Adaptive Pairwise Comparison for Educational Measurement

Elise Crompvoets\textsuperscript{1,2} & Anton Béguin\textsuperscript{2}

\textsuperscript{1}Tilburg University, Netherlands
\textsuperscript{2}Cito, Netherlands

E-mail: E.A.V.Crompvoets@uvt.nl / elisecrompvoets@cito.nl
What is pairwise comparison
What is pairwise comparison
What is pairwise comparison

Thurstone’s model (Thurstone, 1927)

\[ P(i > j|\theta_i, \theta_j) = \phi(\theta_i - \theta_j) \]
Why education needs pairwise comparison
Current issue with pairwise comparison

- 10 objects: 45 comparisons
- 20 objects: 190 comparisons
Current issue with pairwise comparison

10 objects

20 objects
Our proposition

Adaptive Bayesian algorithm to let judges make the most informative comparison
Our proposition

Adaptive Bayesian algorithm to let judges make the most informative comparison

More information for objects closer together on the attribute scale.
Our proposition

Adaptive Bayesian algorithm to let judges make the most informative comparison

- More information for objects closer together on the attribute scale
- But need to take uncertainty of object location into account
Our proposition

Adaptive Bayesian algorithm to let judges make the most informative comparison

1. Draw attribute values $\theta$ from conditional posterior distribution $\theta_i | \theta, Z, X$
Our proposition

Adaptive Bayesian algorithm to let judges make the most informative comparison

1. Draw attribute values $\theta$ from conditional posterior distribution
2. Draw augmented $Z_{ij}$-values for every combination of objects $i$ and $j$ from the posterior distribution conditional on $\theta$
Our proposition

Adaptive Bayesian algorithm to let judges make the most informative comparison

1. Draw attribute values $\theta$ from conditional posterior distribution
2. Draw augmented $Z_{ij}$-values for every combination of objects $i$ and $j$ from the posterior distribution conditional on $\theta$
3. Compute $|Z_{ij}$-values|$
Our proposition

Adaptive Bayesian algorithm to let judges make the most informative comparison

1. Draw attribute values $\theta$ from conditional posterior distribution
2. Draw augmented $Z_{ij}$-values for every combination of objects $i$ and $j$ from the posterior distribution conditional on $\theta$
3. Compute $|Z_{ij}$-values$|
4. Select the pair of objects with the lowest absolute $Z_{ij}$-value
Our proposition

Adaptive Bayesian algorithm to let judges make the most informative comparison

1. Draw attribute values $\theta$ from conditional posterior distribution
2. Draw augmented $Z_{ij}$-values for every combination of objects $i$ and $j$ from the posterior distribution conditional on $\theta$
3. Compute $|Z_{ij}$-values$|
4. Select the pair of objects with the lowest absolute $Z_{ij}$-value
5. Compare object $i$ with object $j$
Our proposition

Adaptive Bayesian algorithm to let judges make the most informative comparison

1. Draw attribute values $\theta$ from conditional posterior distribution
2. Draw augmented $Z_{ij}$-values for every combination of objects $i$ and $j$ from the posterior distribution conditional on $\theta$
3. Compute $|Z_{ij}$-values$|
4. Select the pair of objects with the lowest absolute $Z_{ij}$-value
5. Compare object $i$ with object $j$
6. Repeat until a stopping criterion is reached
Simulation study setup

**Conditions**

- **Algorithm**
  - {semi-random, adaptive}
- **Number of objects**
  - {20, 30}
- **Number of comparisons per object**
  - {10, 15, 20, 25, 30}
- **Number of judges**
  - {2, 3, 5}
Simulation study setup

Evaluation criteria

- Standard errors
  - \( SE(\hat{\theta}_i) \)
- Accuracy of rank order
  - Spearman’s rank coefficient
- Reliability
  - Benchmark reliability
  - Scale Separation Reliability (SSR)
Results: Standard errors
Results: Spearman’s rank coefficient

![Graphs showing Spearman's rank coefficient for 20 and 30 students.](image)
Results: Reliability
Conclusion

- The BSA shows better results compared to the SSA on all three evaluation criteria
- The SSR coefficient can be trusted as a reliability estimate when the BSA is used

Further research
- Improve BSA for bigger gains
- Investigate the influence of rater agreement on BSA performance
Thank you for your attention

Feedback is welcome

E-mail: E.A.V.Crompvoets@uvt.nl / elisecrompvoets@cito.nl