

# 1 Appendix

The model joint probability

$$p(\mathbf{Y}, \mathbf{x}, \beta_i, \sigma, \Delta, s, w) = p(\mathbf{Y} \mid \mathbf{x}, \beta_i, \sigma) p(\beta_i \mid \Delta, s) p(\Delta \mid 0, w) p(\sigma) p(w) p(s)$$

where the individual factors are:

$$\begin{aligned} p(\mathbf{Y} \mid \mathbf{x}, \beta_i, \sigma) &= \prod_{i=1}^I \prod_{c=1}^C \mathcal{N}(y_c \mid x_c \beta_i, \sigma^{-1}) \\ p(\beta_i \mid \Delta, s) &= \prod_{i=1}^I \mathcal{N}(\beta_i \mid \Delta, s^{-1}) \\ p(\Delta \mid 0, w) &= \mathcal{N}(\Delta \mid 0, w^{-1}) \\ p(\sigma) &= \text{Gamma}(\sigma \mid a_0, b_0) \\ p(s) &= \text{Gamma}(s \mid c_0, d_0) \\ p(w) &= \text{Gamma}(w \mid e_0, f_0) \end{aligned}$$

We want to infer the posterior distribution:

$$q(\beta_i, \Delta, \sigma, s, w) = p(\beta_i, \Delta, \sigma, s, w \mid \mathbf{D})$$

we assume that the posterior distribution factorizes into independent factors:

$$q(\beta_i, \Delta, \sigma, s, w) = q(\beta_i) q(\Delta) q(\sigma) q(s) q(w)$$

The true posterior distribution does not in fact factor this way and we will obtain an approximation.

## 2 Derivation of $q(\beta_i)$

$$\begin{aligned}
\ln q_{\beta_i}^*(\beta_i) &= \mathbb{E}_{\Delta, \sigma, s} [\ln p(\mathbf{Y} \mid \mathbf{x}\beta_i, \sigma) + \ln p(\beta_i \mid \Delta, s)] + C \\
&= \mathbb{E}_{\sigma} \left[ \sum_{c=1}^C \frac{1}{2} \ln \sigma - \frac{\sigma}{2} (y_c - x_c \beta_i)^2 \right] + \mathbb{E}_{\Delta, s} \left[ \frac{1}{2} \ln s - \frac{s}{2} (\Delta - \beta_i)^2 \right] + C \\
&= \mathbb{E}_{\sigma} \left[ \sum_{c=1}^C -\frac{\sigma}{2} (y_c - x_c)^2 \right] + \mathbb{E}_{\Delta, s} \left[ -\frac{s}{2} (\Delta - \beta_i)^2 \right] + C \\
&= \mathbb{E}_{\sigma} \left[ -\frac{\sigma}{2} \sum_{c=1}^C y_c^2 + \sigma \beta_i \sum_{c=1}^C y_c x_c - \frac{\sigma \beta_i^2 \sum_{c=1}^C x_c^2}{2} \right] + \mathbb{E}_{\Delta, s} \left[ -\frac{s \Delta^2}{2} + s \Delta \beta_i - \frac{s \beta_i^2}{2} \right] + C \\
&= \mathbb{E}_{\sigma} \left[ \sigma \beta_i \sum_{c=1}^C y_c x_c - \frac{\sigma \beta_i^2 \sum_{c=1}^C x_c^2}{2} \right] + \mathbb{E}_{\Delta, s} \left[ s \Delta \beta_i - \frac{s \beta_i^2}{2} \right] + C \\
&= (\mathbb{E}_{\sigma} [\sigma] \sum_{c=1}^C x_c y_c + \mathbb{E}_s [s] \mathbb{E}_{\Delta} [\Delta]) \beta_i - \frac{1}{2} (\mathbb{E}_{\sigma} [\sigma] \sum_{c=1}^C x_c^2 + \mathbb{E}_s [s]) \beta_i^2 + C
\end{aligned}$$

$$\begin{aligned}
\lambda_{\beta, i} &= \frac{a_n}{b_n} \sum_{c=1}^C x_c^2 + \frac{c_n}{d_n} \\
\beta_i &= \frac{\frac{a_n}{b_n} \sum_{c=1}^C x_c y_c + \frac{c_n}{d_n} \Delta}{\lambda_{\beta, i}}
\end{aligned}$$

### 3 Derivation of $q(\Delta)$

$$\begin{aligned}
\ln q_{\Delta}^*(\Delta) &= \mathbb{E}_{\beta_i, s} [\ln p(\beta_i \mid \Delta, s) + \ln p(\Delta \mid \mu_0, w)] + C \\
&= \mathbb{E}_{\beta_i, s} \left[ \sum_{i=1}^I \frac{1}{2} \ln s - \frac{s}{2} (\Delta - \beta_i)^2 \right] + \mathbb{E}_w \left[ \frac{1}{2} \ln w - \frac{w}{2} (\Delta - \mu_0)^2 \right] + C \\
&= \mathbb{E}_{\beta_i, s} \left[ \sum_{i=1}^I -\frac{s}{2} (\Delta - \beta_i)^2 \right] + \mathbb{E}_w \left[ -\frac{w}{2} (\Delta - \mu_0)^2 \right] + C \\
&= \mathbb{E}_{\beta_i, s} \left[ -\frac{Is\Delta^2}{2} + s\Delta \sum_{i=1}^I \beta_i \right] + \mathbb{E}_w \left[ w\Delta - \frac{w\Delta^2}{2} \right] + C \\
&= \mathbb{E}_{\beta_i, s} \left[ -\frac{1}{2} (Is + w)\Delta^2 \right] + \mathbb{E}_{w, s, \beta_i} \left[ \Delta \left( s \sum_{i=1}^I \beta_i + w\mu_0 \right) \right] + C
\end{aligned}$$

$$\begin{aligned}
\lambda_{\Delta} &= I \frac{c_n}{d_n} + \frac{e_n}{f_n} \\
\Delta &= \frac{\frac{c_n}{d_n} \sum_i \beta_i + \frac{e_n}{f_n} \mu_0}{\lambda_{\Delta}}
\end{aligned}$$

### 4 Derivation of $q(\sigma)$

$$\begin{aligned}
\ln q_{\sigma}^*(\sigma) &= \mathbb{E}_{\beta_i} [\ln p(\mathbf{Y} \mid \mathbf{x}\beta_i, \sigma) + \ln p(\sigma \mid a, b)] + C \\
&= \mathbb{E}_{\beta_i} \left[ \sum_{i=1}^I \sum_{c=1}^C \frac{1}{2} \ln \sigma - \frac{\sigma}{2} (y_{i,c} - \mathbf{x}\beta_i)^2 \right] + [(a-1) \ln \sigma - b\sigma] + C \\
&= \mathbb{E}_{\beta_i} \left[ -\frac{\sigma}{2} \sum_{i=1}^I \sum_{c=1}^C (y_{i,c} - \mathbf{x}\beta_i)^2 \right] + (a-1) \ln \sigma - b\sigma + \frac{\sum_{i=1}^I \sum_{c=1}^C 1}{2} \ln \sigma + C \\
&= \mathbb{E}_{\beta_i} \left[ -\frac{\sigma}{2} \sum_{i=1}^I \sum_{c=1}^C y_{i,c}^2 - 2y_{i,c} \mathbf{x}\beta_i + \mathbf{x}\beta_i^2 \right] + ((a-1) + \frac{\sum_{i=1}^I \sum_{c=1}^C 1}{2}) \ln \sigma - b\sigma + C \\
&= \mathbb{E}_{\beta_i} \left[ -\left( \left( \sum_{i=1}^I \mathbf{x}\beta_i^2 - 2 \sum_{i=1}^I \sum_{c=1}^C y_{i,c} \mathbf{x}\beta_i + \sum_{i=1}^I \sum_{c=1}^C y_{i,c}^2 \right) \frac{1}{2} + b \right) \sigma \right] + ((a-1) + \frac{\sum_{i=1}^I \sum_{c=1}^C 1}{2}) \ln \sigma + C
\end{aligned}$$

$$\begin{aligned}
N &= \sum_{i=1}^I \sum_{c=1}^C 1 \\
a_n &= a + \frac{N}{2} \\
b_n &= b + \frac{1}{2} \left( \sum_{i=1}^I \sum_{c=1}^C (y_{i,c} - \mathbf{x} \beta_i)^2 + \sum_i^I \text{trace}(\lambda_{\beta,i} X_i^T X_i) \right)
\end{aligned}$$

## 5 Derivation of $q(s)$

$$\begin{aligned}
\ln q_s^*(s) &= E_{\beta_i, \Delta} [\ln p(\beta_i \mid \Delta, s) + \ln p(s \mid c, d)] + C \\
&= E_{\beta_i, \Delta} \left[ \sum_{i=1}^I \sum_{d=1}^D \frac{1}{2} \ln s - \frac{s}{2} (\beta_{i,d} - \Delta_d)^2 \right] + [(c-1) \ln s - ds] + C \\
&= E_{\beta_i, \Delta} \left[ \sum_{i=1}^I \sum_{d=1}^D -\frac{s}{2} (\beta_{i,d} - \Delta_d)^2 \right] + ((c-1) + \frac{\sum_{i=1}^I \sum_{d=1}^D 1}{2}) \ln s - ds + C \\
&= E_{\beta_i, \Delta} \left[ -\left( \left( \sum_{i=1}^I \sum_{d=1}^D \beta_{i,d}^2 - 2 \sum_{i=1}^I \sum_{d=1}^D \beta_{i,d} \Delta_d + \sum_{d=1}^D \Delta_d^2 \right) \frac{1}{2} + d \right) s \right] + ((c-1) + \frac{ID}{2}) \ln s + C
\end{aligned}$$

$$\begin{aligned}
c_n &= c + \frac{DI}{2} \\
d_n &= d + \frac{1}{2} \left( \sum_{i=1}^I (\beta_i - \delta)^2 + \text{trace}(\lambda_\Delta) + \sum_i^I \text{trace}(\lambda_{\beta,i} X_i^T X_i) \right)
\end{aligned}$$

## 6 Derivation of $q(w)$

$$\begin{aligned}
\ln q_w^*(w) &= E_{\Delta} [\ln p(\Delta \mid 0, w) + \ln p(w \mid e, f)] + C \\
&= E_{\Delta} \left[ \sum_{d=1}^D \frac{1}{2} \ln w - \frac{w}{2} (\delta_d)^2 \right] + [(e-1) \ln w - fw] + C \\
&= E_{\Delta} \left[ \sum_{d=1}^D -\frac{w}{2} (\delta_d)^2 \right] + ((e-1) + \frac{D}{2}) \ln w - fw + C \\
&= E_{\Delta} \left[ -w \left( \frac{1}{2} \sum_{d=1}^D (-\delta_d)^2 + f \right) \right] + ((e-1) + \frac{D}{2}) \ln w + C
\end{aligned}$$

$$\begin{aligned}
e_n &= e + \frac{D}{2} \\
f_n &= f + \frac{1}{2} (\Delta^T \Delta + \text{trace}(\lambda_{\Delta}))
\end{aligned}$$