# Joint Confidence Regions for Rankings based on Correlated Estimates

Matala, Shaine Rosewel

University of the Philippines

November 26, 2025



2025-11-26

Joint Confidence Regions for Rankings

Joint Confidence Regions for Rankings based on Correlated Estimates

Matala, Shaine Rosewel

University of the Philippines

November 26, 2025

# Background of the Study

- In the problem of estimating ranks of several unknown real-valued parameters $\theta_1, \theta_2, \ldots, \theta_K$, it is desired to rank these $K$ parameters from smallest to largest, $\theta_{(1)} < \theta_{(2)} < \ldots < \theta_{(K)}$.





Background of the Study

- In the problem of estimating ranks of several unknown real-valued parameters $\theta_1, \theta_2, \ldots, \theta_K$, it is desired to rank these $K$ parameters from smallest to largest, $\theta_{(1)} < \theta_{(2)} < \ldots < \theta_{(K)}$.

# Background of the Study

- In the problem of estimating ranks of several unknown real-valued parameters $\theta_1, \theta_2, \ldots, \theta_K$, it is desired to rank these $K$ parameters from smallest to largest, $\theta_{(1)} < \theta_{(2)} < \ldots < \theta_{(K)}$.
- Given estimates in a table without an explicit ranking, readers tend to compare units by looking for smallest or largest estimates and for relative standings between the units. Such tables as motivate "implicit" rankings.





Background of the Study

- In the problem of estimating ranks of several unknown real-valued parameters $\theta_1, \theta_2, \ldots, \theta_K$, it is desired to rank these $K$ parameters from smallest to largest, $\theta_{(1)} < \theta_{(2)} < \ldots < \theta_{(K)}$.
- Given estimates in a table without an explicit ranking, readers tend to compare units by looking for smallest or largest estimates and for relative standings between the units. Such tables as motivate "implicit" rankings.

# Background of the Study

- In the problem of estimating ranks of several unknown real-valued parameters $\theta_1, \theta_2, \ldots, \theta_K$, it is desired to rank these $K$ parameters from smallest to largest, $\theta_{(1)} < \theta_{(2)} < \ldots < \theta_{(K)}$.
- Given estimates in a table without an explicit ranking, readers tend to compare units by looking for smallest or largest estimates and for relative standings between the units. Such tables as motivate "implicit" rankings.
- Because rankings based on the observed values of $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K$ can vary because of sampling variability, widely understood statements of uncertainty should accompany each released ranking.





Background of the Study

- In the problem of estimating ranks of several unknown real-valued parameters $\theta_1, \theta_2, \ldots, \theta_K$, it is desired to rank these $K$ parameters from smallest to largest, $\theta_{(1)} < \theta_{(2)} < \ldots < \theta_{(K)}$.
- Given estimates in a table without an explicit ranking, readers tend to compare units by looking for smallest or largest estimates and for relative standings between the units. Such tables as motivate "implicit" rankings.
- Because rankings based on the observed values of $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K$ can vary because of sampling variability, widely understood statements of uncertainty should accompany each released ranking.

# Background of the Study

- In the problem of estimating ranks of several unknown real-valued parameters $\theta_1, \theta_2, \ldots, \theta_K$, it is desired to rank these $K$ parameters from smallest to largest, $\theta_{(1)} < \theta_{(2)} < \ldots < \theta_{(K)}$.
- Given estimates in a table without an explicit ranking, readers tend to compare units by looking for smallest or largest estimates and for relative standings between the units. Such tables as motivate "implicit" rankings.
- Because rankings based on the observed values of $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K$ can vary because of sampling variability, widely understood statements of uncertainty should accompany each released ranking.
- The margin of error gives uncertainty in the estimate $\hat{\theta}_k$ for each unit separately





Background of the Study
- In the problem of estimating ranks of several unknown real-valued parameters $\theta_1, \theta_2, \ldots, \theta_K$, it is desired to rank these $K$ parameters from smallest to largest, $\theta_{(1)} < \theta_{(2)} < \ldots < \theta_{(K)}$.
- Given estimates in a table without an explicit ranking, readers tend to compare units by looking for smallest or largest estimates and for relative standings between the units. Such tables as motivate "implicit" rankings.
- Because rankings based on the observed values of $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K$ can vary because of sampling variability, widely understood statements of uncertainty should accompany each released ranking.
- The margin of error gives uncertainty in the estimate $\hat{\theta}_k$ for each unit separately

# Background of the Study

- In the problem of estimating ranks of several unknown real-valued parameters $\theta_1, \theta_2, \ldots, \theta_K$, it is desired to rank these $K$ parameters from smallest to largest, $\theta_{(1)} < \theta_{(2)} < \ldots < \theta_{(K)}$.
- Given estimates in a table without an explicit ranking, readers tend to compare units by looking for smallest or largest estimates and for relative standings between the units. Such tables as motivate “implicit” rankings.
- Because rankings based on the observed values of $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K$ can vary because of sampling variability, widely understood statements of uncertainty should accompany each released ranking.
- The margin of error gives uncertainty in the estimate $\hat{\theta}_k$ for each unit separately
- A direct assessment of the uncertainty in the estimated overall ranking would simultaneously involve all units and their relative standing to each other



2025-11-26

Joint Confidence Regions for Rankings

└─Background of the Study

Background of the Study

- In the problem of estimating ranks of several unknown real-valued parameters $\theta_1, \theta_2, \ldots, \theta_K$, it is desired to rank these $K$ parameters from smallest to largest, $\theta_{(1)} < \theta_{(2)} < \ldots < \theta_{(K)}$.
- Given estimates in a table without an explicit ranking, readers tend to compare units by looking for smallest or largest estimates and for relative standings between the units. Such tables as motivate “implicit” rankings.
- Because rankings based on the observed values of $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K$ can vary because of sampling variability, widely understood statements of uncertainty should accompany each released ranking.
- The margin of error gives uncertainty in the estimate $\hat{\theta}_k$ for each unit separately
- A direct assessment of the uncertainty in the estimated overall ranking would simultaneously involve all units and their relative standing to each other

# Background of the Study

- Let $r_1, r_2, \ldots, r_K$ be the true unknown ranks of $\theta_1, \theta_2, \ldots, \theta_K$. A mathematical definition of $r_k$ is as follows:

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





Background of the Study

- Let $r_1, r_2, \ldots, r_K$ be the true unknown ranks of $\theta_1, \theta_2, \ldots, \theta_K$. A mathematical definition of $r_k$ is as follows:

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

# Background of the Study

- Let $r_1, r_2, \ldots, r_K$ be the true unknown ranks of $\theta_1, \theta_2, \ldots, \theta_K$. A mathematical definition of $r_k$ is as follows:

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

- Goal





Background of the Study

- Let $r_1, r_2, \ldots, r_K$ be the true unknown ranks of $\theta_1, \theta_2, \ldots, \theta_K$. A mathematical definition of $r_k$ is as follows:

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

- Goal

# Background of the Study

- Let $r_1, r_2, \ldots, r_K$ be the true unknown ranks of $\theta_1, \theta_2, \ldots, \theta_K$. A mathematical definition of $r_k$ is as follows:

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

- Goal
  1. derive a joint confidence region for $\theta_1, \theta_2, \ldots, \theta_K$





Background of the Study

- Let $r_1, r_2, \ldots, r_K$ be the true unknown ranks of $\theta_1, \theta_2, \ldots, \theta_K$. A mathematical definition of $r_k$ is as follows:

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

- Goal
  1. derive a joint confidence region for $\theta_1, \theta_2, \ldots, \theta_K$

# Background of the Study

- Let $r_1, r_2, \ldots, r_K$ be the true unknown ranks of $\theta_1, \theta_2, \ldots, \theta_K$. A mathematical definition of $r_k$ is as follows:

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

- Goal
  1. derive a joint confidence region for $\theta_1, \theta_2, \ldots, \theta_K$
  2. obtain joint confidence intervals for the $r_1, r_2, \ldots, r_K$ using a result from Klein et al. (2020)).





Background of the Study

- Let $r_1, r_2, \ldots, r_K$ be the true unknown ranks of $\theta_1, \theta_2, \ldots, \theta_K$. A mathematical definition of $r_k$ is as follows:

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

- Goal
  1. derive a joint confidence region for $\theta_1, \theta_2, \ldots, \theta_K$
  2. obtain joint confidence intervals for the $r_1, r_2, \ldots, r_K$ using a result from Klein et al. (2020)).

# Klein's

Suppose that for each $k \in \{1, 2, \ldots, K\}$ there exists values $L_k$ and $U_k$ ST

$$\theta_k \in (L_k, U_k), k = 1, 2, \ldots, K. \tag{2}$$

If the condition in (2) holds, the main result from Klein et al. (2020) gives a range for the value of $r_k$ for each $k \in \{1, 2, \ldots, K\}$ as follows:

$$r_k \in \{|\Lambda_{Lk}| + 1, |\Lambda_{Lk}| + 2, |\Lambda_{Lk}| + 3, \ldots, |\Lambda_{Lk}| + |\Lambda_{Ok}| + 1\} \tag{3}$$





Klein's

Suppose that for each $k \in \{1, 2, \ldots, K\}$ there exists values $L_k$ and $U_k$ ST

$$\theta_k \in (L_k, U_k), k = 1, 2, \ldots, K. \tag{2}$$

If the condition in (2) holds, the main result from Klein et al. (2020) gives a range for the value of $r_k$ for each $k \in \{1, 2, \ldots, K\}$ as follows:

$$r_k \in \{|\Lambda_{Lk}| + 1, |\Lambda_{Lk}| + 2, |\Lambda_{Lk}| + 3, \ldots, |\Lambda_{Lk}| + |\Lambda_{Ok}| + 1\} \tag{3}$$

# Klein's

Suppose that for random quantities $L_k$ and $U_k$ the event defined in (2) satisfies the following probability condition:

$$P\left[\bigcap_{k=1}^{K}\{\theta_k \in (L_k, U_k)\}\right] \geq 1-\alpha, \tag{4}$$

then, by the result of Klein et al. (2020), it also follows that

$$P\left[\bigcap_{k=1}^{K}\{r_k \in \{|\Lambda_{Lk}|+1, |\Lambda_{Lk}|+2, \ldots, |\Lambda_{Lk}|+|\Lambda_{Ok}|+1\}\}\right] \geq 1-\alpha. \tag{5}$$





Klein's
Suppose that for random quantities $L_k$ and $U_k$ the event defined in (2) satisfies the following probability condition:
$$P\left[\bigcap_{k=1}^{K}\{\theta_k \in (L_k, U_k)\}\right] \geq 1-\alpha, \tag{4}$$
then, by the result of Klein et al. (2020), it also follows that
$$P\left[\bigcap_{k=1}^{K}\{r_k \in \{|\Lambda_{Lk}|+1, |\Lambda_{Lk}|+2, \ldots, |\Lambda_{Lk}|+|\Lambda_{Ok}|+1\}\}\right] \geq 1-\alpha. \tag{5}$$

# Other related studies





Other related studies

# Motivation

- Problem: 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.





Motivation

- Problem: 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.

# Motivation

- Problem: 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.
- Aim: develop a procedure capable of handling such dependencies while maintaining coverage close to the nominal level and producing relatively narrow joint confidence intervals.





Motivation

- Problem: 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.
- Aim: develop a procedure capable of handling such dependencies while maintaining coverage close to the nominal level and producing relatively narrow joint confidence intervals.

# Motivation

## Political setting — David & Legara (2015)

- Name recall is a powerful predictor of likely victory in elections.





Motivation
Political setting — David & Legara (2015)
- Name recall is a powerful predictor of likely victory in elections.

- 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
- top-ranked candidates is composed of people who can take the most advantage of name recall: All belong to at least one of the following types: media celebrity, member of political dynasty, or had prior experience in the Senate (labeled henceforth Celebrities and Dynasties). Of the eight candidates, three are former movie and television actors, three are offspring of former presidents and senators, and six had prior experience in the Senate. They come from different political parties and different tickets
- This set of candidates was aggressively campaigned alongside candidate for President Aquino, who was popular throughout the election season and eventually won by a 12

# Motivation

**Political setting — David & Legara (2015)**

- Name recall is a powerful predictor of likely victory in 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.





Motivation

Political setting — David & Legara (2015)

- Name recall is a powerful predictor of likely victory in 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.

- 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
- top-ranked candidates is composed of people who can take the most advantage of name recall: All belong to at least one of the following types: media celebrity, member of political dynasty, or had prior experience in the Senate (labeled henceforth Celebrities and Dynasties). Of the eight candidates, three are former movie and television actors, three are offspring of former presidents and senators, and six had prior experience in the Senate. They come from different political parties and different tickets
- This set of candidates was aggressively campaigned alongside candidate for President Aquino, who was popular throughout the election season and eventually won by a 12

# Motivation

**Measurement across geographies — Klein et al. (2020)**

| Travel Time | Population Density | Common Locations |
|---|---|---|
| Shorter mean | Large unpopulated land areas; fewer high-density population centers | Mountain region and Central region states |
| Longer mean | Highly urbanized areas with large populations and dense population centers | East Coast states |





Motivation
Measurement across geographies — Klein et al. (2020)
Travel Time | Population Density | Common Locations
Shorter mean | Large unpopulated land areas; fewer high-density population centers | Mountain region and Central region states
Longer mean | Highly urbanized areas with large populations and dense population centers | East Coast states

- @klein 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.
- 2019—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.





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.

# 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.





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.

# 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.





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 $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K)'$ and assume that $\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_K)'$ is unknown and $\boldsymbol{\Sigma}$ is a known $K \times K$ positive definite matrix. The diagonal elements of $\boldsymbol{\Sigma}$ are $\sigma_1^2, \ldots, \sigma_K^2$.





Definitions and Assumptions

- Define $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K)'$ and assume that $\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_K)'$ is unknown and $\boldsymbol{\Sigma}$ is a known $K \times K$ positive definite matrix. The diagonal elements of $\boldsymbol{\Sigma}$ are $\sigma_1^2, \ldots, \sigma_K^2$.

# Definitions and Assumptions

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





Definitions and Assumptions

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

# Procedure

1. Derive simultaneous confidence intervals for $\theta_1, \theta_2, \ldots, \theta_K$ of the form

$$\mathfrak{R}_1 = [\hat{\theta}_1 \pm t \times \sigma_1] \times [\hat{\theta}_2 \pm t \times \sigma_2] \times \cdots \times [\hat{\theta}_K \pm t \times \sigma_K]. \quad (6)$$





Procedure

1. Derive simultaneous confidence intervals for $\theta_1, \theta_2, \ldots, \theta_K$ of the form

$$\mathfrak{R}_1 = [\hat{\theta}_1 \pm t \times \sigma_1] \times [\hat{\theta}_2 \pm t \times \sigma_2] \times \cdots \times [\hat{\theta}_K \pm t \times \sigma_K]. \quad (6)$$

# Procedure

1. Derive simultaneous confidence intervals for $\theta_1, \theta_2, \ldots, \theta_K$ of the form

$$\mathfrak{R}_1 = [\hat{\theta}_1 \pm t \times \sigma_1] \times [\hat{\theta}_2 \pm t \times \sigma_2] \times \cdots \times [\hat{\theta}_K \pm t \times \sigma_K]. \qquad (6)$$

2. Once the confidence intervals in (6) have been obtained, we can then use the result of Klein et al. (2020) in (5) to get the lower and upper bounds on the ranks $r_k, k = 1, 2, \ldots, K$. That is, we also get a joint confidence region for $r_1, r_2, ..., r_K$.





Procedure

1. Derive simultaneous confidence intervals for $\theta_1, \theta_2, \ldots, \theta_K$ of the form

$$\mathfrak{R}_1 = [\hat{\theta}_1 \pm t \times \sigma_1] \times [\hat{\theta}_2 \pm t \times \sigma_2] \times \cdots \times [\hat{\theta}_K \pm t \times \sigma_K]. \qquad (6)$$

2. Once the confidence intervals in (6) have been obtained, we can then use the result of Klein et al. (2020) in (5) to get the lower and upper bounds on the ranks $r_k, k = 1, 2, \ldots, K$. That is, we also get a joint confidence region for $r_1, r_2, ..., r_K$.

# Proposed methodology to compute the joint confidence region for the unordered parameters: Algorithm 1

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

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

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

3: Compute

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

4: **end for**

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

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

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




Proposed methodology to compute the joint confidence region for the unordered parameters: Algorithm 1
Proposed methodology to compute the joint confidence region for the unordered parameters: Algorithm 1

# Algorithm 1: Quantile Calculation

We want the joint confidence region in (6) to satisfy the following probability condition:

$$P\left(\hat{\theta}_k - t \cdot \sigma_k \leq \theta_k \leq \hat{\theta}_k + t \cdot \sigma_k,\ \forall\ k = 1, 2, \ldots, K\right) = 1 - \alpha. \tag{7}$$

Equivalently, we require

$$P\left(\max_{k=1,2,\ldots,K} \left|\frac{\hat{\theta}_k - \theta_k}{\sigma_k}\right| \leq t\right) = 1 - \alpha. \tag{8}$$





Algorithm 1: Quantile Calculation

We want the joint confidence region in (6) to satisfy the following probability condition:

$$P\left(\hat{\theta}_k - t \cdot \sigma_k \leq \theta_k \leq \hat{\theta}_k + t \cdot \sigma_k,\ \forall\ k = 1, 2, \ldots, K\right) = 1 - \alpha. \tag{7}$$

Equivalently, we require

$$P\left(\max_{k=1,2,\ldots,K} \left|\frac{\hat{\theta}_k - \theta_k}{\sigma_k}\right| \leq t\right) = 1 - \alpha. \tag{8}$$

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

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

2: Generate $\hat{\boldsymbol{\theta}}_b^* = \left(\hat{\theta}_{b1}^*, \hat{\theta}_{b2}^*, \ldots, \hat{\theta}_{bK}^*\right)' \sim N_K\left(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}\right)$ and let $\hat{\theta}_{b(1)}^*, \hat{\theta}_{b(2)}^*, \ldots, \hat{\theta}_{b(K)}^*$ be the corresponding ordered values

3: Compute $\hat{\sigma}_{b(k)}^*$ using:

- asymptotic variance definition
- second-level bootstrap

4: Compute

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

5: **end for**

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

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

$$\mathfrak{R}_2 = \left[\hat{\theta}_{(1)} \pm \hat{t} \times \hat{\sigma}_{(1)}\right] \times \left[\hat{\theta}_{(2)} \pm \hat{t} \times \hat{\sigma}_{(2)}\right] \times \cdots \times \left[\hat{\theta}_{(K)} \pm \hat{t} \times \hat{\sigma}_{(K)}\right]$$

# Algorithm 2: Asymptotic Definition of Variance

$$\hat{\sigma}^*_{b(k)} = \sqrt{\left[\text{kth ordered value among } \left\{\hat{\theta}^{*2}_{b1} + \sigma^2_1, \ldots, \hat{\theta}^{*2}_{bK} + \sigma^2_K\right\}\right] - \hat{\theta}^{*2}_{(k)}}$$





Algorithm 2: Asymptotic Definition of Variance

$$\hat{\sigma}^*_{b(k)} = \sqrt{\left[\text{kth ordered value among } \left\{\hat{\theta}^{*2}_{b1} + \sigma^2_1, \ldots, \hat{\theta}^{*2}_{bK} + \sigma^2_K\right\}\right] - \hat{\theta}^{*2}_{(k)}}$$

# Algorithm 3: Variance from Second-Level Bootstrap

**for** $c = 1, 2, \ldots, C$ **do**

  Generate $\hat{\boldsymbol{\theta}}^{**}_{bc} = \left(\hat{\theta}^{**}_{bc1}, \hat{\theta}^{**}_{bc2}, \ldots, \hat{\theta}^{**}_{bcK}\right) \sim N_K\left(\hat{\boldsymbol{\theta}}^{*}_{b}, \boldsymbol{\Sigma}\right)$ and let $\hat{\theta}^{**}_{bc(1)}, \hat{\theta}^{**}_{bc(2)}, \ldots, \hat{\theta}^{**}_{bc(K)}$ be the corresponding ordered values of $\hat{\theta}^{**}_{bc1}, \hat{\theta}^{**}_{bc2}, \ldots, \hat{\theta}^{**}_{bcK}$

  Compute $\hat{\sigma}^{*}_{b(k)} = \dfrac{\sum_{c=1}^{C}\left(\hat{\theta}^{**}_{bc(k)} - \bar{\hat{\theta}}^{**}_{b\cdot(k)}\right)^2}{C-1}, \quad \bar{\hat{\theta}}^{**}_{b\cdot(k)} = \dfrac{1}{C}\sum_{c=1}^{C}\hat{\theta}^{**}_{bc(k)}$

**end for**





Algorithm 3: Variance from Second-Level Bootstrap
1: for $c = 1, 2, \ldots, C$ do
2: Generate $\hat{\boldsymbol{\theta}}^{**}_{bc} = \left(\hat{\theta}^{**}_{bc1}, \hat{\theta}^{**}_{bc2}, \ldots, \hat{\theta}^{**}_{bcK}\right) \sim N_K\left(\hat{\boldsymbol{\theta}}^{*}_{b}, \boldsymbol{\Sigma}\right)$ and let $\hat{\theta}^{**}_{bc(1)}, \hat{\theta}^{**}_{bc(2)}, \ldots, \hat{\theta}^{**}_{bc(K)}$ be the corresponding ordered values of $\hat{\theta}^{**}_{bc1}, \hat{\theta}^{**}_{bc2}, \ldots, \hat{\theta}^{**}_{bcK}$
3: Compute $\hat{\sigma}^{*}_{b(k)} = \dfrac{\sum_{c=1}^{C}\left(\hat{\theta}^{**}_{bc(k)} - \bar{\hat{\theta}}^{**}_{b\cdot(k)}\right)^2}{C-1}, \quad \bar{\hat{\theta}}^{**}_{b\cdot(k)} = \dfrac{1}{C}\sum_{c=1}^{C}\hat{\theta}^{**}_{bc(k)}$
4: end for

# Evaluation Algorithm

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

1: **for** replications $= 1, 2, \ldots, 5000$ **do**
2: Generate $\hat{\boldsymbol{\theta}} \sim N_K(\boldsymbol{\theta}, \mathbf{\Sigma})$
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 $(\theta_1, \theta_2, \ldots, \theta_K) \in \mathfrak{R}_1$ and compute $T_1$, $T_2$, and $T_3$. For the ordered parameters, check if $\left(\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(K)}\right) \in \mathfrak{R}_2$
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 $T_3$.



2025-11-26
Joint Confidence Regions for Rankings
└Evaluation Algorithm

# Covariance Matrix $\boldsymbol{\Sigma}$

- The covariance matrix $\boldsymbol{\Sigma}$ need not be a diagonal matrix.
- We assume that $V(\hat{\boldsymbol{\theta}}) = \boldsymbol{\Sigma}$ is known and express $\boldsymbol{\Sigma}$ as in (9), where $\mathbf{R}$ is the population correlation matrix.

$$\boldsymbol{\Sigma} = \boldsymbol{\Delta}^{1/2}\mathbf{R}\boldsymbol{\Delta}^{1/2}; \quad \boldsymbol{\Delta} = \operatorname{diag}\left\{\sigma_1^2, \sigma_2^2, \ldots, \sigma_K^2\right\}. \tag{9}$$

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





Covariance Matrix $\boldsymbol{\Sigma}$

- The covariance matrix $\boldsymbol{\Sigma}$ need not be a diagonal matrix.
- We assume that $V(\hat{\boldsymbol{\theta}}) = \boldsymbol{\Sigma}$ is known and express $\boldsymbol{\Sigma}$ as in (9), where $\mathbf{R}$ is the population correlation matrix.

$$\boldsymbol{\Sigma} = \boldsymbol{\Delta}^{1/2}\mathbf{R}\boldsymbol{\Delta}^{1/2}; \quad \boldsymbol{\Delta} = \operatorname{diag}\left\{\sigma_1^2, \sigma_2^2, \ldots, \sigma_K^2\right\}. \tag{9}$$

- The diagonal elements of $\boldsymbol{\Sigma}$, which are $\sigma_k^2 = V(\hat{\theta}_k)$ for $k = 1, 2, \ldots, K$, 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.

# 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.

# 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 $j \neq k \in \{1, \ldots, K\}$.

$$\mathbf{R}_{\text{eq}} = (1-\rho)\,\mathbf{I}_K + \rho \mathbf{1}_K \mathbf{1}_K' = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}_{K \times K} \tag{10}$$





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 $j \neq k \in \{1, \ldots, K\}$.

$$\mathbf{R}_{\text{eq}} = (1-\rho)\,\mathbf{I}_K + \rho \mathbf{1}_K \mathbf{1}_K' = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}_{K \times K} \tag{10}$$

# Correlation Structures: Block correlation

- Useful in the context of pre-election surveys



Joint Confidence Regions for Rankings
2025-11-26
Correlation Structures: Block correlation
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

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

where $\rho_g$ 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

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

where $\rho_g$ 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

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

  where $\rho_g$ 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 $g \neq g' \in \{1, \ldots, G\}$





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
$$\mathbf{R}_{\text{eq,g}} = (1 - \rho_g)\,\mathbf{I}_{n_g} + \rho_g \mathbf{1}_{n_g} \mathbf{1}'_{n_g}$$
where $\rho_g$ 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 $g \neq g' \in \{1, \ldots, G\}$

## 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

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

  where $\rho_g$ 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 $g \neq g' \in \{1, \ldots, G\}$
- The full block correlation matrix can be expressed as in (11).

$$\mathbf{R}_{\text{block}} = \begin{bmatrix} \mathbf{R}_{eq,1} & \mathbf{C}_{12} & \cdots & \mathbf{C}_{1G} \\ \mathbf{C}_{21} & \mathbf{R}_{eq,2} & \cdots & \mathbf{C}_{2G} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}_{G1} & \mathbf{C}_{G2} & \cdots & \mathbf{R}_{eq,G} \end{bmatrix}_{K\times K} \tag{11}$$





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
$$\mathbf{R}_{\text{eq,g}} = (1 - \rho_g)\,\mathbf{I}_{n_g} + \rho_g \mathbf{1}_{n_g}\mathbf{1}'_{n_g}$$
where $\rho_g$ 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 $g \neq g' \in \{1, \ldots, G\}$
- The full block correlation matrix can be expressed as in (11).
$$\mathbf{R}_{\text{block}} = \begin{bmatrix} \mathbf{R}_{eq,1} & \mathbf{C}_{12} & \cdots & \mathbf{C}_{1G} \\ \mathbf{C}_{21} & \mathbf{R}_{eq,2} & \cdots & \mathbf{C}_{2G} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}_{G1} & \mathbf{C}_{G2} & \cdots & \mathbf{R}_{eq,G} \end{bmatrix}_{K\times K} \tag{11}$$

# 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).

# 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

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

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





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

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

where $\parallel \cdot \parallel$ denotes the Euclidean distance and $K_{\nu}$ is the second kind of the modified Bessel function. It has a scale parameter $\kappa > 0$ and a smoothness parameter $\nu > 0$. $\rho_{\text{matern}}$ 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!!



Joint Confidence Regions for Rankings
2025-11-26
└─Code blocks

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

# This works, incremental bullets

- Bullet 1
- Bullet 2





This works, incremental bullets
- Bullet 1
- Bullet 2

# This nests, but does not increment

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





This nests, but does not increment

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

# This increments

- Bullet 1





This increments

- Bullet 1

# This increments

- Bullet 1
- Bullet 2





This increments
- Bullet 1
- Bullet 2

# This increments

- Bullet 1
- Bullet 2
    - subbullet 1





This increments

- Bullet 1
- Bullet 2
    - subbullet 1

# This increments

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





This increments

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

# This increments too

- Bullet 1





This increments too
Bullet 1

# This increments too

- Bullet 1
- Bullet 2



2025-11-26



└─This increments too

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

# This increments too

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





This increments too

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

# Algo Font Size Adjustments sample

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

Generate $\hat{\boldsymbol{\theta}}_b^* = \left(\hat{\theta}_{b1}^*, \hat{\theta}_{b2}^*, \ldots, \hat{\theta}_{bK}^*\right)' \sim N_K\left(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}\right)$ and let $\hat{\theta}_{b(1)}^*, \hat{\theta}_{b(2)}^*, \ldots, \hat{\theta}_{b(K)}^*$ be the corresponding ordered values

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

Compute

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

**end for**

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

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

$$\mathfrak{R}_2 = \left[\hat{\theta}_{(1)} \pm \hat{t} \times \hat{\sigma}_{(1)}\right] \times \left[\hat{\theta}_{(2)} \pm \hat{t} \times \hat{\sigma}_{(2)}\right] \times \cdots \times \left[\hat{\theta}_{(K)} \pm \hat{t} \times \hat{\sigma}_{(K)}\right]$$

where

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

# Text

Today is going to be a great day

**Today is. . .**

A new beginning.

**Your Reference**

- slideshare
- themes
- incremental bullets

David, C., & Legara, E. F. (2015). *How voters combine candidates on the ballot: The case of the philippine senatorial elections.*

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.





Text

Today is going to be a great day

Today is. . .

A new beginning.

Your Reference

- slideshare
- themes
- incremental bullets

David, C., & Legara, E. F. (2015). *How voters combine candidates on the ballot: The case of the philippine senatorial elections.*

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.