

bayesGT: An R Package for Bayesian Regression Models on Pooled Testing Data

Caleb Frankenberger

Department of Mathematics and Statistics, Radford University

Objectives

The **bayesGT** package was developed to:

- Estimate regression coefficients and their precision using Bayesian models
- Handle a wide range of input, including multiple link functions and any type of pooled data configuration
- Implement parallel computing to significantly reduce program runtime
- Include standard diagnostic tools such as traceplots and histograms

Introduction

Pooled testing is a cost-effective public health method for screening individuals for infectious diseases. By combining samples (e.g., blood, urine, swabs) from multiple people, testing costs are substantially reduced.

Statistical methods have been developed to estimate individual disease probabilities from pooled data using covariates such as age, weight, and symptoms. These methods are challenging because test results are observed for pools, while covariates are measured for individuals.

The **bayesGT** R package implements advanced Bayesian regression models to estimate coefficients in non-linear, binary regression models. Unlike likelihood-based approaches, our Bayesian methods can incorporate historical data to improve precision.

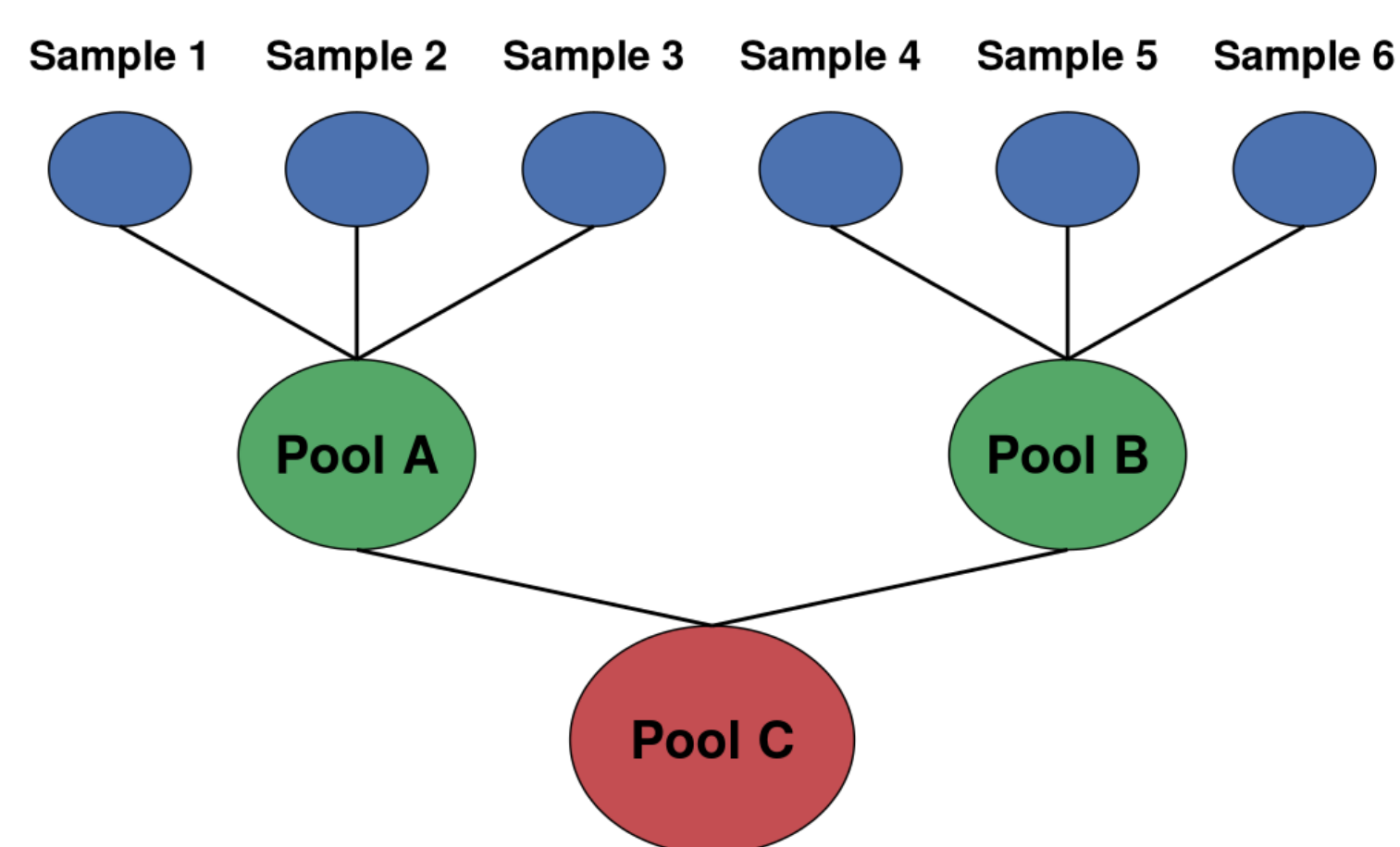


Figure 1: Sample Pooling Diagram (Two-Stage)

Methods

At the core of **bayesGT** is a Gibbs sampler, a Markov Chain Monte Carlo (MCMC) technique. Using a combination of Fortran and R code, we iteratively update parameter estimates in three stages:

- Update the latent individual disease statuses (Y_i) based on pool results (Z)
- Sample the regression coefficients (β) using a Metropolis-Hastings algorithm
- Sample test sensitivity (Se) and specificity (Sp) from their full conditional Beta distributions (if not assumed known)

This process is repeated for thousands of iterations to generate posterior distributions for inference.

Method Validation

To ensure our estimations are reliable, we validated the method against data simulated using the **groupTesting** R package. This approach is not possible with real-world data, where the true parameter values are never known.

The posterior estimates show excellent agreement with the true parameter values used in the simulations. Coverage probabilities are close to the nominal 95% level, and the estimated and empirical standard deviations align closely, demonstrating both accuracy and reliability of the method. Test accuracies are recovered almost exactly, confirming that the model effectively estimates unknown Se/Sp values in a hierarchical pooled testing design.

Simulation Settings: The study used 500 replicates, each with $N = 5000$ individuals. For each replicate, 6000 Gibbs samples were drawn, with 1000 discarded as burn-in. The total runtime for all replicates was ≈ 3 minutes (8.2s / replicate). This simulation used a non-adaptive random walk Metropolis-Hastings—the simplest option in the package—which still produced excellent estimates. Other available MH options include weighted least squares and adaptive random walk.

Optimized Performance

Throughout development, code sections were profiled to identify bottlenecks. Computationally intensive parts were optimized, and parallel computing was implemented to run replicates simultaneously across multiple threads. Overall, **bayesGT** achieved up to 12 \times faster runtimes compared to previous implementations.

Diagnostics

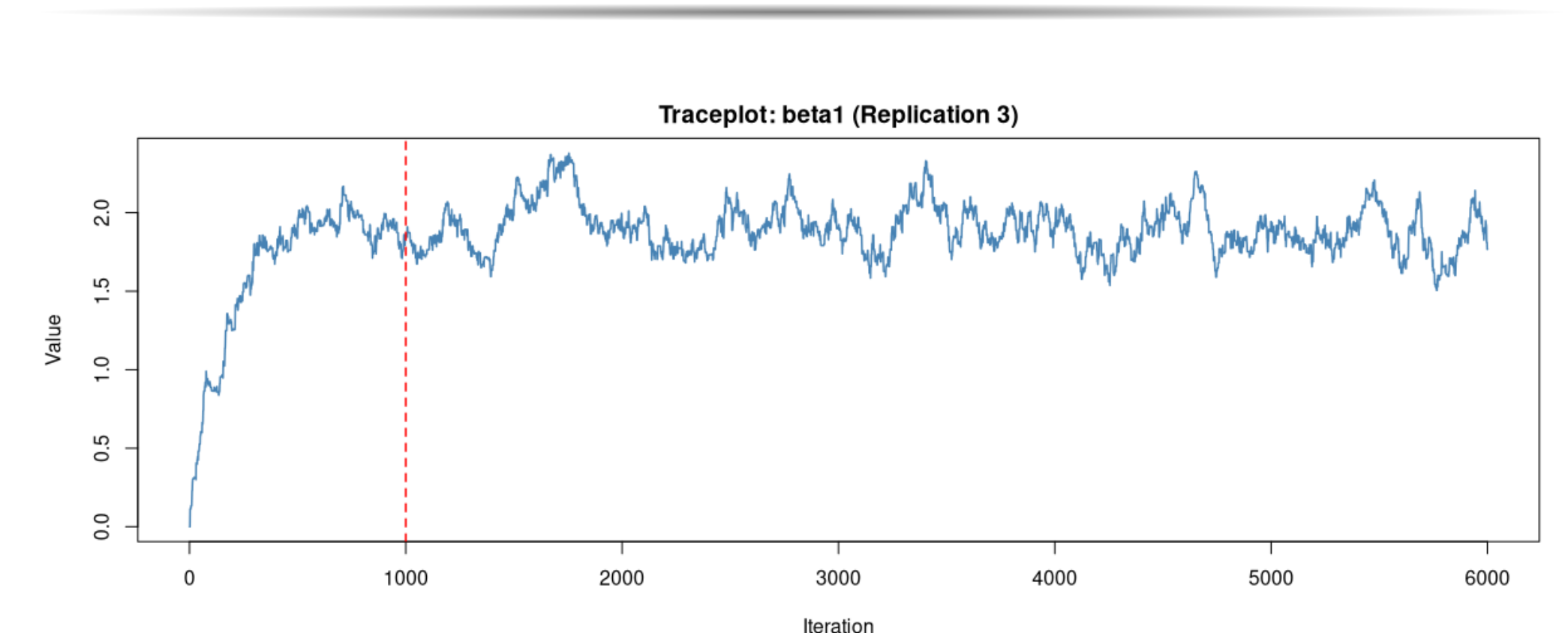


Figure 2: Traceplot for β_1 (replicate 3)

Posterior samples converge quickly post burn-in, indicating reliable estimation of β_1 . Additional diagnostics, such as posterior histograms and ACF plots, are available to further assess mixing and convergence.

References

- Christopher S. McMahan, Joshua M. Tebbs, Timothy E. Hanson, and Christopher R. Bilder. Bayesian regression for group testing data. *Biometrics*, 73(4):1443–1452, 2017. <https://doi.org/10.1111/biom.12704>.
- Md S. Warasi. *groupTesting: Simulating and Modeling Group (Pooled) Testing Data*, 2024. R package version 1.3.0, <https://CRAN.R-project.org/package=groupTesting>.

Acknowledgements

I am extremely grateful to Dr. Shamim Sarker for mentorship throughout this project and to the Radford University SURF program for funding this research.

Source Code

The source code for **bayesGT** is publicly available on GitHub. Explore the package, run your own analyses, or review the methods:

github.com/caleb-ives/bayesGT