Assessment of Simple and Alternative Bayesian Ranking Methods Utilizing Parallel Computing

UMBC REU Site: Interdisciplinary Program in High Performance Computing (www.umbc.edu/hpcf)
Team members: Samantha Allen¹, Dorothy Kirlew², Neil Obetz³, Derek Wade⁴
Graduate assistant: April Albertine⁵ Faculty mentor: Dr. Nagaraj Neerchal⁵
Client: Dr. Martin Klein⁶

¹High Point Univ. ²Hood College ³Millersville Univ. ⁴Boise State Univ. ⁵UMBC ⁶U.S. Census Bureau

Problem Statement
The U.S. Census Bureau (USCB) assists the federal government in distributing over $400 billion of aid by ranking the states according to certain criteria, such as average poverty level. The current ranking algorithm is based on sample estimates which are associated with a certain amount of error. Klein and Wright of the USCB have compared the performance of non-informative Bayesian techniques to the USCB’s current method. We expand on this work to add informed Bayesian and regression models to the comparison. By employing moderation techniques, we obtain excellent probabilities of correct rankings.

The Model
Suppose \( x_i \) is sample estimate, and \( \sigma_i \) is a known standard deviation. The population model used is:
\[
x_i | \theta_i \sim \text{N}(\theta_i, \sigma_i^2), \quad i = 1, \ldots, k;
\]
where \( \theta_i \) is an unknown population mean. The current method, SI, sorts the estimates and ranks them accordingly. The alternative Bayesian ranking methods include: P1EB, P2EB, PMEB [Klein and Wright, 2011], and MXPR [Tech Report HPCF–2011–11]. These alternative procedures incorporate the following prior distributions:
- Non-informative:
  \[
  \theta_1, \ldots, \theta_k | \mu, \tau \sim \text{N}(\mu, \tau^2).
  \]
- Regression Informed Prior (RIP):
  \[
  \theta_i = \beta_0 + \beta_1 m_i + \varepsilon_i
  \]
  where \( \beta_0, \beta_1, \varepsilon_i \) are random, \( m_i \) is the previous data.
- Fully Informed Prior (FIP):
  \[
  \theta_i \sim \text{N}(m_i, \tau_i^2)
  \]

Simulation Comparison: Below is a comparison of the SI method and the two versions of the informed P1EB method: FIP and RIP. These informed methods obtain comparable results until \( \theta_{\text{prev}} \) and \( \theta \) have dissimilar orderings. With small shifts in the previous data, FIP does slightly better, but when shifts are dramatic, RIP produces much better probabilities of a correct ranking and consistently outperforms the SI method.

Comparing Various Priors

Parallel Bootstrap
In practice, the accuracy of the estimated ranks is computed by a bootstrap. We implement this computationally expensive procedure in parallel.

Conclusions
- We are most confident in the P1EB method with RIP, which produces higher probabilities of correct rankings, even with varied \( \theta_{\text{prev}} \) settings.
- Utilizing previous data allows us to specify a more realistic model for the prior distribution of the current data.

References
For more information, refer to:
- Tech Report HPCF–2011–11
  www.umbc.edu/hpcreu/2011/projects/team1.html

Acknowledgments: This research was conducted during Summer 2011 in the REU Site: Interdisciplinary Program in High Performance Computing (www.umbc.edu/hpcreu) in the UMBC Department of Mathematics and Statistics, funded by the National Science Foundation (grant no. DMS-0851749). This program is also supported by UMBC, the Department of Mathematics and Statistics, the Center for Interdisciplinary Research and Consulting (CIRC), and the UMBC High Performance Computing Facility (HPCF). The computational hardware in HPCF (http://www.umbc.edu/hpcreu) is partially funded by the National Science Foundation through the MRI program (grant no. CNS-0821258) and the SCREMS program (grant no. DMS-0821311), with additional substantial support from UMBC.