Dynamic generalized linear model derivations using Polya-gamma data augmentation

1 Logistic regression

Consider the standard logistic regression model.

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

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, Scott and Windle, 2013), which allows for Gibbs draws. The details are provided below.

1.1 Derivations

The full conditional posterior distribution of the regression coefficients is proportional to the following:

$$
\begin{aligned}
p(\boldsymbol{\beta}|\boldsymbol{y}) &\propto p(\boldsymbol{\beta}) \prod_{i=1}^{n} p(y_i|\boldsymbol{\beta}) \\
&\propto p(\boldsymbol{\beta}) \prod_{i=1}^{n} \pi_i^{y_i} (1 - \pi_i)^{n_i - y_i}, \qquad \pi_i = \frac{\exp(\boldsymbol{x}_i' \boldsymbol{\beta})}{1 + \exp(\boldsymbol{x}_i' \boldsymbol{\beta})}
\end{aligned} \tag{1}
$$

This can be rewritten as the following:

$$
\begin{aligned}
p(\boldsymbol{\beta}|\boldsymbol{y}) &\propto p(\boldsymbol{\beta}) \prod_{i=1}^{n} \left(\frac{\exp(\boldsymbol{x}_i' \boldsymbol{\beta})}{1 + \exp(\boldsymbol{x}_i' \boldsymbol{\beta})} \right)^{y_i} \left(1 - \frac{\exp(\boldsymbol{x}_i' \boldsymbol{\beta})}{1 + \exp(\boldsymbol{x}_i' \boldsymbol{\beta})} \right)^{n_i - y_i} \\
&= p(\boldsymbol{\beta}) \prod_{i=1}^{n} \frac{\exp(\boldsymbol{x}_i' \boldsymbol{\beta})^{y_i}}{(1 + \exp(\boldsymbol{x}_i' \boldsymbol{\beta}))^{n_i}}
\end{aligned} \tag{2}
$$

Theorem one of (Polson et al., 2013) states that for $b > 0$,

$$
\frac{(e^{\psi})^a}{(1 + e^{\psi})^b} = 2^{-b} e^{\kappa \psi} \int_0^{\infty} e^{-\omega \psi^2 / 2} p(\omega) d\omega,
$$

for $\kappa = a - b/2$ and $\omega \sim \text{PG}(b, 0)$, where PG denotes the Polya-gamma density. Therefore, revisiting (2), conditioning on the Polya-gamma latents, and letting $\psi_i = \boldsymbol{x}_i'\boldsymbol{\beta}$ we have:

$$
\begin{aligned}
p(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{\omega}) &\propto p(\boldsymbol{\beta}) \prod_{i=1}^{n} \frac{(e^{\psi_i})^{y_i}}{(1+e^{\psi_i})^{n_i}} \\
&= p(\boldsymbol{\beta}) \prod_{i=1}^{n} \exp(\kappa_i \psi_i - \omega_i \psi_i^2/2) \\
&\propto p(\boldsymbol{\beta}) \prod_{i=1}^{n} \exp\left(-\frac{\omega_i}{2}\left(z_i - \psi_i\right)^2\right) \\
&= p(\boldsymbol{\beta}) \exp\left\{-\frac{1}{2}(\boldsymbol{z} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{\Omega}(\boldsymbol{z} - \boldsymbol{X}\boldsymbol{\beta})\right\},
\end{aligned}
\tag{3}
$$

where $\kappa_i = y_i - \frac{n_i}{2}$, $z_i = \frac{\kappa_i}{\omega_i}$, and $\boldsymbol{\Omega} = \text{diag}(\omega_1, ..., \omega_n)$. From (3), $\boldsymbol{z}$ is conditionally Gaussian. That is,

$$
\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Omega}^{-1}) \tag{4}
$$

Therefore, placing a $\mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ prior on $\boldsymbol{\beta}$ results in the following full conditional distribution:

$$
\begin{aligned}
p(\boldsymbol{\beta}|\boldsymbol{z}, \boldsymbol{\Omega}) &\propto p(\boldsymbol{z}|\boldsymbol{\beta}, \boldsymbol{\Omega}) \cdot p(\boldsymbol{\beta}) \\
&\propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{z} - \boldsymbol{X}\boldsymbol{\beta}\right)'\boldsymbol{\Omega}\left(\boldsymbol{z} - \boldsymbol{X}\boldsymbol{\beta}\right)\right\} \exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta} - \boldsymbol{\mu}_0\right)'\boldsymbol{\Sigma}_0^{-1}\left(\boldsymbol{\beta} - \boldsymbol{\mu}_0\right)\right\} \\
&= \exp\left\{-\frac{1}{2}\left(\boldsymbol{z}'\boldsymbol{\Omega}\boldsymbol{z} - 2\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{\Omega}\boldsymbol{z} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{\Omega}\boldsymbol{X}\boldsymbol{\beta}\right)\right\} \exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta}'\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\beta} - 2\boldsymbol{\beta}'\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 + \boldsymbol{\mu}_0'\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(-2\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{\Omega}\boldsymbol{z} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{\Omega}\boldsymbol{X}\boldsymbol{\beta}\right)\right\} \exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta}'\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\beta} - 2\boldsymbol{\beta}'\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0\right)\right\} \\
&= \exp\left\{-\frac{1}{2}\left(-2\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{\Omega}\boldsymbol{z} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{\Omega}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\beta} - 2\boldsymbol{\beta}'\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0\right)\right\} \\
&= \exp\left\{-\frac{1}{2}\left(-2\boldsymbol{\beta}'\left(\boldsymbol{X}'\boldsymbol{\Omega}\boldsymbol{z} + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0\right) + \boldsymbol{\beta}'\left(\boldsymbol{X}'\boldsymbol{\Omega}\boldsymbol{X} + \boldsymbol{\Sigma}_0^{-1}\right)\boldsymbol{\beta}\right)\right\}
\end{aligned}
\tag{5}
$$

And now, a quick note on identifying kernels of a multivariate normal distribution. Suppose $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then

$$
\begin{aligned}
p(\boldsymbol{\beta}) &\propto \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu})\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta} - 2\boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right)\right\}
\end{aligned}
\tag{6}
$$

Therefore, based on (6), (5) implies that

$$\boldsymbol{\beta} | \boldsymbol{z}, \boldsymbol{\Omega} \sim \mathcal{N}(\boldsymbol{m}, \boldsymbol{V}), \tag{7}$$

where $\boldsymbol{V} = \left(\boldsymbol{X}'\boldsymbol{\Omega}\boldsymbol{X} + \boldsymbol{\Sigma}_0^{-1}\right)^{-1}$ and $\boldsymbol{m} = \boldsymbol{V}\left(\boldsymbol{X}'\boldsymbol{\Omega}\boldsymbol{z} + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0\right)$. Finally, we note that $\boldsymbol{\Omega}\boldsymbol{z} = \boldsymbol{\kappa}$.

Finally, (Polson et al., 2013) note that the full conditional distribution of $\boldsymbol{\Omega}$ is also in the Polya-gamma family, and given by the following:

$$\omega_i | \boldsymbol{\beta} \sim \mathrm{PG}(n_i, \psi_i) \tag{8}$$

where $\psi_i = \boldsymbol{x}_i'\boldsymbol{\beta}$. We omit the derivation here.

1.2 Implementation

As a motivating example, consider the famed Donner party dataset. The MLEs for an additive model with sex and age are given below.

```r
data(case2001, package = 'Sleuth3')
summary(glm(Status ~ Sex + Age, family = 'binomial', data = case2001))
```

```
##
## Call:
## glm(formula = Status ~ Sex + Age, family = "binomial", data = case2001)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7445  -1.0441  -0.3029   0.8877   2.0472
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.23041    1.38686   2.329   0.0198 *
## SexMale     -1.59729    0.75547  -2.114   0.0345 *
## Age         -0.07820    0.03728  -2.097   0.0359 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 61.827  on 44  degrees of freedom
## Residual deviance: 51.256  on 42  degrees of freedom
## AIC: 57.256
##
## Number of Fisher Scoring iterations: 4
```

We now fit this model with a Gibbs sampler using the strategy described above.

```r
# response, size, and design
y <- with(Sleuth3::case2001, as.numeric(Status) - 1) # died = 0, survived = 1
size <- rep(1, nrow(Sleuth3::case2001))
```

```r
n <- nrow(Sleuth3::case2001)
X <- model.matrix(~ Sex + Age, data = Sleuth3::case2001)

# precompute kappa
kappa <- y - size/2

# setup sampler and priors
num.mcmc <- 10000
p <- ncol(X)
beta.mcmc <- matrix(0, num.mcmc, p);colnames(beta.mcmc) <- colnames(X)

mu0 <- matrix(0, nrow = p, ncol = 1)
Sigma0.inv <- solve(16*diag(p))
prior.prod <- Sigma0.inv %*% mu0

# initialize
beta <- matrix(rnorm(p), ncol = 1)

# sampler
for(i in 2:num.mcmc){
  # update latent omegas
  eta <- c(X %*% beta)
  omega <- BayesLogit::rpg(n, size, eta)
  Omega <- diag(omega)

  # update beta
  V <- solve(t(X) %*% Omega %*% X + Sigma0.inv)
  m <- V %*% (t(X) %*% kappa + prior.prod)
  beta <- matrix(mvtnorm::rmvnorm(1, m, V), ncol = 1)

  # store
  beta.mcmc[i, ] <- c(beta)
}
est <- cbind(
  colMeans(beta.mcmc),
  apply(beta.mcmc, 2, sd)
); colnames(est) <- c('mean', 'sd')
est
```

```
##                     mean        sd
## (Intercept)  3.19539419 1.3008378
## SexMale     -1.57176687 0.7477836
## Age         -0.07852955 0.0357323
```

These estimates and standard deviations are consistent with those of the MLEs, which reflects our weakly informative priors.

2 Dynamic logistic regression

We now allow the regression coefficients to change over time.

$$
\begin{aligned}
y_t &\sim \text{binomial}(n_t, \pi_t), & \pi_t &= \frac{\exp(\boldsymbol{x}_t'\boldsymbol{\beta}_t)}{1+\exp(\boldsymbol{x}_t'\boldsymbol{\beta}_t)} \\
\boldsymbol{\beta}_t &= \boldsymbol{\beta}_{t-1} + \boldsymbol{V}_t, & \boldsymbol{V}_t &\sim \mathcal{N}(\mathbf{0}, \tau^2 \boldsymbol{I})
\end{aligned}
$$

References

Polson, N.G., Scott, J.G. and Windle, J. (2013) Bayesian inference for logistic models using pólya–gamma latent variables. *Journal of the American Statistical Association*, **108**, 1339–1349.