0 Introduction

The goal of this document is to fully characterize the Dunson and Stanford day-specific probabilities model. In its current state it tries to provide full detail of the derivations described in *Bayesian Inferences on Predictors of Conception Probabilities*.

1 The day-specific probabilities model

1.1 Model specification

We wish to model the probability of a woman becoming pregnant for a given menstrual cycle as a function of her covariate status across the days of the cycle. Consider a study cohort and let us index

woman i, $\quad i = 1, \dots, n$

cycle j, $\quad j = 1, \dots, n_i$

day k, $\quad k = 1, \dots, K$

where day k refers to the k^{th} day out of a total of K days in the fertile window. Let us write day i, j, k as a shorthand for individual i, cycle j, and day k and similarly for cycle j, k. Then define

Y_{ij} an indicator of conception for woman i, cycle j

V_{ijk} an indicator of conception for woman i, cycle j, day k

X_{ijk} an indicator of intercourse for woman i, cycle j, day k

Then writing $\boldsymbol{X}_{ij} = (X_{ij1}, \dots, X_{ijK})$, we observe that

$$
\begin{aligned}
&\mathbb{P}\Big(Y_{ij} = 1 \,\Big|\, \boldsymbol{X}_{ij},\ Y_{i1} = 0, \dots, Y_{i,j-1} = 0\Big) \\
&\quad = 1 - \mathbb{P}\Big(Y_{ij} = 0 \,\Big|\, \boldsymbol{X}_{ij},\ Y_{i1} = 0, \dots, Y_{i,j-1} = 0\Big) \\
&\quad = 1 - \mathbb{P}\Big(V_{ijk} = 0,\ k = 1, \dots, K \,\Big|\, \boldsymbol{X}_{ij},\ Y_{i1} = 0, \dots, Y_{i,j-1} = 0\Big) \\
&\quad = 1 - \prod_{k=1}^{K} \mathbb{P}\Big(V_{ijk} = 0 \,\Big|\, X_{ijk},\ Y_{i1} = 0, \dots, Y_{i,j-1} = 0,\ V_{ij1} = 0, \dots, V_{i,k-1} = 0\Big) \\
&\quad = 1 - \prod_{k=1}^{K} \Big\{1 - \mathbb{P}\Big(V_{ijk} = 1 \,\Big|\, X_{ijk},\ Y_{i1} = 0, \dots, Y_{i,j-1} = 0,\ V_{ij1} = 0, \dots, V_{i,k-1} = 0\Big)\Big\} \\
&\quad = 1 - \prod_{k=1}^{K} \Big\{1 - X_{ijk}\, \mathbb{P}\Big(V_{ijk} = 1 \,\Big|\, Y_{i1} = 0, \dots, Y_{i,j-1} = 0,\ V_{ij1} = 0, \dots, V_{i,k-1} = 0\Big)\Big\} \\
&\quad = 1 - \prod_{k=1}^{K} \Big\{1 - \mathbb{P}\Big(V_{ijk} = 1 \,\Big|\, Y_{i1} = 0, \dots, Y_{i,j-1} = 0,\ V_{ij1} = 0, \dots, V_{i,k-1} = 0\Big)\Big\}^{X_{ijk}}
\end{aligned}
$$

With this result in mind, we now consider the Dunson and Stanford day-specific probabilities model. Using the same indexing scheme as above, define

$\boldsymbol{u}_{ijk}$	a covariate vector of length q for woman i, cycle j, day k
$\boldsymbol{\beta}$	a vector of length q of regression coefficients
ξ_i	woman-specific random effect

Then writing $\boldsymbol{U}_{ij} = \left(\boldsymbol{u}'_{ijk}, \ldots, \boldsymbol{u}'_{ijk}\right)'$, Dunson and Stanford propose the model:

$$\begin{aligned}
\mathbb{P}\left(Y_{ij} = 1 \mid \xi_i, \boldsymbol{X}_{ij}, \boldsymbol{U}_{ij}\right) &= 1 - \prod_{k=1}^{K} (1 - \lambda_{ijk})^{X_{ijk}} \\
\lambda_{ijk} &= 1 - \exp\left\{-\xi_i \exp\left(\boldsymbol{u}'_{ijk}\boldsymbol{\beta}\right)\right\} \\
\xi_i &\sim \text{Gamma}(\phi, \phi)
\end{aligned} \tag{1}$$

From our previous derivation, we see that we may interpret λ_{ijk} as the day-specific probability of conception in cycle j from couple i given that conception has not already occured, or in the language of Dunson and Stanford, given intercourse only on day k.

Delving further, we see that λ_{ijk} is strictly increasing in $u_{ijkh}\beta_h$, where we are denoting u_{ijkh} to be the h^{th} term in $\boldsymbol{u}_{ijk}$ and similarly for β_h. When $\beta_h = 0$ then the h^{th} covariate has no effect on the day-specific probability of conception.

λ_{ijk} is also strictly increasing in ξ_i which as Dunson and Stanford suggest may be interpreted as a woman-specific random effect. The authors state that specifying the distribution of the ξ_i with a common parameters prevents nonidentifiability between $\mathbb{E}[\xi_i]$ and the day-specific parameters. Since $\text{Var}[\xi_i] = 1/\phi$ it follows that ϕ may be interpreted as a measure of variability across women.

1.1.1 Computation consideration

As an aside, we note that it may be more computationally convenient to calculate

$$\begin{aligned}
&\mathbb{P}\left(Y_{ij} = 1 \mid \xi_i, \boldsymbol{X}_{ij}, \boldsymbol{U}_{ij}\right) \\
&\quad = 1 - \prod_{k=1}^{K} (1 - \lambda_{ijk})^{X_{ijk}} \\
&\quad = 1 - \prod_{k=1}^{K} \Big[\exp\left\{-\xi_i \exp\left(\boldsymbol{u}'_{ijk}\boldsymbol{\beta}\right)\right\} \Big]^{X_{ijk}} \\
&\quad = 1 - \prod_{k=1}^{K} \exp\left\{-X_{ijk}\xi_i \exp\left(\boldsymbol{u}'_{ijk}\boldsymbol{\beta}\right)\right\} \\
&\quad = 1 - \exp\left\{-\sum_{k=1}^{K} X_{ijk}\xi_i \exp\left(\boldsymbol{u}'_{ijk}\boldsymbol{\beta}\right)\right\}
\end{aligned}$$

1.2 Marginal probability of conception

The marginal probability of conception, obtained by integrating out the couple-specific frailty ξ_i, has form as follows.

$$
\begin{aligned}
&\mathbb{P}(Y_{ij}=1\,|\,\boldsymbol{X}_{ij},\boldsymbol{U}_{ij}) \\
&\quad = \int_0^\infty \mathbb{P}\left(Y_{ij},\xi_i\,|\,\boldsymbol{X}_{ij},\boldsymbol{U}_{ij}\right) d\xi_i \\
&\quad = \int_0^\infty \mathbb{P}\left(Y_{ij},\xi_i\,|\,\boldsymbol{X}_{ij},\boldsymbol{U}_{ij}\right) \mathcal{G}(\xi_i;\,\phi,\phi)\, d\xi_i \\
&\quad = \int_0^\infty \left[1-\prod_{k=1}^K (1-\lambda_{ijk})^{X_{ijk}}\right] \mathcal{G}(\xi_i;\,\phi,\phi)\, d\xi_i \\
&\quad = 1-\int_0^\infty \prod_{k=1}^K (1-\lambda_{ijk})^{X_{ijk}}\, \mathcal{G}(\xi_i;\,\phi,\phi)\, d\xi_i \\
&\quad = 1-\int_0^\infty \prod_{k=1}^K \left[\exp\left\{-\xi_i \exp\left(\boldsymbol{u}'_{ijk}\boldsymbol{\beta}\right)\right\}\right]^{X_{ijk}} \mathcal{G}(\xi_i;\,\phi,\phi)\, d\xi_i \\
&\quad = 1-\int_0^\infty \prod_{k=1}^K \exp\left\{-\xi_i X_{ijk} \exp\left(\boldsymbol{u}'_{ijk}\boldsymbol{\beta}\right)\right\} \mathcal{G}(\xi_i;\,\phi,\phi)\, d\xi_i \\
&\quad = 1-\int_0^\infty \exp\left\{-\xi_i \sum_{k=1}^K X_{ijk} \exp\left(\boldsymbol{u}'_{ijk}\boldsymbol{\beta}\right)\right\} \mathcal{G}(\xi_i;\,\phi,\phi)\, d\xi_i \\
&\quad = 1-\left[\frac{\phi}{\phi+\sum_{k=1}^K X_{ijk}\exp\left(\boldsymbol{u}'_{ijk}\boldsymbol{\beta}\right)}\right]^{\phi}
\end{aligned}
$$

since

$$
\begin{aligned}
&\int_0^\infty \exp\left\{-\xi_i \sum_{k=1}^K X_{ijk} \exp\left(\boldsymbol{u}'_{ijk}\boldsymbol{\beta}\right)\right\} \mathcal{G}(\xi_i;\,\phi,\phi)\, d\xi_i \\
&\quad = \int_0^\infty \exp\left\{-\xi_i \sum_{k=1}^K X_{ijk} \exp\left(\boldsymbol{u}'_{ijk}\boldsymbol{\beta}\right)\right\} \frac{\phi^\phi}{\Gamma(\phi)}\, \xi_i^{\phi-1} d\xi_i \\
&\quad = \int_0^\infty \frac{\phi^\phi}{\Gamma(\phi)}\, \xi_i^{\phi-1} \exp\left\{-\xi_i \left[\phi+\sum_{k=1}^K X_{ijk} \exp\left(\boldsymbol{u}'_{ijk}\boldsymbol{\beta}\right)\right]\right\} d\xi_i \\
&\quad = \left[\frac{\phi}{\phi+\sum_{k=1}^K X_{ijk}\exp\left(\boldsymbol{u}'_{ijk}\boldsymbol{\beta}\right)}\right]^{\phi} \int_0^\infty \frac{\left[\phi+\sum_{k=1}^K X_{ijk}\exp\left(\boldsymbol{u}'_{ijk}\boldsymbol{\beta}\right)\right]^{\phi}}{\Gamma(\phi)} \\
&\qquad\qquad \times \xi_i^{\phi-1} \exp\left\{-\xi_i \left[\phi+\sum_{k=1}^K X_{ijk} \exp\left(\boldsymbol{u}'_{ijk}\boldsymbol{\beta}\right)\right]^{\phi}\right\} d\xi_i
\end{aligned}
$$

and the function inside the integral is a gamma density function.

1.2.1 Day-specific marginal probability of conception

Dunson and Stanford also point out the following remarkable result. The marginal day-specific probability of conception in a cycle with intercourse only on day k and with predictors $\boldsymbol{u}$ is given by

$$\mathbb{P}(Y=1\,|\,\boldsymbol{u}) = 1 - \left(\frac{\phi}{\phi + \exp(\boldsymbol{u}'\boldsymbol{\beta})}\right)^{\phi}$$

which is in the form of the Aranda-Ordaz generalized linear model, and reduces to a logistic regression model for $\phi = 1$.

1.3 Prior specification

Define

$\mathcal{G}_{\mathcal{A}_h}(\bullet)$	density function of a gamma distribution truncated to the region $\mathcal{A}_h \subset (0,\infty)$
γ_h	$\exp(\beta_h)$

Then the Dunson and Stanford model chooses priors of the form

$$\pi(\boldsymbol{\gamma}) = \prod_{h=1}^{q} \Big\{ p_h I\big(\gamma_h = 1\big) \;+\; (1-p_h) I\big(\gamma_h \neq 1\big) \mathcal{G}_{\mathcal{A}_h}\big(\gamma_h;\; a_h, b_h\big) \Big\}$$

$$\pi(\phi) = \mathcal{G}\left(\phi;\; c_1, c_2\right)$$

where

p_h	prior probability that $\gamma_h = 1$, a hyperparameter
a_h, b_h	shape and rate hyperparameters for gamma distribution of γ_h
c_1, c_2	shape and rate hyperparameters for gamma distribution of ϕ

Values of $\gamma_h = 1$ correspond to $\beta_h = 0$ and the h^{th} predictor in $\boldsymbol{u}_{ijk}$ being dropped from the model. Thus assigning the prior for each of the γ_h to be a mixture distribution between a point mass at one and a gamma distribution allows the model to drop terms from the regression component with nonzero probability.

Typical constraints for the γ_h are $\mathbb{R}^+$, $(0,1)$, and $(1,\infty)$ which correspond to no constraint, a negative effect on probability of conception, and a positive effect on probability of conception, respectively. Thus a priori knowledge of the direction of association of the predictor variables can be incorporated into the model to decrease posterior uncertainty.

1.3.1 Monotone effects

Consider a model where the list of covariates includes an ordered categorical variable with types $1,\ldots,t$. Let $\boldsymbol{s}_{ijk} = \big(s_{ijk,2},\ldots,s_{ijk,t}\big)$ be a vector of length $(t-1)$ for each day i,j,k where

$$\begin{aligned}
s_{ijk,2} &= I\Big(\text{categorical variable for day } i,j,k \text{ is type 2}\Big) \\
s_{ijk,3} &= I\Big(\text{categorical variable for day } i,j,k \text{ is type 2 or 3}\Big) \\
\vdots \;\; & \;\; \vdots \qquad\qquad\qquad \vdots \\
s_{ijk,t} &= I\Big(\text{categorical variable for day } i,j,k \text{ is type 2 or 3 or} \ldots \text{or } t\Big)
\end{aligned}$$

Next, let us partition each covariate vector $\boldsymbol{u}_{ijk} = \left(\boldsymbol{r}_{ijk}, \boldsymbol{s}_{ijk}\right)$ so that $\boldsymbol{r}_{ijk}$ is a vector of the remaining covariate terms. Furthermore let $\boldsymbol{\beta} = \left(\boldsymbol{\tau}, \boldsymbol{\alpha}\right)$ be the corresponding partition of covariate coefficients where $\boldsymbol{\alpha} = \left(\alpha_2, \ldots, \alpha_t\right)$. Then for person i, cycle j, and day k with categorical variable type d where $d \in \{1, \ldots, t\}$, then

$$\begin{aligned}
\lambda_{ijk} &= 1 - \exp\left\{-\xi_i \exp\left(\boldsymbol{u}'_{ijk}\boldsymbol{\beta}\right)\right\} \\
&= 1 - \exp\left\{-\xi_i \exp\left(\boldsymbol{r}'_{ijk}\boldsymbol{\tau} + \boldsymbol{s}'_{ijk}\boldsymbol{\alpha}\right)\right\} \\
&= 1 - I\left(d=1\right)\exp\left\{-\xi_i \exp\left(\boldsymbol{r}'_{ijk}\boldsymbol{\tau}\right)\right\} - I\left(d \geq 2\right)\exp\left\{-\xi_i \exp\left(\boldsymbol{r}'_{ijk}\boldsymbol{\tau} + \sum_{m=2}^{d}\alpha_m\right)\right\}
\end{aligned}$$

From this form we can see that when $\alpha_m \geq 0,\ m = 2, \ldots, t$ then λ_{ijk} is nondecreasing in m. It follows that a monotone increasing categorical variable can be created by coding the variable in the format as described above, and constraining the corresponding parameters of γ_h to be greater than or equal to one (corresonding to $\beta_h \geq 0$ for each of the corresonding h). Similarly, a monotone decreasing categorical variable can be created by coding the variable as described above, and constraining the corresponding parameters of γ_h to be less than or equal to one.

2 Posterior computation

Express the data augmentation model as

$$\begin{aligned}
&Y_{ij} = I\left(\sum_{k=1}^{K} X_{ijk} Z_{ijk} > 0\right), \\
&Z_{ijk} \sim \text{Poisson}\left(\xi_i \exp\left(\boldsymbol{u}'_{ijk}\boldsymbol{\beta}\right)\right), \qquad k = 1, \ldots, K
\end{aligned} \tag{2}$$

Let us further define $W_{ijk} = X_{ijk}Z_{ijk}$ for all i, j, k.

2.1 Verifying the equivalence of the data augmentation model

Under (2), $Y_{ij} = 0$ if and only if $W_{ij1}, \ldots, W_{ijK}$ are identically 0. It follows that

$$\begin{aligned}
&\mathbb{P}\left(Y_{ij} = 0 \mid \xi_i, \boldsymbol{X}_{ij}, \boldsymbol{U}_{ij}\right) \\
&\quad = \prod_{k:X_{ijk}=1} \mathbb{P}\left(W_{ijk} = 0 \mid \xi_i, \boldsymbol{u}_{ijk}\right) \\
&\quad = \prod_{k=1}^{K} \left[\mathbb{P}\left(W_{ijk} = 0 \mid \xi_i, \boldsymbol{u}_{ijk}\right)\right]^{X_{ijk}} \\
&\quad = \prod_{k=1}^{K} \left[\exp\left\{\xi_i \exp\left(\boldsymbol{u}'_{ijk}\boldsymbol{\beta}\right)\right\}\right]^{X_{ijk}} \\
&\quad = \prod_{k=1}^{K} \left(1 - \lambda_{ijk}\right)^{X_{ijk}}
\end{aligned}$$

which is the model in (1).

2.2 The full likelihood

Let $\boldsymbol{Y}$ be a random variable representing all of the potential pregnancy indicators Y_{ij}, let $\boldsymbol{W}$ be a random variable representing all of the latent variables W_{ijk}, and let ξ be a random variable representing all of the woman-specific random effects ξ_i. Then

$$
\begin{aligned}
&\pi\big(\boldsymbol{Y}, \boldsymbol{W}, \boldsymbol{\gamma}, \xi, \phi \mid \text{data}\big) \\
&\quad = \pi\big(\boldsymbol{Y} \mid \boldsymbol{W}, \boldsymbol{\gamma}, \xi, \phi, \text{data}\big)\, \pi\big(\boldsymbol{W} \mid \boldsymbol{\gamma}, \xi, \phi, \text{data}\big)\, \pi\big(\xi \mid \boldsymbol{\gamma}, \phi, \text{data}\big)\, \pi\big(\boldsymbol{\gamma} \mid \phi, \text{data}\big)\, \pi\big(\phi \mid \text{data}\big) \\
&\quad = \pi\big(\boldsymbol{Y} \mid \boldsymbol{W}\big)\, \pi\big(\boldsymbol{W} \mid \boldsymbol{\gamma}, \xi, \text{data}\big)\, \pi\big(\xi \mid \phi\big)\, \pi\big(\boldsymbol{\gamma}\big)\, \pi\big(\phi\big) \\
&\quad = \Bigg(\prod_{i,j} \pi\big(Y_{ij} \mid \boldsymbol{W}_{ij}\big)\Bigg)\Bigg(\prod_{i,j,k:X_{ijk}=1} \pi\big(W_{ijk} \mid \boldsymbol{\gamma}, \xi\big)\Bigg)\Bigg(\prod_{i=1}^{n} \pi\big(\xi_i \mid \phi\big)\Bigg)\Bigg(\prod_{\ell=1}^{q} \pi\big(\gamma_h\big)\Bigg)\, \pi\big(\phi\big) \\
&\quad = \Bigg\{\prod_{i,j}\Bigg[\, I\Bigg(\sum_{k=1}^{K} W_{ijk} > 0\Bigg) Y_{ij} + I\Bigg(\sum_{k=1}^{K} W_{ijk} = 0\Bigg)\big(1 - Y_{ij}\big)\Bigg]\Bigg\} \\
&\qquad\qquad \times \Bigg(\prod_{i,j,k:X_{ijk}=1} \frac{1}{W_{ijk}!}\Bigg[\xi_i \exp\Bigg(\sum_{\ell=1}^{q} u_{ijk\ell} \log \gamma_\ell\Bigg)\Bigg]^{W_{ijk}} \exp\Bigg\{-\xi_i \exp\Bigg(\sum_{\ell=1}^{q} u_{ijk\ell} \log \gamma_\ell\Bigg)\Bigg\}\Bigg) \\
&\qquad\qquad \times \Bigg(\prod_{i=1}^{n} \frac{\phi^{\phi}}{\Gamma(\phi)}\, \xi_i^{\phi-1} \exp\big(-\phi \xi_i\big)\Bigg) \\
&\qquad\qquad \times \Bigg(\prod_{\ell=1}^{q}\Big[p_h I\big(\gamma_h = 1\big) \;+\; (1-p_h)\, I\big(\gamma_h \neq 1\big)\, \mathcal{G}_{\mathcal{A}_h}(\gamma_h;\, a_h, b_h)\Big]\Bigg) \\
&\qquad\qquad \times \frac{c_2^{c_1}}{\Gamma(c_1)}\, \phi^{c_1-1} \exp\big(-c_2 \phi\big)
\end{aligned}
$$

2.3 The full conditional distributions

Step 1. Writing $\boldsymbol{W}_{ij} = \big(W_{ij1}, \ldots, W_{ijK}\big)$ and letting $\boldsymbol{m} = \big(m_1, \ldots, m_K\big)$ be a vector of realized outcomes for $\boldsymbol{W}_{ij}$, we see first that for $Y_{ij} = 0$ we have

$$
\mathbb{P}\Big(\boldsymbol{W}_{ij} = \boldsymbol{m} \;\Big|\; Y_{ij} = 0,\, \boldsymbol{\beta}, \phi, \boldsymbol{\xi}, \text{data}\Big) \;=\; \begin{cases} 1, & \boldsymbol{m} = \boldsymbol{0} \\ 0, & \text{else} \end{cases}
$$

Next, for $Y_{ij} = 1$ we have

$$
\begin{aligned}
&\mathbb{P}\Big(\boldsymbol{W}_{ij} = \boldsymbol{m} \;\Big|\; Y_{ij} = 1,\, \boldsymbol{\beta}, \phi, \boldsymbol{\xi}, \text{data}\Big) \\
&\quad = \sum_{s=0}^{\infty} \mathbb{P}\Big(\boldsymbol{W}_{ij} = \boldsymbol{m},\; \textstyle\sum_k W_{ijk} = s \;\Big|\; Y_{ij} = 1,\, \boldsymbol{\beta}, \phi, \boldsymbol{\xi}, \text{data}\Big) \\
&\quad = \mathbb{P}\Big(\boldsymbol{W}_{ij} = \boldsymbol{m},\; \textstyle\sum_k W_{ijk} = \sum_k m_k \;\Big|\; Y_{ij} = 1,\, \boldsymbol{\beta}, \phi, \boldsymbol{\xi}, \text{data}\Big) \\
&\quad = \mathbb{P}\Big(\boldsymbol{W}_{ij} = \boldsymbol{m} \;\Big|\; \textstyle\sum_k W_{ijk} = \sum_k m_k,\; Y_{ij} = 1,\, \boldsymbol{\beta}, \phi, \boldsymbol{\xi}, \text{data}\Big) \\
&\qquad\qquad \times \mathbb{P}\Big(\textstyle\sum_k W_{ijk} = \sum_k m_k \;\Big|\; Y_{ij} = 1,\, \boldsymbol{\beta}, \phi, \boldsymbol{\xi}, \text{data}\Big)
\end{aligned}
$$

Furthermore,

$$\begin{aligned}
&\pi\left(\sum_{k=1}^{K} W_{ijk} \;\middle|\; Y_{ij}=1, \boldsymbol{\beta}, \phi, \boldsymbol{\xi}, \text{data}\right) \\
&\quad = \pi\left(\sum_{k=1}^{K} W_{ijk} \;\middle|\; \sum_{k=1}^{K} W_{ijk} \geq 1, \boldsymbol{\beta}, \phi, \boldsymbol{\xi}, \text{data}\right) \\
&\quad \sim \text{Poisson}\left(\xi_i \sum_{k:\, X_{ijk}=1} \exp\left(\boldsymbol{u}'_{ijk}\boldsymbol{\beta}\right)\right) \text{ truncated so that } \sum_{k=1}^{K} W_{ijk} \geq 1
\end{aligned}$$

and

$$\begin{aligned}
&\pi\left(\boldsymbol{W}_{ij} \;\middle|\; \sum_{k=1}^{K} W_{ijk},\; Y_{ij}=1, \boldsymbol{\beta}, \phi, \boldsymbol{\xi}, \text{data}\right) \\
&\quad \sim \text{Multinomial}\left(\sum_{k=1}^{K} W_{ijk};\; \frac{X_{ij1}\xi_i \exp\left(\boldsymbol{u}'_{ij1}\boldsymbol{\beta}\right)}{\xi_i \sum_{k:\, X_{ijk}=1} \exp\left(\boldsymbol{u}'_{ijk}\boldsymbol{\beta}\right)}, \;\ldots,\; \frac{X_{ijK}\xi_i \exp\left(\boldsymbol{u}'_{ijK}\boldsymbol{\beta}\right)}{\xi_i \sum_{k:\, X_{ijk}=1} \exp\left(\boldsymbol{u}'_{ijk}\boldsymbol{\beta}\right)}\right)
\end{aligned}$$

Step 2. Define the following terms which will be of use in the following derivation. Denote

$$\begin{array}{ll}
\widetilde{a}_h & a_h + \sum_{i,j,k} u_{ijkh} W_{ijk} \\
\widetilde{b}_h & b_h + \displaystyle\sum_{\substack{i,j,k:X_{ijk}=1,\\ u_{ijkh}=1}} \xi_i \prod_{\ell \neq h} \gamma_\ell^{u_{ijk\ell}} \\
d_1 & p_h \exp\left\{-(\widetilde{b}_h - b_h)\right\} \\
d_2 & (1-p_h)\,\dfrac{C(a_h, b_h)\int_{\mathcal{A}_h} \mathcal{G}(\gamma;\, \widetilde{a}_h, \widetilde{b}_h)\, d\gamma}{C(\widetilde{a}_h, \widetilde{b}_h)\int_{\mathcal{A}_h} \mathcal{G}(\gamma;\, a_h, b_h)\, d\gamma} \\
\widetilde{p}_h & \dfrac{d_1}{d_1 + d_2}
\end{array}$$

Then for the case when the explanetory variables are all categorical, we have

$$\begin{aligned}
&\pi\left(\gamma_h \,|\, \boldsymbol{\gamma}_{(-h)}, \phi, \boldsymbol{\xi}, \boldsymbol{W}, \text{data}\right) \\
&\quad \propto \pi\left(\boldsymbol{W} \mid \boldsymbol{\xi}, \gamma, \text{data}\right) \pi\left(\gamma_h\right) \\
&\quad = \left(\prod_{i=1}^{n}\prod_{j=1}^{n_i}\prod_{k:X_{ijk}=1} \pi\left(W_{ijk} \,|\, \boldsymbol{\xi}_i, \gamma, \text{data}\right)\right) \pi\left(\gamma_h\right) \\
&\quad \propto \left(\prod_{i=1}^{n}\prod_{j=1}^{n_i}\prod_{k:X_{ijk}=1} \left[\exp\left(u_{ijkh}\log\gamma_h\right)\right]^{W_{ijk}} \exp\left\{-\xi_i \exp\left(\sum_{\ell=1}^{q} u_{ijk\ell}\log\gamma_\ell\right)\right\}\right) \pi\left(\gamma_h\right)
\end{aligned}$$

$$
\begin{aligned}
&= \left(\prod_{i=1}^{n} \prod_{j=1}^{n_i} \prod_{k: X_{ijk}=1} \gamma_h^{u_{ijkh} W_{ijk}} \exp\left\{ -\xi_i \prod_{\ell=1}^{q} \gamma_\ell^{u_{ijk\ell}} \right\} \right) \pi(\gamma_h) \\
&= \gamma_h^{\sum_{i,j,k} u_{ijkh} W_{ijk}} \exp\left\{ - \sum_{i,j,k: X_{ijk}=1} \xi_i \prod_{\ell=1}^{q} \gamma_\ell^{u_{ijk\ell}} \right\} \pi(\gamma_h) \\
&= \gamma_h^{\sum_{i,j,k} u_{ijkh} W_{ijk}} \exp\left\{ - \sum_{\substack{i,j,k: X_{ijk}=1, \\ u_{ijkh}=0}} \xi_i \prod_{\ell \neq h} \gamma_\ell^{u_{ijk\ell}} - \gamma_h \sum_{\substack{i,j,k: X_{ijk}=1, \\ u_{ijkh}=1}} \xi_i \prod_{\ell \neq h} \gamma_\ell^{u_{ijk\ell}} \right\} \pi(\gamma_h) \\
&\propto \gamma_h^{\sum_{i,j,k} u_{ijkh} W_{ijk}} \exp\left\{ -\gamma_h \sum_{\substack{i,j,k: X_{ijk}=1, \\ u_{ijkh}=1}} \xi_i \prod_{\ell \neq h} \gamma_\ell^{u_{ijk\ell}} \right\} \pi(\gamma_h) \\
&= \gamma_h^{\sum_{i,j,k} u_{ijkh} W_{ijk}} \exp\left\{ -\gamma_h \sum_{\substack{i,j,k: X_{ijk}=1, \\ u_{ijkh}=1}} \xi_i \prod_{\ell \neq h} \gamma_\ell^{u_{ijk\ell}} \right\} \\
&\qquad\qquad \times \left[p_h I(\gamma_h = 1) + (1 - p_h) I(\gamma_h \neq 1) \mathcal{G}_{\mathcal{A}_h}(\gamma_h; a_h, b_h) \right] \\
&= p_h I(\gamma_h = 1) \exp\left\{ - \sum_{\substack{i,j,k: X_{ijk}=1, \\ u_{ijkh}=1}} \xi_i \prod_{\ell \neq h} \gamma_\ell^{u_{ijk\ell}} \right\} \\
&\qquad + (1 - p_h) I(\gamma_h \neq 1) \gamma_h^{\sum_{i,j,k} u_{ijkh} W_{ijk}} \exp\left\{ -\gamma_h \sum_{\substack{i,j,k: X_{ijk}=1, \\ u_{ijkh}=1}} \xi_i \prod_{\ell \neq h} \gamma_\ell^{u_{ijk\ell}} \right\} \mathcal{G}_{\mathcal{A}_h}(\gamma_h; a_h, b_h) \\
&= p_h I(\gamma_h = 1) \exp\left\{ - \sum_{\substack{i,j,k: X_{ijk}=1, \\ u_{ijkh}=1}} \xi_i \prod_{\ell \neq h} \gamma_\ell^{u_{ijk\ell}} \right\} \\
&\qquad\qquad + (1 - p_h) I(\gamma_h \neq 1) \frac{C(a_h, b_h)}{\int_{\mathcal{A}_h} \mathcal{G}(\gamma; a_h, b_h)\, d\gamma} \gamma_h^{a_h + \sum_{i,j,k} u_{ijkh} W_{ijk} - 1} \\
&\qquad\qquad \times \exp\left\{ -\gamma_h \left[b_h + \sum_{\substack{i,j,k: X_{ijk}=1, \\ u_{ijkh}=1}} \xi_i \prod_{\ell \neq h} \gamma_\ell^{u_{ijk\ell}} \right] \right\} \\
&= p_h I(\gamma_h = 1) \exp\{ -(\widetilde{b}_h - b_h) \} + (1 - p_h) I(\gamma_h \neq 1) \frac{C(a_h, b_h)}{\int_{\mathcal{A}_h} \mathcal{G}(\gamma; a_h, b_h)\, d\gamma} \gamma_h^{\widetilde{a}_h - 1} \exp\{ -\widetilde{b}_h \gamma_h \} \\
&= p_h I(\gamma_h = 1) \exp\{ -(\widetilde{b}_h - b_h) \} + (1 - p_h) I(\gamma_h \neq 1)
\end{aligned}
$$

$$
\times \; \frac{C(a_h,b_h)\int_{\mathcal{A}_h}\mathcal{G}(\gamma;\widetilde{a}_h,\widetilde{b}_h)\,d\gamma}{C(\widetilde{a}_h,\widetilde{b}_h)\int_{\mathcal{A}_h}\mathcal{G}(\gamma;a_h,b_h)\,d\gamma}\,\frac{C(\widetilde{a}_h,\widetilde{b}_h)}{\int_{\mathcal{A}_h}\mathcal{G}(\gamma;\widetilde{a}_h,\widetilde{b}_h)\,d\gamma}\gamma_h^{\widetilde{a}_h-1}\exp\{-\widetilde{b}_h\gamma_h\}
$$

$$
\begin{aligned}
&= p_h\, I\big(\gamma_h=1\big)\exp\{-(\widetilde{b}_h-b_h)\} + (1-p_h)\,I\big(\gamma_h\neq 1\big)\\
&\qquad\qquad \times \frac{C(a_h,b_h)\int_{\mathcal{A}_h}\mathcal{G}(\gamma;\widetilde{a}_h,\widetilde{b}_h)\,d\gamma}{C(\widetilde{a}_h,\widetilde{b}_h)\int_{\mathcal{A}_h}\mathcal{G}(\gamma;a_h,b_h)\,d\gamma}\,\mathcal{G}_{\mathcal{A}_h}\big(\gamma;\widetilde{a}_h,\widetilde{b}_h\big)\\
&= d_1 I\big(\gamma_h=1\big) + d_2 I\big(\gamma_h\neq 1\big)\mathcal{G}_{\mathcal{A}_h}\big(\gamma;\widetilde{a}_h,\widetilde{b}_h\big)\\
&\propto \frac{d_1}{d_1+d_2} I\big(\gamma_h=1\big) + \frac{d_2}{d_1+d_2} I\big(\gamma_h\neq 1\big)\mathcal{G}_{\mathcal{A}_h}\big(\gamma;\widetilde{a}_h,\widetilde{b}_h\big)\\
&= \widetilde{p}_h I\big(\gamma_h=1\big) + (1-\widetilde{p}_h) I\big(\gamma_h\neq 1\big)\mathcal{G}_{\mathcal{A}_h}\big(\gamma;\widetilde{a}_h,\widetilde{b}_h\big)
\end{aligned}
$$

Step 3.

$$
\begin{aligned}
&\pi\left(\xi_i \;\middle|\; \boldsymbol{\beta},\phi,\boldsymbol{W},\text{data}\right)\\
&\quad\propto \pi\left(\boldsymbol{W}_i \;\middle|\; \boldsymbol{\beta},\xi_i,\text{data}\right)\pi\left(\xi_i\,|\,\phi,\text{data}\right)\\
&\quad= \left(\prod_{j,k:X_{ijk}=1}\pi\left(\boldsymbol{W}_{ijk} \;\middle|\; \boldsymbol{\beta},\xi_i,\text{data}\right)\right)\pi\left(\xi_i\,|\,\phi,\text{data}\right)\\
&\quad\propto \left(\prod_{j,k:X_{ijk}=1}\xi_i^{W_{ijk}}\exp\left\{-\xi_i\exp\left(\boldsymbol{u}'_{ijk}\boldsymbol{\beta}\right)\right\}\right)\xi_i^{\phi-1}\exp\left\{-\phi\,\xi_i\right\}\\
&\quad= \left(\xi_i^{\sum_{j,k}W_{ijk}}\exp\left\{-\xi_i\sum_{j,k:X_{ijk}=1}\exp\left(\boldsymbol{u}'_{ijk}\boldsymbol{\beta}\right)\right\}\right)\xi_i^{\phi-1}\exp\left\{-\phi\,\xi_i\right\}\\
&\quad= \xi_i^{\phi+\sum_{j,k}W_{ijk}-1}\exp\left\{-\xi_i\left[\phi+\sum_{j,k:X_{ijk}=1}\exp\left(\boldsymbol{u}'_{ijk}\boldsymbol{\beta}\right)\right]\right\}\\
&\quad\sim \text{Gamma}\left(\phi+\sum_{j,k}W_{ijk},\quad \phi+\sum_{j,k:X_{ijk}=1}\exp\left(\boldsymbol{u}'_{ijk}\boldsymbol{\beta}\right)\right)
\end{aligned}
$$

Step 4. Sampling ϕ can be achieved via the Metropolis algorithms. Let $\phi^{(s)}$ denote the value of ϕ for the s^{th} scan of the MCMC algorithm, and let ϕ^* denote a proposed value of ϕ for the $(s+1)^{\text{th}}$ scan of the algorithm. We consider the following two proposal distributions where δ is a tuning parameter with value greater than 0.

(i) $J\left(\phi^*\,|\,\phi^{(s)}\right)\sim\left|N\left(\phi^{(s)},\delta^2\right)\right|$

(ii) $J\left(\phi^* \mid \phi^{(s)}\right) \sim \left|\text{Uniform}\left(\phi^{(s)} - \delta,\ \phi^{(s)} + \delta\right)\right|$

Now,

$$
\begin{aligned}
&\pi\left(\phi \mid \boldsymbol{Y}, \boldsymbol{W}, \boldsymbol{\beta}, \xi, \text{data}\right) \\
&\quad = \frac{\pi\left(\boldsymbol{Y}, \boldsymbol{W}, \boldsymbol{\beta}, \xi, \phi, \text{data}\right)}{\pi\left(\boldsymbol{Y}, \boldsymbol{W}, \boldsymbol{\beta}, \xi, \text{data}\right)} \\
&\quad = \frac{1}{\pi\left(\boldsymbol{Y}, \boldsymbol{W}, \boldsymbol{\beta}, \xi, \text{data}\right)}\, \pi\left(\boldsymbol{Y} \mid \boldsymbol{W}, \boldsymbol{\beta}, \xi, \phi, \text{data}\right) \pi\left(\boldsymbol{W} \mid \boldsymbol{\beta}, \xi, \phi, \text{data}\right) \\
&\qquad\qquad\qquad\qquad\qquad \times\ \pi\left(\xi \mid \phi, \text{data}\right) \pi\left(\phi \mid \text{data}\right) \\
&\quad = \frac{1}{\pi\left(\boldsymbol{Y}, \boldsymbol{W}, \boldsymbol{\beta}, \xi, \text{data}\right)}\, \pi\left(\boldsymbol{Y} \mid \boldsymbol{W}\right) \pi\left(\boldsymbol{W} \mid \boldsymbol{\beta}, \xi, \text{data}\right) \pi\left(\xi \mid \phi\right) \pi\left(\phi\right) \\
&\quad = \left(\prod_{i=1}^{n} \pi\left(\xi_i \mid \phi\right)\right) \pi\left(\phi\right) \frac{\pi\left(\boldsymbol{Y} \mid \boldsymbol{W}\right) \pi\left(\boldsymbol{W} \mid \boldsymbol{\beta}, \xi, \text{data}\right)}{\pi\left(\boldsymbol{Y}, \boldsymbol{W}, \boldsymbol{\beta}, \xi, \text{data}\right)}
\end{aligned}
$$

It follows that the acceptance ratio is given by $\min(r, 1)$ where

$$
r = \frac{\pi\left(\phi^* \mid \boldsymbol{Y}, \boldsymbol{W}, \boldsymbol{\beta}, \xi, \text{data}\right)}{\pi\left(\phi^{(s)} \mid \boldsymbol{Y}, \boldsymbol{W}, \boldsymbol{\beta}, \xi, \text{data}\right)} = \frac{\left(\prod_{i=1}^{n} \pi\left(\xi_i \mid \phi^*\right)\right) \pi\left(\phi^*\right)}{\left(\prod_{i=1}^{n} \pi\left(\xi_i \mid \phi^{(s)}\right)\right) \pi\left(\phi^{(s)}\right)}
$$

2.3.1 Symmetric distributions verfication

Recall the proposal distributions from Step 4 of the MCMC algorithm:

(i) $J\left(\phi^* \mid \phi^{(s)}\right) \sim \left|N\left(\phi^{(s)}, \delta^2\right)\right|$

(ii) $J\left(\phi^* \mid \phi^{(s)}\right) \sim \left|\text{Uniform}\left(\phi^{(s)} - \delta,\ \phi^{(s)} + \delta\right)\right|$

To see that (i) is indeed a symmetric distribution, consider the following. Let $X \sim \text{Normal}(\mu, \delta^2)$, and let $Y = |X|$. Define

$$
\begin{aligned}
&A_0 = \{0\} \\
&A_1 = (-\infty, 0) \qquad g_1(x) = -x \qquad g_1^{-1}(x) = -x \\
&A_2 = (0, \infty) \qquad g_2(x) = x \qquad g_2^{-1}(x) = x
\end{aligned}
$$

Then

$$
\begin{aligned}
\pi_Y(y) &= \sum_{i=1}^{2} f_X\left(g_i^{-1}(y)\right) \left|\frac{d}{dy} g_i^{-1}(y)\right| \\
&= \frac{1}{\sqrt{2\pi\delta^2}} \exp\left\{-\frac{1}{\delta^2}(-y-\mu)^2\right\} |-1| + \frac{1}{\sqrt{2\pi\delta^2}} \exp\left\{-\frac{1}{\delta^2}(y-\mu)^2\right\} |1|
\end{aligned}
$$

Letting $\pi_{J\,(\text{i})}\,(x|y)$ denote the density function of (i), it follows that

$$\pi_{J\,(\text{i})}\left(\phi^*\,|\,\phi^{(s)}\right) = \frac{1}{\sqrt{2\pi\delta^2}}\,\exp\left\{-\frac{1}{\delta^2}\left(-\phi^* - \phi^{(s)}\right)^2\right\} + \frac{1}{\sqrt{2\pi\delta^2}}\,\exp\left\{-\frac{1}{\delta^2}\left(\phi^* - \phi^{(s)}\right)^2\right\}$$

and that

$$\pi_{J\,(\text{i})}\left(\phi^{(s)}\,|\,\phi^*\right) = \frac{1}{\sqrt{2\pi\delta^2}}\,\exp\left\{-\frac{1}{\delta^2}\left(-\phi^{(s)} - \phi^*\right)^2\right\} + \frac{1}{\sqrt{2\pi\delta^2}}\,\exp\left\{-\frac{1}{\delta^2}\left(\phi^{(s)} - \phi^*\right)^2\right\}$$

which are readily seen to be equivalent.

To see that (ii) is indeed a symmetric distribution, consider the following. Let $X \sim \text{Uniform}(a,b)$, and let $Y = |X|$. Then for $a < y < b$,

$$\begin{aligned}
F_Y(y) &= \mathbb{P}(Y \le y) \\
&= \mathbb{P}(|X| \le y) \\
&= \mathbb{P}(-y \le X \le y) \\
&= F_X(y) - F_X(-y) \\
&= \frac{y-a}{b-a} - \frac{-y-a}{b-a}\,I\,(a < -y)
\end{aligned}$$

so that

$$\pi_Y(y) = \frac{1}{b-a} + \frac{1}{b-a}\,I\,(a < -y)$$

Letting $\pi_{J\,(\text{ii})}\,(x|y)$ denote the density function of (ii), it follows that for $\phi^{(s)} < \phi^* < \phi^{(s)}$,

$$\begin{aligned}
\pi_{J\,(\text{ii})}\left(\phi^*\,|\,\phi^{(s)}\right) &= \frac{1}{(\phi^{(s)}+\delta)-(\phi^{(s)}-\delta)} + \frac{1}{(\phi^{(s)}+\delta)-(\phi^{(s)}-\delta)}\,I\,(\phi^{(s)} - \delta < -\phi^*) \\
&= \frac{1}{2\delta} + \frac{1}{2\delta}\,I\big(\phi^{(s)} - \delta < -\phi^*\big)
\end{aligned}$$

and similarly that for $\phi^* < \phi^{(s)} < \phi^*$,

$$\begin{aligned}
\pi_{J\,(\text{ii})}\left(\phi^{(s)}\,|\,\phi^*\right) &= \frac{1}{(\phi^*+\delta)-(\phi^*-\delta)} + \frac{1}{(\phi^*+\delta)-(\phi^*-\delta)}\,I\,(\phi^* - \delta < -\phi^{(s)}) \\
&= \frac{1}{2\delta} + \frac{1}{2\delta}\,I\big(\phi^* - \delta < -\phi^{(s)}\big)
\end{aligned}$$

which after rearranging terms are seen to be equivalent.

2.3.2 Computational considerations

(i) **Sampling from a truncated gamma distribution**

Consider a set (a, b) and let X be a continuous random variable with support $\mathcal{A}$ such that $(a, b) \subset \mathcal{A}$. Define

F_X	the distribution function of X
$U(d_1, d_2)$	a uniform random variable with support on (d_1, d_2)
V	a random variable defined by $V = F_X^{-1}\Big(U\big(F_X(a), F_X(b)\big)\Big)$

Then

$$\begin{aligned}
\mathbb{P}(V \le v) &= \mathbb{P}\left\{F_X^{-1}\Big(U\big(F_X(a), F_X(b)\big)\Big) \le v\right\} \\
&= \mathbb{P}\left\{U\big(F_X(a), F_X(b)\big) \le F_X(v)\right\} \\
&= \begin{cases} 0, & v \le F_X(a) \\ \dfrac{F_X(v) - F_X(a)}{F_X(b) - F_X(a)}, & F_X(a) < v < F_X(b) \\ 1, & F_X(b) \le v \end{cases}
\end{aligned}$$

which is the distribution function of X truncated to (a, b). Thus by choosing F_X to be the distribution function of some desired gamma distribution, we may sample from the truncated gamma distribution by sampling $u \sim U\big(F_X(a), F_X(b)\big)$ and then calculating $F_X^{-1}(u)$.

(ii) **Sampling from a truncated Poisson distribution**

Let X be a random variable with support $\{x_1, x_2, \dots\}$ where $x_i < x_j$ for all $i < j$. Let F_X denote the distribution function of X, and let $G_X\colon (0,1) \mapsto \{x_1, x_2, \dots\}$ be a pseudo-inverse of F_X defined by

$$G_X(p) = \min\left\{x_i \in \{x_1, x_2, \dots\} \colon F_X(x_i) \ge p\right\}$$

Next, let $j_1, j_2, k \in \mathbb{N}$ with $j_1 < j_2$. Note that this implies that $F_X(x_{j_1}) < F_X(x_{j_2})$ if we assume $\mathbb{P}(X = x_i) > 0$ for all $i \in \mathbb{N}$. Define $U(d_1, d_2)$ to be a uniform random variable with support on (d_1, d_2), then

$$\begin{aligned}
&\mathbb{P}\left\{G_X\Big(U\left(F_X(x_{j_1-1}), F_X(x_{j_2})\right)\Big) = x_k\right\} \\
&\quad = \mathbb{P}\left\{U\left(F_X(x_{j_1-1}), F_X(x_{j_2})\right) \in \Big(F_X(x_{k-1}),\, F_X(x_k)\Big)\right\} \\
&\quad = I\left(x_{j_1} \le x_k,\, x_k \le x_{j_2}\right) \int_{F_X(x_{k-1})}^{F_X(x_k)} \frac{1}{F_X(x_{j_2}) - F_X(x_{j_1-1})}\, dy \\
&\quad = I\left(x_{j_1} \le x_k,\, x_k \le x_{j_2}\right) \frac{F_X(x_k) - F_X(x_{k-1})}{F_X(x_{j_2}) - F_X(x_{j_1-1})} \\
&\quad = I\left(x_{j_1} \le x_k,\, x_k \le x_{j_2}\right) \frac{\mathbb{P}(X = x_k)}{F_X(x_{j_2}) - F_X(x_{j_1-1})}
\end{aligned}$$

which is the probability mass function of X truncated to $\{x_{j_1}, \dots, x_{j_2}\}$. Notice that we may replace $F_X(x_{j_2})$ with 1 throughout to obtain the pmf of X truncated to $\{x_{j_1}, x_{j_1+1}, \dots\}$. Thus by choosing F_X to be the distribution of a Poisson distribution with mean λ, we may sample from the Poisson distribution truncated to be greater than or equal to 1 by sampling $u \sim U\big(F_X(0),\, 1\big) \overset{d}{=} U\big(e^{-\lambda}, 1\big)$ and then calculating $G_X(u)$.

3 Extending the algorithm to continuous predictors

Suppose for ease of exposition that the model has a single continuous variable - the extension to multiple continuous variables is straightforward. Let α_{ijk} denote the value for the continuous variable for the $(i, j, k)^{\text{th}}$ day, and let θ be the exponentiated value of the corresponding coefficient. Then we can express the model as

$$\mathbb{P}\Big(Y_{ij} = 1 \;\Big|\; \boldsymbol{\beta}, \theta, \boldsymbol{\xi}, \text{data}\Big) = 1 - \prod_{k=1}^{K} (1 - \lambda_{ijk})^{X_{ijk}}$$

where

$$\lambda_{ijk} = 1 - \exp\Big\{-\xi_i \exp\Big(\boldsymbol{u}'_{ijk}\boldsymbol{\beta} + \alpha_{ijk}\log(\theta)\Big)\Big\}$$

When a predictor variable is not categorical, then we do not have a closed-form expression for the full conditional posterior distribution of the corresonding coefficient. To express the model in a form that is more amenable to sampling via the Metropolis-Hastings algorithm, we cast the prior distribution of θ as a mixture distribution by defining

$$\begin{aligned}
&\theta \,|\, (M = 1) = 1 \\
&\theta \,|\, (M = 0) \sim \mathcal{G}_{\mathcal{A}}(a, b) \\
&M \sim \text{Bern}\,(p)
\end{aligned}$$

and we assume that $(\theta, M) \perp \boldsymbol{\gamma}$. Then we observe that

$$\begin{aligned}
&\pi\Big(M \;\Big|\; \boldsymbol{Y}, \boldsymbol{W}, \boldsymbol{\gamma}, \theta, \boldsymbol{\xi}, \phi, \text{data}\Big) \\
&\quad \propto \pi\Big(\boldsymbol{Y} \;\Big|\; \boldsymbol{W}, M, \boldsymbol{\gamma}, \theta, \boldsymbol{\xi}, \phi, \text{data}\Big)\, \pi\Big(\boldsymbol{W} \;\Big|\; M, \boldsymbol{\gamma}, \theta, \boldsymbol{\xi}, \boldsymbol{\phi}, \text{data}\Big) \\
&\qquad\quad \times\; \pi\Big(\boldsymbol{\gamma} \;\Big|\; M, \theta, \boldsymbol{\xi}, \phi, \text{data}\Big)\, \pi\Big(\theta \;\Big|\; M, \boldsymbol{\xi}, \phi, \text{data}\Big) \pi\Big(M \;\Big|\; \boldsymbol{\xi}, \phi, \text{data}\Big)\, \pi\Big(\boldsymbol{\xi}, \phi \;\Big|\; \text{data}\Big) \\
&\quad = \pi\big(\boldsymbol{Y} \,|\, \boldsymbol{W}\big)\, \pi\Big(\boldsymbol{W} \;\Big|\; \boldsymbol{\gamma}, \theta, \boldsymbol{\xi}, \text{data}\Big)\, \pi\big(\boldsymbol{\gamma}\big)\, \pi\big(\theta \,|\, M\big)\, \pi\big(M\big)\, \pi\big(\boldsymbol{\xi}, \phi\big) \\
&\quad \propto \pi\,(M \,|\, \theta) \\
&\quad = I(\theta = 1)\, M + I\,(\theta \neq 1)\,(1 - M)
\end{aligned}$$

From this we see that $M = 0$ and $M = 1$ are both absorbing states and we need to proceed with a different tack. We adapt the data augmentation approach proposed by Carlin and Chib (1995) to suit our needs. Define

$$\begin{aligned}
&\theta_1 \,|\, (M = 1) = 1 \\
&\theta_0 \,|\, (M = 0) \sim \mathcal{G}_{\mathcal{A}}(a, b)
\end{aligned}$$

and let

$$\mathbb{P}\left(Y_{ij}=1 \;\middle|\; M,\boldsymbol{\beta},\theta_0,\theta_1,\boldsymbol{\xi},\text{data}\right)=1-\prod_{k=1}^{K}(1-\lambda_{ijk})^{X_{ijk}}$$

where

$$\lambda_{ijk}=\begin{cases} 1-\exp\left\{-\xi_i\exp\left(\boldsymbol{u}_{ijk}'\boldsymbol{\beta}+\alpha_{ijk}\log(\theta_0)\right)\right\}, & M=0 \\ 1-\exp\left\{-\xi_i\exp\left(\boldsymbol{u}_{ijk}'\boldsymbol{\beta}+\alpha_{ijk}\log(\theta_1)\right)\right\}, & M=1 \end{cases}$$

Of course $M=1$ corresponds to α_{ijk} being dropped from the model since $\log(\theta_1)=0$. We also see that Y is independent of $\theta_{k\neq j}$ given that $M=j,\ j=0,1$. Usually here we would assume that $(\theta_0\,|\,M)\perp(\theta_1\,|\,M)$, but in this case it is automatic. In order to complete the prior specification it remains to specify the linking distributions $\theta_1\,|\,(M=0)$ and $\theta_0\,|\,(M=1)$, but we will defer this for a moment. Instead, we observe that

$$\begin{aligned}
&\pi\left(\boldsymbol{Y},\boldsymbol{W},M=j,\boldsymbol{\gamma},\theta_0,\theta_1,\boldsymbol{\xi},\phi \;\middle|\; \text{data}\right)\\
&\quad=\pi\left(\boldsymbol{Y}\;\middle|\;\boldsymbol{W},M=j,\theta_0,\theta_1,\boldsymbol{\xi},\phi,\text{data}\right)\\
&\qquad\times\ \pi\left(\boldsymbol{W}\;\middle|\;M=j,\theta_0,\theta_1,\boldsymbol{\xi},\phi,\text{data}\right)\\
&\qquad\times\ \pi\left(\theta_0,\theta_1\;\middle|\;M=j,\boldsymbol{\xi},\phi,\text{data}\right)\\
&\qquad\times\ \mathbb{P}\left(M=j\;\middle|\;\boldsymbol{\xi},\phi,\text{data}\right)\\
&\qquad\times\ \pi\left(\boldsymbol{\xi},\phi\;\middle|\;\text{data}\right)\\
&\quad=\pi\left(\boldsymbol{Y}\,|\,\boldsymbol{W}\right)\pi\left(\boldsymbol{W}\;\middle|\;M=j,\boldsymbol{\gamma},\theta_j,\boldsymbol{\xi},\text{data}\right)\left[\prod_{k=0}^{1}\pi\left(\theta_k\,|\,M=j\right)\right]\mathbb{P}\left(M=j\right)\pi\left(\boldsymbol{\xi},\phi\right)
\end{aligned}$$

Now we consider the auxiliary variables model. We have

$$\begin{aligned}
&\pi\left(\theta_j\;\middle|\;\boldsymbol{Y},\boldsymbol{W},M=j,\boldsymbol{\gamma},\theta_{k\neq j},\boldsymbol{\xi},\phi,\text{data}\right)\\
&\quad=\frac{\pi\left(\boldsymbol{Y},\boldsymbol{W},M=j,\boldsymbol{\gamma},\theta_0,\theta_1,\boldsymbol{\xi},\phi\;\middle|\;\text{data}\right)}{\int\pi\left(\boldsymbol{Y},\boldsymbol{W},M=j,\boldsymbol{\gamma},\theta_0,\theta_1,\boldsymbol{\xi},\phi\;\middle|\;\text{data}\right)d\theta_j}\\
&\quad=\frac{\pi\left(\boldsymbol{Y}\,|\,\boldsymbol{W}\right)\pi\left(\boldsymbol{W}\;\middle|\;M=j,\boldsymbol{\gamma},\theta_j,\boldsymbol{\xi},\text{data}\right)\left[\prod_{k=0}^{1}\pi\left(\theta_k\,|\,M=j\right)\right]\mathbb{P}\left(M=j\right)\pi\left(\boldsymbol{\xi},\phi\right)}{\int\pi\left(\boldsymbol{Y}\,|\,\boldsymbol{W}\right)\pi\left(\boldsymbol{W}\;\middle|\;M=j,\boldsymbol{\gamma},\theta_j,\boldsymbol{\xi},\text{data}\right)\left[\prod_{k=0}^{1}\pi\left(\theta_k\,|\,M=j\right)\right]\mathbb{P}\left(M=j\right)\pi\left(\boldsymbol{\xi},\phi\right)d\theta_j}\\
&\quad=\frac{\pi\left(\boldsymbol{W}\;\middle|\;M=j,\boldsymbol{\gamma},\theta_j,\boldsymbol{\xi},\text{data}\right)\pi\left(\theta_j\,|\,M=j\right)}{\int\pi\left(\boldsymbol{W}\;\middle|\;M=j,\boldsymbol{\gamma},\theta_j,\boldsymbol{\xi},\text{data}\right)\pi\left(\theta_j\,|\,M=j\right)d\theta_j}\\
&\quad=\pi\left(\theta_j\;\middle|\;\boldsymbol{W},M=j,\boldsymbol{\gamma},\theta_j,\boldsymbol{\xi},\text{data}\right)
\end{aligned}$$

so we see that the full conditional distribution for $\theta_j,\ j = 0, 1$ remains unchanged under the auxiliary variables model. Next, for $k \neq j$,

$$\begin{aligned}
&\pi\Big(\theta_k \,\Big|\, \boldsymbol{Y}, \boldsymbol{W}, M = j, \boldsymbol{\gamma}, \theta_{k\neq j}, \xi, \phi, \text{data}\Big) \\
&\quad = \frac{\pi\Big(\boldsymbol{Y}, \boldsymbol{W}, M = j, \boldsymbol{\gamma}, \theta_0, \theta_1, \xi, \phi \,\Big|\, \text{data}\Big)}{\displaystyle\int \pi\Big(\boldsymbol{Y}, \boldsymbol{W}, M = j, \boldsymbol{\gamma}, \theta_0, \theta_1, \xi, \phi \,\Big|\, \text{data}\Big)\, d\theta_k} \\
&\quad = \frac{\pi\big(\boldsymbol{Y} \,|\, \boldsymbol{W}\big)\, \pi\Big(\boldsymbol{W} \,\Big|\, M = j, \boldsymbol{\gamma}, \theta_j, \xi, \text{data}\Big) \Big[\prod_{i=0}^{1} \pi\big(\theta_i \,|\, M = j\big)\Big]\, \mathbb{P}\big(M = j\big)\, \pi\big(\xi, \phi\big)}{\displaystyle\int \pi\big(\boldsymbol{Y} \,|\, \boldsymbol{W}\big)\, \pi\Big(\boldsymbol{W} \,\Big|\, M = j, \boldsymbol{\gamma}, \theta_j, \xi, \text{data}\Big) \Big[\prod_{i=0}^{1} \pi\big(\theta_i \,|\, M = j\big)\Big]\, \mathbb{P}\big(M = j\big)\, \pi\big(\xi, \phi\big)\, d\theta_k} \\
&\quad = \frac{\pi\big(\theta_k \,|\, M = j\big)}{\displaystyle\int \pi\big(\theta_k \,|\, M = j\big)\, d\theta_k} \\
&\quad = \pi\big(\theta_k \,|\, M = j\big)
\end{aligned}$$

Thus the full conditional distribution for $\theta_{k\neq j}$ when $M = j$ is just the linking density. It is easy to verify that the full conditional distributions of $\boldsymbol{\gamma}, \xi$, and ϕ remain unchanged under the auxiliary variables model. Next we observe that

$$\begin{aligned}
&\mathbb{P}\Big(M = 1 \,\Big|\, \boldsymbol{Y}, \boldsymbol{W}, \boldsymbol{\gamma}, \theta_0, \theta_1, \xi, \phi, \text{data}\Big) \\
&\quad = \frac{\pi\Big(\boldsymbol{Y}, \boldsymbol{W}, M = 1, \boldsymbol{\gamma}, \theta_0, \theta_1, \xi, \phi \,\Big|\, \text{data}\Big)}{\sum_{j=0}^{1} \pi\Big(\boldsymbol{Y}, \boldsymbol{W}, M = j, \boldsymbol{\gamma}, \theta_0, \theta_1, \xi, \phi \,\Big|\, \text{data}\Big)} \\
&\quad = \frac{\pi\Big(\boldsymbol{W} \,\Big|\, M = 1, \boldsymbol{\gamma}, \theta_1, \xi, \text{data}\Big) \Big[\prod_{k=0}^{1} \pi\big(\theta_k \,|\, M = 1\big)\Big]\, \mathbb{P}\big(M = 1\big)}{\sum_{j=0}^{1} \pi\Big(\boldsymbol{W} \,\Big|\, M = j, \boldsymbol{\gamma}, \theta_j, \xi, \text{data}\Big) \Big[\prod_{k=0}^{1} \pi\big(\theta_k \,|\, M = j\big)\Big]\, \mathbb{P}\big(M = j\big)}
\end{aligned}$$

In light this result, we now choose the prior distributions for $\theta_1 \,|\, (M = 0)$ and $\theta_0 \,|\, (M = 1)$. Clearly we should specify $\theta_1 \,|\, (M = 0) = 1$. Furthermore, out of convenience, we propose specifying $\theta_0 \,|\, (M = 1) \overset{d}{=} \theta_0 \,|\, (M = 0)$. Under this specification, we obtain that

$$\mathbb{P}\Big(M = 1 \,\Big|\, \boldsymbol{Y}, \boldsymbol{W}, \boldsymbol{\gamma}, \theta_0, \theta_1, \xi, \phi, \text{data}\Big) = \frac{\pi\Big(\boldsymbol{W} \,\Big|\, M = 1, \boldsymbol{\gamma}, \theta_1, \xi, \text{data}\Big)\, \mathbb{P}\big(M = 1\big)}{\sum_{j=0}^{1} \pi\Big(\boldsymbol{W} \,\Big|\, M = j, \boldsymbol{\gamma}, \theta_j, \xi, \text{data}\Big)\, \mathbb{P}\big(M = j\big)}$$

In conclusion, we may incorporate a continuous covariate into the MCMC sampler by using the data augmentation approach detailed above. When $M = 0$ then θ_0 is updated via a Metropolis step.

4 Posterior inference