## PRIORS FOR $M_j$

### 1. Parametric Gamma Prior

1.1. **Initialization.** The initial values for the model parameters and latent variables is obtained by a method-of-moments (MoM) procedure. MoM works by setting the population moment equal to the sample moment. We present the initial parameter estimates and provide the derivations here.

Since the distribution is $M_j \sim \text{Gamma}(a, b)$. The first and second population moments are

$$E[M_j] = ab, \tag{1}$$

$$E[{M_j}^2] = b^2 a(a+1). \tag{2}$$

The first and second sample moments are $m_1 = \frac{1}{J}\sum_{j=1}^{J} M_j$ and $m_2 = \frac{1}{J}\sum_{j=1}^{J} M_j^2$. Setting the population moments equal to the sample moments and we get

$$a = \frac{{m_1}^2}{m_2 - {m_1}^2} \tag{3}$$

$$b = \frac{m_2 - {m_1}^2}{m_1} \tag{4}$$

1.2. **Sampling from $p(M_j|a, b, \theta_{ji}, \mu_j)$.** The posterior distribution over $M_j$ given its Markov blanket is

$$p(M_j|a, b, \theta_{ji}, \mu_j) \propto p(\theta_{ji}|\mu_j, M_j)p(M_j|a, b) \tag{5}$$

We have $\theta_{ji} \sim \text{Beta}(\mu_j, M_j)$, and $M_j \sim \text{Gamma}(a, b)$. Assuming there is only one replicate,

$$p(M_j|a,b,\theta_{ji},\mu_j) \propto \frac{\Gamma(\mu_j M_j)}{\Gamma(\mu_j)\Gamma(M_j)} {\theta_j}^{\mu_j M_j - 1} {(1-\theta)_j}^{(1-\mu_j)M_j - 1} \frac{b^a}{\Gamma(a)} M_j^{a-1} e^{-bM_j} \tag{6}$$

Since we cannot give an analytical form for this, so we sample from the posterior distribution using the Metropolis-Hastings algorithm.

## 2. Jefferys Priors

Jeffreys prior is proposed for invariance by Harold Jeffreys (1946), and defined in terms of the Fisher information. For our ploblem,

$$\pi(M_j) = I(M_j)^{\frac{1}{2}} \tag{7}$$

Since $\theta_{ji} \sim Beta(\mu_j, M_j)$, the Fisher information is given by

$$I(M_j) = E_{M_j}\left[ -\frac{\delta^2 \log p(\theta_j|\mu_j, M_j)}{\delta M_j^2} \right] \tag{8}$$

Now to calculate the equations to acquire the Fisher information, we assume there is only one replicate,

$$p(\theta_j) = \frac{\Gamma(M_j)}{\Gamma(\mu_j M_j)\Gamma((1-\mu_j)M_j)} {\theta_j}^{\mu_j M_j - 1} {(1-\theta)_j}^{(1-\mu_j)M_j - 1} \tag{9}$$

$$\begin{aligned} \log p(\theta_j|\mu_j, M_j) &= \log\Gamma(M_j) - \log\Gamma(\mu_j, M_j) - \log\Gamma(1-\mu_j, M_j) \\ &\quad + (\mu_j M_j - 1)\log\theta_j + ((1-\ u_j)M_j - 1)\log(1-\theta_j) \end{aligned} \tag{10}$$

$$\frac{\delta \log p(\theta_j)}{\delta M_j} = \Psi(M_j) - \Psi(\mu_j M_j)\mu_j - \Psi((1-\mu_j)M_j)(1-\mu_j) + \mu_j \log\theta_j + (1-\mu_j)\log(1-\theta_j) \tag{11}$$

$$\frac{\delta^2 \log p(\theta_j)}{\delta M_j^2} = \Psi_1(M_j) - \Psi_1(\mu_j M_j)\mu_j^2 - \Psi_1((1-\mu_j)M_j)(1-\mu_j)^2 \tag{12}$$

Now we have the Jeffreys' prior for $M_j$:

$$\pi\left(M_j\right) = [-\left(\Psi_1(M_j) - \Psi_1(\mu_j M_j)\mu_j^2 - \Psi_1((1-\mu_j)M_j)(1-\mu_j)^2\right)]^{\frac{1}{2}} \tag{13}$$

## 3. Reference Prior

We want to make the prior as less informative as possible, so to maximize a certain distance measure of prior and posterior is considered using Kullback-Leibler divergence.

3.1. **KL-Divergence.** For distributions $p$ and $q$, their KL-divergence is defined as $KL\left(p||\right) = E_p[\log\left(p/q\right)] = int_p\left(x\right)\log\frac{p(x)}{q(x)}dx$

3.2. **Definition for Reference Prior.** Suppose posterior is $p\left(\theta|x\right)$, depending on pror $p\left(\theta\right)$, marginal is $p\left(x\right)$, then $p\left(\theta\right)$ is called the reference pror if it maximizes $max_{p(\theta)}E_{p(x)}KL\left(p\left(\theta|x\right)||p\left(\theta\right)\right)$ The Jeffreys prior is proved to be reference prior in one dimension (by Michael I. Jordan).

## 4. Log-normal Prior

In log-normal distribution, the parameters denoted $\mu$ and $\sigma$, the mean and standard deviation respectively. Log-normal prior was used on $M_j$.

4.1. **Initialization.** The initial values is also obtained by a method-of-moments (MoM) procedure. We present the initial parameter estimates below.

Since the distribution is $M_j \sim$ log-normal$(\mu, \sigma)$. The first and second population moments are

$$E[M_j] = \qquad e^{\mu+\frac{\sigma^2}{2}}, \tag{14}$$

$$E[{M_j}^2] = \left(e^{\sigma^2}-1\right)[E[M_j]]^2. \tag{15}$$

We can acquire the sample moments, then parameters $\mu$ and $\sigma$ can be obtained

$$\mu = \ln\left(E[M_j]\right) - \frac{\sigma^2}{2} \tag{16}$$

$$\sigma^2 = \ln\left(1 + \frac{E[{M_j}^2]}{E[M_j]^2}\right) \tag{17}$$

4.2. **Sampling from $p\left(M_j|\mu,\sigma,\theta_{ji},\mu_j\right)$.** The posterior distribution over $M_j$ given its Markov blanket is

$$p(M_j|\mu,\sigma,\theta_{ji},\mu_j) \propto p(\theta_{ji}|\mu_j, M_j)p(M_j|\mu,\sigma) \tag{18}$$

We have $\theta_{ji} \sim$ Beta$(\mu_j, M_j)$, and $M_j \sim$ log-normal$(\mu,\sigma)$. Since we cannot give an analytical form for this, so we sample from the posterior distribution using the Metropolis-Hastings algorithm.