Appendix A: Details on simulating from the predictive

In this section we provide additional on the simulation from the predictive distribution depending on the choice of the prior on F_Y.

A.1. Empirical CDF

It might be the case that n is so large that simply using the empirical CDF is a reasonable approximation. In this case, the inversion of the cdf becomes trivial, and MCMC is only required for the β posterior sample.

- Simulate $\beta^{(j)}$ from the partial posterior
- $Z' \sim \text{Exp}(\lambda_{\beta^{(j)}}(x'))$
- Calculate $u^{(j)} := 1 - \frac{1}{n}\sum_{i=1}^{n} e^{-Z'\lambda_{\beta^{(j)}}(x_i)}$
- Set $Y'^{(j)} = y_{(\lceil nu^{(j)} \rceil)}$

A.1.1. Bayesian Bootstrap

An alternative approach might be to use a Bayesian Bootstrap on F_Y. This works out similarly to using the empirical CDF, but we must simulate the Dirichlet weights for each of the atoms. The sampling scheme becomes:

- Simulate $\beta^{(j)}$ from the partial posterior
- Sample $Z' \sim \text{Exp}(\lambda_{\beta^{(j)}}(x'))$
- Calculate $u^{(j)} := 1 - \frac{1}{n}\sum_{i=1}^{n} e^{-Z'\lambda_{\beta^{(j)}}(x_i)}$
- Simulate $(W_1, W_2, \ldots, W_n) \sim \text{Dirichlet}(1, 1, \ldots, 1)$
- Set $Y'^{(j)} = y_{(\min\{k : \sum_{i=1}^{k} W_i \geq u^{(j)}\})}$

A.1.2. Pólya Trees

Under our composite likelihood scheme, the posterior for F_Y is also a Pólya tree, due to conjugacy of the Pólya tree prior. Simulation then proceeds as follows:

- Simulate $\beta^{(j)}$ from the partial posterior
- $Z' \sim \text{Exp}(\lambda_{\beta^{(j)}}(x'))$
- Calculate $u^{(j)} := 1 - \frac{1}{n}\sum_{i=1}^{n} e^{-Z'\lambda_{\beta^{(j)}}(x_i)}$

Then all we need to calculate is $F_Y^{-1(j)}(u^{(j)})$. Pólya trees make this easy. A sample from a Pólya tree distribution is a random probability measure, which is constructed by assigning random masses to each branch in a partition tree of the space. So, given the first partition point in the tree, we can simply generate the random mass either side of this point, and trivially deduce which branch $F_Y^{-1(j)}(u^{(j)})$ lies in. We repeat this process down the tree until we reach the truncation point often used when using Pólya trees.

Formally, given a Pólya tree truncated at level M, let $a_0 = 0, B_0 = 1, \epsilon_0 = \emptyset$ and for k from 1 to M:

- $\theta_{\epsilon_k 0} \sim \text{Beta}(\alpha_{\epsilon_k 0}, \alpha_{\epsilon_k 1})$
- if $u \in [a_k, \theta_k(b_{k-1} - a_{k-1}) + a_{k-1}]$
  - $\epsilon_k = \epsilon_{k-1}0$
  - $a_k = a_{k-1}$
  - $b_k = a_{k-1} + \theta_k(b_{k-1} - a_{k-1})$
- Otherwise
  - $\epsilon_k = \epsilon_{k-1}1$
  - $a_k = a_{k-1} + \theta_k(b_{k-1} - a_{k-1})$
  - $b_k = a_{k-1}$

A.1.3. Dirichlet Process Mixture models

The difficulty here becomes the inversion of F_Y, since this has no closed form. A simple Monte Carlo approximation can be used to approximate this inversion for each posterior sample.

- Simulate $\beta^{(j)}$ from the partial posterior
- Sample $Z' \sim \text{Exp}(\lambda_{\beta^{(j)}}(x'))$
- Simulate $F_Y^{(j)}$ from the partial posterior
- Simulate $Y_k'^{(j)} \sim F_Y^{(j)}$ for $k = 1, \dots, N$
- Calculate $u^{(j)} := 1 - \frac{1}{n}\sum_{i=1}^{n} e^{-Z'\lambda_{\beta^{(j)}}(x_i)}$
- Set $Y'^{(j)} = Y'^{(j)}_{(\lceil N u^{(j)} \rceil)}$