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

Omer Ozturk

The Ohio State University

June 24, 2011

Outline

Outline

Outline

Outline

Outline

Outline

Outline

Outline

Outline

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 $X_{[1]i}, ..., X_{[m]i}$ are called a <u>cycle</u>.
- Note: Process repeated $i = 1, \cdots, n$ cycle to get nm observations. These nm observations are called a <u>standard ranked set sample</u>.

Ranked Set Sample Diagram

Let m=3 and n=2

Cycle	Judgment Rank 1	2	3
1	$X_{[1]1}$	$X_{[2]1}$	$X_{[3]1}$
	$X_{[1]1}$	$X_{[2]1}$	$X_{[3]1}$
	$X_{[1]1}$	$X_{[2]1}$	$X_{[3]1}$
2	$X_{[1]2}$	$X_{[2]2}$	$X_{[3]2}$
	$X_{[1]2}$	$X_{[2]2}$	$X_{[3]2}$
	$X_{[1]2}$	$X_{[2]2}$	$X_{[3]2}$

$X_{[1]1}, \cdots, X_{[3]2}$ is called a ranked set sample.

- In each set, colored unit is selected for gold standard measurement.
- $X_{[i]j}, i = 1, \cdots, m, j = 1, \cdots, n$ are all independent, but not identically distributed.
- For each fixed i, $X_{[i]j}, j = 1, \cdots, n$ are iid with judgment class cdf $F_{[i]}$.
- If there is no ranking error, the judgment order statistic $X_{[i]j}$ 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

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

- 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.

- Type I error Rates of MWW test

Corr	n	m	α
1	5	2	0.060
		3	0.053
	10	2	0.054
		3	0.059
0.5	5	2	0.112
		3	0.147
	10	2	0.097
		3	0.144

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.

- Type I error Rates of MWW test

Corr	n	m	α
1	5	2	0.060
		3	0.053
	10	2	0.054
		3	0.059
0.5	5	2	0.112
		3	0.147
	10	2	0.097
		3	0.144

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.

- $\boldsymbol{P} = (p_{i,j})$ is a doubly stochastic matrix.
- One parameter model, Frey (2007): Judgment ranking probabilities, $p_{i,j}$, expressed as a function of a single parameter, η,

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

- In these models, we are interested in the estimation of $\boldsymbol{P}$ (or $\boldsymbol{P}(\eta)$) 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.
- $\boldsymbol{P} = (p_{i,j})$ is a doubly stochastic matrix.
- One parameter model, Frey (2007): Judgment ranking probabilities, $p_{i,j}$, expressed as a function of a single parameter, η,

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

- In these models, we are interested in the estimation of $\boldsymbol{P}$ (or $\boldsymbol{P}(\eta)$) 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.
- $\boldsymbol{P} = (p_{i,j})$ is a doubly stochastic matrix.
- One parameter model, Frey (2007): Judgment ranking probabilities, $p_{i,j}$, expressed as a function of a single parameter, η,

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

- In these models, we are interested in the estimation of $\boldsymbol{P}$ (or $\boldsymbol{P}(\eta)$) and the underlying distribution function F.

Model:Continued

- Dell and Clutter Model (1972): Ranking is performed based on perceived values of experimental units.
  1. We generate a set of m observations, $\boldsymbol{Y} = (Y_1, \cdots, Y_m)$, from a distribution F with mean θ and variance σ^2
  2. We generate another independent random vector, $\boldsymbol{w} = (w_1, \cdots, w_m)$ from a normal distribution with mean zero and variance τ^2. We add $\boldsymbol{Y}$ and $\boldsymbol{w}$ to obtain $\boldsymbol{X} = \boldsymbol{Y} + \boldsymbol{w}$
  3. We sort the vector $\boldsymbol{X}$ and select the $Y_{[j]}$ as the j-th judgment order statistics that corresponds to the j-th position in the sorted vector $\boldsymbol{X}$.
  4. Quality of judgment ranking is controlled by the correlation coefficient between X and Y, $\rho = corr(X, Y) = \frac{\sigma}{\sqrt{\sigma^2+\tau^2}}$.

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

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

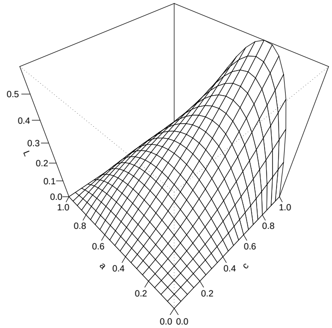
- The parameter space:

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

- In this model we wist to estimate $\boldsymbol{P}$ and ϕ.
- The likelihood function $L(\boldsymbol{I}, \phi)$ is considered by Kvam and Samaniego (1994), where $\boldsymbol{I}$ is identity matrix.

A Simple Example: $X_{[1]} < X_{[2]}$

Likelihood surface



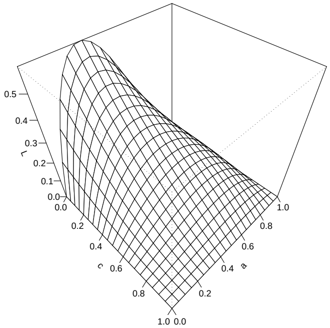
- Let $a = dF(X_{[1]}), b = dF(X_{[2]})$ and

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

- 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 surface



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

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

- 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 ϕ exists for any **P** and is unique for some **P**, $\mathbf{P} \in \mathcal{P}$.*

- For a fixed value $\boldsymbol{P}$, NPMLE of ϕ is obtained as a solution of the following estimating equation

$$
\begin{aligned}
\frac{\sum_{s=1}^{m} A_{1,s}\left\{\frac{s-1}{\phi_1} - \frac{m-s}{1-\phi_1}\right\}}{\sum_{s=1}^{m} A_{1,s}} + \frac{1}{\phi_1} - \frac{1}{\bar{\phi}_2} &= 0 \\
\frac{\sum_{s=1}^{m} A_{i,s}\left\{\frac{s-1}{\phi_i} - \frac{m-s}{1-\phi_i}\right\}}{\sum_{s=1}^{m} A_{i,s}} + \frac{1}{\bar{\phi}_i} - \frac{1}{\bar{\phi}_{i+1}} &= 0, i = 2, \cdots, N-1 \\
\frac{\sum_{s=1}^{m} A_{N,s}\left\{\frac{s-1}{\phi_N} - \frac{m-s}{1-\phi_N}\right\}}{\sum_{s=1}^{m} A_{N,s}} + \frac{1}{\bar{\phi}_N} &= 0, \qquad (1)
\end{aligned}
$$

Missing data model

- Let $\boldsymbol{Y}_{[j]}$ be the vector of m within-set order statistics

$$\boldsymbol{Y}_{[j]}^{\top} = (Y_{(1)j} < \cdots < Y_{(m)j}).$$

- Let $\boldsymbol{Z}_{[r_j]}^{\top} = (z_{1j}, \cdots, z_{mj})$ be a multinomial random vector with parameter 1 and $\boldsymbol{p}_{r_j}$, where $\boldsymbol{p}_{r_j} = (p_{r_j,1}, \cdots, p_{r_j,m})$ is the r_j-th row of $\boldsymbol{P}$.
- The complete data then can be expressed as

$$(\boldsymbol{Y}_{[j]}, \boldsymbol{Z}_{[r_j]}), j = 1, \cdots, N.$$

- For each r_j, based on BW model with parameter $\boldsymbol{P}$, we observe the r_j-th judgment order statistic, $X_{[r_j]j} = \boldsymbol{Z}_{[r_j]}^{\top} \boldsymbol{Y}_{[j]}$.

EM-Algorithm

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

$$M_{\boldsymbol{Y}}(t) = \sum_{j=1}^{N} \sum_{i=1}^{m} I(Y_{(i)j} \leq t).$$

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

$$M_{\boldsymbol{X}}^{(k+1)}(t) = E_{F^{(k)}} M_{\boldsymbol{Y}}(t) | \boldsymbol{X}, F^{(k)}$$

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

$$F^{(k+1)} = \frac{1}{Nm} M_{\boldsymbol{X}}^{(k+1)}(t)$$

- 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 $\boldsymbol{P}$*, the sequence of estimator* $(F^{(1)}, F^{(2)}, F^{(3)}, \cdots)$ *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 $\boldsymbol{P}$ as long as $\boldsymbol{P}$ 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

$$
\begin{aligned}
L(\boldsymbol{P},\phi) &= \sum_{j=1}^{N}\sum_{i=1}^{m} z_{ij} log(p_{r_j,i}) + \sum_{j=1}^{N}\sum_{i=1}^{m} z_{ij} log(\mathcal{L}_{j,i}(\phi_j)) \qquad (2)\\
\mathcal{L}_{j,i}(\phi_j) &= m\binom{m-1}{i-1}\phi_j^{i-1}\{1-\phi_j\}^{m-i}(\phi_j-\phi_{j-1}). \qquad (3)
\end{aligned}
$$

- We need to maximize this likelihood function over $\boldsymbol{P}$ and ϕ.
- We again use EM-algorithm to find the maximizer.

EM-algorithm

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

Quadratic minimization, Ozturk (2008)

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

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

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

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

Quadratic minimization, Ozturk (2008)

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

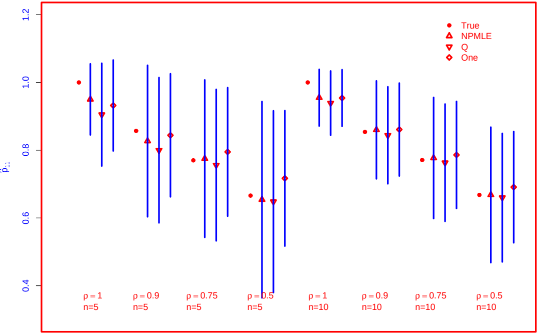
$$d(\boldsymbol{P}) = \sum_{t=1}^{N} \sum_{j=1}^{m} \left\{ \hat{F}_{[j]}(X_{(t)}) - \sum_{s=1}^{m} p_{js} B(u_t, s, m+1-s) \right\}^2,$$

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

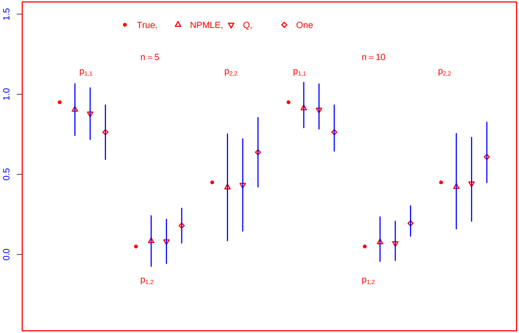
$$F_{[j]}(u) = \sum_{s=1}^{m} \hat{p}_{j,s} B(F(u), s, m+1-s).$$

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.

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

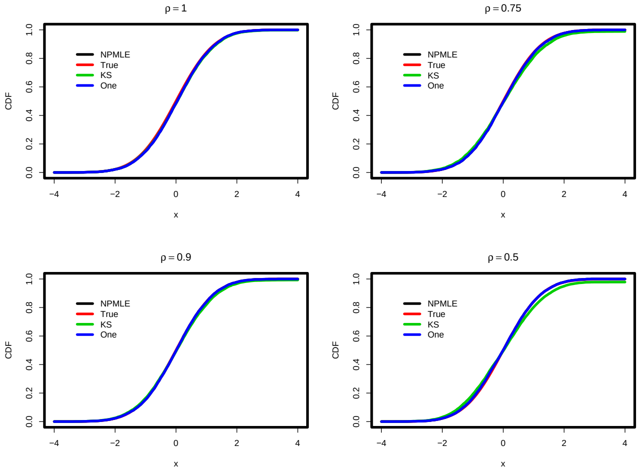


- Data is generated from

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

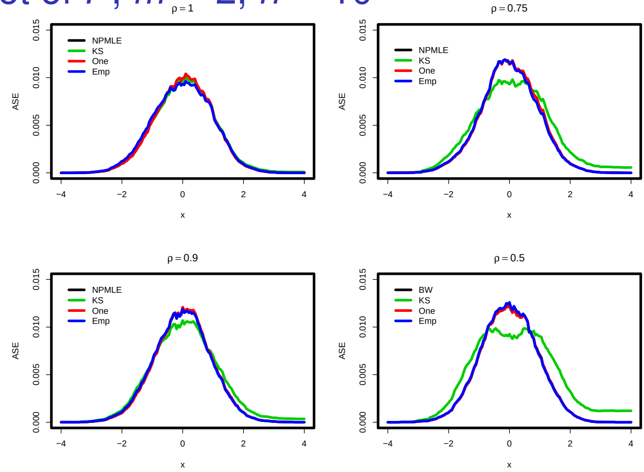
- The NPMLE and Q-estimator have very little bias.
- The Q-estimator has smaller variance than the NPMLE.
- One parameter NPMLE has large bias, but it has slightly smaller variance.

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



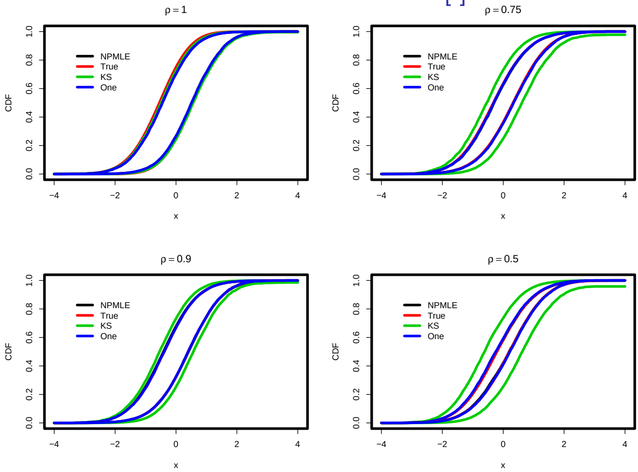
- When $\rho = 1$, all estimators appear to be unbiased.
- When $\rho < 1$, the KS estimator is not a CDF.

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



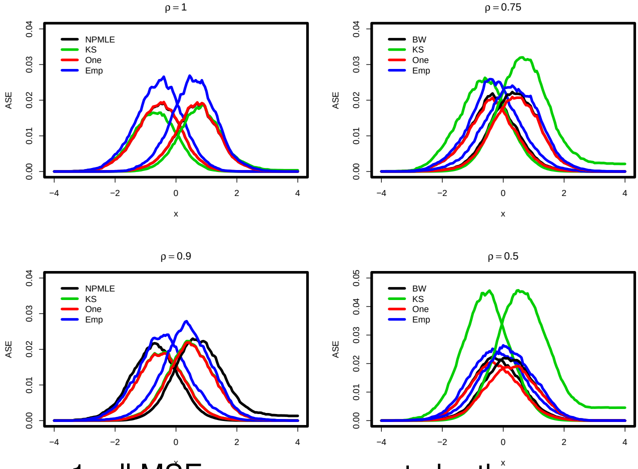
- When $\rho = 1$, all MSE curves appear to be the same.
- When $\rho < 1$, the MSE curve of KS estimator has heavier tail on the right.

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).

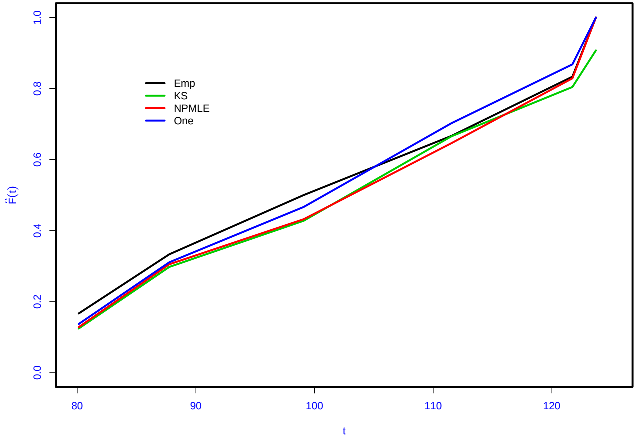
Rank=1	Rank=2	Rank=3
80.12	87.76	111.54
99.08	123.71	121.73

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

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

- 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 $\boldsymbol{P}(\hat{\eta})$ 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 $\sigma^2_{\bar{T}} = \xi_{1,0}/\lambda + \xi_{0,1}/(1 - \lambda)$,

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

- 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

n	m	Est	$\rho = 0.5$	$\rho = 0.75$	$\rho = 1.00$
5	2	NPMLE	0.040	0.046	0.034
		One	0.036	0.036	0.036
		Perf	0.086	0.088	0.064
5	3	NPMLE	0.076	0.070	0.052
		One	0.056	0.044	0.050
		Perf	0.170	0.102	0.060
10	2	NPMLE	0.048	0.058	0.042
		One	0.032	0.052	0.038
		Perf	0.102	0.066	0.052
10	3	NPMLE	0.068	0.074	0.054
		One	0.052	0.050	0.052
		Perf	0.148	0.110	0.058

- 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

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

- 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.