Bernoulli **25**(1), 2019, 26–46
https://doi.org/10.3150/17-BEJ930

Leading the field: Fortune favors the bold in Thurstonian choice models

STEVEN N. EVANS[1,*], RONALD L. RIVEST[2] and PHILIP B. STARK[1,**]

[1]*Department of Statistics, University of California, Berkeley, CA 94720-3860, USA. E-mail:* [*]*evans@stat.berkeley.edu;* [**]*stark@stat.berkeley.edu*
[2]*Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. E-mail: rivest@mit.edu*

Schools with the highest average student performance are often the smallest schools; localities with the highest rates of some cancers are frequently small; and the effects observed in clinical trials are likely to be largest for the smallest numbers of subjects. Informal explanations of this "small-schools phenomenon" point to the fact that the sample means of smaller samples have higher variances. But this cannot be a complete explanation: If we draw two samples from a diffuse distribution that is symmetric about some point, then the chance that the smaller sample has larger mean is 50%. A particular consequence of results proved below is that if one draws three or more samples of different sizes from the same normal distribution, then the sample mean of the smallest sample is most likely to be highest, the sample mean of the second smallest sample is second most likely to be highest, and so on; this is true even though for any pair of samples, each one of the pair is equally likely to have the larger sample mean. The same effect explains why heteroscedasticity can result in misleadingly small nominal p-values in nonparametric tests of association.

Our conclusions are relevant to certain stochastic choice models, including the following generalization of Thurstone's Law of Comparative Judgment. There are n items. Item i is preferred to item j if $Z_i < Z_j$, where Z is a random n-vector of preference scores. Suppose $\mathbb{P}\{Z_i = Z_j\} = 0$ for $i \neq j$, so there are no ties. Item k is the favorite if $Z_k < \min_{i \neq k} Z_i$. Let p_i denote the chance that item i is the favorite. We characterize a large class of distributions for Z for which $p_1 > p_2 > \cdots > p_n$. Our results are most surprising when $\mathbb{P}\{Z_i < Z_j\} = \mathbb{P}\{Z_i > Z_j\} = \frac{1}{2}$ for $i \neq j$, so neither of any two items is likely to be preferred over the other in a pairwise comparison. Then, under suitable assumptions, $p_1 > p_2 > \cdots > p_n$ when the variability of Z_i decreases with i in an appropriate sense. Our conclusions echo the proverb "Fortune favors the bold."

Keywords: coupling; discrete choice models; extreme value; maximum (or minimum) of random variables; most dangerous equation; order statistic; preference scores; small schools phenomenon; stochastic domination; test of association; Thurstone; winning probability

1. Introduction

When an achievement test is administered to all students of a particular age in a U.S. state, it is typically observed that there is a disproportionate number of small schools among those with the highest average scores [26]. This "small-schools phenomenon" is to be expected even if the scores of individual students at small schools come from the same population as those at other schools: the standard deviation of the average score at a school with n students is proportional to $\frac{1}{\sqrt{n}}$; averages at small schools will thus be more variable than those at larger schools; and hence small schools are likely to be disproportionately represented among the highest performing (and

1350-7265

lowest performing) schools. Wainer [26] lists several more examples of this effect such as small communities having what seem to be unusually high rates of kidney cancer and small cities appearing to be safer than larger ones.

The results we establish here bear on how the probability that a school has the highest average depends on its size under the assumption that student performances are drawn from a common population. Suppose that the $n \geq 3$ schools are listed in order of increasing size and $Z_1, Z_2, \ldots, Z_n$ are the average test scores. Assume that the Z_k are independent and symmetrically distributed around a common mean μ and that $|Z_i - \mu|$ stochastically dominates $|Z_j - \mu|$ when $i < j$ (for example, this will hold approximately when each Z_k is approximately Gaussian because then Z_k is approximately of the form $\frac{\sigma}{\sqrt{N_k}} Y_k$, where σ is the standard deviation for an individual student's score, N_k is the number of students at the kth school, and $Y_1, Y_2, \ldots, Y_n$ are independent standard normal random variables). It follows from the results we establish that $\mathbb{P}\{Z_k$ is the largest of $Z_1, \ldots, Z_n\}$ is decreasing in k – the smaller a school is the more likely it is to have the highest average test score – even though no school has an advantage over any other in a "head-to-head" competition ($\mathbb{P}\{Z_i > Z_j\} = \mathbb{P}\{Z_j > Z_i\}$ for any pair $i \neq j$).

We can describe this conclusion a little more picturesquely. Consider a group of $n \geq 3$ independent individuals with equal skill, in the sense that each individual's performance is symmetrically distributed about some common mean, so that in a head-to-head contest between any two there is an equal chance that either will win. For each individual k, let $\Pr_k(x)$ be the chance that the absolute value of the difference between his or her performance and the shared expected performance exceeds x. Suppose the individuals are well ordered by these probabilities: for all $x > 0$, $\Pr_1(x) > \Pr_2(x) > \cdots > \Pr_n(x)$. Under these assumptions, individual 1 has the highest probability of having the best performance, individual 2 has the second-highest, and so on. If a greater chance of extreme performance results from deliberate risk-taking, then individual 1 is the "boldest" and the most likely to perform best. In this sense, fortune favors the bold. (Of course, symmetry dictates that individual 1 is also most likely to perform worst!)

To make our results mathematically precise and to connect them to the literature on stochastic models for ranking and ordering, we require the following notation and terminology. Label the n items with the set of integers $[n] := \{1, 2, \ldots, n\}$. An individual's preferences can be represented in two related ways: either we report the *order vector* $(w(1), w(2), \ldots, w(n))$, where $w(1) \in [n]$ is the label of the most favored item, $w(2)$ is the label of the second-most favored item, etc., or we report the corresponding *rank vector* $(y(1), y(2), \ldots, y(n))$, where $w(y(i)) = i$ for $i \in [n]$.

The order vector and rank vector are permutations of the set $[n]$. There is a huge literature on models of random permutations, much of which attempts to capture features of how individuals actually go about assigning orders or rankings using whatever information they have at their disposal. The standard reference is Marden [15], with Diaconis [7]; Fligner and Verducci [9] as useful adjuncts.

The earliest model for assigning orders is due to Thurstone [24,25]. In Thurstone's model, the item labeled i is associated with a (real-valued) random variable Z_i, where the random vector $Z = (Z_1, \ldots, Z_n)$ is such that $\mathbb{P}\{Z_i = Z_j\} = 0$ for $i \neq j$, and the resulting order vector is $(i_1, i_2, \ldots, i_n)$ if $Z_{i_1} < Z_{i_2} < \cdots < Z_{i_n}$. One may interpret $-Z_i$ as the desirability of item i measured on a one-dimensional scale, so that items are ordered in decreasing order of their desirability.

In many applications, it is more natural to consider Z_i rather than $-Z_i$ to be the desirability. For example, one might model an election by letting Z_i be the number of voters who will vote for candidate i (see, for example, Laslier [13]). The candidate who garners the *most* votes wins the election. As another example, consider *Thompson Sampling* for the multi-armed bandit problem in machine learning. The random variable Z_i models the benefit resulting from pulling arm i. The Z_i are sampled and the arm with *maximum* Z_i is pulled [1,23]. Nonetheless, we shall continue here to follow the tradition of Thurstone, and let $-Z_i$ model the desirability of item i.

Let Z_* denote $\min\{Z_i : 1 \le i \le n\}$, the smallest Z_i value, and let I_* denote $\arg\min\{Z_i : 1 \le i \le n\}$, the index of the minimum Z_i value. Let p_i denote the probability that the rank of item i is 1 (i.e., that $i = I_*$).

Given a specification of Z, there are three closely-related problems to consider:

1. Finding the distribution of Z_*. See Gumbel [11]; Kotz and Nadarajah [12]; de Haan and Ferreira [6]; Resnick [18] for a sample of the extensive body of work in this area.
2. Determining which i is most likely to be I_*.
3. Given i, deriving an expression for p_i.

We do not consider problem 1 here; our focus is on problem 2, particularly when, as is usually the case, solving 2 by solving 3 seems intractable. Most generally, we are concerned with finding conditions that imply that $p_1 > p_2 > \cdots > p_n$.

We briefly survey various distributional assumptions on $Z = (Z_1, \dots, Z_n)$ that have been considered in this context.

Thurstone proposed taking $Z = (Z_1, \dots, Z_n)$ to have a non-degenerate Gaussian distribution. Despite its conceptual simplicity, this model is rather daunting computationally. Here p_i is the probability that Z falls in the region $R := \{z \in \mathbb{R}^n : z_i < z_j, j \ne i\}$, an intersection of half-spaces $\bigcap_{j \ne i} \{z \in \mathbb{R}^n : z_i - z_j < 0\}$. We can write $Z = \mu + X\Sigma^{\frac{1}{2}}$, where μ is the mean vector of Z, Σ is the variance-covariance matrix of Z, $\Sigma^{\frac{1}{2}}$ is the positive definite symmetric square root of Σ, and X is a vector with independent standard normal entries. We are therefore interested in the probability that X falls in the polyhedral region $(R - \mu)\Sigma^{-\frac{1}{2}}$. It is usually not possible to express such probabilities in a simple closed form, but there is a large literature on approximating them numerically using various ingenious recursive schemes – see, for example, Gassmann et al. [10]; Miwa et al. [16]; Craig [4].

Appendix A shows that when $\{Z_i\}_{i=1}^n$ are independent Gaussian random variables, computing the probabilities $\{p_i\}_{i=1}^n$ explicitly is somewhat complex even when $n = 3$. Appendix B shows that this is also true when $\{Z_i\}_{i=1}^n$ is a vector of independent random variables with bilateral exponential distributions. However, if all one cares about is the the ordering of the p_i's, then the results of the present paper may apply to cases where explicitly computing $\{p_i\}$ is intractable.

Daniels [5] suggested taking the random vector Z in Thurstone's general model to be of the form $(\theta_1 + X_1, \dots, \theta_n + X_n)$, where $\theta_1, \dots, \theta_n$ are real-valued parameters and $X_1, \dots, X_n$ are independent and identically distributed (IID) random variables. Equivalently (by exponentiating), one can take Z to be of the form $(\gamma_1 Y_1, \dots, \gamma_n Y_n)$, where $\gamma_1, \dots, \gamma_n$ are positive parameters and $Y_1, \dots, Y_n$ are IID positive random variables. It is a consequence of our results here that if $\theta_i < \theta_j$ (or $\gamma_i < \gamma_j$), then i is at least as likely as j to have rank 1, and this inequality is strict under mild conditions. Savage [20] provides a number of other results about the dependence on the parameters of various other probabilities related to the order and rank vectors.

A particularly tractable example of the multiplicative version of Daniels' type of Thurstonian model is when $(Z_1, \ldots, Z_n) = (\gamma_1 Y_1, \ldots, \gamma_n Y_n)$ with $Y_1, \ldots, Y_n$ IID exponential random variables. In this case the probability of a given order vector $(i_1, \ldots, i_n)$ can be computed explicitly: it is

$$\frac{\lambda_{i_1}}{\sum_j \lambda_j} \frac{\lambda_{i_2}}{\sum_{j \neq i_1} \lambda_j} \frac{\lambda_{i_3}}{\sum_{j \neq i_1, i_2} \lambda_j} \cdots,$$

where $\lambda_i := \gamma_i^{-1}$ for $1 \leq i \leq n$. This model is due to Plackett [17] and Luce [14], and was studied in Silverberg [21,22] as the *vase* model: if we imagine a vase containing n types of balls with balls of type i being in proportion $\lambda_i / (\sum_j \lambda_j)$ and we remove balls one-by-one uniformly without replacement, then the order in which the n types first appear is given by this model. The Plackett and Luce model is the only Thurstonian model of the Daniels type that satisfies the axioms laid out in Luce [14] for a rational choice procedure – see Yellott [29] for a discussion.

The Plackett and Luce model is also the stationary distribution of a discrete-time Markov chain that is sometimes called the *Tsetlin library process* or the *move-to-the-front self-organizing list*. Here the items are pictured as books and an order vector $(i_1, \ldots, i_n)$ corresponds to a stack with the book labeled i_n on the bottom and the book labeled i_1 on top. In each step of the chain, book i is chosen with probability proportional to λ_i, removed from its current position in the stack, and placed on top of the stack. See, for example, Rivest [19] for early work on this process, and Fill [8] for a detailed analysis of this Markov chain and an extensive review of the related literature.

Thurstonian models based on random vectors with much more complex structure are discussed in Böckenholt [2,3].

Section 5.1 presents a third, more involved, example that illustrates a model of a more complex type that is not built from IID random variables, but where the assumptions of our main result, Theorem 3.1, giving the ordering of $\{p_i\}$, still applies. This example is cast in terms of the times taken by three workers to complete three randomly assigned tasks. The expected time for a worker to complete a task is the same for every (worker, task) pair, but the performance of the first worker is more variable than that of the second, which is in turn more variable than that of the third. Again, computing $\{p_i\}$ is tedious and complex, but Theorem 3.1 easily allows one to find their ordering without explicit computation and to conclude that the first worker has the highest probability of finishing first and the second worker has the second highest probability of finishing first.

This paper investigates how to determine, in Thurstonian models, the ordering of the probabilities that each of the given items will be the most preferred, without having to explicitly compute these probabilities.

In other words, we study the distribution of the first entry in the order vector or, equivalently, the distribution of the label of the item with rank one, and we seek conditions on the distribution of the random vector $(Z_1, \ldots, Z_n)$ such that if p_i is the probability that the item labeled i has rank one, then $p_1 > p_2 > \cdots > p_n$ or at least $p_1 \geq p_2 \geq \cdots \geq p_n$. As we have already remarked, we show that the chain of weak inequalities holds in the Daniels model if $\theta_1 < \theta_2 < \cdots < \theta_n$ in the additive case and $\gamma_1 < \gamma_2 < \cdots < \gamma_n$ in the multiplicative case.

The strict inequalities also hold under suitable assumptions. To see that extra assumptions are necessary, suppose we are in the additive case with $n = 3$ and the common distribution of X_1, X_2, X_3 is uniform on the interval $[0, 1]$, with $\theta_1 = 0 < \theta_2 = 1 < \theta_3 = 2$. Then $p_1 = 1 >$

$p_2 = 0 = p_3$, so only weak and not strict inequalities hold in general. The conclusion $p_1 > p_2 > \cdots > p_n$ can be verified by direct computation for the Plackett and Luce model, where $p_i = \lambda_i/(\lambda_1 + \cdots + \lambda_n)$ with $\lambda_i = \gamma_i^{-1}$.

The plan of the remainder of the paper is as follows. In Section 2 we consider a Thurstonian model with $(Z_1, \ldots, Z_n) = (\sigma_1 X_1, \ldots, \sigma_n X_n)$, where the σ_i are positive constants and $(X_1, \ldots, X_n)$ is a random vector with IID standard Gaussian entries. Of course, if $n = 2$, then $p_1 = p_2 = \frac{1}{2}$ by the symmetry of the Gaussian distribution, but we show in Section 2 that if $n \geq 3$ and $\sigma_1 > \sigma_2 > \cdots > \sigma_n$, then $p_1 > p_2 > \cdots > p_n$. In Appendix A we compute $\{p_1, p_2, p_3\}$ for $n = 3$ to emphasize the difficulty of establishing by direct computation that such an ordering holds for general n.

One way to think about this result is that a choice is being made among n individuals based on their responses to a set of stimuli. The IID random variables $\{|X_1|, \ldots, |X_n|\}$ represent the random stimuli given to the individuals. The response of individual k to the stimulus $|X_k|$ is $S_k \eta_k(|X_k|)$, where $\eta_k(y) = \sigma_k y$ and S_k is the sign of X_k, a $\{-1, +1\}$-valued random variable that is independent of $|X_k|$ and equally likely to be -1 or $+1$. For each k, the function η_k happens to be increasing – but as we shall see, that is irrelevant for a conclusion like that above. What is important is that $\eta_i(y) > \eta_j(y)$ for all y and $1 \leq i < j \leq n$, so that if individuals i and j receive the same stimulus, the response of individual i will be more extreme than that of individual j. The expected responses $\mathbb{E}[Z_k]$, $1 \leq k \leq n$, are all zero and $\mathbb{P}\{Z_i > Z_j\} = \mathbb{P}\{Z_i < Z_j\} = \frac{1}{2}$, $1 \leq i \neq j \leq n$, so that individual i has no advantage over individual j in a head-to-head contest, and yet $p_1 > p_2 > \cdots > p_n$.

These observations suggest that a similar result might hold if

$$(Z_1, \ldots, Z_n) = \bigl(S_1 \eta_1(Y_1), \ldots, S_n \eta_n(Y_n)\bigr),$$

where $(S_1, \ldots, S_n)$ is a suitable exchangeable $\{-1, +1\}^n$-valued random vector (recall that a random vector is exchangeable if its joint distribution is unchanged by any permutation of the coordinates), $(Y_1, \ldots, Y_n)$ is an exchangeable E^n-valued random vector for some measurable space E, and the functions $\eta_k : E \to \mathbb{R}_+$ have the property that $\eta_i(y) > \eta_j(y)$ for all $y \in E$ and $1 \leq i < j \leq n$ (so that the response Z_i is "bolder" than the response Z_j). We show in Section 3 that this conclusion is indeed valid under appropriate assumptions (e.g., the ordering of the p_k would not hold if $S_k = +1$ with probability one for all k; to rule this sort of situation out, we require

$$\mathbb{P}\bigl\{\#\bigl\{k \in [n] : S_k = -1\bigr\} = 2\bigr\} \geq \binom{n}{2} \mathbb{P}\bigl\{\#\bigl\{k \in [n] : S_k = -1\bigr\} = 0\bigr\},$$

which holds, for example, when $\{S_k\}$ are IID with individual probability at least $\frac{1}{2}$ of taking the value -1).

In Section 4, we look at the special case in which $\{Y_1, \ldots, Y_n\}$ and $\{S_1, \ldots, S_n\}$ are both IID.

We give two applications of our results in Section 5. In Section 5.1, we consider a model for randomized experiments where n treatments are assigned uniformly at random to n individuals. The distribution of the response of individual j to treatment i is symmetrically distributed about zero. For a fixed individual j the distribution of the magnitude of the effect of treatment i is stochastically nonincreasing in i: Lower numbered treatments are more likely to have larger magnitude effects than higher numbered ones. We will show that treatment 1 is most likely to have the greatest effect, treatment 2 is second most likely to have the greatest effect, and so on, even though no treatment causes any systematic benefit or harm to any individual.

In Section 5.2, we use our results to show that heteroscedasticity can distort the p-value of a permutation-based test for association between two series to make it appear that there is positive or negative association between the two series when there is no such systematic relationship.

Appendix C sketches an approach for removing the "small-school bias" in a way that is both *fair* (equally likely to choose as best any school, when the schools have the same effect on student scores) and *valid* (most likely to choose as best the school that increases student scores the most).

2. Motivating Gaussian example

Our interest in the general topic of this paper was piqued by the following observation about a Gaussian version of the Thurstone model we mentioned in the Introduction.

Proposition 2.1. *Suppose $n \geq 3$ and $(Z_1, \dots, Z_n) = (\sigma_1 X_1, \dots, \sigma_n X_n)$, where $\sigma_i > 0$ for $1 \leq i \leq n$ and the entries of the random vector $(X_1, \dots, X_n)$ are independent standard Gaussian random variables. If $\sigma_1 > \sigma_2 > \cdots > \sigma_n$, then $p_1 > p_2 > \cdots > p_n$.*

Proof. Let $\bigwedge\{\cdot\}$ denote the minimum of a set of real numbers and $\bigvee\{\cdot\}$ denote the maximum. Note that

$$\begin{aligned} p_i &= \mathbb{P}\{\sigma_i X_i < \sigma_k X_k, k \neq i\} \\ &= \mathbb{P}\left\{\sigma_i X_i < \bigwedge_{k \neq i} \sigma_k X_k\right\} \\ &= \mathbb{P}\left\{\bigvee_{k \neq i} (\sigma_i X_i - \sigma_k X_k) < 0\right\}, \end{aligned} \tag{1}$$

for $1 \leq i \leq n$.

Let ϕ and Φ denote the standard Gaussian probability density function and cumulative distribution function, respectively. Then (by conditioning on X_i in the first integral, integrating by parts in the second, and applying the chain rule in the third),

$$\begin{aligned} p_i &= \int_{-\infty}^{\infty} \prod_{j \neq i} \left(1 - \Phi\left(\frac{x}{\sigma_j}\right)\right) \frac{\partial}{\partial x} \Phi\left(\frac{x}{\sigma_i}\right) dx \\ &= -\int_{-\infty}^{\infty} \Phi\left(\frac{x}{\sigma_i}\right) \frac{\partial}{\partial x} \prod_{j \neq i} \left(1 - \Phi\left(\frac{x}{\sigma_j}\right)\right) dx \\ &= -\int_{-\infty}^{\infty} \Phi\left(\frac{x}{\sigma_i}\right) \sum_{j \neq i} \phi\left(\frac{x}{\sigma_j}\right) \frac{1}{\sigma_j} \prod_{k \neq i, j} \left(1 - \Phi\left(\frac{x}{\sigma_k}\right)\right) dx, \end{aligned} \tag{2}$$

and so

$$\begin{aligned}\frac{\partial p_i}{\partial \sigma_i} &= -\sum_{j\neq i}\int_{-\infty}^{\infty}\phi\left(\frac{x}{\sigma_i}\right)\left(-\frac{x}{\sigma_i^2}\right)\phi\left(\frac{x}{\sigma_j}\right)\frac{1}{\sigma_j}\prod_{k\neq i,j}\left(1-\Phi\left(\frac{x}{\sigma_k}\right)\right)dx \\ &= \sum_{j\neq i}\int_{0}^{\infty}\phi\left(\frac{x}{\sigma_i}\right)\left(\frac{x}{\sigma_i^2}\right)\phi\left(\frac{x}{\sigma_j}\right)\frac{1}{\sigma_j} \\ &\quad\times\left[\prod_{k\neq i,j}\left(1-\Phi\left(\frac{x}{\sigma_k}\right)\right)-\prod_{k\neq i,j}\left(1-\Phi\left(\frac{-x}{\sigma_k}\right)\right)\right]dx \\ &> 0,\end{aligned} \tag{3}$$

where we used the facts that $\phi(z) = \phi(-z)$ for all $z \in \mathbb{R}$ and that the function Φ is increasing. It follows that p_i is an increasing function of σ_i, and, because $p_i = p_j$ when $\sigma_i = \sigma_j$, it is clear that if $\sigma_1 > \sigma_2 > \cdots > \sigma_n$, then $p_1 > p_2 > \cdots > p_n$. $\square$

Remark 2.2. We show in Appendix A that when $n = 3$

$$\begin{aligned}p_1 &= \frac{1}{2\pi}\arccos\left(-\frac{\sigma_1^2}{\sqrt{(\sigma_2^2+\sigma_1^2)(\sigma_3^2+\sigma_1^2)}}\right) \\ > p_2 &= \frac{1}{2\pi}\arccos\left(-\frac{\sigma_2^2}{\sqrt{(\sigma_1^2+\sigma_2^2)(\sigma_3^2+\sigma_2^2)}}\right) \\ > p_3 &= \frac{1}{2\pi}\arccos\left(-\frac{\sigma_3^2}{\sqrt{(\sigma_1^2+\sigma_3^2)(\sigma_2^2+\sigma_3^2)}}\right),\end{aligned} \tag{4}$$

but finding such explicit expressions for the p_i and establishing the ordering claimed in Proposition 2.1 becomes increasingly complex for larger values of n. Moreover, Proposition 2.1 holds, with essentially the same proof, if the common distribution of $X_1, \ldots, X_n$ is an arbitrary symmetric distribution possessing a density, whereas it is typically impossible to find explicit closed form expressions for the p_i in this case. We observe in Appendix B that even for a symmetric distribution as tractable as the bilateral exponential, the formulae for the p_i are already somewhat formidable for $n = 3$ and establishing an ordering analogous to that claimed in Proposition 2.1 requires a certain amount of algebraic manipulation.

3. Main theorem

This section presents our main theorem, giving the most general conditions we have found so far that imply $p_1 \geq p_2 \geq \cdots \geq p_n$.

Theorem 3.1. *Let $(Z_1, \ldots, Z_n)$ be an $\mathbb{R}^n$-valued random vector given by $Z_k = S_k\eta_k(Y_k)$, $1 \leq k \leq n$, where*:

- *$(Y_1, \ldots, Y_n)$ is an exchangeable E^n-valued random vector for some measurable space $(E, \mathcal{E})$;*
- *$\eta_1, \ldots, \eta_n$ are measurable functions from E to $\mathbb{R}_+$ with the property that $\eta_i(y) \geq \eta_j(y)$ for all $y \in E$ and $1 \leq i < j \leq n$;*
- *$(S_1, \ldots, S_n)$ is an exchangeable $\{-1, +1\}^n$-valued random vector;*
- *$(Y_1, \ldots, Y_n)$ and $(S_1, \ldots, S_n)$ are independent;*
- *$\mathbb{P}\{S_1 = \cdots = S_{n-2} = +1;\, S_{n-1} = -1\} \geq \mathbb{P}\{S_1 = \cdots = S_{n-1} = +1\}$.*

Define

$$p_k := \mathbb{P}\left\{Z_k < \bigwedge_{\ell \neq k} Z_\ell\right\}.$$

Then, $p_1 \geq p_2 \geq \cdots \geq p_n$.

Proof. Let $(T_1, \ldots, T_n)$ be a vector of independent random variables that is independent of the pair of random vectors $(Y_1, \ldots, Y_n)$ and $(S_1, \ldots, S_n)$ and such that each random variable T_k has an exponential distribution with mean 1. Set $Z_k^\varepsilon = S_k(\eta_k(Y_k) + \varepsilon T_k)$ for $1 \leq k \leq n$ and $\varepsilon > 0$. It is clear that p_k is the limit as $\varepsilon \downarrow 0$ of

$$p_k^\varepsilon := \mathbb{P}\left\{Z_k^\varepsilon < \bigwedge_{\ell \neq k} Z_\ell^\varepsilon\right\}$$

for $1 \leq k \leq n$, so it suffices to show that $p_1^\varepsilon \geq p_2^\varepsilon \geq \cdots \geq p_n^\varepsilon$.

Set

$$q(m) := \begin{cases} \mathbb{P}\{S_1 = \cdots = S_m = +1;\, S_{m+1} = -1\}, & 0 \leq m < n-1, \\ \mathbb{P}\{S_1 = \cdots = S_{n-1} = +1\}, & m = n-1. \end{cases}$$

By the assumptions of the theorem, for $0 \leq m < n-1$,

$$q(m) = \mathbb{P}\{S_{k_1} = \cdots = S_{k_m} = +1;\, S_{k_{m+1}} = -1\}$$

for any subset $\{k_1, \ldots, k_{m+1}\} \subseteq [n]$ of cardinality $m+1$, and

$$q(n-1) = \mathbb{P}\{S_{k_1} = \cdots = S_{k_{n-1}} = +1\}$$

for any subset $\{k_1, \ldots, k_{n-1}\} \subseteq [n]$ of cardinality $n-1$. Thus, $q(0) \geq q(1) \geq \cdots \geq q(n-2)$ and, by assumption, $q(n-2) \geq q(n-1)$.

Suppose that $a_1, \ldots, a_n \in \mathbb{R}_+$ are distinct. If $a_k \neq \bigwedge_\ell a_\ell$, then

$$\mathbb{P}\left\{S_k a_k < \bigwedge_{\ell \neq k} S_\ell a_\ell\right\} = \mathbb{P}\big(\{S_k = -1\} \cap \{S_\ell = +1,\ \forall \ell \neq k \text{ such that } a_\ell > a_k\}\big);$$

whereas if $a_k = \bigwedge_\ell a_\ell$, then

$$\mathbb{P}\left\{S_k a_k < \bigwedge_{\ell \neq k} S_\ell a_\ell\right\} = \mathbb{P}\{S_\ell = +1, \ell \neq k\}.$$

In either case,

$$\mathbb{P}\left\{S_k a_k < \bigvee_{\ell \neq k} S_\ell a_\ell\right\} = q\big(\#\{1 \leq \ell \leq n : a_\ell > a_k\}\big).$$

The values of $|Z_1^\varepsilon|, \ldots, |Z_n^\varepsilon|$ are almost surely distinct. For $1 \leq k \leq n$ set

$$M_k := \#\big\{1 \leq \ell \leq n : \big|Z_\ell^\varepsilon\big| > \big|Z_k^\varepsilon\big|\big\}.$$

We must show that

$$\mathbb{E}\big[q(M_i)\big] \geq \mathbb{E}\big[q(M_j)\big]$$

for $1 \leq i < j \leq n$; or, equivalently after summing by parts, that

$$q(0)\mathbb{P}\{M_i \geq 0\} + \sum_{m=0}^{n-2}\big[q(m+1) - q(m)\big]\mathbb{P}\{M_i \geq m+1\}$$

$$\geq q(0)\mathbb{P}\{M_j \geq 0\} + \sum_{m=0}^{n-2}\big[q(m+1) - q(m)\big]\mathbb{P}\{M_j \geq m+1\}.$$

Since $\mathbb{P}\{M_i \geq 0\} = \mathbb{P}\{M_j \geq 0\} = 1$ and $q(0) \geq q(1) \geq \cdots \geq q(n-1)$, it suffices to show that

$$\mathbb{P}\{M_i \geq m\} \leq \mathbb{P}\{M_j \geq m\}$$

for $1 \leq m \leq n-1$.

Fix $1 \leq i < j \leq n$. Note that

$$\mathbb{P}\{M_i \geq m\} = \mathbb{P}\big\{\exists k_1, \ldots, k_m \neq i : \big|Z_{k_h}^\varepsilon\big| > \big|Z_i^\varepsilon\big|, 1 \leq h \leq m\big\}$$

and $\mathbb{P}\{M_j \geq m\}$ is given by a similar expression. Define functions $\tilde{\eta}_k$, $1 \leq k \leq n$, by $\tilde{\eta}_i = \eta_j$, $\tilde{\eta}_j = \eta_i$, and $\tilde{\eta}_k = \eta_k$, $k \notin \{i, j\}$. Observe that

$$\begin{aligned}
&\big\{\exists k_1, \ldots, k_m \neq i : \big|Z_{k_h}^\varepsilon\big| > \big|Z_i^\varepsilon\big|, 1 \leq h \leq m\big\} \\
&\quad = \big\{\exists k_1, \ldots, k_m \neq i : \eta_{k_h}(Y_{k_h}) + \varepsilon T_{k_h} > \eta_i(Y_i) + \varepsilon T_i, 1 \leq h \leq m\big\} \\
&\quad \subseteq \big\{\exists k_1, \ldots, k_m \neq i : \tilde{\eta}_{k_h}(Y_{k_h}) + \varepsilon T_{k_h} > \tilde{\eta}_i(Y_i) + \varepsilon T_i, 1 \leq h \leq m\big\}
\end{aligned}$$

because $\tilde{\eta}_i(y) = \eta_j(y) \leq \eta_i(y)$ and $\tilde{\eta}_k(y) \geq \eta_k(y)$ for $k \neq i$ (with equality unless $k = j$). Define random variables $\tilde{Y}_k$, $1 \leq k \leq n$, by $\tilde{Y}_i = Y_j$, $\tilde{Y}_j = Y_i$, and $\tilde{Y}_k = Y_k$, $k \notin \{i, j\}$. Define $\tilde{T}_k$, $1 \leq k \leq n$, similarly. By exchangeability, $(Y_1, \ldots, Y_n)$ and $(\tilde{Y}_1, \ldots, \tilde{Y}_n)$ have the same distribution. Of course, $(T_1, \ldots, T_n)$ and $(\tilde{T}_1, \ldots, \tilde{T}_n)$ have the same distribution. Therefore,

$$\begin{aligned}
&\mathbb{P}\big\{\exists k_1, \ldots, k_m \neq i : \tilde{\eta}_{k_h}(Y_{k_h}) + \varepsilon T_{k_h} > \tilde{\eta}_i(Y_i) + \varepsilon T_i, 1 \leq h \leq m\big\} \\
&\quad = \mathbb{P}\big\{\exists k_1, \ldots, k_m \neq i : \tilde{\eta}_{k_h}(\tilde{Y}_{k_h}) + \varepsilon \tilde{T}_{k_h} > \tilde{\eta}_i(\tilde{Y}_i) + \varepsilon \tilde{T}_i, 1 \leq h \leq m\big\}.
\end{aligned}$$

Now

$$
\begin{aligned}
&\big\{\exists k_1,\dots,k_m \neq i : \tilde{\eta}_{k_h}(\tilde{Y}_{k_h}) + \varepsilon \tilde{T}_{k_h} > \tilde{\eta}_i(\tilde{Y}_i) + \varepsilon \tilde{T}_i,\ 1 \le h \le m\big\} \\
&\quad = \big\{\exists k_1,\dots,k_m \neq j : \eta_{k_h}(Y_{k_h}) + \varepsilon T_{k_h} > \eta_j(Y_j) + \varepsilon T_j,\ 1 \le h \le m\big\} \\
&\quad = \{M_j \ge m\}.
\end{aligned}
$$

Putting the above together gives $\mathbb{P}\{M_i \ge m\} \le \mathbb{P}\{M_j \ge m\}$ as required. □

Remark 3.2. Assume the hypotheses of Theorem 3.1. Note that

$$
\begin{aligned}
&\mathbb{P}\{S_1 = \cdots = S_{n-2} = +1;\ S_{n-1} = -1\} \\
&\quad = \mathbb{P}\{S_1 = \cdots = S_{n-2} = +1;\ S_{n-1} = -1;\ S_n = +1\} \\
&\qquad + \mathbb{P}\{S_1 = \cdots = S_{n-2} = +1;\ S_{n-1} = -1;\ S_n = -1\}
\end{aligned}
\tag{5}
$$

and

$$
\begin{aligned}
&\mathbb{P}\{S_1 = \cdots = S_{n-1} = +1\} \\
&\quad = \mathbb{P}\{S_1 = \cdots = S_{n-1} = +1;\ S_n = +1\} \\
&\qquad + \mathbb{P}\{S_1 = \cdots = S_{n-1} = +1;\ S_n = -1\} \\
&\quad = \mathbb{P}\{S_1 = \cdots = S_{n-1} = +1;\ S_n = +1\} \\
&\qquad + \mathbb{P}\{S_1 = \cdots = S_{n-2} = +1;\ S_{n-1} = -1;\ S_n = +1\},
\end{aligned}
\tag{6}
$$

by the exchangeability hypothesis, so the hypothesis that

$$
\mathbb{P}\{S_1 = \cdots = S_{n-2} = +1;\ S_{n-1} = -1\} \ge \mathbb{P}\{S_1 = \cdots = S_{n-1} = +1\}
$$

is equivalent to the hypothesis that

$$
\mathbb{P}\{S_1 = \cdots = S_{n-2} = +1;\ S_{n-1} = S_n = -1\} \ge \mathbb{P}\{S_1 = \cdots = S_n = +1\}.
$$

Again using exchangeability, the latter is equivalent to

$$
\frac{1}{\binom{n}{2}} \mathbb{P}\big\{\#\big\{k \in [n] : S_k = -1\big\} = 2\big\} \ge \mathbb{P}\big\{\#\big\{k \in [n] : S_k = -1\big\} = 0\big\}.
$$

Remark 3.3. Suppose in addition to the hypothesis of Theorem 3.1 that $\mathbb{P}\{S_i = S_j = +1\} = \mathbb{P}\{S_i = S_j = -1\}$ for $i \neq j$. Then, by exchangeability,

$$
\begin{aligned}
\mathbb{P}\{Z_i < Z_j\} &= \mathbb{P}\big\{\eta_i(Y_i) < \eta_j(Y_j)\big\}\mathbb{P}\{S_i = S_j = +1\} \\
&\quad + \mathbb{P}\big\{\eta_i(Y_i) > \eta_j(Y_j)\big\}\mathbb{P}\{S_i = S_j = -1\} \\
&\quad + \mathbb{P}\{S_i = -1;\ S_j = +1\}
\end{aligned}
\tag{7}
$$

$$
\begin{aligned}
&= \mathbb{P}\bigl\{\eta_i(Y_i) < \eta_j(Y_j)\bigr\}\mathbb{P}\{S_i = S_j = -1\} \\
&\quad + \mathbb{P}\bigl\{\eta_i(Y_i) > \eta_j(Y_j)\bigr\}\mathbb{P}\{S_i = S_j = +1\} \\
&\quad + \mathbb{P}\{S_i = +1; S_j = -1\} \\
&= \mathbb{P}\{Z_i > Z_j\}.
\end{aligned}
$$

Theorem 3.1 is especially interesting in this case, because then Z_i is not systematically smaller than Z_j for $i < j$, and yet $p_1 \geq p_2 \geq \cdots \geq p_n$.

Remark 3.4. Theorem 3.1 gives a sufficient condition for the weak inequalities $p_1 \geq p_2 \geq \cdots \geq p_n$ but not the strict inequalities $p_1 > p_2 > \cdots > p_n$. Examining the proof indicates how the hypotheses can be strengthened to yield the latter conclusion. Suppose that $\mathbb{P}\{Z_i = Z_j\} = 0$ for $1 \leq i \neq j \leq n$. It is clear from the proof of the theorem that $p_i > p_j$ for a given pair $1 \leq i < j \leq n$ if and only if there exists $0 \leq m \leq n-2$ such that $q(m+1) < q(m)$ and

$$
\begin{aligned}
&\mathbb{P}\bigl\{\exists k_1, \ldots, k_{m+1} \neq i : \eta_{k_h}(Y_{k_h}) > \eta_i(Y_i), 1 \leq h \leq m+1\bigr\} \\
&\quad < \mathbb{P}\bigl\{\exists k_1, \ldots, k_{m+1} \neq j : \eta_{k_h}(Y_{k_h}) > \eta_j(Y_j), 1 \leq h \leq m+1\bigr\}
\end{aligned}
$$

for that m. For example, if $n \geq 3$ and $q(0) = \mathbb{P}\{S_1 = -1\} > \mathbb{P}\{S_1 = +1; S_2 = -1\} = q(1)$, then it suffices that $\mathbb{P}\{\exists k \neq i : \eta_k(Y_k) > \eta_i(Y_i)\} < \mathbb{P}\{\exists k \neq j : \eta_k(Y_k) > \eta_j(Y_j)\}$ or, equivalently by exchangeability,

$$
\begin{aligned}
\mathbb{P}\Biggl\{\bigvee_{k \notin \{i,j\}} \eta_k(Y_k) \vee \eta_j(Y_j) > \eta_i(Y_i)\Biggr\} &< \mathbb{P}\Biggl\{\bigvee_{k \notin \{i,j\}} \eta_k(Y_k) \vee \eta_i(Y_i) > \eta_j(Y_j)\Biggr\} \\
&= \mathbb{P}\Biggl\{\bigvee_{k \notin \{i,j\}} \eta_k(Y_k) \vee \eta_i(Y_j) > \eta_j(Y_i)\Biggr\}.
\end{aligned}
$$

Because $\eta_j(Y_j) \leq \eta_i(Y_j)$ and $\eta_i(Y_i) \geq \eta_j(Y_i)$ it further suffices to have

$$
0 < \mathbb{P}\Biggl\{\bigvee_{k \notin \{i,j\}} \eta_k(Y_k) \vee \eta_i(Y_j) > \eta_j(Y_i), \bigvee_{k \notin \{i,j\}} \eta_k(Y_k) \vee \eta_j(Y_j) \leq \eta_i(Y_i)\Biggr\}. \tag{8}
$$

4. Independent random variables

Theorem 3.1 has the following consequence when the entries of $(Z_1, \ldots, Z_n)$ are independent.

Corollary 4.1. *Suppose that $n \geq 3$. Let $(Z_1, \ldots, Z_n)$ be an $\mathbb{R}^n$-valued random vector given by $Z_k = S_k W_k$, $1 \leq k \leq n$, where:*

- *$W_1, \ldots, W_n$ are independent $\mathbb{R}$-valued random variables;*
- *W_i stochastically dominates W_j for $1 \leq i < j \leq n$ (that is, $\mathbb{P}\{W_i > w\} \geq \mathbb{P}\{W_j > w\}$ for all $w \in \mathbb{R}_+$);*

- *$S_1, \ldots, S_n$ are IID $\{-1, +1\}$-valued random variables with $\mathbb{P}\{S_k = +1\} \le \mathbb{P}\{S_k = -1\}$;*
- *$(W_1, \ldots, W_n)$ and $(S_1, \ldots, S_n)$ are independent.*

Define

$$p_k := \mathbb{P}\left\{Z_k < \bigwedge_{\ell \neq k} Z_\ell\right\}.$$

Then, $p_1 \ge p_2 \ge \cdots \ge p_n$.

Proof. It is possible to write $W_k = \eta_k(Y_k)$, where $Y_1, \ldots, Y_n$ are IID random variables that each have the uniform distribution on the interval $[0, 1]$ and

$$\eta_k(y) := \inf\left\{w \in \mathbb{R}_+ : \mathbb{P}\{W_k \le w\} \ge y\right\}, \qquad y \in [0, 1].$$

It follows from the stochastic ordering assumption on $W_1, \ldots, W_n$ that $\eta_i(y) \ge \eta_j(y)$ for $y \in [0, 1]$ and $1 \le i < j \le n$.

Also, if we write p for the common value of $\mathbb{P}\{S_k = +1\}$, then

$$\begin{aligned} \mathbb{P}\{S_1 = \cdots = S_{n-2} = +1; S_{n-1} = -1\} &= p^{n-2}(1-p) \\ &\ge p^{n-1} = \mathbb{P}\{S_1 = \cdots = S_{n-1} = +1\}. \end{aligned} \tag{9}$$

The result now follows from Theorem 3.1. □

Remark 4.2. A simple consequence of Corollary 4.1 is that if $n \ge 3$, $V_1, \ldots, V_n$ are IID random variables that are symmetrically distributed (that is, the common distribution of V_k is the same as that of $-V_k$) and $c_1 \ge c_2 \ge \cdots \ge c_n > 0$ are nonnegative constants, then

$$\mathbb{P}\left\{c_i V_i < \bigwedge_{k \neq i} c_k V_k\right\} \ge \mathbb{P}\left\{c_j V_j < \bigwedge_{k \neq j} c_k V_k\right\} \tag{10}$$

for $1 \le i < j \le n$.

The discussion in Remark 3.4 addresses when inequality in (10) will be strict. Assume that $n \ge 3$ and $c_1 > c_2 > \cdots > c_n > 0$. Writing $V_k = S_k |V_k|$, $1 \le k \le n$, where $(S_1, \ldots, S_n)$ is IID $\{-1, +1\}$-valued random variables that are independent of $(|V_1|, \ldots, |V_n|)$ with $\mathbb{P}\{S_k = \pm 1\} = \frac{1}{2}$, we have

$$\mathbb{P}\{S_1 = -1\} = \frac{1}{2} > \frac{1}{4} = \mathbb{P}\{S_1 = +1; S_2 = -1\}.$$

Suppose that the common distribution of V_k, $1 \le k \le n$, is diffuse and that 0 is in the support of this distribution. Then

$$\mathbb{P}\left\{\bigvee_{k \notin \{i,j\}} c_k |V_k| \vee c_i |V_j| > c_j |V_i|, \bigvee_{k \notin \{i,j\}} c_k |V_k| \vee c_j |V_j| \le c_i |V_i|\right\} > 0 \tag{11}$$

for $1 \le i < j \le n$, which is the special case in the present setting of the sufficient condition (8) for strict inequality. To see this, note first that for all $\varepsilon > 0$ sufficiently small we have

$$\begin{aligned}
&\mathbb{P}\bigl\{c_i|V_j| > c_j|V_i|, c_j|V_j| \le c_i|V_i|, c_i|V_j| > c_j|V_j| > \varepsilon\bigr\} \\
&\quad = \mathbb{P}\left\{\frac{c_j}{c_i} < \frac{|V_j|}{|V_i|} \le \frac{c_i}{c_j}, c_i|V_j| > c_j|V_j| > \varepsilon\right\} > 0
\end{aligned}$$

whereas

$$\mathbb{P}\left\{\bigvee_{k \notin \{i,j\}} c_k|V_k| \le \varepsilon\right\} > 0$$

for all $\varepsilon > 0$. In particular, we recover Proposition 2.1

It is worth noting that (10) doesn't hold with a strict inequality under just the assumption that $V_1, \dots, V_n$ are IID random variables with a diffuse, symmetric common distribution. For example, assume that $n = 3$ and $c_1 > c_2 > c_3 > 0$ are given. Suppose that the common distribution of $|V_k|$, $1 \le k \le 3$, is supported on an interval $[a, b]$ where the intervals $c_1[a, b]$, $c_2[a, b]$, $c_3[a, b]$ are pairwise disjoint. Then

$$\begin{aligned}
&\mathbb{P}\{c_1V_1 < c_2V_2 \wedge c_2V_3\} = \mathbb{P}\{V_1 < 0\} = \frac{1}{2}, \\
&\mathbb{P}\{c_2V_2 < c_1V_1 \wedge c_3V_3\} = \mathbb{P}\{V_1 > 0, V_2 < 0\} = \frac{1}{4},
\end{aligned}$$

and

$$\mathbb{P}\{c_3V_3 < c_1V_1 \wedge c_2V_2\} = \mathbb{P}\{V_1 > 0, V_2 > 0\} = \frac{1}{4}.$$

5. Applications

5.1. Randomized experiments

Suppose we are interested in comparing n treatments. We will test each treatment on one of n individuals, which might be people, families, banks, local or national economies, or plots of land, for instance. Treatments are assigned uniformly at random to individuals: All $n!$ assignments are equally likely. The distribution of the response of individual j to treatment i is a distribution P_{ij} that is symmetric about zero, so that no treatment causes any systematic benefit or harm to any individual. Suppose for each fixed $j \in [n]$ and all $y > 0$ that $P_{ij}\{x \in \mathbb{R} : |x| > y\}$ is nonincreasing in i, so that the magnitude of the responses of a fixed individual to the various treatments are stochastically nonincreasing in the treatment number (i.e., low numbered treatments are more likely to have effects with a large magnitude than high numbered treatments). Suppose further that given the assignment of treatments to individuals the responses of the individuals are conditionally independent.

Table 1. Time for each of three workers to complete each of three tasks

	Worker $\mathcal{A}$	Worker $\mathcal{B}$	Worker $\mathcal{C}$
Task 1	$T \pm A$	$T \pm B$	$T \pm C$
Task 2	$T \pm a$	$T \pm b$	$T \pm c$
Task 3	$T \pm \alpha$	$T \pm \beta$	$T \pm \gamma$

We can represent the response to treatment i as $Z_i = S_i \eta_i(\Pi_i, U_i)$, where $S_1, \ldots, S_n$ are IID $\{-1, +1\}$-valued random variables with $\mathbb{P}\{S_i = -1\} = \mathbb{P}\{S_i = +1\} = \frac{1}{2}$; $(\Pi_1, \ldots, \Pi_n)$ is a uniform random permutation of $[n]$; $U_1, \ldots, U_n$ are IID random variables with a uniform distribution on the interval $[0, 1]$; and $\eta_i(j, \cdot)$ is the inverse of the function $y \mapsto P_{ij}\{x \in \mathbb{R} : |x| > y\}$, that is,

$$\eta_i(j, u) := \sup\bigl\{y \geq 0 : P_{ij}\bigl\{x \in \mathbb{R} : |x| \leq y\bigr\} < u\bigr\}.$$

By assumption, $\eta_1(j, u) \geq \cdots \geq \eta_n(j, u)$, and it follows from Theorem 3.1 that $p_1 \geq p_2 \geq \cdots \geq p_n$. Hence, if we think of low values of the response as desirable, then low numbered treatments are likely to appear to be the most desirable in a single instance of the experiment, even though they are also likely to appear to be the least desirable.

In order to give a simple, concrete example of this phenomenon, consider a situation in which there are three tasks of comparable difficulty that have to be completed and three workers available to do them. In terms of the setting above, the tasks are the "individuals" and the workers are the "treatments."

Number the tasks 1, 2 and 3, and designate the workers by the letters $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$. The tasks are assigned to the workers at random, with the $3! = 6$ possible allocations being equally likely. On average, the workers are equally rapid at completing a given task, but the performance of Worker $\mathcal{A}$ is more variable than that of Worker $\mathcal{B}$, which is more variable than that of Worker $\mathcal{C}$.

We model this very simply by assuming that the time taken to perform Task 1 by Worker $\mathcal{A}$ (respectively, Workers $\mathcal{B}$ and $\mathcal{C}$) is either $T - A$ or $T + A$ (respectively, $T - B$ or $T + B$, and $T - C$ or $T + C$) with equal probability, where A, B, C are positive constants. Similarly, the respective times taken by the three workers to perform Tasks 2 and 3 are $T \pm a$, $T \pm b$, $T \pm c$ and $T \pm \alpha$, $T \pm \beta$, $T \pm \gamma$, with the two alternatives in each case always being equally likely. We assume that the times taken by the workers are conditionally independent given the random allocation of tasks (that is, all $2^3 = 8$ possible choices of sign are equally likely for any particular allocation).

The relative variability of the workers' performance is modeled by taking $A > B > C$, $a > b > c$, and $\alpha > \beta > \gamma$. The ordering among these nine quantities is otherwise arbitrary. We thus have an instance of the general situation considered above with the inconsequential difference that the responses are symmetric about T rather than 0. We will explore how the probability that a particular worker finishes first depends on the ordering in detail.

Suppose the ordering is $A > B > C > a > b > c > \alpha > \beta > \gamma > 0$. Then worker $\mathcal{A}$ finishes first in the following scenarios:

1. All signs are negative and $\mathcal{A}$ is assigned task 1 (2 of 48)

2. Only the first and second signs are negative and $\mathcal{A}$ is assigned task 1, or $\mathcal{A}$ is assigned task 2 and $\mathcal{B}$ is assigned task 3 (3 of 48)
3. Only the first and third signs are negative and $\mathcal{A}$ is assigned task 1, or $\mathcal{A}$ is assigned task 2 and $\mathcal{C}$ is assigned task 3 (3 of 48)
4. Only the first sign is negative (6 of 48)
5. All signs are positive and $\mathcal{A}$ is assigned task 3 (2 of 48)

These comprise $16/48 = 1/3$ of the equally likely possibilities, so the chance that $\mathcal{A}$ finishes first is $1/3$. Similarly, worker $\mathcal{B}$ finishes first in the following scenarios:

1. All signs are negative and $\mathcal{B}$ is assigned task 1 (2 of 48)
2. Only the first and second signs are negative and $\mathcal{B}$ is assigned task 1, or $\mathcal{B}$ is assigned task 2 and $\mathcal{A}$ is assigned task 3 (3 of 48)
3. Only the second and third signs are negative and $\mathcal{B}$ is assigned task 1, or $\mathcal{B}$ is assigned task 2 and $\mathcal{C}$ is assigned task 3 (3 of 48)
4. Only the third sign is negative (6 of 48)
5. All signs are positive and $\mathcal{B}$ is assigned task 3 (2 of 48)

Again, these comprise $1/3$ of the possibilities, so the chance that $\mathcal{B}$ finishes first is $1/3$; the same is true for $\mathcal{C}$.

However, if the ordering is $A > a > \alpha > B > b > \beta > C > c > \gamma > 0$, then $\mathcal{A}$ finishes first if and only if the first sign is negative, which has chance $1/2$. For this ordering, $\mathcal{B}$ finishes first if the first sign is positive and the second is negative, which has chance $1/4$. Worker $\mathcal{C}$ finishes first if the first two signs are positive, which also has chance $1/4$.

It is possible to consider the various other possibilities that are not the same as one of these two after a relabeling of the tasks; for example, if $A > a > b > c > B > \alpha > \beta > \gamma > C > 0$, then the probability that Worker $\mathcal{A}$ finishes first is $\frac{5}{12}$, whereas the probabilities that Workers $\mathcal{B}$ and $\mathcal{C}$ finish first are both $\frac{7}{24}$. We do not present an exhaustive list of the results.

5.2. Heteroscedasticity and nonparametric tests of association

The null hypothesis for standard nonparametric (permutation-based) tests for association between two series, such as the Spearman rank correlation test, amounts to the hypothesis that one series is conditionally exchangeable given the other. Heteroscedasticity can make that null hypothesis false, even when there is no positive (resp. negative) association between the series, where by positive (resp. negative) association we mean that, in some sense, larger values of one variable tend to occur in conjunction with larger (resp. smaller) values of the other. Our results show qualitatively that this can distort the apparent p-value of permutation tests for association.

Consider a decreasing deterministic sequence $x = (x_1, \dots, x_n)$ and a sequence $Z = (Z_1, \dots, Z_n)$ whose components are independent and symmetrically distributed, but such that $|Z_i|$ stochastically dominates $|Z_j|$ for $1 \leq i < j \leq n$. We haven't given a rigorous definition of association, but x and Z are not associated in any intuitively reasonable sense of the term. However, Corollary 4.1 shows that the first component of Z is most likely to be the largest; when that occurs, the rank of the largest component of Z is aligned with the rank of the largest component of x. The full distributional details are complicated, but one might expect that an extension of this phenomenon will tend to make the Spearman rank correlation coefficient r_S take more extreme values than it would be if the null hypothesis of exchangeability held.

The following simple example from Walther [27,28] shows that the quantitative difference in probabilities can be quite striking. Let $x = (4, 3, 2, 1)$ and

$$Z = (\sigma_1 Y_1, \sigma_2 Y_2, \sigma_3 Y_3, \sigma_4 Y_4),$$

where $\{Y_i\}$ are IID standard Gaussian variables, $\sigma_1 = 2$, and $\sigma_2 = \sigma_3 = \sigma_4 = 1$. The chance that $r_S = 1$ is the chance that $Z_1 > Z_2 > Z_3 > Z_4$. If $\{Z_j\}$ were exchangeable, then that chance would be $1/24 \approx 4.17\%$. Simulation shows that in the heteroscedastic (non-exchangeable) model,

$$\mathbb{P}\big\{r_S(X, Y) = 1\big\} \approx 7\%,$$

about 68% higher. Calibrating the Spearman rank correlation test using the null hypothesis of exchangeability is misleading, because heteroscedasticity alone makes the components of Z tend to be closer to ordered than they would be under random permutations.

We can illustrate the phenomenon even more concretely with the three workers and three tasks example from Section 5.1. Note that if $A > a > \alpha > B > b > \beta > C > c > \gamma > 0$, then the distribution of the order in which the workers $\mathcal{A}, \mathcal{B}, \mathcal{C}$ finish is uniform over the four possibilities $(\mathcal{A}, \mathcal{B}, \mathcal{C})$, $(\mathcal{A}, \mathcal{C}, \mathcal{B})$, $(\mathcal{B}, \mathcal{C}, \mathcal{A})$, $(\mathcal{C}, \mathcal{B}, \mathcal{A})$ and the distribution of the Spearman rank correlation r_S between the vector of finish times for the three workers and the vector $(1, 2, 3)$ is

$$\mathbb{P}\{r_S = -1\} = \mathbb{P}\left\{r_S = -\frac{1}{2}\right\} = \mathbb{P}\left\{r_S = +\frac{1}{2}\right\} = \mathbb{P}\{r_S = +1\} = \frac{1}{4},$$

whereas if the random vector of finish times were exchangeable (that is, if we were in the usual null situation for the Spearman rank correlation test), then the distribution of r_S would be

$$\mathbb{P}\{r_S = -1\} = \frac{1}{6}, \qquad \mathbb{P}\left\{r_S = -\frac{1}{2}\right\} = \mathbb{P}\left\{r_S = +\frac{1}{2}\right\} = \frac{1}{3}, \qquad \mathbb{P}\{r_S = +1\} = \frac{1}{6},$$

so performing a Spearman rank correlation test would be likely to result in the conclusion that there is a positive (or negative) association between a worker's label and the worker's finish time.

Our results do not predict the magnitude of the distortion of the null distribution of r_S, but they do suggest that there will be such a distortion quite generally when one sequence is heteroscedastic with an ordering of the degree of dispersion that matches the ordering of magnitudes of the other, even when the components of the first sequence are independent and have equal means.

6. Discussion and conclusions

We have presented general conditions on a random vector $(Z_1, Z_2, \ldots, Z_n)$ that guarantee that the probabilities $p_i := \mathbb{P}\{Z_i < \bigwedge_{j \neq i} Z_j\}$ satisfy $p_1 \geq p_2 \geq \cdots \geq p_n$; that is, that the probability the ith coordinate is the smallest is decreasing in i. Analogous results hold for the the probability that the ith coordinate is the largest. The general conclusion is that "Fortune favors the bold," and that even if $\mathbb{P}\{Z_i > Z_j\} = \mathbb{P}\{Z_i < Z_j\}$ for $1 \leq i \neq j \leq n$, so that no coordinate is systematically larger than another, we can still have situations in which such an ordering will occur because the variability of Z_i decreases with i. Our results give technical precision to the intuition embodied by the proverb. We emphasize that our results do not require the explicit computation of the probability that Z_i is extreme.

Presumably, even more general conditions that determine the ranks of the probabilities that each random variable will be extremal could be derived. Similarly, we have considered inequalities among the probabilities that different items will be most favored, but it should also be possible to derive inequalities among the probabilities that various subsets of the items will have various subsets of the ranks, not just the chances that each individual item is best. These remain open problems.

Appendix A: Three independent Gaussians

Suppose that X, Y, Z are independent zero mean Gaussian random vectors with variances $\alpha^2 > \beta^2 > \gamma^2 > 0$. Observe that $\mathbb{P}\{X < Y \wedge Z\} = \mathbb{P}\{(Y - X, Z - X) \in Q\}$, where Q is the positive quadrant $\{(s, t) \in \mathbb{R}^2 : s > 0, t > 0\}$. The variance-covariance matrix of the pair $(Y - X, Z - X)$ is

$$\Sigma := \begin{pmatrix} \beta^2 + \alpha^2 & \alpha^2 \\ \alpha^2 & \gamma^2 + \alpha^2 \end{pmatrix}.$$

We can write

$$(Y - X, Z - X) = (V, W)\Sigma^{\frac{1}{2}},$$

where $\Sigma^{\frac{1}{2}}$ is the positive definite square root of the matrix Σ and (U, V) is a pair of independent standard Gaussian random variables. The image of the quadrant Q under the linear map defined by $\Sigma^{-\frac{1}{2}}$ is a wedge with boundary given by the images of the two positive coordinate axes. Some algebra shows that

$$\begin{aligned} &\frac{((1,0)\Sigma^{-\frac{1}{2}}) \cdot ((0,1)\Sigma^{-\frac{1}{2}})}{\sqrt{((1,0)\Sigma^{-\frac{1}{2}}) \cdot ((1,0)\Sigma^{-\frac{1}{2}})}\sqrt{((0,1)\Sigma^{-\frac{1}{2}}) \cdot ((0,1)\Sigma^{-\frac{1}{2}})}} \\ &\quad = -\frac{\alpha^2}{\sqrt{(\beta^2 + \alpha^2)(\gamma^2 + \alpha^2)}}, \end{aligned} \tag{12}$$

where we use $a \cdot b$ to denote the usual inner product of two vectors a and b.

It follows from the rotational symmetry of the distribution of (U, V) that

$$\mathbb{P}\{X < Y \wedge Z\} = \frac{1}{2\pi} \arccos\left(-\frac{\alpha^2}{\sqrt{(\beta^2 + \alpha^2)(\gamma^2 + \alpha^2)}}\right).$$

A similar formula holds for $\mathbb{P}\{Y < X \wedge Z\}$ (resp. $\mathbb{P}\{Z < X \wedge Y\}$) by interchanging the roles of α^2 and β^2 (resp. α^2 and γ^2).

Some more algebra shows that

$$\frac{\alpha^4}{(\beta^2+\alpha^2)(\gamma^2+\alpha^2)} - \frac{\beta^4}{(\alpha^2+\beta^2)(\gamma^2+\beta^2)} = \frac{(\alpha^2-\beta^2)(\alpha^2\beta^2+\beta^2\gamma^2+\alpha^2\gamma^2)}{(\alpha^2+\beta^2)(\alpha^2+\gamma^2)(\beta^2+\gamma^2)} > 0,$$

and so

$$\mathbb{P}\{X < Y \wedge Z\} > \mathbb{P}\{Y < X \wedge Z\}.$$

Similarly,

$$\mathbb{P}\{Y < X \wedge Z\} > \mathbb{P}\{Z < X \wedge Y\}.$$

Appendix B: The minimum of three bilateral exponentials

Given a dispersion parameter $\theta > 0$, write $f_\theta(x) := \frac{1}{2\theta} e^{-\frac{|x|}{\theta}}$ for the density of the corresponding bilateral exponential distribution. Note that

$$\int_x^\infty f_\theta(t)\,dt = \begin{cases} \frac{1}{2}\left(1 - e^{-\frac{|x|}{\theta}}\right) + \frac{1}{2}, & x < 0, \\ \frac{1}{2} e^{-\frac{|x|}{\theta}}, & x \geq 0. \end{cases}$$

Suppose that X, Y, Z are independent real-valued random variables with respective bilateral exponential densities f_a, f_b, f_c, where the parameters satisfy $a > b > c > 0$, so that X is more dispersed than Y, which is more dispersed than Z.

An explicit integration shows that

$$\begin{aligned} \mathbb{P}\{X < Y \wedge Z\} &= \int_{-\infty}^\infty \mathbb{P}\{Y > x\}\mathbb{P}\{Z > x\}\mathbb{P}\{X \in dx\} \\ &= \frac{2a^3b + a^2b^2 + 2a^3c + 5a^2bc + 2ab^2c + a^2c^2 + 2abc^2 + b^2c^2}{4(a+b)(a+c)(ab+bc+ac)}. \end{aligned} \tag{13}$$

A similar expression for $\mathbb{P}\{Y < X \wedge Z\}$ (resp. $\mathbb{P}\{Z < X \wedge Y\}$) follows by interchanging the roles of a and b (resp. a and c).

It follows that

$$\mathbb{P}\{X < Y \wedge Z\} - \mathbb{P}\{Y < X \wedge Z\} = \frac{(a-b)(b^2c^2 + a^2(b+c)^2 + abc(2b+3c))}{4(a+b)(a+c)(b+c)(ab+bc+ac)} > 0$$

and

$$\mathbb{P}\{Y < X \wedge Z\} - \mathbb{P}\{Z < X \wedge Y\} = \frac{(b-c)(b^2c^2 + 2abc(b+c) + a^2(b^2+3bc+c^2))}{4(a+b)(a+c)(b+c)(ab+bc+ac)} > 0,$$

so

$$\mathbb{P}\{X < Y \wedge Z\} > \mathbb{P}\{Y < X \wedge Z\} > \mathbb{P}\{Z < X \wedge Y\}.$$

Appendix C: Avoiding small-school bias

We consider how one might correct for small-school bias in a model problem involving standardized testing.

There are n schools of different sizes. The schools draw their students at random, independently, from the same infinite population. At the beginning of the school year, the scores students would get on the standardized test are modeled as IID. Attending school i for the year increases the expected value of a student's test score by s_i, $i = 1, \dots, n$. Let S_{ij} be the score of the jth student at school i at the end of the year. In this model, $\{S_{ij} - s_i\}$ are IID.

We wish to award a "best school" prize to exactly one school, based on student scores on the standardized test. We want the scheme to be *fair*, in that if $s_1 = s_2 = \cdots = s_n$, then all schools are equally likely to win.

We want the scheme to be *valid* in the sense that if if $s_i > s_j$, then school i is more likely to be picked as "best school" than s_j.

The proposed solution (suggested to us by Alex Rivest) is both fair and valid.

Let m be the smallest school size. The summary score for school i is the average test score of a random sample of m students at school i. The prize is awarded to the school with the highest summary score.

The method is fair, since the summary score for each school is determined by a random size-m set of students: If $\{s_i\}$ are equal, the summary scores of the n schools are IID, and every school is equally likely to rank first. The method is valid, since the score of school i is stochastically larger than the score for school j if $s_i > s_j$.

While this method is fair and valid, it relies on a subsample, so it might not maximize the probability that the prize is awarded to the school with the largest s_i among all far and valid methods. Finding a better method that is both fair and valid is an open problem.

Acknowledgments

SNE supported in part by NSF grants DMS-09-07630 and DMS-15-12933 and NIH grant 1R01GM109454, RLR supported in part by NSF Science & Technology Center grant CCF-0939370. We thank Alex Rivest for suggesting the procedure given in Appendix C for correcting small-school bias.

References

[1] Agrawal, S. and Goyal, N. (2011). Analysis of Thompson sampling for the multi-armed bandit problem. *CoRR*. Available at arXiv/abs/1111.1797.

[2] Böckenholt, U. (1993). Applications of Thurstonian models to ranking data. In *Probability Models and Statistical Analyses for Ranking Data* (*Amherst*, *MA*, 1990). *Lect. Notes Stat.* **80** 157–172. New York: Springer. MR1237206

[3] Böckenholt, U. (2006). Thurstonian-based analyses: Past, present, and future utilities. *Psychometrika* **71** 615–629. MR2312235

[4] Craig, P. (2008). A new reconstruction of multivariate normal orthant probabilities. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **70** 227–243. MR2412640

[5] Daniels, H.E. (1950). Rank correlation and population models. *J. Roy. Statist. Soc. Ser. B.* **12** 171–181. MR0040629

[6] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory. Springer Series in Operations Research and Financial Engineering: An Introduction*. New York: Springer. MR2234156

[7] Diaconis, P. (1988). *Group Representations in Probability and Statistics. Institute of Mathematical Statistics Lecture Notes – Monograph Series* **11**. Hayward, CA: IMS. MR0964069

[8] Fill, J.A. (1996). An exact formula for the move-to-front rule for self-organizing lists. *J. Theoret. Probab.* **9** 113–160. MR1371073

[9] Fligner, M.A. and Verducci, J.S., eds. (1993) *Probability Models and Statistical Analyses for Ranking Data. Lecture Notes in Statistics* **80**. New York: Springer. Papers from the conference held at the University of Massachusetts, Amherst, Massachusetts, June 8–13, 1990. MR1237197

[10] Gassmann, H.I., Deák, I. and Szántai, T. (2002). Computing multivariate normal probabilities: A new look. *J. Comput. Graph. Statist.* **11** 920–949. MR1944268

[11] Gumbel, E.J. (2004). *Statistics of Extremes*. Mineola, NY: Dover. MR2080344

[12] Kotz, S. and Nadarajah, S. (2000). *Extreme Value Distributions. Theory and Applications*. London: Imperial College Press. MR1892574

[13] Laslier, J.-F. (2006). *Strategic Approval Voting in a Large Electorate*. Marseille, France: Institut D'Economie Publique. IDEP Working Papers No. 405. Available at http://halshs.archives-ouvertes.fr/docs/00/12/17/51/PDF/stratapproval4.pdf.

[14] Luce, R.D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley; London: Chapman & Hall. MR0108411

[15] Marden, J.I. (1995). *Analyzing and Modeling Rank Data. Monographs on Statistics and Applied Probability* **64**. London: Chapman & Hall. MR1346107

[16] Miwa, T., Hayter, A.J. and Kuriki, S. (2003). The evaluation of general non-centred orthant probabilities. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **65** 223–234. MR1959823

[17] Plackett, R.L. (1975). The analysis of permutations. *J. Roy. Statist. Soc. Ser. C Appl. Statist.* **24** 193–202. MR0391338

[18] Resnick, S.I. (2008). *Extreme Values, Regular Variation and Point Processes. Springer Series in Operations Research and Financial Engineering*. New York: Springer. Reprint of the 1987 original. MR2364939

[19] Rivest, R. (1976). On self-organizing sequential search heuristics. *Comm. ACM* **19** 63–67. MR0408303

[20] Savage, I.R. (1957). Contributions to the theory of rank order statistics – the "trend" case. *Ann. Math. Statist* **28** 968–977. MR0093876

[21] Silverberg, A.R. (1980). *Statistical Models for Q-Permutations*. Ann Arbor, MI: ProQuest LLC. Thesis (Ph.D.) – Princeton University. MR2630352

[22] Silverberg, A.R. (1984). Statistical models for q-permutations. *Proc. Biopharm. Sec. Amer. Statist. Assoc.* 107–112.

[23] Thompson, W.R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika* **25** 285–294.

[24] Thurstone, L.L. (1927). A law of comparative judgement. *Psychological Reviews* **34** 273–286.

[25] Thurstone, L.L. (1931). Rank order as a psycho-physical method. *J. Exp. Psychol.* **14** 187–201.

[26] Wainer, H. (2007). The most dangerous equation. *Am. Sci.* **95** 249–256. Available at http://www.americanscientist.org/issues/feature/2007/3/the-most-dangerous-equation/1.

[27] Walther, G. (1997). Absence of correlation between the solar neutrino flux and the sunspot number. *Phys. Rev. Lett.* **79** 4522–4524.

[28] Walther, G. (1999). On the solar-cycle modulation of the Homestake solar neutrino capture rate and the shuffle test. *Ap*. *J*. **513** 990–996.
[29] Yellott, J.I. Jr. (1977). The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution. *J*. *Math*. *Psych*. **15** 109–144. MR0449795

Received July 2014 and revised December 2016