Dynamic generalized linear model derivations using Polya-gamma data augmentation

1 Logistic Regression

Consider the standard logistic regression model.

$$y_i \sim \text{binomial}(n_i, \pi_i)$$
$$\pi_i = \frac{\exp(\boldsymbol{x}_i' \boldsymbol{\beta})}{1 + \exp(\boldsymbol{x}_i' \boldsymbol{\beta})}$$

To sample the joint posterior distribution of $(\boldsymbol{\beta})$, we place multivariate normal priors on $\boldsymbol{\beta}$ and implement the Polya-gamma data augmentation strategy described by [Polson et al., 2013].

2 References

[Polson et al., 2013] Polson, N. G., Scott, J. G., and Windle, J. (2013). Bayesian inference for logistic models using plyagamma latent variables. *Journal of the American Statistical Association*, 108(504):13391349.

3 Multi-Scale Occupancy Models

Consider a study designed to determine the presence/absence of some substance of interest (parasites, *e coli.*, fish DNA, etc.). In such a study, we sample multiple sites; from each site, we collect multiple samples; and finally, from each sample, we run multiple tests for the presence/absence of that substance. However, our detection method is not perfect, nor are we guaranteed to capture the substance in each sample from a site, even if it is present. To account for this uncertainty, consider the following multi-scale occupancy model.

3.1 Occurrence at site i

Let Z_i be a binary valued random variable denoting the presence or absence of the substance at site i. That is,

$$Z_i = \begin{cases} 1 & \text{if the substance is present at site } i \\ 0 & \text{if the substance is not present at site } i \end{cases}$$

If we think of the presence or absence of the substance being related to site specific covariates $\boldsymbol{x}_i$, we can model this as:

$$Z_i \sim \text{Bernoulli}\left(\psi_i\right), \quad \psi_i = \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}{1 + \exp(\boldsymbol{x}_i'\boldsymbol{\beta})}$$

3.2 Occurrence in sample j at site i

Even if the substance is present at site i, it may not be present in every sample we collect from that site. Let A_{ij} be a binary valued random variable denoting the presence or absence of the substance in sample j from site i. That is,

$$A_{ij} = \begin{cases} 1 & \text{if the substance is present in sample } j \text{ at site } i \\ 0 & \text{if the substance is not present in sample } j \text{ at site } i \end{cases}$$

If we think of the presence or absence of the substance being related to sample specific covariates

$\boldsymbol{w}_{ij}$, we can model this as:

$$A_{ij}|z_i \sim \text{Bernoulli}\left(z_i\theta_{ij}\right), \quad \theta_{ij} = \frac{\exp(\boldsymbol{w}'_{ij}\boldsymbol{\alpha})}{1 + \exp(\boldsymbol{w}'_{ij}\boldsymbol{\alpha})}$$

3.3 Detection in replicate k from sample j at site i

Even if the substance is present in sample j, we may not detect it in every replicate we take from that sample. Let Y_{ijk} be a binary valued random variable denoting the detection of the substance in replicate k from sample j at site i. That is,

$$Y_{ijk} = \begin{cases} 1 & \text{if the substance is detected in replicate } k \text{ from sample } j \text{ at site } i \\ 0 & \text{if the substance is not detected in replicate } k \text{ from sample } j \text{ at site } i \end{cases}$$

It is typically reasonable to assume that the probability of detection is constant across the replicates within a sample, and that the replicates themselves are independent of one another (assuming adequate mixing of the sample before replicates are taken). If we think of this probability of detection as being related to sample specific covariates $\boldsymbol{v}_{ij}$ and letting k_{ij} denote the number of replicates taken from sample j at site i, we can model this as:

$$Y_{ij}|a_{ij} \sim \text{Binomial}\left(k_{ij}, a_{ij}p_{ij}\right), \quad p_{ij} = \frac{\exp(\boldsymbol{v}'_{ij}\boldsymbol{\delta})}{1 + \exp(\boldsymbol{v}'_{ij}\boldsymbol{\delta})}$$

4 MCMC Procedures

The joint posterior distribution is the following:

$$p(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{z}, \boldsymbol{A}|\boldsymbol{y}) \propto p(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\delta}) \times \prod_{i=1}^{M} \psi^{z_i}(1-\psi)^{1-z_i} \prod_{i=1}^{M}\prod_{j=1}^{J_i}(z_i\theta_{ij})^{a_{ij}}(1-z_i\theta_{ij})^{1-a_{ij}}\binom{k_{ij}}{y_{ij}}(a_{ij}p_{ij})^{y_{ij}}(1-a_{ij}p_{ij})^{k_{ij}-y_{ij}}$$

To implement MCMC procedures, we need the full conditional distributions. We derive them below.

4.1 Latent variables

Each of these are from Dorazio and Erickson (2017).

4.1.1 Full conditional of Z

$$Z_i|\cdot \sim \begin{cases} \text{Bernoulli}(1) & \text{if } \sum_{j=1}^{J_i} a_{ij} > 0 \\ \text{Bernoulli}\left(\frac{\psi_i \prod_{j=1}^{J_i}(1-\theta_{ij})}{\psi_i \prod_{j=1}^{J_i}(1-\theta_{ij})+1-\psi_i}\right) & \text{if } \sum_{j=1}^{J_i} a_{ij} = 0 \end{cases}$$

4.1.2 Full conditional of A

$$A_{ij}|\cdot \sim \begin{cases} \text{Bernoulli}(1) & \text{if } y_{ij} > 0 \\ \text{Bernoulli}\left(\frac{z_i\theta_{ij}(1-p_{ij})^{k_{ij}}}{z_i\theta_{ij}(1-p_{ij})^{k_{ij}}+1-z_i\theta_{ij}}\right) & \text{if } y_{ij} = 0 \end{cases}$$

4.2 Regression coefficients

To efficiently sample the regression coefficients, we implement a data augmentation strategy based on Polya-gamma random variables.

4.2.1 Full conditional of $\boldsymbol{\beta}$ and ψ

Constant ψ

The full conditional distribution of $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta}|\cdot \propto p(\boldsymbol{\beta}) \prod_{i=1}^{M} \psi_i^{z_i} (1 - \psi_i)^{1-z_i}$$

where $\psi_i = \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}$. If we assume that ψ_i is constant across sites ($\psi_i = \psi$), we can efficiently sample ψ from a beta distribution by placing a semi-conjugate beta prior on it. Let $\psi \sim \text{beta}(a, b)$. Then,

$$\psi|\cdot \propto \psi^{a-1}(1 - \psi)^{b-1} \prod_{i=1}^{M} \psi^{z_i} (1 - \psi)^{1-z_i}$$

$$\propto \psi^{a-1}(1 - \psi)^{b-1} \psi^{\sum_{i=1}^{M} z_i} (1 - \psi)^{M - \sum_{i=1}^{M} z_i}$$

$$\propto \psi^{a-1+\sum_{i=1}^{M} z_i} (1 - \psi)^{b-1+M-\sum_{i=1}^{M} z_i}$$

We recognize this as the kernel of a $\text{beta}\left(a + \sum_{i=1}^{M} z_i, b + M - \sum_{i=1}^{M} z_i\right)$ distribution. Therefore,

$$\psi|\cdot \sim \text{beta}\left(a + \sum_{i=1}^{M} z_i, b + M - \sum_{i=1}^{M} z_i\right)$$

Choosing $a = b = 1$ places a uniform prior distribution on ψ.

Non-constant ψ

Consider the likelihood contribution of z_i, denoted $L_i(\boldsymbol{\beta}|z_i)$. Also, let $\eta_i = \boldsymbol{x}_i'\boldsymbol{\beta}$. From theorem 1

of Polson et al. (2012), we have

$$L_i(\boldsymbol{\beta}|z_i) \propto \psi_i^{z_i}(1-\psi_i)^{1-z_i} = \left(\frac{\exp(\eta_i)}{1+\exp(\eta_i)}\right)^{z_i}\left(1-\frac{\exp(\eta_i)}{1+\exp(\eta_i)}\right)^{1-z_i}$$

$$\propto \frac{(\exp(\eta_i))^{z_i}}{(1+\exp(\eta_i))^1}$$

$$\propto e^{(z_i-\frac{1}{2})\eta_i}\int e^{-\omega_i\eta_i^2/2}p(\omega_i)d\omega_i$$

where ω_i is a Polya-gamma random variable. Therefore, conditional on ω, and letting $\kappa_i = z_i - \frac{1}{2}$, we have

$$L_i(\boldsymbol{\beta}|z_i,\omega_i) \propto e^{\kappa_i\eta_i - \frac{\omega_i}{2}\eta_i^2}$$

$$\propto e^{-\frac{\omega_i}{2}(\eta_i^2 - \frac{2\kappa_i}{\omega_i}\eta_i)}$$

$$\propto e^{-\frac{\omega_i}{2}(\frac{\kappa_i}{\omega_i}-\eta_i)^2}$$

Therefore, letting $z_i^* = \frac{\kappa_i}{\omega_i} = \frac{1}{\omega_i}(z_i - \frac{1}{2})$, we see that $z_i^* \sim N(\eta_i, \frac{1}{\omega_i})$. And now, letting $\boldsymbol{z}^*$ denote the stacked vector of z_i^*'s and $\Omega = \text{diag}(\omega_i)$, we have that

$$\boldsymbol{z}^* \sim \mathcal{N}(\boldsymbol{X}\boldsymbol{\beta}, \Omega^{-1})$$

Consequently, we express the full conditional of $\boldsymbol{\beta}$ as

$$p(\boldsymbol{\beta}|\cdot) \propto p(\boldsymbol{\beta})\prod_{i=1}^{M}\psi_i^{z_i}(1-\psi_i)^{1-z_i}$$

$$\propto p(\boldsymbol{\beta})\exp\left\{-\frac{1}{2}(\boldsymbol{z}^* - \boldsymbol{X}\boldsymbol{\beta})'\Omega(\boldsymbol{z}^* - \boldsymbol{X}\boldsymbol{\beta})\right\}$$

Therefore, placing a MVN(μ_0, Σ_0) prior on the regression coefficients, we get

$$p(\boldsymbol{\beta}|\cdot) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\mu_0)'\Sigma_0^{-1}(\boldsymbol{\beta}-\mu_0)\right\}\exp\left\{-\frac{1}{2}(\boldsymbol{z}^* - \boldsymbol{X}\boldsymbol{\beta})'\Omega(\boldsymbol{z}^* - \boldsymbol{X}\boldsymbol{\beta})\right\}$$

$$\propto \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}'\Sigma_0^{-1}\boldsymbol{\beta} - 2\boldsymbol{\beta}'\Sigma_0^{-1}\mu_0) + -\frac{1}{2}(\boldsymbol{\beta}\boldsymbol{X}'\Omega\boldsymbol{X}\boldsymbol{\beta} - 2\boldsymbol{\beta}'\boldsymbol{X}\Omega\boldsymbol{z}^*)\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta}'(\boldsymbol{X}'\Omega\boldsymbol{X} + \Sigma_0^{-1})\boldsymbol{\beta} - 2\boldsymbol{\beta}'(\boldsymbol{X}'\Omega\boldsymbol{z}^* + \Sigma_0^{-1}\mu_0)\right)\right\}$$

Putting this all together, we get $\boldsymbol{\beta}|\cdot \sim \mathcal{N}\left(m, V^{-1}\right)$ where

$$m = V^{-1}(\boldsymbol{X}'\Omega\boldsymbol{z}^* + \Sigma_0^{-1}\mu_0)$$

$$V^{-1} = (\boldsymbol{X}'\Omega\boldsymbol{X} + \Sigma_0^{-1})^{-1}$$

4.2.2 Full conditional of $\boldsymbol{\alpha}$ and θ

Constant θ across sites

The full conditional distribution of $\boldsymbol{\alpha}$ is

$$\boldsymbol{\alpha}|\cdot \propto p(\boldsymbol{\alpha}) \prod_{i=1,z_i=1}^{M} \prod_{j=1}^{J_i} (z_i\theta_{ij})^{a_{ij}}(1 - z_i\theta_{ij})^{1-a_{ij}}$$

where $\theta_{ij} = \frac{\exp(\boldsymbol{w}_{ij}'\boldsymbol{\alpha})}{1+\exp(\boldsymbol{w}_{ij}'\boldsymbol{\alpha})}$. If we assume that θ_{ij} is constant across sites ($\theta_{ij} = \theta_i$), we can efficiently sample θ_i from a beta distribution by placing a semi-conjugate beta prior on it. Let $\theta_i \sim \text{beta}(c,d)$. Then,

$$\theta_i|z_i = 1,\cdot \propto \theta_i^{c-1}(1 - \theta_i)^{d-1} \prod_{j=1}^{J_i} \theta_i^{a_{ij}}(1 - \theta_i)^{1-a_{ij}}$$

$$\propto \theta_i^{c-1}(1 - \theta_i)^{d-1}\theta_i^{\sum_{j=1}^{J_i} a_{ij}}(1 - \theta_i)^{J_i - \sum_{j=1}^{J_i} a_{ij}}$$

$$\propto \theta_i^{c-1}(1 - \theta_i)^{d-1}\theta_i^{\sum_{j=1}^{J_i} a_{ij}}(1 - \theta_i)^{J_i - \sum_{j=1}^{J_i} a_{ij}}$$

$$\propto \theta_i^{c-1+\sum_{j=1}^{J_i} a_{ij}}(1 - \theta_i)^{d-1+J_i - \sum_{j=1}^{J_i} a_{ij}}$$

To account for the dependence on $z_i = 1$, we rewrite this as:

$$\propto \theta_i^{c-1+z_i\sum_{j=1}^{J_i} a_{ij}}(1 - \theta_i)^{d-1+z_i(J_i - \sum_{j=1}^{J_i} a_{ij})}$$

We recognize this as the kernel of a beta $\left(c + z_i\sum_{j=1}^{J_i} a_{ij}, d + z_i(J_i - \sum_{j=1}^{J_i} a_{ij})\right)$ distribution. Therefore,

$$\theta_i|\cdot \sim \text{beta}\left(c + z_i\sum_{j=1}^{J_i} a_{ij}, d + z_i(J_i - \sum_{j=1}^{J_i} a_{ij})\right)$$

Choosing $c = d = 1$ places a uniform prior distribution on θ_i. Note that this prior also implies a uniform posterior distribution over (0,1) when $z_i = 0$.

Non-constant θ_{ij}

Consider the likelihood contribution of a_{ij}, denoted $L_i(\boldsymbol{\alpha}|a_{ij}, z_i = 1)$ (we condition on $z_i = 1$, because if $z_i = 0$, then $a_{ij} = 0 \ \forall \ j$). Also, let $\nu_{ij} = \boldsymbol{w}'_{ij}\boldsymbol{\alpha}$. From theorem 1 of Polson et al. (2012), we have

$$L_i(\boldsymbol{\alpha}|a_{ij}, z_i = 1) \propto (\theta_{ij})^{a_{ij}}(1 - \theta_{ij})^{1-a_{ij}} = \left(\frac{\exp(\nu_{ij})}{1 + \exp(\nu_{ij})}\right)^{a_{ij}}\left(1 - \frac{\exp(\nu_{ij})}{1 + \exp(\nu_{ij})}\right)^{1-a_{ij}}$$

$$\propto \frac{(\exp(\nu_{ij}))^{a_{ij}}}{(1 + \exp(\nu_{ij}))^1}$$

$$\propto e^{(a_{ij}-\frac{1}{2})\nu_{ij}}\int e^{-\omega_{ij}\nu_{ij}^2/2}p(\omega_{ij})d\omega_{ij}$$

where ω_{ij} is a Polya-gamma random variable. Therefore, conditional on ω_{ij}, and letting $\kappa_{ij} = a_{ij} - \frac{1}{2}$, we have

$$L_i(\boldsymbol{\beta}|a_{ij}, \omega_{ij}, z_i = 1) \propto e^{\kappa_{ij}\nu_{ij}-\frac{\omega_{ij}}{2}\nu_{ij}^2}$$

$$\propto e^{-\frac{\omega_{ij}}{2}(\nu_{ij}^2-\frac{2\kappa_{ij}}{\omega_{ij}}\nu_{ij})}$$

$$\propto e^{-\frac{\omega_{ij}}{2}(\frac{\kappa_{ij}}{\omega_{ij}}-\nu_{ij})^2}$$

Therefore, letting $a_{ij}^* = \frac{\kappa_{ij}}{\omega_{ij}} = \frac{1}{\omega_{ij}}(a_{ij} - \frac{1}{2})$, we see that $a_{ij}^* \sim \mathrm{N}(\nu_{ij}, \frac{1}{\omega_{ij}})$. And now, letting $\boldsymbol{a}^*$ denote the stacked vector of a_{ij}^*'s and $\Omega = \mathrm{diag}(\omega_{ij})$, we have that

$$\boldsymbol{a}^* \sim \mathcal{N}(\boldsymbol{W}\boldsymbol{\alpha}, \Omega^{-1})$$

Recall that we are still conditioning on $z_i = 1$. Therefore, let $\boldsymbol{W}_{red}$, $\boldsymbol{a}^*_{red}$, and Ω^{-1}_{red} denote the reduced design matrix, reduced vector of latent a_{ij}'s, and reduced matrix of ω_{ij}'s including only the samples for which $z_i = 1$. We express the full conditional of $\boldsymbol{\alpha}$ as

$$p(\boldsymbol{\alpha}|\cdot) \propto p(\boldsymbol{\alpha}) \prod_{i=1,z_i=1}^{M} \prod_{j=1}^{J_i}(z_i\theta_{ij})^{a_{ij}}(1 - z_i\theta_{ij})^{1-a_{ij}}$$

$$\propto p(\boldsymbol{\alpha})\exp\left\{-\frac{1}{2}(\boldsymbol{a}^*_{red} - \boldsymbol{W}_{red}\boldsymbol{\alpha})'\Omega(\boldsymbol{a}^*_{red} - \boldsymbol{W}_{red}\boldsymbol{\alpha})\right\}$$

Therefore, placing a MVN(μ_0, Σ_0) prior on the regression coefficients, we get

$$p(\boldsymbol{\beta}|\cdot) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{\alpha}-\mu_0)'\Sigma_0^{-1}(\boldsymbol{\alpha}-\mu_0)\right\}\exp\left\{-\frac{1}{2}(\boldsymbol{a}_{red}^*-\boldsymbol{W}_{red}\boldsymbol{\alpha})'\Omega(\boldsymbol{a}_{red}^*-\boldsymbol{W}_{red}\boldsymbol{\alpha})\right\}$$

$$\propto \exp\left\{-\frac{1}{2}(\boldsymbol{\alpha}'\Sigma_0^{-1}\boldsymbol{\alpha}-2\boldsymbol{\alpha}'\Sigma_0^{-1}\mu_0) + -\frac{1}{2}(\boldsymbol{\alpha}\boldsymbol{W}_{red}'\Omega_{red}\boldsymbol{W}_{red}\boldsymbol{\alpha}-2\boldsymbol{\alpha}'\boldsymbol{W}_{red}\Omega_{red}\boldsymbol{a}_{red}^*)\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{\alpha}'(\boldsymbol{W}_{red}'\Omega_{red}\boldsymbol{W}_{red}+\Sigma_0^{-1})\boldsymbol{\alpha}-2\boldsymbol{\alpha}'(\boldsymbol{W}_{red}'\Omega_{red}\boldsymbol{a}_{red}^*+\Sigma_0^{-1}\mu_0))\right)\right\}$$

Putting this all together, we get $\boldsymbol{\alpha}|\cdot \sim \mathcal{N}\left(m, V^{-1}\right)$ where

$$m = V^{-1}(\boldsymbol{W}_{red}'\Omega_{red}\boldsymbol{a}_{red}^* + \Sigma_0^{-1}\mu_0)$$

$$V^{-1} = (\boldsymbol{W}_{red}'\Omega_{red}\boldsymbol{W}_{red} + \Sigma_0^{-1})^{-1}$$

4.2.3   Full conditional of $\boldsymbol{\delta}$ and p

Constant p

The full conditional distribution of $\boldsymbol{\delta}$ is

$$\boldsymbol{\delta}|\cdot \propto p(\boldsymbol{\delta}) \prod_{i=1}^{M} \prod_{j=1, a_{ij}=1}^{J_i} p_{ij}^{y_{ij}}(1-p_{ij})^{k_{ij}-y_{ij}}$$

where $p_{ij} = \frac{exp(\boldsymbol{v}_{ij}'\boldsymbol{\delta})}{1+exp(\boldsymbol{v}_{ij}'\boldsymbol{\delta})}$. If we assume that p is constant across sites and samples ($p_{ij} = p$), we can efficiently sample p from a beta distribution by placing a semi-conjugate beta(e, f) prior on it. From Dorazio and Erickson (2017), we get the following full conditional distribution.

$$p|\cdot \sim \text{beta}\left(e + \sum_{i=1}^{M}\sum_{j=1}^{J_i} a_{ij}y_{ij}, f + \sum_{i=1}^{M}\sum_{j=1}^{J_i} a_{ij}(k_{ij}-y_{ij})\right)$$

Choosing $e = f = 1$ places a uniform prior on p. Furthermore, this implies a uniform posterior for the samples where $a_{ij} = 0$.

Non-constant p