

Joint Confidence Regions for Rankings based on Correlated Estimates

Matala, Shaine Rosewel

University of the Philippines

November 25, 2025

Background of the Study I

Klein et al. (2020) illustrates the problem of ranking the US states in terms of travel time to work. In this context, θ_k is the true mean travel time to work for state k , and we wish to find a joint confidence statement about r_1, r_2, \dots, r_K .

Rank r_k

A mathematical definition of r_k is as follows:

$$r_k = \sum_{j=1}^K I(\theta_j \leq \theta_k) = 1 + \sum_{j:j \neq k} I(\theta_j \leq \theta_k), \quad \text{for } k = 1, 2, \dots, K. \quad (1)$$

Background of the Study II

Estimated rank

Let \hat{r}_k denote the estimated rank of the k th observation.

$$\hat{r}_k = 1 + \sum_{j:j \neq k} I(\hat{\theta}_j \leq \hat{\theta}_k), \quad \text{for } k = 1, 2, \dots, K. \quad (2)$$

Assuming independence when constructing joint confidence regions for estimators that are, in fact, correlated may lead to overly conservative and thus wider intervals, implying greater uncertainty.

Political setting

- David & Legara (2015) demonstrated that candidates with a name-recall advantage, such as media celebrities, incumbents, and members of dynastic families, received majority of the votes in the 2010 senatorial elections.

Political setting

- David & Legara (2015) demonstrated that candidates with a name-recall advantage, such as media celebrities, incumbents, and members of dynastic families, received majority of the votes in the 2010 senatorial elections.
- In weak-party systems, candidates who belong to the same political alliance or ticket commonly co-occur in ballots and hence perform with similarity.

Measurement across geographies

- Klein et al. (2020) also noted that states with large unpopulated land areas and relatively few high-density population centers tend to report shorter travel times while longer travel times are typically observed in highly urbanized states with large populations and high population densities.

Measurement across geographies

- Klein et al. (2020) also noted that states with large unpopulated land areas and relatively few high-density population centers tend to report shorter travel times while longer travel times are typically observed in highly urbanized states with large populations and high population densities.
- Many states with shorter travel times are located in the Mountain and Central regions, whereas majority of those with longer travel times are concentrated along the East Coast.

Objective

This research aims to do the following:

- Develop a procedure to construct joint confidence intervals for the ranks and the ranked parameters when the estimates to be ranked may be correlated.
- Evaluate the performance of the proposed approaches under various parameter settings.
- Apply the proposed approaches to a real-life example.

Definitions and Assumptions

- Define $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_K)'$ and assume that $\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ where $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_K)'$ is unknown and $\boldsymbol{\Sigma}$ is a known $K \times K$ positive definite matrix. The diagonal elements of $\boldsymbol{\Sigma}$ are $\sigma_1^2, \dots, \sigma_K^2$.
- Note that in the literature on inferences on the ranks, it is customary to assume that the variances are known.

Procedure I

- 1 Derive simultaneous confidence intervals for $\theta_1, \theta_2, \dots, \theta_K$ of the form

$$\mathfrak{R}_1 = [\hat{\theta}_1 \pm t \times \sigma_1] \times [\hat{\theta}_2 \pm t \times \sigma_2] \times \cdots \times [\hat{\theta}_K \pm t \times \sigma_K]. \quad (3)$$

- 2 Once the confidence intervals in (3) have been obtained, we can then use the result of Klein et al. (2020) in (??) to get the lower and upper bounds on the ranks $r_k, k = 1, 2, \dots, K$. That is, we also get a joint confidence region for r_1, r_2, \dots, r_K .

Proposed methodology to compute the joint confidence region for the unordered parameters:

Algorithm 1

Let the data be represented by $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_K)'$ and suppose that $\boldsymbol{\Sigma}$ is known

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

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

3: Compute

$$t_b^* = \max_{1 \leq k \leq K} \left| \frac{\hat{\theta}_{bk}^* - \hat{\theta}_k}{\sigma_k} \right|$$

4: **end for**

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

6: The joint confidence region for $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_K)'$ is given by

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

Algorithm 1: Quantile Calculation

We want the joint confidence region in (3) to satisfy the following probability condition:

$$P\left(\hat{\theta}_k - t \cdot \sigma_k \leq \theta_k \leq \hat{\theta}_k + t \cdot \sigma_k, \forall k = 1, 2, \dots, K\right) = 1 - \alpha. \quad (4)$$

Equivalently, we require

$$P\left(\max_{k=1,2,\dots,K} \left| \frac{\hat{\theta}_k - \theta_k}{\sigma_k} \right| \leq t\right) = 1 - \alpha. \quad (5)$$

Proposed methodology to compute a joint confidence region for the ordered parameters

- 1: **for** $b = 1, 2, \dots, B$ **do**
- 2: Generate $\hat{\theta}_b^* = (\hat{\theta}_{b(1)}^*, \hat{\theta}_{b(2)}^*, \dots, \hat{\theta}_{b(K)}^*)' \sim N_K(\hat{\theta}, \Sigma)$ and let $\hat{\theta}_{b(1)}^*, \hat{\theta}_{b(2)}^*, \dots, \hat{\theta}_{b(K)}^*$ be the corresponding ordered values
- 3: Compute $\hat{\sigma}_{b(k)}^*$ using:
 - asymptotic variance definition
 - second-level bootstrap
- 4: Compute

$$t_b^* = \max_{1 \leq k \leq K} \left| \frac{\hat{\theta}_{b(k)}^* - \hat{\theta}_k}{\hat{\sigma}_{b(k)}^*} \right|$$

- 5: **end for**
- 6: Compute the $(1 - \alpha)$ -sample quantile of t_1^*, \dots, t_B^* , call this \hat{t} .
- 7: The joint confidence region of $\theta_{(1)}, \theta_{(2)}, \dots, \theta_{(K)}$ is given by
$$\mathfrak{R}_2 = [\hat{\theta}_{(1)} \pm \hat{t} \times \hat{\sigma}_{(1)}] \times [\hat{\theta}_{(2)} \pm \hat{t} \times \hat{\sigma}_{(2)}] \times \dots \times [\hat{\theta}_{(K)} \pm \hat{t} \times \hat{\sigma}_{(K)}]$$

Algorithm 2: Asymptotic Definition of Variance

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

Algorithm 3: Variance from Second-Level Bootstrap

- 1: **for** $c = 1, 2, \dots, C$ **do**
- 2: Generate $\hat{\theta}_{bc}^{**} = (\hat{\theta}_{bc1}^{**}, \hat{\theta}_{bc2}^{**}, \dots, \hat{\theta}_{bcK}^{**}) \sim N_K(\hat{\theta}_b^*, \Sigma)$ and let $\hat{\theta}_{bc(1)}^{**}, \hat{\theta}_{bc(2)}^{**}, \dots, \hat{\theta}_{bc(K)}^{**}$ be the corresponding ordered values of $\hat{\theta}_{bc1}^{**}, \hat{\theta}_{bc2}^{**}, \dots, \hat{\theta}_{bcK}^{**}$
- 3: Compute $\hat{\sigma}_{b(k)}^* = \frac{\sum_{c=1}^C (\hat{\theta}_{bc(k)}^{**} - \bar{\hat{\theta}}_{b \cdot (k)}^{**})^2}{C - 1}$, $\bar{\hat{\theta}}_{b \cdot (k)}^{**} = \frac{1}{C} \sum_{c=1}^C \hat{\theta}_{bc(k)}^{**}$
- 4: **end for**

Evaluation Algorithm

For given values of Σ and $\theta_1, \theta_2, \dots, \theta_K$ (with corresponding $\theta_{(1)}, \theta_{(2)}, \dots, \theta_{(K)}$ for rank-based methods)

- 1: **for** replications = 1, 2, ..., 5000 **do**
- 2: Generate $\hat{\theta} \sim N_K(\theta, \Sigma)$
- 3: Compute the confidence region \mathfrak{R}_1 for the unordered parameters using Algorithm 1 and the confidence region for the ordered parameters \mathfrak{R}_2 using Algorithms 2 and 3.
- 4: For the unordered parameters, check if $(\theta_1, \theta_2, \dots, \theta_K) \in \mathfrak{R}_1$ and compute T_1, T_2 , and T_3 . For the ordered parameters, check if $(\theta_{(1)}, \theta_{(2)}, \dots, \theta_{(K)}) \in \mathfrak{R}_2$
- 5: **end for**
- 6: Compute the proportion of times that the condition in line 4 is satisfied and the average of T_1, T_2 , and T_3 .

Covariance Matrix Σ

- The covariance matrix Σ need not be a diagonal matrix.
- We assume that $V(\hat{\theta}) = \Sigma$ is known and express Σ as in (6), where \mathbf{R} is the population correlation matrix.

$$\Sigma = \Delta^{1/2} \mathbf{R} \Delta^{1/2}; \quad \Delta = \text{diag} \left\{ \sigma_1^2, \sigma_2^2, \dots, \sigma_K^2 \right\}. \quad (6)$$

- The diagonal elements of Σ , which are $\sigma_k^2 = V(\hat{\theta}_k)$ for $k = 1, 2, \dots, K$, are treated as known quantities in practice.

Correlation Structures: Equicorrelation

- We assume certain correlation structures among the $\hat{\theta}$ s.

Correlation Structures: Equicorrelation

- We assume certain correlation structures among the $\hat{\theta}$ s.
- Equicorrelation is considered for simplicity.

Correlation Structures: Equicorrelation

- We assume certain correlation structures among the $\hat{\theta}$ s.
- Equicorrelation is considered for simplicity.
- This assumes that the k variables are equally correlated, i.e., that $\rho_{jk} = \rho$ where $\rho \in [-1, 1]$ for $j \neq k \in \{1, \dots, K\}$.

$$\mathbf{R}_{\text{eq}} = (1 - \rho) \mathbf{I}_K + \rho \mathbf{1}_K \mathbf{1}_K' = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}_{K \times K} \quad (7)$$

Correlation Structures: Block correlation

- Useful in the context of pre-election surveys

Correlation Structures: Block correlation

- Useful in the context of pre-election surveys
- In a block correlation matrix \mathbf{R}_{block} with G blocks, each diagonal block represents an equicorrelation structure within group g , denoted by

$$\mathbf{R}_{eq,g} = (1 - \rho_g) \mathbf{I}_{n_g} + \rho_g \mathbf{1}_{n_g} \mathbf{1}_{n_g}'$$

where ρ_g is the within-block correlation and n_g is the number of variables in block g such that $\sum_{g=1}^G n_g = K$.

Correlation Structures: Block correlation

- Useful in the context of pre-election surveys
- In a block correlation matrix \mathbf{R}_{block} with G blocks, each diagonal block represents an equicorrelation structure within group g , denoted by

$$\mathbf{R}_{eq,g} = (1 - \rho_g) \mathbf{I}_{n_g} + \rho_g \mathbf{1}_{n_g} \mathbf{1}_{n_g}'$$

where ρ_g is the within-block correlation and n_g is the number of variables in block g such that $\sum_{g=1}^G n_g = K$.

- The off-diagonal blocks capture between-block correlations, represented by $\mathbf{C}_{g'g} = \mathbf{C}_{gg'} = \rho_{gg'} \mathbf{1}_{n_g} \mathbf{1}_{n_{g'}}'$ where $g \neq g' \in \{1, \dots, G\}$

Correlation Structures: Block correlation

- Useful in the context of pre-election surveys
- In a block correlation matrix \mathbf{R}_{block} with G blocks, each diagonal block represents an equicorrelation structure within group g , denoted by

$$\mathbf{R}_{eq,g} = (1 - \rho_g) \mathbf{I}_{n_g} + \rho_g \mathbf{1}_{n_g} \mathbf{1}_{n_g}'$$

where ρ_g is the within-block correlation and n_g is the number of variables in block g such that $\sum_{g=1}^G n_g = K$.

- The off-diagonal blocks capture between-block correlations, represented by $\mathbf{C}_{g'g} = \mathbf{C}_{gg'} = \rho_{gg'} \mathbf{1}_{n_g} \mathbf{1}_{n_{g'}}'$ where $g \neq g' \in \{1, \dots, G\}$
- The full block correlation matrix can be expressed as in (8).

$$\mathbf{R}_{block} = \begin{bmatrix} \mathbf{R}_{eq,1} & \mathbf{C}_{12} & \cdots & \mathbf{C}_{1G} \\ \mathbf{C}_{21} & \mathbf{R}_{eq,2} & \cdots & \mathbf{C}_{2G} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}_{G1} & \mathbf{C}_{G2} & \cdots & \mathbf{R}_{eq,G} \end{bmatrix}_{K \times K} \quad (8)$$

Correlation Structures: Distance-based correlation

- Useful in the mean travel time from the study of Klein et al. (2020).

Correlation Structures: Distance-based correlation

- Useful in the mean travel time from the study of Klein et al. (2020).
- Spatial dependence can be modeled using a stationary Matérn correlation function, which for two locations \mathbf{s}_i and \mathbf{s}_j is expressed as

$$\rho_{\text{matern}} = \frac{2^{1-\nu}}{\Gamma(\nu)} (\kappa \|\mathbf{s}_i - \mathbf{s}_j\|)^\nu K_\nu(\kappa \|\mathbf{s}_i - \mathbf{s}_j\|)$$

where $\|\cdot\|$ denotes the Euclidean distance and K_ν is the second kind of the modified Bessel function. It has a scale parameter $\kappa > 0$ and a smoothness parameter $\nu > 0$. ρ_{matern} reduces to the exponential correlation when $\nu = 0.5$ and to Gaussian correlation function when $\nu = \infty$.

Code blocks

Alert block

$$E = mc^2$$

Examples

Example blocks are automatically green in color

Blue block

- happens with level 2, 3 headings
- this is only true for 'Madrid' theme in R Markdown!!

This works, incremental bullets

- Bullet 1

This works, incremental bullets

- Bullet 1
- Bullet 2

This nests, but does not increment

- Bullet 1
- Bullet 2
 - subbullet 1
 - subbullet 2

This increments

- Bullet 1

This increments

- Bullet 1
- Bullet 2

This increments

- Bullet 1
- Bullet 2
 - subbullet 1

This increments

- Bullet 1
- Bullet 2
 - subbullet 1
 - subbullet 2

This increments too

- Bullet 1

This increments too

- Bullet 1
- Bullet 2

This increments too

- Bullet 1
- Bullet 2
 - subbullet 1

This increments too

- Bullet 1
- Bullet 2
 - subbullet 1
 - subbullet 2

Algo Font Size Adjustments sample

- 1: **for** $b = 1, 2, \dots, B$ **do**
- 2: Generate $\hat{\theta}_b^* = (\hat{\theta}_{b1}^*, \hat{\theta}_{b2}^*, \dots, \hat{\theta}_{bK}^*)' \sim N_K(\hat{\theta}, \Sigma)$ and let $\hat{\theta}_{b(1)}^*, \hat{\theta}_{b(2)}^*, \dots, \hat{\theta}_{b(K)}^*$ be the corresponding ordered values
- 3: Compute
$$\hat{\sigma}_{b(k)}^* = \frac{\hat{\theta}_{b(k)}^*}{\sqrt{[\text{kth ordered value among } \{\hat{\theta}_{b1}^{*2} + \sigma_1^2, \dots, \hat{\theta}_{bK}^{*2} + \sigma_K^2\}] - \hat{\theta}_{(k)}^{*2}}}$$
- 4: Compute

$$t_b^* = \max_{1 \leq k \leq K} \left| \frac{\hat{\theta}_{b(k)}^* - \hat{\theta}_k^*}{\hat{\sigma}_{b(k)}^*} \right|$$

- 5: **end for**
- 6: Compute the $(1 - \alpha)$ -sample quantile of t_1^*, \dots, t_B^* , call this \hat{t} .
- 7: The joint confidence region of $\theta_{(1)}, \theta_{(2)}, \dots, \theta_{(K)}$ is given by
$$\mathfrak{R}_2 = [\hat{\theta}_{(1)} \pm \hat{t} \times \hat{\sigma}_{(1)}] \times [\hat{\theta}_{(2)} \pm \hat{t} \times \hat{\sigma}_{(2)}] \times \dots \times [\hat{\theta}_{(K)} \pm \hat{t} \times \hat{\sigma}_{(K)}]$$

where

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

Today is going to be a great day

Today is...

A new beginning.

Your Reference

- slideshare
- themes
- incremental bullets

Klein, M., Wright, T., & Wieczorek, J. (2020). A joint confidence region for an overall ranking of populations. *Journal of the Royal Statistical Society*, 589–606.