Likelihood function and posterior distribution

For clarity purposes, in this section, we review the main components of our model, which we use in the MCMC steps detailed in the next section.

Under model
\[ Y_{ν}^* = X_{ν}^* \beta_{ν} + \varepsilon_{ν}^* \]  \[ \varepsilon_{ν}^* \sim N(0, \Sigma_{ν}^*) \]

the likelihood function is
\[ f(Y^*|Θ) \propto V^{−1} \prod_{ν=1}^{V} |Σ_{ν}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2}(Y_{ν}^* - X_{ν}^* \beta_{ν})^{T} (φ_{ν} Σ_{ν}^{-1} (Y_{ν}^* - X_{ν}^* \beta_{ν})) \right] \]

where \( Θ = (β, γ, λ, φ, α) \), \( Σ_{ν} \) is a diagonal matrix with each element set as \((2^{α_{ν}})^{-m}\).

Let \( φ = (ψ, α) \). The full posterior distribution function is obtained via Bayes theorem as
\[ π(Θ|Y^*) \propto f(Y^*|Θ)π(β|γ)π(γ)π(λ)π(φ), \]

where
\[ π(β_{ν}|γ_{ν} = 1) \propto τ^{-\frac{1}{2}} \exp \left( \frac{β_{ν}^2}{2τ} \right), \]
\[ π(β_{ν}|γ_{ν} = 0) = δ_{ν0}, \]
\[ π(γ_{ν}|γ_{ν-}) \propto \exp(γ_{ν}(d + e \sum_{k \in N_{ν}} γ_{k})), \]
\[ π(λ_{ν}) = \frac{1}{u_{2} - u_{1}} I_{(u_{1}, u_{2})}(λ_{ν}), \]
\[ π(φ_{ν}|φ_{ν-}) = \frac{η}{η + V - 1} G_{0} + \frac{1}{η + V - 1} \sum_{j=1}^{m} n_{j} δ_{φ_{ν}j}. \]

In equation (2), \( φ_{1}^*, ..., φ_{m}^* \) are the unique values of \( φ_{i}^* \)'s for \( i \neq ν \), and \( n_{j} \) are the frequencies of \( φ_{i}^* \) in the vector \( (φ_{1}^*, ..., φ_{ν-1}^*, φ_{ν+1}^*, ..., φ_{V}^*) \).
**MCMC algorithm**

**Update step for \((\beta, \gamma)\)**

The full conditional distribution of \(\beta_v\) is

\[
\begin{align*}
\beta_v | Y^*_v, \psi_v, \alpha_v, \gamma_v = 1 & \sim N(\mu_v, \sigma^2_v), \\
\beta_v | Y^*_v, \psi_v, \alpha_v, \gamma_v = 0 & \sim \delta_0,
\end{align*}
\]

where

\[
\mu_v = \frac{Y^*_v (\psi_v \Sigma_{\alpha_v})^{-1} X_v \tau_v}{X_v^T (\psi_v \Sigma_{\alpha_v})^{-1} X_v \tau_v + 1},
\]

\[
\sigma^2_v = \frac{\tau_v}{X_v^T (\psi_v \Sigma_{\alpha_v})^{-1} X_v \tau_v + 1}.
\]

We perform Add, Delete, Swap steps as in (1) to jointly update \((\beta, \gamma)\):

1) Randomly choose among the three moves below.

i) **Add**: set \(\gamma_v^* = 1\) and sample \(\beta_v^*\) from a \(N(\mu_v, \sigma^2_v)\) proposal. Here \(\mu_v\) and \(\sigma^2_v\) are as shown in Equations (3) and (4). Position \(v\) is randomly chosen from the set of \(v's\) where \(\gamma_v = 0\) at the previous iteration.

ii) **Delete**: set \(\gamma_v^* = 0, \beta_v^* = 0\). This results in voxel \(v\) being excluded in the current iteration. Position \(v\) is randomly chosen from among those included in the model at the previous iteration.

iii) **Swap**: perform both an Add and Delete move.

The proposed value \((\gamma^*, \beta^*)\) is accepted with probability

\[
\alpha_{\beta, \gamma} = \min \left\{ 1, \frac{\pi(\gamma^*, \beta^*) | Y^*, X^*, \psi, \alpha) q(\gamma, \beta | \gamma^*, \beta^*)}{\pi(\gamma, \beta | Y^*, X^*, \psi, \alpha) q(\gamma^*, \beta^* | \gamma, \beta)} \right\} = \min \left\{ 1, \frac{f(Y^* | \beta^*, \gamma^*, ..., \gamma, \beta) \pi(\gamma^*)}{f(Y^* | \beta, \gamma, ..., \gamma, \beta) \pi(\gamma)} \right\}.
\]

2) Repeat step (1) \(m\) times.

3) Sampling from \(N(\mu_v, \sigma^2_v)\) for \(\beta_v\)'s such that \(\gamma_v = 1\).

When \((\beta^*, \gamma^*)\) is obtained via the Add or Delete move, then

\[
\frac{\pi(\gamma^*)}{\pi(\gamma)} = \frac{\pi(\gamma^*_v | y_k, k \in N_v)}{\pi(\gamma_v | y_k, k \in N_v)}
\]

where \(N_v\) is the neighbor of voxel \(v\) which is updated in the Add or Delete move.

When \((\beta^*, \gamma^*)\) is obtained via the Swap move, if, say, voxels \(j\) and \(l\) are the ones to be updated, then

\[
\frac{\pi(\gamma^*)}{\pi(\gamma)} = \frac{\pi(\gamma^*_j | y_k, k \in N_j) \pi(\gamma^*_l | y_k, k \in N_{l(-j)})}{\pi(\gamma_j | y_k, k \in N_j) \pi(\gamma_l | y_k, k \in N_{l(-j)})},
\]

where \(N_j\) is the neighbor of voxel \(j\) and \(N_{l(-j)}\) is the neighbor of voxel \(l\) excluding voxel \(j\).

**Update step for \(\lambda_v\)**

The full conditional distribution of \(\lambda_v\) is

\[
\lambda_v | Y^*_v, \beta_v, \psi_v, \alpha_v \propto \exp \left[ -\frac{1}{2} Y^*_v (\psi_v \Sigma_{\alpha_v})^{-1} Y^*_v - 2 Y^*_v \beta_v (\psi_v \Sigma_{\alpha_v})^{-1} \beta_v \right] I_{(\alpha_1, \alpha_2)}(\lambda_v).
\]

We propose \(\lambda^*_v \sim U(\lambda_v - h_v, \lambda_v + h_v)\), and the proposed value is accepted with the acceptance probability

\[
a_{\lambda} = \min \left\{ 1, \frac{\pi(\lambda^*_v | Y^*_v, \beta_v, \psi_v, \alpha_v, \lambda)}{\pi(\lambda_v | Y^*_v, \beta_v, \psi_v, \alpha_v, \lambda)} \right\}.
\]
Update step for \((\psi, \alpha)\)

Let \(\psi = (\psi_v, \alpha_v)\). The full conditional distribution of \(\psi_v\) is

\[
\pi(\psi_v | Y_1^* , \ldots , Y_n^* , \beta, \lambda, \alpha) = \begin{cases} 
\phi_{j_v}^* & \text{w.p. } b\ n_j \ f(Y_v^* | \lambda_v, \beta_v, \alpha_v, \psi_v = \phi_{j_v}^*) \\
\hat{h}(\psi_v | Y_v^*, \beta_v, \lambda_v, \alpha_v) & \text{w.p. } b\ \eta q_0
\end{cases}
\]

where

\[
b = \frac{1}{\eta q_0 + \sum_{i=1}^{n} n_j f(Y_v^* | \lambda_v, \beta_v, \psi_v = \phi_{j_v}^*)},
\]

\[
q_0 = \int G_0(\psi_v, \alpha_v) f(Y_v^* | \psi_v, \alpha_v, \beta_v, \lambda_v) d\psi_v d\alpha_v
\]

\[
= \int \left(2\pi \right)^{-\frac{n}{2}} \prod_{i=1}^{n} \left| \Sigma_{\alpha_i} \right|^{-\frac{1}{2}} \left(1 - \alpha_c \right)^{-\frac{1}{2}} \left(1 - \alpha_c \right)^{-\frac{1}{2}} \prod_{i=1}^{n} \left( \Gamma(\alpha_0 + \frac{n}{2}) \Gamma(\alpha_0 + b_i) \right)^{-\frac{1}{2}} \prod_{i=1}^{n} \left( \Sigma_{\alpha_i} \right)^{-\frac{1}{2}} d\alpha_v,
\]

and

\[
\hat{h}(\psi_v | Y_v^*, \beta_v, \lambda_v, \alpha_v) \propto \psi^{-\frac{n}{2}} \prod_{i=1}^{n} \left(1 - \alpha_c \right)^{-\frac{1}{2}} \prod_{i=1}^{n} \left( \Sigma_{\alpha_i} \right)^{-\frac{1}{2}} \exp \left\{ -b_0 - \frac{1}{2} (Y_v^* - X_v^* \beta_v)^T \Sigma_{\alpha_i}^{-1} (Y_v^* - X_v^* \beta_v) \right\},
\]

Here \(\phi_{j_v}^*\)’s and \(n_j\) are as defined in Equation 23.

Since \(q_0\) cannot be computed in closed form, we use algorithm 8 proposed by ?? to update \((\psi, \alpha)\), which can be described as follows:

We introduce an auxiliary parameter \(c_v\) indicating which “latent cluster” is associated with \(\phi_v\). Suppose the state of the Markov chain consist of \(c = (c_1, \ldots, c_V)\), and \(\phi = (\phi_c : c \in \{1, \ldots, c_V\})\), where \(\phi_c = (\phi_v, \alpha_v), c\) and \(\phi\) can be updated as follows:

a) For \(v = 1, \ldots, V\): let \(k^v\) be the number of distinct \(c_{v-j}\), where \(c_{v-j}\) is a set of \(c_j\)’s for \(j \neq v\), if \(c_v \in c_{v-j}\), relabel these \(c_{v-j}\) with values in \(\{1, \ldots, k-v\}\), then draw values independently from \(G_0\) for \(\phi_{k^v+1}\); if \(c_v \notin c_{v-j}\), relabel these \(c_{v-j}\) with values in \(\{1, \ldots, k-v\}\), let \(c_{v+k^v} = \phi_v\). Draw a new value for \(c_v\) from \(\{1, \ldots, k^v+1\}\) using the following probabilities:

\[
\pi(c_v = c | Y_v^*, c_{v-j}, \beta_v, \lambda_v, \phi_1, \ldots, \phi_{k^v+1}) = \begin{cases} 
\frac{n_{c-v}}{b_0} F(Y_v^*, \phi_c) & \text{for } 1 \leq c \leq k^v \\
\frac{b_0}{\int F(Y_v^*, \phi_c)} & \text{for } c = k^v + 1
\end{cases}
\]

where \(n_{c-v}\) is the number of \(c_j\) for \(j \neq v\) that are equal to \(c\), and \(b\) is the appropriate normalizing constant.

b) For all \(c \in \{c_1, \ldots, c_V\}\): sample a new value from \(\phi_c\) all \(Y_v^*\) for which \(c_v = c\), that is, from the posterior distribution based on the prior \(G_0\) and all the data points currently associated with latent cluster \(c\). Since the posterior distribution of \(\phi_c\) is not in a closed form, we perform Metropolis-Hasting algorithm in this part to approximate the distribution of \(\phi_c\). The procedure is as below: suppose the number of repeats of value \(\phi_c\) is \(n_c\), all the data points associated with class \(c\) is \(Y_c^* = (Y_{v_1}, \ldots, Y_{v_{n_c}})^T, \beta_c, \lambda_c\) are defined similarly, the posterior distribution is

\[
\pi(\alpha_c, \psi_c | Y_c^*, \beta_c, \lambda_c) \propto \psi^{-\frac{n_c}{2}} - \alpha_c^{n_c-1} (1 - \alpha_c)^{\frac{n_c}{2}} \prod_{i=1}^{n_c} \left( \Sigma_{\alpha_i} \right)^{-\frac{1}{2}} \exp \left\{ -b_0 - \frac{1}{2} \sum_{i=1}^{n_c} (Y_{v_i}^* - X_v^* \beta_{v_i})^T \Sigma_{\alpha_i}^{-1} (Y_{v_i}^* - X_v^* \beta_{v_i}) \right\}.
\]
Propose $\alpha'_c \sim N(\alpha_c, \sigma^2_0)$, and $\psi'_c \sim N(\psi_c, \sigma^2_0)$, then compute the acceptance rate

$$a_{\psi,\alpha} = \min \left\{ 1, \frac{\pi(\psi'_c, \alpha'_c | Y^*_c, \beta_c, \lambda_c) q(\psi'_c, \alpha'_c | \psi_c, \alpha_c)}{\pi(\psi_c, \alpha_c | Y^*_c, \beta_c, \lambda_c) q(\psi'_c, \alpha'_c | \psi_c, \alpha_c)} \right\}$$

The proposed value $(\psi'_c, \alpha'_c)$ is accepted with probability $a_{\psi,\alpha}$. 
Event-related Design

Figure 1 shows the posterior activation map, and the posterior mean estimates for the parameters $\beta$ and $\lambda$ for the event-related design in the simulation described in section 3.1 of the main text. Plot (a) in Figure 2 shows the resulting clustering of the voxels for the event-related design. There are 3 clusters in plot (a). The posterior mean maps for the parameter $\psi$ and $\alpha$ are shown in plots (b) and (c) of Figure 2 for event-related design. We find the good estimates to the true values of all the parameters.

Figure 1: Simulated data with event-related design: (a) True map of the activation indicators $\gamma$; (b) Posterior activation map obtained by assigning value 1 to those voxels with $p(\gamma = 1 | y) > 0.8$ and value 0 otherwise; (c) Posterior mean map of $\beta$; (d) Scatter plot of posterior mean estimates vs. true values for $\beta$; (e) Posterior mean map for $\lambda$; (f) same as (d) for $\lambda$. 

5
A case study for fMRI data

Figure 2: Simulated data with event-related design: (a) Posterior clustering map - different colors correspond to different clustering allocations; (b) Posterior mean map of $\psi$; (c) Posterior mean map of $\alpha$.

Figure 3: Real fMRI data: Time series fitting for one active voxel on (a) V1, (b) V5, and (c) PP. The continuous black curves represent the real time series and the dashed black curves represent the fitted response.