

Replicate Variation

Michael L. Fugate, CCS-6; Timothy C. Wallstrom, T-13

Over the past year, we have been working on the problem of estimating replicate variation, which is the variance in some parameter arising from unmodelled differences in nominally identical experiments, or replicates. Such estimates are essential for many purposes, such as comparing computational models with experimental data.

If all of the experiments are of the same type, the problem is very simple: we simply use the standard variance estimate. Frequently, however, our data come from many different types of experiments, and the variance may differ from one group of experiments to the next.

To understand the nature of the problem, consider two different models we might use to solve this problem. In the first model, we assume that the variance of the groups is completely independent, so we estimate the variance separately for each group. However, there may be few datapoints for each group, so the uncertainty in each of the variance estimates may be so large as to render them almost worthless. If the groups are related, then data from one may give information about others, and we are throwing out a lot of information in each of the individual estimates. In the second model, we assume that the variances of the groups are identical, and we pool all of the data. We now have many datapoints, and the statistical

uncertainty is greatly reduced. However, we may be inappropriately assigning the same variance to groups whose variances are actually quite different. Some middle ground is needed.

The middle ground is provided through the use of Bayesian methods, using an approach known as hierarchical modeling. We introduce different variance parameters for each group. However, we do not allow these estimates to vary freely; they must come from a “population distribution” that is determined, in typical Bayesian fashion, from a combination of “prior information” and data, where the data are from all the groups combined. In this way, the data itself are able to influence whether the estimates for different groups should be constrained to agree closely, or allowed to be independent.

In Fig. 1, we show how the data from several groups of replicates can be used to improve the estimate of the variance in any particular group. The replicate variation is assumed to be described by an “inverse gamma” distribution, parameterized by unknown α and β . The key parameter is α : the larger the value of α , the better we know σ . The figure shows the posterior probability of α and β , given 20 groups of 6 replicates each, where the prior probability was assumed constant. The peak occurs at about $\alpha = 10$, which is close to the value of 15 which was used to generate the data. We can use this information to obtain better estimates of the replicate

variance for any particular group, even though we did not assume that the variance was the same for every group.

For more information contact Timothy Wallstrom at tcw@lanl.gov.

Funding Acknowledgements

NNSA's Advanced Simulation and Computing (ASC), Verification and Validation Program.

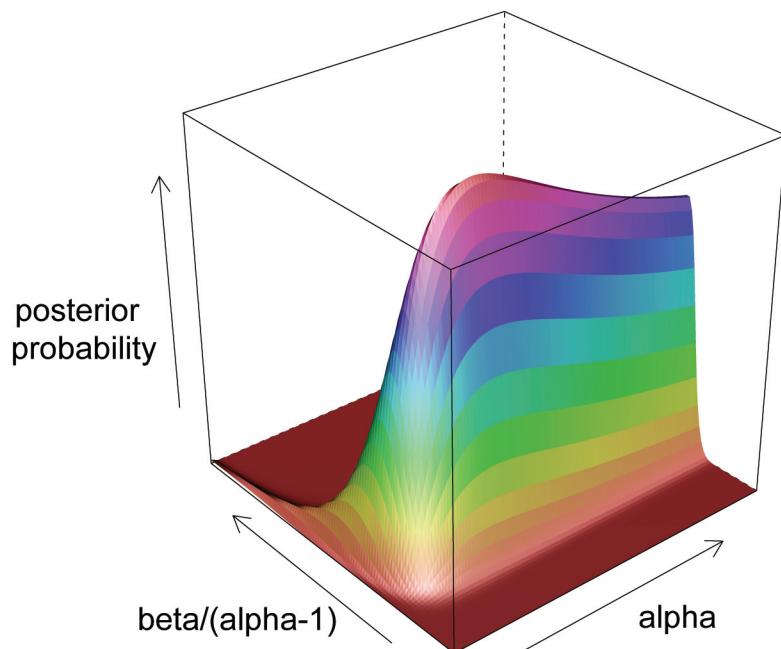


Fig. 1.
Posterior probability
of α and β .