Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 137 (2007) 1838 – 1850
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Asymptotic efficiency of majority rule relative to rank-sum method for selecting the best population

Samuel S. Wu[a, b, *], David H. Annis[c]

[a] College of Medicine, University of Florida, Gainesville, FL 32610, USA
[b] Malcom Randall VA Medical Center, Gainesville, FL 32608, USA
[c] Operations Research Department, Naval Postgraduate School, Monterey, CA 93943, USA

Received 27 May 2005; received in revised form 9 February 2006; accepted 28 March 2006
Available online 21 June 2006

Abstract

The ranking and selection problem has been well-studied in the case of continuous responses. In this paper, we address the situation in which continuous responses are replaced by discrete orderings. When individuals in the population provide exhaustive rank-orderings of the alternatives, two common decision rules are the majority rule and the rank-sum method. In the former case, the alternative receiving the most first-place votes is declared superior, while in the latter, the alternative with the smallest rank-sum is deemed the best. Both the Pitman efficiencies and the lower bounds on Bahadur efficiencies of the majority rule relative to the rank-sum method are derived, assuming that the rank data are generated from either the Plackett–Luce or the translative strengths models. In addition, finite sample properties of the two methods are compared with the maximum likelihood approach through simulation studies. Our results suggest two things. First, when it is substantially more difficult to obtain a complete rank-ordering than simply the top choice, the majority rule performs adequately and efforts would be better spent asking many voters to provide top choice rather than fewer voters to provide complete orderings. Second, the rank-sum rule compares favorably to, and is substantially more robust than, the maximum likelihood approach.

Keywords: Asymptotic efficiency; Majority rule; Rank sum method; Plackett–Luce model; Translative strengths model

1. Introduction

The selection and ranking problem has been extensively studied since the 1950s, including pioneering works of Bechhofer (1954) and Gupta (1956). That these problems have been studied in such detail points to their wide-spread relevance to many applied problems. In many analyses, the observed responses (on which ranking and selection are based) are recorded on a continuous scale. In this paper, we address the situation in which continuous responses are replaced by their corresponding discrete rank-orders. In this case, suppose a committee of size n is charged with choosing the best among m potential alternatives. To this end, each voter submits an exhaustive rank-ordering of the alternatives, with ties prohibited (i.e., the ranking is strict). There are many potential ways to determine a winner in such a circumstance. For example, the alternative receiving the most first-place votes may be chosen. Such a system is commonly referred to as the majority (MJ) rule and is in wide use. On the other hand, when the number of alternatives,

* Corresponding author. College of Medicine, University of Florida, Gainesville, FL 32610, USA.
E-mail addresses: samwu@biostat.ufl.edu (S.S. Wu), annis@nps.edu (D.H. Annis).

0378-3758/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2006.03.012

m, is large, the rank-sum (RS) method may be preferable. This procedure selects the alternative that has the smallest rank sum over all votes. (Low ranks are considered preferable to higher ones—i.e., first-place corresponds to the best alternative, while mth-place corresponds to the worst).

Throughout, we assume that each of the alternatives possesses an inherent desirability or appeal, and that these determine how likely a particular alternative is to be preferred to another (or a group of others). In this paper, we investigate the asymptotic efficiency of the MJ rule relative to the RS method in determining the correct winner (i.e., the alternative whose desirability is largest), first under the Plackett–Luce model (Luce, 1959; Plackett, 1975) and then under the translative strengths model (Stern, 1990; Marden, 1995).

We derive two criteria to quantify asymptotic efficiency. The first is a Pitman (1948) type asymptotic efficiency which measures the asymptotic ratio of sample sizes required to achieve equal limiting power against the same sequence of alternative strengths whose differences tend to zero. The second measure is a Bahadur (1971) type efficiency, namely, the limiting ratio of sample sizes such that the probability of selecting the best alternative tends to 1 at the same rate, for any fixed alternative strengths.

The remainder of the paper is organized as follows. Section 2 introduces the Plackett–Luce and translative strengths models. Sections 3 and 4 derive the Pitman efficiencies and the lower bounds on Bahadur efficiencies of the MJ rule relative to the RS method first under the Plackett–Luce model and then assuming a translative strengths formulation. Section 5 presents results of a series of simulations comparing the aforementioned approaches for various values of n (voting committee size) and m (alternatives under consideration). A discussion of main results is given in Section 6. Appendix A states a few properties of the Plackett–Luce model. Longer proofs are left for Appendix B, and Appendix C derives quantities related to Pitman efficiency for some common translative strengths models.

2. Modeling rank data

Let $\pi(\cdot)$ denote a permutation of the integers $\{1, 2, \ldots, m\}$ which corresponds to a rank-ordering of the alternatives. In particular, $\pi(r) = s$ denotes alternative s receiving rank r. The complete rank-ordering is expressed as $\pi(1) \to \pi(2) \to \cdots \to \pi(m)$, indicating that the alternative receiving the first-place vote, $\pi(1)$, is preferred to the alternative receiving the second-place vote, $\pi(2)$, and so forth. Implicit in this notation is the idea of transitivity of an individual's preference, namely that together $i \to j$ and $j \to k$ are sufficient to guarantee $i \to k$.

The n voters generate their rank-orderings as follows. Let $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_m)' \in \mathscr{R}_+^m$ denote the desirabilities of alternatives. Without loss of generality, we assume that the m alternatives are labeled according to their desirability ($\gamma_1 \geqslant \gamma_2 \geqslant \cdots \geqslant \gamma_m$) throughout. Suppose $Y_j = (Y_{1,j}, Y_{2,j}, \ldots, Y_{m,j})'$, $j = 1, \ldots, n$ are i.i.d. random vectors taking values in $\mathscr{R}^m$ and consisting of independent components distributed according to $P_{\gamma_1}, P_{\gamma_2}, \ldots, P_{\gamma_m}$, respectively. The data we observe are orderings $\pi_j = \kappa(Y_j)$, where κ is a measurable mapping from $\mathscr{R}^m$ to Π, the set of all $m!$ permutations of alternatives.

If P_{γ_i} is the exponential distribution with rate γ_i (mean $1/\gamma_i$) and κ is given by the *increasing* order of the $Y_{i,j}$'s (that is, $\pi_j = (i_1, i_2, \ldots, i_m)$ if and only if $Y_{i_1,j} < Y_{i_2,j} < \cdots < Y_{i_m,j}$), then the ordering vectors follow the Plackett–Luce model (Luce, 1959; Plackett, 1975), which is perhaps the most common characterization of permutation probabilities. This model, denoted by PL($\boldsymbol{\gamma}$), assumes that each voter independently submits a ranking of all alternatives, and the probability assigned to a particular ordering π, is

$$P_{\boldsymbol{\gamma}}[\pi(1) \to \pi(2) \to \cdots \to \pi(m)] = \prod_{i=1}^{m} \frac{\gamma_{\pi(i)}}{\gamma_{\pi(i)} + \cdots + \gamma_{\pi(m)}}. \tag{1}$$

It is easy to verify that $P_{\boldsymbol{\gamma}}[\pi] = P_{c\boldsymbol{\gamma}}[\pi]$ for any constant $c > 0$. In other words, the probability of the ordering is solely dependent on the *relative* (rather than absolute) desirability of an alternative. To ensure a unique solution, we always assume $\sum_{i=1}^{m} \gamma_i = 1$ for this model. Our motivation of the PL($\boldsymbol{\gamma}$) model is not unique. Indeed, Stern (1990) shows that this model can be obtained in many different ways. In addition, the Plackett–Luce model possesses an appealing internal consistency property. Furthermore, it satisfies Luce's choice axiom and has a connection to the proportional hazard model (Cox, 1972). More details are given in Appendix A.

Another class of models for rank data involves a location-family with possibly different location parameters (Daniel, 1950; Marden, 1995). In this case, $P_{\gamma_1}, P_{\gamma_2}, \ldots, P_{\gamma_m}$ have cumulative distribution functions $F(y - \gamma_1), F(y - \gamma_2), \ldots, F(y - \gamma_m)$ and κ is a one-to-one mapping given by the *decreasing* order of the $Y_{i,j}$'s, that is,

$\pi_j = (i_1, i_2, \ldots, i_m)$ if and only if $Y_{i_1,j} > Y_{i_2,j} > \cdots > Y_{i_m,j}$. This class of models is quite broad, encompassing, among others, the normal, logistic, exponential and uniform distributions. We refer to these models, denoted by $\mathrm{TS}(\gamma, F)$, as transitive strengths models to reflect that the distributions governing the permutations follow a base distribution F and differ only by translation determined by γ.

3. Asymptotic efficiency under Plackett–Luce model

3.1. Pitman asymptotic relative efficiency

In the decision context, determining the correct choice (i.e., alternative 1, whose desirability exceeds all others') is analogous to rejecting a null hypothesis assuming equality of alternatives in favor of a hypothesis that alternative 1 is the best among its peers. Therefore, one measure of effectiveness of a decision rule is the power of the procedure. Clearly, as the sample sizes tend to infinity, both the MJ rule and the RS method are capable of detecting increasingly smaller differences between alternatives. Pitman relative efficiency (PRE) measures test performance against local hypotheses (i.e., those which are very "close" to the null in some sense). Specifically, it is the ratio of sample sizes required to achieve a fixed power level against a sequence of hypotheses approaching to the null.

Before deriving the asymptotic efficiencies of the MJ rule relative to the RS method, we present results concerning the limiting behavior of the number of first-place votes and the rank-sums when rank-orderings follow the Plackett–Luce model. Let $\boldsymbol{\gamma}_s = (\gamma_{s,1}, \gamma_{s,2}, \ldots, \gamma_{s,m})'$ be a sequence of desirability parameters satisfying $\sum_{i=1}^m \gamma_{s,i} = 1$, and define $\boldsymbol{\delta}_s = (\delta_{s,2}, \ldots, \delta_{s,m})'$ such that $\delta_{s,i} = \gamma_{s,1} - \gamma_{s,i}$ and $\boldsymbol{\eta}_s = (\eta_{s,2}, \ldots, \eta_{s,m})'$ where $\eta_{s,i} = (\gamma_{s,1} - \gamma_{s,i}) \sum_{k=1}^m \{\gamma_{s,k}/[(\gamma_{s,1} + \gamma_{s,k})(\gamma_{s,i} + \gamma_{s,k})]\}$. Suppose we observe n_s independent votes governed by $\mathrm{PL}(\boldsymbol{\gamma}_s)$. Let $\boldsymbol{W}_s = (W_{s,2}, W_{s,3}, \ldots, W_{s,m})'$ be the $(m-1)$-vector whose entries contain the differences in first-place votes earned by alternative 1 and alternatives 2 through m, respectively. Furthermore, define $\boldsymbol{D}_s = (D_{s,2}, D_{s,3}, \ldots, D_{s,m})'$ as the $(m-1)$-vectorwhose entries contain the differences in rank-sums assigned to alternatives 2 through m, respectively, and alternative 1. Then the differences in first-place votes, $\boldsymbol{W}_s$, and rank sums, $\boldsymbol{D}_s$, have limiting multivariate normal distributions given by Lemma 1.

Lemma 1. *Suppose* $\lim_{s\to\infty} n_s = \infty$ *and there exist positive constants* $\theta_i, 2 \leqslant i \leqslant m$ *such that* $\lim_{s\to\infty} b_s \delta_{s,i} = \theta_i$ *for a fixed sequence* $b_s \to \infty$, *then, as* $s \to \infty$, *we have*: (a) $n_s^{-1/2}(\boldsymbol{W}_s - n_s \boldsymbol{\delta}_s) \to N(\mathbf{0}, (1/m)\Sigma)$, *and* (b) $n_s^{-1/2}(\boldsymbol{D}_s - n_s \boldsymbol{\eta}_s) \to N(\mathbf{0}, ((m(m+1))/12)\Sigma)$, *where*

$$\Sigma = \begin{pmatrix} 2 & 1 & \cdots & 1 \\ 1 & 2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \cdots & 1 & 2 \end{pmatrix}. \tag{2}$$

A proof is provided in Appendix B. Using Lemma 1, we are ready to prove the following theorem.

Theorem 1. *Let* $n_s^{\mathrm{MJ}}, n_s^{\mathrm{RS}}$ *be the sample sizes needed to achieve a given power under the Plackett–Luce model* $\mathrm{PL}(\boldsymbol{\gamma}_s)$ *for the* MJ *rule and the* RS *method, respectively. Then, the Pitman relative efficiency* (PRE) *of the procedures is* $\lim_{s\to\infty} (n_s^{\mathrm{RS}}/n_s^{\mathrm{MJ}}) = 4(m+1)/3m^2$, *where m is the number of alternatives.*

Proof. For any given power $\beta > (1/m)$ of identifying the correct choice, let $t > 0$ be the number such that $P(Z_i \leqslant t\theta_i, 2 \leqslant i \leqslant m) = \beta$, where $(Z_2, \ldots, Z_m)' \sim N(\mathbf{0}, \Sigma)$ and Σ is defined in Eq. (2). Such t must be unique because this probability increases monotonically from $1/m$ to 1, as t goes from 0 to ∞.

With a voting committee of size n_s, the MJ rule selects the correct alternative precisely when the difference in first-place votes between alternatives 1 and i, $W_{s,i}$, is positive for all $2 \leqslant i \leqslant m$. Note that

$$\begin{aligned} P(W_{s,i} > 0, 2 \leqslant i \leqslant m) &= P\left[(-W_{s,i} + n_s\delta_{s,i})\sqrt{m/n_s} - \sqrt{mn_s}\delta_{s,i} < 0,\ 2 \leqslant i \leqslant m\right] \\ &= P\left[(-W_{s,i} + n_s\delta_{s,i})\sqrt{m/n_s} + \left(t\theta_i - \sqrt{mn_s}\delta_{s,i}\right) < t\theta_i, 2 \leqslant i \leqslant m\right]. \end{aligned} \tag{3}$$

Now, by Lemma 1, $(-\boldsymbol{W}_s + n_s\boldsymbol{\delta}_s)\sqrt{m/n_s} \to N(\mathbf{0}, \Sigma)$, so when this probability is β, we have $n_s^{\mathrm{MJ}} = t^2 b_s^2/m + \mathrm{o}(b_s^2)$.

The RS method selects the correct alternative precisely when $D_{s,i}$ is positive for all $2 \leqslant i \leqslant m$. In particular, this probability is given by

$$P(D_{s,i} > 0, 2 \leqslant i \leqslant m) = P\left[(-D_{s,i} + n_s\eta_{s,i})\sqrt{\frac{12}{m(m+1)n_s}} - \sqrt{\frac{12n_s}{m(m+1)}}\eta_{s,i} < 0 \ \forall i\right]$$
$$= P\left[(-D_{s,i} + n_s\eta_{s,i})\sqrt{\frac{12}{m(m+1)n_s}} + \left(t\theta_i - \sqrt{\frac{12n_s}{m(m+1)}}\eta_{s,i}\right) < t\theta_i \ \forall i\right]. \quad (4)$$

It is easy to check that $\lim_{s\to\infty} b_s\eta_{s,i} = m^2\theta_i/4$. Also, $(-\boldsymbol{D}_s + n_s\boldsymbol{\eta}_s)\sqrt{12/m(m+1)n_s} \to N(\mathbf{0}, \Sigma)$ by Lemma 1. These facts imply that when this probability is β, we have $n_s^{\mathrm{RS}} = ((4(m+1))/3m^3)t^2b_s^2 + \mathrm{o}(b_s^2)$. Comparing the two, we see that $\lim_{s\to\infty}(n_s^{\mathrm{RS}}/n_s^{\mathrm{MJ}}) = (4(m+1))/3m^2$, which completes the proof. □

3.2. Bahadur asymptotic relative efficiency

While Pitman efficiency measures relative sample sizes required to achieve a fixed power against an arbitrarily small deviation from the null, the Bahadur type efficiency measures, for a fixed vector of desirabilities $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_m)'$, the limiting ratio of sample sizes such that the power associated with selecting the correct alternative will converge to 1 at the same rate. Though perhaps more natural, the Bahadur efficiency is much more difficult to compute explicitly, and therefore we give a lower bound on its value. A fundamental result on which the remaining arguments rely is given by Chernoff (1952) and stated below.

Theorem 2 (*Chernoff, 1952*). *Let $M(t)$ be the moment generating function* (m.g.f.) *of a random variable X, and define $f(a) = \sup\{at - \log M(t) : t \geqslant 0$ and $M(t) < \infty\}$. Suppose $X_1, X_2, \ldots$ are independent and identically distributed on $[-\infty, \infty)$ with m.g.f. $M(t)$. Then if $a_n \to a \in (-\infty, \infty)$ as $n \to \infty$ such that $P(X > a) > 0$, then* $\lim_{n\to\infty} n^{-1}\log P(\sum_{k=1}^n X_k/n \geqslant a_n) = -f(a)$.

Let $W_i^{(j)}$ be the difference in first-place votes cast by voter j between alternative 1 and alternative i. Since the m.g.f. of $-W_i^{(j)}$ is $M_W(t) = E[\mathrm{e}^{-tW_i^{(j)}}] = \mathrm{e}^{-t}\gamma_1 + \mathrm{e}^t\gamma_i + (1 - \gamma_1 - \gamma_i)$, it follows that $-f(0) = \inf[\log M_W(t)] = \log[1 - (\sqrt{\gamma_1} - \sqrt{\gamma_i})^2]$. Therefore, by Chernoff's Theorem, $n^{-1}\log P(-\sum_{j=1}^n W_i^{(j)} \geqslant 0) \to \log[1 - (\sqrt{\gamma_1} - \sqrt{\gamma_i})^2]$. Hence, the probability that alternative i has more first-place votes than alternative 1 is $P(\sum_{j=1}^n W_i^{(j)} \leqslant 0) = \mathrm{O}\{[1 - (\sqrt{\gamma_1} - \sqrt{\gamma_i})^2]^n\}$, which is of the highest order when $i = 2$. In summary, we have the following Theorem.

Theorem 3. *Under the Plackett–Luce model* PL($\boldsymbol{\gamma}$), *the* MJ *rule selects the correct alternative with probability* $1 - \mathrm{O}\{[1 - (\sqrt{\gamma_1} - \sqrt{\gamma_2})^2]^n\}$, *as $n \to \infty$.*

For the rank-sum method, we let $D_i^{(j)}$ be the difference in rank-sums assigned by voter j to alternative i and alternative 1. In this case, a closed form for the m.g.f. of $-D_i^{(j)}$ is unwieldy. However, for $t \geqslant 0$, $M_D(t) = E[\mathrm{e}^{-tD_i^{(j)}}] \geqslant P(D_i^{(j)} < 0) = P(Y_{1,j} > Y_{i,j}) = \gamma_i/(\gamma_i + \gamma_1)$. In fact, a stricter lower bound is given by

$$M_D(t) \geqslant P(Y_{1,j} > Y_{i,j})\mathrm{e}^t + P(Y_{1,j} < Y_{i,j})\mathrm{e}^{-(m-1)t} = \frac{\gamma_i}{\gamma_i + \gamma_1}\mathrm{e}^t + \frac{\gamma_1}{\gamma_i + \gamma_1}\mathrm{e}^{-(m-1)t}$$
$$\geqslant m(m-1)^{1/m-1}\left(\frac{\gamma_i}{\gamma_i + \gamma_1}\right)^{1-1/m}\left(\frac{\gamma_1}{\gamma_i + \gamma_1}\right)^{1/m}.$$

Thus, by Chernoff's Theorem,

$$\lim_{n\to\infty} n^{-1}\log P\left(-\sum_{j=1}^n D_i^{(j)} \geqslant 0\right) \geqslant \log\left[m(m-1)^{1/m-1}\left(\frac{\gamma_i}{\gamma_i + \gamma_1}\right)^{1-1/m}\left(\frac{\gamma_1}{\gamma_i + \gamma_1}\right)^{1/m}\right].$$

Hence, the probability that alternative i has smaller rank sum than alternative 1 is at least of order $[m(m-1)^{1/m-1}(\gamma_i/(\gamma_i+\gamma_1))^{1-1/m}(\gamma_1/(\gamma_i+\gamma_1))^{1/m}]^n$, which is of the highest order when $i=2$. Thus the probability of selecting the best alternative using the rank-sum method converges to 1 slower than $1-\mathrm{O}\{[m(m-1)^{1/m-1}(\gamma_2/(\gamma_2+\gamma_1))^{1-1/m}(\gamma_1/(\gamma_2+\gamma_1))^{1/m}]^n\}$. Combining this result with Theorem 3 yields the following lower-bound for Bahadur asymptotic efficiency.

Theorem 4. *Under the Plackett–Luce model* PL($\boldsymbol{\gamma}$), *the Bahadur relative efficiency* (BRE) *of the* MJ *rule to* RS *method satisfies*

$$\mathrm{BRE} \geqslant \log\left[1-\left(\sqrt{\gamma_1}-\sqrt{\gamma_2}\right)^2\right] \Big/ \log\left[m(m-1)^{1/m-1}\left(\frac{\gamma_2}{\gamma_2+\gamma_1}\right)^{1-1/m}\left(\frac{\gamma_1}{\gamma_2+\gamma_1}\right)^{1/m}\right]. \tag{5}$$

Note that the power of choosing the best alternative converges to 1 exponentially fast for both decision rules, and when m is large, BRE is in the order of $1/m$. Unfortunately, when $\gamma_1 \approx \gamma_2$, the lower bound given in Eq. (5) is overly conservative and therefore of little practical use.

4. Asymptotic efficiency under translative strengths model

4.1. Pitman asymptotic relative efficiency

Let $\boldsymbol{\gamma}_s$ be a sequence of desirabilities satisfying $\lim_{s\to\infty} b_s(\gamma_{s,1}-\gamma_{s,i})=\theta_i$ for a fixed sequence $b_s\to\infty$, where θ_i, $2\leqslant i\leqslant m$ are positive constants. Under the translative strengths model TS($\boldsymbol{\gamma}_s$, F), we observe n_s rank-orderings of the m alternatives. We assume that $\lim_{s\to\infty} n_s=\infty$. Similar to Section 3.1, we define $\boldsymbol{\delta}_s=(\gamma_{s,1}-\gamma_{s,2},\ldots,\gamma_{s,1}-\gamma_{s,m})'$ as the vector of desirability differences between alternative 1 and alternatives 2 through m, respectively; $\boldsymbol{W}_s=(W_{s,2},W_{s,3},\ldots,W_{s,m})'$ as the vector whose entries contain the differences in first-place votes earned by alternative 1 and other alternatives; and $\boldsymbol{D}_s=(D_{s,2},D_{s,3},\ldots,D_{s,m})'$ as the vector whose entries contain the differences in rank-sums assigned to alternatives 2 through m and alternative 1. The following lemma gives their limiting multivariate normal distributions under the TS($\boldsymbol{\gamma}_s$, F) model. (Compare, for example, to the asymptotic distributions under the PL($\boldsymbol{\gamma}_s$) model given by Lemma 1).

Lemma 2. *Under the preceding assumptions, if* $f(y)=F'(y)$, *then as* $s\to\infty$: (a) $n_s^{-1/2}(\boldsymbol{W}_s-n_s\boldsymbol{\delta}_s m\int f^2F^{m-2})\to N(\boldsymbol{0},(1/m)\Sigma)$, *and* (b) $n_s^{-1/2}(\boldsymbol{D}_s-n_s\boldsymbol{\delta}_s m\int f^2)\to N(\boldsymbol{0},((m(m+1))/12)\Sigma)$, *where* Σ *is given in Eq.* (2).

A proof of the lemma is provided in Appendix B. Therefore, following similar arguments as given in the proof of Theorem 1, we can show that the sample sizes needed to achieve a given power are $n_s^{\mathrm{MJ}}=(1/(m^3[\int f^2F^{m-2}]^2))t^2b_s^2+\mathrm{o}(b_s^2)$ for the MJ rule and $n_s^{\mathrm{RS}}=((m+1)/12m[\int f^2]^2)t^2b_s^2+\mathrm{o}(b_s^2)$ for the RS method, respectively.

Combining the above results yields the following theorem concerning Pitman asymptotic relative efficiency for translative strengths models.

Theorem 5. *If* n_s^{MJ} *and* n_s^{RS} *are the sample sizes needed to achieve a given power under the translative strengths model* TS($\boldsymbol{\gamma}_s$, F) *for the* MJ *rule and the* RS *method, respectively, then* $\lim_{s\to\infty}(n_s^{\mathrm{RS}}/n_s^{\mathrm{MJ}})=((m+1)m^2[\int f^2F^{m-2}]^2)/12([\int f^2]^2)$.

Derivations of $\int f^2F^{m-2}$ are provided in Appendix C for selected translative strengths models. Table 1 gives Pitman relative efficiencies. Except for the Uniform distribution, all other models considered herein have PRE less than one.

4.2. Bahadur asymptotic relative efficiency

In this section, we deal with a fixed vector of desirabilities $\boldsymbol{\gamma}=(\gamma_1,\gamma_2,\ldots,\gamma_m)'$. We assume that, associated with the n i.i.d. votes, there are i.i.d. unobserved random vectors $Y_j=(Y_{1,j},Y_{2,j},\ldots,Y_{m,j})'$, $1\leqslant j\leqslant n$, with

Table 1
Pitman relative efficiency for selected distributions

Distribution	Pre(MJ, RS)
Uniform	$\frac{(m+1)m^2}{12(m-1)^2}$
Logistic	$\frac{3}{m+1}$
Exponential	$\frac{4(m+1)}{3(m-1)^2}\left[1-\frac{1}{2^{m-1}}\right]^2$
Normal ($m=3$)	$\frac{3}{4}$
Normal ($m=4$)	$\frac{20}{3}\left[\frac{\arctan(\sqrt{2})}{\pi}\right]^2 \approx 0.62$
Normal ($m=5$)	$\frac{25}{2}\left[\frac{6\arctan(\sqrt{2})-\pi}{4\pi}\right]^2 \approx 0.53$

independent components with distributions $F(y-\gamma_1), F(y-\gamma_2), \ldots, F(y-\gamma_m)$. Define the probabilities u_i and v_i as follows:

$$u_i = P(Y_{i,j} > Y_{k,j}\ \forall k \neq i) = \int f(y-\gamma_i)\prod_{k\neq 1} F(y-\gamma_k) = \int f(y)\prod_{k\neq i} F(y+\gamma_i-\gamma_k),$$

$$v_i = P(Y_{i,j} > Y_{1,j}) = \int f(y)F(y+\gamma_i-\gamma_1).$$

Now, consistent with the notation established in Section 3.2, let $W_i^{(j)}$ be the difference in first-place votes cast by voter j between alternative 1 and alternative i, and $D_i^{(j)}$ be the difference in rank-sums assigned by voter j to alternative i and alternative 1. Then, $M_W(t) = E[\mathrm{e}^{-tW_i^{(j)}}] = \mathrm{e}^{-t}u_1 + \mathrm{e}^t u_i + (1-u_1-u_i)$, and

$$\begin{aligned} M_D(t) = E[\mathrm{e}^{-tD_i^{(j)}}] &\geqslant P(Y_{i,j} > Y_{1,j})\mathrm{e}^t + P(Y_{i,j} < Y_{1,j})\mathrm{e}^{-(m-1)t} = v_i\mathrm{e}^t + (1-v_i)\mathrm{e}^{-(m-1)t} \\ &\geqslant m(m-1)^{1/m-1}v_i^{1-1/m}(1-v_i)^{1/m}. \end{aligned}$$

Therefore, by Chernoff's Theorem, $P(\sum_{j=1}^n W_i^{(j)} \leqslant 0) = \mathrm{O}\{[1-(\sqrt{u_1}-\sqrt{u_i})^2]^n\}$, which implies that the MJ rule selects the correct alternative with probability $1-\mathrm{O}\{[1-(\sqrt{u_1}-\sqrt{u_2})^2]^n\}$, as $n \to \infty$. For the RS method, application of Chernoff's Theorem shows $P(-\sum_{j=1}^n D_i^{(j)} \geqslant 0)$ is at least of order $[m(m-1)^{1/m-1}v_i^{1-1/m}(1-v_i)^{1/m}]^n$, and is of the highest order when $i=2$. Thus the probability of selecting the best alternative using the RS method converges to 1 slower than $1-\mathrm{O}\{[m(m-1)^{1/m-1}v_2^{1-1/m}(1-v_2)^{1/m}]^n\}$. Hence, we have the following theorem.

Theorem 6. *Under the translative strengths model* TS(γ_s, F), *the Bahadur efficiency of the* MJ *rule relative to* RS *method is bounded from below by*

$$\log\left[1-(\sqrt{u_1}-\sqrt{u_2})^2\right] / \log[m(m-1)^{1/m-1}v_2^{1-1/m}(1-v_2)^{1/m}]. \tag{6}$$

5. Simulation results

In this section, we report results from two simulation studies. Our first study compares the probabilities of choosing the correct alternative under the Plackett–Luce model using both the MJ rule and RS method. Also included are results based on the maximum likelihood estimates (MLE), which selects the alternative with the largest corresponding point estimate of desirability. More specifically, the MLE of the desirability parameters ($\hat{\gamma}$) maximizes $\prod_{j=1}^n P_\gamma(\pi_j)$ as defined

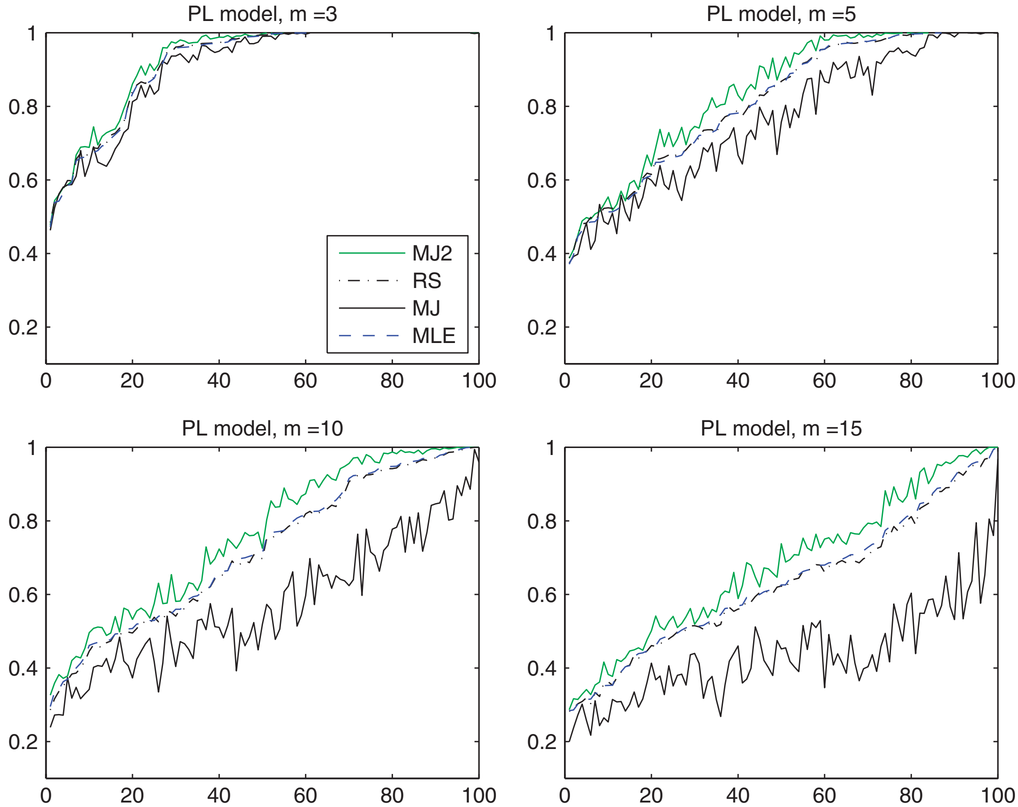

Fig. 1. Power comparisons for MJ, RS and MLE under Plackett–Luce model ($n = 200$). The horizontal axis index gives 100 randomly simulated desirability configurations with $\gamma_1 = 1$ and $\gamma_i \sim \text{Uniform}(0,1)$ for $i = 2, 3, \ldots, m$, sorted by the power of the MLE method. When the size of voting committee is enlarged by 1/PRE, the efficiency-adjusted majority rule (MJ2) has power slightly larger than RS and MLE methods.

in Eq. (1), and can be computed using the MM algorithm of Hunter (2004), which is guaranteed to converge to the unique maximum likelihood estimator under mild conditions. Fig. 1 presents the power results from the Plackett–Luce model with a voting committee of size $n = 200$. The four panels correspond to $m = 3$, 5, 10 and 15 competing alternatives, respectively. Because closed-form power calculations are intractable, the power associated with each decision rule is approximated based on 2000 simulations for each random draw of $\boldsymbol{\gamma}$. In each plot, the horizontal axis index gives 100 randomly simulated desirability configurations sorted such that they are increasing in the power of the MLE approach. The vertical axis gives the probability of a correct selection (i.e., choosing the most deserving candidate). For comparison, the plots also include the probability of a correct selection using MJ2, the efficiency-adjusted MJ rule evaluated with $n^* = n/\text{PRE}$ voters. The figure shows that for all values of m considered, the RS method and MLE-based method result in approximately equal probability of a correct decision. The MJ rule performs worse than either the RS or MLE method. The MJ2 rule (using n^* votes) always has probability of correct selection slightly larger than those corresponding to the RS and MLE methods. In addition, our simulation studies indicate that the difference between MJ2 and RS methods become smaller as the sample size n increases, which is expected as n^* is based on asymptotic considerations.

The second study compares the MJ, RS, and MLE methods under the translative strength models. Both Fig. 2 ($m = 3$) and Fig. 3 ($m = 5$) compare the probability of a correct selection for four common distributions: uniform, logistic, exponential and normal. Each panel displays the results of 100 randomly simulated desirability configurations (again sorted so that the power of the MLE-based approach is increasing), with $n = 20$ and $n^* = n/\text{PRE}$. In this study, the MLE is intentionally misspecified and evaluated under the Plackett–Luce model. In many cases, the MLE approach is worse than even the MJ rule when the translative strengths model is misspecified.

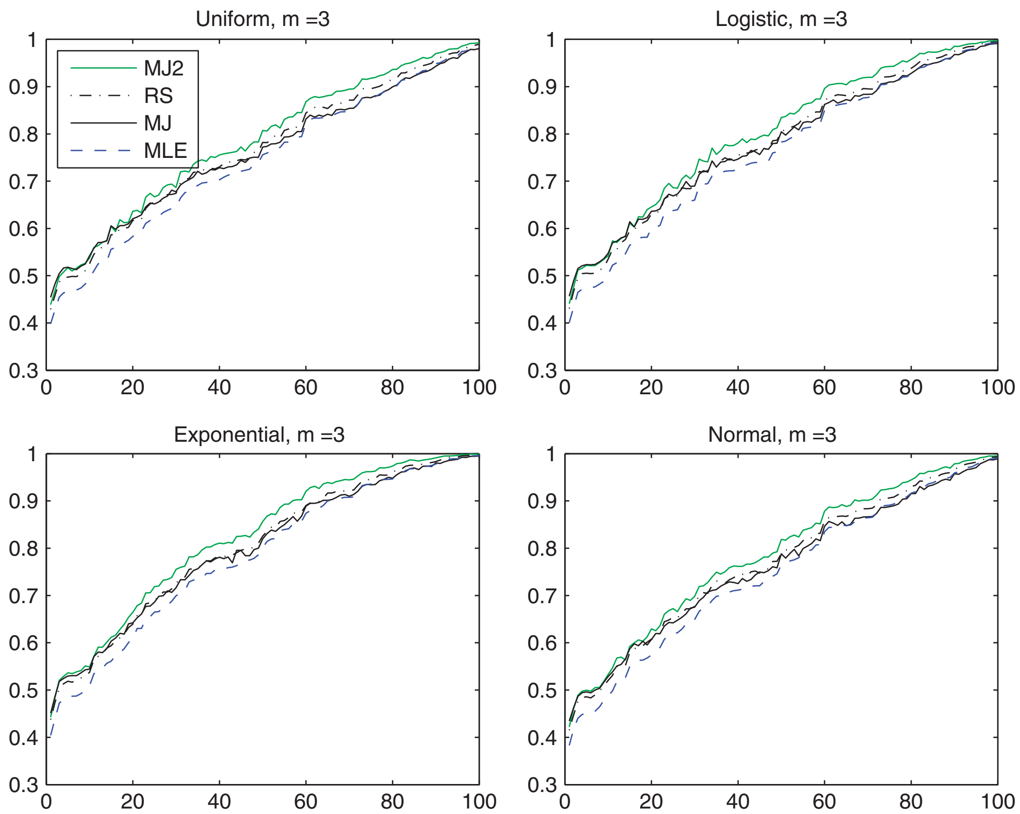

Fig. 2. Power comparisons among MJ, RS and MLE under translative strengths model; the size of voting committee $n = 20$ and number of alternative $m = 3$. The horizontal axis index gives 100 randomly simulated strength (location parameter) configurations (sorted): $\gamma_1 = \sigma$ and $\gamma_i = u_i\sigma$, where σ^2 is the variance of the f distribution and $u_i \sim \text{Uniform}(0, 1)$ for $i = 2, 3, \ldots, m$. MJ2 is the efficiency-adjusted majority rule using $n^* = n/\text{PRE}$ votes.

6. Conclusions

In many applications, a panel selects and ranks many alternatives. We investigate a variant of the problem where, in lieu of continuous responses, data consist of discrete orderings of the alternatives under consideration. As such, the discrete case can be thought of as a censored-data version of the original decision problem. For two classes of data-generating models (Plackett–Luce and translative strengths), we derived both the Pitman efficiency and lower bounds on the Bahadur efficiency of the MJ rule relative to the RS method.

The case where there are only two alternatives is trivial since the two decision rules are equivalent. However, as the number of alternatives, m, increases the rank-sum rule makes more use of the data at hand and is more efficient than the majority rule. Interestingly, when there are three competing alternatives, the Pitman relative efficiency of the uniform, logistic, exponential and normal translative strengths models are identical ($\text{PRE} = \frac{3}{4}$). These results follow immediately from Table 1. By contrast, when $m = 3$ and the rank-orderings follow a Plackett–Luce model, $\text{PRE} = \frac{16}{27}$. These results suggest that the power of the rank-sum method can be reached by the majority rule if the sample size is increased to $n^* = n/\text{PRE}$. Therefore, when it is substantially more difficult to obtain a complete rank-ordering than an additional partial-ordering, the majority rule performs adequately and efforts would be better spent asking many voters to provide top choice rather than fewer voters to provide complete orderings.

For comparison purposes, the decision rule based on maximum likelihood estimates was also given. In all cases, the presumed underlying model was the Plackett–Luce because in reality a true model is never known. Furthermore, this encompasses both a best-case (true model specified correctly) and worst-case (wrong model family chosen) scenario which allows for investigation into the MLE's robustness. Simulation studies suggest that the RS method compares

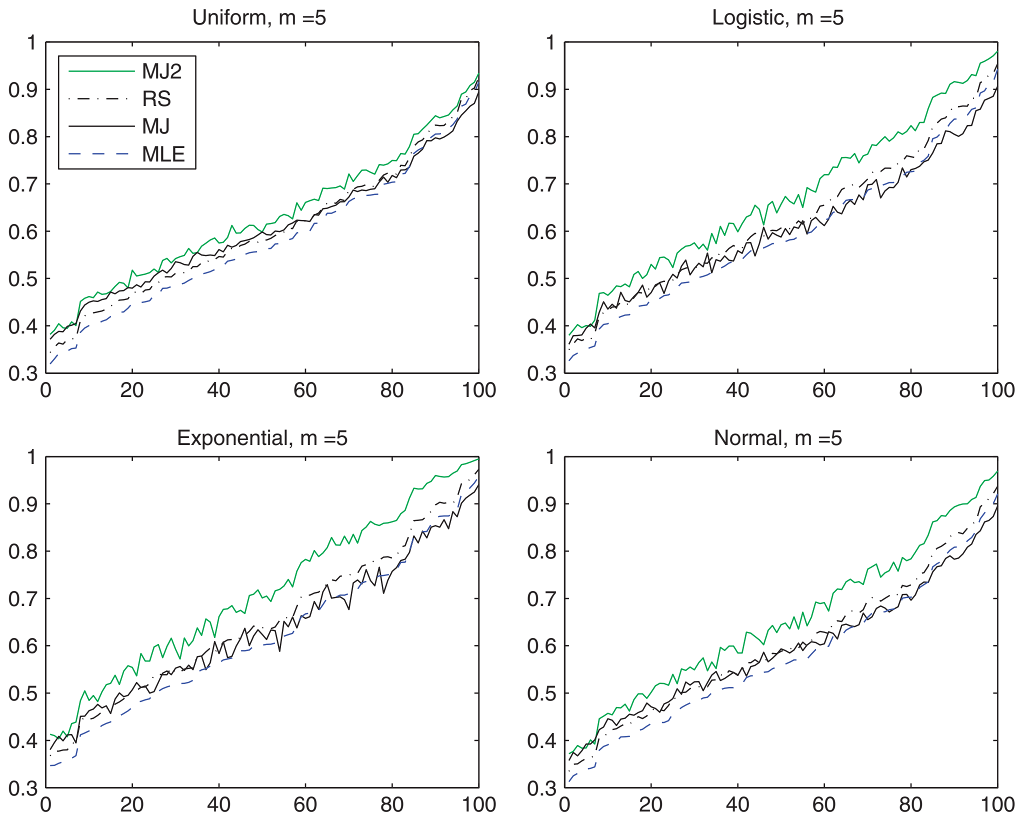

Fig. 3. Power comparisons among MJ, RS and MLE under translative strengths model ($n = 20$ and $m = 5$). The horizontal axis index is the same as in Fig. 2. Once again, MJ2 is the efficiency-adjusted majority rule using $n^* = n/\text{PRE}$ votes.

favorably to the maximum likelihood estimate in both cases. The MLE method is not robust in the sense that it becomes worse than even the majority rule when the translative strengths model is misspecified as the Plackett–Luce model. This lack of robustness is not surprising since the MLE is parametric while the MJ rule and RS procedure are non-parametric.

In the preceding arguments, no voter manipulation was allowed. That is, all members of the population submitted genuine rank-orderings reflecting their preferences of the alternatives. In a real situation, however, a manipulative voter may wish to increase the chances of his preferred alternative being chosen by submitting a vote on which other desirable alternatives are ranked much lower than their merits would suggest. In such a case, the preceding robustness and efficiency arguments do not apply, and one may wish to simply use the majority rule to prevent manipulation. A potential area of future work is to determine the relative amount of manipulative votes the RS procedure can withstand before it loses its efficiency over the MJ rule.

Acknowledgments

The authors are very grateful to the coordinating editor Dr. Klaus Miescke and two anonymous referees for their careful reading of the earlier versions of this paper and helpful comments.

Appendix A. Properties of the Plackett–Luce model

In the case of two competing alternatives, denoted i and j, Bradley and Terry (1952) posit that the probability i is preferred to j is given by $p_{i,j} = \gamma_i/(\gamma_i + \gamma_j)$. Aside from its simplicity, the Bradley–Terry formulation enjoys wide-spread use owing, in part, to the variety of situations from which it may be derived. While not stated as such, the Bradley–Terry formulation assigns probabilities to (degenerate) permutations. In situations where $m > 2$ items are to be ranked, the Plackett–Luce model is the natural extension of the Bradley–Terry model, assigning probabilities to all $m!$ orderings of alternatives.

An appealing property of the Plackett–Luce model is its internal consistency. (This property is sometimes referred to as indifference to irrelevant alternatives.) More formally, consider a proper subset $B \subset \{1, 2, \ldots, m\}$ whose cardinality is $m_b < m$ and whose elements are $\{b_1, b_2, \ldots, b_{m_b}\}$. If $P_B(i)$ is the probability that alternative i is ranked highest among those contained in B, then under the Plackett–Luce model,

$$P[b_1 \to b_2 \to \cdots \to b_{m_b}] = \prod_{i=1}^{m_b} \frac{\gamma_{b_i}}{\gamma_{b_i} + \cdots + \gamma_{m_b}}, \quad \text{and consequently } P_B(i) = \frac{\gamma_i}{\sum_{j \in B} \gamma_j}. \tag{7}$$

In other words, this probability only depends on the relative merits of the alternatives in B and is unaffected by the remaining alternatives not in B. Another attractive property of the model is the variety of ways in which it may be derived, some of which follow.

Luce's choice axiom: Consider a situation in which every alternative has a positive probability of being preferred to any other alternative in a pairwise comparison (i.e., $\gamma_i > 0;\ \forall i$). If $A = \{a_1, a_2, \ldots, a_{m_a}\}$ is a subset of B consisting of m_a elements and $P_B(A) = \sum_{i \in A} P_B(i)$ is the probability that a member of A is ranked highest among those in B, then Luce's choice axiom (Luce, 1959) states

$$P_B(i) = P_B(A) P_A(i) \quad \forall i \in A \subseteq B. \tag{8}$$

This condition can be interpreted using a two-stage argument as follows. First, if $i \in A \subseteq B$ is to be preferred to all other alternatives in B, then it is a weaker condition to require that at least *some* member of A is the most preferred alternative in B. Subsequently, if this alternative is to be i, then i must also be most preferred among those in A. Luce showed that the above axiom is equivalent to Eq. (1) for positive-valued parameters $\boldsymbol{\gamma}$.

Plackett's hierarchical approach: Plackett (1975) presents a general framework for modeling permutations motivated by considering the results of a horse race. Such results may be viewed in a sequential fashion by determining which horse finishes the race first, which of the remaining horses finishes second, and so forth. This paradigm led him to propose a hierarchical structure for which the probabilities of each remaining horse finishing next (given those which have not already finished) are modeled. Plackett's first-stage logistic model, which includes no interactions, results in permutation probabilities that coincide with Luce's.

Although Luce (1959) and Plackett (1975) arrive at the model given by Eq. (1) from different axioms, theirs are not the only motivations. Stern (1990) develops a model based on independent Gamma waiting times. In his formulation the winner is the first competitor to accumulate a fixed number, r, of points. The second-place finisher is the next to accumulate r, etc. When $r = 1$ and the waiting times are exponentially distributed, the permutation probabilities follow Eq. (1). Finally, we present a connection between the Plackett–Luce model and the proportional hazards model (Cox, 1972).

Relationship to proportional hazards model: In survival analysis, the hazards is defined as $h(y) = f(y)/[1 - F(y)]$, where $f(y)$ is the density of survival times and $S(y) = [1 - F(y)]$ is the survival function (i.e., the probability that the survival time exceeds y). A common, simplifying assumption for parametric survival modeling is to assume that hazards are proportional and that the covariates determine their ratio, i.e., $h_i(y) = h_0(y)\gamma_i$ for some *baseline hazard* $h_0(\cdot)$. In such a situation, the following theorem shows that relative order of survival times follow a Plackett–Luce model.

Theorem 7. *Suppose $Y_1, Y_2, \ldots, Y_m$ are independent survival times and follow a proportional hazards model with hazard functions $h_i(y) = h_0(y)\gamma_i$, $i = 1, \ldots, m$, then the rank-orderings of the survival times follows a Plackett–Luce formulation, $P[Y_1 < Y_2 < \cdots < Y_m] = \prod_{i=1}^{m} (\gamma_i/(\gamma_i + \gamma_{i+1} + \cdots + \gamma_m))$.*

The proof is straightforward and omitted.

Appendix B. Proofs of lemmas

Proof of Lemma 1. Let $\boldsymbol{W}_s^{(j)} = (W_{s,2}^{(j)}, W_{s,3}^{(j)}, \ldots, W_{s,m}^{(j)})'$ represent the differences in first-place votes awarded by voter j to alternative 1 and alternatives 2 through m, respectively. Then $W_{s,i}^{(j)}$ is a difference of indicator functions, $W_{s,i}^{(j)} = I\{\pi_j(1) = 1\} - I\{\pi_j(1) = i\}$ and has expectation $E(W_{s,i}^{(j)}) = \gamma_{s,1} - \gamma_{s,i}$ (since $\sum_{i=1}^m \gamma_{s,i} = 1$). Furthermore,

$$
\begin{aligned}
E(W_{s,i}^{(j)} W_{s,i'}^{(j)}) &= E\left[(I\{\pi_j(1) = 1\} - I\{\pi_j(1) = i\})(I\{\pi_j(1) = 1\} - I\{\pi_j(1) = i'\})\})\right] \\
&= \begin{cases} \gamma_{s,1} + \gamma_{s,i} & \text{if } i = i', \\ \gamma_{s,1} & \text{if } i \neq i'. \end{cases}
\end{aligned}
$$

Hence, we have $E(\boldsymbol{W}_s^{(j)}) = \boldsymbol{\delta}_s$ and $\lim_{s\to\infty} \operatorname{Cov}(\boldsymbol{W}_s^{(j)}) = (1/m)\Sigma$. Finally, since $\boldsymbol{W}_s = \sum_{j=1}^{n_s} \boldsymbol{W}_s^{(j)}$ is the sum of n_s independent and identically distributed random variables with finite moments, the central limit theorem for a double array guarantees that $n_s^{-1/2}(\boldsymbol{W}_s - n_s\boldsymbol{\delta}_s) \to N(\boldsymbol{0}, (1/m)\Sigma)$.

Turning to rank-sums, define $\boldsymbol{D}_s^{(j)} = (D_{s,2}^{(j)}, D_{s,3}^{(j)}, \ldots, D_{s,m}^{(j)})'$ as the rank differences between alternatives 2 through m and alternative 1 by voter j. Note that the rank difference may be expressed as

$$
D_{s,i}^{(j)} = \sum_{k=1}^{m} [I(Y_{k,j} < Y_{i,j}) - I(Y_{k,j} < Y_{1,j})] = \sum_{k=1}^{m} [I(Y_{k,j} > Y_{1,j}) - I(Y_{k,j} > Y_{i,j})],
$$

which has expectation

$$
E(D_{s,i}^{(j)}) = \sum_{k=1}^{m} \left[\frac{\gamma_{s,1}}{\gamma_{s,1} + \gamma_{s,k}} - \frac{\gamma_{s,i}}{\gamma_{s,i} + \gamma_{s,k}}\right] = (\gamma_{s,1} - \gamma_{s,i}) \sum_{k=1}^{m} \frac{\gamma_{s,k}}{(\gamma_{s,1} + \gamma_{s,k})(\gamma_{s,i} + \gamma_{s,k})}.
$$

In addition,

$$
E(D_{s,i}^{(j)} D_{s,i'}^{(j)}) = \sum_{k=1}^{m} \sum_{l=1}^{m} E([I(Y_{k,j} > Y_{1,j}) - I(Y_{k,j} > Y_{i,j})][I(Y_{l,j} > Y_{1,j}) - I(Y_{l,j} > Y_{i',j})]).
$$

Therefore, $E(\boldsymbol{D}_s^{(j)}) = \boldsymbol{\eta}_s$ and $\lim_{s\to\infty} \operatorname{Cov}(\boldsymbol{D}_s^{(j)}) = ((m(m+1))/12)\Sigma$. Once again, since $\boldsymbol{D}_s = \sum_{j=1}^{n_s} \boldsymbol{D}_s^{(j)}$ is the sum of independent and identically distributed random variables with finite moments, the central limit theorem for a double array guarantees that $n_s^{-1/2}(\boldsymbol{D}_s - n_s\boldsymbol{\eta}_s) \to N(\boldsymbol{0}, ((m(m+1))/12)\Sigma)$. □

Proof of Lemma 2. We define $\boldsymbol{W}_s^{(j)}$ and $\boldsymbol{D}_s^{(j)}$ same as in the Proof of Lemma 1. Then, under the translative strengths model $\text{TS}(\boldsymbol{\gamma}_s, F)$, $\boldsymbol{W}_s = \sum_{j=1}^{n_s} \boldsymbol{W}_s^{(j)}$ and $\boldsymbol{D}_s = \sum_{j=1}^{n_s} \boldsymbol{D}_s^{(j)}$ remain as sum of independent and identically distributed random variables with finite moments.

Next, if we let $f(y) = F'(y)$, then the probability of alternative k being preferred to alternative i by a voter under model $\text{TS}(\boldsymbol{\gamma}_s, F)$ satisfies

$$
\begin{aligned}
\lim_{s\to\infty} b_s \left[P_{\boldsymbol{\gamma}_s}(Y_{k,j} > Y_{i,j}) - \frac{1}{2}\right] &= \lim_{s\to\infty} b_s \left[\int F(y - \gamma_{s,i}) f(y - \gamma_{s,k}) - \frac{1}{2}\right] \\
&= \lim_{s\to\infty} b_s \left[\int F(y + \gamma_{s,k} - \gamma_{s,i}) f(y) - \int F(y) f(y)\right] \\
&= \lim_{s\to\infty} b_s (\gamma_{s,k} - \gamma_{s,i}) \int f(y)^2 = (\theta_i - \theta_k) \int f(y)^2.
\end{aligned}
$$

Consequently, the rank difference between alternative i and alternative 1 assigned by voter j under model TS(γ_s, F) satisfies

$$\lim_{s\to\infty} b_s E_{\gamma_s}[D_{s,i}^{(j)}] = \lim_{s\to\infty} b_s E_{\gamma_s}\left\{\sum_{k\neq i}\left[I(Y_{k,j} > Y_{i,j}) - \frac{1}{2}\right] - \sum_{k\neq 1}\left[I(Y_{k,j} > Y_{1,j}) - \frac{1}{2}\right]\right\}$$
$$= \lim_{s\to\infty} b_s\left[\sum_{k\neq i}(\gamma_{s,k} - \gamma_{s,i}) - \sum_{k\neq 1}(\gamma_{s,k} - \gamma_{s,1})\right]\int f(y)^2 = m\theta_i \int f^2. \tag{9}$$

On the other hand, for the majority rule, the probability that 1 is the most preferred alternative by voter j satisfies

$$\lim_{s\to\infty} b_s\left[P_{\gamma_s}(Y_{1,j} > Y_{i,j}\ \forall i \neq 1) - \frac{1}{m}\right] = \lim_{s\to\infty} b_s\left[\int f(y-\gamma_{s,1})\prod_{i\neq 1} F(y-\gamma_{s,i}) - \frac{1}{m}\right]$$
$$= \lim_{s\to\infty} b_s\left[\int f(y)\prod_{i\neq 1} F(y+\gamma_{s,1}-\gamma_{s,i}) - \int f(y)F(y)^{m-1}\right]$$
$$= \sum_{i\neq 1}\theta_i \int f(y)^2 F(y)^{m-2}.$$

Hence, the difference in first place votes awarded to alternative 1 and alternative i by voter j satisfies

$$\lim_{s\to\infty} b_s E_{\gamma_s}[W_{s,i}^{(j)}] = \lim_{s\to\infty} b_s E\gamma_s\left\{\left[I(Y_{1,j} > Y_{k,j}\ \forall k \neq 1) - \frac{1}{m}\right] - \left[I(Y_{i,j} > Y_{k,j}\ \forall k \neq i) - \frac{1}{m}\right]\right\}$$
$$= \lim_{s\to\infty} b_s\left[\sum_{k\neq 1}(\gamma_{s,1} - \gamma_{s,k}) - \sum_{k\neq i}(\gamma_{s,i} - \gamma_{s,k})\right]\int f(x)^2 F(x)^{m-2}$$
$$= m\theta_i \int f^2 F^{m-2}. \tag{10}$$

Thus, we have $\lim_{s\to\infty}[E(\boldsymbol{W}_s) - n_s\boldsymbol{\delta}_s m\int f^2F^{m-2}] = 0$ and $\lim_{s\to\infty}[E(\boldsymbol{D}_s) - n_s\boldsymbol{\delta}_s m\int f^2] = 0$. Furthermore, as $s\to\infty$, $\mathrm{Cov}(\boldsymbol{W}_s^{(j)})$ and $\mathrm{Cov}(\boldsymbol{D}_s^{(j)})$ converge to the same null covariance matrices as in the Plackett–Luce model. Therefore, Lemma 2 follows from the central limit theorem for a double array. □

Appendix C. $\int fF^{m-2}$ for common translative strengths models

First, for the logistic distribution, $f(y) = \mathrm{e}^{-y}/(1+\mathrm{e}^{-y})^2$ and $F(y) = 1/(1+\mathrm{e}^{-y})$. Therefore,

$$\int f(y)^2 F(y)^{m-2} = \int \mathrm{e}^{-2y}(1+\mathrm{e}^{-y})^{-m-2}\,\mathrm{d}y = \frac{1}{(m+1)m}.$$

Next, for the exponential distribution, we have $f(y) = \mathrm{e}^{-|y|}/2$ and $F(y) = \mathrm{e}^{y}/2$ when $y < 0$, and $F(y) = 1 - \mathrm{e}^{-y}/2$ when $y \geqslant 0$. Thus,

$$\int f(y)^2 F(y)^{m-2} = \int_{-\infty}^{0}\frac{1}{2^m}\mathrm{e}^{my}\,\mathrm{d}y + \int_0^\infty \frac{1}{4}\mathrm{e}^{-2y}[1-\mathrm{e}^{-y}/2]^{m-2}\,\mathrm{d}y = \frac{1}{m(m-1)}\left[1 - \frac{1}{2^{m-1}}\right].$$

Lastly, for the normal distribution, following Bose and Gupta (1959), we define $I_n(a) = \int [\Phi(ay)]^n e^{-y^2}$. Then, their results imply that $\int \phi^2 = I_0(1)/(2\pi) = 1/(2\sqrt{\pi})$, and

$$\int \phi^2 \Phi^{m-2} = I_{m-2}(1)/(2\pi) = \begin{cases} \dfrac{1}{4\sqrt{\pi}}, & m = 3, \\ \dfrac{\arctan(\sqrt{2})}{2\pi^{3/2}}, & m = 4, \\ \dfrac{6\arctan(\sqrt{2}) - \pi}{8\pi^{3/2}}, & m = 5. \end{cases}$$

References

Bahadur, R.R., 1971. Some Limit Theorems in Statistics. Society for Industrial and Applied Mathematics, Philadelphia.

Bechhofer, R.E., 1954. A single-sample multiple decision procedure for ranking means of normal populations with known variances. Ann. Math. Statist. 25, 16–39.

Bose, R.C., Gupta, S.S., 1959. Moments of order statistics from a normal population. Biometrika 46, 433–440.

Bradley, R.A., Terry, M.E., 1952. Rank analysis of incomplete block designs: I. The method of paired comparisons. Biometrika 39 (3/4), 324–345.

Chernoff, H., 1952. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. Ann. Math. Statist. 23 (4), 493–507.

Cox, D.R., 1972. Regression models and life tables (with discussion). J. Roy. Statist. Soc. B 30, 187–220.

Daniel, H.E., 1950. Rank correlation and population methods. Biometrika 33, 129–135.

Gupta, S.S., 1956. On a decision rule for a problem in ranking means. Ph.D. Dissertation, University of North Carolina, Chapel Hill.

Hunter, D.R., 2004. MM algorithms for generalized Bradley–Terry models. Ann. Statist. 32, 384–406.

Luce, R.D., 1959. Individual Choice Behavior. Wiley, New York.

Marden, J.I., 1995. Analyzing and Modeling Rank Data. Chapman & Hall, New York.

Pitman, E.J.G., 1948. Notes on Non-parametric Statistical Inference. Columbia University, New York.

Plackett, R.L., 1975. The analysis of permutations. Appl. Statist. 24, 193–202.

Stern, H., 1990. Models for distributions on permutations. J. Amer. Statist. Assoc. 85 (410), 558–564.