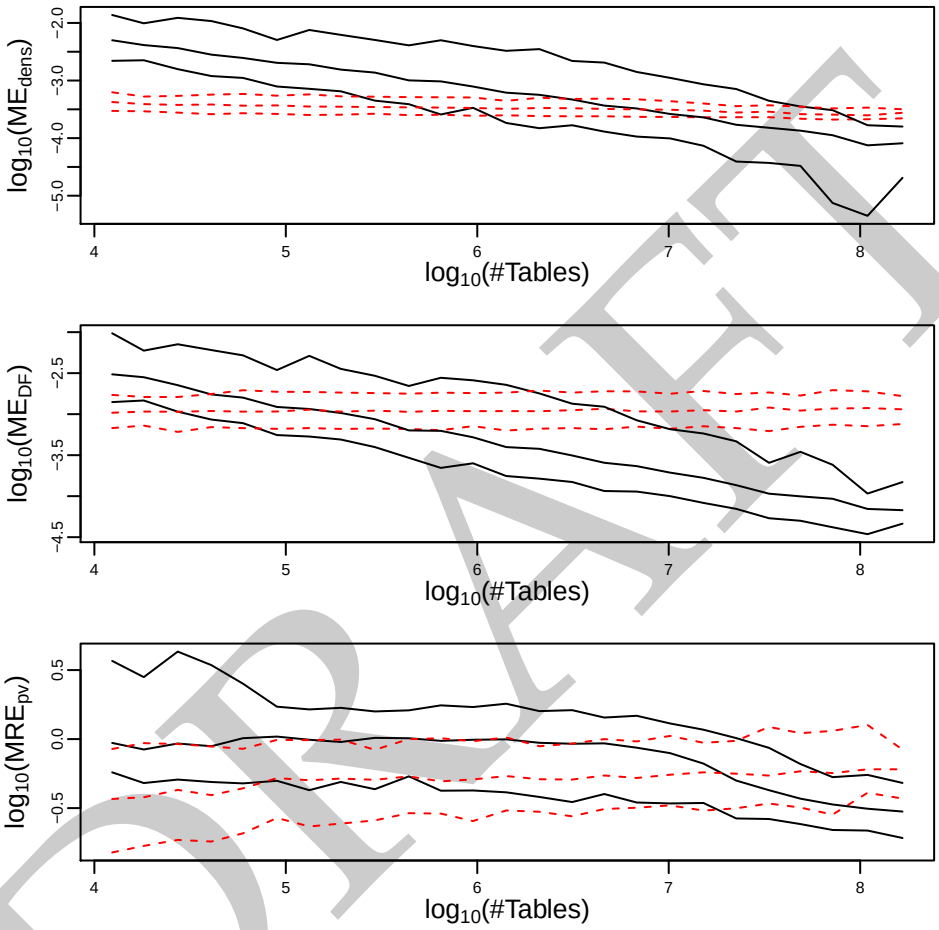

Figure 15.2: Comparing the $\log_{10}$(errors) of the Edgeworth approximation with $L = 2$ to estimation with 500,000 simulations. The horizontal axis is the $\log_{10}(\#\mathcal{T})$. The three panels graph the errors in the density, distribution function, and relative p-values, respectively. In each plot, the solid lines represent the $5^{\text{th}}, 50^{\text{th}}$, and 95^{th} percentiles for the Edgeworth approximations; the dotted lines represent the same percentiles for the simulations.

then multiply by -4, which we can then write as a function of signs:

$$\begin{aligned} U \equiv u(\boldsymbol{W}, \boldsymbol{Z}) &= -4\,(d^{A}_{Ken}(\boldsymbol{W}, \boldsymbol{Z}) - \mu) \\ &= \sum_{1 \leqslant i \neq j \leqslant m} \sum S_{ij} T_{ij}, \end{aligned} \tag{15.31}$$

where

$$S_{ij} = \mathrm{Sign}(W_i - W_j) \;\text{ and }\; T_{ij} = \mathrm{Sign}(Z_i - Z_j), \tag{15.32}$$

and $\mathrm{Sign}(a) = -1, 0$, or $+1$ as $a < 0, = 0$, or > 0, as below (1.13). The factor "4" arises since for each term in the summation of d_{Ken} we have

$$2(I[(w_i - w_j)(z_i - z_j) > 0] - 1) = -\mathrm{Sign}(w_i - w_j)\mathrm{Sign}(z_i - z_j), \tag{15.33}$$

and in (15.32) we are summing over $i \neq j$, which is double the sum over $j < i$. Thus for the N^{th} central moment we have

$$E[(d^{A}_{Ken}(\boldsymbol{W}, \boldsymbol{Z}) - \mu)^N] = (-4)^N\, E[U^N]. \tag{15.34}$$

To find the N^{th} moment of U, we write it out as a $2N$-degree sum:

$$E[U^N] = \sum_{1 \leqslant i_1 \neq j_1 \leqslant m} \sum \cdots \sum_{1 \leqslant i_N \neq j_N \leqslant m} \sum \mu_W(i_1 j_1, \ldots, i_N j_N) \mu_Z(i_1 j_1, \ldots, i_N j_N), \tag{15.35}$$

where

$$\mu_W(i_1 j_1, \ldots, i_N j_N) = E[S_{i_1 j_1} \cdots S_{i_N j_N}] \;\text{ and }\; \mu_Z(i_1 j_1, \ldots, i_N j_N) = E[T_{i_1 j_1} \cdots T_{i_N j_N}]. \tag{15.36}$$

The expectations in the summands depend on the equalities among the indices. We partition the summation according to the pattern of equalities, which will be defined in such a way that all sets of indices with a given pattern yield the same expected value for their summand. Note that the means μ_W and μ_Z are invariant under permutation of the pairs (i_k, j_k), i.e., the action $(i_1, j_1, \ldots, i_N, j_N) \to (i_{\pi_1}, j_{\pi_1}, \ldots, i_{\pi_N}, j_{\pi_N})$ for π a permutation of $1, \ldots, N$. Also, making the switch $(i_k, j_k) \to (j_k, i_k)$ just changes the sign for both means, hence their product remains invariant. Thus any pattern is defined to be invariant under those actions. For example, suppose $N = 4$ and consider the pattern $P_t = (12, 13, 24, 34)$, which means $i_1 = i_2$, $j_1 = i_3$, $j_2 = i_3$, and $j_3 = j_4$, and i_1, j_1, j_2, and j_3 are distinct. This pattern could also be represented as $(24, 13, 12, 34)$, or $(21, 13, 24, 34)$, or $(12, 14, 26, 46)$, etc.

Suppose $P_1, \ldots, P_T$ are the patterns. Then by the independence of $\boldsymbol{W}$ and $\boldsymbol{Z}$,

$$E[U^N] = \sum_{t=1}^{T} n(P_t) \mu_W(P_t) \mu_Z(P_t), \tag{15.37}$$

where $n(P_t)$ is the number of sets of indices with pattern P_t, and $\mu_Z(P_t)$ and $\mu_W(P_t)$ are as in (15.36) for any set of indices with pattern P_t , as long as it is the same representative for both

$\boldsymbol{W}$ and $\boldsymbol{Z}$. Let $d(P_t)$ be the number of distinct indices in pattern P_t. Then with $d = d(P_t)$, we can write

$$\begin{aligned}\mu_W(P_t) &= \frac{\sigma_W(P_t)}{(m)_d} \text{ where}\\ \sigma_W(P_t) &= \sum_{\substack{1\leqslant k_1,\ldots,k_d\leqslant m,\\ \text{distinct}}} \cdots \sum S_{P_t}(w_{k_1},\ldots,w_{k_d}),\end{aligned} \tag{15.38}$$

and $S_{P_t}(a_1,\ldots,a_d)$ is the value of the summand for the indices with equalities given by P_t. For example, consider $P_t = (12,13,24,34)$, so that $d(12,13,24,34) = 4$. Then we have

$$S_{(12,13,24,34)}(a_1,\ldots,a_4) = S(a_1-a_2)S(a_1-a_3)S(a_2-a_4)S(a_3-a_4). \tag{15.39}$$

See (4.18) for Pochhammer's symbol $(m)_k$. An equivalent expression to (15.37) that we find useful is then

$$E[U^N] = \sum_{t=1}^{T} \frac{n(P_t)}{(m)_d}\frac{\sigma_W(P_t)\sigma_Z(P_t)}{(m)_d}. \tag{15.40}$$

If $d(P_t) > m$, then $n(P_t) = 0$, in which case we use the convention that $n(P_t)/(m)_d = 0$.

The main steps for finding the N^{th} moment are to first find each pattern P_t of equalities and their $n(P_t)$'s, then to calculate the $\sigma_W(P_t)$ and $\sigma_Z(P_t)$. Some preliminaries, and details for the N^{th} moments, $N = 2$, 3, and 4, are in the following sections.

15.4.1 Some useful formulas

A pattern specifies equalities among the indices; we also deal with equalities among the values of the W_i's. Let the pattern of ties for the variable W be $\boldsymbol{m} = (m_1,\ldots,m_L)$, where each $m_i > 0$ and $m_1+\cdots+m_L = m$. For given P_t, we write the summand S_{P_t} as

$$S_{P_t}(W_1,\ldots,W_d) = S_{P_t}(W_1,\ldots,W_d)\prod_{1\leqslant i<j\leqslant d}(I_{ij}+\bar{I}_{ij}), \tag{15.41}$$

where we let

$$I_{ij} = I[W_i = W_j] \text{ and } \bar{I}_{ij} = I[W_i \neq W_j]. \tag{15.42}$$

By multiplying out the indicator functions, the right-hand side of (15.41) can be written as a sum of 2^d terms. Many of these are zero, either because the equalities in the W_i's yield $S_{P_t}(W_1,\ldots,W_d) = 0$, or the combination of indicator functions is impossible, e.g., $I_{12}I_{23}\bar{I}_{13} \equiv 0$. Let $C(W_1,\ldots,W_d)$ be a set of conditions on the W_i's, and for function g define

$$\sigma_d\{g(W_1,\ldots,W_d)\,|\,C(W_1,\ldots,W_d)\} = \sum_{\substack{1\leqslant i_1,\ldots,i_d\leqslant m\\ i_1,\ldots,i_d \text{ distinct}\\ C(w_{i_1},\ldots,w_{i_d}) \text{ holds}}} \cdots \sum g(w_{i_1},\ldots,w_{i_d}). \tag{15.43}$$

Then (15.41) is a sum of such σ_d's with $g = S_{P_t}$. These conditions end up equating certain of the W_i's, and otherwise specifying their values are distinct. For each such term, we further decompose the sum into parts based on the order of the distinct values of W_i.

For example, suppose $d = 6$ and the equality/inequality conditions yield the conditions $W_1 = W_2 = W_4$, $W_3 = W_6$, and W_1, W_3, W_5 are distinct. Then we can write

$$\begin{aligned}
&\sigma_6\{S_{P_t}(W_1,\dots,W_6) \,|\, W_1 = W_2 = W_4, W_3 = W_6, \ \&\ W_1, W_3, W_5 \text{ distinct}\} \\
&\quad = \sum_{\pi\in\mathcal{P}_3} \sigma_6\{S_{P_t}(W_1,\dots,W_6) \,|\, W_1 = W_2 = W_4, W_3 = W_6, \ \&\ \operatorname{rank}(W_1, W_3, W_5) = \pi\} \\
&\quad = \sum_{\pi\in\mathcal{P}_3} S_{P_t}(\pi_1,\pi_1,\pi_2,\pi_1,\pi_3,\pi_2)\sigma_6\{1 \,|\, W_1 = W_2 = W_4, W_3 = W_6, \ \&\ \operatorname{rank}(W_1, W_3, W_5) = \pi\},
\end{aligned} \tag{15.44}$$

the final equality following since the functions S_{P_t} are invariant under strict monotone functions of the W_i's. This last representation is useful because the final σ_6 function depends only on the multiplicities of the W_i's and their order. We will use the shorthand "$W_{i_1}^{(n_1)}, \dots, W_{i_q}^{(n_q)}$" to mean that W_{i_1} is equal to n_1-1 other W_i's, W_{i_2} is equal to n_2-1 other W_i's, etc. Thus the condition in the final summation of (15.44) would be written $\{\operatorname{rank}(W_1^{(3)}, W_3^{(2)}, W_5^{(1)}) = \pi\}$. More generally, we have

$$\begin{aligned}
&\sigma_d\{S_{P_t}(W_1,\dots,W_d) \,|\, W_1^{(n_1)},\dots,W_q^{(n_q)}, \text{ d.}\} \\
&\qquad = \sum_{\pi\in\mathcal{P}_q} S_{P_t}(\pi_1,\dots,\pi_q,\dots)\sigma_d\{1 \,|\, \operatorname{rank}(W_1^{(n_1)},\dots,W_q^{(n_q)}) = \pi\},
\end{aligned} \tag{15.45}$$

where $d = n_1 + \cdots + n_q$, and the second "..." in the argument for S_{P_t} consists of replacing the W_i, $i > q$, with the π_j such that $W_i = W_j$, $1 \leqslant j \leqslant q$.

Now define

$$\rho_{n_1,n_2,\dots,n_q} = \sigma_d\{1 \,|\, W_1^{(n_1)} < W_2^{(n_2)} < \cdots < W_q^{(n_q)}\}. \tag{15.46}$$

Note that the value is invariant under permutation of the W_i's, so that it doesn't matter which of them are equal to each other, just the resulting n_j's. Summing over permutations of the n_j, we obtain the number of ways for the $W_1, \dots, W_q$ to be distinct, which we denote by ϵ:

$$\begin{aligned}
\epsilon_{n_1,n_2,\dots,n_q} &\equiv \sigma_d\{1 \,|\, W_1^{(n_1)}, W_2^{(n_2)},\dots,W_q^{(n_q)} \text{ d.}\} \ \text{(“d.” means “distinct”)} \\
&= \sum_{\pi\in\mathcal{P}_q} \rho_{n_{\pi_1},n_{\pi_2},\dots,n_{\pi_q}}.
\end{aligned} \tag{15.47}$$

Note in particular that if there are q subscripts, all equal to r, since the ρ's are then all the same, we have

$$\epsilon_{rr\cdots r} = q!\rho_{rr\cdots r}. \tag{15.48}$$

Thus from (15.45),

$$\begin{aligned}
\sigma_d\{S_{P_t}(W_1,\dots,W_d) \,|\, W_1^{(r)},\dots,W_q^{(r)}, \text{ distinct}\} &= \sum_{\pi\in\mathcal{P}_q} S_{P_t}(\pi_1,\dots,\pi_q,\dots)\rho_{rr\cdots r} \\
&= \epsilon_{rr\cdots r}\frac{1}{q!}\sum_{\pi\in\mathcal{P}_q} S_{P_t}(\pi_1,\dots,\pi_q,\dots).
\end{aligned} \tag{15.49}$$

If $q-1$ of the multiplicities are equal, e.g., $(r_1,\ldots,r_q)=(r,s,\ldots,s)$, with $q-1$ s's, then we can collect terms depending on the value of π_1:

$$\sigma_d\{S_{P_t}(W_1,\ldots,W_d)\,|\,W_1^{(r)},W_2^{(s)},\ldots,W_q^{(s)},\ \text{distinct}\}=\sum_{k=1}^{q}\rho_{s\cdots srs\cdots s}\sum_{\substack{\pi\in\mathcal{P}_q\\ \pi_1=k}}S_{P_t}(\pi_1,\ldots,\pi_q,\ldots), \tag{15.50}$$

where in the subscript of ρ, the r is in the k^{th} slot.

A special case that is easy to deal with is exemplified by the summand $S_{12}S_{13}S_{45}$. If the W_4 and W_5 are equal to the same number of other W_i's, then by symmetry the sum will be zero since $S_{ab}=-S_{ba}$. Note that the key is that W_4 and W_5 appear only in that one S_{ij}. That is,

$$\sigma_d\{S_{i_1j_1}\cdots S_{i_Nj_N}\,|\,C(W_1,\ldots,W_d))=0 \tag{15.51}$$

if for some k

$$\{i_k,j_k\}\cap(\{i_1,i_2,\ldots,i_N,j_N\}-\{i_k,j_k\})=\emptyset, \tag{15.52}$$

and only requirement the condition $C(W_1,\ldots,W_d)$ places on W_{i_k} and W_{j_k} is that they have the same multiplicity.

Another formula we find useful is for summands of the form $S_{12}S_{13}I[W_2\neq W_3]$ with multiplicities r_1,r_2,r_3, basing it on (15.45). The terms in the sum are based on the permutations $\pi\in\mathcal{P}_3$, where the condition translates the ranking as an ordering:

π	$S(\pi_1-\pi_2)S(\pi_1-\pi_3)$	Condition
$(1,2,3)$	$+1$	$W_1^{(r_1)}<W_2^{(r_2)}<W_3^{(r_3)}$
$(1,3,2)$	$+1$	$W_1^{(r_1)}<W_3^{(r_3)}<W_2^{(r_2)}$
$(2,1,3)$	-1	$W_2^{(r_2)}<W_1^{(r_1)}<W_3^{(r_3)}$
$(2,3,1)$	-1	$W_3^{(r_3)}<W_1^{(r_1)}<W_2^{(r_2)}$
$(3,1,2)$	$+1$	$W_2^{(r_2)}<W_3^{(r_3)}<W_1^{(r_1)}$
$(3,2,1)$	$+1$	$W_3^{(r_3)}<W_2^{(r_2)}<W_1^{(r_1)}$

(15.53)

Thus by (15.45) and (15.46),

$$\sigma_d\{S_{12}S_{13}\,|\,W_1^{(r_1)},W_2^{(r_2)},W_3^{(r_3)},\ d.\}=\rho_{r_1r_2r_3}+\rho_{r_1r_3r_2}+\rho_{r_2r_3r_1}+\rho_{r_3r_2r_1}-\rho_{r_3r_1r_2}-\rho_{r_2r_1r_3}. \tag{15.54}$$

A special case is when two of the multiplicities are equal, in which case the sum depends on whether the single multiplicity is in the first slot or not. That is,

$$\begin{aligned}\sigma_d\{S_{12}S_{13}\,|\,W_1^{(r)},W_2^{(s)},W_3^{(s)},\ d.\}&=2(\rho_{rss}-\rho_{srs}+\rho_{ssr});\\ \sigma_d\{S_{12}S_{13}\,|\,W_1^{(s)},W_2^{(r)},W_3^{(s)},\ d.\}&=2\rho_{srs}.\end{aligned} \tag{15.55}$$

Consider the extension to four distinct variable, where the fourth is uninvolved in the $S_{12}S_{13}$ part. Then if we fix the rank of the fourth variable at k, the summation is the same as in

(15.54), but with the fourth multiplicity inserted between the $(k-1)^{st}$ and $(k+1)^{st}$ subscript. Thus

$$\begin{aligned}\sigma_d\{S_{12}S_{13}\,|\,W_1^{(r_1)},W_2^{(r_2)},W_3^{(r_3)},W_4^{(r_4)},\ d.\} &= (\rho_{r_4r_1r_2r_3}+\rho_{r_1r_4r_2r_3}+\rho_{r_1r_2r_4r_3}+\rho_{r_1r_2r_3r_4})\\ &+(\rho_{r_4r_1r_3r_2}+\rho_{r_1r_4r_3r_2}+\rho_{r_1r_3r_4r_2}+\rho_{r_1r_3r_2r_4})+(\rho_{r_4r_2r_3r_1}+\rho_{r_2r_4r_3r_1}+\rho_{r_2r_3r_4r_1}+\rho_{r_2r_3r_1r_4})\\ &+(\rho_{r_4r_3r_2r_1}+\rho_{r_3r_4r_2r_1}+\rho_{r_3r_2r_4r_1}+\rho_{r_3r_2r_1r_4})-(\rho_{r_4r_3r_1r_2}+\rho_{r_3r_4r_1r_2}+\rho_{r_3r_1r_4r_2}+\rho_{r_3r_1r_2r_4})\\ &-(\rho_{r_4r_2r_1r_3}+\rho_{r_2r_4r_1r_3}+\rho_{r_2r_1r_4r_3}+\rho_{r_2r_1r_3r_4}).\end{aligned}\tag{15.56}$$

Paralleling (15.55), we have the case with three multiplicities equal, which has three possibilities:

$$\begin{aligned}\sigma_d\{S_{12}S_{13}\,|\,W_1^{(r)},W_2^{(s)},W_3^{(s)},W_4^{(s)},\ d.\} &= 6\rho_{rsss}-2\rho_{srss}-2\rho_{ssrs}+6\rho_{sssr};\\ \sigma_d\{S_{12}S_{13}\,|\,W_1^{(s)},W_2^{(r)},W_3^{(s)},W_4^{(s)},\ d.\} &= 4(\rho_{srss}+\rho_{ssrs});\\ \sigma_d\{S_{12}S_{13}\,|\,W_1^{(s)},W_2^{(s)},W_3^{(s)},W_4^{(r)},\ d.\} &= 6(\rho_{rsss}+\rho_{srss}+\rho_{ssrs}+\rho_{sssr})=\epsilon_{rsss}.\end{aligned}\tag{15.57}$$

Most of the $\sigma_d\{S_{P_t}\}$'s are functions just of a number of such ϵ's, though a few need specific ρ's as well. We now derive some specific formulas we'll need for ρ and ϵ. We first define some basic functions we need. Starting with $\boldsymbol{m}=(m_1,\ldots,m_K)$, the pattern of ties for $\boldsymbol{W}$, we define for positive integer r,

$$\boldsymbol{m}_{(r)}=((m_1)_r,\ldots,(m_K)_r).\tag{15.58}$$

We need the sum and sum of products of these, leading us to define

$$\Sigma_r=\sum_{a=1}^{K}(m_a)_r\ \text{ and, more generally, }\ \Sigma_{r_1\cdots r_s}=\sum_{a=1}^{K}(m_a)_{r_1}\cdots(m_a)_{r_s}.\tag{15.59}$$

Note that $\Sigma_1=\sum m_a(=m)$, $\Sigma_{11}=\sum m_a^2$, etc. Partial sums we denote as follows:

$$m_{<a}=\sum_{i=1}^{a-1}m_i,\ m_{<a(r)}=\sum_{i=1}^{a-1}(m_i)_r,\ m_{>a}=\sum_{i=a+1}^{m}m_i,\ \text{and}\ m_{>a(r)}=\sum_{i=a+1}^{m}(m_i)_r.\tag{15.60}$$

Consider $\rho_{rs}=\sigma_{r+s}\{1\,|\,W_1^{(r)}<W_2^{(s)}\}$. Thus we have $r+s$ W_i's, r of which are equal to W_1 and s of which are equal to W_2. Fix $a<b$ with $W_1=a$ and $W_2=b$. We need to choose, without replacement, r of the $w_i=a$'s, of which there are m_a, to assign to the W_i's equalling W_1. Thus there are $(m_a)_r$ possibilities. Likewise, there are $(m_b)_s$ ways to choose the w_i that equal b. Multiplying those counts, and summing over $a<b$, we have that

$$\begin{aligned}\rho_{rs}&=\sum_{a=1}^{K-1}\sum_{b=a+1}^{K}(m_a)_r(m_b)_s\\ &=\sum_{a=1}^{K-1}(m_a)_r\sum_{b=a+1}^{K}(m_b)_s\\ &=\sum_{a=1}^{K-1}(m_a)_r m_{>a(s)}.\end{aligned}\tag{15.61}$$

Note that $\rho_{rs} = \sum_{a=2}^{K} m_{<a(r)}(m_a)_s$ would work, too. With three subscripts, we set $b < a < c$:

$$\begin{aligned}\rho_{rst} &= \sum_{a=2}^{K-1}\sum_{b=1}^{a-1}\sum_{c=a+1}^{K}(m_b)_r(m_a)_s(m_c)_t \\ &= \sum_{a=2}^{K-1} m_{<a(r)}(m_a)_s m_{>a(t)}. \end{aligned} \tag{15.62}$$

For the general case, we have

$$\rho_{r_1\cdots r_d} = \sum_{1\leqslant a_1<a_2<\cdots<a_d\leqslant K}\cdots\sum (m_{a_1})_{r_1}(m_{a_2})_{r_2}\cdots(m_{a_d})_{r_d}. \tag{15.63}$$

In the special case that the r_i's are equal to r, the value is an elementary symmetric polynomial of degree d in the elements of $\boldsymbol{m}_{(r)}$. See Wikipedia contributors (2019b) for Newton's identities, which find the $\rho_{rr\cdots r}$ in terms of the $\Sigma_{r\cdots r}$ (where there are d "r" in the subscript).

Turn to ϵ's in (15.47). For the general case, we build up from $d = 2$ subscripts, finding the ϵ's as functions of the $\Sigma_{r_1\cdots r_d}$ of (15.59). The formula for ϵ_{rs} is like that for ρ_{rs} in (15.61), but with an inequality. Using (15.59), we find

$$\begin{aligned}\epsilon_{rs} &= \sum_{a=1}^{K}(m_a)_r\sum_{b\neq a}(m_b)_s \\ &= \sum_{a=1}^{K}(m_a)_r(\Sigma_r - (m_a)_s) \\ &= \sum_{a=1}^{K}(m_a)_r\Sigma_r - \sum_{a=1}^{K}(m_a)_r(m_a)_s \\ &= \Sigma_r\Sigma_s - \Sigma_{rs}. \end{aligned} \tag{15.64}$$

Now

$$\epsilon_{rst} = \sum_{a=1}^{K}(m_a)_r\sum_{b\neq a}(m_b)_s\sum_{c\neq a,b}(m_c)_t. \tag{15.65}$$

For fixed a, the final two summations equal ϵ_{st}, but using $\boldsymbol{m}$ without the a^{th} component, which is as in (15.64) with the Σ's leaving out m_a. Thus,

$$\begin{aligned}\epsilon_{rst} &= \sum_{a=1}^{K}(m_a)_r\left((\Sigma_s - (m_a)_s)(\Sigma_t - (m_a)_t) - (\Sigma_{st} - (m_a)_s(m_a)_t)\right) \\ &= \sum_{a=1}^{K}(m_a)_r\left(\Sigma_s\Sigma_t - (m_a)_s\Sigma_t - \Sigma_s(m_a)_t + (m_a)_s(m_a)_t - \Sigma_{st} + (m_a)_s(m_a)_t\right) \\ &= \Sigma_r\Sigma_s\Sigma_t - \Sigma_{rs}\Sigma_t - \Sigma_{rt}\Sigma_s - \Sigma_{st}\Sigma_r + 2\Sigma_{rst}. \end{aligned} \tag{15.66}$$

A similar approach will show that

$$\begin{aligned}\epsilon_{rstu} = {} & \Sigma_r\Sigma_s\Sigma_t\Sigma_u - \Sigma_r\Sigma_s\Sigma_{tu} - \Sigma_r\Sigma_t\Sigma_{su} - \Sigma_r\Sigma_u\Sigma_{st} - \Sigma_s\Sigma_t\Sigma_{ru} - \Sigma_s\Sigma_u\Sigma_{rt} - \Sigma_t\Sigma_u\Sigma_{rs} \\ & + 2\Sigma_r\Sigma_{stu} + 2\Sigma_s\Sigma_{rtu} + 2\Sigma_t\Sigma_{rsu} + 2\Sigma_u\Sigma_{rst} + \Sigma_{rs}\Sigma_{tu} + \Sigma_{rt}\Sigma_{su} + \Sigma_{ru}\Sigma_{st} - 6\Sigma_{rstu}. \quad (15.67)\end{aligned}$$

For an arbitrary number d of subscripts, we wish to find $\epsilon_{r_1\cdots r_d}$. Note that in the above formulas, there is a term for each partition of the set of subscripts for the ϵ. The sign in front of the term depends on the number of subsets in the partition, and the magnitude of its coefficient depends on the numbers of elements in the subsets. Specifically, letting $(A_1, \ldots, A_J)$ be a partition of the subscripts, and for set $A = \{s_1, \ldots, s_k\}$, defining $\Sigma_A = \Sigma_{s_1\cdots s_k}$, we have

$$\epsilon_{r_1\cdots r_d} = \sum_{\text{Partitions } (A_1,\ldots,A_J)} (-1)^{d-J}(\#A_1 - 1)! \cdots (\#A_J - 1)! \Sigma_{A_1} \cdots \Sigma_{A_J}. \quad (15.68)$$

For $d = 5$, we need only ϵ_{11112} and ϵ_{11111}, which are, respectively,

$$\begin{aligned}\epsilon_{11112} = {} & m^4\Sigma_2 - 4m^3\Sigma_{12} - 6m^2\Sigma_{11}\Sigma_2 + 12m^2\Sigma_{112} + 12m\Sigma_{11}\Sigma_{12} + 3\Sigma_{11}^2\Sigma_2 + 8m\Sigma_{111}\Sigma_2 \\ & - 24m\Sigma_{1112} - 12\Sigma_{11}\Sigma_{112} - 8\Sigma_{111}\Sigma_{12} - 6\Sigma_{1111}\Sigma_2 + 24\Sigma_{11112}, \quad (15.69)\end{aligned}$$

and

$$\epsilon_{11111} = m^5 - 10m^3\Sigma_{11} + 20m^2\Sigma_{111} + 15m\Sigma_{11}^2 - 20\Sigma_{11}\Sigma_{111} - 30m\Sigma_{1111} + 24\Sigma_{11111}. \quad (15.70)$$

(Recall $m = \Sigma_1$.) The only representative we need for $d = 6$ is that with all ones:

$$\begin{aligned}\epsilon_{111111} = {} & m^6 - 15m^4\Sigma_{11} + 40m^3\Sigma_{111} + 45m^2\Sigma_{11}^2 - 90m^2\Sigma_{1111} - 120m\Sigma_{11}\Sigma_{111} \\ & - 15\Sigma_{11}^3 + 144m\Sigma_{11111} + 90\Sigma_{11}\Sigma_{1111} + 40\Sigma_{111}^2 - 120\Sigma_{111111} \quad (15.71)\end{aligned}$$

15.4.2 The variance

For the variance, we deal with pairs of pairs of indices: ij, kl, where $i \neq j$ and $k \neq l$. There are three patterns, depending on the number of distinct indices in the two pairs. The patterns and their $n(P_t)/(m)_d$ and σ_W's are given next:

d = #distinct indices	P_t	$n(P_t)/(m)_d$	σ_W
2	12, 12	2	ϵ_{11}
3	12, 13	4	$\epsilon_{12} + \frac{1}{3}\epsilon_{111}$
4	12, 34	1	0

(15.72)

The pattern for two distinct indices is $(12, 12)$, which includes both (ij, ij) and (ij, ji). There are $(m)_2$ ways to choose the i and j, hence $n(12, 12) = 2(m)_2$. The summand is $S_{12}^2 = I[W_1 \neq W_2]$, hence we need the sum $\sigma_2\{1 \mid W_1^{(1)} \neq W_2^{(1)}\}$, which by (15.47) is ϵ_{11}.

Three distinct indices yields the pattern $(12, 13)$. There are four ways to arrange three indices, depending on where the two equal ones are: $(ij, ik), (ij, ki), (ji, ik), (ji, ki)$. Then there are $(m)_3$ ways to choose the i, j, k. To find the summation, we use (15.41). The summand is

$S_{12}S_{13} = S(W_1 - W_2)S(W_1 - W_3)$, hence is zero if $W_1 = W_2$ or W_3. Thus we need only deal with whether or not $W_2 = W_3$:

$$\begin{aligned} S_{12}S_{13}(I_{23} + \bar{I}_{23}) &= S_{12}^2 + S_{12}S_{13}\bar{I}_{23} \\ &= \bar{I}_{12} + S_{12}S_{13}\bar{I}_{23}. \end{aligned} \tag{15.73}$$

Since $W_2 = W_3$, the first term yields the sum

$$\sigma_3\{1 \,|\, W_1^{(1)} \neq W_2^{(2)}\} = \epsilon_{12}, \tag{15.74}$$

again by (15.47). For the second term, we can use (15.49) because the three multiplicities (for W_1, W_2, W_3) are one. It is not hard to see that

$$\sum_{\pi \in \mathcal{P}_3} S(\pi_1 - \pi_2)S(\pi_1 - \pi_3) = 2, \tag{15.75}$$

because the summand is $+1$ if the rank of W_1 is 1 or 3, and -1 otherwise. Thus, since $d! = 6$,

$$\sigma_3\{S_{12}S_{13} \,|\, W_1, W_2, W_3 \text{ d.}\} = \frac{1}{3}\epsilon_{111}. \tag{15.76}$$

With no equalities, i.e., $d = 4$, we have the pattern $(12, 34)$, and all multiplicities of one, so that by (15.38), $\sigma_W(12, 34) = 0$.

Thus from (15.40)

$$E[U^2] = 4\,\frac{\sigma_W(12,13)\sigma_Z(12,13)}{(m)_3} + 2\,\frac{\sigma_W(12,12)\sigma_Z(12,12)}{(m)_2}, \tag{15.77}$$

and by (15.37),

$$Var[d^A_{Ken}(\boldsymbol{W}, \boldsymbol{Z})] = \frac{1}{16}\,E[U^2]. \tag{15.78}$$

There are many equivalent ways to express this variance. Using (15.64) and (15.66), for $\boldsymbol{W}$, since $\Sigma_1 = m$, we have

$$\begin{aligned} \epsilon_{11} &= m^2 - \sum m_a^2 = (m)_2 - \sum (m_a)_2, \\ \epsilon_{12} &= m \sum (m_a)_2 - \sum m_a (m_a)_2, \text{ and} \\ \epsilon_{111} &= m^3 - 3m\Sigma_{11} + \Sigma_{111} = m^3 - 3m \sum m_a^2 + 2 \sum m_a^3. \end{aligned} \tag{15.79}$$

Then some further manipulation shows that

$$\sigma_W(12, 13) = \epsilon_{12} + \frac{1}{3}\epsilon_{111} = \frac{1}{3}((m)_3 - \sum (m_a)_3). \tag{15.80}$$

Thus

$$\begin{aligned} Var[d^A_{Ken}(\boldsymbol{W}, \boldsymbol{Z})] = &\frac{((m)_3 - \sum (m_a)_3)((m)_3 - \sum (n_a)_3)}{36(m)_3} \\ &+ \frac{((m)_2 - \sum (m_a)_2)((m)_2 - \sum (n_a)_2)}{8(m)_2}. \end{aligned} \tag{15.81}$$

It's convenient to have the formulas for the MWW statistic, i.e., $K = 2$. Then

$$(m)_2 - \sum (m_a)_2 = 2m_1m_2 \text{ and } (m)_3 - \sum (m_a)^3 = 3m_1m_2(m-2). \tag{15.82}$$

The higher moments are found similarly, but the formulas become more complicated.

15.4.3 The third moment

Here we deal with triples of indices, (ij, kl, pq). Table (15.83) displays the patterns P_t, their counts $n(P_t)$, and their σ_W's:

d	P_t	$n(P_t)/(m)_d$	σ_W
2	12, 12, 12	4	0
3	12, 12, 13	24	$\rho_{21} - \rho_{12}$
	12, 13, 23	8	0
4	12, 12, 34	6	0
	12, 13, 14	8	$\rho_{31} - \rho_{13}$
	12, 14, 34	24	0
5	12, 13, 45	12	0
6	12, 34, 56	1	0

(15.83)

Since the $n(P_t)/(m)_d$ does not depend on m, as long as $m \geqslant d$, we found the values here for $m = 6$ by using an algorithm to run over all (ij, kl, pq)'s, and counting the number which conformed to each pattern. One check of these counts is to note that the total number of triples is $(m(m-1))^3$, to which the third column times $(m)_d$ should (and does) sum.

We give the details for (15.83). First, since S_{ij} is either 0 or ± 1, $S_{ij}^3 = S_{ij}$, hence $\sigma_W(12, 12, 12)$ is zero by (15.51). That equation also shows the sum is zero for patterns $(12, 12, 34)$, $(12, 13, 45)$, and $(12, 34, 56)$.

For $(12, 12, 13)$, the summand is $\bar{I}_{12} S_{13}$. We proceed similar to (15.73), finding

$$\sigma_3\{\bar{I}_{12} S_{13}\} = \sigma_3\{S_{12} \,|\, W_1^{(1)}, W_2^{(2)}\} + \sigma_3\{S_{13} \,|\, W_1, W_2, W_3 \text{ d.}\}. \tag{15.84}$$

The second term is zero, again by (15.51). We use (15.45) for the first term, where there are just the two permutations $(1, 2)$ and $(2, 1)$. S_{12} is -1 if $W_1 < W_2$ and $+1$ if $W_1 > W_2$, thus with multiplicities 1 and 2 for W_1 and W_2, respectively, we obtain

$$\sigma_3\{S_{12} \,|\, W_1^{(1)}, W_2^{(2)}\} = -\rho_{12} + \rho_{21}. \tag{15.85}$$

For the pattern $(12, 13, 23)$, we need to sum over just the indices with W_1, W_2, and W_3 distinct. For each of the six permutations, it is easy to find $S_{12} S_{13} S_{23}$, of which there are three permutations that yield $+1$ and three that yield -1. Since the multiplicities are all one, by (15.49) the summation is zero.

The pattern $(12, 13, 14)$ has summand zero if W_1 is equal to any of the other three variables. Thus to utilize (15.41) we need to consider the three equalities among W_2, W_3, W_4. If the three are distinct, then we can use (15.49) again. The sum over the permutations turns out to be zero, and the multiplicities are all equal to one, hence the sum is zero. Next suppose $W_2 = W_3 \neq W_4$. Then we have

$$\sigma_4\{S_{14} \,|\, W_1^{(1)}, W_2^{(2)}, W_4^{(1)}, \text{ d.}\} = 0 \tag{15.86}$$

by (15.51). By symmetry, the value is also zero for $W_2 = W_4 \neq W_3$ and $W_3 = W_4 \neq W_2$. The other possibility is $W_2 = W_3 = W_4$, in which case we have

$$\sigma_4\{S_{12} \,|\, W_1^{(1)}, W_2^{(3)}\} = \rho_{31} - \rho_{13}, \tag{15.87}$$

d	Pattern P_t	$n(P_t)/(m)_d$	σ_W
2	12, 12, 12, 12	8	ϵ_{11}
3	12, 12, 12, 13	64	$\epsilon_{12} + \frac{1}{3}\epsilon_{111}$
	12, 12, 13, 13	48	$\epsilon_{12} + \epsilon_{111}$
	12, 13, 23, 23	96	$\frac{1}{3}\epsilon_{111}$
4	12, 12, 13, 14	96	$\epsilon_{13} + 4\rho_{121} + \epsilon_{112} + \frac{1}{3}\epsilon_{1111}$
	12, 13, 23, 34	192	0
	12, 12, 12, 34	16	0
	12, 12, 13, 24	96	$-\epsilon_{22} - 2(\rho_{112} - \rho_{121} + \rho_{211})$
	12, 12, 13, 34	192	$-\epsilon_{22} - \epsilon_{112} - 2(\rho_{112} + \rho_{211}) - \frac{1}{3}\epsilon_{1111}$
	12, 12, 34, 34	12	$2\epsilon_{22} + 4\epsilon_{112} + \epsilon_{1111}$
	12, 13, 24, 34	48	$2\,\epsilon_{112} + \epsilon_{22} + \frac{1}{3}\epsilon_{1111}$
5	12, 13, 14, 15	16	$\epsilon_{14} + 8\rho_{131} + 3\epsilon_{122} + 2\epsilon_{1112} + \frac{1}{5}\epsilon_{11111}$
	12, 12, 13, 45	96	0
	12, 13, 23, 45	32	0
	12, 13, 24, 25	192	$\epsilon_{23} + \epsilon_{122} + 2(\rho_{311} + \rho_{113}) + \frac{2}{3}\epsilon_{1112} + \frac{1}{15}\epsilon_{11111}$
	12, 12, 34, 35	48	$2\epsilon_{23} + 2\epsilon_{113} + 4\epsilon_{122} + 3\epsilon_{e1112} + \frac{1}{3}\epsilon_{11111}$
	12, 13, 24, 35	192	$\epsilon_{23} + \epsilon_{113} + 2\epsilon_{122} + \frac{4}{3}\epsilon_{1112} + \frac{2}{15}\epsilon_{11111}$
6	12, 12, 34, 56	12	0
	12, 13, 24, 56	96	0
	12, 13, 14, 56	32	0
	12, 13, 45, 46	48	$\epsilon_{24} + \epsilon_{114} + \epsilon_{33} + 6\epsilon_{123} + \frac{5}{3}\epsilon_{222} + 2\epsilon_{1113}$ $+5\epsilon_{1122} + \frac{5}{3}\epsilon_{11112} + \frac{1}{9}\epsilon_{111111}$
7	12, 13, 45, 67	24	0
8	12, 34, 56, 78	1	0

Table 15.1: The patterns for the fourth moment, and their corresponding sums.

using the same reasoning as for (15.85).

Finally, turn to the pattern $(12, 14, 34)$. If we make the switches $1 \leftrightarrow 4$ and $2 \leftrightarrow 3$, we have $S_{43}S_{41}S_{21} = -S_{12}S_{14}S_{34}$. Since the W_i's are interchangeable, both summands $S_{43}S_{41}S_{21}$ and $S_{12}S_{14}S_{34}$ have the same sum, which thus must be zero.

As in (15.40) we have we have

$$E[U^3] = 24\,\frac{\sigma_W(12,12,13)\sigma_Z(12,12,13)}{(m)_3} + 8\,\frac{\sigma_W(12,13,14)\sigma_Z(12,13,14)}{(m)_4}, \tag{15.88}$$

and by (15.37),

$$E[(d^A_{Ken}(\boldsymbol{W}, \boldsymbol{Z}) - \mu)^3] = -\frac{1}{64}\,E[U^3]. \tag{15.89}$$

15.4.4 The fourth moment

For the fourth moment, we have to deal with four pairs of indices. Table (15.1) contains the results. The patterns $(12,12,12,12)$ and $(12,12,12,13)$ have the same σ_W's as the patterns $(12,12)$ and $(12,13)$ in (15.72), respectively, since $S_{ij}^3 = S_{ij}$. The patterns with $\sigma_W(P_t) = 0$, except for $(12,13,23,34)$, have at least one of their pairs of indices with neither index in common with any of the other pairs, hence have summation zero by (15.51).

We treat the other patterns individually.

- $(12,12,13,13)$. Here we consider whether $W_2 = W_3$ or not. If so, then the summand is $\bar{I}_{12}$, and if not, it is that all three W_i's are distinct. Thus we can write

$$\begin{aligned}\sigma_W(12,12,13,13) &= \sigma_3\{1 \,|\, W_1^{(1)} \neq W_2^{(2)}\} + \sigma_3\{1 \,|\, W_1, W_2, W_3, \text{ distinct}\} \\ &= \epsilon_{12} + \epsilon_{111},\end{aligned} \tag{15.90}$$

  by the definition in (15.47).

- $(12,13,23,23)$. Here the summand is zero unless the three W_i are distinct. Thus we have the same sum as in (15.76), i.e., $\epsilon_{111}/3$.

- $(12,12,13,14)$. We deal with equalities/inequalities among W_2, W_3, W_4. If $W_2 = W_3 = W_4$, the summand is $\bar{I}_{12}$, and since the multiplicities are 1 and 3, the sum is ϵ_{13}. If $W_2 = W_3 \neq W_4$, we have the summand $S_{12}S_{14}$ with condition $W_1^{(1)}, W_2^{(2)}, W_4^{(1)}$, distinct. Thus by (15.54) we have

$$\sigma_4\{S_{12}S_{14} \,|\, W_1^{(1)}, W_2^{(2)}, W_4^{(1)} \text{ d.}\} = 2\rho_{121}. \tag{15.91}$$

  The summand for $W_2 = W_4 \neq W_3$ is also $2\rho_{121}$. For $W_3 = W_4 \neq W_2$, the sum is

$$\sigma_3\{1 \,|\, W_1^{1)}, W_2^{(1)}, W_3^{(2)}\} = \epsilon_{112}. \tag{15.92}$$

  Finally, consider the case W_2, W_3, W_4 are distinct. Using (15.49), we have $d! = 24$ and the sum over the $\pi \in \mathcal{P}_4$ turns out to be 8, hence

$$\sigma_4\{S_{13}S_{14} \,|\, W_1, W_2, W_3, W_4 \text{ d.}\} = \frac{1}{3}\,\epsilon_{1111}. \tag{15.93}$$

  Summing the various components yields the result in Table 15.1.

- $(12,13,23,34)$. If we switch W_1 and W_2 in the summand, we go from $S_{12}S_{13}S_{23}S_{34}$ to $S_{21}S_{23}S_{13}S_{34}$, which is its opposite. Thus the summation is zero.

- $(12,12,13,24)$. The equalities we deal with here are $W_1 = W_4$, $W_2 = W_3$, and $W_3 = W_4$. If all hold, the summand is zero. Looking at sets of two of the equalities, the only one not zero is $W_1 = W_4$, $W_2 = W_3$, but $W_3 \neq W_4$, which has summand $S_{12}^3 S_{21} = -\bar{I}_{12}$. The multiplicities are both two, hence

$$\sigma_4\{-\bar{I}_{12} \,|\, W_1^{(2)}, W_2^{(2)}\} = -\epsilon_{22}. \tag{15.94}$$

The sums for the single equalities are next.

$$\begin{array}{llll} \mathrm{I}_{14} & \bar{\mathrm{I}}_{23} & \bar{\mathrm{I}}_{34}: & \sigma_4\{-S_{12}S_{13} \,|\, W_1^{(2)}, W_2^{(1)}, W_3^{(1)} \text{ d.}\} \\ \bar{\mathrm{I}}_{14} & \mathrm{I}_{23} & \bar{\mathrm{I}}_{34}: & \sigma_4\{-S_{21}S_{24} \,|\, W_1^{(1)}, W_2^{(2)}, W_4^{(1)} \text{ d.}\} \\ \bar{\mathrm{I}}_{14} & \bar{\mathrm{I}}_{23} & \mathrm{I}_{34}: & \sigma_4\{S_{31}S_{32} \,|\, W_1^{(1)}, W_2^{(1)}, W_3^{(2)} \text{ d.}\} \end{array} \tag{15.95}$$

In each case, the variable with multiplicity of two is the one appearing in both S_{ij}'s, so that the three summations are the same except for sign. The sum of the three is thus the same as the first one, which by (15.54) is

$$\sigma_4\{-S_{12}S_{13} \,|\, W_1^{(2)}, W_2^{(1)}, W_3^{(1)}\} = -2\rho_{211} - 2\rho_{112} + 2\rho_{121}. \tag{15.96}$$

The case that all W_i are distinct yields zero, since the sum over the $\boldsymbol{\pi} \in \mathcal{P}_4$ of the summand is zero. Thus the total is the sum of (15.94) and (15.96).

- $(12, 12, 13, 34)$. We look at the equalities (among $W_1 = W_4, W_2 = W_3, W_2 = W_4$) that do not lead to a summand of zero:

(In)equalities			Summand	Condition	Sum
I_{14}	I_{23}	$\bar{\mathrm{I}}_{24}$	$-\bar{\mathrm{I}}_{12}$	$W_1^{(2)}, W_2^{(2)}$ d.	$-\epsilon_{22}$
I_{14}	$\bar{\mathrm{I}}_{23}$	$\bar{\mathrm{I}}_{24}$	$-\bar{\mathrm{I}}_{12}\bar{\mathrm{I}}_{13}$	$W_1^{(2)}, W_2^{(1)}, W_3^{(1)}$ d.	$-\epsilon_{112}$
$\bar{\mathrm{I}}_{14}$	I_{23}	$\bar{\mathrm{I}}_{24}$	$-S_{21}S_{24}$	$W_1^{(1)}, W_2^{(2)}, W_4^{(1)}$ d.	$-2\rho_{211} - 2\rho_{112} + 2\rho_{121}$
$\bar{\mathrm{I}}_{14}$	$\bar{\mathrm{I}}_{23}$	I_{24}	$-S_{31}S_{32}$	$W_1^{(1)}, W_2^{(2)}, W_3^{(1)}$ d.	$-2\rho_{121}$
$\bar{\mathrm{I}}_{14}$	$\bar{\mathrm{I}}_{23}$	$\bar{\mathrm{I}}_{24}$	$-S_{31}S_{34}\bar{\mathrm{I}}_{12}$	W_1, W_2, W_3, W_4 d.	$-\frac{1}{3}\epsilon_{1111}$

(15.97)

The first two sums follow directly from (15.47). The third is the same as (15.96), and the fourth is the negative of (15.91). The final one is the negative of (15.93). Summing the final column yields the result in Table 15.1

- $(12, 12, 34, 34)$. The extra equalities are $W_1 = W_3, W_1 = W_4, W_2 = W_3, W_2 = W_4$. If three or more of these hold, the summand is zero. The only pairs of these equalities that yield nonzero summands are $W_1 = W_3, W_2 = W_4$ and $W_1 = W_4, W_2 = W_3$. Both pairs yield a sum $\sigma_4\{1 \,|\, W_1^{(2)}, W_2^{(2)}\} \equiv \epsilon_{22}$. All four single equalities yield the same sum, which is equivalent to $\sigma_4\{1 \,|\, W_1^{(2)}, W_2^{(1)}, W_4^{(1)}\} \equiv \epsilon_{112}$. Finally, if the four variables are distinct, the sum is ϵ_{1111}. Thus the total sum is $2\epsilon_{22} + 4\epsilon_{112} + \epsilon_{1111}$.

- $(12, 13, 24, 34)$. We consider case depending on whether $W_1 = W_4$ and/or $W_2 = W_3$.

(In)equalities		Summand	Condition	Sum
I_{14}	I_{23}	$\bar{\mathrm{I}}_{12}$	$W_1^{(2)}, W_2^{(2)}$ d.	ϵ_{22}
I_{14}	$\bar{\mathrm{I}}_{23}$	$\bar{\mathrm{I}}_{12}\bar{\mathrm{I}}_{13}$	$W_1^{(2)}, W_2^{(1)}, W_3^{(1)}$ d.	ϵ_{112}
$\bar{\mathrm{I}}_{14}$	I_{23}	$\bar{\mathrm{I}}_{12}\bar{\mathrm{I}}_{24}$	$W_1^{(1)}, W_2^{(2)}, W_4^{(1)}$ d.	ϵ_{112}
$\bar{\mathrm{I}}_{14}$	$\bar{\mathrm{I}}_{23}$	$S_{12}S_{13}S_{24}S_{34}$	W_1, W_2, W_3, W_4 d.	$\frac{1}{3}\epsilon_{1111}$

(15.98)

The first three sums follow directly from (15.47). The final one uses (15.49), where the sum of the summand over the $\pi \in \mathcal{P}_4$ is 8, hence with d! = 24, we obtain $\epsilon_{1111}/3$. Thus the total sums up the last column.

- $(12, 13, 14, 15)$. We need to consider all possible equalities/inequalities among W_2, W_3, W_4, W_5, of which there are six. We work according to how many of the six are equalities. In each case, there are zero, one, or several equivalent sets of equalities that have nonzero summands. We give one archtype example for each (if there are any).

Equalities	Archtype	#	Summand	Condition	Sum
0	All distinct	1	$S_{12}S_{13}S_{14}S_{15}$	$W_1, \dots, W_5$ d.	$\frac{1}{5}\epsilon_{11111}$
1	$I_{23}\bar{I}_{24}\bar{I}_{25}\bar{I}_{34}\bar{I}_{35}\bar{I}_{45}$	6	$S_{14}S_{15}$	$W_1^{(1)}, W_2^{(2)}, W_4^{(1)}, W_5^{(1)}$ d.	$\frac{1}{3}\epsilon_{1112}$
2	$I_{23}\bar{I}_{24}\bar{I}_{25}\bar{I}_{34}\bar{I}_{35}I_{45}$	3	$\bar{I}_{12}\bar{I}_{14}$	$W_1^{(1)}, W_2^{(2)}, W_4^{(2)}$ d.	ϵ_{122}
3	$I[W_2{=}W_3{=}W_4{\neq}W_5]$	4	$S_{12}S_{15}$	$W_1^{(1)}, W_2^{(3)}, W_5^{(1)}$ d.	$2\rho_{131}$
4		0			
5		0			
6	$I[W_2{=}W_3{=}W_4{=}W_5]$	1	$\bar{I}_{12}$	$W_1^{(1)}, W_2^{(4)}$	ϵ_{14}

(15.99)

If none of the equalities hold, then all five variables are distinct, so we can use (15.49). The sum over the π of the summand is 24, and the q! = 120, hence the sum is $\epsilon_{11111}/5$. There are six ways exactly on equality can hold, and by symmetry they all have the same sum. Taking $W_2 = W_3$, the summand is $S_{12}^2 S_{14}S_{15}$. Since otherwise all the variables are distinct, we can use the summand $S_{14}S_{15}$. We use (15.45). Since the multiplicity of W_2 is two, while the others' is one, we arrange the sum depending on the rank of W_2 (setting $(W_1, W_2, W_4, W_5) = (\pi_1, \pi_2, \pi_3, \pi_4)$):

$$\sigma_d\{S_{14}S_{15} \,|\, W_1^{(1)}, W_2^{(2)}, W_4^{(1)}, W_5^{(1)} \text{ d.}\} = \sum_{j=1}^{4} \sum_{\pi \in \mathcal{P}_4, \pi_2 = j} S(\pi_1 - \pi_3) S(\pi_1 - \pi_4) \rho_{1\dots2\dots1}, \quad (15.100)$$

where the "2" in the "1…2…1" is in the j^{th} slot. But the summand does not depend on π_2, so by (15.75) the inner sum of summand is 2 for each fixed π_2. Thus,

$$\sigma_d\{S_{14}S_{15} \,|\, W_1^{(1)}, W_2^{(2)}, W_4^{(1)}, W_5^{(1)} \text{ d.}\} = 2(\rho_{2111} + \rho_{1211} + \rho_{1121} + \rho_{1112}) \quad (15.101)$$

$$= \frac{1}{3}\epsilon_{1112}. \quad (15.102)$$

There are $\binom{6}{2} = 15$ pairs of equalities to consider. If the indices in the two equalities intersect, e.g., $W_2 = W_3$ and $W_3 = W_5$, then since the inequalities would include $W_2 \neq W_5$, we have an impossibility. There are twelve such pairs. The remaining three are $I_{23}I_{45}, I_{14}I_{35}$, and $I_{25}I_{34}$, which all yields the same sum. Taking the first pair, the summand is $\bar{I}_{12}\bar{I}_{14}$, which with the condition $W_1^{(1)}, W_2^{(2)}, W_4^{(2)}$ distinct yields a sum of ϵ_{122} by (15.47).

Every triple of equalities has at least one pair for which the indices overlap. Thus as in the previous paragraph, for the case to be possible, the other equality has to overlap the others. For example, $I_{23}I_{25}I_{35}\bar{I}_{24}\bar{I}_{34}\bar{I}_{45} = I[W_2 = W_3 = W_5 \neq W_4]$ is nonzero, and the other nonzero cases have three W_i's equal to each other, but not equal to the fourth. Thus there are four possibilities. Each is equal to

$$\sigma_5\{S_{12}S_{15} \,|\, W_1^{(1)}, W_2^{(3)}, W_5^{(1)} \text{ d.}\} = 2\rho_{131}, \tag{15.103}$$

using the same idea as in (15.91).

Any four or more of the equalities will imply that $W_2 = W_3 = W_4 = W_5$, hence any left-over inequalities will produce an impossibility. Thus there are no nonzero sums for four or five equalities. For six, we have the four all equal, and the summand of $\bar{I}_{12}$ with multiplicities (1,4), hence the sum is ϵ_{14}. The total for this pattern is found by multiplying the "#" and "Sum' columns in (15.99), then adding.

- $(12, 13, 24, 25)$. We proceed much as we did for the previous pattern, though the lack of as much symmetry requires more cases. The extra possible (in)equalities to consider are $14, 15, 23, 34, 35$, and 45. If all of these are inequalities, then we use (15.49) to find

$$\sigma_5\{S_{12}S_{13}S_{24}S_{25} \,|\, W_1, \ldots, W_5 \text{ d.}\} = \frac{1}{15}\epsilon_{11111} \tag{15.104}$$

since the sum over the π is 8, and q! = 120. For just one inequality, we note that by symmetry $W_1 = W_4$ and $W_1 = W_5$ yield the same sum, as do $W_3 = W_4$ and $W_3 = W_5$. In each case, we have multiplicities with three 1's and one 2. We use the idea in (15.100), where the inner sum fixes the rank of the variable with multiplicity 2. We obtain the following:

Equalities	Summand	Condition	Sum
I_{14} or I_{15}	$S_{13}S_{25}$	$W_1^{(2)}, W_2, W_3, W_5$ d.	0
I_{23}	$S_{24}S_{25}$	$W_1, W_2^{(2)}, W_4, W_5$ d.	$6\rho_{2111} - 2\rho_{1211} - 2\rho_{1121} + 6\rho_{1112}$
I_{34} or I_{35}	$S_{12}S_{13}S_{23}S_{25}$	$W_1, W_2, W_3^{(2)}, W_5$ d.	$-2\rho_{2111} + 2\rho_{1211} + 2\rho_{1121} - 2\rho_{1112}$
I_{45}	$S_{12}S_{13}$	$W_1, W_2, W_3, W_4^{(2)}$ d.	$2\rho_{2111} + 2\rho_{1211} + 2\rho_{1121} + 2\rho_{1112}$
Total			$4(\rho_{2111} + \rho_{1211} + \rho_{1121} + \rho_{1112}) = \frac{2}{3}\epsilon_{1112}$

(15.105)

For the cases with two possible equalities, we indicate the summand and condition for the first mentioned. All other relevant pairs are inequalities. The final row in (15.105) sums the sums, multiplying the third sum by two. Using (15.47), we see that ϵ_{1112} is the sum of those four ρ's times six, hence we obtain $2\epsilon_{1112}/3$.

Consider the pairs of equalities. Again if a pair's subscripts intersect, then either the summand becomes zero, or with the inequalities the situation is impossible. That leaves five pairs of equalities. The pairs (I_{14}, I_{23}) and (I_{15}, I_{23}) yield the same sum, as do the pairs (I_{14}, I_{35}) and (I_{15}, I_{34}). These sums are

$$(I_{14}, I_{23}) : \sigma_5\{S_{21}S_{25} \,|\, W_1^{(2)}, W_2^{(2)}, W_5 \text{ d.}\} \;\;\&\;\; (I_{14}, I_{35}) : \sigma_5\{-S_{31}S_{32} \,|\, W_1^{(2)}, W_2, W_3^{(2)} \text{ d.}\} \tag{15.106}$$

If in the latter expression we change 3→2 and 2→5, we see that the two expressions in (15.106) are opposites, hence the sum of the sums for these four pairs is zero.The other nonzero pair is $(\mathrm{I}_{23}, \mathrm{I}_{45})$, which renders the summand $\bar{\mathrm{I}}_{12}\bar{\mathrm{I}}_{24}$ with condition $W_1, W_2^{(2)}, W_4^{(2)}$ distinct. Thus by (15.47) its sum is ϵ_{122}.

There are only two triples of equalities that have nonzero sums, given next:

$$\begin{aligned}\mathrm{I}_{14}, \mathrm{I}_{15}, \mathrm{I}_{45} &: \sigma_5\{S_{12}S_{13} \,|\, W_1^{(3)}, W_2, W_3\} = 2\rho_{311} - 2\rho_{131} + 2\rho_{311}\\ \mathrm{I}_{34}, \mathrm{I}_{35}, \mathrm{I}_{45} &: \sigma_5\{S_{12}S_{13} \,|\, W_1, W_2, W_3^{(3)}\} = 2\rho_{131}.\end{aligned} \tag{15.107}$$

See (15.54), or (15.96) (multiplied by -1, with multiplicity "2" replaced with "3") and (15.103).

The single nonzero quadruple of equalities is $(\mathrm{I}_{14}, \mathrm{I}_{15}, \mathrm{I}_{45}, \mathrm{I}_{23})$, which has summand $\bar{\mathrm{I}}_{12}$ and multiplicities 3, 2, hence sum ϵ_{23}. All sets of five equalities, and the set of six, has zero sum. Thus we find the overall sum for this pattern by summing the results in (15.104), (15.105), (15.107), and the ϵ_{122} for the pairs, and ϵ_{23} for the quadruples.

- $(12, 12, 34, 35)$. The extra equalities we have here are $\mathrm{I}_{13}, \mathrm{I}_{14}\mathrm{I}_{15}, \mathrm{I}_{23}, \mathrm{I}_{24}, \mathrm{I}_{25}, \mathrm{I}_{45}$. If none of those equalities hold, then using (15.49) we find the sum is $\epsilon_{11111}/3$. For single equalities, we have three different sums, as follows, where the summand and condition is for the first mentioned possibility:

Equalities	Summand	Condition	Sum
I_{13} or I_{23}	$S_{13}S_{15}$	$W_1^{(2)}, W_2, W_4, W_5$ d.	$6\rho_{2111} - 2\rho_{1211} - 2\rho_{1121} + 6\rho_{1112}$
I_{14} or I_{15} or I_{24} or I_{25}	$S_{31}S_{35}$	$W_1, W_2^{(2)}, W_3, W_5$ d.	$4\rho_{1211} + 4\rho_{1121}$
I_{45}	$\bar{\mathrm{I}}_{12}\bar{\mathrm{I}}_{34}$	$W_1, W_2, W_3, W_4^{(2)}$ d.	ϵ_{1112}
Total			$3\epsilon_{1112}$

(15.108)

The first two lines use (15.57), and the final one uses (15.47). The sum of the sums for the first six equalities is

$$\begin{aligned}2(6\rho_{2111} - 2\rho_{1211} - 2\rho_{1121} + 6\rho_{1112}) + 16(\rho_{1121} + \rho_{1211}) &= 12(\rho_{2111} + \rho_{1211} + \rho_{1121} + \rho_{1112})\\ &= 2\epsilon_{1112}.\end{aligned} \tag{15.109}$$

Most pairs of equalities yield a sum of zero. The next table summarizes the nonzero pairs.

Equalities	Summand	Condition	Sum
$(\mathrm{I}_{13}, \mathrm{I}_{45}), (\mathrm{I}_{23}, \mathrm{I}_{45})$	$\bar{\mathrm{I}}_{12}\bar{\mathrm{I}}_{14}$	$W_1^{(2)}, W_2, W_4^{(2)}$ d.	ϵ_{122}
$(\mathrm{I}_{13}, \mathrm{I}_{24}), (\mathrm{I}_{13}, \mathrm{I}_{25})$ $(\mathrm{I}_{14}, \mathrm{I}_{23}), (\mathrm{I}_{15}, \mathrm{I}_{23})$	$S_{12}S_{15}$	$W_1^{(2)}, W_2^{(2)}, W_5$ d.	$2\rho_{212}$
$(\mathrm{I}_{14}, \mathrm{I}_{25}), (\mathrm{I}_{15}, \mathrm{I}_{24})$	$S_{31}S_{32}$	$W_1^{(2)}, W_2^{(2)}, W_3$ d.	$2(\rho_{122} - \rho_{212} + \rho_{221})$
Total			$3\epsilon_{122}$

(15.110)

The second and third sums add to $8\rho_{212} + 4(\rho_{122} - \rho_{212} + \rho_{221}) = 2\epsilon_{122}$, hence the total is $3\epsilon_{122}$.

There are two nonzero triples of equalities (with the same sum), and two nonzero quadruples (also with the same sum), as below:

$$\begin{aligned}(I_{14}, I_{15}, I_{45}) \text{ or } (I_{24}, I_{25}, I_{45}) &: \sigma_5\{\bar{I}_{12}\bar{I}_{13} \,|\, W_1^{(3)}, W_2, W_3\} = \epsilon_{113};\\ (I_{14}, I_{15}, I_{23}, I_{45}) \text{ or } (I_{13}, I_{24}, I_{25}, I_{45}) &: \sigma_5\{\bar{I}_{12} \,|\, W_1^{(3)}, W_2^{(2)}\} = \epsilon_{23}.\end{aligned} \tag{15.111}$$

For the total we add the $\epsilon_{11111}/3$ from the first paragraph to the totals in (15.108), (15.110), and (15.111).

- $(12, 13, 24, 35)$. Now the extra equalities are $I_{14}, I_{15}, I_{23}, I_{25}, I_{34}, I_{45}$. Using (15.49) for the case that none of those qualities hold, we find the sum over $\boldsymbol{\pi}$ to be 16, hence with $q! = 120$, the overall sum is $2\epsilon_{11111}/15$. The next table has the results for single equalities holding:

Equalities	Summand	Condition	Sum
I_{14} or I_{15}	$S_{31}S_{35}$	$W_1^{(2)}, W_2, W_3, W_5$ d.	$4\rho_{1211} + 4\rho_{1121}$
I_{23}	$S_{24}S_{25}$	$W_1, W_2^{(2)}, W_4, W_5$ d.	$6\rho_{2111} - 2\rho_{1211} - 2\rho_{1121} + 6\rho_{1112}$
I_{25} or I_{34}	$S_{12}S_{13}S_{24}S_{32}$	$W_1, W_2^{(2)}, W_3, W_4$ d.	0
I_{45}	$S_{12}S_{13}S_{24}S_{34}$	$W_1, W_2, W_3, W_4^{(2)}$ d.	$\frac{1}{3}\epsilon_{1112}$
Total			$\frac{4}{3}\epsilon_{1112}$

(15.112)

The first two lines use (15.57), and the last two need the more general (15.45). For the third, fixing the $\pi_4 = k$, the sum of the summand over the other π_i's is zero for each k. For the fourth, the sum is 2 for each k.

The pairs of equalities where the equality's subscripts intersect have sums of zero. The others are next:

Equalities	Summand	Condition	Sum
$(I_{14}, I_{23}), (I_{15}, I_{23})$	$S_{21}S_{25}$	$W_1^{(2)}, W_2^{(2)}, W_5$ d.	$2\rho_{212}$
$(I_{14}, I_{25}), (I_{15}, I_{34})$	$S_{31}S_{32}$	$W_1^{(2)}, W_2^{(2)}, W_3$ d.	$2(\rho_{122} - \rho_{212} + \rho_{221})$
(I_{23}, I_{45})	$\bar{I}_{12}\bar{I}_{24}$	$W_1, W_2^{(2)}, W_4^{(2)}$ d.	ϵ_{122}
(I_{25}, I_{34})	$-S_{12}S_{13}$	$W_1, W_2^{(2)}, W_3^{(2)}$ d.	$-2(\rho_{122} - \rho_{212} + \rho_{221})$
Total			$2\epsilon_{122}$

(15.113)

The only set of three equalities that has a nonzero sum is (I_{14}, I_{15}, I_{45}), which has the sum $\sigma_5\{\bar{I}_{12}\bar{I}_{13} \,|\, W_1^{(3)}, W_2, W_3\} = \epsilon_{113}$. The only quadruple adds an equality to that triple: $(I_{14}, I_{15}, I_{45,23})$, which has sum ϵ_{23}. No set of more than four equalities has a nonzero sum. Hence with (15.112), (15.113), and the $2\epsilon_{11111}/15$ in the first paragraph, we obtain the answer in Table 15.1.

- $(12, 13, 45, 46)$. This pattern is the only one we need to handle that has six variables, which does make it more complicated. There is a good deal of symmetry. There are now eleven extra equalities to consider: $I_{14}, I_{15}, I_{16}, I_{23}, I_{24}, I_{25}, I_{26}, I_{34}, I_{35}, I_{36}$, and I_{56}. With none of them holding, we use (15.49) to find

$$\sigma_6\{S_{12}S_{13}S_{45}S_{46} \,|\, W_1, W_2, W_3, W_4, W_5\} = 80\rho_{111111} = \frac{1}{9}\epsilon_{111111}, \tag{15.114}$$

since $q! = 720$. The results for the single equalities follow, where the coefficients (a, b, c, d, e) mean the sum for that equality is

$$a\,\rho_{21111} + b\,\rho_{12111} + c\,\rho_{11211} + d\,\rho_{11121} + e\rho_{11112}: \tag{15.115}$$

Equality	Summand	Condition					
I_{14}	$S_{12}S_{13}S_{15}S_{16}$	$W_1^{(2)}, W_2, W_3, W_5, W_6$ d.	24	−24	24	−24	24
I_{15} or I_{16} or I_{24} or I_{34}	$S_{12}S_{13}S_{41}S_{46}$	$W_1^{(2)}, W_2, W_3, W_4, W_6$ d.	0	12	−16	12	0
I_{23} or I_{56}	$S_{45}S_{46}$	$W_1, W_2^{(2)}, W_4, W_5, W_6$ d.	8	8	8	8	8
I_{25} or I_{26} or I_{35} or I_{36}	$S_{12}S_{13}S_{42}S_{46}$	$W_1, W_2^{(2)}, W_3, W_4, W_6$ d.	0	0	16	0	0
Total			40	40	40	40	40

(15.116)

The total is 40 times the sum of the five ρ's. Since ϵ_{11112} is 24 times the sum, the total is $5\epsilon_{11112}/3$.

Again the pairs of equalities whose subscripts intersect have zero sums. There are 25 remaining pairs, which can be grouped by equivalence into six groups. The next table has the results for one representative from each group. The coefficients are for, respectively, $\rho_{2211}, \rho_{2121}, \rho_{2112}, \rho_{1221}, \rho_{1212}, \rho_{1122}$.

Equalities	#	Summand	Condition						
(I_{14}, I_{23})	6	$S_{15}S_{16}$	$W_1^{(2)}, W_2^{(2)}, W_5, W_6$ d.	4	0	4	−4	0	4
(I_{15}, I_{23})	4	$S_{41}S_{46}$	$W_1^{(2)}, W_2^{(2)}, W_4, W_6$ d.	0	2	0	4	2	0
(I_{15}, I_{24})	4	$-S_{13}S_{26}$	$W_1^{(2)}, W_2^{(2)}, W_3, W_6$ d.	−4	0	4	4	0	−4
(I_{15}, I_{26})	8	$S_{12}S_{13}S_{41}S_{42}$	$W_1^{(2)}, W_2^{(2)}, W_3, W_4$ d.	0	2	−4	0	2	0
(I_{23}, I_{56})	1	1	$W_1, W_2^{(2)}, W_4, W_5^{(2)}$ d.	4	4	4	4	4	4
(I_{25}, I_{36})	2	$S_{12}S_{13}S_{42}S_{43}$	$W_1, W_2^{(2)}, W_3^{(2)}, W_4$ d.	4	−4	4	4	−4	4
Total				20	20	20	20	20	20

(15.117)

Since 4 times the sum of the six ρ's is ϵ_{1122}, the total from this table is $5\epsilon_{1122}$.

The nonzero triples of equalities sort into two types: Those with no intersections in their subscripts, e.g., (I_{14}, I_{23}, I_{56}), and those which imply the equality of three W_i's,

such as (I_{15}, I_{16}, I_{56}). Below we have summarized the results, where again we have one representative from each equivalent type of sum:

Equalities	#	Summand	Condition	Sum
(I_{14}, I_{23}, I_{56})	3	1	$W_1^{(2)}, W_2^{(2)}, W_5^{(2)}$ d.	ϵ_{222}
(I_{15}, I_{24}, I_{36})	4	$-S_{31}S_{32}$	$W_1^{(2)}, W_2^{(2)}, W_3^{(2)}$ d.	$-\frac{1}{3}\epsilon_{222}$
(I_{15}, I_{16}, I_{56})	2	$S_{12}S_{13}$	$W_1^{(3)}, W_2, W_3$ d.	$6\rho_{3111} - 2\rho_{1311} - 2\rho_{1131} + 6\rho_{1113}$
(I_{23}, I_{25}, I_{35})	4	$S_{12}S_{13}$	$W_1, W_2^{(3)}, W_4$ d.	$4\rho_{1311} + 4\rho_{1131}$
Total				$\frac{5}{3}\epsilon_{222} + 2\epsilon_{1113}$

(15.118)

The sum of the last two sums is $12(\rho_{3111} + \rho_{1311} + \rho_{1131} + \rho_{1113}) = 2\epsilon_{1113}$.

There are 14 quadruples of equalities with nonzero sums. They are of four types, one representative of each being presented below. The coefficients for the ρ's in the order $(\rho_{123}, \rho_{132}, \rho_{213}, \rho_{231}, \rho_{312}, \rho_{321})$.

Equalities	#	Summand	Condition	Coefficients					
$(I_{15}, I_{16}, I_{23}, I_{56})$	2	1	$W_1^{(3)}, W_2^{(2)}, W_4$ d.	1	1	1	1	1	1
$(I_{14}, I_{23}, I_{26}, I_{36})$	4	$S_{12}S_{15}$	$W_1^{(2)}, W_2^{(3)}, W_5$ d.	-1	1	1	1	1	-1
$(I_{15}, I_{16}, I_{56}, I_{24})$	4	$S_{12}S_{13}$	$W_1^{(3)}, W_2^{(2)}, W_3$ d.	1	-1	1	-1	1	1
$(I_{24}, I_{35}, I_{36}, I_{56})$	4	$S_{12}S_{13}$	$W_1, W_2^{(2)}, W_3^{(3)}$ d.	1	1	-1	1	-1	1
Total				6	6	6	6	6	6

(15.119)

The sum of the six ρ's is ϵ_{123}, hence the total is $6\epsilon_{123}$.

There are three higher-order sets of equalities yielding nonzero sums, two sextets and one septet, as below:

Equalities	Summand	Condition	Sum
$(I_{15}, I_{16}, I_{23}, I_{24}, I_{34}, I_{56})$	1	$W_1^{(3)}, W_2^{(3)}$ d.	ϵ_{33}
$(I_{23}, I_{25}, I_{26}, I_{35}, I_{36}, I_{56})$	1	$W_1, W_2^{(4)}, W_4$ d.	ϵ_{114}
$(I_{14}, I_{23}, I_{25}, I_{26}, I_{35}, I_{36}, I_{56})$	1	$W_1^{(2)}, W_2^{(4)}$ d.	ϵ_{24}

(15.120)

Collecting the various terms, we have the result given in Table 15.1.

Chapter 16

Incomplete rankings

Here we allow that only a subset of the objects are ranked. If J of the m objects are ranked, then the ranked ones will have values in $\boldsymbol{z}$ of $1, \ldots, J$, while the non-ranked objects values will be "$*$". E.g., if there are $m = 5$ objects, a possible ranking of just objects $2, 3$, and 5 is $\boldsymbol{z} = (*, 3, 1, *, 2)$. Let $\mathcal{O}$ be the indices for the ranked objects, and for any vector $\boldsymbol{x}$, let $\boldsymbol{x}_{\mathcal{O}}$ be the subvector of x_i with $i \in \mathcal{O}$. In our example, $\mathcal{O} = \{2, 3, 5\}$ and $\boldsymbol{z}_{\mathcal{O}} = (3, 1, 2)$. For given $\mathcal{O}$, the space of $\boldsymbol{z}$ is

$$\mathcal{Z}_{\mathcal{O}} = \{\boldsymbol{z} \text{ is } 1 \times m \,|\, \boldsymbol{z}_{\mathcal{O}} \in \mathcal{P}_J \text{ and } z_i = * \text{ for } i \notin \mathcal{O}\}. \tag{16.1}$$

Then for given $\boldsymbol{z} \in \mathcal{Z}$, the set of compatible full rankings is

$$C(\boldsymbol{z}) = \{\boldsymbol{y} \in \mathcal{P}_m \,|\, \text{rank}(\boldsymbol{y}_{\mathcal{O}}) = \boldsymbol{z}_{\mathcal{O}}\}. \tag{16.2}$$

We start by finding the conditional distributions of Y_i given the $\boldsymbol{Z} = \boldsymbol{z}$, so that we can find the averaged versions of Spearman's, other Hoeffding, and Kendall distances. The result is due to Alvo & Cabilio (1995).

Lemma 16.1 (Alvo and Cabilio)**.** *Given $\mathcal{O} \subset \{1, \ldots, m\}$, let $J \equiv \#\mathcal{O}$, and suppose $\boldsymbol{z} \in \mathcal{Z}_{\mathcal{O}}$. Then we have the following:*

(a) For $1 \leqslant t \leqslant m$ and $1 \leqslant r \leqslant J$, we have

$$P[Y_i = t \,|\, \textit{rank}(\boldsymbol{Y}_{\mathcal{O}}) = \boldsymbol{z}_{\mathcal{O}}] = \begin{cases} g_J(t \,|\, r) & \textit{if } z_i = r, \\ \frac{1}{m} & \textit{if } i \notin \mathcal{O}, \end{cases} \tag{16.3}$$

where

$$g_J(t \,|\, r) = \frac{\binom{t-1}{r-1}\binom{m-t}{J-r}}{\binom{m}{J}} \textit{ if } t = r, \ldots, m - J + r, \textit{ and } g_J(t \,|\, r) = 0 \textit{ otherwise.} \tag{16.4}$$

Furthermore,

$$E[\boldsymbol{Y}_{\mathcal{O}} \,|\, \textit{rank}(\boldsymbol{Y}_{\mathcal{O}}) = \boldsymbol{z}_{\mathcal{O}}] = \frac{m+1}{J+1}\boldsymbol{z}_{\mathcal{O}} \textit{ and } E[Y_i \,|\, \textit{rank}(\boldsymbol{Y}_{\mathcal{O}}) = \boldsymbol{z}_{\mathcal{O}}] = \frac{m+1}{2} \textit{ if } i \notin \mathcal{O}. \tag{16.5}$$

(b)

$$P[Y_i > Y_j \,|\, \mathit{rank}(\boldsymbol{Y}_{\mathcal{O}}) = \boldsymbol{z}_{\mathcal{O}}] = \begin{cases} I[z_i > z_j] & \text{if } i,j \in \mathcal{O}, \\ \frac{1}{2} & \text{if } i,j \notin \mathcal{O}, \\ \frac{z_i}{J+1} & \text{if } i \in \mathcal{O}, j \notin \mathcal{O}, \\ 1 - \frac{z_j}{J+1} & \text{if } i \notin \mathcal{O}, j \in \mathcal{O}. \end{cases} \tag{16.6}$$

Proof. (a) Consider the first condition in (16.3). Since $z_i = r$, the y_i has to be the r^{th} largest among the observed rankings, that is, among $\{y_j \,|\, j \in \mathcal{O}\}$. Thus t must be at least r. Also, there have to be at least $J - r$ of the y_j larger than y_i, to match with the z_j larger than z_i. Thus $t \leqslant m - J + r$. For t in the valid range, we argue that

$$\#\{\boldsymbol{y} \in \mathcal{P}_m \,|\, y_i = t, \mathrm{rank}(\boldsymbol{y}_{\mathcal{O}}) = \boldsymbol{z}_{\mathcal{O}}\} = \binom{t-1}{r-1}\binom{m-t}{J-r}(m-J)!. \tag{16.7}$$

We need $y_i = t$ to be the r^{th} largest value in $\boldsymbol{y}_{\mathcal{O}}$, hence $r-1$ of them need to be chosen from the $t-1$ values less than t, and the other $J-r$ must be chosen from the $m-t$ values greater than t. Hence the two binomial coefficients in (16.7). The remaining y_j for $j \notin \mathcal{O}$ have the $m-J$ values left unchosen, and can be in any order, hence the $(m-J)!$ term. For the denominator, we need $\#\{\boldsymbol{y} \in \mathcal{Y} \,|\, \mathrm{rank}(\boldsymbol{y}_{\mathcal{O}}) = \boldsymbol{z}_{\mathcal{O}}\}$. By symmetry, the number of compatible $\boldsymbol{y}$ for each $\boldsymbol{z} \in \mathcal{Z}$ is the same, being equal to $\#\mathcal{P}_m/\#\mathcal{P}_J = m!/J!$, hence dividing (16.7) by $m!/J!$ yields the $g_J(t\,|\,r)$ in (16.4). If $t < r$ or $J - r > m - t$, there are no compatible $\boldsymbol{y}$, yielding $g_J(t\,|\,r) = 0$.

Next, if $i \notin \mathcal{O}$, then there is no constraint on y_i, hence it is equally likely to be anything between 1 and m.

We could calculate the mean directly via (16.4), but instead we will use a representation described by Feller (1968), in exercise 15 of chapter IX. Let $G_1, \ldots, G_{J+1}$ be iid Geometric(p) random variables, with density $P[G = u] = (1-p)p^u$, $u = 0, 1, 2, \ldots$. Then $G_1 + \ldots + G_K \sim$ Negative Binomial(K, p), with density

$$P[G_1 + \ldots + G_K = u] = \binom{K+u-1}{K-1}(1-p)^K p^u, \quad u = 0, 1, 2, \ldots. \tag{16.8}$$

Then we can express (16.3) as the conditional distribution

$$g_J(t\,|\,r) = P[G_1 + \ldots + G_r = t - r \,|\, G_1 + \ldots + G_{J+1} = m - J]. \tag{16.9}$$

Conditionally, the $(G_1, \ldots, G_{J+1})$ are exchangeable given their sum, hence each element has the same conditional expectation. Thus

$$\begin{aligned} E[G_i \,|\, G_1 + \ldots + G_{J+1} = m - J] &= \frac{1}{J+1} E[G_1 + \ldots + G_{J+1} \,|\, G_1 + \ldots + G_{J+1} = m - J] \\ &= \frac{m-J}{J+1}. \end{aligned} \tag{16.10}$$

The mean of $g_J(t\,|\,r)$ is found by multiplying (16.10) by r, then adding r, yielding $(m+1)r/(J+1)$ as in the first case of (16.5). If $i \notin \mathcal{O}$, then since Y_i is equally likely to be anything from 1 to m, the mean is $(m+1)/2$, completing (16.5).

(b) If i and j are both observed, since $\text{rank}(\boldsymbol{y}_{\mathcal{O}}) = \boldsymbol{z}_{\mathcal{O}}$, $I[y_i > y_j] = I[z_i > z_j]$. If neither are observed, since the non-observed y_i's can be in any order, $P[Y_i > Y_j] = \frac{1}{2}$. Now suppose $i \in \mathcal{O}$ and $j \notin \mathcal{O}$. Conditioning on $y_i = t$ and $\boldsymbol{z}_{\mathcal{O}}$, we know that of the $t-1$ values smaller than y_i, $r-1$ must be assigned to observed indices, where $r = z_i$. Thus the other $t-r$ are not observed, i.e.,

$$E[\sum_{k \neq \mathcal{O}} I[Y_i > Y_k] \,|\, Y_i = t, \text{rank}(\boldsymbol{Y}_{\mathcal{O}}) = \boldsymbol{z}_{\mathcal{O}}] = t - r. \tag{16.11}$$

Then by symmetry, for any given $j \notin \mathcal{O}$,

$$E[I[Y_i > Y_j] \,|\, Y_i = t, \text{rank}(\boldsymbol{Y}_{\mathcal{O}}) = \boldsymbol{z}_{\mathcal{O}}] = \frac{t-r}{m-J}. \tag{16.12}$$

Taking a further conditional expected value over Y_i, we have

$$P[Y_i > Y_j \,|\, \text{rank}(\boldsymbol{Y}_{\mathcal{O}}) = \boldsymbol{z}_{\mathcal{O}}] = \frac{E[Y_i \,|\, \text{rank}(\boldsymbol{Y}_{\mathcal{O}}) = \boldsymbol{z}_{\mathcal{O}}] - r}{m-J} = \frac{r}{J+1'}, \tag{16.13}$$

using the conditional mean from (16.5). The last case in (16.6) then follows from the third. □

16.1 Ties, too

Here we consider ties as well as incomplete rankings, so that the observed rankings may have ties. Thus in addition to $\mathcal{O}$, the set of the indices of the J observed rankings, we have a pattern of ties $\boldsymbol{J} = (J_1, \ldots, J_L)$, where $\sum J_l = J$ and each $J_l \geqslant 1$. Thus there are J_l values of l among the observed ranking, as in Chapter 12. Similarly, the pattern of ties for the observed ranks in $\boldsymbol{W}$ is given by $\boldsymbol{I} = (I_1, \ldots, I_K)$. The space of vectors is then

$$\boldsymbol{Z}_{\mathcal{O}} = \{\boldsymbol{z} \text{ is } 1 \times m \,|\, \boldsymbol{z}_{\mathcal{O}} \text{ has pattern of ties } \boldsymbol{J},\ z_i = * \text{ if } i \notin \mathcal{O}\}. \tag{16.14}$$

For $\boldsymbol{z} \in \mathcal{Z}$, we wish to find the corresponding set of compatible complete rankings. We can do this in two steps: First find the $1 \times J$ rank vectors compatible with the observed part, $\boldsymbol{z}_{\mathcal{O}}$, then find the $1 \times m$ complete rankings compatible with each of those. That is, let

$$C^*(\boldsymbol{z}_0) = \{\boldsymbol{u}_{\mathcal{O}} \in \mathcal{P}_J \,|\, \boldsymbol{u}_{\mathcal{O}} \text{ is compatible with } \boldsymbol{z}_{\mathcal{O}}\}, \tag{16.15}$$

so that the complete rankings compatible with $\boldsymbol{z}$ is

$$C(\boldsymbol{z}) = \cup_{\boldsymbol{u}_{\mathcal{O}} \in C^*(\boldsymbol{z}_{\mathcal{O}})} C'(\boldsymbol{u}_{\mathcal{O}}), \quad \text{where} \quad C'(\boldsymbol{u}_{\mathcal{O}}) = \{\boldsymbol{y} \in \mathcal{P}_m \,|\, \text{rank}(\boldsymbol{y}_{\mathcal{O}}) = \boldsymbol{u}_{\mathcal{O}}\}. \tag{16.16}$$

Note that the sets $C'(\boldsymbol{u}_{\mathcal{O}})$ in the union are disjoint, since each has a distinct $\text{rank}(\boldsymbol{y}_{\mathcal{O}})$, and by symmetry they all have the same number of elements. Thus finding the average of any function over all elements in $C(\boldsymbol{z})$ is the same as finding the average of the individual averages of the elements of the $C'(\boldsymbol{u}_{\mathcal{O}})$. For $\boldsymbol{Y}$ itself, we have

$$E[\boldsymbol{Y} \,|\, \boldsymbol{Y} \in C(\boldsymbol{z})] = \frac{1}{\#C^*(\boldsymbol{z}_{\mathcal{O}})} \sum_{\boldsymbol{u}_{\mathcal{O}} \in C^*(\boldsymbol{z}_{\mathcal{O}})} E[\boldsymbol{Y} \,|\, \text{rank}(\boldsymbol{Y}_{\mathcal{O}}) = \boldsymbol{u}_{\mathcal{O}}]. \tag{16.17}$$

As in (16.5), the inner expected value is given by

$$\mathrm{E}[\boldsymbol{Y}_{\mathcal{O}} \,|\, \mathrm{rank}(\boldsymbol{Y}_{\mathcal{O}}) = \boldsymbol{u}_{\mathcal{O}}] = \frac{m+1}{J+1}\,\boldsymbol{u}_{\mathcal{O}} \;\text{ and }\; \mathrm{E}[Y_i \,|\, \mathrm{rank}(\boldsymbol{Y}_{\mathcal{O}}) = \boldsymbol{u}_{\mathcal{O}}] = \frac{m+1}{2} \;\text{ if }\; i \notin \mathcal{O}. \tag{16.18}$$

As in (12.24), averaging the $\boldsymbol{u}_{\mathcal{O}}$ over the $C(\boldsymbol{z}_{\mathcal{O}})$ yields the midranks of $\boldsymbol{z}_{\mathcal{O}}$. Also, for $i \notin \mathcal{O}$, the further average is still $(m+1)/2$. Thus

$$\mathrm{E}[\boldsymbol{Y}_{\mathcal{O}} \,|\, \boldsymbol{Y} \in C(\boldsymbol{z})] = \frac{m+1}{J+1}\,\mathrm{rank}(\boldsymbol{z}_{\mathcal{O}}) \;\text{ and }\; \mathrm{E}[Y_i \,|\, \boldsymbol{Y} \in C(\boldsymbol{z})] = \frac{m+1}{2} \;\text{ if }\; i \notin \mathcal{O}. \tag{16.19}$$

A similar approach can be used for $I[Y_i > Y_j]$, so that

$$\mathrm{P}[Y_i > Y_j \,|\, \boldsymbol{Y} \in C(\boldsymbol{z})] = \frac{1}{\#C^*(\boldsymbol{z}_{\mathcal{O}})} \sum_{\boldsymbol{u}_{\mathcal{O}} \in C^*(\boldsymbol{z}_{\mathcal{O}})} \mathrm{P}[Y_i > Y_j \,|\, \mathrm{rank}(\boldsymbol{Y}_{\mathcal{O}}) = \boldsymbol{u}_{\mathcal{O}}]. \tag{16.20}$$

Here, the inner probability is as in (16.6) with $\boldsymbol{u}_{\mathcal{O}}$ instead of $\boldsymbol{z}$. If $i, j \in \mathcal{O}$, then taking further expected value over $C^*(\boldsymbol{z}_{\mathcal{O}})$ is $I[z_i > z_j]$ if z_i and z_j are not tied, and $\frac{1}{2}$ otherwise. If $i, j \notin \mathcal{O}$, then the mean will stay $\frac{1}{2}$. If $i \in \mathcal{O}$ and $j \notin \mathcal{O}$, the further expected value of U_i is the midrank, divided by $J+1$. Likewise for the final case. Thus

$$\mathrm{P}[Y_i > Y_j \,|\, \boldsymbol{Y} \in C(\boldsymbol{z})] = \begin{cases} I[z_i > z_j] & \text{if } i, j \in \mathcal{O},\ z_i \neq z_j, \\ \frac{1}{2} & \text{if } i, j \in \mathcal{O},\ z_i = z_j, \\ \frac{1}{2} & \text{if } i, j \notin \mathcal{O}, \\ \frac{\mathrm{rank}(\boldsymbol{z})_i}{J+1} & \text{if } i \in \mathcal{O}, j \notin \mathcal{O}, \\ 1 - \frac{\mathrm{rank}(\boldsymbol{z})_j}{J+1} & \text{if } i \notin \mathcal{O}, j \in \mathcal{O}, \end{cases} \tag{16.21}$$

where

$$\mathrm{rank}(\boldsymbol{z})_{\mathcal{O}} = \mathrm{rank}(\boldsymbol{z}_{\mathcal{O}}) \;\text{ and }\; \mathrm{rank}(\boldsymbol{z})_i = * \;\text{ if }\; i \notin \mathcal{O}. \tag{16.22}$$

16.2 Null distribution

There are several possible null distributions that are reasonable to consider for cases in which one or both rank vectors $\boldsymbol{W}$ and $\boldsymbol{Z}$ have missing values. Suppose the numbers of observed rankings in $\boldsymbol{W}$ is I, and in $\boldsymbol{Z}$ is J. Some choices to make include whether to consider one of the vectors fixed, or both random; and if random, whether the observed indices are fixed or random. A further choice would be to have the I and J random, but we will not deal with that possibility. Also, the overall assumption is that whether a ranking is missing is independent of the rank it would have been assigned if not missing.

Suppose $\boldsymbol{Z}$ is not fixed, and let $\mathcal{O} \subset \{1, \ldots, m\}$ be the observed indices, with $\#\mathcal{O} = J$. Then two possible distributions for $\boldsymbol{Z}$ depending on whether we condition on the observed indices are

$$\begin{aligned} \text{Conditionally} &: \boldsymbol{Z} \sim \mathrm{Uniform}(\mathcal{Z}_{\mathcal{O}}); \\ \text{Unconditionally} &: \boldsymbol{Z} \sim \mathrm{Uniform}(\cup\{\mathcal{Z}_{\mathcal{O}'} \,|\, \mathcal{O}' \subset \{1, \ldots, m\}, \#\mathcal{O}' = J\}) \\ &\Leftrightarrow \quad \boldsymbol{Z} \sim \mathrm{Uniform}(\mathrm{Permutations}(\boldsymbol{z})), \end{aligned} \tag{16.23}$$

where $\boldsymbol{z}$ is any member of $\mathcal{Z}_{\mathcal{O}}$. We define the two distributions similarly for $\boldsymbol{W}$, where $\mathcal{A}$ is the set of observed indices, and $I = \#\mathcal{A}$. Alvo & Cabilio (1995) define the null hypothesis $\mathcal{H}_1$ to be when $I \geqslant J$, $\boldsymbol{W} = \boldsymbol{w}$ is fixed, and $\boldsymbol{Z}$ has the conditional distribution (although one could remove the $I \geqslant J$ condition); and $\mathcal{H}_2$ when one or both $\boldsymbol{W}$ and $\boldsymbol{Z}$ have the unconditional distribution. Another possibility is that both $\boldsymbol{W}$ and $\boldsymbol{Z}$ have a conditional distribution, which we will call $\mathcal{H}_3$.

We next treat the Spearman and Kendall distances separately.

16.3 Spearman

Recall from (13.1) that we can write Spearman's distance adjusting for ties in terms of the midranks. Similarly, we have here

$$d^{A}_{\text{Spear}}(\boldsymbol{w}, \boldsymbol{z}) = \mu_{\text{Spear}}(m) - 2\sum_{i=1}^{m}(r^*_i - \nu)(s^*_i - \nu), \tag{16.24}$$

where $\nu = (m+1)/2$ again, and the analogs of the midranks are as in (16.19):

$$\boldsymbol{r}^* = E[\boldsymbol{X} \,|\, \boldsymbol{X} \in C(\boldsymbol{w})] = \frac{m+1}{I+1}\,\boldsymbol{r} \;\text{ and }\; \boldsymbol{s}^* = E[\boldsymbol{Y} \,|\, \boldsymbol{Y} \in C(\boldsymbol{z})] = \frac{m+1}{J+1}\,\boldsymbol{s}. \tag{16.25}$$

We chose the factors so that

$$\boldsymbol{r}_{\mathcal{A}} = \text{rank}(\boldsymbol{w}_{\mathcal{A}}) \;\text{ and }\; \boldsymbol{s}_{\mathcal{O}} = \text{rank}(\boldsymbol{z}_{\mathcal{O}}), \tag{16.26}$$

the midranks of the respective vector of observed ranks. For the unobserved ranks, we have

$$r_i = \nu_I \equiv \frac{I+1}{2}, i \notin \mathcal{A} \;\text{ and }\; s_i = \nu_J \equiv \frac{J+1}{2}, i \notin \mathcal{O}. \tag{16.27}$$

Thus the summation part of (16.24) is

$$\sum_{i=1}^{m}(r^*_i - \nu)(s^*_i - \nu) = \frac{(m+1)^2}{(I+1)(J+1)}\sum_{i=1}^{m}(r_i - \nu_I)(s_i - \nu_J). \tag{16.28}$$

Note that for any index i, the summand is zero unless at both of the vectors' ranks are observed, i.e., unless $i \in \mathcal{A} \cap \mathcal{O}$.

Turn to the mean. If either vector has the unconditional distribution, the mean of (16.24) is $\mu_{\text{Spear}}(m)$, by definition of averaging. If $\boldsymbol{Z}$ has the conditional distribution, then the observed vector

$$\boldsymbol{S}_{\mathcal{O}} \sim \text{Uniform}(\text{Permutations}(\boldsymbol{s}_{\mathcal{O}})), \tag{16.29}$$

which has mean $\nu_J \mathbf{1}_J$, hence $E[S_i] = \nu_J$ for all i. Thus again the adjusted distance has the usual mean. That is, if either vector has either the conditional or unconditional distribution, the mean is of the distance is $\mu_{\text{Spear}}(m)$.

The variance, and higher moments, depend on the specific assumed distribution. Consider hypothesis $\mathcal{H}_1$, where $\boldsymbol{W} = \boldsymbol{w}$, and $\boldsymbol{Z}$ has the conditional distribution. Then the set of observed indices for $\boldsymbol{Z}$ is fixed at $\mathcal{O}$. Now

$$\begin{aligned} \mathrm{Var}\left[\sum_{i=1}^{m}(r_i-\nu_I)(S_i-\nu_J)\right] &= \mathrm{Var}\left[\sum_{i\in\mathcal{O}}(r_i-\nu_I)(S_i-\nu_J)\right] \\ &= \frac{\sum_{i\in\mathcal{O}}(r_i-\bar{r}_{\mathcal{O}})^2\sum_{i\in\mathcal{O}}(s_i-\nu_J)^2}{J-1}, \text{ where } \bar{r}_{\mathcal{O}}=\frac{\sum_{i\in\mathcal{O}} r_i}{J}, \end{aligned} \tag{16.30}$$

the final step following as in (13.35), but with the $1\times J$ vector $\boldsymbol{S}_{\mathcal{O}}$ in (16.29). Note that the r_i and s_i are not treated the same, since we are dealing with the indices observed by the $\boldsymbol{z}$, which in general do not coincide with those observed by the $\boldsymbol{w}$.

For $\mathcal{H}_2$, we assume that $\boldsymbol{W} = \boldsymbol{w}$ again, but $\boldsymbol{Z}$ has the unconditional distribution as in (16.23). Then the entire $1\times m$ vector $\boldsymbol{S}\sim$ Uniform(Permutations($\boldsymbol{s}$)), hence

$$\begin{aligned} \mathrm{Var}\left[\sum_{i=1}^{m}(r_i-\nu_I)(S_i-\nu_J)\right] &= \frac{\sum_{i=1}^{m}(r_i-\nu_I)^2\sum_{i=1}^{m}(s_i-\nu_J)^2}{m-1} \\ &= \frac{\sum_{i\in\mathcal{A}}(r_i-\nu_I)^2\sum_{i\in\mathcal{O}}(s_i-\nu_J)^2}{m-1}. \end{aligned} \tag{16.31}$$

Finally, turn to $\mathcal{H}_3$, where $\boldsymbol{R}_{\mathcal{A}}$ and $\boldsymbol{S}_{\mathcal{O}}$ are independent, with distribution (16.29) for the latter, and Uniform(Permutations($\boldsymbol{r}_{\mathcal{O}}$)) for the former. Then

$$\begin{aligned} \mathrm{Var}\left[\sum_{i=1}^{m}(R_i-\nu_I)(S_i-\nu_J)\right] &= \mathrm{E}\left[\mathrm{Var}\left[\sum_{i=1}^{m}(R_i-\nu_I)(S_i-\nu_J)\,|\,\boldsymbol{R}\right]\right] \\ &= \mathrm{E}\left[\frac{\sum_{i\in\mathcal{O}}(R_i-\overline{R}_{\mathcal{O}})^2\sum_{i\in\mathcal{O}}(s_i-\nu_J)^2}{J-1}\right], \end{aligned} \tag{16.32}$$

where the first equality follows since the conditional expected value is identically zero, and the second equality is from (16.30). Consider $\boldsymbol{R}_{\mathcal{O}}$, which has (potentially) some elements which are observed, and some which are not, hence equal ν_I. We can rearrange the indices so that the observed come first in the vector. That is, with $H = \#(\mathcal{A}\cap\mathcal{O})$,

$$\boldsymbol{R}_{\mathcal{O}} = (R_1,\ldots,R_H,\nu_I,\ldots,\nu_I), \text{ with } J-H \text{ of the } \nu_I'\text{s}. \tag{16.33}$$

Then

$$\sum_{i\in\mathcal{O}}(R_i-\nu_I)^2 = \sum_{i\in\mathcal{O}}(R_i-\overline{R}_{\mathcal{O}})^2 + J(\overline{R}_{\mathcal{O}}-\nu_I)^2. \tag{16.34}$$

Since

$$\overline{R}_{\mathcal{O}} = \frac{\sum_{i=1}^{H}R_i+(J-H)\nu_I}{J} = \frac{H}{J}\overline{R}_H + \frac{J-H}{J}\nu_I, \text{ where } \overline{R}_H=\frac{\sum_{i=1}^{H}R_i}{H}, \tag{16.35}$$

(16.34) implies that

$$\sum_{i\in\mathcal{O}}(R_i-\overline{R}_{\mathcal{O}})^2 = \sum_{i\in\mathcal{O}}(R_i-\nu_I)^2 - \frac{H^2}{J}(\overline{R}_H-\nu_I)^2. \tag{16.36}$$

Taking expectations, note that the final term involves $Var[\overline{R}_H]$, which is the sample variance of a mean of H observations taken without replacement from the I observed R_i's, hence with the finite sample correction,

$$Var[\overline{R}_H] = \frac{I-H}{I-1}\frac{Var[R_1]}{H}. \tag{16.37}$$

Thus

$$\begin{aligned} E\left[\sum_{i\in\mathcal{O}}(R_i-\overline{R}_{\mathcal{O}})^2\right] &= H\,Var[R_1] - \frac{H^2}{J}\frac{I-H}{I-1}\frac{Var[R_1]}{H} \\ &= H\left(1-\frac{I-H}{J(I-1)}\right)Var[R_1]. \end{aligned} \tag{16.38}$$

With (16.32), and some manipulations, we obtain

$$\begin{aligned} Var\left[\sum_{i=1}^{m}(R_i-\nu_I)(S_i-\nu_J)\right] &= H\left(1-\frac{I-H}{J(I-1)}\right)\frac{\sum_{i\in\mathcal{A}}(r_i-\nu_I)^2\sum_{i\in\mathcal{O}}(s_i-\nu_J)^2}{I}\frac{}{J-1} \\ &= \frac{H}{IJ}\left(1+\frac{H-1}{(I-1)(J-1)}\right)\sum_{i\in\mathcal{A}}(r_i-\nu_I)^2\sum_{i\in\mathcal{O}}(s_i-\nu_J)^2. \end{aligned} \tag{16.39}$$

Let

$$const = 4\left(\frac{(m+1)^2}{(I+1)(J+1)}\right)^2. \tag{16.40}$$

Then

Hypothesis	$Var[d^A_{Spear}(\boldsymbol{W},\boldsymbol{Z})]$
$\mathcal{H}_1$	$\frac{const}{J-1}\sum_{i\in\mathcal{O}}(r_i-\overline{r}_{\mathcal{O}})^2\sum_{i\in\mathcal{O}}(s_i-\nu_J)^2$
$\mathcal{H}_1'$	$\frac{const}{I-1}\sum_{i\in\mathcal{A}}(r_i-\nu_I)^2\sum_{i\in\mathcal{A}}(s_i-\overline{s}_{\mathcal{A}})^2$
$\mathcal{H}_2$	$\frac{const}{m-1}\sum_{i\in\mathcal{A}}(r_i-\nu_I)^2\sum_{i\in\mathcal{O}}(s_i-\nu_J)^2$
$\mathcal{H}_3$	$const\,\frac{H}{IJ}\left(1+\frac{H-1}{(I-1)(J-1)}\right)\sum_{i\in\mathcal{A}}(r_i-\nu_I)^2\sum_{i\in\mathcal{O}}(s_i-\nu_J)^2$

(16.41)

Note that for $\mathcal{H}_1$ and $\mathcal{H}_1'$, the variance depends on values of the observed ranks in the fixed vector corresponding to the observed ranks in the random vector (those with indices in $\mathcal{A}\cap\mathcal{O}$), while for $\mathcal{H}_3$ the variance depends on just the number of ranks observed in both vectors simultaneously, $\#(\mathcal{A}\cap\mathcal{O})$. For $\mathcal{H}_2$, the variance does not depend on the overlap at all.

Turn to asymptotic normality. Yu, Lam, & Alvo (2016) prove asymptotic normality under conditions for hypotheses $\mathcal{H}_1$ and $\mathcal{H}_2$. We present these results here, with slightly modified conditions, starting with $\mathcal{H}_1$. We do not have an asymptotic result for $\mathcal{H}_3$.

Theorem 16.2. *Suppose* $\boldsymbol{W} = \boldsymbol{w}$ *fixed and* $\boldsymbol{Z} \sim$ *Uniform*$(\mathcal{Z}_{\mathcal{O}})$, *and let* $I_x = \max\{I_k\}$ *and* $J_z = \max\{J_l\}$. *If*

$$\frac{\sum_{i\in\mathcal{O}}(r_i-\overline{r}_{\mathcal{O}})^2}{I^2}\frac{J-J_z}{J} \longrightarrow \infty, \quad \textit{where } \overline{r}_{\mathcal{O}} = \frac{1}{J}\sum_{i\in\mathcal{O}} r_i, \tag{16.42}$$

then

$$\frac{d^A_{Spear}(\boldsymbol{w}, \boldsymbol{Z}) - \mu_{Spear}(m)}{\sqrt{Var[d^A_{Spear}(\boldsymbol{w}, \boldsymbol{Z})]}} \longrightarrow^{\mathcal{D}} N(0,1), \tag{16.43}$$

where the variance is given in the $\mathcal{H}_1$ line of (16.41).

If there are no ties in the $\boldsymbol{w}$, then condition (16.42) is implied by

$$\frac{H^3}{I^2}\frac{J - J_z}{J} \longrightarrow \infty. \tag{16.44}$$

Alternatively, if the r_i for $i \in \mathcal{O}$ are asymptotically "similar" to the entire set of r_i's, in the sense that

$$\liminf \frac{\sum_{i\in\mathcal{O}}(r_i - \bar{r}_{\mathcal{O}})^2/J}{\sum_{i=1}^{m}(r_i - \nu_I)^2/m} > 0, \tag{16.45}$$

then condition (16.42) is implied by

$$\frac{(I - I_x)(J - J_z)}{m} \longrightarrow \infty. \tag{16.46}$$

We further note that a sufficient condition for (16.44) to hold is that $H \to \infty$, $\liminf H/I > 0$, and $\limsup J_z/J < 1$.

Now to H_2. The result is the same whether $\boldsymbol{W}$ is fixed and $\boldsymbol{Z}$ has the unconditional distribution, $\boldsymbol{Z}$ is fixed and $\boldsymbol{W}$ has the unconditional distribution, or they are independent and both have the unconditional distributions.

Theorem 16.3. *Suppose $\boldsymbol{W} = \boldsymbol{w}$ is fixed and $\boldsymbol{Z} \sim$ Uniform(Permutations($\boldsymbol{z}$)). Then*

$$\frac{(I - I_z)(J - J_x)}{m} \longrightarrow \infty \tag{16.47}$$

implies (16.43), where now the variance is given in the $\mathcal{H}_2$ line of (16.41).

Proof of Theorem 16.2. By (16.24) and (16.28), the asymptotic normality of Spearman's distance is equivalent to the asymptotic distribution of $\sum(r_i - \nu_I)(S_i - \nu_J)$. By (16.27), many of the summands are zero, hence we just need to prove the asymptotic normality of

$$\sum_{i\in\mathcal{O}}(r_i - \nu_I)(S_i - \nu_j), \quad \boldsymbol{S}_{\mathcal{O}} \sim \text{Uniform}(\text{Permutations}(\boldsymbol{s}_{\mathcal{O}})). \tag{16.48}$$

Now the Hoeffding (1951) permutation central limit theorem shows that (16.48) is asymptotically normal if

$$\frac{1}{J}\frac{\sum_{i\in\mathcal{O}}(r_i - \bar{r}_{\mathcal{O}})^2}{\max\{(r_i - \bar{r}_{\mathcal{O}})^2 \,|\, i \in \mathcal{O}\}}\frac{\sum_{i\in\mathcal{O}}(s_i - \nu_J)^2}{\max\{(s_i - \nu_J)^2 \,|\, i \in \mathcal{O}\}} \longrightarrow \infty. \tag{16.49}$$

See (16.30) for $\bar{r}_{\mathcal{O}}$.

As in (13.60), but for $\boldsymbol{J}$, the ratio based on the $\boldsymbol{s}$ is asymptotically equivalent to $J - J_z$. Since all the r_i are between 1 and I,

$$\max\{(r_i - \bar{r}_{\mathcal{O}})^2 \,|\, i \in \mathcal{O}\} \leqslant I^2. \tag{16.50}$$

Thus the term based on the $\boldsymbol{r}$ is bounded from below by $\sum_{i\in\mathcal{O}}(r_i-\bar{r}_{\mathcal{O}})^2/I^2$, implying (16.43).

For the special case that there are no ties in the $\boldsymbol{w}$, the $\sum_{i\in\mathcal{O}}(r_i-\bar{r}_{\mathcal{O}})^2$ is smallest if the H observed values of r_i for $i\in\mathcal{O}$ are bunched around $\nu_I=(I+1)/2$, in which case the sum of squares is asymptotically $H^3/12$. More precisely, if we let $a=\lfloor H/2\rfloor$, $b=\lceil H/2\rceil$, $\alpha=\lfloor\nu_I\rfloor$, and $\beta=\lceil\nu_I\rceil$, we have

$$\begin{aligned}\sum_{i\in\mathcal{A}\cap\mathcal{O}}(r_i-\nu_I)^2 &\geqslant \sum_{i=\alpha-a+1}^{\alpha-1}(i-\alpha)^2+\sum_{i=\beta+1}^{\beta+b-1}(i-\beta)^2\\ &=\sum_{j=1}^{a-1}j^2+\sum_{j=1}^{b-1}j^2\\ &=\frac{(a-1)a(2a-1)+(b-1)b(2b-1)}{6}\\ &\approx\frac{H^3}{12}.\end{aligned}\tag{16.51}$$

Thus (16.42) turns into (16.44).

Finally, suppose (16.45) holds, so that (16.42) holds if

$$\frac{\sum_{i=1}^m(r_i-\nu_I)^2}{I^2}\frac{J-J_z}{m}\longrightarrow\infty.\tag{16.52}$$

But $r_i-\nu_I=0$ if $i\notin\mathcal{A}$, and using (13.60) again shows that $\sum_{i\in\mathcal{A}}(r_i-\nu_I)^2/I^2$ is asymptotically equivalent to $I-I_x$. Thus (16.46) implies (16.42). □

Proof of Theorem 16.3. Using Hoeffding's theorem again, but on the entire vector, since $\boldsymbol{S}\sim$ Uniform(Permutations($\boldsymbol{s}$)), asymptotical normality follows from

$$\frac{1}{m}\frac{\sum_{i=1}^m(r_i-\nu_I)^2}{\max\{(r_i-\nu_I)^2\,|\,i=1,\ldots,m\}}\frac{\sum_{i=1}^m(s_i-\nu_J)^2}{\max\{(s_i-\nu_J)^2\,|\,i=1,\ldots,m\}}\longrightarrow\infty.\tag{16.53}$$

SInce $r_i-\nu_I=0$ if $i\notin\mathcal{A}$ and $s_i-\nu_J=0$ if $i\notin\mathcal{O}$, the two ratios in (16.53) are asymptotically equivalent to $I-I_x$ and $J-J_z$, respectively, yielding the theorem. □

16.4 Kendall

We start with the result for hypothesis $\mathcal{H}_1$, so that $\boldsymbol{W}=\boldsymbol{w}$ is fixed, and $\boldsymbol{Z}$ has the conditional distribution (16.23). This hypothesis fixes the observed indices for $\boldsymbol{Z}$, hence we can reorder the indices so that the observed ones for $\boldsymbol{Z}$ are the first J, i.e, $\mathcal{O}=\{1,\ldots,J\}$. Then Kendall's distance, minus its mean, adjusted for the incomplete data and possible ties, can be written

$$d^A_{Ken}(\boldsymbol{w},\boldsymbol{z})-\mu_{Ken}(m)=-2\sum_{1\leqslant i<j\leqslant m}\sum a_{ij}d_{ij},\tag{16.54}$$

where for $i \neq j$, using (16.6),

$$a_{ij} = \mathrm{E}[\mathrm{I}[X_i > X_j] \,|\, \boldsymbol{X} \in \mathrm{C}(\boldsymbol{w})] - \frac{1}{2},$$
$$d_{ij} = \mathrm{E}[\mathrm{I}[Y_i > Y_j] \,|\, \boldsymbol{Y} \in \mathrm{C}(\boldsymbol{z})] - \frac{1}{2} = \begin{cases} \mathrm{I}[z_i > z_j] + \frac{1}{2}\,\mathrm{I}[z_i = z_j] - \frac{1}{2} & \text{if } 1 \leqslant i < j \leqslant J \\ \frac{1}{J+1}(s_i - \nu_J) & \text{if } 1 \leqslant i \leqslant J < j \leqslant m\,, \\ 0 & \text{if } J < i < j \leqslant m \end{cases} \tag{16.55}$$

and we set $a_{ii} = d_{ii} = 0$. As in (16.25) and (16.26), we let $\boldsymbol{s}_{\mathcal{O}}$ denote $\mathrm{rank}(\boldsymbol{z}_{\mathcal{O}})$, the midranks of the observed elements of $\boldsymbol{z}$.

Thus we can write

$$\begin{aligned} d^{\mathrm{A}}_{\mathrm{Ken}}(\boldsymbol{w}, \boldsymbol{z}) - \mu_{\mathrm{Ken}}(m) &= -2\left(\sum_{1 \leqslant i < j \leqslant J} \sum a_{ij} d_{ij} + \frac{1}{J+1} \sum_{i=1}^{J} (s_i - \nu_J) \sum_{j=J+1}^{m} a_{ij}\right) \\ &= -2\left(\boldsymbol{a}^* \boldsymbol{d}^{*\prime} + \frac{1}{J+1}\, \boldsymbol{u}^{(2)}_{\mathcal{O}} (\boldsymbol{s}_{\mathcal{O}} - \nu_J \mathbf{1}_J)'\right). \end{aligned} \tag{16.56}$$

Here, $\boldsymbol{a}^*$ and $\boldsymbol{d}^*$ are the $1 \times \binom{J}{2}$ vectors with elements $1 \leqslant i < j \leqslant J$, arranged as

$$\boldsymbol{a}^* = (a_{12}, \ldots, a_{1J}, a_{23}, \ldots, a_{2J}, \ldots, a_{J-1,J}) \tag{16.57}$$

and $\boldsymbol{d}^*$ similarly; and we define $\boldsymbol{u}^{(1)}_{\mathcal{O}}$ and $\boldsymbol{u}^{(2)}_{\mathcal{O}}$ to be $1 \times J$ vectors, $\boldsymbol{u}^{(k)}_{\mathcal{O}} = (u^{(k)}_1, \ldots, u^{(k)}_J)$, with

$$u^{(1)}_i = \sum_{j=1}^{J} a_{ij} \;\text{ and }\; u^{(2)}_i = \sum_{j=J+1}^{m} a_{ij}. \tag{16.58}$$

Theorem 16.4. *Suppose hypothesis $\mathcal{H}_1$ holds, so that $\boldsymbol{W} = \boldsymbol{w}$ is fixed, and $\boldsymbol{Z}$ has the conditional distribution (16.23), $\boldsymbol{Z}_{\mathcal{O}} \sim$ Uniform(Permutations($\boldsymbol{z}_{\mathcal{O}}$)). Then*

$$\begin{aligned} \sigma^2_{Ken}(m) \equiv \mathrm{Var}[d^A_{Ken}(\boldsymbol{w}, \boldsymbol{Z})] = \frac{1}{3}\Bigg(&\frac{\omega_J}{J+1} \|\boldsymbol{r}^*_{\mathcal{O}} - \bar{r}^*_{\mathcal{O}} \mathbf{1}_J\|^2 + \|\boldsymbol{a}^*\|^2 (1 - 3\gamma_2 + 2\gamma_3) \\ &+ 3\,\|\boldsymbol{u}^{(1)}_{\mathcal{O}}\|^2 \frac{\gamma_2 - \gamma_3}{J+1} - \frac{\omega_J}{(J+1)^2}\, \|\boldsymbol{u}^{(2)}_{\mathcal{O}} - \bar{u}^{(2)}_{\mathcal{O}} \mathbf{1}_J\|^2\Bigg), \end{aligned} \tag{16.59}$$

where

$$\gamma_c = \sum_{b=1}^{L} \frac{(J_b)_c}{(J)_c}, \;\; \bar{r}^*_{\mathcal{O}} = \frac{\sum_{i=1}^{J} r^*_i}{J}, \;\; \bar{u}^{(2)}_{\mathcal{O}} = \frac{\sum_{i=1}^{J} u^{(2)}_i}{J}, \tag{16.60}$$

and

$$\omega_J = \frac{J^3 - \sum_{b=1}^{L} J_b^3}{J(J-1)} = 3(1-\gamma_2) + (J-2)(1-\gamma_3). \tag{16.61}$$

If there are no ties in the $\boldsymbol{z}_{\mathcal{O}}$, then

$$\sigma^2_{Ken}(m) = \frac{1}{3}\left(\|\boldsymbol{r}^*_{\mathcal{O}} - \bar{r}^*_{\mathcal{O}} \mathbf{1}_J\|^2 + \|\boldsymbol{a}^*\|^2 - \frac{1}{J+1}\, \|\boldsymbol{u}^{(2)}_{\mathcal{O}} - \bar{u}^{(2)}_{\mathcal{O}} \mathbf{1}_J\|^2\right). \tag{16.62}$$

The result for asymptotic normality is the same as for Spearman's distance in Theorem 16.2.

Theorem 16.5. *Consider the same conditions as in Theorem 16.4. If furthermore*

$$\frac{\sum_{i\in\mathcal{O}}(r_i-\bar{r}_{\mathcal{O}})^2}{I^2}\frac{J-J_z}{J}\longrightarrow\infty, \tag{16.63}$$

then

$$\frac{d^A_{Ken}(\boldsymbol{w},\boldsymbol{Z})-\mu_{Ken}(m)}{\sqrt{Var[d^A_{Ken}(\boldsymbol{w},\boldsymbol{Z})]}}\longrightarrow^{\mathcal{D}} N(0,1). \tag{16.64}$$

The conditions (16.44) through (16.46) for Spearman are thus also relevant for Kendall. The formulas for the variance for hypotheses $\mathcal{H}_2$ and $\mathcal{H}_3$ are too unwieldy to present in a few lines, so we give them Sections 16.4.3 and 16.4.2, respectively. The asymptotic normality for $\mathcal{H}_2$ is the same as for Spearman in Theorem 16.3.

Theorem 16.6. *Suppose $\boldsymbol{W}=\boldsymbol{w}$ is fixed and $\boldsymbol{Z}\sim$ Uniform(Permutations($\boldsymbol{z}$)). Then*

$$\frac{(I-I_z)(J-J_x)}{m}\longrightarrow\infty \tag{16.65}$$

implies (16.64), where now the variance is given in Section 16.4.3.

16.4.1 Proofs

Proof of Theorem 16.4. Recall that we are taking $\mathcal{O}=\{1,\ldots,J\}$, and note that Kendall's distance in (16.56) depends on just the first J values, $\boldsymbol{z}_{\mathcal{O}}$. Thus $\boldsymbol{d}^*$ and $\boldsymbol{s}_{\mathcal{O}}$ can be found using J-dimensional vectors. Let $\boldsymbol{Y}^*_{\mathcal{O}}\sim$ Uniform($\mathcal{P}_J$), and $C(\boldsymbol{z}_O)$ be the vectors in $\mathcal{P}_J$ compatible with $\boldsymbol{z}_{\mathcal{O}}$. Then

$$d_{ij}=E[I[Y^*_i>Y^*_j]\,|\,\boldsymbol{Y}^*_{\mathcal{O}}\in C(\boldsymbol{z}_{\mathcal{O}})]-\tfrac{1}{2},1\leqslant i,j\leqslant J,\ \text{ and }\ \boldsymbol{s}_{\mathcal{O}}=E[\boldsymbol{Y}^*_{\mathcal{O}}\,|\,\boldsymbol{Y}^*_{\mathcal{O}}\in C(\boldsymbol{z}_{\mathcal{O}})]. \tag{16.66}$$

We first find the variances and covariances for $\boldsymbol{D}^*$ and $\boldsymbol{S}_{\mathcal{O}}$.

Suppose the indices $1\leqslant i,j,k,l,\leqslant J$ are distinct. Now

$$\begin{aligned}Var[D_{ij}]&=Var[E[I[Y^*_i>Y^*_j]\,|\,\boldsymbol{Y}^*_{\mathcal{O}}\in C(\boldsymbol{z})]]\\&=Var[I[Y^*_i>Y^*_j]]-E[Var[I[Y^*_i>Y^*_j]\,|\,\boldsymbol{Y}^*_{\mathcal{O}}\in C(\boldsymbol{z})]].\end{aligned} \tag{16.67}$$

For fixed $\boldsymbol{z}$, if $z_i\neq z_j$, then $I[Y^*_i>Y^*_j]$ is fixed at $I[z_i>z_j]$ for any compatible $\boldsymbol{Y}^*_{\mathcal{O}}$. If $z_i=z_j$, then Y^*_i and Y^*_j are equally likely to be in either order. Thus

$$Var[I[Y^*_i>Y^*_j]\,|\,\boldsymbol{Y}^*_{\mathcal{O}}\in C(\boldsymbol{z})]=\frac{1}{4}I[z_i=z_j]. \tag{16.68}$$

Now $\gamma_2\equiv E[I[Z_i=Z_j]]$ in (16.60) is the chance both Z_i and Z_j are chosen from one of the groups defined by the ties, so that

$$Var[D_{ij}]=\frac{1}{4}(1-\gamma_2). \tag{16.69}$$

Similarly for $\mathrm{Cov}[D_{ij}, D_{ik}]$, where now at least one of $I[Y_i^* > Y_j^*]$ and $I[Y_i^* > Y_k^*]$ is conditionally fixed unless $Z_i = Z_j = Z_k$, hence

$$\mathrm{Cov}[D_{ij}, D_{ik}] = \frac{1}{12}(1-\gamma_3). \tag{16.70}$$

For $\mathrm{Cov}[D_{ij}, D_{kl}]$, at least one of $I[Y_i^* > Y_j^*]$ and $I[Y_k^* > Y_l^*]$ is conditionally fixed unless $Z_i = Z_j$ and $Z_k = Z_l$, in which case the conditional covariance is zero anyway. Let $\boldsymbol{\Gamma}_J$ be the $\binom{J}{2} \times J$, with rows indexed by $(ij), 1 \leqslant i < j \leqslant J$, ordered as in (16.57), and with elements

$$(\boldsymbol{\Gamma}_J)_{(ij)k} = \begin{cases} 1 & \text{if } i = k \\ -1 & \text{if } j = k \\ 0 & \text{otherwise} \end{cases}. \tag{16.71}$$

A little manipulation then shows that

$$\mathrm{Cov}[\boldsymbol{D}^*] = \frac{1}{12}\left(\boldsymbol{I}_{\binom{J}{2}}(1-3\gamma_2+2\gamma_3) + \boldsymbol{\Gamma}_J\boldsymbol{\Gamma}_J'(1-\gamma_3)\right). \tag{16.72}$$

For $\boldsymbol{S}_\mathcal{O}$, write

$$\begin{aligned} \sum_{j=1}^{J} d_{ij} &= \mathrm{E}\left[\sum_{1\leqslant j\leqslant J, j\neq i} I[Y_i^* > Y_j^*] \,|\, \boldsymbol{Y}_\mathcal{O}^* \in C(\boldsymbol{z}_\mathcal{O})\right] - \frac{J-1}{2} \\ &= \mathrm{E}\left[Y_i^* - 1 \,|\, \boldsymbol{Y}_\mathcal{O}^* \in C(\boldsymbol{z}_\mathcal{O})\right] - \frac{J-1}{2} \\ &= s_i - \nu_J, \end{aligned} \tag{16.73}$$

which, because $d_{ij} = -d_{ji}$, lets us write

$$\boldsymbol{S}_\mathcal{O} - \nu_J \mathbf{1}_J = \boldsymbol{D}^* \boldsymbol{\Gamma}_J. \tag{16.74}$$

Thus

$$\begin{aligned} \mathrm{Cov}[\boldsymbol{D}^*, \boldsymbol{S}_\mathcal{O}] = \mathrm{Cov}[\boldsymbol{D}^*]\,\boldsymbol{\Gamma}_J &= \frac{1}{12}\left(\boldsymbol{\Gamma}_J(1-3\gamma_2+2\gamma_3) + \boldsymbol{\Gamma}_J\boldsymbol{\Gamma}_J'\boldsymbol{\Gamma}_J(1-\gamma_3)\right) \\ &= \frac{\omega_J}{12}\boldsymbol{\Gamma}_J, \end{aligned} \tag{16.75}$$

since $\boldsymbol{\Gamma}_J'\boldsymbol{\Gamma}_J = J\boldsymbol{H}_J$ and $\boldsymbol{\Gamma}_J\boldsymbol{H}_J = \boldsymbol{\Gamma}_J$, and we can show that

$$1-3\gamma_2+2\gamma_3+J(1-\gamma_3) = \frac{J^3 - \sum_{b=1}^{L} J_b^3}{J(J-1)} = \omega_J \tag{16.76}$$

in (16.61). We now can express the covariance of $\boldsymbol{S}_O$ as

$$\mathrm{Cov}[\boldsymbol{S}_\mathcal{O}] = \boldsymbol{\Gamma}_J'\,\mathrm{Cov}[\boldsymbol{D}^*]\,\boldsymbol{\Gamma}_J = \frac{J\omega_J}{12}\boldsymbol{H}_J. \tag{16.77}$$

Then using (16.56), we have

$$\begin{aligned}\sigma^2_{\text{Ken}}(m) &= 4\left(\boldsymbol{a}^*\text{Cov}[\boldsymbol{D}^*]\boldsymbol{a}^{*\prime} + \frac{2}{J+1}\,\boldsymbol{a}^*\text{Cov}[\boldsymbol{D}^*,\boldsymbol{S}_{\mathcal{O}}]\boldsymbol{u}^{(2)\prime} + \frac{1}{(J+1)^2}\,\boldsymbol{u}^{(2)}\text{Cov}[\boldsymbol{S}_{\mathcal{O}}]\boldsymbol{u}^{(2)\prime}\right)\\ &= \frac{1}{3}\left(\|\boldsymbol{a}^*\|^2(1-3\gamma_2+2\gamma_3) + \boldsymbol{a}^*\boldsymbol{\Gamma}_J\boldsymbol{\Gamma}_J'\boldsymbol{a}^{*\prime}(1-\gamma_3)\right.\\ &\qquad\left. + \frac{2\omega_J}{J+1}\,\boldsymbol{a}^*\boldsymbol{\Gamma}_J\boldsymbol{u}_{\mathcal{O}}^{(2)\prime} + \frac{J\omega_J}{(J+1)^2}\,\boldsymbol{u}_{\mathcal{O}}^{(2)}\boldsymbol{H}_J\boldsymbol{u}_{\mathcal{O}}^{(2)\prime}\right)\\ &= \frac{1}{3}\left(\|\boldsymbol{a}^*\|^2(1-3\gamma_2+2\gamma_3) + \|\boldsymbol{u}_{\mathcal{O}}^{(1)}\|^2(1-\gamma_3)\right.\\ &\qquad\left. + \frac{2\omega_J}{J+1}\,\boldsymbol{u}_{\mathcal{O}}^{(1)}\boldsymbol{u}_{\mathcal{O}}^{(2)\prime} + \frac{J\omega_J}{(J+1)^2}\,\boldsymbol{u}_{\mathcal{O}}^{(2)}\boldsymbol{H}_J\boldsymbol{u}_{\mathcal{O}}^{(2)\prime}\right).\end{aligned} \tag{16.78}$$

since by (16.58), $\boldsymbol{u}_{\mathcal{O}}^{(1)} = \boldsymbol{a}^*\boldsymbol{\Gamma}_J$. We also have, for $1 \leqslant i \leqslant J$,

$$\begin{aligned}u_i^{(1)} + u_i^{(2)} = \sum_{j=1}^{m} a_{ij} &= E\left[\sum_{j\neq i} I[X_i > X_j] \mid \boldsymbol{X} \in C(\boldsymbol{w})\right] - \frac{m-1}{2}\\ &= E\left[X_i - 1 \mid \boldsymbol{X} \in C(\boldsymbol{w})\right] - \frac{m-1}{2}\\ &= r_i^* - \nu_m.\end{aligned} \tag{16.79}$$

Since $\boldsymbol{\Gamma}_J\boldsymbol{H}_J = \boldsymbol{\Gamma}_J$, the sample mean of the $u_i^{(1)}$ is zero, hence

$$\|\boldsymbol{u}_{\mathcal{O}}^{(1)}\|^2 + 2\boldsymbol{u}_{\mathcal{O}}^{(1)}\boldsymbol{u}_{\mathcal{O}}^{(2)\prime} + \boldsymbol{u}_{\mathcal{O}}^{(2)}\boldsymbol{H}_J\boldsymbol{u}_{\mathcal{O}}^{(2)\prime} = (\boldsymbol{u}_{\mathcal{O}}^{(1)} + \boldsymbol{u}_{\mathcal{O}}^{(2)})\boldsymbol{H}_J(\boldsymbol{u}_{\mathcal{O}}^{(1)} + \boldsymbol{u}_{\mathcal{O}}^{(2)})' = \|\boldsymbol{r}_{\mathcal{O}}^* - \bar{r}_{\mathcal{O}}^*\mathbf{1}_J\|^2. \tag{16.80}$$

Now we add and subtract $\omega_J/(J+1)$ to the coefficients of $\|\boldsymbol{u}^{(1)}\|^2$ and $\boldsymbol{u}_{\mathcal{O}}^{(2)}\boldsymbol{H}_J\boldsymbol{u}_{\mathcal{O}}^{(2)\prime}$ in (16.78), and use (16.76) to obtain (16.59).

If there are no ties in $\boldsymbol{Z}_{\mathcal{O}}$, then $\gamma_2 = \gamma_3 = 0$, and since $J_b \equiv 1$, $\omega_J = J+1$, yielding (16.62). □

Proof of Theorem 16.5. We proceed as we did for complete rankings, obtaining the asymptotic result for Kendall from that for Spearman, where we used Lemma 15.2. Thus we first find the covariance of Kendall and Spearman. Using (16.24), and noting that $s_i^* = (m+1)s_i/(J+1)$, we can write

$$d^A_{\text{Spear}}(\boldsymbol{w},\boldsymbol{z}) - \mu_{\text{Spear}}(m) = -2\,\frac{m+1}{J+1}\,(\boldsymbol{r}_{\mathcal{O}}^* - \nu\mathbf{1}_J)(\boldsymbol{s}_{\mathcal{O}} - \nu_J\mathbf{1}_J). \tag{16.81}$$

Thus with Kendall's distance as in (16.56), we have that the covariance is

$$\begin{aligned}\mathrm{Cov}[\mathrm{d}^{\mathrm{A}}_{\mathrm{Spear}}(\boldsymbol{w},\boldsymbol{Z}),\mathrm{d}^{\mathrm{A}}_{\mathrm{Ken}}(\boldsymbol{w},\boldsymbol{Z})] &= 4\,\frac{m+1}{J+1}\Big(\mathrm{Cov}[\boldsymbol{a}^*\boldsymbol{D}^{*\prime},(\boldsymbol{r}^*_{\mathcal{O}}-\nu\mathbf{1}_J)\boldsymbol{S}_{\mathcal{O}}]\\ &\qquad +\frac{1}{J+1}\,\mathrm{Cov}[\boldsymbol{u}^{(2)}_{\mathcal{O}}\boldsymbol{S}_{\mathcal{O}},(\boldsymbol{r}^*_{\mathcal{O}}-\nu\mathbf{1}_J)\boldsymbol{S}_{\mathcal{O}}]\Big)\\ &= \frac{\omega_J}{3}\frac{m+1}{J+1}\left(\boldsymbol{a}^*\boldsymbol{\Gamma}_J(\boldsymbol{r}^*_{\mathcal{O}}-\nu\mathbf{1}_J)'+\frac{J}{J+1}\boldsymbol{u}^{(2)}_{\mathcal{O}}\boldsymbol{H}_J(\boldsymbol{r}^*_{\mathcal{O}}-\nu\mathbf{1}_J)\right)\\ &= \frac{\omega_J}{3}\frac{m+1}{J+1}\left(\boldsymbol{u}^{(1)}_{\mathcal{O}}(\boldsymbol{r}^*_{\mathcal{O}}-\bar{r}^*_{\mathcal{O}}\mathbf{1}_J)'+\frac{J}{J+1}\boldsymbol{u}^{(2)}_{\mathcal{O}}\boldsymbol{H}_J(\boldsymbol{r}^*_{\mathcal{O}}-\nu\mathbf{1}_J)\right)\\ &= \frac{\omega_J}{3}\frac{m+1}{J+1}\left(\|\boldsymbol{r}^*_{\mathcal{O}}-\bar{r}^*_{\mathcal{O}}\mathbf{1}_J\|^2-\frac{1}{J+1}\boldsymbol{u}^{(2)}_{\mathcal{O}}\boldsymbol{H}_J(\boldsymbol{r}^*_{\mathcal{O}}-\nu\mathbf{1}_J)\right). \quad (16.82)\end{aligned}$$

From (16.41), we have

$$\begin{aligned}\sigma^2_{\mathrm{Spear}}(m) &= \frac{4}{J-1}\frac{(m+1)^2}{(J+1)^2}\,\|\boldsymbol{r}^*_{\mathcal{O}}-\bar{r}^*_{\mathcal{O}}\mathbf{1}_J\|^2\|\boldsymbol{s}_{\mathcal{O}}-\nu_J\mathbf{1}_J\|^2\\ &= \frac{\omega_J}{3}\frac{J(m+1)^2}{(J+1)^2}\,\|\boldsymbol{r}^*_{\mathcal{O}}-\bar{r}^*_{\mathcal{O}}\mathbf{1}_J\|^2. \quad (16.83)\end{aligned}$$

For constant κ_J, the above with (16.59) gives

$$\begin{aligned}\Delta_m &\equiv \mathrm{E}\left[\Big(\kappa_J(\mathrm{d}^{\mathrm{A}}_{\mathrm{Spear}}(\boldsymbol{w},\boldsymbol{Z})-\mu_{\mathrm{Spear}}(m))-(\mathrm{d}^{\mathrm{A}}_{\mathrm{Ken}}(\boldsymbol{w},\boldsymbol{Z})-\mu_{\mathrm{Ken}}(m))\Big)^2\right]\\ &= \kappa_J^2\frac{\omega_J}{3}\frac{J(m+1)^2}{(J+1)^2}\,\|\boldsymbol{r}^*_{\mathcal{O}}-\bar{r}^*_{\mathcal{O}}\mathbf{1}_J\|^2-2\kappa_J\frac{\omega_J}{3}\frac{m+1}{J+1}\left(\|\boldsymbol{r}^*_{\mathcal{O}}-\bar{r}^*_{\mathcal{O}}\mathbf{1}_J\|^2-\frac{1}{J+1}\boldsymbol{u}^{(2)}_{\mathcal{O}}\boldsymbol{H}_J(\boldsymbol{r}^*_{\mathcal{O}}-\nu\mathbf{1}_J)\right)\\ &\qquad +\frac{1}{3}\left(\|\boldsymbol{a}^*\|^2(1-3\gamma_2+2\gamma_3)+3\,\|\boldsymbol{u}^{(1)}_{\mathcal{O}}\|^2\frac{\gamma_2-\gamma_3}{J+1}\right.\\ &\qquad\qquad \left.+\frac{\omega_J}{J+1}\left(\|\boldsymbol{r}^*_{\mathcal{O}}-\bar{r}^*_{\mathcal{O}}\mathbf{1}_J\|^2-\frac{1}{J+1}\,\boldsymbol{u}^{(2)}_{\mathcal{O}}\boldsymbol{H}_J\boldsymbol{u}^{(2)\prime}_{\mathcal{O}}\right)\right). \quad (16.84)\end{aligned}$$

We collect the coefficients of $\|\boldsymbol{r}^*_{\mathcal{O}}-\bar{r}^*_{\mathcal{O}}\mathbf{1}_J\|^2$:

$$\kappa_J^2\frac{\omega_J}{3}\frac{J(m+1)^2}{(J+1)^2}-2\kappa_J\frac{\omega_J}{3}\frac{m+1}{J+1}+\frac{\omega_J}{3}\frac{1}{J+1}=\frac{\omega_J}{3}\frac{1}{J+1}\left(\kappa_J^2\frac{J(m+1)^2}{J+1}-2\kappa_J(m+1)+1\right). \quad (16.85)$$

We can set the coefficient to zero by taking

$$\kappa_J=\frac{1}{m+1}\frac{J+1}{J}\left(1-\frac{1}{\sqrt{J}}\right). \quad (16.86)$$

With that choice,

$$\begin{aligned}\Delta_m &= \frac{2}{3}\left(1-\frac{1}{\sqrt{J}}\right)\frac{\omega_J}{J(J+1)}\boldsymbol{u}^{(2)}_{\mathcal{O}}(\boldsymbol{r}^*_{\mathcal{O}}-\bar{r}^*_{\mathcal{O}}\mathbf{1}_J)'+\frac{1}{3}\,\|\boldsymbol{a}^*\|^2(1-3\gamma_2+2\gamma_3)\\ &\qquad +\|\boldsymbol{u}^{(1)}_{\mathcal{O}}\|^2\frac{\gamma_2-\gamma_3}{J+1}-\frac{1}{3}\,\frac{\omega_J}{(J+1)^2}\boldsymbol{u}^{(2)}_{\mathcal{O}}\boldsymbol{H}_J\boldsymbol{u}^{(2)\prime}_{\mathcal{O}}. \quad (16.87)\end{aligned}$$

Let

$$\beta_m = \kappa_J \sigma_{\text{Spear}}(m) \text{ and } \alpha_m = \frac{\sigma_{\text{Ken}}(m)}{\beta_m}, \tag{16.88}$$

so that by (16.84), we obtain

$$\frac{\Delta_m}{\beta_m^2} = E\left[\left(\frac{d^A_{\text{Spear}}(\boldsymbol{w},\boldsymbol{Z}) - \mu_{\text{Spear}}(m)}{\sigma_{\text{Spear}}(m)} - \alpha_m \frac{d^A_{\text{Ken}}(\boldsymbol{w},\boldsymbol{Z}) - \mu_{\text{Ken}}(m)}{\sigma_{\text{Ken}}(m)}\right)^2\right]. \tag{16.89}$$

Then using (16.86) and (16.83),

$$\begin{aligned}\beta_m^2 &= \frac{1}{(m+1)^2}\frac{(J+1)^2}{J^2}\left(1-\frac{1}{\sqrt{J}}\right)^2 \frac{\omega_J}{3}\frac{J(m+1)^2}{(J+1)^2}\|\boldsymbol{r}^*_{\mathcal{O}} - \bar{r}^*_{\mathcal{O}}\mathbf{1}_J\|^2 \\ &= \frac{1}{3}\left(1-\frac{1}{\sqrt{J}}\right)^2 \|\boldsymbol{r}^*_{\mathcal{O}} - \bar{r}^*_{\mathcal{O}}\mathbf{1}_J\|^2 \frac{J^3 - \sum_{b=1}^{L} J_b^3}{J^2(J-1)}\end{aligned} \tag{16.90}$$

As in (13.59),

$$J^2(J-J_z) \leqslant J^3 - \sum_{b=1}^{L} J_b^3 \leqslant 3\,J^2(J-J_z). \tag{16.91}$$

By assumption (16.66), since $r_i^* = (m+1)r_i/(I+1)$,

$$\frac{\|\boldsymbol{r}^*_{\mathcal{O}} - \bar{r}^*_{\mathcal{O}}\mathbf{1}_J\|^2}{m^2}\,\frac{J-J_z}{J} \longrightarrow \infty. \tag{16.92}$$

Then

$$\frac{\beta_m^2}{m^2} \longrightarrow \infty. \tag{16.93}$$

We wish to show that Δ_m in (16.87) is of order m^2. By (16.69), we have $0 \leqslant \gamma_c \leqslant 1$, hence by (16.76), $\omega_J = 3(1-\gamma_2) + (J-2)(1-\gamma_3) \leqslant J+1$. By (16.55), we have that each $|a_{ij}| \leqslant \frac{1}{2}$. Now $\boldsymbol{a}^*$ has $\binom{J}{2}$ of the a_{ij}'s, and $u_i^{(1)}$, $\boldsymbol{u}^{(2)}_{\mathcal{O}}$, and r_i^* are sums of, respectively, $J, m-J$, and m of them, and the $\boldsymbol{u}^{(k)}_{\mathcal{O}}$ and $\boldsymbol{r}^*_{\mathcal{O}}$ have J elements, hence

$$\begin{aligned}\|\boldsymbol{a}^*\|^2 &\leqslant \frac{J(J-1)}{8}, \\ \|\boldsymbol{u}^{(1)}_{\mathcal{O}}\|^2 &\leqslant \frac{J^3}{4}, \\ \boldsymbol{u}^{(2)}_{\mathcal{O}}\boldsymbol{H}_J\boldsymbol{u}^{(2)\prime}_{\mathcal{O}} &\leqslant \frac{J(m-J)^2}{4}, \text{ and} \\ \boldsymbol{u}^{(2)}_{\mathcal{O}}(\boldsymbol{r}^*_{\mathcal{O}} - \bar{r}^*_{\mathcal{O}}\mathbf{1}_J)' &\leqslant \frac{Jm(m-J)}{4}.\end{aligned} \tag{16.94}$$

Thus

$$\begin{aligned}\Delta_m &\leqslant \frac{2}{3J}\,\boldsymbol{u}^{(2)}_{\mathcal{O}}(\boldsymbol{r}^*_{\mathcal{O}} - \bar{r}^*_{\mathcal{O}}\mathbf{1}_J)' + 2\,\|\boldsymbol{a}^*\|^2 + \frac{2}{J+1}\|\boldsymbol{u}^{(1)}_{\mathcal{O}}\|^2 + \frac{1}{3}\,\frac{1}{J+1}\boldsymbol{u}^{(2)}_{\mathcal{O}}\boldsymbol{H}_J\boldsymbol{u}^{(2)\prime}_{\mathcal{O}} \\ &\leqslant \frac{m(m-J)}{6} + \frac{J(J-1)}{4} + \frac{J^2}{2} + \frac{(m-J)^2}{12} \\ &\leqslant m^2.\end{aligned} \tag{16.95}$$

Hence by (16.93), $\Delta_m/\beta_m^2 \to 0$. Theorem 16.2 states that (16.92) implies normality for Spearman's distance, hence (16.89) shows that

$$\alpha_m \frac{d^A_{\text{Ken}}(\boldsymbol{w}, \boldsymbol{Z}) - \mu_{\text{Ken}}(m)}{\sigma_{\text{Ken}}(m)} \longrightarrow^{\mathcal{D}} N(0,1). \tag{16.96}$$

Now using the inequalities in (16.94) and above, and (16.83), we can write

$$\sigma^2_{\text{Ken}}(m) = \frac{J+1}{J(m+1)^2}\, \sigma^2_{\text{Spear}}(m) + \delta_m, \tag{16.97}$$

where

$$\begin{aligned}\delta_m &= \frac{1}{3}\left(\|\boldsymbol{a}^*\|^2(1-3\gamma_2+2\gamma_3) + 3\,\|\boldsymbol{u}_{\mathcal{O}}^{(1)}\|^2 \frac{\gamma_2-\gamma_3}{J+1} - \frac{\omega_J}{(J+1)^2}\, \boldsymbol{u}_{\mathcal{O}}^{(2)} \boldsymbol{H}_J \boldsymbol{u}_{\mathcal{O}}^{(2)\prime} \right) \\ &\leqslant \frac{J(J-1)}{4} + \frac{J^2}{2} + \frac{(m-J)^2}{12} \\ &\leqslant m^2. \end{aligned} \tag{16.98}$$

Thus

$$\begin{aligned}\alpha_m^2 = \frac{\sigma^2_{\text{Ken}}(m)}{\beta_m^2} &= \frac{J+1}{J(m+1)^2}\, \frac{1}{\kappa_J^2} + \frac{\delta_m}{\beta_m^2} \\ &= \frac{J}{J+1}\left(1 - \frac{1}{\sqrt{J}}\right)^{-2} + \frac{\delta_m}{\beta_m^2} \longrightarrow 1. \end{aligned} \tag{16.99}$$

The conclusion (16.64) then follows from (16.97). □

Proof of Theorem 16.6. Consider (15.20), but with $\boldsymbol{w}$ replaced by $\boldsymbol{X}$, so that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent and Uniform($\mathcal{P}_m$). Then the bound on the right-hand side is $m(m-1)/12$. Using Jensen's inequality as in (15.21), we can replace the distances with those in the theorem, thus achieving the same bound. As we did for the $\mathcal{H}_1$ case in (16.88), we set

$$\beta_m = \kappa_m \sigma_{\text{Spear}}(m) \;\text{ and }\; \alpha_m = \frac{\sigma_{\text{Ken}}(m)}{\beta_m}, \;\text{ where }\; \kappa_m = \frac{1}{m+1+\sqrt{m+1}}. \tag{16.100}$$

Then to show that the asymptotic normality of Spearman's distance under assumption (16.47) given in Theorem 16.3 implies the same for Kendall, we need to show that

$$\frac{\beta_m^2}{m^2} \longrightarrow \infty \;\text{ and }\; \alpha_m \longrightarrow 1. \tag{16.101}$$

From (16.41) for $\mathcal{H}_2$, we have

$$\beta_m^2 = \frac{4}{(m+1+\sqrt{m+1})^2}\left(\frac{(m+1)^2}{(I+1)(J+1)}\right)^2 \frac{1}{m-1} \sum_{i\in\mathcal{A}} (r_i - \nu_I)^2 \sum_{i\in\mathcal{O}} (s_i - \nu_J)^2. \tag{16.102}$$

□

16.4.2 The variance for Kendall under $\mathcal{H}_3$

Suppose both $\boldsymbol{W}$ and $\boldsymbol{Z}$ have the conditional distribution, so that $\mathcal{A}$ and $\mathcal{O}$ are fixed, $\boldsymbol{W}_{\mathcal{A}} \sim$ Uniform(Permutations($\boldsymbol{w}_{\mathcal{A}}$)), $\boldsymbol{Z}_{\mathcal{O}} \sim$ Uniform(Permutations($\boldsymbol{z}_{\mathcal{O}}$)), and $\boldsymbol{W}_{\mathcal{A}}$ and $\boldsymbol{Z}_{\mathcal{O}}$ are independent. We know the variance if $\boldsymbol{W}$ is fixed, and

$$\begin{aligned}\mathrm{Var}[d^{A}_{\mathrm{Ken}}(\boldsymbol{W},\boldsymbol{Z})] &= \mathrm{Var}[\mathrm{E}[d^{A}_{\mathrm{Ken}}(\boldsymbol{W},\boldsymbol{Z}) \,|\, \boldsymbol{W}=\boldsymbol{w}]] + \mathrm{E}[\mathrm{Var}[d^{A}_{\mathrm{Ken}}(\boldsymbol{W},\boldsymbol{Z}) \,|\, \boldsymbol{W}=\boldsymbol{w}]] \\ &= \mathrm{E}[\mathrm{Var}[d^{A}_{\mathrm{Ken}}(\boldsymbol{w},\boldsymbol{Z})],\end{aligned} \tag{16.103}$$

since the conditional expected value is a constant. Thus by (16.59), we need the expected value over $\boldsymbol{Z}$ of the quantities $\|\boldsymbol{R}^*_{\mathcal{O}} - \overline{R}^*_{\mathcal{O}}\mathbf{1}_J\|^2$, $\boldsymbol{A}^*$, $\|\boldsymbol{U}^{(1)}_{\mathcal{O}}\|^2$, and $\|\boldsymbol{U}^{(2)}_{\mathcal{O}} - \overline{U}^{(2)}_{\mathcal{O}}\mathbf{1}_J\|^2$. We will take $\mathcal{O} = \{1,\ldots,J\}$, and as in (16.33), the first H of the w_i's to be the observed rankings among the first J.

Using (16.38), since $r^*_i = (m+1)r_i/(I+1)$,

$$\mathrm{E}\left[\|\boldsymbol{R}^*_{\mathcal{O}} - \overline{R}^*_{\mathcal{O}}\mathbf{1}_J\|^2\right] = H\,\frac{(m+1)^2}{(I+1)^2}\left(1 - \frac{I-H}{J(I-1)}\right)\mathrm{Var}[R_a]. \tag{16.104}$$

Here, R_a is any R_i with $i \in \mathcal{A}$. As in (16.55), $\mathrm{E}[A^2_{ij}]$ depends on which of i and j are observed. Using (16.69) for the first possibility, we have

$$\mathrm{E}[A^2_{ij}] = \begin{cases} \frac{1}{4}(1-\overline{\gamma}_2) & \text{if } i,j \in \mathcal{A}, \\ \frac{1}{(I+1)^2}\,\mathrm{Var}[R_a] & \text{if } i \in \mathcal{A}, j \notin \mathcal{A} \text{ or } i \notin \mathcal{A}, j \in \mathcal{A} \\ 0 & \text{if } i,j \notin \mathcal{A}. \end{cases}$$

where as in (16.60), $\overline{\gamma}_c = \sum_{a=1}^{K}(I_a)_c/(I)_c$. Since in $\boldsymbol{A}^*$ there are H observed r_i's and $J-H$ unobserved, and $1 \leqslant i < j \leqslant J$,

$$\mathrm{E}[\|\boldsymbol{A}^*\|^2] = \frac{H(H-1)}{8}\,(1-\overline{\gamma}_2) + \frac{H(J-H)}{(I+1)^2}\,\mathrm{Var}[R_a]. \tag{16.105}$$

The $\boldsymbol{u}^{(1)}$ is defined in (16.58). Since the first H rankings are observed, and the next $J-H$ are not,

$$u^{(1)}_i = \sum_{j=1}^{J} a_{ij} = \begin{cases} \sum_{j=1}^{H} a_{ij} + (J-H)\left(\frac{r_i}{I+1} - \frac{1}{2}\right) & \text{if } 1 \leqslant i \leqslant H, \\ -\sum_{j=1}^{H}\left(\frac{r_j}{I+1} - \frac{1}{2}\right) & \text{if } H < i \leqslant J. \end{cases} \tag{16.106}$$

Since all the individual terms have mean zero, and the $U^{(1)}_1, \ldots, U^{(1)}_H$ are identically distributed,

$$\begin{aligned}\mathrm{E}\left[\|\boldsymbol{U}^{(1)}_{\mathcal{O}}\|^2\right] &= H\,\mathrm{Var}\left[\sum_{j=2}^{H} A_{1j}\right] + H(J-H)^2\,\mathrm{Var}\left[\frac{R_1}{I+1} - \frac{1}{2}\right] \\ &\quad + 2\,H(J-H)\,\mathrm{Cov}\left[\sum_{j=2}^{H} A_{1j}, \frac{R_1}{I+1} - \frac{1}{2}\right] + (J-H)\mathrm{Var}\left[\sum_{j=1}^{H}\left(\frac{R_j}{I+1} - \frac{1}{2}\right)\right].\end{aligned} \tag{16.107}$$

Let $\boldsymbol{A}^\dagger$ be the $1 \times \binom{I}{2}$ vector with elements A_{ij}, $i, j \in \mathcal{A}$ and $i < j$, arranged in a vector like (16.57). Then as in (16.72),

$$\mathrm{Cov}[\boldsymbol{A}^\dagger] = \frac{1}{12}\left(\boldsymbol{I}_{\binom{I}{2}}(1-3\overline{\gamma}_2+2\overline{\gamma}_3) + \boldsymbol{\Gamma}_I\boldsymbol{\Gamma}_I'(1-\overline{\gamma}_3)\right), \tag{16.108}$$

and

$$\begin{aligned}\mathrm{Var}\left[\sum_{j=2}^{H} A_{1j}\right] &= \frac{1}{12}\left(\mathbf{1}_{H-1}, \mathbf{0}_{\binom{I}{2}-H+1}\right)\mathrm{Cov}[\boldsymbol{A}^\dagger]\left(\mathbf{1}_{H-1}, \mathbf{0}_{\binom{I}{2}-H+1}\right)' \\ &= \frac{1}{12}\left((H-1)(1-3\overline{\gamma}_2+2\overline{\gamma}_3) + H(H-1)(1-\overline{\gamma}_3)\right).\end{aligned} \tag{16.109}$$

Using (16.37), we have

$$\mathrm{Var}\left[\sum_{j=1}^{H}\left(\frac{R_j}{I+1}-\frac{1}{2}\right)\right] = \frac{H^2}{(I+1)^2}\mathrm{Var}[\overline{R}_H] = \frac{H(I-H)}{(I+1)^2(I-1)}\mathrm{Var}[R_1]. \tag{16.110}$$

For the covariance, as in (16.75),

$$\mathrm{Cov}[\boldsymbol{D}^\dagger, \boldsymbol{R}_{\mathcal{A}}] = \frac{\overline{\omega}_I}{12}\boldsymbol{\Gamma}_I = \frac{\mathrm{Var}[R_a]}{I-1}\boldsymbol{\Gamma}_I, \quad \overline{\omega}_I = 1-3\overline{\gamma}_2+2\overline{\gamma}_3+I(1-\overline{\gamma}_3), \tag{16.111}$$

hence,

$$\mathrm{Cov}\left[\sum_{j=2}^{H} A_{1j}, \frac{R_1}{I+1}-\frac{1}{2}\right] = \frac{H-1}{(I+1)(I-1)}\mathrm{Var}[R_1]. \tag{16.112}$$

Using (16.109) through (16.112) in (16.107), and simplifying, we obtain

$$\begin{aligned}\mathrm{E}\left[\|\boldsymbol{U}_{\mathcal{O}}^{(1)}\|^2\right] &= \frac{H(H-1)}{12}(3(1-\overline{\gamma}_2)+(H-2)(1-\overline{\gamma}_3)) \\ &\quad + \frac{H(J-H)(H(I+2)+(I-1)(J-1)-3)}{(I+1)^2(I-1)}\mathrm{Var}[R_1].\end{aligned} \tag{16.113}$$

Turn to $\boldsymbol{U}_{\mathcal{O}}^{(2)}$, whose elements are given in (16.58). We will take the elements of $\boldsymbol{w}$ to be arranged so that $\mathcal{A}\cap\mathcal{O} = \{1,\ldots,H\}$ as before, and $\mathcal{A}\cap\mathcal{O}^c = \{J+1,\ldots,J+\overline{H}\}$, where $\overline{H} = I-H$. Also, let $\overline{J} = m-J$. Then similar to (16.106), we have

$$u_i^{(2)} = \sum_{j=J+1}^{m} a_{ij} = \begin{cases}\sum_{j=J+1}^{J+\overline{H}} a_{ij} + (\overline{J}-\overline{H})\left(\frac{r_i}{I+1}-\frac{1}{2}\right) & \text{if } 1 \leqslant i \leqslant H, \\ -\sum_{j=J+1}^{J+\overline{H}}\left(\frac{r_j}{I+1}-\frac{1}{2}\right) & \text{if } H < i \leqslant J.\end{cases} \tag{16.114}$$

Thus

$$\begin{aligned}\mathrm{E}[\|\boldsymbol{U}_{\mathcal{O}}^{(2)}\|^2 &= H\,\mathrm{Var}\left[\sum_{j=J+1}^{J+\overline{H}} A_{1j}\right] + H(\overline{J}-\overline{H})^2\,\mathrm{Var}\left[\frac{R_1}{I+1}-\frac{1}{2}\right] \\ &\quad + 2\,H(\overline{J}-\overline{H})\,\mathrm{Cov}\left[\sum_{j=J+1}^{J+\overline{H}} A_{1j}, \frac{R_1}{I+1}-\frac{1}{2}\right] + (J-H)\mathrm{Var}\left[\sum_{j=J+1}^{J+\overline{H}}\left(\frac{R_j}{I+1}-\frac{1}{2}\right)\right].\end{aligned} \tag{16.115}$$

As in (16.109), but with $\overline{H}$ instead of $H-1$,

$$\mathrm{Var}\left[\sum_{j=J+1}^{J+\overline{H}} A_{1j}\right] = \frac{\overline{H}}{12}\,(\overline{H}+2-3\overline{\gamma}_2-(\overline{H}-1)\overline{\gamma}_3). \tag{16.116}$$

Likewise for the covariance as in (16.112),

$$\mathrm{Cov}\left[\sum_{j=J+1}^{J+\overline{H}} A_{1j}, \frac{R_1}{I+1}-\frac{1}{2}\right] = \frac{\overline{H}}{(I+1)(I-1)}\,\mathrm{Var}[R_1]. \tag{16.117}$$

The final variance on the right-hand side of (16.115) is the same as, but with $\overline{H}$ in place of H. Thus we now have

$$\begin{aligned}\mathrm{E}[\|\boldsymbol{U}_{\mathcal{O}}^{(2)}\|^2] = {}& \frac{H\overline{H}\,(\overline{H}+2-3\overline{\gamma}_2-(\overline{H}-1)\overline{\gamma}_3)}{12} \\ & + \frac{H}{I+1}\left(\frac{(\bar{J}-\overline{H})^2}{I+1}+2\,\frac{(\bar{J}-\overline{H})\overline{H}}{I-1}+\frac{(J-H)\overline{H}}{(I+1)(I-1)}\right)\mathrm{Var}[R_1].\end{aligned} \tag{16.118}$$

Finally,

$$\|\boldsymbol{U}_{\mathcal{O}}^{(2)}-\overline{U}_{\mathcal{O}}^{(2)}\mathbf{1}_J\|^2 = \|\boldsymbol{U}_{\mathcal{O}}^{(2)}\|^2 - J(\overline{U}_{\mathcal{O}}^{(2)})^2. \tag{16.119}$$

By (16.79), $r_i^*-\nu_m = u_i^{(1)}+u_i^{(2)}$, and the mean over $\mathcal{O}$ of the $u_i^{(1)}$ is zero, hence using (16.35) and (16.37),

$$\begin{aligned}\mathrm{E}[(\overline{U}_{\mathcal{O}}^{(2)})^2] = \mathrm{Var}[\overline{R}_{\mathcal{O}}^*] &= \frac{(m+1)^2}{(I+1)^2}\,\mathrm{Var}[\overline{R}_{\mathcal{O}}] \\ &= \frac{(m+1)^2}{(I+1)^2}\frac{H^2}{J^2}\,\mathrm{Var}[\overline{R}_H] \\ &= \frac{(m+1)^2}{(I+1)^2}\frac{H}{J^2}\frac{I-H}{I-1}\,\mathrm{Var}[R_1].\end{aligned} \tag{16.120}$$

$$\begin{aligned}\mathrm{E}[\|\boldsymbol{U}_{\mathcal{O}}^{(2)}-\overline{U}_{\mathcal{O}}^{(2)}\mathbf{1}_J\|^2] = {}& \frac{H\overline{H}\,(3(1-\overline{\gamma}_2)+(\overline{H}-1)(1-\overline{\gamma}_3))}{12} \\ & + \frac{H}{I+1}\left(\frac{(\bar{J}-\overline{H})^2}{I+1}+2\,\frac{(\bar{J}-\overline{H})\overline{H}}{I-1}+\frac{(J-H)\overline{H}}{(I+1)(I-1)}-\frac{(m+1)^2(I-H)}{(I+1)(I-1)J}\right)\mathrm{Var}[R_1].\end{aligned} \tag{16.121}$$

16.4.3 The variance for Kendall under $\mathcal{H}_2$

Now we take the expected value of the quantities in Section **??** over H, where H is can hypergeometric:

$$\mathrm{P}[H=h] = \frac{\binom{J}{h}\binom{m-J}{I-h}}{\binom{m}{I}} = \frac{\binom{I}{h}\binom{m-I}{J-h}}{\binom{m}{J}}. \tag{16.122}$$

From (16.104),

$$\begin{aligned}\mathsf{E}\left[\mathsf{E}\left[\|\boldsymbol{R}^*_{\mathcal{O}} - \overline{\mathsf{R}}^*_{\mathcal{O}}\mathbf{1}_J\|^2 \,|\, \mathsf{H}\right]\right] &= \frac{(m+1)^2}{(I+1)^2}\left(\mathsf{E}[\mathsf{H}] - \frac{\mathsf{E}[\mathsf{H}(I-\mathsf{H})]}{J(I-1)}\right)\mathrm{Var}[\mathsf{R}_a] \\ &= \frac{(m+1)^2}{(I+1)^2}\left(\frac{IJ}{m} - \frac{J(m-J)}{J(I-1)}\frac{I(I-1)}{m(m-1)}\right)\mathrm{Var}[\mathsf{R}_a] \\ &= \frac{(m+1)^2}{m-1}\frac{I(J-1)}{(I+1)^2}\mathrm{Var}[\mathsf{R}_a]\end{aligned} \tag{16.123}$$

Next, from (16.105),

$$\mathsf{E}[\mathsf{E}[\|\boldsymbol{A}^*\|^2 \,|\, \mathsf{H}]] = \frac{\mathsf{E}[\mathsf{H}(\mathsf{H}-1)]}{8}(1-\overline{\gamma}_2) + \frac{\mathsf{E}[\mathsf{H}(J-\mathsf{H})]}{(I+1)^2}\mathrm{Var}[\mathsf{R}_a]. \tag{16.124}$$

Note that

$$\begin{aligned}\mathsf{E}[(\mathsf{H})_c] &= \frac{(I)_c(J)_c}{(m)_c}, \\ \mathsf{E}[(\mathsf{H})_a(J-\mathsf{H})_b] &= \frac{(I)_a(m-I)_b(J)_{a+b}}{(m)_{a+b}}, \text{ and} \\ \mathsf{E}[(\mathsf{H})_a(I-\mathsf{H})_b] &= \frac{(J)_a(m-J)_b(I)_{a+b}}{(m)_{a+b}}.\end{aligned} \tag{16.125}$$

Then

$$\mathsf{E}[\mathsf{E}[\|\boldsymbol{A}^*\|^2 \,|\, \mathsf{H}]] = \frac{I(J)_2}{(m)_2}\left(\frac{(I-1)}{8}(1-\overline{\gamma}_2) + \frac{(m-I)}{(I+1)^2}\mathrm{Var}[\mathsf{R}_a]\right). \tag{16.126}$$

Next, consider $\|\boldsymbol{U}^{(1)}_{\mathcal{O}}\|^2$ in (16.113). We need another expected value:

$$\begin{aligned}\mathsf{E}[\mathsf{H}^2(J-\mathsf{H})] &= \mathsf{E}[\mathsf{H}(\mathsf{H}-1)(J-\mathsf{H})] + \mathsf{E}[\mathsf{H}(J-\mathsf{H})] \\ &= \frac{(I)_2(J)_3(m-I)}{(m)_3} + \frac{I(J)_2(m-I)}{(m)_2}\end{aligned} \tag{16.127}$$

Then

$$\mathrm{E}\left[\mathrm{E}\left[\|\boldsymbol{U}_{\mathcal{O}}^{(1)}\|^2 \,|\, \mathrm{H}\right]\right] = \frac{(\mathrm{I})_2(\mathrm{J})_2}{4(m)_2}(1-\overline{\gamma}_2) + \frac{(\mathrm{I})_3(\mathrm{J})_3}{12(m)_3}(1-\overline{\gamma}_3)$$

$$+\left(\frac{(\mathrm{I})_2(\mathrm{J})_3(\mathrm{I}+2)(m-\mathrm{I})}{(m)_3} + \frac{\mathrm{I}(\mathrm{J})_2(\mathrm{I}+2)(m-\mathrm{I})}{(m)_2} + \frac{(\mathrm{I})_2(\mathrm{J})_2(\mathrm{J}-1)(m-\mathrm{I})}{(m)_2} - 3\,\frac{\mathrm{I}(\mathrm{J})_2(m-\mathrm{I})}{(m)_2}\right)$$
$$\times \frac{\mathrm{Var}[\mathrm{R}_1]}{(\mathrm{I}+1)^2(\mathrm{I}-1)}$$
$$= \frac{(\mathrm{I})_2(\mathrm{J})_2}{12(m)_2}\left(3(1-\overline{\gamma}_2) + \frac{(\mathrm{I}-2)(\mathrm{J}-2)}{m-2}(1-\overline{\gamma}_3)\right)$$
$$+\left(\frac{(\mathrm{J}-2)(\mathrm{I}+2)}{(m-2)} + \mathrm{J}\right)\frac{(\mathrm{I})_2(\mathrm{J})_2(m-\mathrm{I})}{(m)_2(\mathrm{I}+1)^2(\mathrm{I}-1)}\,\mathrm{Var}[\mathrm{R}_1]. \quad (16.128)$$

For $\|\boldsymbol{U}_{\mathcal{O}}^{(2)}\|^2$,

$$\mathrm{E}[(\mathrm{H})_a(\mathrm{I}-\mathrm{H})_b] = \frac{(\mathrm{J})_a(m-\mathrm{J})_b(\mathrm{I})_{a+b}}{(m)_{a+b}}. \quad (16.129)$$

Next,

$$\begin{aligned}\mathrm{E}[\mathrm{H}(\mathrm{I}-\mathrm{H})(\mathrm{J}-\mathrm{H})] &= (\mathrm{J}-1)\mathrm{E}[\mathrm{H}(\mathrm{I}-\mathrm{H})] + \mathrm{E}[(\mathrm{H})_2(\mathrm{I}-\mathrm{H})] \\ &= (\mathrm{J}-1)\frac{\mathrm{J}(m-\mathrm{J})(\mathrm{I})_2}{(m)_2} + \frac{(\mathrm{J})_2(m-\mathrm{J})(\mathrm{I})_3}{(m)_3} \\ &= \frac{(\mathrm{I})_2(\mathrm{J})_2(m-\mathrm{J})(m-\mathrm{I})}{(m)_3}; \end{aligned} \quad (16.130)$$

since $\bar{\mathrm{J}} - \overline{\mathrm{H}} = (m-\mathrm{I}) - (\mathrm{J}-\mathrm{H})$,

$$\begin{aligned}\mathrm{E}[\mathrm{H}(\mathrm{I}-\mathrm{H})(\bar{\mathrm{J}}-\overline{\mathrm{H}})] &= (m-\mathrm{I})\mathrm{E}[\mathrm{H}(\mathrm{I}-\mathrm{H})] - \mathrm{E}[\mathrm{H}(\mathrm{I}-\mathrm{H})(\mathrm{J}-\mathrm{H})] \\ &= (m-\mathrm{I})\frac{\mathrm{J}(m-\mathrm{J})(\mathrm{I})_2}{(m)_2} - \frac{(\mathrm{I})_2(\mathrm{J})_2(m-\mathrm{J})(m-\mathrm{I})}{(m)_3} \\ &= \frac{(\mathrm{I})_2\mathrm{J}(m-\mathrm{J})(m-\mathrm{I})}{(m)_2}\left(1-\frac{\mathrm{J}-1}{m-2}\right) \\ &= \frac{(\mathrm{I})_2\mathrm{J}(m-\mathrm{J})_2(m-\mathrm{I})}{(m)_3}. \end{aligned} \quad (16.131)$$

Also, using (16.125) and the idea in (16.127),

$$\begin{aligned}
\mathrm{E}[\mathrm{H}(\overline{\mathrm{J}}-\overline{\mathrm{H}})^2] &= (m-I)^2\,\mathrm{E}[\mathrm{H}] - 2(m-I)\,\mathrm{E}[\mathrm{H}(\mathrm{J}-\mathrm{H})] + \mathrm{E}[\mathrm{H}(\mathrm{J}-\mathrm{H})^2] \\
&= (m-I)^2\frac{IJ}{m} - 2(m-I)^2\,\frac{I(J)_2}{(m)_2} + \frac{I(m-I)_2(J)_3}{(m)_3} + \frac{I(m-I)(J)_2}{(m)_2} \\
&= \frac{(m-I)IJ}{m}\left((m-I) - 2\,\frac{(m-I)(J-1)}{m-1} + \frac{(m-I-1)(J-1)(J-2)}{(m-1)(m-2)} + \frac{J-1}{m-1}\right) \\
&= \frac{(m-I)IJ}{m}\left((m-I)\left(1-2\,\frac{J-1}{m-1} + \frac{(J-1)(J-2)}{(m-1)(m-2)}\right) - \frac{(J-1)(J-2)}{(m-1)(m-2)} + \frac{J-1}{m-1}\right) \\
&= \frac{IJ(m-I)(m-J)}{(m)_3}\left((m-I)(m-J-1) + (J-1)\right) \\
&= \frac{IJ(m-I)^2(m-J)_2}{(m)_3} + \frac{I(J)_2(m-I)(m-J)}{(m)_3}. \qquad (16.132)
\end{aligned}$$

Then from (16.121),

$$\begin{aligned}
\mathrm{E}\left[\mathrm{E}\left[\|\boldsymbol{U}_{\mathcal{O}}^{(2)}\|^2 \,|\, \mathrm{H}\right]\right] &= \frac{1}{12}\frac{J(m-J)(I)_2}{(m)_2}\left(3\,(1-\overline{\gamma}_2) + \frac{(m-J-1)(I-2)}{m-2}(1-\overline{\gamma}_3)\right) \\
&\quad + \left(\frac{1}{I+1}\left(\frac{IJ(m-I)^2(m-J)_2}{(m)_3} + \frac{I(J)_2(m-I)(m-J)}{(m)_3}\right)\right. \\
&\quad \left. +2\,\frac{(I)_2J(m-J)_2(m-I)}{(I-1)(m)_3} + \frac{(I)_2(J)_2(m-J)(m-I)}{(I+1)(I-1)(m)_3} - \frac{(m+1)^2J(m-J)(I)_2}{(I+1)(I-1)J(m)_2}\right)\frac{\mathrm{Var}[\mathrm{R}_1]}{I+1} \\
&= \frac{1}{12}\frac{J(m-J)(I)_2}{(m)_2}\left(3\,(1-\overline{\gamma}_2) + \frac{(m-J-1)(I-2)}{m-2}(1-\overline{\gamma}_3)\right) \\
&\quad + \left(\frac{IJ(m-I)^2(m-J)_2}{(I+1)(m)_3} + 2\,\frac{I(J)_2(m-I)(m-J)}{(I+1)(m)_3}\right. \\
&\quad \left. +2\,\frac{IJ(m-J)_2(m-I)}{(m)_3} - \frac{(m+1)^2(m-J)I}{(I+1)(m)_2}\right)\frac{\mathrm{Var}[\mathrm{R}_1]}{I+1} \\
&= \frac{1}{12}\frac{J(m-J)(I)_2}{(m)_2}\left(3\,(1-\overline{\gamma}_2) + \frac{(m-J-1)(I-2)}{m-2}(1-\overline{\gamma}_3)\right) \\
&\quad + \left(\frac{IJ(m-I)(m-J)}{(I+1)(m)_3}((m+I+2)(m-J-1) + 2\,(J-1)) - \frac{(m+1)^2(m-J)I}{(I+1)(m)_2}\right)\frac{\mathrm{Var}[\mathrm{R}_1]}{I+1} \\
&= \frac{1}{12}\frac{J(m-J)(I)_2}{(m)_2}\left(3\,(1-\overline{\gamma}_2) + \frac{(m-J-1)(I-2)}{m-2}(1-\overline{\gamma}_3)\right) \\
&\quad + \frac{I(m-J)\,\mathrm{Var}[\mathrm{R}_1]}{(I+1)^2(m)_2}\left(\frac{J(m-I)((m+I+2)(m-J-1)+2\,(J-1))}{m-2} - (m+1)^2\right). \qquad (16.133)
\end{aligned}$$

For the variance of Kendall, we find the expected value of the variance in (16.59) over the

$\mathcal{H}_2$ hypothesis.

$$\mathrm{Var}[\mathrm{d}^{\mathrm{A}}_{\mathrm{Ken}}(\boldsymbol{w}, \boldsymbol{Z})] = \frac{1}{3}\left(\frac{\omega_J}{J+1}\,\mathrm{E}\left[\mathrm{E}\left[\|\boldsymbol{R}^*_{\mathcal{O}} - \overline{\mathrm{R}}^*_{\mathcal{O}}\mathbf{1}_J\|^2 \,|\, \mathrm{H}\right]\right] + \mathrm{E}\left[\mathrm{E}\left[\|\boldsymbol{A}^*\|^2 \,|\, \mathrm{H}\right]\right](1 - 3\gamma_2 + 2\gamma_3)\right.$$
$$\left. +3\,\mathrm{E}\left[\mathrm{E}\left[\|\boldsymbol{U}^{(1)}_{\mathcal{O}}\|^2 \,|\, \mathrm{H}\right]\right]\frac{\gamma_2 - \gamma_3}{J+1} - \frac{\omega_J}{(J+1)^2}\,\mathrm{E}\left[\mathrm{E}\left[\|\boldsymbol{U}^{(2)}_{\mathcal{O}} - \overline{\mathrm{U}}^{(2)}_{\mathcal{O}}\mathbf{1}_J\|^2 \,|\, \mathrm{H}\right]\right]\right). \quad (16.134)$$

References

Aldous, D., & Diaconis, P. (1999). Longest increasing subsequences: From patience sorting to the baik-deift-johansson theorem. *Bulletin of the American Mathematical Society*, *36*, 413–432.

Alvo, M., & Cabilio, P. (1991). On the balanced incomplete block design for rankings. *Annals of Statistics*, *19*, 1597 – 1613.

Alvo, M., & Cabilio, P. (1995). Rank correlation methods for missing data. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, *23*(4), 345–358.

Alvo, M., & Yu, P. L. (2014). *Statistical methods for ranking data*. Springer Publishing Company, Incorporated.

Baer, R. M., & Brock, P. (1968). Natural sorting over permutation spaces. *Mathematics of Computation*, *22*, 385–410.

Baik, J., Deift, P., & Johansson, K. (1999). On the distribution of the length of the longest increasing subsequence of random permutations. *Journal of the American Mathematical Society*, *12*(4), 1119–1178.

Best, D. J., & Roberts, D. E. (1975). Algorithm as 89: The upper tail probabilities of spearman's rho. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *24*(3), 377–379.

Blinnikov, S., & Moessner, R. (1998). Expansions for nearly Gaussian distributions. *Astron. Astrophys. Suppl. Ser.*, *130*, 193-205.

Bornemann, F. (2010). On the numerical evaluation of distributions in random matrix theory: A review with an invitation to experimental mathematics. *Markov Processes and Related Fields*, *16*(4), 803–866.

Brown, B. (1988). Kendall's tau and contingency tables. *Australian Journal of Statistics*, *30*(3), 276-291.

Chiani, M. (2014). Distribution of the largest eigenvalue for real wishart and gaussian random matrices and a simple approximation for the tracy-widom distribution. *Journal of Multivariate Analysis*, *129*, 69 – 81.

Cramér, H. (1946). *Mathematical methods of statistics*. Princeton University Press.

Critchlow, D. E. (1985). *Metric methods for analyzing partially ranked data.* New York.: Springer–Verlag.

David, S. T., Kendall, M. G., & Stuart, A. (1951). Some questions of distribution in the theory of rank correlation. *Biometrika*, *38*(1-2), 131-140.

Diaconis, P. (1988). *Group representations in probability and statistics.* Hayward, California.: Institute of Mathematical Statistics.

Diaconis, P., & Gangolli, A. (1995). Rectangular arrays with fixed margins. In *Discrete probability and algorithms. Proceedings of the workshops "Probability and algorithms" and "The finite Markov chain renaissance" held at IMA, University of Minnesota, Minneapolis, MN, USA, 1993* (pp. 15–41). New York, NY: Springer-Verlag.

di Bruno, C. F. F. (1855). Sullo sviluppo delle funzioni. *Annali di Scienze Matematiche e Fisiche*, *6*.

Esseen, C.-G. (1945). Fourier analysis of distribution functions. a mathematical study of the laplace-gaussian law. *Acta Mathematica*, *77*, 1–125.

Feller, W. (1968). *An introduction to probability theory and its applications.* (Vol. I). New York.: Wiley.

Frame, J. S., de B. Robinson, G., & Thrall, R. M. (1954). The hook graphs of the symmetric group. *Canadian Journal of Mathematics*, *6*, 316–324.

Franklin, L. (1988). The complete exact null distribution of spearman's rho for n = 12(1)18. *Journal of Statistical Computation and Simulation*, *29*(3), 578-580.

Gibbons, J., & Chakraborti, S. (2010). *Nonparametric statistical inference* (Fifth ed.). CRC Press.

Good, I. J. (1976). On the application of symmetric dirichlet distributions and their mixtures to contingency tables. *The Annals of Statistics*, *4*(6), 1159–1189.

Greene, C., Nijenhuis, A., & Wilf, H. S. (1979). A probabilistic proof of a formula for the number of young tableaux of a given shape. *Advances in Mathematics*, *31*(1), 104 - 109.

Hammersley, J. M. (1972). A few seedlings of research. In *Proceedings of the sixth berkeley symposium on mathematical statistics and probability, volume 1: Theory of statistics* (pp. 345–394). Berkeley, Calif.: University of California Press.

Henery, R. J. (1981). Permutation probabilities as models for horse races. *Journal of the Royal Statistical Society B*, *43*, 86 – 91.

Herstein, I. N. (1964). *Topics in algebra.* Massachusetts.: Blaisdell.

Hoeffding, W. (1951). A combinatorial central limit theorem. *Annals of Mathematical Statistics*, *22*, 558 – 566.

Hotelling, H., & Pabst, M. R. (1936, 03). Rank correlation and tests of significance involving no assumption of normality. *The Annals of Mathematical Statistics*, *7*(1), 29–43.

Johnstone, I. M., Ma, Z., Perry, P. O., & Shahram, M. (2014). Rmtstat: Distributions, statistics and tests derived from random matrix theory [Computer software manual]. (R package version 0.3)

Jonckheere, A. R. (1954). A distribution-free *k*-sample test against ordered alternatives. *Biometrika*, *41*(1/2), 133-145. doi: 10.1093/biomet/41.1-2.133

Kendall, M. (1948). *Rank correlation methods*. London: Griffin.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, *30*(1/2), 81–93.

Kendall, M. G., & Gibbons, J. D. (1990). *Rank correlation methods*. London.: Edward Arnold.

Kim, J. H. (1996). On increasing subsequences of random permutations. *Journal of Combinatorial Theory, Series A*, *76*(1), 148 - 155.

Kleinecke, D., Ury, H., & Wagner, L. (1962). *Spearman's footrule: An alternative rank statistic* (Tech. Rep. No. Report CDRP-182-114). University of California, Berkeley: Civil Defense Research Project, Institute of Engineering Research.

Kolassa, J. E., & McCullagh, P. (1990, 06). Edgeworth series for lattice distributions. *The Annals of Statistics*, *18*(2), 981–985.

Kou, S. G., & Ying, Z. (1996). Asymptotics for a 2 × 2 table with fixed margins. *Statistica Sinica*, *6*(4), 809–829.

Logan, B., & Shepp, L. (1977). A variational problem for random young tableaux. *Advances in Mathematics*, *26*(2), 206 - 222.

Maciak, W. (2009). Exact null distribution for n25 and probability approximations for spearman's score in an absence of ties. *Journal of Nonparametric Statistics*, *21*(1), 113-133.

Mann, H. B., & Whitney, D. R. (1947, 03). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, *18*(1), 50–60. doi: 10.1214/aoms/1177730491

Marden, J. (1996). *Analyzing and modeling rank data*. Taylor & Francis.

Marengo, J. E., Farnsworth, D. L., & Stefanic, L. (2017). A geometric derivation of the irwin-hall distribution. *International Journal of Mathematics and Mathematical Sciences*.

Mehta, C. R., & Patel, N. R. (1983). A network algorithm for performing fisher's exact test in r x c contingency tables. *Journal of the American Statistical Association*, *78*(382), 427–434.

Metropolis, N., & Ulam, S. (1949). The monte carlo method. *Journal of the American Statistical Association*, *44*(247), 335–341.

Moran, P. A. P. (1950). Recent developments in ranking theory. *Journal of the Royal Statistical Society. Series B (Methodological)*, *12*(2), 153–162.

Odlyzko, A. M., & Rains, E. M. (2000). On longest increasing subsequences in random permutations. In *Analysis, geometry, number theory: the mathematics of Leon Ehrenpreis (Philadelphia, PA, 1998)* (Vol. 251, pp. 439–451). American Mathematical Society.

Pearson, K. (1907). *On further methods of determining correlation* (No. v. 16). Cambridge University Press.

Romik, D. (2015). *The Surprising Mathematics of Longest Increasing Subsequences*. Cambridge University Press.

Salama, I. A., & Quade, D. (1990). A note on Spearman's footrule. *Communications in Statistics - Simulation and Computation*, *19*(2), 591-601.

Sen, P. K., & Salama, I. A. (1983). The Spearman footrule and a Markov chain property. *Statistics & Probability Letters*, *1*(6), 285–289.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics.* New York.: Wiley.

Silverberg, A. R. (1980). *Statistical models for* q*-permutations.* (Unpublished doctoral dissertation). Department of Statistics, Princeton University.

Silverstone, H. (1950). A note on the cumulants of Kendall's S-distribution. *Biometrika*, *37*(3-4), 231-235.

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, *15*(1), 72–101.

Stockmal, F. (1962). Algorithm 95: Generation of partitions in part-count form. *Commun. ACM*, *5*(6), 344.

Stojmenović, I., & Zoghbi, A. (1998). Fast algorithms for generating integer partitions. *International Journal of Computer Mathematics*, *70*(2), 319-332.

Terpstra, T. J. (1952). The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking. *Indagationes Mathematicae (Proceedings)*, *55*, 327–333.

Terpstra, T. J. (1953). The exact probability distribution of the t statistic for testing against trend and its normal approximation. *Indagationes Mathematicae (Proceedings)*, *56*, 433–436.

Tracy, C. A., & Widom, H. (1994). Level-spacing distributions and the airy kernel. *Communications in Mathematical Physics*, *159*(1), 151–174.

Ulam, S. M. (1961). Monte carlo calculations in problems of mathematical physics. In E. F. Beckenbach (Ed.), *Modern mathematics for the engineer: Second series.* McGraw-Hill.

Ury, H., & Kleinecke, D. (1979). Tables of the distribution of spearman's footrule. *Applied Statistics*, *28*, 271 – 275.

van de Wiel, M., & Bucchianico, A. (2001). Fast computation of the exact null distribution of spearman's and page's l statistic for samples with and without ties. *Journal of Statistical Planning and Inference*, *92*(1), 133 - 145.

Verbeek, A., & Kroonenberg, P. M. (1985). A survey of algorithms for exact distributions of test statistics in r x c contingency tables with fixed margins. *Computational Statistics & Data Analysis*, *3*, 159 - 185.

Vershik, A. M., & Kerov, S. V. (1977). Asymptotics of the plancherel measure of the symmetric group and the limiting form of young tableaux. *Soviet Mathematics – Doklady*, *18*, 527–531. (English translation of Doklady Akademii Nauk SSSR, 32, 1024–1027.)

Weisstein, E. W. (2018a). *Hermite polynomial.* (From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/StirlingNumberoftheSecondKind.html)

Weisstein, E. W. (2018b). *Stirling number of the second kind.* (From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/HermitePolynomial.html)

Wellner, J. A. (2002, April). *On longest increasing subsequences and random young tableaux: Experimental results and recent theorems* (Tech. Rep.). Seattle, Washington: University of Washington. https://www.stat.washington.edu/jaw/RESEARCH/PAPERS/lis.pdf.

Wichura, M. J. (2001). Statistics 304: Autumn 2001 distribution theory lecture notes [Computer software manual]. Department of Statistics, University of Chicago. (https://galton.uchicago.edu/~wichura/Stat304/handouts.html)

Wikipedia contributors. (2018a). *Riemann zeta function — Wikipedia, The Free Encyclopedia.* Retrieved from https://en.wikipedia.org/w/index.php?title=Riemann_zeta_function&oldid=863170052 ([Online; accessed 18-October-2018])

Wikipedia contributors. (2018b). *Tracy-Widom distribution — Wikipedia, The Free Encyclopedia.* https://en.wikipedia.org/wiki/Tracy-Widom_distribution. ([Online; accessed 22-June-2018])

Wikipedia contributors. (2019a). *Irwin–hall distribution — Wikipedia, the free encyclopedia.* Retrieved from https://en.wikipedia.org/w/index.php?title=Irwin%E2%80%93Hall_distribution&oldid=920954059 ([Online; accessed 16-December-2019])

Wikipedia contributors. (2019b). *Newton's identities — Wikipedia, the free encyclopedia.* Retrieved from https://en.wikipedia.org/w/index.php?title=Newton%27s_identities&oldid=889902778 ([Online; accessed 16-April-2019])

Wilcoxon, F. (1945, December). Individual comparisons by ranking methods. *Biometrics Bulletin*, *1*(6), 80–83. doi: 10.2307/3001968

Yu, P. L., Lam, K., & Alvo, M. (2016, Apr.). Nonparametric rank tests for independence in opinion surveys. *Austrian Journal of Statistics*, *31*(4), 279 – 290.