Some notes on the Metropolis-Hastings Implementation

MCMC for BART

According to Gelman p.291, the Metropolis-Hastings algorithm differs from the Metropolis algorithm because you need to consider the ratio of ratios:

$$r = \frac{\dfrac{\mathbb{P}\left(\theta^* \mid Y\right)}{J_t(\theta^* \mid \theta^{t-1})}}{\dfrac{\mathbb{P}\left(\theta^{t-1} \mid Y\right)}{J_t(\theta^{t-1} \mid \theta^*)}} = \frac{J_t(\theta^{t-1} \mid \theta^*)}{J_t(\theta^* \mid \theta^{t-1})} \frac{\mathbb{P}\left(\theta^* \mid Y\right)}{\mathbb{P}\left(\theta^{t-1} \mid Y\right)}$$

Where we accept if a random draw from a uniform is less than the ratio above *i.e.* $X \sim$ U $(0,\ 1) \rightsquigarrow x < r$.

The parameters of interest in our case are the new tree (which is created from one of three types of proposals), which we denote T^*, and the original tree, denoted by T. For jump notation which is J, we denote this just using the regular probability symbol (to be boring). We denote the residual left over from Y, the vector of random variables, as $\boldsymbol{R} := [R_1, \ldots, R_n]^\top$. We denote σ^2 as the random variable estimating the homoskedastic noise in the model which is sampled after all the trees are sampled. We do not notate all the hyperparameters to avoid notational messiness. The M-H ratio then becomes:

$$r = \frac{\mathbb{P}\left(T^* \to T\right)}{\mathbb{P}\left(T \to T^*\right)} \frac{\mathbb{P}\left(T^* \mid \boldsymbol{R}, \sigma^2\right)}{\mathbb{P}\left(T \mid \boldsymbol{R}, \sigma^2\right)}$$

My goal is to come up with an exact way of calculating r for all possible tree proposals.

Throughout this document, we use the following notation:

- $H_{\text{internals}}$ — the collection of internal nodes in tree T

- H_{leaves} — the collection of leaves in tree T

- η — a node $\in H_{\text{internals}} \cup H_{\text{leaves}}$.

- d_η — the depth of the ηth node. The root node is defined as having depth 0, its first child has depth 1, etc.

- b — the number of terminal nodes / leaves in the tree T, $|H_{\text{leaves}}|$. These are the nodes that can potentially be "grown." For example, the number of terminal nodes in the T^* tree is $b+1$ if T was grown and $b-1$ if T was pruned.

- w_2 — the number of 2nd generation internal nodes in tree T, $|H_{\text{internals}}|$ *i.e.* the number of nodes who only have two children. These are the nodes that can potentially be "pruned."

- w_2^* — the number of 2nd generation internal nodes in tree T^*, $|H_{\text{internals}}^*|$. This can be equal to w or $w+1$ so it has to be recalculated. Draw a few pictures of trees and grow steps and you'll see why this unfortunate fact is so.

- ℓ — the index of the terminal node which we've "picked" to grow from it is a number $\in \{1, \ldots, b\}$.

- ℓ_L and ℓ_R — represent the new left and right node indices in tree T^* which are "grown" from the ℓth node of tree T

Likelihood Calculation

It is imperative that we can calculate $\mathbb{P}\left(T \mid \boldsymbol{R}, \sigma^2\right)$ for the calculation of r. Let's analyze carefully what it means to get the likelihood of a tree.

First, note that the likelihood for T given the data is not defined in our model, so we use Bayes Rule to obtain something that is tractable in our model:

$$\mathbb{P}\left(T \mid \boldsymbol{R}, \sigma^2\right) = \frac{\mathbb{P}\left(\boldsymbol{R} \mid T, \sigma^2\right) \mathbb{P}\left(T \mid \sigma^2\right)}{\mathbb{P}\left(\boldsymbol{R} \mid \sigma^2\right)}$$

What's in the numerator? The likelihood of the data given the tree and the variance times the probability of the tree given the variance. The probability of the tree is based on probabilities of splits and rules and is not dependent on the variance, $\mathbb{P}\left(T \mid \sigma^2\right) = \mathbb{P}\left(T\right)$, so we can already simplify to:

$$\mathbb{P}\left(T \mid \boldsymbol{R}, \sigma^2\right) = \frac{\mathbb{P}\left(\boldsymbol{R} \mid T, \sigma^2\right) \mathbb{P}\left(T\right)}{\mathbb{P}\left(\boldsymbol{R} \mid \sigma^2\right)}$$

What about the denominator — the probability of the data? The probability of the data is weighted over every possible tree configuration:

$$\mathbb{P}\left(\boldsymbol{R} \mid \sigma^2\right) = \int_{T \in \mathcal{T}} \mathbb{P}\left(\boldsymbol{R} \mid T, \sigma^2\right) \mathbb{P}\left(T\right) dT$$

and removing the dependency on T becomes:

$$\mathbb{P}\left(\boldsymbol{R} \mid \sigma^2\right) = \int_{T \in \mathcal{T}} \prod_{\ell=1}^{b_T} \mathbb{P}\left(R_{\ell_1}, \ldots, R_{\ell_{n_\ell}} \mid \sigma^2\right) \mathbb{P}\left(T\right) dT$$

which of course is arrived at via the margining out of the means of each leaf:

$$\mathbb{P}\left(\boldsymbol{R}\,\middle|\,\sigma^2\right) = \int_{T\in\mathcal{T}}\left(\prod_{\ell=1}^{b_T}\int_{\mathbb{R}}\mathbb{P}\left(R_{\ell_1},\ldots,R_{\ell_{n_\ell}}\,\middle|\,\mu,\sigma^2\right)\mathbb{P}\left(\mu;\sigma_\mu^2\right)d\mu\right)\mathbb{P}\left(T\right)dT$$

The point being is that this quantity is the same for all data $\boldsymbol{R}$. This is useful since we're creating ratios where we're using the same data, and this quantity will cancel.

Now let's look at the ratio of the likelihoods which is what we care about for the calculation of r. I identify three pieces which we will use for the next couple of sections:

$$\begin{aligned}
r &= \frac{\mathbb{P}\left(T^*\to T\right)}{\mathbb{P}\left(T\to T^*\right)}\underbrace{\frac{\mathbb{P}\left(T^*\,\middle|\,\boldsymbol{R},\sigma^2\right)}{\mathbb{P}\left(T\,\middle|\,\boldsymbol{R},\sigma^2\right)}}_{} \\[2em]
&\qquad\qquad\frac{\dfrac{\mathbb{P}\left(\boldsymbol{R}\,\middle|\,T^*,\sigma^2\right)\mathbb{P}\left(T^*\right)}{\cancel{\mathbb{P}\left(\boldsymbol{R}\,\middle|\,\sigma^2\right)}}}{\dfrac{\mathbb{P}\left(\boldsymbol{R}\,\middle|\,T,\sigma^2\right)\mathbb{P}\left(T\right)}{\cancel{\mathbb{P}\left(\boldsymbol{R}\,\middle|\,\sigma^2\right)}}} \\[2em]
&= \underbrace{\frac{\mathbb{P}\left(T^*\to T\right)}{\mathbb{P}\left(T\to T^*\right)}}_{\text{transition ratio}}\quad\times\quad\underbrace{\frac{\mathbb{P}\left(\boldsymbol{R}\,\middle|\,T^*,\sigma^2\right)}{\mathbb{P}\left(\boldsymbol{R}\,\middle|\,T,\sigma^2\right)}}_{\text{likelihood ratio}}\quad\times\quad\underbrace{\frac{\mathbb{P}\left(T^*\right)}{\mathbb{P}\left(T\right)}}_{\text{tree structure ratio}}
\end{aligned}$$

Grow Proposal

Let's pretend we're transitioning from $T\to T^*$ using a GROW step and let's analyze each of the above expressions one-by-one.

Transition Ratio

So $\mathbb{P}\left(T\to T^*\right)$ means the probability of transitioning from T into the new tree proposal T^*. This would have to be equal to the following:

$$\begin{aligned}
\mathbb{P}\left(T\to T^*\right) =\ &\mathbb{P}\left(\text{GROW}\right)\mathbb{P}\left(\text{selecting the }\ell\text{th node to grow from}\right)\times \\
&\mathbb{P}\left(\text{selecting the }j\text{th attribute to split on}\right)\mathbb{P}\left(\text{selecting the }i\text{th value to split on}\right)
\end{aligned}$$

We're picking from one of the terminal nodes, and then we're picking an attribute and split point, this becomes:

$$\mathbb{P}\left(T^*\,\middle|\,T\right) =\ \mathbb{P}\left(\text{GROW}\right)\frac{1}{b}\frac{1}{p_{\text{adj}}(\eta^*)}\frac{1}{n_{j\cdot\text{adj}}(\eta^*)}$$

Now, $p_{\text{adj}}(\eta^*)$ is the number of predictors left available to split on. This is *from the perspective* of the ℓth node in tree T. Why would this be less than p? Because if you look up into the node's lineage, you may have already used all available split values for some attributes. Those would no longer be available to split from.

$n_{j\cdot\text{adj}}(\eta^*)$ is the number of *unique* values left in the pth column after adjusting for previous splits. We can obtain this adjusted row count by looking at the node's lineage for any splits on

j and then taking the minimum of those split values, and then finding the subset of the design matrix whose values are less than that minimum for column j.

So now $\mathbb{P}(T^* \to T)$ is the probability of transitioning from the new tree back to the old tree which would be:

$$
\begin{aligned}
\mathbb{P}(T^* \to T) &= \mathbb{P}(\text{PRUNE})\,\mathbb{P}(\text{selecting the } \ell\text{th node to prune from}) \\
&= \mathbb{P}(\text{PRUNE})\,\frac{1}{w_2^*}
\end{aligned}
$$

Thus, the transition ratio will be:

$$
\frac{\mathbb{P}(T^* \to T)}{\mathbb{P}(T \to T^*)} = \frac{\mathbb{P}(\text{PRUNE})\,\frac{1}{w_2^*}}{\mathbb{P}(\text{GROW})\,\frac{1}{b}\,\frac{1}{p_{\text{adj}}(\eta^*)}\,\frac{1}{n_{j\cdot\text{adj}}(\eta^*)}} = \frac{\mathbb{P}(\text{PRUNE})\overset{?}{\cancel{b}}\,p_{\text{adj}}(\eta^*)\,n_{j\cdot\text{adj}}(\eta^*)}{\cancel{\mathbb{P}(\text{GROW})}\,w_2^*}
$$

Why don't the probabilities of prune and grow cancel? Well, under the case where we cannot grow anymore (if we use all split variables), the probability of growth will be 0. Thus, those steps *cannot* be considered since the ratio would be undefined. As long as they can be considered, that ratio will cancel. In log form, this becomes:

$$
\ln(b) + \ln\left(p_{\text{adj}}(\eta^*)\right) + \ln\left(n_{j\cdot\text{adj}}(\eta^*)\right) - \ln(w_2^*)
$$

Likelihood Ratio

What about the likelihood given the tree? The only reason you need the T information is so we know which R values fall in which of the leaves:

$$
\mathbb{P}\left(R_1, \ldots, R_n \,\middle|\, T, \sigma^2\right) = \prod_{\ell=1}^{b} \underbrace{\mathbb{P}\left(R_{\ell_1}, \ldots, R_{\ell_{n_\ell}} \,\middle|\, \sigma^2\right)}
$$

The r.h.s underbraces is the likelihood of each leaf separately which is *not* dependent on T anymore. We can multiply the likelihoods of each leaf because we assume the leaves are independent. The R_ℓ's are the data in the ℓth leaf and there is n_ℓ of them, the portion of n in the leaf. Obviously, $n = \sum_{\ell=1}^{b} n_\ell$.

Let's look at the likelihood of a single leaf more carefully. We know that if we knew the mean at the leaf, which we denote μ_ℓ, we would have:

$$
R_{\ell_1}, \ldots, R_{\ell_{n_\ell}} | \mu_\ell, \sigma^2 \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)
$$

This means that if we can margin out μ, we can arrive at the expression that is needed for the calculation. Recall that one of the BART model assumptions is a prior on the average value of $\mu \sim \mathcal{N}\left(0, \sigma_\mu^2\right)$ and thus:

$$\mathbb{P}\left(R_{\ell_1},\ldots,R_{\ell_{n_\ell}} \,\middle|\, \sigma^2\right) = \int_{\mathbb{R}} \mathbb{P}\left(R_{\ell_1},\ldots,R_{\ell_{n_\ell}} \,\middle|\, \mu,\sigma^2\right) \mathbb{P}\left(\mu;\sigma_\mu^2\right) d\mu$$

Since the likelihoods are solely determined by the terminal nodes, the proposal tree differs from the original tree by only the ℓth node in the original becoming the ℓ_L and ℓ_R nodes in the proposal. Hence, the likelihood ratio becomes only:

$$\frac{\mathbb{P}\left(\boldsymbol{R} \,\middle|\, T^*, \sigma^2\right)}{\mathbb{P}\left(\boldsymbol{R} \,\middle|\, T, \sigma^2\right)} = \frac{\mathbb{P}\left(R_{\ell_{L,1}},\ldots,R_{\ell_{L,n_{\ell,L}}} \,\middle|\, \sigma^2\right) \mathbb{P}\left(R_{\ell_{R,1}},\ldots,R_{R,\ell_{n_{\ell,R}}} \,\middle|\, \sigma^2\right)}{\mathbb{P}\left(R_{\ell_1},\ldots,R_{\ell_{n_\ell}} \,\middle|\, \sigma^2\right)}$$

$$= \frac{\int_{\mathbb{R}} \mathbb{P}\left(R_{\ell_{L,1}},\ldots,R_{\ell_{L,n_{\ell,L}}} \,\middle|\, \mu,\sigma^2\right) \mathbb{P}\left(\mu\right) d\mu \;\; \int_{\mathbb{R}} \mathbb{P}\left(R_{\ell_{R,1}},\ldots,R_{R,\ell_{n_{\ell,R}}} \,\middle|\, \mu,\sigma^2\right) \mathbb{P}\left(\mu\right) d\mu}{\int_{\mathbb{R}} \mathbb{P}\left(R_{\ell_1},\ldots,R_{\ell_{n_\ell}} \,\middle|\, \mu,\sigma^2\right) \mathbb{P}\left(\mu\right) d\mu} \quad (1)$$

I now present three strategies to perform the margining.

Strategy #1 — Convolution

Let's take a more careful look at the non-margined leaf-likelihood expression and recall from basic mathematical statistics the result that the sample average, $\bar{R}_\ell$, is a sufficient statistic for μ via the factorization theorem:

$$\begin{aligned}
\mathbb{P}\left(R_{\ell_1},\ldots,R_{\ell_{n_\ell}} \,\middle|\, \mu,\sigma^2\right) &= \frac{1}{(2\pi\sigma^2)^{n_\ell/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_\ell}\left(R_{\ell_i}-\mu\right)^2\right) \\
&= \frac{1}{(2\pi\sigma^2)^{n_\ell/2}} \exp\left(-\frac{1}{2\sigma^2}\left(n_\ell\left(\bar{R}_\ell-\mu\right)^2 + \sum_{i=1}^{n_\ell}\left(R_{\ell_i}-\bar{R}_\ell\right)^2\right)\right) \\
&= \frac{1}{(2\pi\sigma^2)^{n_\ell/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_\ell}\left(R_{\ell_i}-\bar{R}_\ell\right)^2\right) \exp\left(-\frac{n_\ell}{2\sigma^2}\left(\bar{R}_\ell-\mu\right)^2\right) \\
&= \underbrace{\frac{1}{\sqrt{n_\ell}}\frac{1}{(2\pi\sigma^2)^{\frac{n_\ell-1}{2}}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_\ell}\left(R_{\ell_i}-\bar{R}_\ell\right)^2\right)}_{h} \sqrt{\frac{n_\ell}{2\pi\sigma^2}} \exp\left(-\frac{n_\ell}{2\sigma^2}\left(\bar{R}_\ell-\mu\right)^2\right)
\end{aligned}$$

So now we margin by integrating:

$$\begin{aligned}
\mathbb{P}\left(R_{\ell_1},\ldots,R_{\ell_{n_\ell}} \,\middle|\, \sigma^2\right) &= \int_{\mathbb{R}} \mathbb{P}\left(R_{\ell_1},\ldots,R_{\ell_{n_\ell}} \,\middle|\, \mu_\ell,\sigma^2\right) \mathbb{P}\left(\mu_\ell;\sigma_\mu^2\right) d\mu_\ell \\
&= h \underbrace{\int_{\mathbb{R}} \sqrt{\frac{n_\ell}{2\pi\sigma^2}} \exp\left(-\frac{n_\ell}{2\sigma^2}\left(\bar{R}_\ell-\mu_\ell\right)^2\right) \frac{1}{\sqrt{2\pi\sigma_\mu^2}} \exp\left(-\frac{1}{2\sigma_\mu^2}\mu^2\right) d\mu_\ell}_{\text{the definition of a convolution}}
\end{aligned}$$

5

$$= h\,\mathcal{N}\left(0,\,\frac{\sigma^2}{n_\ell}\right)\star\mathcal{N}\left(0,\,\sigma_\mu^2\right)$$

$$= h\,\mathcal{N}\left(0,\,\frac{\sigma^2}{n_\ell}+\sigma_\mu^2\right)\quad\text{(with free variable }\bar{R}_\ell\text{)}$$

$$= h\,\frac{1}{\sqrt{2\pi\left(\frac{\sigma^2}{n_\ell}+\sigma_\mu^2\right)}}\,\exp\left(-\frac{1}{2\left(\frac{\sigma^2}{n_\ell}+\sigma_\mu^2\right)}\bar{R}_\ell^2\right)$$

So now we substitue the above result into equation 1:

$$\frac{\mathbb{P}\left(\boldsymbol{R}\,\middle|\,T^*,\sigma^2\right)}{\mathbb{P}\left(\boldsymbol{R}\,\middle|\,T,\sigma^2\right)}$$

$$= \frac{h_L\,\dfrac{1}{\sqrt{2\pi\left(\frac{\sigma^2}{n_{\ell_L}}+\sigma_\mu^2\right)}}\,\exp\left(-\dfrac{1}{2\left(\frac{\sigma^2}{n_{\ell_L}}+\sigma_\mu^2\right)}\bar{R}_{\ell_L}^2\right)\;h_R\,\dfrac{1}{\sqrt{2\pi\left(\frac{\sigma^2}{n_{\ell_R}}+\sigma_\mu^2\right)}}\,\exp\left(-\dfrac{1}{2\left(\frac{\sigma^2}{n_{\ell_R}}+\sigma_\mu^2\right)}\bar{R}_{\ell_R}^2\right)}{h\,\dfrac{1}{\sqrt{2\pi\left(\frac{\sigma^2}{n_\ell}+\sigma_\mu^2\right)}}\,\exp\left(-\dfrac{1}{2\left(\frac{\sigma^2}{n_\ell}+\sigma_\mu^2\right)}\bar{R}_\ell^2\right)}$$

$$= \frac{h_L h_R}{h}\sqrt{\frac{2\pi\left(\frac{\sigma^2}{n_\ell}+\sigma_\mu^2\right)}{2\pi\left(\frac{\sigma^2}{n_{\ell_L}}+\sigma_\mu^2\right)2\pi\left(\frac{\sigma^2}{n_{\ell_R}}+\sigma_\mu^2\right)}}\;\frac{\exp\left(-\dfrac{1}{2\left(\frac{\sigma^2}{n_{\ell_L}}+\sigma_\mu^2\right)}\bar{R}_{\ell_L}^2\right)\,\exp\left(-\dfrac{1}{2\left(\frac{\sigma^2}{n_{\ell_R}}+\sigma_\mu^2\right)}\bar{R}_{\ell_R}^2\right)}{\exp\left(-\dfrac{1}{2\left(\frac{\sigma^2}{n_\ell}+\sigma_\mu^2\right)}\bar{R}_\ell^2\right)}$$

$$= \frac{h_L h_R}{h}\sqrt{\frac{\left(\frac{\sigma^2}{n_\ell}+\sigma_\mu^2\right)}{\left(\frac{\sigma^2}{n_{\ell_L}}+\sigma_\mu^2\right)2\pi\left(\frac{\sigma^2}{n_{\ell_R}}+\sigma_\mu^2\right)}}\,\exp\left(\frac{1}{2\left(\frac{\sigma^2}{n_\ell}+\sigma_\mu^2\right)}\bar{R}_\ell^2-\frac{1}{2\left(\frac{\sigma^2}{n_{\ell_L}}+\sigma_\mu^2\right)}\bar{R}_{\ell_L}^2-\frac{1}{2\left(\frac{\sigma^2}{n_{\ell_R}}+\sigma_\mu^2\right)}\bar{R}_{\ell_R}^2\right)$$

Let's examine the ratio of the h functions:

$$\frac{h_L h_R}{h}=\frac{\dfrac{1}{\sqrt{n_{\ell_L}}}\dfrac{1}{(2\pi\sigma^2)^{\frac{n_{\ell_L}-1}{2}}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_{\ell_L}}\left(R_{L,i}-\bar{R}_{\ell_L}\right)^2\right)\dfrac{1}{\sqrt{n_{\ell_R}}}\dfrac{1}{(2\pi\sigma^2)^{\frac{n_{\ell_R}-1}{2}}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_{\ell_R}}\left(R_{\ell_R,i}-\bar{R}_{\ell,R}\right)^2\right)}{\dfrac{1}{\sqrt{n_\ell}}\dfrac{1}{(2\pi\sigma^2)^{\frac{n_\ell-1}{2}}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_\ell}\left(R_{\ell_i}-\bar{R}_\ell\right)^2\right)}$$

Now note that:

$$\frac{\dfrac{1}{(2\pi\sigma^2)^{\frac{n_{\ell_L}-1}{2}}}\dfrac{1}{(2\pi\sigma^2)^{\frac{n_{\ell_R}-1}{2}}}}{\dfrac{1}{(2\pi\sigma^2)^{\frac{n_\ell-1}{2}}}}=\frac{(2\pi\sigma^2)^{\frac{n_\ell-1}{2}}}{(2\pi\sigma^2)^{\frac{n_\ell-2}{2}}}=\sqrt{2\pi\sigma^2}$$

Grouping the rest of the terms inside and outside of the exponentiation:

$$\frac{h_L h_R}{h} = \sqrt{\frac{2\pi\sigma^2 n_\ell}{n_{\ell_L} n_{\ell_R}}} \exp\left(-\frac{1}{2\sigma^2}\left(\underbrace{\sum_{i=1}^{n_{\ell_L}}\left(R_{\ell_L,i}-\bar{R}_{\ell_L}\right)^2 + \sum_{i=1}^{n_{\ell_R}}\left(R_{\ell_R,i}-\bar{R}_{\ell,R}\right)^2 - \sum_{i=1}^{n_\ell}\left(R_{\ell_i}-\bar{R}_\ell\right)^2}\right)\right)$$

Now let's take a look at the underbraced quantity:

$$= \sum_{i=1}^{n_{\ell_L}} R_{\ell_L,i}^2 - 2n_{\ell_L}\bar{R}_{\ell_L}^2 + n_{\ell_L}\bar{R}_{\ell_L}^2 + \sum_{i=1}^{n_{\ell_R}} R_{\ell_R,i}^2 - 2n_{\ell_R}\bar{R}_{\ell_R}^2 + n_{\ell_R}\bar{R}_{\ell_R}^2 - \sum_{i=1}^{n_\ell} R_{\ell,i}^2 + 2n_\ell\bar{R}_\ell^2 - n_\ell\bar{R}_\ell^2$$

$$= \sum_{i=1}^{n_{\ell_L}} \cancel{R_{\ell_L,i}^2} - 2n_{\ell_L}\bar{R}_{\ell_L}^2 + n_{\ell_L}\bar{R}_{\ell_L}^2 + \sum_{i=1}^{n_{\ell_R}} \cancel{R_{\ell_R,i}^2} - 2n_{\ell_R}\bar{R}_{\ell_R}^2 + n_{\ell_R}\bar{R}_{\ell_R}^2 - \sum_{i=1}^{n_\ell} \cancel{R_{\ell,i}^2} + 2n_\ell\bar{R}_\ell^2 - n_\ell\bar{R}_\ell^2$$

$$= n_\ell\bar{R}_\ell^2 - n_{\ell_L}\bar{R}_{\ell_L}^2 - n_{\ell_R}\bar{R}_{\ell_R}^2$$

Thus, we can arrive at the h ratio's final expression:

$$\frac{h_L h_R}{h} = \sqrt{\frac{2\pi\sigma^2 n_\ell}{n_{\ell_L} n_{\ell_R}}} \exp\left(-\frac{1}{2\sigma^2}\left(n_\ell\bar{R}_\ell^2 - n_{\ell_L}\bar{R}_{\ell_L}^2 - n_{\ell_R}\bar{R}_{\ell_R}^2\right)\right)$$

Now, putting it all together:

$$\frac{\mathbb{P}\left(\boldsymbol{R}\,\middle|\,T^*,\sigma^2\right)}{\mathbb{P}\left(\boldsymbol{R}\,\middle|\,T,\sigma^2\right)}$$

$$= \sqrt{\frac{2\pi\sigma^2 n_\ell}{n_{\ell_L} n_{\ell_R}}} \sqrt{\frac{\left(\frac{\sigma^2}{n_\ell}+\sigma_\mu^2\right)}{\left(\frac{\sigma^2}{n_{\ell_L}}+\sigma_\mu^2\right)2\pi\left(\frac{\sigma^2}{n_{\ell_R}}+\sigma_\mu^2\right)}} \exp\left(-\frac{1}{2\sigma^2}\left(n_\ell\bar{R}_\ell^2 - n_{\ell_L}\bar{R}_{\ell_L}^2 - n_{\ell_R}\bar{R}_{\ell_R}^2\right)\right) \times$$

$$\exp\left(\frac{1}{2}\left(\frac{1}{\left(\frac{\sigma^2}{n_\ell}+\sigma_\mu^2\right)}\bar{R}_\ell^2 - \frac{1}{\left(\frac{\sigma^2}{n_{\ell_L}}+\sigma_\mu^2\right)}\bar{R}_{\ell_L}^2 - \frac{1}{\left(\frac{\sigma^2}{n_{\ell_R}}+\sigma_\mu^2\right)}\bar{R}_{\ell_R}^2\right)\right)$$

$$= \sqrt{\frac{\sigma^2 n_\ell}{n_{\ell_L} n_{\ell_R}} \frac{\left(\frac{\sigma^2}{n_\ell}+\sigma_\mu^2\right)}{\left(\frac{\sigma^2}{n_{\ell_L}}+\sigma_\mu^2\right)\left(\frac{\sigma^2}{n_{\ell_R}}+\sigma_\mu^2\right)}} \exp\left(\frac{1}{2\sigma^2}\left(n_{\ell_L}\bar{R}_{\ell_L}^2 + n_{\ell_R}\bar{R}_{\ell_R}^2 - n_\ell\bar{R}_\ell^2\right)\right) \times$$

$$\exp\left(\frac{1}{2\sigma^2}\left(\frac{1}{\left(\frac{1}{n_\ell}+\frac{\sigma_\mu^2}{\sigma^2}\right)}\bar{R}_\ell^2 - \frac{1}{\left(\frac{1}{n_{\ell_L}}+\frac{\sigma_\mu^2}{\sigma^2}\right)}\bar{R}_{\ell_L}^2 - \frac{1}{\left(\frac{1}{n_{\ell_R}}+\frac{\sigma_\mu^2}{\sigma^2}\right)}\bar{R}_{\ell_R}^2\right)\right)$$

$$= \underbrace{\sqrt{\frac{\sigma^2\left(\sigma^2+\sigma_\mu^2 n_{\ell_L}\right)}{\left(\sigma^2+\sigma_\mu^2 n_{\ell_L}\right)\left(\sigma^2+\sigma_\mu^2 n_{\ell_R}\right)}}}_{c} \times$$

7

$$\exp\left(\frac{1}{2\sigma^2}\left(\left(\frac{1}{\left(\frac{1}{n_\ell}+\frac{\sigma_\mu^2}{\sigma^2}\right)}-n_\ell\right)\bar{R}_\ell^2-\left(\frac{1}{\left(\frac{1}{n_{\ell_L}}+\frac{\sigma_\mu^2}{\sigma^2}\right)}+n_{\ell_L}\right)\bar{R}_{\ell_L}^2-\left(\frac{1}{\left(\frac{1}{n_{\ell_R}}+\frac{\sigma_\mu^2}{\sigma^2}\right)}+n_{\ell_R}\right)\bar{R}_{\ell_R}^2\right)\right)$$

$$=\ c\exp\left(\frac{1}{2\sigma^2}\left(-\left(\frac{\frac{\sigma_\mu^2 n_\ell}{\sigma^2}}{\left(\frac{1}{n_\ell}+\frac{\sigma_\mu^2}{\sigma^2}\right)}\right)\bar{R}_\ell^2+\left(\frac{\frac{\sigma_\mu^2 n_{\ell_L}}{\sigma^2}}{\left(\frac{1}{n_{\ell_L}}+\frac{\sigma_\mu^2}{\sigma^2}\right)}\right)\bar{R}_{\ell_L}^2+\left(\frac{\frac{\sigma_\mu^2 n_{\ell_R}}{\sigma^2}}{\left(\frac{1}{n_{\ell_R}}+\frac{\sigma_\mu^2}{\sigma^2}\right)}\right)\bar{R}_{\ell_R}^2\right)\right)$$

$$=\ c\exp\left(\frac{1}{2\sigma^2}\left(\left(\frac{\sigma_\mu^2 n_{\ell_L}}{\left(\frac{\sigma^2}{n_{\ell_L}}+\sigma_\mu^2\right)}\right)\bar{R}_{\ell_L}^2+\left(\frac{\sigma_\mu^2 n_{\ell_R}}{\left(\frac{\sigma^2}{n_{\ell_R}}+\sigma_\mu^2\right)}\right)\bar{R}_{\ell_R}^2\right)-\left(\frac{\sigma_\mu^2 n_\ell}{\left(\frac{\sigma^2}{n_\ell}+\sigma_\mu^2\right)}\right)\bar{R}_\ell^2\right)$$

$$=\ \sqrt{\frac{\sigma^2\left(\sigma^2+\sigma_\mu^2 n_{\ell_L}\right)}{\left(\sigma^2+\sigma_\mu^2 n_{\ell_L}\right)\left(\sigma^2+\sigma_\mu^2 n_{\ell_R}\right)}}\exp\left(\frac{\sigma_\mu^2}{2\sigma^2}\left(\frac{\left(\sum_{i=1}^{n_{\ell_L}}R_{\ell,i}\right)^2}{\left(\sigma^2+n_{\ell_L}\sigma_\mu^2\right)}+\frac{\left(\sum_{i=1}^{n_{\ell_R}}R_{\ell,i}\right)^2}{\left(\sigma^2+n_{\ell_R}\sigma_\mu^2\right)}-\frac{\left(\sum_{i=1}^{n_\ell}R_{\ell,i}\right)^2}{\left(\sigma^2+n_\ell\sigma_\mu^2\right)}\right)\right)$$

Strategy #2 — Integral Computation in Mathematica

Let's factorize the $\overset{iid}{\sim}$ normal likelihood like the following:

$$\mathbb{P}\left(R_{\ell_1},\ldots,R_{\ell_{n_\ell}}|\mu,\sigma^2\right)\ =\ \frac{1}{(2\pi\sigma^2)^{n_\ell/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_\ell}\left(R_{\ell_i}-\mu\right)^2\right)$$

$$=\ \frac{1}{(2\pi\sigma^2)^{n_\ell/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_\ell}R_{\ell_i}^2\right)\underbrace{\exp\left(-\frac{1}{2\sigma^2}\left(-2n_\ell\mu\bar{R}_\ell+n_\ell\mu^2\right)\right)}_{g}$$

Now, let's try to build the ratio:

$$\frac{\mathbb{P}\left(\boldsymbol{R}\,|\,T^*,\sigma^2\right)}{\mathbb{P}\left(\boldsymbol{R}\,|\,T,\sigma^2\right)}$$

$$=\ \frac{\mathbb{P}\left(R_{\ell_L,1},\ldots,R_{\ell_{L,n_{\ell,L}}}\,|\,\sigma^2\right)\mathbb{P}\left(R_{\ell_R,1},\ldots,R_{R,\ell_{n_{\ell,R}}}\,|\,\sigma^2\right)}{\mathbb{P}\left(R_{\ell_1},\ldots,R_{\ell_{n_\ell}}\,|\,\sigma^2\right)}$$

$$=\ \frac{\int_{\mathbb{R}}\mathbb{P}\left(R_{\ell_L,1},\ldots,R_{\ell_{L,n_{\ell,L}}}\,|\,\mu,\sigma^2\right)\mathbb{P}\left(\mu\right)\,\mathrm{d}\mu\ \int_{\mathbb{R}}\mathbb{P}\left(R_{\ell_R,1},\ldots,R_{R,\ell_{n_{\ell,R}}}\,|\,\mu,\sigma^2\right)\mathbb{P}\left(\mu\right)\,\mathrm{d}\mu}{\int_{\mathbb{R}}\mathbb{P}\left(R_{\ell_1},\ldots,R_{\ell_{n_\ell}}\,|\,\mu,\sigma^2\right)\mathbb{P}\left(\mu\right)\,\mathrm{d}\mu}$$

$$=\ \frac{\frac{1}{(2\pi\sigma^2)^{n_\ell L/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_{\ell_L}}R_{\ell_{Li}}^2\right)\left(\int_{\mathbb{R}}g_{\ell_L}\mathbb{P}\left(\mu\right)\,\mathrm{d}\mu\right)\ \frac{1}{(2\pi\sigma^2)^{n_\ell R/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_{\ell_R}}R_{\ell_{Ri}}^2\right)\left(\int_{\mathbb{R}}g_{\ell_R}\mathbb{P}\left(\mu\right)\,\mathrm{d}\mu\right)}{\frac{1}{(2\pi\sigma^2)^{n_\ell/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_\ell}R_{\ell_i}^2\right)\left(\int_{\mathbb{R}}g_\ell\mathbb{P}\left(\mu\right)\,\mathrm{d}\mu\right)}$$

$$=\ \frac{\cancel{\frac{1}{(2\pi\sigma^2)^{n_\ell L/2}}}\cancel{\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_{\ell_L}}R_{\ell_{Li}}^2\right)}\left(\int_{\mathbb{R}}g_{\ell_L}\mathbb{P}\left(\mu\right)\,\mathrm{d}\mu\right)\ \cancel{\frac{1}{(2\pi\sigma^2)^{n_\ell R/2}}}\cancel{\exp\left(-\frac{1}{2\sigma^2}\right)\sum_{i=1}^{n_{\ell_R}}R_{\ell_{Ri}}^2}\left(\int_{\mathbb{R}}g_{\ell_R}\mathbb{P}\left(\mu\right)\,\mathrm{d}\mu\right)}{\cancel{\frac{1}{(2\pi\sigma^2)^{n_\ell/2}}}\cancel{\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_\ell}R_{\ell_i}^2\right)}\left(\int_{\mathbb{R}}g_\ell\mathbb{P}\left(\mu\right)\,\mathrm{d}\mu\right)}$$

$$=\ \frac{\int_{\mathbb{R}}g_{\ell_L}\mathbb{P}\left(\mu\right)\,\mathrm{d}\mu\int_{\mathbb{R}}g_{\ell_R}\mathbb{P}\left(\mu\right)\,\mathrm{d}\mu}{\int_{\mathbb{R}}g_\ell\mathbb{P}\left(\mu\right)\,\mathrm{d}\mu}$$

$$= \frac{\int_{\mathbb{R}} g_{\ell_L} \frac{1}{\sqrt{2\pi\sigma_\mu^2}} \exp\left(-\frac{1}{2\sigma_\mu^2}\mu^2\right) \mathrm{d}\mu \int_{\mathbb{R}} g_{\ell_R} \frac{1}{\sqrt{2\pi\sigma_\mu^2}} \exp\left(-\frac{1}{2\sigma_\mu^2}\mu^2\right) \mathrm{d}\mu}{\int_{\mathbb{R}} g_\ell \frac{1}{\sqrt{2\pi\sigma_\mu^2}} \exp\left(-\frac{1}{2\sigma_\mu^2}\mu^2\right) \mathrm{d}\mu}$$

$$= \frac{1}{\sqrt{2\pi\sigma_\mu^2}} \frac{\int_{\mathbb{R}} \exp\left(-\frac{1}{2\sigma^2}\left(-2n_{\ell_L}\mu\bar{R}_{\ell_L} + n_{\ell_L}\mu^2\right)\right) \exp\left(-\frac{1}{2\sigma_\mu^2}\mu^2\right) \mathrm{d}\mu \int_{\mathbb{R}} \exp\left(-\frac{1}{2\sigma^2}\left(-2n_{\ell_L}\mu\bar{R}_{\ell_L} + n_{\ell_L}\mu^2\right)\right) \exp\left(-\frac{1}{2\sigma_\mu^2}\mu^2\right) \mathrm{d}\mu}{\underbrace{\int_{\mathbb{R}} \exp\left(-\frac{1}{2\sigma^2}\left(-2n_\ell\mu\bar{R}_\ell + n_\ell\mu^2\right)\right) \exp\left(-\frac{1}{2\sigma_\mu^2}\mu^2\right) \mathrm{d}\mu}}$$

Let's evaluate the underbraced integral using Mathematica:

$$\int_{\mathbb{R}} g(\bar{R}_\ell \mid \mu, \sigma^2)\mathbb{P}\left(\mu; \sigma_\mu^2\right) d\mu = \int_{\mathbb{R}} \exp\left(-\frac{1}{2\sigma^2}\left(-2n_\ell\mu\bar{R}_\ell + n_\ell\mu^2\right)\right) \exp\left(-\frac{1}{2\sigma_\mu^2}\mu^2\right) d\mu$$

$$= \sqrt{\frac{2\pi}{\frac{n_\ell}{\sigma^2} + \frac{1}{\sigma_\mu^2}}} \exp\left(\frac{n_\ell^2 \bar{R}_\ell^2 \sigma_\mu^2}{2\sigma^2\left(\sigma^2 + n_\ell\sigma_\mu^2\right)}\right)$$

So now we have:

$$\frac{\mathbb{P}\left(\boldsymbol{R} \mid T^*, \sigma^2\right)}{\mathbb{P}\left(\boldsymbol{R} \mid T, \sigma^2\right)}$$

$$= \frac{1}{\sqrt{2\pi\sigma_\mu^2}} \frac{\sqrt{\frac{2\pi}{\frac{n_{\ell_L}}{\sigma^2} + \frac{1}{\sigma_\mu^2}}} \exp\left(\frac{n_{\ell_L}^2 \bar{R}_{\ell_L}^2 \sigma_\mu^2}{2\sigma^2\left(\sigma^2 + n_{\ell_L}\sigma_\mu^2\right)}\right) \sqrt{\frac{2\pi}{\frac{n_{\ell_R}}{\sigma^2} + \frac{1}{\sigma_\mu^2}}} \exp\left(\frac{n_{\ell_R}^2 \bar{R}_{\ell_R}^2 \sigma_\mu^2}{2\sigma^2\left(\sigma^2 + n_{\ell_R}\sigma_\mu^2\right)}\right)}{\sqrt{\frac{2\pi}{\frac{n_\ell}{\sigma^2} + \frac{1}{\sigma_\mu^2}}} \exp\left(\frac{n_\ell^2 \bar{R}_\ell^2 \sigma_\mu^2}{2\sigma^2\left(\sigma^2 + n_\ell\sigma_\mu^2\right)}\right)}$$

$$= \sqrt{\frac{\frac{n_\ell}{\sigma^2} + \frac{1}{\sigma_\mu^2}}{\sigma_\mu^2\left(\frac{n_{\ell_L}}{\sigma^2} + \frac{1}{\sigma_\mu^2}\right)\left(\frac{n_{\ell_R}}{\sigma^2} + \frac{1}{\sigma_\mu^2}\right)}} \exp\left(\frac{\sigma_\mu^2}{2\sigma^2}\left(\frac{n_{\ell_L}^2 \bar{R}_{\ell_L}^2}{\sigma^2 + n_{\ell_L}\sigma_\mu^2} + \frac{n_{\ell_R}^2 \bar{R}_{\ell_R}^2}{\sigma^2 + n_{\ell_R}\sigma_\mu^2} - \frac{n_\ell^2 \bar{R}_\ell^2}{\sigma^2 + n_\ell\sigma_\mu^2}\right)\right)$$

$$= \sqrt{\frac{\sigma^2\left(\sigma^2 + n_\ell\sigma_\mu^2\right)}{\left(\sigma^2 + n_{\ell_L}\sigma_\mu^2\right)\left(\sigma^2 + n_{\ell_R}\sigma_\mu^2\right)}} \exp\left(\frac{\sigma_\mu^2}{2\sigma^2}\left(\frac{\left(\sum_{i=1}^{n_{\ell_L}} R_{L,i}\right)^2}{\sigma^2 + n_{\ell_L}\sigma_\mu^2} + \frac{\left(\sum_{i=1}^{n_{\ell_R}} R_{R,i}\right)^2}{\sigma^2 + n_{\ell_R}\sigma_\mu^2} - \frac{\left(\sum_{i=1}^{n_\ell} R_{\ell,i}\right)^2}{\sigma^2 + n_\ell\sigma_\mu^2}\right)\right)$$

Strategy #3 — Completing the Square

Recall that:

$$\mathbb{P}\left(R_{\ell_1}, \ldots, R_{\ell_{n_\ell}} \mid \mu, \sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{n_\ell/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_\ell}\left(R_{\ell_i} - \mu\right)^2\right)$$

$$= \frac{1}{(2\pi\sigma^2)^{n_\ell/2}} \exp\left(-\frac{1}{2\sigma^2}\left(n_\ell\left(\bar{R}_\ell - \mu\right)^2 + \sum_{i=1}^{n_\ell}\left(R_{\ell_i} - \bar{R}_\ell\right)^2\right)\right)$$

9

$$= \frac{1}{(2\pi\sigma^2)^{n_\ell/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_\ell}\left(R_{\ell_i}-\bar{R}_\ell\right)^2\right)\exp\left(-\frac{n_\ell}{2\sigma^2}\left(\bar{R}_\ell-\mu\right)^2\right)$$

$$= \frac{1}{(2\pi\sigma^2)^{n_\ell/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_\ell}\left(R_{\ell_i}-\bar{R}_\ell\right)^2\right)\exp\left(-\frac{n_\ell}{2\sigma^2}\left(\bar{R}_\ell^2-2\bar{R}_\ell\mu+\mu^2\right)\right)$$

which means that:

$$\mathbb{P}\left(R_{\ell_1},\ldots,R_{\ell_{n_\ell}}\,\big|\,\sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{n_\ell/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_\ell}\left(R_{\ell_i}-\bar{R}_\ell\right)^2\right)\times$$

$$\int_{\mathbb{R}}\exp\left(-\frac{n_\ell}{2\sigma^2}\left(\bar{R}_\ell^2-2\bar{R}_\ell\mu+\mu^2\right)\right)\frac{1}{\sqrt{2\pi\sigma_\mu^2}}\exp\left(-\frac{1}{2\sigma_\mu^2}\mu^2\right)\,\mathrm{d}\mu$$

$$= \frac{1}{(2\pi\sigma^2)^{n_\ell/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_\ell}\left(R_{\ell_i}-\bar{R}_\ell\right)^2\right)\times$$

$$\frac{1}{\sqrt{2\pi\sigma_\mu^2}}\int_{\mathbb{R}}\exp\left(-\frac{n_\ell}{2\sigma^2}\left(\bar{R}_\ell^2-2\bar{R}_\ell\mu+\mu^2\right)-\frac{1}{2\sigma_\mu^2}\mu^2\right)\,\mathrm{d}\mu$$

$$= \frac{1}{(2\pi\sigma^2)^{n_\ell/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_\ell}\left(R_{\ell_i}-\bar{R}_\ell\right)^2\right)\times$$

$$\frac{1}{\sqrt{2\pi\sigma_\mu^2}}\int_{\mathbb{R}}\exp\left(\underbrace{\left(-\frac{n_\ell}{2\sigma^2}-\frac{1}{2\sigma_\mu^2}\right)\mu^2+\left(\frac{\bar{R}_\ell n_\ell}{\sigma^2}\right)\mu-\frac{n_\ell\bar{R}_\ell^2}{2\sigma^2}}\right)\,\mathrm{d}\mu$$

Now let's complete the square of the underbraced above. We want it to look like $c\left(d-\mu\right)^2 = c\mu^2-2cd\mu+cd^2$. Already we have c, the factor multiplying μ^2, so then d becomes (using the μ term above):

$$d = \frac{\bar{R}_\ell n_\ell}{-2c\sigma^2} = \frac{\bar{R}_\ell n_\ell}{-2\left(-\frac{n_\ell}{2\sigma^2}-\frac{1}{2\sigma_\mu^2}\right)\sigma^2} = \frac{\bar{R}_\ell n_\ell}{n_\ell+\frac{\sigma^2}{\sigma_\mu^2}}$$

and the final cd^2 term becomes:

$$cd^2 = c\left(\frac{\bar{R}_\ell n_\ell}{-2c\sigma^2}\right)^2 = \frac{\bar{R}_\ell^2 n_\ell^2}{4c\sigma^4} = \frac{\bar{R}_\ell^2 n_\ell^2}{4\left(-\frac{n_\ell}{2\sigma^2}-\frac{1}{2\sigma_\mu^2}\right)\sigma^4} = \frac{\bar{R}_\ell^2 n_\ell^2}{-2\left(\frac{n_\ell}{\sigma^2}+\frac{1}{\sigma_\mu^2}\right)\sigma^4} = \frac{\bar{R}_\ell^2 n_\ell^2}{-2\sigma^2\left(n_\ell+\frac{\sigma^2}{\sigma_\mu^2}\right)}$$

Returning to our margined expression, we subtract and add the cd^2 term inside the exponentiation:

$$\mathbb{P}\left(R_{\ell_1},\ldots,R_{\ell_{n_\ell}}\,\big|\,\sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{n_\ell/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_\ell}\left(R_{\ell_i}-\bar{R}_\ell\right)^2\right)\frac{1}{\sqrt{2\pi\sigma_\mu^2}}\times$$

$$\int_{\mathbb{R}} \exp\left(\underbrace{\left(-\frac{n_\ell}{2\sigma^2} - \frac{1}{2\sigma_\mu^2}\right)\mu^2 + \left(\frac{\bar{R}_\ell}{\sigma^2}\right)\mu - \frac{\bar{R}_\ell^2 n_\ell^2}{2\sigma^2\left(n_\ell + \frac{\sigma^2}{\sigma_\mu^2}\right)}} + \frac{\bar{R}_\ell^2 n_\ell^2}{2\sigma^2\left(n_\ell + \frac{\sigma^2}{\sigma_\mu^2}\right)} - \frac{n_\ell \bar{R}_\ell^2}{2\sigma^2}\right)\mathrm{d}\mu$$

So the underbraced is the "square" and we pull the last two terms out (because they are not functions of μ) to arrive at:

$$\mathbb{P}\left(R_{\ell_1},\ldots,R_{\ell_{n_\ell}}\,\middle|\,\sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{n_\ell/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_\ell}\left(R_{\ell_i} - \bar{R}_\ell\right)^2\right)\frac{1}{\sqrt{2\pi\sigma_\mu^2}}\times$$

$$\exp\left(\frac{\bar{R}_\ell^2 n_\ell^2}{2\sigma^2\left(n_\ell + \frac{\sigma^2}{\sigma_\mu^2}\right)} - \frac{n_\ell \bar{R}_\ell^2}{2\sigma^2}\right)\int_{\mathbb{R}}\exp\left(\left(\underbrace{-\frac{n_\ell}{2\sigma^2} - \frac{1}{2\sigma_\mu^2}}\right)\left(\mu - \frac{\bar{R}_\ell}{n_\ell + \frac{\sigma^2}{\sigma_\mu^2}}\right)^2\right)\mathrm{d}\mu$$

Note how the integral is now the Gaussian integral with the $-\frac{1}{2\sigma_*^2}$ parameter underbraced. Hence it will integrate to $\sqrt{2\pi\sigma_*^2}$ and the expression becomes:

$$\mathbb{P}\left(R_{\ell_1},\ldots,R_{\ell_{n_\ell}}\,\middle|\,\sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{n_\ell/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_\ell}\left(R_{\ell_i} - \bar{R}_\ell\right)^2\right)\times$$

$$\frac{1}{\sqrt{2\pi\sigma_\mu^2}}\exp\left(\frac{\bar{R}_\ell^2 n_\ell^2}{2\sigma^2\left(n_\ell + \frac{\sigma^2}{\sigma_\mu^2}\right)} - \frac{n_\ell \bar{R}_\ell^2}{2\sigma^2}\right)\sqrt{2\pi\frac{1}{\frac{n_\ell}{\sigma^2} + \frac{1}{\sigma_\mu^2}}}$$

$$= \frac{1}{(2\pi\sigma^2)^{n_\ell/2}}\sqrt{\frac{\sigma^2}{\sigma^2 + n_\ell\sigma_\mu^2}}\times$$

$$\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_\ell}\left(R_{\ell_i} - \bar{R}_\ell\right)^2\right)\exp\left(\frac{\bar{R}_\ell^2 n_\ell^2}{2\sigma^2\left(n_\ell + \frac{\sigma^2}{\sigma_\mu^2}\right)} - \frac{n_\ell \bar{R}_\ell^2}{2\sigma^2}\right)$$

So now we're ready to calculate the likelihood ratio.

$$\frac{\mathbb{P}\left(\boldsymbol{R}\,\middle|\,T^*,\sigma^2\right)}{\mathbb{P}\left(\boldsymbol{R}\,\middle|\,T,\sigma^2\right)}$$

$$= \frac{\mathbb{P}\left(R_{\ell_{L,1}},\ldots,R_{\ell_{L,n_{\ell,L}}}\,\middle|\,\sigma^2\right)\mathbb{P}\left(R_{\ell_{R,1}},\ldots,R_{R,\ell_{n_{\ell,R}}}\,\middle|\,\sigma^2\right)}{\mathbb{P}\left(R_{\ell_1},\ldots,R_{\ell_{n_\ell}}\,\middle|\,\sigma^2\right)}$$

$$= \frac{\frac{1}{\cancel{(2\pi\sigma^2)^{n_{\ell_L}/2}}}\exp\left(\frac{\bar{R}_{\ell_L}^2 n_\ell^2}{2\sigma^2\left(n_{\ell_L} + \frac{\sigma^2}{\sigma_\mu^2}\right)} - \frac{n_{\ell_L}\bar{R}_{\ell_L}^2}{2\sigma^2}\right)\sqrt{\frac{\cancel{\sigma^2}}{\sigma^2 + n_{\ell_L}\sigma_\mu^2}}\frac{1}{\cancel{(2\pi\sigma^2)^{n_{\ell_R}/2}}}\exp\left(\frac{\bar{R}_{\ell_R}^2 n_\ell^2}{2\sigma^2\left(n_{\ell_R} + \frac{\sigma^2}{\sigma_\mu^2}\right)} - \frac{n_{\ell_R}\bar{R}_{\ell_R}^2}{2\sigma^2}\right)\sqrt{\frac{\sigma^2}{\sigma^2 + n_{\ell_R}\sigma_\mu^2}}}{\frac{1}{\cancel{(2\pi\sigma^2)^{n_\ell/2}}}\exp\left(\frac{\bar{R}_\ell^2 n_\ell^2}{2\sigma^2\left(n_\ell + \frac{\sigma^2}{\sigma_\mu^2}\right)} - \frac{n_\ell\bar{R}_\ell^2}{2\sigma^2}\right)\sqrt{\frac{\cancel{\sigma^2}}{\sigma^2 + n_\ell\sigma_\mu^2}}}\times$$

$$\exp\left(-\frac{1}{2\sigma^2}\left(\underbrace{\sum_{i=1}^{n_{\ell_L}}\left(R_{\ell_L,i}-\bar{R}_{\ell_L}\right)^2+\sum_{i=1}^{n_{\ell_R}}\left(R_{\ell_R,i}-\bar{R}_{\ell,R}\right)^2-\sum_{i=1}^{n_\ell}\left(R_{\ell_i}-\bar{R}_\ell\right)^2}\right)\right)$$

We've calculate the underbraced before in the strategy #1 section. We now substitute, factor, and collect terms:

$$=\sqrt{\underbrace{\frac{\sigma^2\left(\sigma^2+n_\ell\sigma_\mu^2\right)}{\left(\sigma^2+n_{\ell_L}\sigma_\mu^2\right)\left(\sigma^2+n_{\ell_R}\sigma_\mu^2\right)}}_{c}}\times$$

$$\exp\left(\frac{1}{2\sigma^2}\left(\frac{\bar{R}_{\ell_L}^2 n_{\ell_L}^2}{\left(n_{\ell_L}+\frac{\sigma^2}{\sigma_\mu^2}\right)}-n_{\ell_L}\bar{R}_{\ell_L}^2+\frac{\bar{R}_{\ell_R}^2 n_{\ell_R}^2}{\left(n_{\ell_R}+\frac{\sigma^2}{\sigma_\mu^2}\right)}-n_{\ell_R}\bar{R}_{\ell_R}^2-\frac{\bar{R}_\ell^2 n_\ell^2}{\left(n_\ell+\frac{\sigma^2}{\sigma_\mu^2}\right)}+n_\ell\bar{R}_\ell^2\right)\right)\times$$

$$\exp\left(\frac{1}{2\sigma^2}\left(n_{\ell_L}\bar{R}_{\ell_L}^2+n_\ell\bar{R}_\ell^2-n_\ell\bar{R}_\ell^2\right)\right)$$

$$=\quad c\,\exp\left(\frac{\sigma_\mu^2}{2\sigma^2}\left(\left(\frac{n_{\ell_L}^2}{\sigma_\mu^2 n_{\ell_L}+\sigma^2}\right)\bar{R}_{\ell_L}^2+\left(\frac{n_{\ell_R}^2}{\sigma_\mu^2 n_{\ell_R}+\sigma^2}\right)\bar{R}_{\ell_R}^2-\left(\frac{n_\ell^2}{\sigma_\mu^2 n_\ell+\sigma^2}\right)\bar{R}_\ell^2\right)\right)$$

$$=\quad\sqrt{\frac{\sigma^2\left(\sigma^2+n_\ell\sigma_\mu^2\right)}{\left(\sigma^2+n_{\ell_L}\sigma_\mu^2\right)\left(\sigma^2+n_{\ell_R}\sigma_\mu^2\right)}}\quad\exp\left(\frac{\sigma_\mu^2}{2\sigma^2}\left(\frac{\left(\sum_{i=1}^{n_{\ell_L}}R_{\ell_L,i}\right)^2}{\sigma^2+n_{\ell_L}\sigma_\mu^2}+\frac{\left(\sum_{i=1}^{n_{\ell_R}}R_{\ell_R,i}\right)^2}{\sigma^2+n_{\ell_R}\sigma_\mu^2}-\frac{\left(\sum_{i=1}^{n_\ell}R_{\ell,i}\right)^2}{\sigma^2+n_\ell\sigma_\mu^2}\right)\right)$$

We've now computed this using three different methods, so it must be correct?

In log form, this becomes:

$$\frac{1}{2}\left(\ln\left(\sigma^2\right)+\ln\left(\sigma^2+n_\ell\sigma_\mu^2\right)-\ln\left(\sigma^2+n_{\ell_L}\sigma_\mu^2\right)-\ln\left(\sigma^2+n_{\ell_R}\sigma_\mu^2\right)\right)+$$

$$\frac{\sigma_\mu^2}{2\sigma^2}\left(\frac{\left(\sum_{i=1}^{n_{\ell_L}}R_{\ell_L,i}\right)^2}{\sigma^2+n_{\ell_L}\sigma_\mu^2}+\frac{\left(\sum_{i=1}^{n_{\ell_R}}R_{\ell_R,i}\right)^2}{\sigma^2+n_{\ell_R}\sigma_\mu^2}-\frac{\left(\sum_{i=1}^{n_\ell}R_{\ell,i}\right)^2}{\sigma^2+n_\ell\sigma_\mu^2}\right)$$

Tree Structure Ratio

Remember the prior form on trees:

$$\mathbb{P}\left(T\right)\quad=\quad\prod_{\eta\in H_{\text{leaves}}}\left(1-\mathbb{P}_{\text{SPLIT}}\left(\eta\right)\right)\prod_{\eta\in H_{\text{internals}}}\mathbb{P}_{\text{SPLIT}}\left(\eta\right)\prod_{\eta\in H_{\text{internals}}}\mathbb{P}_{\text{RULE}}\left(\eta\right)$$

Remember from CGM98 that the probability of splitting on a given node η is driven by two hyperparameters, α and β in the following way where d_η is the depth of the η node:

$$\mathbb{P}_{\text{SPLIT}}\left(\eta\right) = \frac{\alpha}{\left(1 + d_\eta\right)^\beta}$$

The probability of assigning the specific rule, $\mathbb{P}\left(\text{rule } \eta\right)$, is just picking from all available attributes and then from all available unique splits with the modifications discussed in the previous section. Once again, the proposal tree differs from the original tree by only the ℓth node in the original becoming the ℓ_L and ℓ_R nodes in the proposal. This simplifies the tree structure ratio to just:

$$
\begin{aligned}
\frac{\mathbb{P}\left(T^*\right)}{\mathbb{P}\left(T\right)} &= \frac{\left(1 - \mathbb{P}_{\text{SPLIT}}\left(\eta_L\right)\right)\left(1 - \mathbb{P}_{\text{SPLIT}}\left(\eta_R\right)\right)\mathbb{P}_{\text{SPLIT}}\left(\eta\right)\mathbb{P}_{\text{RULE}}\left(\eta\right)}{\left(1 - \mathbb{P}_{\text{SPLIT}}\left(\eta\right)\right)} \\[2mm]
&= \frac{\left(1 - \dfrac{\alpha}{\left(1 + d_{\eta_L}\right)^\beta}\right)\left(1 - \dfrac{\alpha}{\left(1 + d_{\eta_R}\right)^\beta}\right)\dfrac{\alpha}{\left(1 + d_\eta\right)^\beta}\dfrac{1}{p_{\text{adj}}(\eta)}\dfrac{1}{n_{j \cdot \text{adj}}(\eta)}}{1 - \dfrac{\alpha}{\left(1 + d_\eta\right)^\beta}} \\[2mm]
&= \alpha\frac{\left(1 - \frac{\alpha}{(2 + d_\eta)^\beta}\right)^2}{\left(\left(1 + d_\eta\right)^\beta - \alpha\right)p_{\text{adj}}(\eta)n_{j \cdot \text{adj}}(\eta)}
\end{aligned}
$$

With the last line following from the fact that the depth of the child nodes is just the depth of the parent node incremented by 1. In log form this becomes:

$$\ln\left(\alpha\right) + 2\ln\left(1 - \frac{\alpha}{\left(2 + d_\eta\right)^\beta}\right) - \ln\left(\left(1 + d_\eta\right)^\beta - \alpha\right) - \ln\left(p_{\text{adj}}(\eta)\right) - \ln\left(n_{j \cdot \text{adj}}(\eta)\right)$$

Now we have a way of calculating r for grow proposals by multiplying all three above results.

Prune Proposal

Transition Ratio

For prune proposals, we move in the opposite direction. We need to hack off a node:

$$
\begin{aligned}
\mathbb{P}\left(T \to T^*\right) &= \mathbb{P}\left(\text{PRUNE}\right)\mathbb{P}\left(\text{selecting the } \ell\text{th node to prune from}\right) \\
&= \mathbb{P}\left(\text{PRUNE}\right)\frac{1}{w_2}
\end{aligned}
$$

To go the opposite direction, we need to make sure we grow the exact same node so $p_{\text{adj}}(\eta), n_{j \cdot \text{adj}}(\eta)$ need to be calculated based on whatever the ℓth node originally was:

$$
\begin{aligned}
\mathbb{P}\left(T^* \to T\right) &= \mathbb{P}\left(\text{GROW}\right)\mathbb{P}\left(\text{selecting the } \ell\text{th node to grow from}\right) \times \\
&\quad \mathbb{P}\left(\text{selecting the original attribute to split on}\right) \times
\end{aligned}
$$

13

$$\mathbb{P}\left(\text{selecting the original value to split on}\right)$$
$$= \ \mathbb{P}\left(\text{GROW}\right)\frac{1}{b-1}\frac{1}{p_{\text{adj}}(\eta)}\frac{1}{n_{j\cdot\text{adj}}(\eta)}$$

We're using $b-1$ here because the proposed tree has one less terminal node due to the pruning than the original tree T had.

Thus, the transition ratio becomes:

$$\frac{\mathbb{P}\left(T^* \to T\right)}{\mathbb{P}\left(T \to T^*\right)} = \frac{\mathbb{P}\left(\text{GROW}\right)\frac{1}{b-1}\frac{1}{p_{\text{adj}}(\eta^*)}\frac{1}{n_{j\cdot\text{adj}}(\eta^*)}}{\mathbb{P}\left(\text{PRUNE}\right)\frac{1}{w_2}} = \frac{\cancel{\mathbb{P}\left(\text{GROW}\right)}^{\,?}}{\cancel{\mathbb{P}\left(\text{PRUNE}\right)}}\frac{w_2}{(b-1)p_{\text{adj}}(\eta^*)n_{j\cdot\text{adj}}(\eta^*)}$$

and in log form this becomes:

$$\ln\left(w_2\right) - \ln\left(b-1\right) - \ln\left(p_{\text{adj}}(\eta^*)\right) - \ln\left(n_{j\cdot\text{adj}}(\eta^*)\right)$$

Once again, we need to bookkeep and make sure we can actually prune this node. If it's just a root node, then we can't even consider this step at all. Otherwise, they'll cancel.

Likelihood Ratio

It is pretty obvious this is the inverse of the grow step's likelihood ratio. Now the tree proposal has just one collapsed node where the original has a left and right component:

$$\frac{\mathbb{P}\left(\boldsymbol{R}\,\middle|\,T^*,\sigma^2\right)}{\mathbb{P}\left(\boldsymbol{R}\,\middle|\,T,\sigma^2\right)}$$

$$= \left(\sqrt{\frac{\sigma^2\left(\sigma^2 + n_\ell\sigma_\mu^2\right)}{\left(\sigma^2 + n_{\ell_L}\sigma_\mu^2\right)\left(\sigma^2 + n_{\ell_R}\sigma_\mu^2\right)}}\ \exp\left(\frac{\sigma_\mu^2}{2\sigma^2}\left(\frac{\left(\sum_{i=1}^{n_{\ell_L}}R_{\ell_L,i}\right)^2}{\sigma^2 + n_{\ell_L}\sigma_\mu^2} + \frac{\left(\sum_{i=1}^{n_{\ell_R}}R_{\ell_R,i}\right)^2}{\sigma^2 + n_{\ell_R}\sigma_\mu^2} - \frac{\left(\sum_{i=1}^{n_\ell}R_{\ell,i}\right)^2}{\sigma^2 + n_\ell\sigma_\mu^2}\right)\right)\right)^{-1}$$

$$= \sqrt{\frac{\left(\sigma^2 + n_{\ell_L}\sigma_\mu^2\right)\left(\sigma^2 + n_{\ell_R}\sigma_\mu^2\right)}{\sigma^2\left(\sigma^2 + n_\ell\sigma_\mu^2\right)}}\ \exp\left(\frac{\sigma_\mu^2}{2\sigma^2}\left(\frac{\left(\sum_{i=1}^{n_\ell}R_{\ell,i}\right)^2}{\sigma^2 + n_\ell\sigma_\mu^2} - \frac{\left(\sum_{i=1}^{n_{\ell_L}}R_{\ell_L,i}\right)^2}{\sigma^2 + n_{\ell_L}\sigma_\mu^2} - \frac{\left(\sum_{i=1}^{n_{\ell_R}}R_{\ell_R,i}\right)^2}{\sigma^2 + n_{\ell_R}\sigma_\mu^2}\right)\right)$$

and in log form:

$$\frac{1}{2}\left(\ln\left(\sigma^2 + n_{\ell_L}\sigma_\mu^2\right) + \ln\left(\sigma^2 + n_{\ell_R}\sigma_\mu^2\right) - \ln\left(\sigma^2\right) - \ln\left(\sigma^2 + n_\ell\sigma_\mu^2\right)\right) +$$
$$\frac{\sigma_\mu^2}{2\sigma^2}\left(\frac{\left(\sum_{i=1}^{n_\ell}R_{\ell,i}\right)^2}{\sigma^2 + n_\ell\sigma_\mu^2} - \frac{\left(\sum_{i=1}^{n_{\ell_L}}R_{\ell_L,i}\right)^2}{\sigma^2 + n_{\ell_L}\sigma_\mu^2} - \frac{\left(\sum_{i=1}^{n_{\ell_R}}R_{\ell_R,i}\right)^2}{\sigma^2 + n_{\ell_R}\sigma_\mu^2}\right)$$

Tree Structure Ratio

It is also clear this is just the inverse of the tree structure ratio for the grow step.

$$
\begin{aligned}
\frac{\mathbb{P}\left(T^{*}\right)}{\mathbb{P}\left(T\right)} &= \left(\alpha \frac{\left(1 - \frac{\alpha}{(2+d_\eta)^\beta}\right)^2}{\left((1+d_\eta)^\beta - \alpha\right) p_{\mathrm{adj}}(\eta^*) n_{j \cdot \mathrm{adj}}(\eta^*)}\right)^{-1} \\
&= \frac{\left((1+d_\eta)^\beta - \alpha\right) p_{\mathrm{adj}}(\eta^*) n_{j \cdot \mathrm{adj}}(\eta^*)}{\alpha \left(1 - \frac{\alpha}{(2+d_\eta)^\beta}\right)^2}
\end{aligned}
$$

In log form this becomes:

$$
\ln\left((1+d_\eta)^\beta - \alpha\right) + \ln\left(p_{\mathrm{adj}}(\eta^*)\right) + \ln\left(n_{j \cdot \mathrm{adj}}(\eta^*)\right) - \ln(\alpha) - 2\ln\left(1 - \frac{\alpha}{(2+d_\eta)^\beta}\right)
$$

Change (for any internal node)

Transition Ratio

We compute the probabilities for changing the η_* node which then affects all children. Note that the number of split points is different now because we cannot make splits which ostensibly render child nodes empty. This means that if there's a child on the left with a certain split rule for the variable selected, we can only have split points *higher* than that rule; if there's a child on the right with a certain split rule for the variable selected, we can only have split points *lower* than that rule.

$$
\begin{aligned}
\mathbb{P}\left(T^* \to T\right) &= \mathbb{P}\left(\mathrm{CHANGE}\right) \frac{1}{|H_{\mathrm{internals}}|} \frac{1}{p_{\mathrm{adj}}(\eta^*)} \frac{1}{n_{j \cdot \mathrm{adj} \cdot \mathrm{ch}}(\eta^*)} \\
\mathbb{P}\left(T \to T^*\right) &= \mathbb{P}\left(\mathrm{CHANGE}\right) \frac{1}{|H_{\mathrm{internals}}|} \frac{1}{p_{\mathrm{adj}}(\eta)} \frac{1}{n_{j \cdot \mathrm{adj} \cdot \mathrm{ch}}(\eta)}
\end{aligned}
$$

Everything cancels except for the ratio of n's.

$$
\frac{\mathbb{P}\left(T^* \to T\right)}{\mathbb{P}\left(T \to T^*\right)} = \frac{n_{j \cdot \mathrm{adj} \cdot \mathrm{ch}}(\eta^*)}{n_{j \cdot \mathrm{adj} \cdot \mathrm{ch}}(\eta)}
$$

and in log form:

$$
\ln\left(n_{j \cdot \mathrm{adj} \cdot \mathrm{ch}}(\eta^*)\right) - \ln\left(n_{j \cdot \mathrm{adj} \cdot \mathrm{ch}}(\eta)\right)
$$

15

Likelihood Ratio

Recall the following result which we can further simplify:

$$
\begin{aligned}
\mathbb{P}\left(R_{\ell_1},\ldots,R_{\ell_{n_\ell}}\,\middle|\,\sigma^2\right) &= \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_\ell}\left(R_{\ell_i}-\bar{R}_\ell\right)^2\right)\exp\left(\frac{\bar{R}_\ell^2 n_\ell^2}{2\sigma^2\left(n_\ell+\frac{\sigma^2}{\sigma_\mu^2}\right)}-\frac{n_\ell\bar{R}_\ell^2}{2\sigma^2}\right) \\[2mm]
&= \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n_\ell}\left(R_{\ell_i}-\bar{R}_\ell\right)^2-\frac{\bar{R}_\ell^2 n_\ell^2}{n_\ell+\frac{\sigma^2}{\sigma_\mu^2}}+n_\ell\bar{R}_\ell^2\right)\right) \\[2mm]
&= \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n_\ell}\left(R_{\ell_i}\right)^2-2n_\ell\bar{R}_\ell^2+n_\ell\bar{R}_\ell^2-\frac{\bar{R}_\ell^2 n_\ell^2}{n_\ell+\frac{\sigma^2}{\sigma_\mu^2}}+n_\ell\bar{R}_\ell^2\right)\right) \\[2mm]
&= \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n_\ell}\left(R_{\ell_i}\right)^2-\frac{\bar{R}_\ell^2 n_\ell^2}{n_\ell+\frac{\sigma^2}{\sigma_\mu^2}}\right)\right)
\end{aligned}
$$

Now, we need to account for the likelihood ratio between the leaves that changed *i.e.* the change leaf and everything below (which we call $\eta_* \in H_{\text{leaves}}^*$ in this case):

$$
\frac{\mathbb{P}\left(\boldsymbol{R}\,\middle|\,T^*,\sigma^2\right)}{\mathbb{P}\left(\boldsymbol{R}\,\middle|\,T,\sigma^2\right)} = \frac{\prod_{\eta_*\in H_{\text{leaves}\cdot\text{below}}^*}\mathbb{P}\left(R_{\eta_*,1},\ldots,R_{\eta_*,n_{\eta_*}}\,\middle|\,\sigma^2\right)}{\prod_{\eta\in H_{\text{leaves}\cdot\text{below}}}\mathbb{P}\left(R_{\eta,1},\ldots,R_{\eta,n_\eta}\,\middle|\,\sigma^2\right)}
$$

In two equations above, we have a sum of R_i^2 term. Note that these R_i's have to be apportioned among the leaves. During a change, they'll go to different leaves, but if you aggregate the terms, they'll still sum up to the total sum of squared. Hence, that term will cancel, and you are left with a ratio that looks like something below (in log form):

$$
\frac{1}{2\sigma^2}\left(\sum_{\eta_*\ \in\ H_{\text{leaves}\cdot\text{below}}^*}\frac{\left(\sum_{i=1}^{n_{\eta_*}}R_{n_{\eta_*},i}\right)^2}{n_{\eta_*}+\frac{\sigma^2}{\sigma_\mu^2}} - \sum_{\eta\ \in\ H_{\text{leaves}\cdot\text{below}}}\frac{\left(\sum_{i=1}^{n_\eta}R_{n_\eta,i}\right)^2}{n_\eta+\frac{\sigma^2}{\sigma_\mu^2}}\right)
$$

Tree Structure Ratio

Since both the proposal and original tree have the same structure, we just have to examine the rules:

$$
\frac{\mathbb{P}\left(T^*\right)}{\mathbb{P}\left(T\right)} = \frac{\prod_{\eta_*\ \in\ H_{\text{leaves}\cdot\text{below}}^*}\mathbb{P}_{\text{RULE}}\left(\eta_*\right)}{\prod_{\eta\ \in\ H_{\text{leaves}\cdot\text{below}}}\mathbb{P}_{\text{RULE}}\left(\eta\right)} = \frac{\prod_{\eta\ \in\ H_{\text{leaves}\cdot\text{below}}}p_{\text{adj}}(\eta)n_{j\cdot\text{adj}}(\eta)}{\prod_{\eta_*\ \in\ H_{\text{leaves}\cdot\text{below}}^*}p_{\text{adj}}(\eta^*)n_{j\cdot\text{adj}}(\eta^*)}
$$

Change (for just a singly internal node)

Change would be picking one father node (a node with two leaves and no nodes of their own) and picking a new rule (*i.e.* a new split attribute and new split value). Let's establish some new notation for this section:

- η, η_* — the node destined to be changed on the original tree T and the proposed tree T^*.

- η_1, η_2 — the left and right nodes on tree T whose parent was chosen to be changed.

- η_{1*}, η_{2*} — the left and right nodes on tree T^* after the parent was changed.

- $R_{1,1}, \ldots, R_{1,n_1}$ — the residuals in the left leaf on tree T whose parent was chosen to be changed.

- $R_{2,1}, \ldots, R_{2,n_2}$ — the residuals in the right leaf on tree T whose parent was chosen to be changed.

- $R_{1*,1}, \ldots, R_{1*,n_{1*}}$ — the residuals in the left leaf on tree T^* after the parent was changed.

- $R_{2*,1}, \ldots, R_{2*,n_{2*}}$ — the residuals in the right leaf on tree T^* after the parent was changed.

- $\bar{R}_1$ — the mean of the residuals on the left node on tree T whose parent was chosen to be changed.

- $\bar{R}_2$ — the mean of the residuals on the right node on tree T whose parent was chosen to be changed.

- $\bar{R}_{1*}$ — the mean of the residuals on the left node on tree T^* after the parent was changed.

- $\bar{R}_{2*}$ — the mean of the residuals on the right node on tree T^* after the parent was changed.

Transition Ratio

We're looking at two simple transitions:

$$
\begin{aligned}
\mathbb{P}\left(T \to T^*\right) &= \mathbb{P}\left(\text{CHANGE}\right)\mathbb{P}\left(\text{selecting } \eta \text{ to change}\right) \times \\
&\quad \mathbb{P}\left(\text{selecting the new attribute to split on}\right) \times \\
&\quad \mathbb{P}\left(\text{selecting the new value to split on}\right) \\
&= \mathbb{P}\left(\text{CHANGE}\right)\frac{1}{b}\frac{1}{p_{\text{adj}}(\eta^*)}\frac{1}{n_{j\cdot\text{adj}}(\eta^*)} \\
\mathbb{P}\left(T^* \to T\right) &= \mathbb{P}\left(\text{CHANGE}\right)\mathbb{P}\left(\text{selecting } \eta_* \text{ to change}\right) \times \\
&\quad \mathbb{P}\left(\text{selecting the original attribute to split on}\right) \times \\
&\quad \mathbb{P}\left(\text{selecting the original value to split on}\right) \\
&= \mathbb{P}\left(\text{CHANGE}\right)\frac{1}{b}\frac{1}{p_{\text{adj}}(\eta^*)}\frac{1}{n_{j\cdot\text{adj}}(\eta^*)}
\end{aligned}
$$

It is obvious that the majority of the above will cancel below to produce the ratio:

$$
\frac{\mathbb{P}\left(T^* \to T^*\right)}{\mathbb{P}\left(T \to T^*\right)} = \frac{n_{j\cdot\text{adj}}(\eta^*)}{n_{j\cdot\text{adj}}(\eta)}
$$

Likelihood Ratio

The proposal tree differs from the original by just having two different leaves where the residuals are apportioned differently. We begin with direct substitution to quantities derived previously:

$$\frac{\mathbb{P}\left(\boldsymbol{R}\,|\,T^*,\sigma^2\right)}{\mathbb{P}\left(\boldsymbol{R}\,|\,T,\sigma^2\right)} = \frac{\mathbb{P}\left(R_{1^*,1},\ldots,R_{1^*,n_{1^*}}\,|\,\sigma^2\right)\mathbb{P}\left(R_{2^*,1},\ldots,R_{2^*,n_{2^*}}\,|\,\sigma^2\right)}{\mathbb{P}\left(R_{1,1},\ldots,R_{1,n_1}\,|\,\sigma^2\right)\mathbb{P}\left(R_{2,1},\ldots,R_{2,n_2}\,|\,\sigma^2\right)}$$

$$= c\,\frac{\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_{1^*}}\left(R_{1^*,i}-\bar{R}_{1^*}\right)^2\right)\exp\left(\frac{\bar{R}_{1^*}^2 n_{1^*}^2}{2\sigma^2\left(n_{1^*}+\frac{\sigma^2}{\sigma_\mu^2}\right)}-\frac{n_{1^*}\bar{R}_{1^*}^2}{2\sigma^2}\right)}{\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_1}\left(R_{1,i}-\bar{R}_1\right)^2\right)\exp\left(\frac{\bar{R}_1^2 n_1^2}{2\sigma^2\left(n_1+\frac{\sigma^2}{\sigma_\mu^2}\right)}-\frac{n_1\bar{R}_1^2}{2\sigma^2}\right)} \times$$

$$\frac{\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_{2^*}}\left(R_{2^*,i}-\bar{R}_{2^*}\right)^2\right)\exp\left(\frac{\bar{R}_{2^*}^2 n_{2^*}^2}{2\sigma^2\left(n_{2^*}+\frac{\sigma^2}{\sigma_\mu^2}\right)}-\frac{n_{2^*}\bar{R}_{2^*}^2}{2\sigma^2}\right)}{\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n_2}\left(R_{1,i}-\bar{R}_2\right)^2\right)\exp\left(\frac{\bar{R}_2^2 n_1^2}{2\sigma^2\left(n_2+\frac{\sigma^2}{\sigma_\mu^2}\right)}-\frac{n_2\bar{R}_2^2}{2\sigma^2}\right)}$$

where:

$$c := \sqrt{\frac{\left(\frac{\sigma^2}{\sigma_\mu^2}+n_1\right)\left(\frac{\sigma^2}{\sigma_\mu^2}+n_2\right)}{\left(\frac{\sigma^2}{\sigma_\mu^2}+n_1^*\right)\left(\frac{\sigma^2}{\sigma_\mu^2}+n_2^*\right)}}$$

The first exponential term can be combined among all four quantities and the second exponential term can be combined among all four quantities to yield:

$$= \exp\left(\frac{1}{2\sigma^2}\left(n_{1^*}\bar{R}_{1^*}^2+n_{2^*}\bar{R}_{2^*}^2-n_1\bar{R}_1^2-n_2\bar{R}_2^2\right)\right) \times$$

$$\exp\left(\frac{\bar{R}_{1^*}^2 n_{1^*}^2}{2\sigma^2\left(n_{1^*}+\frac{\sigma^2}{\sigma_\mu^2}\right)}-\frac{n_{1^*}\bar{R}_{1^*}^2}{2\sigma^2}+\frac{\bar{R}_{2^*}^2 n_{2^*}^2}{2\sigma^2\left(n_{2^*}+\frac{\sigma^2}{\sigma_\mu^2}\right)}-\frac{n_{2^*}\bar{R}_{2^*}^2}{2\sigma^2}\right) \times$$

$$\exp\left(-\frac{\bar{R}_1^2 n_1^2}{2\sigma^2\left(n_1+\frac{\sigma^2}{\sigma_\mu^2}\right)}+\frac{n_1\bar{R}_1^2}{2\sigma^2}-\frac{\bar{R}_2^2 n_1^2}{2\sigma^2\left(n_2+\frac{\sigma^2}{\sigma_\mu^2}\right)}+\frac{n_2\bar{R}_2^2}{2\sigma^2}\right)$$

Factoring out the $\frac{1}{2\sigma^2}$, we see that the top row cancels and we arrive at:

$$c\,\exp\left(\frac{1}{2\sigma^2}\left(\frac{\bar{R}_{1^*}^2 n_{1^*}^2}{n_{1^*}+\frac{\sigma^2}{\sigma_\mu^2}}+\frac{\bar{R}_{2^*}^2 n_{2^*}^2}{n_{2^*}+\frac{\sigma^2}{\sigma_\mu^2}}-\frac{\bar{R}_1^2 n_1^2}{n_1+\frac{\sigma^2}{\sigma_\mu^2}}-\frac{\bar{R}_2^2 n_1^2}{n_2+\frac{\sigma^2}{\sigma_\mu^2}}\right)\right)$$

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

which in log form is just simply:

$$\frac{1}{2}\left(\ln\left(n_1 + \frac{\sigma^2}{\sigma_\mu^2}\right) + \ln\left(n_2 + \frac{\sigma^2}{\sigma_\mu^2}\right) - \ln\left(n_{1*} + \frac{\sigma^2}{\sigma_\mu^2}\right) - \ln\left(n_{2*} + \frac{\sigma^2}{\sigma_\mu^2}\right)\right) +$$

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

Now if the number of data points doesn't change, we have $n_1 = n_1^*$ and $n_2 = n_2^*$ and this simplifies to:

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

Tree Structure Ratio

The proposal tree is the same shape as the original tree. Thus we only need to take into account the probabilities of the proposed split and the original split.

$$\frac{\mathbb{P}(T^*)}{\mathbb{P}(T)} = \frac{(1 - \mathbb{P}(\text{splitting } \eta_{1*}))(1 - \mathbb{P}(\text{splitting } \eta_{2*})))\,\mathbb{P}(\text{splitting } \eta_*)\,\mathbb{P}(\text{rule } \eta_*)}{(1 - \mathbb{P}(\text{splitting } \eta_1))(1 - \mathbb{P}(\text{splitting } \eta_2)))\,\mathbb{P}(\text{splitting } \eta)\,\mathbb{P}(\text{rule } \eta)}$$

The one minus probability of splitting will cancel since it's computed based on α, β and depth of the leaves which are the same. The probabilities of splitting will cancel based on the same argument. Within the probability of the rule, the number of possible attributes will cancel since it's the same so we really only have to consider:

$$\frac{\mathbb{P}(T^*)}{\mathbb{P}(T)} = \frac{n_{j\cdot\text{adj}}(\eta)}{n_{j\cdot\text{adj}}(\eta^*)}$$

Note that this is the inverse of the transition ratio. Thus, only the likelihood ratio needs to be computed to determine r for the change step.

Change & Swap

With m large and enough burn-in, these two steps have nominal effect on our predictions, but they may have an effect on variable selection.