

Homework 1 (due October 4, 2016)

Problem 1: [Numerical example] Simulate data from a Gaussian regression model $Y \sim N(X\beta, I_n)$ with $p = 7$ (including the intercept), $n = 100$, $x_{ij} \sim U(0, 1)$ and the intercept and the slope for x_{i2} as the only non-zero coefficients with signal strength 6. Using a spike and slab mixture prior $\beta_j \mid \tau_j, c_j, \pi \sim (1 - \pi)N(0, \tau_j^2) + \pi N(0, \tau_j^2 c_j^2)$. Set $\pi = 1/2$. Run Gibbs sampler for 5,000 iterations after discarding the first 2,000 as burn-in. Summarize posterior mean, median, credible interval and exclusion probabilities of the parameters and the top 10 highest posterior probability models. Calculate the percentage of the visited models. Increase p to 200 with only the first two active predictors (including intercept) and repeat the analysis. Now use a prior $\pi \sim U(0, 1)$ and repeat the analysis. A couple of important points:

1. Choose τ_j, c_j as in Section 2.2 of GM1993.
2. Report the percentage increase in speed per iteration of the Gibbs sampler if you use the inversion trick from BCM2016 versus if you do not.