Coverage of the β_{g4} parameters using the horseshoe distribution in `fbseq`

June 22, 2016

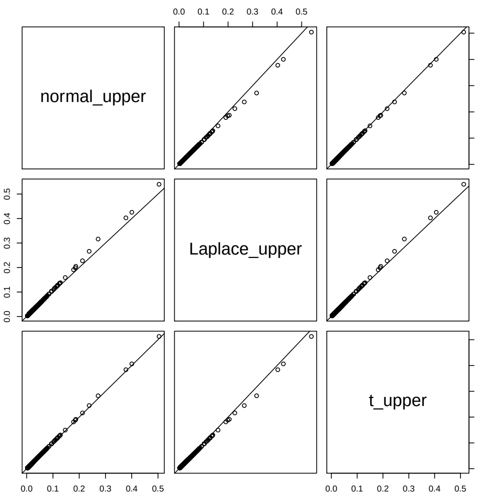
Will Landau

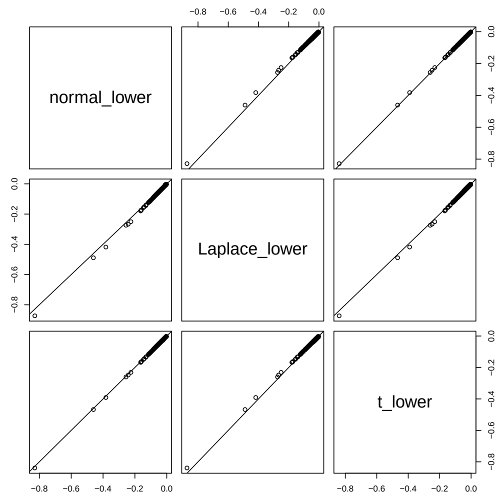
Department of Statistics
Iowa State University

As part of `fbseq` paper 3, the horseshoe distribution was used to analyze datasets generated from the normal model. When credible intervals of the β_{g4}'s were approximated with normal distributions, coverage of the true parameter values was extremely poor. Here, I take one example normal-generated dataset and try different distributions (Student t_6 and Laplace) for approximating credible intervals. Coverage improves only slightly.

	upper_bound_high_enough	lower_bound_low_enough	ci_covers_truth
Laplace	0.71360	0.54193	0.25553
t	0.70863	0.53413	0.24277
normal	0.70677	0.53180	0.23857

The upper and lower bounds were mostly very similar across models.





I suspect that the poor coverage reflects not a bug in the code, but rather a lack of fit quality in our model. From the figures in paper 3, coverage is on par with other models whenever

1. the true β_{g4} values are generated as in the edgeR and Simple scenarios rather than coming from a normal distribution.
2. a non-horseshoe distribution is applied to the β_{g4}'s for analysis.
3. we focus on $\beta_{g\ell}$'s for $\ell \neq 4$.

In addition, convergence using the horseshoe was excellent for non-normal simulations and worse for normal ones. On the real data, no Gelman factors exceeded 1.1, not even the ones for the ϵ_{gn}'s. The same was true for the non-normal simulations in Simulation Study 2 except the edgeR scenario with $N = 32$, where $\beta_{6411,3}$ and $\beta_{6411,1}$ had Gelman factors of around 1.28. For the normal simulations, however, usually 10-20 parameters had high Gelman factors, almost always including σ_4^2 ($\widehat{R} \approx 1.5$).