Nonparametric Maximum Likelihood Estimation of Within-Set Ranking Errors in Ranked Set Sampling

Omer Ozturk

The Ohio State University

June 24, 2011
Outline

1 Motivation
2 Ranking Error
3 Ranking Error Models
4 Likelihood Function
5 Missing Data Approach
6 Estimation of Judgment Ranking Probabilities
7 Simulation Results
8 Application
9 Summary
Outline

1 Motivation
2 Ranking Error
3 Ranking Error Models
4 Likelihood Function
5 Missing Data Approach
6 Estimation of Judgment Ranking Probabilities
7 Simulation Results
8 Application
9 Summary
Outline

1 Motivation
2 Ranking Error
3 Ranking Error Models
4 Likelihood Function
5 Missing Data Approach
6 Estimation of Judgment Ranking Probabilities
7 Simulation Results
8 Application
9 Summary
Outline

1. Motivation
2. Ranking Error
3. Ranking Error Models
4. Likelihood Function
5. Missing Data Approach
6. Estimation of Judgment Ranking Probabilities
7. Simulation Results
8. Application
9. Summary
Outline

1. Motivation
2. Ranking Error
3. Ranking Error Models
4. Likelihood Function
5. Missing Data Approach
6. Estimation of Judgment Ranking Probabilities
7. Simulation Results
8. Application
9. Summary
Outline

1. Motivation
2. Ranking Error
3. Ranking Error Models
4. Likelihood Function
5. Missing Data Approach
6. Estimation of Judgment Ranking Probabilities
7. Simulation Results
8. Application
9. Summary
Outline

1 Motivation
2 Ranking Error
3 Ranking Error Models
4 Likelihood Function
5 Missing Data Approach
6 Estimation of Judgment Ranking Probabilities
7 Simulation Results
8 Application
9 Summary
Motivation

- Main objective is to reduce the cost in a data collection process.
- Instead of making expensive or time consuming gold standard measurements, we make some quick and cheap potential observations on a set of experimental units.
- These potential observations provide subjective forecast on the ranks of small set of experimental units.
- However imperfect these ranks may be, if they are used properly, they often lead to an efficient statistical inference.
Ranked Set Sampling

- Select $m$ units at random from a specified population.
- Rank these $m$ units with some expert judgment without a gold standard measurement.
- Retain the smallest judged unit for gold standard measurement and return the others.
- Select the second $m$ units and retain the second smallest unit judged for a measurement.
- Continue to the process until $m$ ordered units are measured.
- Note: These $m$ ordered observations $X_{[1]i}, \ldots, X_{[m]i}$ are called a cycle.
- Note: Process repeated $i = 1, \cdots, n$ cycle to get $nm$ observations. These $nm$ observations are called a standard ranked set sample.
Let $m=3$ and $n=2$

<table>
<thead>
<tr>
<th>Cycle</th>
<th>Judgment Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$X_1$, $X_2$, $X_3$</td>
</tr>
<tr>
<td>2</td>
<td>$X_1$, $X_2$, $X_3$</td>
</tr>
</tbody>
</table>

$X_1, \ldots, X_3$ is called a ranked set sample.

- In each set, colored unit is selected for gold standard measurement.
- $X_{ij}, i = 1, \ldots, m, j = 1, \ldots, n$ are all independent, but not identically distributed.
- For each fixed $i$, $X_{ij}, j = 1, \ldots, n$ are iid with judgment class cdf $F_i$.
- If there is no ranking error, the judgment order statistic $X_{ij}$ becomes usual order statistics $X_{(i)j}$. 
Why ranked-set sampling?

- Let $X_i, i = 1, \cdots, m$ be a SRS, and let $\bar{X}_{RSS}$ and $\bar{X}_{SRS}$ denote the sample averages based on RSS and SRS.
- It is easy to observe that

$$\text{var}(\bar{X}_{SRS}) = \frac{1}{m^2} \text{var}(\sum_{i=1}^{m} X_i) = \frac{1}{m^2} \text{var}(\sum_{i=1}^{m} X_{(i)})$$

$$= \frac{1}{m^2} \left\{ \sum_{i=1}^{m} \sigma_{(i)}^2 + \sum_{i \neq j} \sigma_{ij} \right\} = \text{var}(\bar{X}_{RSS}) + \text{cov} \text{var}(\bar{X}_{SRS}) \geq \text{var}(\bar{X}_{RSS})$$

- Inequality becomes an equality when the ranking is completely random.

- This improved efficiency result holds for almost all statistical procedures based on RSS.
Impact of Ranking Error

- We are almost certain that there will be ranking error in practice.
- Even though the efficiency gain still holds under imperfect ranking, statistical procedure may not be valid.
- In MWW test, even with a minor ranking error, Type I error rate is inflated.

<table>
<thead>
<tr>
<th>Corr</th>
<th>n</th>
<th>m</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>2</td>
<td>0.060</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>2</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td>0.059</td>
</tr>
<tr>
<td>0.5</td>
<td>5</td>
<td>2</td>
<td>0.112</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td>0.147</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>2</td>
<td>0.097</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td>0.144</td>
</tr>
</tbody>
</table>
Impact of Ranking Error

- We are almost certain that there will be ranking error in practice.
- Even though the efficiency gain still holds under imperfect ranking, statistical procedure may not be valid.
- In MWW test, even with a minor ranking error, Type I error rate is inflated.

<table>
<thead>
<tr>
<th>Corr</th>
<th>n</th>
<th>m</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>2</td>
<td>0.060</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.053</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>2</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.059</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>5</td>
<td>2</td>
<td>0.112</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.147</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>2</td>
<td>0.097</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.144</td>
<td></td>
</tr>
</tbody>
</table>
Model

- Bohn and Wolfe (1994) Model: Judgment class distribution is modeled as a mixture distribution of order statistics.

\[
f_{[i]}(y) = \sum_{j=1}^{m} p_{i,j} f(j)(y), \quad f(j)(y) = m \binom{m-1}{i-1} F^{i-1}(y) \{1 - F(y)\}^{m-i} dF(y),
\]

where \( p_{i,j} \) is the probability that the \( j \)-th order statistic is assigned rank \( i \).

- \( P = (p_{i,j}) \) is a doubly stochastic matrix.

- One parameter model, Frey (2007): Judgment ranking probabilities, \( p_{i,j} \), expressed as a function of a single parameter, \( \eta \),

\[
f_{[i]}(y) = \sum_{j=1}^{m} p_{i,s(\eta)} f(j)(y),
\]

In these models, we are interested in the estimation of \( P \) (or \( P(\eta) \)) and the underlying distribution function \( F \).
Model

- Bohn and Wolfe (1994) Model: Judgment class distribution is modeled as a mixture distribution of order statistics.

\[ f_{[i]}(y) = \sum_{j=1}^{m} p_{i,j} f_{(j)}(y), \quad f_{(i)}(y) = m \binom{m-1}{i-1} F^{-1}(y) \left\{ 1 - F(y) \right\}^{m-i} dF(y), \]

where \( p_{i,j} \) is the probability that the \( j \)-th order statistic is assigned rank \( i \).

- \( P = (p_{i,j}) \) is a doubly stochastic matrix.

- One parameter model, Frey (2007): Judgment ranking probabilities, \( p_{i,j} \), expressed as a function of a single parameter, \( \eta \),

\[ f_{[i]}(y) = \sum_{j=1}^{m} p_{i,s(\eta)} f_{(j)}(y), \]

In these models, we are interested in the estimation of \( P \) (or \( P(\eta) \)) and the underlying distribution function \( F \).
Model

- Bohn and Wolfe (1994) Model: Judgment class distribution is modeled as a mixture distribution of order statistics.

\[ f_{[i]}(y) = \sum_{j=1}^{m} p_{i,j} f_{(j)}(y), \quad f_{(i)}(y) = m \binom{m-1}{i-1} F^{i-1}(y) \{1 - F(y)\}^{m-i} dF(y), \]

where \( p_{i,j} \) is the probability that the \( j \)-th order statistic is assigned rank \( i \).

- \( P = (p_{i,j}) \) is a doubly stochastic matrix.

- One parameter model, Frey (2007): Judgment ranking probabilities, \( p_{i,j} \), expressed as a function of a single parameter, \( \eta \),

\[ f_{[i]}(y) = \sum_{j=1}^{m} p_{i,s(\eta)} f_{(j)}(y), \]

In these models, we are interested in the estimation of \( P \) (or \( P(\eta) \)) and the underlying distribution function \( F \).
Dell and Clutter Model (1972): Ranking is performed based on perceived values of experimental units.

1. We generate a set of $m$ observations, $\mathbf{Y} = (Y_1, \cdots, Y_m)$, from a distribution $F$ with mean $\theta$ and variance $\sigma^2$.

2. We generate another independent random vector, $\mathbf{w} = (w_1, \cdots, w_m)$ from a normal distribution with mean zero and variance $\tau^2$. We add $\mathbf{Y}$ and $\mathbf{w}$ to obtain $\mathbf{X} = \mathbf{Y} + \mathbf{w}$.

3. We sort the vector $\mathbf{X}$ and select the $Y_{[j]}$ as the $j$-th judgment order statistics that corresponds to the $j$-th position in the sorted vector $\mathbf{X}$.

4. Quality of judgment ranking is controlled by the correlation coefficient between $\mathbf{X}$ and $\mathbf{Y}$, $\rho = corr(\mathbf{X}, \mathbf{Y}) = \frac{\sigma}{\sqrt{\sigma^2 + \tau^2}}$. 


Likelihood Function

- Let $X_{[r_j]j}, 1 \leq r_j \leq m, j = 1, \cdots, N$, $N = \sum_{i=1}^{m} n_i$ be a ranked set sample from a continues distribution $F$.
- Let $X_{(1)} < \cdots < X_{(N)}$ be the ordered values of $X_{[r_j]j}, j = 1, \cdots, N$.
- Let $\phi_j = F(X_{(j)})$ and $dF(X_{(j)}) = \bar{\phi}_j = \phi_j - \phi_{j-1}$.
- Log likelihood function, based on BW model, can be written as

$$L(P, \phi) = C + \sum_{i=1}^{n} \log \left\{ \sum_{s=1}^{m} p_{r_i,s} \binom{m-1}{s-1} \phi_i^{s-1} \{1 - \phi_i\}^{m-s} \right\}.$$ 

- The parameter space:

$$\Phi = \{\phi : 0 < \phi_1 < \cdots < \phi_N = 1\} \text{ and } P = \{P : \text{Doubly stoch.}\}$$

- In this model we wish to estimate $P$ and $\phi$.
- The likelihood function $L(I, \phi)$ is considered by Kvam and Samaniego (1994), where $I$ is identity matrix.
A Simple Example: $X_{[1]} < X_{[2]}$

Likelihood surface

- Let $a = dF(X_{[1]})$, $b = dF(X_{[2]})$ and

$$P = \begin{pmatrix} c & 1 - c \\ 1 - c & c \end{pmatrix}.$$ 

- Likelihood is maximized at $c = 1$, $a = 1/3$, $b = 2/3$.

- Empirical CDF $a = 1/2$, $b = 1/2$.

- Kvam-Samaniego Est, $c = 1$, $a = 1/3$, $b = 2/3$. 
A Simple Example: $X_2 < X_1$

Likelihood function

- Let $a = dF(X_1)$, $b = dF(X_2)$ and
  \[
  P = \begin{pmatrix}
  c & 1 - c \\
  1 - c & c
  \end{pmatrix}.
  \]

- Likelihood is maximized at $c = 0$, $a = 1/3$, $b = 2/3$.
- Empirical CDF $a = 1/2$, $b = 1/2$.
- Kvam-Samaniego Est, $c = 1$, $a = 1/2$, $b = 1/4$. 
**Theorem**

For a given doubly stochastic matrix $P$, the NPMLE of $\phi$ exists for any $P$ and is unique for some $P$, $P \in \mathcal{P}$.

- For a fixed value $P$, NPMLE of $\phi$ is obtained as a solution of the following estimating equation

\[
\sum_{s=1}^{m} A_{1,s} \left\{ \frac{s-1}{\phi_1} - \frac{m-s}{1-\phi_1} \right\} + \frac{1}{\phi_1} - \frac{1}{\bar{\phi}_2} = 0
\]

\[
\sum_{s=1}^{m} A_{i,s} \left\{ \frac{s-1}{\phi_i} - \frac{m-s}{1-\phi_i} \right\} + \frac{1}{\phi_i} - \frac{1}{\bar{\phi}_{i+1}} = 0, \quad i = 2, \cdots, N - 1
\]

\[
\sum_{s=1}^{m} A_{N,s} \left\{ \frac{s-1}{\phi_N} - \frac{m-s}{1-\phi_N} \right\} + \frac{1}{\phi_N} = 0,
\]

(1)
Missing data model

- Let \( Y_{[j]} \) be the vector of \( m \) within-set order statistics
  \[
  Y_{[j]}^\top = (Y_{(1)j} < \cdots < Y_{(m)j}).
  \]

- Let \( Z_{[r_j]}^\top = (z_{1j}, \cdots, z_{mj}) \) be a multinomial random vector with parameter 1 and \( p_{r_j} \), where \( p_{r_j} = (p_{r_j,1}, \cdots, p_{r_j,m}) \) is the \( r_j \)-th row of \( P \).

- The complete data then can be expressed as
  \[
  (Y_{[j]}, Z_{[r_j]}), j = 1, \cdots, N.
  \]

- For each \( r_j \), based on BW model with parameter \( P \), we observe the \( r_j \)-th judgment order statistic, \( X_{[r_j]j} = Z_{[r_j]}^\top Y_{[j]} \).
EM-Algorithm

- For a fixed a known value of $P$, we use EM-algorithm to find the NPMLE of $\phi$.
- Let $F^{(0)}$ be an initial estimate of $F$ and

$$M_Y(t) = \sum_{j=1}^{N} \sum_{i=1}^{m} I(Y(i)_{j} \leq t).$$

- E-step: We find the conditional expectation of $M_Y(t)$ given $X$ and $F^{(k)}$

$$M^{(k+1)}_X(t) = E_{F^{(k)}}M_Y(t) | X, F^{(k)}$$

- M-step: We construct the estimator from $M^{(k+1)}_X(t)$.

$$F^{(k+1)} = \frac{1}{Nm} M^{(k+1)}_X(t)$$

- We repeat the E- and M-steps until we have a convergence.
Equivalence Result

Theorem

Suppose that we have a ranked set sample of size $N$. For a given stochastic matrix $P$, the sequence of estimator $(F^{(1)}, F^{(2)}, F^{(3)}, \ldots)$ generated from the EM-algorithm converges to the MLE defined in estimating equations (1).

- The EM-algorithm and estimating equations give the same estimator.
- It appears that the estimator is unique for an arbitrary $P$ as long as $P$ is in the parameter space.
Consistency

Theorem

Suppose that we have a ranked set sample of size $N$ drawn from distribution $F$ with $\lim_{N \to \infty} \frac{n_i}{N} = \epsilon_i > 0$ for $i = 1, \cdots, m$. Assume that $F^{(k)}(t)$ almost surely converges to $F(t)$ as $N$ goes to infinity, then the updated estimator $F^{(k+1)}(t)$ also converges almost surely to $F(t)$.

- If we select a consistent initial value for $F$, then $k$–th iteration of the EM-algorithm will also be consistent.
- We may conjecture from this theorem that NPMLE is a consistent estimator.
- As initial value of $F$, we select

$$F^{(0)} = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{n_i} \sum_{j=1}^{n_i} I(X_{[i]j} \leq t).$$
Likelihood function based on missing data model

- Log-likelihood function for missing data model is given by
  \[
  L(P, \phi) = \sum_{j=1}^{N} \sum_{i=1}^{m} z_{ij} \log(p_{r_j,i}) + \sum_{j=1}^{N} \sum_{i=1}^{m} z_{ij} \log(L_{j,i}(\phi_j))
  \]  
  \(2\)

  \[
  L_{j,i}(\phi_j) = m \binom{m - 1}{i - 1} \phi_j^{i-1} (1 - \phi_j)^{m-i} (\phi_j - \phi_{j-1}).
  \]  
  \(3\)

- We need to maximize this likelihood function over \(P\) and \(\phi\).
- We again use EM-algorithm to find the maximizer.
EM-algorithm

- Let $P^{(0)}$ be an initial value of $P$.
- E-step: For the current value of $P^{(t)}$, we estimate $\phi$ from the EM-algorithm and obtain $\phi^*(t)$. We then evaluate the conditional expectation of log-likelihood function, $Q(P)$, given the observed judgment order statistics $X_{[r]j} (j = 1, \cdots, N)$, $\phi^*(t)$ and $P^{(t)}$, where $Q(P) = E\{L(P, \phi^*(t))|\phi^*(t), P^{(t)}, X_{[r]j}\}$.
- M-step: We find $P^{(t+1)}$ that maximizes $Q(P)$.
- We repeat E- and M-steps until we have a convergence.

Omer Ozturk (OSU)
Quadratic minimization, Ozturk (2008)

A competitive estimator for $(p_{i,j})$ is obtained by minimizing a dispersion function

$$d(P) = \sum_{t=1}^{N} \sum_{j=1}^{m} \left\{ \hat{F}_j(X(t)) - \sum_{s=1}^{m} \hat{p}_{js} B(u_t, s, m + 1 - s) \right\}^2,$$

where $\hat{F}_j(Y^*_t)$ is the empirical cdf of the $j$-th judgment class distribution and $u_t = \hat{F}(X(t))$ is the empirical cdf of $F$ evaluated at $X(t)$.

The estimate of the $j$-th judgment class distribution is then obtained from Bohn-Wolfe model as

$$F_j(u) = \sum_{s=1}^{m} \hat{p}_{j,s} B(F(u), s, m + 1 - s).$$
Quadratic minimization, Ozturk (2008)

- A competitive estimator for \((p_{i,j})\) is obtained by minimizing a dispersion function

\[
d(P) = \sum_{t=1}^{N} \sum_{j=1}^{m} \left\{ \hat{F}_{[j]}(X(t)) - \sum_{s=1}^{m} p_{js} B(u_t, s, m + 1 - s) \right\}^2,
\]

where \(\hat{F}_{[j]}(Y^*_t)\) is the empirical cdf of the \(j\)-th judgment class distribution and \(u_t = \hat{F}(X(t))\) is the empirical cdf of \(F\) evaluated at \(X(t)\).

- The estimate of the \(j\)-th judgment class distribution is then obtained from Bohn-Wolfe model as

\[
F_{[j]}(u) = \sum_{s=1}^{m} \hat{p}_{j,s} B(F(u), s, m + 1 - s).
\]
Simulation Results

Estimation of $p_{1,1}$, $m = 2$

When $\rho = 1$ there is some bias in all estimators.
The bias shrinks when $\rho < 1$.
One parameter model has larger bias, but slightly smaller standard deviation.
Simulation Results

Estimation of $p_{i,j}$, $m = 3$

Data is generated from

$$P = \begin{pmatrix}
0.95 & 0.05 & 0 \\
0.05 & 0.45 & 0.50 \\
0 & 0.50 & 0.50
\end{pmatrix}.$$  

- The NPMLE and Q-estimator have very little bias.
- The Q-estimator has smaller variance than the NPMLE.
- One parameter NPMLE has large bias, but it has slightly smaller variance.
Simulations Results

Estimation of $F$, $m = 2$, $n = 10$

- When $\rho = 1$, all estimators appear to be unbiased.
- When $\rho < 1$, the KS estimator is not a CDF.
Simulation Results

MSE plot of $F, m = 2, n = 10$

- When $\rho = 1$, all MSE curves appear to be the same.
- When $\rho < 1$, the MSE curve of KS estimator has heavier tail on the right.
Simulation Results

Judgment class CDF estimators $F[i]$, $m = 2$, $n = 10$

- When $\rho = 1$, all estimators appear to be unbiased.
- When $\rho < 1$ the KS estimator is biased and $\hat{F}[2]$ is not a cdf.
MSE plot of the judgment class cdf estimators, $m = 2$, $n = 10$

- When $\rho = 1$, all MSE curves appear to be the same.
- When $\rho < 1$, the MSE curve of KS estimator for $\hat{F}_{[2]}$ has heavier tail on the right.
Example: Discharge water

- This data represents the amount of discharge water, in cubic meters per second, for floods on the Nidd River in Yorkshire, England, Kvam and Samaniego (1994).

<table>
<thead>
<tr>
<th>Rank=1</th>
<th>Rank=2</th>
<th>Rank=3</th>
</tr>
</thead>
<tbody>
<tr>
<td>80.12</td>
<td>87.76</td>
<td>111.54</td>
</tr>
<tr>
<td>99.08</td>
<td>123.71</td>
<td>121.73</td>
</tr>
</tbody>
</table>

- The NPMLE of $P$ and $P(\eta)$.

\[
\hat{P} = \begin{pmatrix}
0.951 & 0.049 & 0.000 \\
0.049 & 0.452 & 0.499 \\
0.000 & 0.499 & 0.501 \\
\end{pmatrix},
\quad P(\hat{\eta}) = \begin{pmatrix}
0.736 & 0.226 & 0.037 \\
0.226 & 0.547 & 0.226 \\
0.037 & 0.226 & 0.736 \\
\end{pmatrix}.
\]

- Data suggests that there is not much ranking error between ranking groups 1 and 2, but substantial errors in between groups 2 and 3.
- The estimator $P(\hat{\eta})$ is not flexible enough to explain the ranking structure in the data.
Example (Continued): Estimate of $F$

- All estimators distribute their masses differently.
- The KS estimator, which ignores ranking error, is not a cdf since it does not reach to 1.
Example: Calibration for Two-sample MWW test

Suppose that we wish to test the location shift between $F(y)$ and $G(y) = F(y - \Delta)$. $H_0: \Delta = \theta_F - \theta_G = 0$ against $H_A: \Delta \neq 0$.

We reject the null hypothesis for too large (or too small) values of rank-sum statistics (Bohn and Wolfe, 1992), $\bar{T}$, of a ranked set sample.

The limiting null distribution of $\bar{T}$ is normal with mean zero and variance $\sigma^2_{\bar{T}} = \frac{\xi_{1,0}}{\lambda} + \frac{\xi_{0,1}}{(1 - \lambda)}$,

$$\xi_{0,1} = \frac{1}{3} - \frac{1}{k} \sum_{i=1}^{k} \left\{ \int F_{[i]}(y) dF(y) \right\}^2$$

$$\xi_{1,0} = \frac{1}{3} - \frac{1}{q} \sum_{i=1}^{q} \left\{ \int F_{[i]}(y) dF(y) \right\}^2 .$$

The limiting null distribution is not distribution-free if there is ranking error.

We estimate $\xi_{0,1}$ and $\xi_{1,0}$ by using NPMLE of $F_{[i]}$. 
Empirical type I error rates

<table>
<thead>
<tr>
<th>n</th>
<th>m</th>
<th>Est</th>
<th>ρ = 0.5</th>
<th>ρ = 0.75</th>
<th>ρ = 1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2</td>
<td>NPMLE</td>
<td>0.040</td>
<td>0.046</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td></td>
<td>One</td>
<td>0.036</td>
<td>0.036</td>
<td>0.036</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Perf</td>
<td>0.086</td>
<td>0.088</td>
<td>0.064</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>NPMLE</td>
<td>0.076</td>
<td>0.070</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td></td>
<td>One</td>
<td>0.056</td>
<td>0.044</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Perf</td>
<td>0.170</td>
<td>0.102</td>
<td>0.060</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>NPMLE</td>
<td>0.048</td>
<td>0.058</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td></td>
<td>One</td>
<td>0.032</td>
<td>0.052</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Perf</td>
<td>0.102</td>
<td>0.066</td>
<td>0.052</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>NPMLE</td>
<td>0.068</td>
<td>0.074</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td></td>
<td>One</td>
<td>0.052</td>
<td>0.050</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Perf</td>
<td>0.148</td>
<td>0.110</td>
<td>0.058</td>
</tr>
</tbody>
</table>

- Under perfect ranking the Type I error rates are inflated when $\rho < 1$.
- The one-parameter model provides reasonable calibration for the test.
- When $m = 3$ and $\rho < 1$, the NPMLE slightly overestimate the Type I error rates.
### Empirical coverage probabilities

<table>
<thead>
<tr>
<th>n</th>
<th>m</th>
<th>Est</th>
<th>$\rho = 0.5$</th>
<th>$\rho = 0.75$</th>
<th>$\rho = 1.00$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2</td>
<td>NPMLE</td>
<td>0.956</td>
<td>0.946</td>
<td>0.960</td>
</tr>
<tr>
<td></td>
<td></td>
<td>One</td>
<td>0.962</td>
<td>0.958</td>
<td>0.962</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Perf</td>
<td>0.930</td>
<td>0.918</td>
<td>0.944</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>NPMLE</td>
<td>0.926</td>
<td>0.938</td>
<td>0.954</td>
</tr>
<tr>
<td></td>
<td></td>
<td>One</td>
<td>0.948</td>
<td>0.958</td>
<td>0.956</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Perf</td>
<td>0.824</td>
<td>0.890</td>
<td>0.932</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>NPMLE</td>
<td>0.950</td>
<td>0.942</td>
<td>0.952</td>
</tr>
<tr>
<td></td>
<td></td>
<td>One</td>
<td>0.966</td>
<td>0.948</td>
<td>0.958</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Perf</td>
<td>0.894</td>
<td>0.930</td>
<td>0.948</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>NPMLE</td>
<td>0.932</td>
<td>0.926</td>
<td>0.950</td>
</tr>
<tr>
<td></td>
<td></td>
<td>One</td>
<td>0.948</td>
<td>0.952</td>
<td>0.952</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Perf</td>
<td>0.856</td>
<td>0.890</td>
<td>0.942</td>
</tr>
</tbody>
</table>

- Under perfect ranking coverage probabilities are deflated when $\rho < 1$.
- The one-parameter model provides a reasonable adjustment.
- When $m = 3$ and $\rho < 1$, the NPMLE slightly underestimate the coverage probabilities.
Summary

- We proposed NPMLE for the within-set ranking error probabilities and the cdf of the underlying population.
- The NPMLEs of $p_{i,j}$ have some bias when the true values are at the edge of the parameter space. This bias gets smaller when $p_{i,j}$s stay away from 0 or 1.
- The estimators would be helpful to reduce the impact of ranking errors on statistical procedures based on ranked set sample data.