# TITLE HERE

A Thesis Proposal Presented to
The Faculty of the School of Statistics
univ

In Partial Fulfillment
of the Requirements for the Degree of
Master of Science in Statistics
1st Semester A.Y. 2025-2026

by
Shaine Rosewel Matala

# Contents


**1 Introduction** **3**

**2 Related Literature** **3**

2.1 Joint confidence region for an overall ranking . . . . . . . . . . . . . . . . . . 3

2.2 Joint confidence region construction . . . . . . . . . . . . . . . . . . . . . . . . . 3

2.3 $T_1, T_2, T_3$ . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

**3 Methodology** **3**

3.1 Joint confidence intervals for $\theta_1, \ldots, \theta_K$ by using Parametric Bootstrap . 3

3.2 Joint confidence intervals for $\theta_1, \ldots, \theta_K$ by using Nonrank-based method 4

3.3 Evaluation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5

# 1 Introduction

# 2 Related Literature

## 2.1 Joint confidence region for an overall ranking

## 2.2 Joint confidence region construction

## 2.3 $T_1, T_2, T_3$

# 3 Methodology

This section introduces the proposed methodologies to obtain confidence regions for the unknown overall true ranking. The following cases are tackled: case when items ranked are assumed to have zero and nonzero correlation. Both approaches are based on parametric bootstrap. Sections 3.1 and 3.2 discuss the algorithms for the cases mentioned. Section 3.3 shows the algorithms used to assess the performance of the proposed approaches. This makes use of coverage and metrics to measure the tightness of the estimated confidence regions.

For sections 3.1 and 3.2, let $\theta_1, \theta_2, \ldots, \theta_K$ be the true parameter values and $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K$ be the corresponding estimates.

## 3.1 Joint confidence intervals for $\theta_1, \ldots, \theta_K$ by using Parametric Bootstrap

The rank-based parametric bootstrap approach assumes $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K$ to be independent but not identically distributed estimates, where $\hat{\theta}_k \sim N\left(\theta_k, \sigma_k^2\right)$, $k = 1, 2, \ldots, K$. $\sigma_k^2$ is assumed known. Denote the corresponding ordered values by $\hat{\theta}_{(1)}, \hat{\theta}_{(2)}, \ldots, \hat{\theta}_{(K)}$.

**Algorithm 1** Computation of Joint Confidence Region using Parametric Bootstrap

1: **for** $b = 1, 2, \ldots, B$ **do**

2: Generate $\hat{\theta}^*_{bk} \sim N\left(\hat{\theta}_k, \sigma_k^2\right)$, $k = 1, 2, \ldots, K$ and let $\hat{\theta}_{b(1)}, \hat{\theta}_{b(2)}, \ldots, \hat{\theta}_{b(K)}$ be the corresponding ordered values

| | $k=1$ | $k=2$ | $\ldots$ | $k=K$ |
|---|---|---|---|---|
| $b=1$ | $\hat{\theta}^*_{1(1)}$ | $\hat{\theta}^*_{1(2)}$ | $\ldots$ | $\hat{\theta}^*_{1(K)}$ |
| $b=2$ | $\hat{\theta}^*_{2(1)}$ | $\hat{\theta}^*_{2(2)}$ | $\ldots$ | $\hat{\theta}^*_{2(K)}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\ldots$ | $\vdots$ |
| $b=B$ | $\hat{\theta}^*_{B(1)}$ | $\hat{\theta}^*_{B(2)}$ | $\ldots$ | $\hat{\theta}^*_{B(K)}$ |

3: Compute

$$\hat{\sigma}^*_{b(k)} = \sqrt{\text{kth ordered value among } \left\{\hat{\theta}^{*2}_{b1} + \sigma_1^2, \hat{\theta}^{*2}_{b2} + \sigma_2^2, \ldots, \hat{\theta}^{*2}_{bK} + \sigma_K^2\right\} - \hat{\theta}^{*2}_{(k)}}$$

4: Compute $t^*_b = \max\limits_{1 \le k \le K} \left| \frac{\hat{\theta}^*_{b(k)} - \hat{\theta}^*_k}{\sigma^*_{b(k)}} \right|$

5: **end for**

6: Compute the $(1-\alpha)$-sample quantile of $t^*_1, t^*_2, \ldots, t^*_B$, call this $\hat{t}$.

7: The joint confidence region of $\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(K)}$ is given by

$$\mathfrak{R} = \left[\hat{\theta}_{(1)} \pm \hat{t} \times \hat{\sigma}_{(1)}\right] \times \left[\hat{\theta}_{(2)} \pm \hat{t} \times \hat{\sigma}_{(2)}\right] \times \cdots \times \left[\hat{\theta}_{(K)} \pm \hat{t} \times \hat{\sigma}_{(K)}\right]$$

where $\hat{\sigma}_{(k)}$ is computed as

$$\hat{\sigma}_{(k)} = \sqrt{\text{kth ordered value among } \left\{\hat{\theta}^2_1 + \sigma_1^2, \hat{\theta}^2_2 + \sigma_2^2, \ldots, \hat{\theta}^2_K + \sigma_K^2\right\} - \hat{\theta}^2_{(k)}}$$

## 3.2 Joint confidence intervals for $\theta_1, \ldots, \theta_K$ by using Nonrank-based method

The nonrank-based method assumes that $\hat{\boldsymbol{\theta}} = \left(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K\right) \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$. It accounts for potential correlation among items being ranked. For this case, an exchangeable correlation, $\boldsymbol{\rho}$ (See Equation 3.1.), is assumed and used in the calculation of the variance covariance matrix (See Equation 3.2.).

$$\boldsymbol{\rho} = (1-\rho)\,\mathbf{I}_K + \rho \mathbf{1}_K \mathbf{1}'_K \tag{3.1}$$

$$\boldsymbol{\Sigma} = \boldsymbol{\Delta}^{1/2} \boldsymbol{\rho} \boldsymbol{\Delta}^{1/2} \tag{3.2}$$

where $\boldsymbol{\Delta} = \operatorname{diag}\{\sigma_1^2, \sigma_2^2, \ldots, \sigma_K^2\}$, with known $\sigma_k$'s and $\rho$ is studied for $0.1, 0.5, 0.9$.

**Algorithm 2** Computation of Joint Confidence Region using Nonrank-based Method

Let the data consist of $\hat{\theta}_1, \ldots, \hat{\theta}_K$ and suppose $\mathbf{\Sigma}$ is known

1: **for** $b = 1, 2, \ldots, B$ **do**

2: Generate $\hat{\boldsymbol{\theta}}_b^* \sim N_K\left(\hat{\boldsymbol{\theta}}, \mathbf{\Sigma}\right)$ and write $\hat{\boldsymbol{\theta}}_b^* = \left(\hat{\theta}_{b1}^*, \hat{\theta}_{b2}^*, \ldots, \hat{\theta}_{bK}^*\right)'$

3: Compute $t_b^* = \max\limits_{1 \le k \le K} \left| \frac{\hat{\theta}_{bk}^* - \hat{\theta}_k^*}{\sigma_k} \right|$

4: **end for**

5: Compute the $(1-\alpha)$-sample quantile of $t_1^*, t_2^*, \ldots, t_B^*$, call this $\hat{t}$.

6: The joint confidence region of $\theta_1, \theta_2, \ldots, \theta_K$ is given by

$$\mathfrak{R} = \left[\hat{\theta}_1 \pm \hat{t} \times \sigma_1\right] \times \left[\hat{\theta}_2 \pm \hat{t} \times \sigma_2\right] \times \cdots \times \left[\hat{\theta}_K \pm \hat{t} \times \sigma_K\right]$$

## 3.3 Evaluation

Algorithm 3 is used to calculate the coverage which is defined as the proportion of times that the true parameter values fall within the confidence interval for all $K$ simultaneously. Ideally, this should be equal to 0.90 since $\alpha = 0.1$. It also calculates the average $T_1, T_2$, and $T_3$. Higher values of $T_1$ and $T_2$ indicate wider confidence intervals and are therefore less desirable, whereas higher values of $T_3$ are preferable.

**Algorithm 3** Computation of Coverage Probability for Parametric Bootstrap

For given values of $\theta_1, \theta_2, \ldots, \theta_K$ and thus $\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(K)}$

1: **for** replications $= 1, 2, \ldots, 5000$ **do**

2: Generate $\hat{\theta}_k \sim N(\theta_k, \sigma_k^2)$, for $k = 1, 2, \ldots, K$

3: Compute the rectangular confidence region $\mathfrak{R}$ using Algorithm 1.

4: Check if $\left(\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(K)}\right) \in \mathfrak{R}$ and compute

$$T_1 = \tfrac{1}{K} \textstyle\sum_{k=1}^K \left|\Lambda_{Ok}\right|$$

$$T_2 = \textstyle\prod_{k=1}^K \left|\Lambda_{Ok}\right|$$

$$T_3 = 1 - \frac{K + \sum_{k=1}^K \left|\Lambda_{Ok}\right|}{K^2}$$

5: **end for**

6: Compute the proportion of times that the condition in step 4 is satisfied and the average of $T_1, T_2$, and $T_3$.

Algorithm 4 is similar to Algorithm 3 but computes for the coverage and average $T_1, T_2$, and $T_3$ for the nonrank-based method.

**Algorithm 4** Computation of Coverage Probability for Nonrank-based Method

For given values of $\theta_1, \theta_2, \ldots, \theta_K$ and $\mathbf{\Sigma}$

1: **for** replications $= 1, 2, \ldots, 5000$ **do**

2: Generate $\hat{\boldsymbol{\theta}} \sim N_K(\boldsymbol{\theta}, \mathbf{\Sigma})$

3: Compute the rectangular confidence region $\mathfrak{R}$ using Algorithm 2.

4: Check if $(\theta_1, \theta_2, \ldots, \theta_K) \in \mathfrak{R}$ and compute $T_1, T_2$, and $T_3$.

5: **end for**

6: Compute the proportion of times that the condition in step 4 is satisfied and the average of $T_1, T_2$, and $T_3$.

Klein et al. (2020)

# Bibliography

Klein, M., Wright, T., & Wieczorek, J. (2020). *A joint confidence region for an overall ranking of populations.*