A Appendix

A.1 Proof of Theorem 1

For simplicity, assume $n = N/K$ is the number of observations in each batch and consider θ to be a one-dimensional parameter. We will show Theorem 1's statements separately for LISA and CMC.

LISA:

Proof. Given assumption **A1**, $\forall\ j\ \ w.p.1$:

$$\forall\ \epsilon_1^{(j)} > 0\ \ \exists\ \ M_1 > 0\ \ s.t.\ \ \forall\ n > M_1\quad |\hat{\theta}_{n,L}^{(j)} - \theta_L| < \epsilon_1^{(j)} \tag{1}$$

hence with the continuous assumption in **A3**, we have $\forall\ j\ \ w.p.1$:

$$\forall\ \gamma_1^{(j)} > 0\ \exists\ M_1 > 0\ s.t.\ \forall\ n > M_1\ \ \left|\log\left(\pi_{j,LISA}(\hat{\theta}_{n,L}^{(j)}|Y^{(j)})\right) - \log\left(\pi_{j,LISA}(\theta_L|Y^{(j)})\right)\right| < \gamma_1^{(j)} \tag{2}$$

We know that $\left(\pi_{Full}(\theta|Y_N)\right)^K \propto \prod_{j=1}^K \pi_{j,LISA}(\theta|Y^{(j)})$, hence:

$$\log\left(\pi_{Full}(\theta|Y_N)\right) = \frac{1}{K}\sum_{j=1}^K \log\left(\pi_{j,LISA}(\theta|Y^{(j)})\right) + c \tag{3}$$

where c is a constant. This implies that

$$\log\left(\pi_{Full}(\theta|Y_N)\right)\Bigg|_{\theta=\hat{\theta}_N} = \frac{1}{K}\sum_{j=1}^K \log\left(\pi_{j,LISA}(\hat{\theta}_N|Y^{(j)})\right) + c \tag{4}$$

Since $\hat{\theta}_N$ is the full posterior mode:

$$\left[\frac{1}{K}\sum_{j=1}^K \log\left(\pi_{j,LISA}(\hat{\theta}_N|Y^{(j)})\right)\right] - \left[\frac{1}{K}\sum_{j=1}^K \log\left(\pi_{j,LISA}(\theta_L|Y^{(j)})\right)\right] \geq 0 \tag{5}$$

and because $\hat{\theta}_{n,L}^{(j)}$ is the mode of $\pi_{j,LISA}$:

$$\frac{1}{K}\sum_{j=1}^{K}\log\left(\pi_{j,LISA}(\hat{\theta}_N|Y^{(j)})\right) \leq \frac{1}{K}\sum_{j=1}^{K}\log\left(\pi_{j,LISA}(\hat{\theta}_{n,L}^{(j)}|Y^{(j)})\right) \tag{6}$$

and thus from (5) and (6), we will have:

$$\begin{aligned} 0 &\leq \left[\frac{1}{K}\sum_{j=1}^{K}\log\left(\pi_{j,LISA}(\hat{\theta}_N|Y^{(j)})\right)\right] - \left[\frac{1}{K}\sum_{j=1}^{K}\log\left(\pi_{j,LISA}(\theta_L|Y^{(j)})\right)\right] \\ &\leq \left[\frac{1}{K}\sum_{j=1}^{K}\log\left(\pi_{j,LISA}(\hat{\theta}_{n,L}^{(j)}|Y^{(j)})\right)\right] - \left[\frac{1}{K}\sum_{j=1}^{K}\log\left(\pi_{j,LISA}(\theta_L|Y^{(j)})\right)\right] \end{aligned} \tag{7}$$

Taking absolute values from last inequality in (7) and using the triangle inequality, we have $w.p.1$:

$$\begin{aligned} &\frac{1}{K}\left|\sum_{j=1}^{K}\left[\log\left(\pi_{j,LISA}(\hat{\theta}_N|Y^{(j)})\right) - \log\left(\pi_{j,LISA}(\theta_L|Y^{(j)})\right)\right]\right| \leq \\ &\frac{1}{K}\left|\sum_{j=1}^{K}\left[\log\left(\pi_{j,LISA}(\hat{\theta}_{n,L}^{(j)}|Y^{(j)})\right) - \log\left(\pi_{j,LISA}(\theta_L|Y^{(j)})\right)\right]\right| \leq \\ &\frac{1}{K}\sum_{j=1}^{K}\left|\log\left(\pi_{j,LISA}(\hat{\theta}_{n,L}^{(j)}|Y^{(j)})\right) - \log\left(\pi_{j,LISA}(\theta_L|Y^{(j)})\right)\right| \leq \frac{1}{K}\sum_{j=1}^{K}\gamma_1^{(j)} = \gamma_1 \end{aligned} \tag{8}$$

The last inequality in (8) is followed by (2). From inequality (8) and the fact that the posteriors are unimodal as stated in assumption **A3**, we can conclude $w.p.1$:

$$|\hat{\theta}_N - \theta_L| \longrightarrow 0 \quad \text{as} \quad N \to \infty \tag{9}$$

From (9) and assumption **A1**, we can conclude $\forall\ j \quad w.p.1$:

$$|\hat{\theta}_N - \hat{\theta}_{n,L}^{(j)}| \longrightarrow 0 \quad \text{as} \quad n \to \infty \tag{10}$$

And from (10) and assumption **A3**, $wp.1$, $\forall\ j$:

$$\left| \frac{\partial^2}{\partial\theta^2} \log\left(\pi_{j,LISA}(\theta|Y^{(j)})\right)\Big|_{\theta=\hat{\theta}_N} - \frac{\partial^2}{\partial\theta^2} \log\left(\pi_{j,LISA}(\theta|Y^{(j)})\right)\Big|_{\theta=\hat{\theta}^{(j)}_{n,L}} \right| \longrightarrow 0 \quad \text{as} \quad n \to \infty \tag{11}$$

In addition from (10), we can also conclude that for any i and j such that $i \neq j$:

$$|\hat{\theta}^{(i)}_{n,L} - \hat{\theta}^{(j)}_{n,L}| \longrightarrow 0 \quad \text{as} \quad n \to \infty \tag{12}$$

And thus benefitting from (11), (12), and the structural form of sub-posterior distributions in LISA (or assumption **A2**) for $i \neq j$, we have $w.p.1$:

$$\left| \frac{\partial^2}{\partial\theta^2} \log\left(\pi_{i,LISA}(\theta|Y^{(i)})\right)\Big|_{\theta=\hat{\theta}^{(i)}_{n,L}} - \frac{\partial^2}{\partial\theta^2} \log\left(\pi_{j,LISA}(\theta|Y^{(j)})\right)\Big|_{\theta=\hat{\theta}^{(j)}_{n,L}} \right| \longrightarrow 0 \quad \text{as} \quad n \to \infty \tag{13}$$

Now take the second derivative with respect to θ from both sides of (3) evaluated at $\theta = \hat{\theta}_N$:

$$-\hat{I}_N := \frac{\partial^2}{\partial\theta^2} \log\left(\pi_{Full}(\theta|Y_N)\right)\Big|_{\theta=\hat{\theta}_N} = \frac{1}{K} \sum_{j=1}^{K} \frac{\partial^2}{\partial\theta^2} \log\left(\pi_{j,LISA}(\theta|Y^{(j)})\right)\Big|_{\theta=\hat{\theta}_N} \tag{14}$$

Denoting:

$$-\hat{I}^{(j)}_{n,L} := \frac{\partial^2}{\partial\theta^2} \log\left(\pi_{j,LISA}(\theta|Y^{(j)})\right)\Big|_{\theta=\hat{\theta}^{(j)}_{n,L}} \tag{15}$$

Using (11), (13), and (14), will result in:

$$|\hat{I}_N - \hat{I}^{(j)}_{n,L}| = \left| \frac{\partial^2}{\partial \theta^2} \log\big(\pi_{j,LISA}(\theta|Y^{(j)})\big)\big|_{\theta=\hat{\theta}^{(j)}_{n,L}} - \frac{1}{K}\sum_{i=1}^{K} \frac{\partial^2}{\partial \theta^2} \log\big(\pi_{i,LISA}(\theta|Y^{(i)})\big)\big|_{\theta=\hat{\theta}_N} \right|$$
$$\leq \frac{1}{K}\left| \frac{\partial^2}{\partial \theta^2} \log\big(\pi_{j,LISA}(\theta|Y^{(j)})\big)\big|_{\theta=\hat{\theta}^{(j)}_{n,L}} - \frac{\partial^2}{\partial \theta^2} \log\big(\pi_{j,LISA}(\theta|Y^{(j)})\big)\big|_{\theta=\hat{\theta}_N} \right| +$$
$$\frac{1}{K}\left| \sum_{i \neq j} \left[\frac{\partial^2}{\partial \theta^2} \log\big(\pi_{i,LISA}(\theta|Y^{(i)})\big)\big|_{\theta=\hat{\theta}_N} - \frac{\partial^2}{\partial \theta^2} \log\big(\pi_{j,LISA}(\theta|Y^{(j)})\big)\big|_{\theta=\hat{\theta}^{(j)}_{n,L}} \right] \right| \longrightarrow 0 \quad (16)$$

$w.p.1 \ \forall \ j$.

□

CMC:

Proof. In CMC, since $\pi_{Full}(\theta|Y_N) \propto \prod_{j=1}^{K} \pi_{j,CMC}(\theta|Y^{(j)})$, we will have

$$\log\big(\pi_{Full}(\theta|Y_N)\big) = \sum_{j=1}^{K} \log\big(\pi_{j,CMC}(\theta|Y^{(j)})\big) + c \quad (17)$$

where c is a constant. Thus, using **A1** through **A3** with a similar proof as in LISA, we can show that $w.p.1$:

$$|\hat{\theta}_N - \theta_C| \longrightarrow 0 \quad \text{as} \quad N \to \infty \quad (18)$$

and hence $\forall \ j \quad w.p.1$:

$$|\hat{\theta}_N - \hat{\theta}^{(j)}_{n,C}| \longrightarrow 0 \quad \text{as} \quad n \to \infty \quad (19)$$

$$|\hat{\theta}^{(i)}_{n,C} - \hat{\theta}^{(j)}_{n,C}| \longrightarrow 0 \quad \text{as} \quad n \to \infty \quad \text{for} \ \ i \neq j \quad (20)$$

Similarly, from (19) and assumption **A3**, $wp.1$, $\forall\ j$:

$$\left| \frac{\partial^2}{\partial \theta^2} \log\left(\pi_{j,CMC}(\theta|Y^{(j)})\right)\Big|_{\theta=\hat{\theta}_N} - \frac{\partial^2}{\partial \theta^2} \log\left(\pi_{j,CMC}(\theta|Y^{(j)})\right)\Big|_{\theta=\hat{\theta}^{(j)}_{n,C}} \right| \longrightarrow 0 \quad \text{as} \quad n \to \infty \tag{21}$$

And again benefitting from (20), (21), and the structural form of sub-posterior distributions in CMC (or assumption **A2**), for $i \neq j$, we have $w.p.1$:

$$\left| \frac{\partial^2}{\partial \theta^2} \log\left(\pi_{i,CMC}(\theta|Y^{(i)})\right)\Big|_{\theta=\hat{\theta}^{(i)}_{n,C}} - \frac{\partial^2}{\partial \theta^2} \log\left(\pi_{j,CMC}(\theta|Y^{(j)})\right)\Big|_{\theta=\hat{\theta}^{(j)}_{n,C}} \right| \longrightarrow 0 \quad \text{as} \quad n \to \infty \tag{22}$$

Now taking the second derivative with respect to θ from both sides of (17) evaluated at $\theta = \hat{\theta}_N$:

$$-\hat{I}_N := \frac{\partial^2}{\partial \theta^2} \log\left(\pi_{Full}(\theta|Y_N)\right)\Big|_{\theta=\hat{\theta}_N} = \sum_{j=1}^{K} \frac{\partial^2}{\partial \theta^2} \log\left(\pi_{j,CMC}(\theta|Y^{(j)})\right)\Big|_{\theta=\hat{\theta}_N} \tag{23}$$

Denoting:

$$-\hat{I}^{(j)}_{n,C} := \frac{\partial^2}{\partial \theta^2} \log\left(\pi_{j,CMC}(\theta|Y^{(j)})\right)\Big|_{\theta=\hat{\theta}^{(j)}_{n,C}} \tag{24}$$

Using (21), (22), and (23), will similarly result in:

$$\begin{aligned} |\frac{\hat{I}_N}{K} - \hat{I}^{(j)}_{n,C}| &= \left| \frac{\partial^2}{\partial \theta^2} \log\left(\pi_{j,CMC}(\theta|Y^{(j)})\right)\big|_{\theta=\hat{\theta}^{(j)}_{n,C}} - \frac{1}{K}\sum_{i=1}^{K} \frac{\partial^2}{\partial \theta^2} \log\left(\pi_{i,CMC}(\theta|Y^{(i)})\right)\big|_{\theta=\hat{\theta}_N} \right| \\ &\leq \frac{1}{K}\left| \frac{\partial^2}{\partial \theta^2} \log\left(\pi_{j,CMC}(\theta|Y^{(j)})\right)\big|_{\theta=\hat{\theta}^{(j)}_{n,C}} - \frac{\partial^2}{\partial \theta^2} \log\left(\pi_{j,CMC}(\theta|Y^{(j)})\right)\big|_{\theta=\hat{\theta}_N} \right| + \\ \frac{1}{K}&\left| \sum_{i \neq j} \left[\frac{\partial^2}{\partial \theta^2} \log\left(\pi_{i,CMC}(\theta|Y^{(i)})\right)\big|_{\theta=\hat{\theta}_N} - \frac{\partial^2}{\partial \theta^2} \log\left(\pi_{j,CMC}(\theta|Y^{(j)})\right)\big|_{\theta=\hat{\theta}^{(j)}_{n,C}} \right] \right| \longrightarrow 0 \end{aligned} \tag{25}$$

$w.p.1 \ \forall \ j.$ $\square$

A.2 BART

In this section we will use a similar explanation and notation given by Kapelner and Bleich (2013) to derive the acceptance ratios of the Metropolis-Hastings step in updating trees of BART. We will further extend these calculations for LISA and CMC.

The Metropolis-Hastings algorithm is used to draw samples from conditional distribution given in equation (14)

$$p(T \mid R, \sigma) \quad \propto \quad p(T) \int p(R \mid M, T, \sigma) \; p(M \mid T, \sigma) \; dM$$

Assume we propose T_*, then the acceptance ratio will be:

$$r \;\; = \;\; \underbrace{\frac{P(T_* \to T)}{P(T \to T_*)}}_{\text{transition ratio}} \;\; \times \;\; \underbrace{\frac{P(R \mid T_*, \sigma^2)}{P(R \mid T, \sigma^2)}}_{\text{likelihood ratio}} \;\; \times \;\; \underbrace{\frac{P(T_*)}{P(T)}}_{\text{tree structure ratio}}$$

We will calculate r for each possible proposal:

GROW Proposal:

- **Transition ratio:** Consider growing one of the b terminal nodes of tree T, say node η, to two children nodes. Then we will have:

$$\begin{aligned} P(T \to T_*) = & \; P(GROW) \; P(\text{choosing } \eta) \; P(\text{choosing a predictor to split on}) \times \\ & P(\text{choosing a splitting value}) \\ = & \; P(GROW) \frac{1}{b} \frac{1}{p(\eta)} \frac{1}{n_p(\eta)} \end{aligned}$$

  where $p(\eta)$ denotes the number of predictors left available to split on at node η

(there must be at least two unique values in each predictor to consider), and $n_p(\eta)$ denotes the number of unique splitting values left in the chosen pth attribute.

In addition, we have:

$$P(T_* \to T) = P(PRUNE)\ P(\text{choosing } \eta \text{ to prune}) = P(PRUNE)\frac{1}{w_*}$$

where w_* is the number of nodes with two terminal nodes in the new tree T_*. Hence the transition ratio will be:

$$\frac{P(T_* \to T)}{P(T \to T_*)} = \frac{P(PRUNE)}{P(GROW)} \frac{b\ p(\eta)\ n_p(\eta)}{w_*}$$

- **Likelihood ratio:** For computing the likelihood ratio, we have:

$$P(R_1, ..., R_n \mid T, \sigma^2) = \prod_{l=1}^{b} P(R_{l_1}, ..., R_{l_{n_l}} \mid \sigma^2)$$

since the data are partitioned across all b terminal nodes of tree T. R_{l_j} denotes the j-th data (residual) in the l-th terminal node and n_l is the number of observations in the l-th terminal node. From BART we know that $\mu_l \sim N(0, \sigma_\mu^2)$, hence we will have:

$$P(R_{l_1}, ..., R_{l_{n_l}} \mid \sigma^2) = \int_{\mathbb{R}} P(R_{l_1}, ..., R_{l_{n_l}} \mid \mu_l, \sigma^2)\ P(\mu_l; \sigma_\mu^2)\ d\mu_l.$$

By completion of the square this will equal to:

$$P(R_{l_1}, ..., R_{l_{n_l}} \mid \sigma^2) =$$

$$\frac{1}{(2\pi\sigma^2)^{n_l/2}}\sqrt{\frac{\sigma^2}{\sigma^2 + n_l\sigma_\mu^2}} \;\; \exp\left(- \frac{1}{2\sigma^2}\left[\sum_{i=1}^{n_l}(R_{l_i} - \bar{R}_l)^2 \;-\; \frac{\bar{R}_l^{\,2} n_l^{\,2}}{n_l + \frac{\sigma^2}{\sigma_\mu^2}} \;+\; n_l\bar{R}_l^{\,2}\right]\right), \tag{26}$$

where $\bar{R}_l$ is the average residual at terminal node l. Note that the likelihood is specified by all terminal nodes, and since T differs from T_* only at its l-th terminal node which splits into two terminal children l_L and l_R, the probability terms from other terminal nodes will be canceled in the likelihood ratio which results in (using (26)):

$$\frac{P(R \mid T_*, \sigma^2)}{P(R \mid T, \sigma^2)} = \sqrt{\frac{\sigma^2(\sigma^2 + n_l\sigma_\mu^2)}{(\sigma^2 + n_{l_L}\sigma_\mu^2)(\sigma^2 + n_{l_R}\sigma_\mu^2)}} \;\times$$

$$\exp\left(\frac{\sigma_\mu^2}{2\sigma^2}\left[\frac{(\sum_{i=1}^{n_{l_L}} R_{l_L,i})^2}{\sigma^2 + n_{l_L}\sigma_\mu^2} + \frac{(\sum_{i=1}^{n_{l_R}} R_{l_R,i})^2}{\sigma^2 + n_{l_R}\sigma_\mu^2} - \frac{(\sum_{i=1}^{n_l} R_{l,i})^2}{\sigma^2 + n_l\sigma_\mu^2}\right]\right), \tag{27}$$

where R_{l_L} and R_{l_R} are residuals in the left and right child (respectively) with corresponding number of observations n_{l_L} and n_{l_R}.

- **Tree Structure ratio:** Recall the descriptions given in BART related to the probability that node η at depth d_η is non-terminal:

$$P_{\text{Split}}(\eta) = \frac{\alpha}{(1 + d_\eta)^\beta}$$

with probability of assigning a rule given as:

$$P_{\text{Rule}}(\eta) = \frac{1}{p(\eta)} \frac{1}{n_p(\eta)}$$

Hence, the prior on each tree will be:

$$P(T) = \prod_{\eta \,\in\, \text{non-terminal nodes}} P_{\text{Split}}(\eta)\ P_{\text{Rule}}(\eta) \quad \times \prod_{\eta \,\in\, \text{terminal nodes}} (1 - P_{\text{Split}}(\eta))$$

which will result in the following tree structure ratio:

$$\frac{P(T_*)}{P(T)} = \alpha \frac{(1 - \frac{\alpha}{(2+d_\eta)^\beta})^2}{((1+d_\eta)^\beta - \alpha)\ p(\eta)\ n_p(\eta)}. \tag{28}$$

PRUNE Proposal:

- **Transition ratio:** A similar description as in the GROW step will lead to:

$$\frac{P(T_* \to T)}{P(T \to T_*)} = \frac{P(GROW)}{P(PRUNE)} \frac{w}{(b-1)\ p(\eta^*)\ n_p(\eta^*)}$$

  where w is the number of nodes with two terminal nodes in tree T. Note that tree T_* has one less terminal nodes $(b-1)$.

- **Likelihood ratio:** This is the inverse of the likelihood ratio in the GROW proposal.

- **Tree Structure ratio:** This is also the inverse of the tree structure in the GROW proposal.

CHANGE Proposal:

- **Transition ratio:** As described by Kapelner and Bleich (2013), for simplicity, we will only change the rule assignments for nodes with two terminal children. Hence:

$$P(T \to T_*) = P(CHANGE)\ P(\text{choosing } \eta)\ P(\text{choosing a predictor to split on})\times$$
$$P(\text{choosing a splitting value})$$

with the first three terms canceling in the transition ratio given as:

$$\frac{P(T_* \to T)}{P(T \to T_*)} = \frac{n_{p^*}(\eta^*)}{n_p(\eta)}.$$

- **Likelihood ratio:** T_* differs from T only from the two terminal children effected by the changed rules from their parents. Hence, by canceling the probabilities from other terminal nodes, we will achieve the likelihood ratio:

$$\frac{P(R \mid T_*, \sigma^2)}{P(R \mid T, \sigma^2)} = \sqrt{\frac{(\frac{\sigma^2}{\sigma_\mu^2} + n_1)(\frac{\sigma^2}{\sigma_\mu^2} + n_2)}{(\frac{\sigma^2}{\sigma_\mu^2} + n_1^*)(\frac{\sigma^2}{\sigma_\mu^2} + n_2^*)}} \times$$

$$\exp\left(\frac{1}{2\sigma^2}\left[\frac{(\sum_{i=1}^{n_{1^*}} R_{1^*,i})^2}{\frac{\sigma^2}{\sigma_\mu^2} + n_1^*} + \frac{(\sum_{i=1}^{n_{2^*}} R_{2^*,i})^2}{\frac{\sigma^2}{\sigma_\mu^2} + n_2^*} - \frac{(\sum_{i=1}^{n_1} R_{1,i})^2}{\frac{\sigma^2}{\sigma_\mu^2} + n_1} - \frac{(\sum_{i=1}^{n_2} R_{2,i})^2}{\frac{\sigma^2}{\sigma_\mu^2} + n_2}\right]\right), \quad (29)$$

where subscripts 1 and 2 denote the two terminal children, while the asterisk refers to the proposed tree T_*.

- **Tree Structure ratio:** Following the definition of $P(T)$, we will have:

$$\frac{P(T_*)}{P(T)} = \frac{n_p(\eta)}{n_{p^*}(\eta^*)}.$$

Note that:

$$\frac{P(T_* \to T)}{P(T \to T_*)} \times \frac{P(T_*)}{P(T)} = 1.$$

A.3 LISA for BART

GROW Proposal:

- **Transition ratio:** No change.
- **Likelihood ratio:** Equation (26) changes to:

$$P(R_{l_1}, ..., R_{l_{n_l}} \mid \sigma^2) =$$

$$\frac{1}{(2\pi\sigma^2)^{n_l/2}}\sqrt{\frac{\sigma^2}{\sigma^2 + Kn_l\sigma_\mu^2}} \;\; \exp\left(-\frac{K}{2\sigma^2}\left[\sum_{i=1}^{n_l}(R_{l_i} - \bar{R}_l)^2 \; - \; \frac{K\bar{R}_l^{\,2} n_l^{\,2}}{Kn_l + \frac{\sigma^2}{\sigma_\mu^2}} \; + \; n_l\bar{R}_l^{\,2}\right]\right). \tag{30}$$

Thus the likelihood ratio will change to:

$$\frac{P(R \mid T_*, \sigma^2)}{P(R \mid T, \sigma^2)} = \sqrt{\frac{\sigma^2(\sigma^2 + Kn_l\sigma_\mu^2)}{(\sigma^2 + Kn_{l_L}\sigma_\mu^2)(\sigma^2 + Kn_{l_R}\sigma_\mu^2)}} \;\; \times$$

$$\exp\left(\frac{K^2\sigma_\mu^2}{2\sigma^2}\left[\frac{(\sum_{i=1}^{n_{l_L}} R_{l_L,i})^2}{\sigma^2 + Kn_{l_L}\sigma_\mu^2} + \frac{(\sum_{i=1}^{n_{l_R}} R_{l_R,i})^2}{\sigma^2 + Kn_{l_R}\sigma_\mu^2} - \frac{(\sum_{i=1}^{n_l} R_{l,i})^2}{\sigma^2 + Kn_l\sigma_\mu^2}\right]\right). \tag{31}$$

- **Tree Structure ratio:** No change.

PRUNE Proposal:

- **Transition ratio:** No change.
- **Likelihood ratio:** This is the inverse of the likelihood ratio in the GROW proposal.

- **Tree Structure ratio:** No change.

CHANGE Proposal:

- **Transition ratio:** No change.
- **Likelihood ratio:**

$$\frac{P(R \mid T_*, \sigma^2)}{P(R \mid T, \sigma^2)} = \sqrt{\frac{(\frac{\sigma^2}{\sigma_\mu^2} + Kn_1)(\frac{\sigma^2}{\sigma_\mu^2} + Kn_2)}{(\frac{\sigma^2}{\sigma_\mu^2} + Kn_1^*)(\frac{\sigma^2}{\sigma_\mu^2} + Kn_2^*)}} \times$$

$$\exp\left(\frac{K^2}{2\sigma^2}\left[\frac{(\sum_{i=1}^{n_{1^*}} R_{1^*,i})^2}{\frac{\sigma^2}{\sigma_\mu^2} + Kn_1^*} + \frac{(\sum_{i=1}^{n_{2^*}} R_{2^*,i})^2}{\frac{\sigma^2}{\sigma_\mu^2} + Kn_2^*} - \frac{(\sum_{i=1}^{n_1} R_{1,i})^2}{\frac{\sigma^2}{\sigma_\mu^2} + Kn_1} - \frac{(\sum_{i=1}^{n_2} R_{2,i})^2}{\frac{\sigma^2}{\sigma_\mu^2} + Kn_2}\right]\right). \quad (32)$$

- **Tree Structure ratio:** No change.

The conditional posterior of σ^2 and M_j changes to:

- $\sigma^2 \mid (T_1, M_1), ..., (T_m, M_m), Y, X \quad \propto Inv - Gamma(\rho, \gamma)$

  where $\rho = \frac{\nu+Kn}{2}$ and $\gamma = \frac{1}{2}\ [\ K \sum_{i=1}^{n} (y_i - \sum_{j=1}^{m} g(x_i; M_j, T_j))^2 + \lambda\nu\ ]$.

- For the conditional posterior $M_j \mid T_j, R_j, \sigma$, we have:

$$\mu_{ij} \mid T_j, R_j, \sigma \;\sim\; \mathcal{N}\left(\frac{\frac{\sigma^2}{\sigma_\mu^2}\ \mu_\mu\ +\ Kn_i\bar{R}_{j(i)}}{\frac{\sigma^2}{\sigma_\mu^2}\ +\ Kn_i}, \frac{\sigma^2}{\frac{\sigma^2}{\sigma_\mu^2}\ +\ Kn_i}\right),$$

  where $\bar{R}_{j(i)}$ denotes the average residual (computed without tree j) at terminal node i with total number of data n_i. Note that we can consider $\mu_\mu = 0$.

A.4 CMC for BART

GROW Proposal:

- **Transition ratio:** No change.

- **Likelihood ratio:** Equation (26) changes to:

$$P(R_{l_1}, ..., R_{l_{n_l}} \mid \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n_l/2}} \left(\sqrt{2\pi\sigma_\mu^2}\right)^{1-\frac{1}{K}} \sqrt{\frac{\sigma^2}{\frac{\sigma^2}{K} + n_l\sigma_\mu^2}} \times$$

$$\exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n_l}(R_{l_i} - \bar{R}_l)^2 - \frac{\bar{R}_l{}^2 n_l{}^2}{n_l + \frac{\sigma^2}{K\sigma_\mu^2}} + n_l\bar{R}_l{}^2\right]\right) \quad (33)$$

Thus the likelihood ratio will change to:

$$\frac{P(R \mid T_*, \sigma^2)}{P(R \mid T, \sigma^2)} = (\sqrt{2\pi\sigma_\mu^2})^{1-\frac{1}{K}} \sqrt{\frac{\sigma^2(\frac{\sigma^2}{K} + n_l\sigma_\mu^2)}{(\frac{\sigma^2}{K} + n_{l_L}\sigma_\mu^2)(\frac{\sigma^2}{K} + n_{l_R}\sigma_\mu^2)}} \times$$

$$\exp\left(\frac{\sigma_\mu^2}{2\sigma^2}\left[\frac{(\sum_{i=1}^{n_{l_L}} R_{l_L,i})^2}{\frac{\sigma^2}{K} + n_{l_L}\sigma_\mu^2} + \frac{(\sum_{i=1}^{n_{l_R}} R_{l_R,i})^2}{\frac{\sigma^2}{K} + n_{l_R}\sigma_\mu^2} - \frac{(\sum_{i=1}^{n_l} R_{l,i})^2}{\frac{\sigma^2}{K} + n_l\sigma_\mu^2}\right]\right) \quad (34)$$

- **Tree Structure ratio:** The tree structure ratio will be raised to the power $1/K$:

$$\left[\frac{P(T_*)}{P(T)}\right]^{\frac{1}{K}}.$$

PRUNE Proposal:

- **Transition ratio:** No change.

- **Likelihood ratio:** This is the inverse of the likelihood ratio in the GROW proposal.

- **Tree Structure ratio:** This is also the inverse of the tree structure ratio in the GROW proposal.

CHANGE Proposal:

- **Transition ratio:** No change.

- **Likelihood ratio:**

$$\frac{P(R \mid T_*, \sigma^2)}{P(R \mid T, \sigma^2)} = \sqrt{\frac{(\frac{\sigma^2}{K\sigma_\mu^2} + n_1)(\frac{\sigma^2}{K\sigma_\mu^2} + n_2)}{(\frac{\sigma^2}{K\sigma_\mu^2} + n_1^*)(\frac{\sigma^2}{K\sigma_\mu^2} + n_2^*)}} \times$$

$$\exp\left(\frac{1}{2\sigma^2}\left[\frac{(\sum_{i=1}^{n_{1^*}} R_{1^*,i})^2}{\frac{\sigma^2}{K\sigma_\mu^2} + n_1^*} + \frac{(\sum_{i=1}^{n_{2^*}} R_{2^*,i})^2}{\frac{\sigma^2}{K\sigma_\mu^2} + n_2^*} - \frac{(\sum_{i=1}^{n_1} R_{1,i})^2}{\frac{\sigma^2}{K\sigma_\mu^2} + n_1} - \frac{(\sum_{i=1}^{n_2} R_{2,i})^2}{\frac{\sigma^2}{K\sigma_\mu^2} + n_2}\right]\right). \quad (35)$$

- **Tree Structure ratio:** The tree structure ratio will be raised to the power $1/K$.

Now the product of transition ratio and tree structure ratio is not 1 anymore:

$$\frac{P(T_* \to T)}{P(T \to T_*)} \times \frac{P(T_*)}{P(T)} = n_p(\eta)^{\frac{1}{K}-1} \quad n_{p^*}(\eta^*)^{1-\frac{1}{K}}.$$

The conditional posterior of σ^2 and M_j changes to:

- $\sigma^2 \mid (T_1, M_1), ..., (T_m, M_m), Y, X \quad \propto Inv - Gamma(\rho, \gamma)$

  where $\rho = \frac{\nu+2+K(n-2)}{2K}$ and $\gamma = \frac{1}{2}\left[\ \sum_{i=1}^{n} (y_i - \sum_{j=1}^{m} g(x_i; M_j, T_j))^2 + \frac{\lambda\nu}{K}\ \right]$.

- For the conditional posterior $M_j \mid T_j, R_j, \sigma$, we have:

$$\mu_{ij} \mid T_j, R_j, \sigma \ \sim \ \mathcal{N}(\frac{\frac{\sigma^2}{K\sigma_\mu^2}\ \mu_\mu \ + \ n_i \bar{R}_{j(i)}}{\frac{\sigma^2}{K\sigma_\mu^2} \ + \ n_i}, \frac{\sigma^2}{\frac{\sigma^2}{K\sigma_\mu^2} \ + \ n_i})$$

where we can consider $\mu_\mu = 0$.

References

Adam Kapelner and Justin Bleich. bartmachine: Machine learning with Bayesian additive regression trees. *arXiv preprint arXiv:1312.2171*, 2013.