

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 , $i = 1, \dots, n$

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

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

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

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

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

- \mathbf{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 $\mathbf{U}_{ij} = (\mathbf{u}'_{ijk}, \dots, \mathbf{u}'_{ijk})'$, Dunson and Stanford propose the model:

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

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 occurred, 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 \mathbf{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

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

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.

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

since

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

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 \mathbf{u} is given by

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

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

$$\begin{aligned} \mathcal{G}_{\mathcal{A}_h}(\cdot) & \quad \text{density function of a gamma distribution truncated to the region } \mathcal{A}_h \subset (0, \infty) \\ \gamma_h & \quad \exp(\beta_h) \end{aligned}$$

Then the Dunson and Stanford model chooses priors of the form

$$\begin{aligned} \pi(\boldsymbol{\gamma}) &= \prod_{h=1}^q \left\{ p_h I(\gamma_h = 1) + (1 - p_h) I(\gamma_h \neq 1) \mathcal{G}_{\mathcal{A}_h}(\gamma_h; a_h, b_h) \right\} \\ \pi(\phi) &= \mathcal{G}(\phi; c_1, c_2) \end{aligned}$$

where

$$\begin{aligned} p_h & \quad \text{prior probability that } \gamma_h = 1, \text{ a hyperparameter} \\ a_h, b_h & \quad \text{shape and rate hyperparameters for gamma distribution of } \gamma_h \\ c_1, c_2 & \quad \text{shape and rate hyperparameters for gamma distribution of } \phi \end{aligned}$$

Values of $\gamma_h = 1$ correspond to $\beta_h = 0$ and the h^{th} predictor in \mathbf{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 $1, \dots, t$. Let $\mathbf{s}_{ijk} = (s_{ijk,2}, \dots, s_{ijk,t})$ be a vector of length $(t - 1)$ for each day i, j, k where

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

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

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

From this form we can see that when $\alpha_m \geq 0$, $m = 2, \dots, t$ 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 (corresponding to $\beta_h \geq 0$ for each of the corresponding 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

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

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}, \dots, W_{ijK} are identically 0. It follows that

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

which is the model in (1).

2.2 The full likelihood

Let Y be a random variable representing all of the potential pregnancy indicators Y_{ij} , let 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

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

2.3 The full conditional distributions

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

$$\mathbb{P}(W_{ij} = \mathbf{m} \mid Y_{ij} = 0, \boldsymbol{\beta}, \phi, \xi, \text{data}) = \begin{cases} 1, & \mathbf{m} = \mathbf{0} \\ 0, & \text{else} \end{cases}$$

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

$$\begin{aligned}
& \mathbb{P}(W_{ij} = \mathbf{m} \mid Y_{ij} = 1, \boldsymbol{\beta}, \phi, \xi, \text{data}) \\
&= \sum_{s=0}^{\infty} \mathbb{P}(W_{ij} = \mathbf{m}, \sum_k W_{ijk} = s \mid Y_{ij} = 1, \boldsymbol{\beta}, \phi, \xi, \text{data}) \\
&= \mathbb{P}(W_{ij} = \mathbf{m}, \sum_k W_{ijk} = \sum_k m_k \mid Y_{ij} = 1, \boldsymbol{\beta}, \phi, \xi, \text{data}) \\
&= \mathbb{P}(W_{ij} = \mathbf{m} \mid \sum_k W_{ijk} = \sum_k m_k, Y_{ij} = 1, \boldsymbol{\beta}, \phi, \xi, \text{data}) \\
&\quad \times \mathbb{P}(\sum_k W_{ijk} = \sum_k m_k \mid Y_{ij} = 1, \boldsymbol{\beta}, \phi, \xi, \text{data})
\end{aligned}$$

Furthermore,

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

and

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

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

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

Then

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

$$\begin{aligned}
&= \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_{i,j,k: X_{ijk}=1} \xi_i \prod_{\ell=1}^q \gamma_\ell^{u_{ijk\ell}} \right\} \\
&\quad \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_{i,j,k: X_{ijk}=1} \xi_i \prod_{\ell \neq h} \gamma_\ell^{u_{ijk\ell}} \right\} \\
&\quad + (1-p_h) I(\gamma_h \neq 1) \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\} \mathcal{G}_{\mathcal{A}_h}(\gamma_h; a_h, b_h) \\
&= p_h I(\gamma_h = 1) \exp \left\{ - \sum_{i,j,k: X_{ijk}=1} \xi_i \prod_{\ell \neq h} \gamma_\ell^{u_{ijk\ell}} \right\} \\
&\quad + (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} \\
&\quad \times \exp \left\{ -\gamma_h \left[b_h + \sum_{i,j,k: X_{ijk}=1} \xi_i \prod_{\ell \neq h} \gamma_\ell^{u_{ijk\ell}} \right] \right\} \\
&= p_h I(\gamma_h = 1) \exp \{ -(\tilde{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^{\tilde{a}_h - 1} \exp \{ -\tilde{b}_h \gamma_h \} \\
&= p_h I(\gamma_h = 1) \exp \{ -(\tilde{b}_h - b_h) \} + (1-p_h) I(\gamma_h \neq 1) \\
&\quad \times \frac{C(a_h, b_h) \int_{\mathcal{A}_h} \mathcal{G}(\gamma; \tilde{a}_h, \tilde{b}_h) d\gamma}{C(\tilde{a}_h, \tilde{b}_h) \int_{\mathcal{A}_h} \mathcal{G}(\gamma; a_h, b_h) d\gamma} \frac{C(\tilde{a}_h, \tilde{b}_h)}{\int_{\mathcal{A}_h} \mathcal{G}(\gamma; \tilde{a}_h, \tilde{b}_h) d\gamma} \gamma_h^{\tilde{a}_h - 1} \exp \{ -\tilde{b}_h \gamma_h \} \\
&= p_h I(\gamma_h = 1) \exp \{ -(\tilde{b}_h - b_h) \} + (1-p_h) I(\gamma_h \neq 1) \\
&\quad \times \frac{C(a_h, b_h) \int_{\mathcal{A}_h} \mathcal{G}(\gamma; \tilde{a}_h, \tilde{b}_h) d\gamma}{C(\tilde{a}_h, \tilde{b}_h) \int_{\mathcal{A}_h} \mathcal{G}(\gamma; a_h, b_h) d\gamma} \mathcal{G}_{\mathcal{A}_h}(\gamma; \tilde{a}_h, \tilde{b}_h) \\
&= d_1 I(\gamma_h = 1) + d_2 I(\gamma_h \neq 1) \mathcal{G}_{\mathcal{A}_h}(\gamma; \tilde{a}_h, \tilde{b}_h) \\
&\propto \frac{d_1}{d_1 + d_2} I(\gamma_h = 1) + \frac{d_2}{d_1 + d_2} I(\gamma_h \neq 1) \mathcal{G}_{\mathcal{A}_h}(\gamma; \tilde{a}_h, \tilde{b}_h) \\
&= \tilde{p}_h I(\gamma_h = 1) + (1 - \tilde{p}_h) I(\gamma_h \neq 1) \mathcal{G}_{\mathcal{A}_h}(\gamma; \tilde{a}_h, \tilde{b}_h)
\end{aligned}$$

Version 2

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

$$\begin{aligned}
\tilde{a}_h & a_h + \sum_{i,j,k} u_{ijkh} W_{ijk} \\
\tilde{b}_h & b_h + \sum_{\substack{i,j,k: X_{ijk}=1, \\ u_{ijkh}=1}} \xi_i \prod_{\ell \neq h} \gamma_\ell^{u_{ijk\ell}} \\
d_1 & p_h I(\gamma_h = 1) \exp \left\{ - \sum_{i,j,k: X_{ijk}=1} \xi_i \prod_{\ell \neq h} \gamma_\ell^{u_{ijk\ell}} \right\} \\
d_2 & (1 - p_h) \exp \left\{ - \sum_{\substack{i,j,k: X_{ijk}=1, \\ u_{ijkh}=0}} \xi_i \prod_{\ell \neq h} \gamma_\ell^{u_{ijk\ell}} \right\} \frac{C(a_h, b_h) \int_{\mathcal{A}_h} \mathcal{G}(\gamma; \tilde{a}_h, \tilde{b}_h) d\gamma}{C(\tilde{a}_h, \tilde{b}_h) \int_{\mathcal{A}_h} \mathcal{G}(\gamma; a_h, b_h) d\gamma} \\
\tilde{p}_h & \frac{d_1}{d_1 + d_2}
\end{aligned}$$

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

$$\begin{aligned}
& \pi(\gamma_h | \gamma_{(-h)}, \phi, \xi, \mathbf{W}, \text{data}) \\
& \propto \pi(\mathbf{W} | \xi, \gamma, \text{data}) \pi(\gamma_h) \\
& = \left(\prod_{i=1}^n \prod_{j=1}^{n_i} \prod_{k: X_{ijk}=1} \pi(W_{ijk} | \xi_i, \gamma, \text{data}) \right) \pi(\gamma_h) \\
& \propto \left(\prod_{i=1}^n \prod_{j=1}^{n_i} \prod_{k: X_{ijk}=1} \left[\exp(u_{ijkh} \log \gamma_h) \right]^{W_{ijk}} \exp \left\{ - \xi_i \exp \left(\sum_{\ell=1}^q u_{ijk\ell} \log \gamma_\ell \right) \right\} \right) \pi(\gamma_h) \\
& = \left(\prod_{i=1}^n \prod_{j=1}^{n_i} \prod_{k: X_{ijk}=1} \gamma_h^{u_{ijkh} W_{ijk}} \exp \left\{ - \xi_i \prod_{\ell=1}^q \gamma_\ell^{u_{ijk\ell}} \right\} \right) \pi(\gamma_h) \\
& = \gamma_h^{\sum_{i,j,k} u_{ijkh} W_{ijk}} \exp \left\{ - \sum_{i,j,k: X_{ijk}=1} \xi_i \prod_{\ell=1}^q \gamma_\ell^{u_{ijk\ell}} \right\} \pi(\gamma_h) \\
& = \gamma_h^{\sum_{i,j,k} u_{ijkh} W_{ijk}} \exp \left\{ - \sum_{i,j,k: X_{ijk}=1} \xi_i \prod_{\ell=1}^q \gamma_\ell^{u_{ijk\ell}} \right\} \\
& \quad \times \left[p_h I(\gamma_h = 1) + (1 - p_h) I(\gamma_h \neq 1) \mathcal{G}_{\mathcal{A}_h}(\gamma_h; a_h, b_h) \right] \\
& = p_h I(\gamma_h = 1) \exp \left\{ - \sum_{i,j,k: X_{ijk}=1} \xi_i \prod_{\ell \neq h} \gamma_\ell^{u_{ijk\ell}} \right\}
\end{aligned}$$

$$\begin{aligned}
& + (1-p_h)I(\gamma_h \neq 1)\gamma_h^{\sum_{i,j,k} u_{ijkh}W_{ijk}} \exp \left\{ - \sum_{i,j,k: X_{ijk}=1} \xi_i \prod_{\ell=1}^q \gamma_\ell^{u_{ijk\ell}} \right\} \mathcal{G}_{\mathcal{A}_h}(\gamma_h; a_h, b_h) \\
& = p_h I(\gamma_h = 1) \exp \left\{ - \sum_{i,j,k: X_{ijk}=1} \xi_i \prod_{\ell \neq h} \gamma_\ell^{u_{ijk\ell}} \right\} \\
& \quad + (1-p_h)I(\gamma_h \neq 1)\gamma_h^{\sum_{i,j,k} u_{ijkh}W_{ijk}} \\
& \quad \times \exp \left\{ - \sum_{i,j,k: X_{ijk}=1, \atop u_{ijkh}=0} \xi_i \prod_{\ell \neq h} \gamma_\ell^{u_{ijk\ell}} \right\} \exp \left\{ -\gamma_h \sum_{i,j,k: X_{ijk}=1, \atop u_{ijkh}=1} \xi_i \prod_{\ell \neq h} \gamma_\ell^{u_{ijk\ell}} \right\} \mathcal{G}_{\mathcal{A}_h}(\gamma_h; a_h, b_h) \\
& = p_h I(\gamma_h = 1) \exp \left\{ - \sum_{i,j,k: X_{ijk}=1} \xi_i \prod_{\ell \neq h} \gamma_\ell^{u_{ijk\ell}} \right\} \\
& \quad + (1-p_h)I(\gamma_h \neq 1) \exp \left\{ - \sum_{i,j,k: X_{ijk}=1, \atop u_{ijkh}=0} \xi_i \prod_{\ell \neq h} \gamma_\ell^{u_{ijk\ell}} \right\} \frac{C(a_h, b_h)}{\int_{\mathcal{A}_h} \mathcal{G}(\gamma; a_h, b_h) d\gamma} \\
& \quad \times \gamma_h^{\alpha_h + \sum_{i,j,k} u_{ijkh}W_{ijk} - 1} \exp \left\{ -\gamma_h \left[b_h + \sum_{i,j,k: X_{ijk}=1, \atop u_{ijkh}=1} \xi_i \prod_{\ell \neq h} \gamma_\ell^{u_{ijk\ell}} \right] \right\} \\
& = d_1 I(\gamma_h = 1) + d_2 I(\gamma_h \neq 1) \mathcal{G}_{\mathcal{A}_h}(\gamma_h; \tilde{a}_h, \tilde{b}_h) \\
& = \frac{d_1}{d_1 + d_2} I(\gamma_h = 1) + \frac{d_2}{d_1 + d_2} I(\gamma_h \neq 1) \mathcal{G}_{\mathcal{A}_h}(\gamma_h; \tilde{a}_h, \tilde{b}_h) \\
& = \tilde{p}_h I(\gamma_h = 1) + (1 - \tilde{p}_h) I(\gamma_h \neq 1) \mathcal{G}_{\mathcal{A}_h}(\gamma_h; \tilde{a}_h, \tilde{b}_h)
\end{aligned}$$

Step 3.

$$\begin{aligned}
& \pi(\xi_i \mid \boldsymbol{\beta}, \phi, \mathbf{W}, \text{data}) \\
& \propto \pi(\mathbf{W}_i \mid \xi_i, \boldsymbol{\beta}, \text{data}) \pi(\xi_i \mid \phi, \text{data}) \\
& = \left(\prod_{j,k: X_{ijk}=1} \pi(\mathbf{W}_{ijk} \mid \xi_i, \boldsymbol{\beta}, \text{data}) \right) \pi(\xi_i \mid \phi, \text{data}) \\
& \propto \left(\prod_{j,k: X_{ijk}=1} \xi_i^{W_{ijk}} \exp \left\{ -\xi_i \exp(\mathbf{u}'_{ijk} \boldsymbol{\beta}) \right\} \right) \xi_i^{\phi-1} \exp \{ -\phi \xi_i \}
\end{aligned}$$

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

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) $J(\phi^* | \phi^{(s)}) \sim N(\phi^{(s)}, \delta^2)$ truncated so that $\phi^* > 0$

(ii) $J(\phi^* | \phi^{(s)}) \sim \text{Uniform}(\max(0, \phi^{(s)} - \delta), \phi^{(s)} + \delta)$

Now,

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

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

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

2.3.1 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 $V = F_X^{-1}(U(F_X(a), F_X(b)))$

Then

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

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 $u = U(F_X(a), F_X(b))$ and then calculating $F_X^{-1}(u)$.

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

(iii) **Calculating** $\frac{d_1}{d_1 + d_2}$

(iv) **Sampling from acceptance ratio** r

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

3 Posterior inference