

separate build

Consider the simple two-part model where we model our response as

$$\begin{aligned} p_i &= \frac{1}{1 + e^{-(x_i\gamma_1 + \gamma_0)}} \\ y_i^* &= x_i\beta_1 + \beta_0 + \epsilon_i \quad \text{where} \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2) \\ y_i &= p_i \cdot y_i^* \end{aligned}$$

Under a Bayesian frame we could define a full posterior for all model parameters as

$$\begin{aligned} f(\beta_1, \beta_0, \gamma_1, \gamma_0, \sigma^2 \mid y) &\propto f(y \mid \beta_1, \beta_0, \gamma_1, \gamma_0, \sigma^2) f(\beta_1, \beta_0, \gamma_1, \gamma_0, \sigma^2) \\ &= L(\beta_1, \beta_0, \gamma_1, \gamma_0, \sigma^2 \mid y) f(\beta_1, \beta_0, \gamma_1, \gamma_0, \sigma^2) \\ &= L(\beta_1, \beta_0, \gamma_1, \gamma_0, \sigma^2 \mid y) f(\beta_1) f(\beta_0) f(\gamma_1) f(\gamma_0) f(\sigma^2) \end{aligned}$$

We can expand this by writing out the likelihood more fully based on whether y is zero or not:

$$\begin{aligned} f(\beta_1, \beta_0, \gamma_1, \gamma_0, \sigma^2 \mid y) &\propto \left[\prod_{i:y_i=0} (1 - p_i) \prod_{i:y_i>0} p_i f(y_i \mid \beta_1, \beta_0, \sigma^2) \right] f(\beta_1) f(\beta_0) f(\gamma_1) f(\gamma_0) f(\sigma^2) \\ &= \left[\prod_{i:y_i=0} \left(1 - \frac{1}{1 + e^{-(x_i\gamma_1 + \gamma_0)}} \right) \right] \left[\prod_{i:y_i>0} \left(\frac{1}{1 + e^{-(x_i\gamma_1 + \gamma_0)}} \right) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i - (x_i\beta_1 + \beta_0))^2}{2\sigma^2}} \right] \\ &\quad f(\beta_1) f(\beta_0) f(\gamma_1) f(\gamma_0) f(\sigma^2) \end{aligned}$$

We then can group these terms based on the parameters that they use

$$\begin{aligned} &= \left[\prod_{i:y_i=0} \left(1 - \frac{1}{1 + e^{-(x_i\gamma_1 + \gamma_0)}} \right) \prod_{i:y_i>0} \left(\frac{1}{1 + e^{-(x_i\gamma_1 + \gamma_0)}} \right) \right] f(\gamma_1) f(\gamma_0) \left[\prod_{i:y_i>0} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i - (x_i\beta_1 + \beta_0))^2}{2\sigma^2}} \right] \\ &\quad f(\beta_1) f(\beta_0) f(\sigma^2) \end{aligned}$$

but now, if we look at this closely we can see that what we really have here is a full separation of the likelihoods for the individual models

$$\begin{aligned}
f(\beta_1, \beta_0, \gamma_1, \gamma_0, \sigma^2 \mid y) &\propto \left[L(\gamma_1, \gamma_0 \mid y) f(\gamma_1) f(\gamma_0) \right] \cdot \left[L(\beta_1, \beta_0, \sigma \mid y) f(\beta_1) f(\beta_0) f(\sigma^2) \right] \\
&= \left[f(y \mid \gamma_1, \gamma_0) f(\gamma_1, \gamma_0) \right] \cdot \left[f(y \mid \beta_1, \beta_0, \sigma^2) f(\beta_1, \beta_0, \sigma^2) \right] \\
&= \left[f(\gamma_1, \gamma_0 \mid y) f(y) \right] \left[f(\beta_1, \beta_0, \sigma^2 \mid y) f(y) \right] \\
&\propto f(\gamma_1, \gamma_0 \mid y) f(\beta_1, \beta_0, \sigma^2 \mid y)
\end{aligned}$$

So we've just shown that the full posterior for the model built simultaneously is proportional to the posteriors for each model built separately. What this means for us is that we can fit the models separately, and then combine the results at the end to get our posterior predictive distributions.