

# Some notes on the Metropolis-Hastings Implementation

## Heteroskedastic BART Introduction

We take vanilla BART and update the last step of the Gibbs sampler to include a sampling an estimate for the variance for each observation,  $\sigma_1^2, \dots, \sigma_n^2$ .

$$\begin{array}{l|l} \mathcal{F}_1 & R_1, \sigma_1^2, \dots, \sigma_n^2 \\ M_1 & \mathcal{F}_1, R_1, \sigma_1^2, \dots, \sigma_n^2 \\ \vdots & \\ \mathcal{F}_m & R_m, \sigma_1^2, \dots, \sigma_n^2 \\ M_m & \mathcal{F}_m, R_m, \sigma_1^2, \dots, \sigma_n^2 \\ \sigma_1^2 & E \\ \vdots & \\ \sigma_n^2 & E \end{array}$$

## Prior on the variance

We keep the same prior on each of the observation's variances:

$$\sigma_1^2, \dots, \sigma_n^2 \stackrel{iid}{\sim} \text{InvGamma} \left( \frac{\nu}{2}, \frac{\nu\lambda}{2} \right)$$

We may have to rethink the above, but it seems like a good start. Once again we pick  $\nu = 3$  and we pick  $\lambda$  so that 90% of the prior's density is below  $s^2$ , our estimate from the data which is either just the sample variance, or the RMSE squared from a linear model.

We also can have individual priors:

$$\begin{array}{l} \sigma_1^2 \sim \text{InvGamma} \left( \frac{\nu}{2}, \frac{\nu\lambda_1}{2} \right) \\ \vdots \\ \sigma_n^2 \sim \text{InvGamma} \left( \frac{\nu}{2}, \frac{\nu\lambda_n}{2} \right) \end{array}$$

where  $\lambda_i$  is picked based on a weighted least squares residual.

## Changes in Gibbs Sampler for $M$

The other step that has to change is sampling the leaf  $\mu$ 's. Remember that  $M_t \mid \mathfrak{F}_1, R_1, \sigma_1^2, \dots, \sigma_n^2$  is actually the sampling for all leaves where each leaf is considered independent:

$$\begin{array}{c|c} \mu_{t1} & \mathfrak{F}_t, R_{t1}, \sigma_1^2, \dots, \sigma_n^2 \\ \vdots & \\ \mu_{tb_t} & \mathfrak{F}_t, R_{tb_t}, \sigma_1^2, \dots, \sigma_n^2 \end{array}$$

The subscripts on the  $R$  term indicates we only consider the data that falls into the leaf. So remember that the prior on the leaf value was normal, and the likelihood was assumed normal as well. The only thing that changes is we now have a different variance estimate for each observation.

We derive the correct posterior distribution below. We drop the subscripts just for convenience since it will be the same for each of the above and denote  $k$  as the number of data records that fell into this leaf.  $\sigma_1^2, \dots, \sigma_k^2 \subset \sigma_1^2, \dots, \sigma_n^2$  for the data in this leaf as well:

$$\begin{aligned} \mathbb{P}(\mu \mid R, \sigma_1^2, \dots, \sigma_k^2) &\propto \mathbb{P}(R \mid \mu, \sigma_1^2, \dots, \sigma_k^2) \mathbb{P}(\mu \mid \sigma_1^2, \dots, \sigma_k^2) \\ &= \mathcal{N}_k \left( \mu \mathbf{1}, \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_k^2 \end{bmatrix} \right) \mathcal{N}(0, \sigma_\mu^2) \\ &\quad \vdots \\ &= \mathcal{N} \left( \frac{\frac{k^2 \bar{R}}{\sum_{i=1}^k \sigma_i^2}}{\frac{1}{\sigma_\mu^2} + \frac{1}{\sum_{i=1}^k \sigma_i^2}}, \frac{1}{\frac{1}{\sigma_\mu^2} + \frac{1}{\sum_{i=1}^k \sigma_i^2}} \right) \end{aligned}$$

## Changes in Gibbs Sampler for $\sigma_i^2$

In vanilla BART, the posterior of  $\sigma^2$  was:

$$\sigma^2 \mid e_1, \dots, e_n \sim \text{InvGamma} \left( \frac{\nu + n}{2}, \frac{\lambda\nu + \sum_{i=1}^n e_i^2}{2} \right)$$

Now, we have to sample each  $\sigma_1^2, \dots, \sigma_n^2$ . We estimate each of them based on only one residual:

$$\sigma_i^2 \mid e_i \sim \text{InvGamma} \left( \frac{\nu + 1}{2}, \frac{\lambda\nu + e_i^2}{2} \right)$$