Bayesian inference for logistic models using Polya-gamma latent variables

Nicholas G. Polson, James G. Scott, Jesse Windle

(Slides based on slides by James G. Scott)

Modeling binary data

	Age	AgeGroup	Race	Completed	InsuranceType	Location	PracticeType
515	21	18to26	Black	0	Military	Odenton	FamilyPractice
423	21	18to26	Black	0	PrivatePayer	Odenton	FamilyPractice
388	17	11to17	White	0	PrivatePayer	Odenton	Pediatric
6	11	11to17	Black	0	Medicaid	Odenton	Pediatric
1104	19	18to26	Black	0	Medicaid	Bayview	Pediatric
1412	19	18to26	Black	0	Medicaid	JohnsHopkins	OBGYN
1354	24	18to26	White	0	PrivatePayer	JohnsHopkins	OBGYN
318	18	18to26	Black	1	Military	Odenton	FamilyPractice
768	24	18to26	White	1	PrivatePayer	Odenton	OBGYN
29	13	11to17	Other/Unknown	0	PrivatePayer	Odenton	FamilyPractice
1173	14	11to17	Hispanic	0	PrivatePayer	Bayview	Pediatric
799	24	18to26	White	0	PrivatePayer	Odenton	OBGYN
633	24	18to26	White	1	PrivatePayer	WhiteMarsh	OBGYN
111	13	11to17	Other/Unknown	0	Medicaid	Odenton	Pediatric
69	15	11to17	Black	0	PrivatePayer	Odenton	FamilyPractice
559	12	11to17	Black	0	Military	Odenton	Pediatric
1289	26	18to26	White	1	HospitalBased	Bayview	OBGYN
1127	18	11to17	White	0	Medicaid	Bayview	Pediatric
1250	18	11to17	Black	0	PrivatePayer	Bayview	Pediatric
1098	15	11to17	White	1	Medicaid	Bayview	Pediatric
378	12	11to17	White	1	Military	Odenton	FamilyPractice
702	26	18to26	White	0	PrivatePayer	WhiteMarsh	OBGYN
....							

Idea: $p(y|x) = f(\beta^\top x)$

Modeling binary data

	Age	AgeGroup	Race	Completed	InsuranceType	Location	PracticeType
515	21	18to26	Black	0	Military	Odenton	FamilyPractice
423	21	18to26	Black	0	PrivatePayer	Odenton	FamilyPractice
388	17	11to17	White	0	PrivatePayer	Odenton	Pediatric
6	11	11to17	Black	0	Medicaid	Odenton	Pediatric
1104	19	18to26	Black	0	Medicaid	Bayview	Pediatric
1412	19	18to26	Black	0	Medicaid	JohnsHopkins	OBGYN
1354	24	18to26	White	0	PrivatePayer	JohnsHopkins	OBGYN
318	18	18to26	Black	1	Military	Odenton	FamilyPractice
768	24	18to26	White	1	PrivatePayer	Odenton	OBGYN
29	13	11to17	Other/Unknown	0	PrivatePayer	Odenton	FamilyPractice
1173	14	11to17	Hispanic	0	PrivatePayer	Bayview	Pediatric
799	24	18to26	White	0	PrivatePayer	Odenton	OBGYN
633	24	18to26	White	1	PrivatePayer	WhiteMarsh	OBGYN
111	13	11to17	Other/Unknown	0	Medicaid	Odenton	Pediatric
69	15	11to17	Black	0	PrivatePayer	Odenton	FamilyPractice
559	12	11to17	Black	0	Military	Odenton	Pediatric
1289	26	18to26	White	1	HospitalBased	Bayview	OBGYN
1127	18	11to17	White	0	Medicaid	Bayview	Pediatric
1250	18	11to17	Black	0	PrivatePayer	Bayview	Pediatric
1098	15	11to17	White	1	Medicaid	Bayview	Pediatric
378	12	11to17	White	1	Military	Odenton	FamilyPractice
702	26	18to26	White	0	PrivatePayer	WhiteMarsh	OBGYN

....

Idea: $p(y|x) = f(\beta^\top x)$

Non-Bayesian: logistic regression

Bayesian: probit regression

Why probit?

- Simple auxiliary variable trick (Albert and Chib)

$$y_i \sim \text{Bern}(w_i)\,, \qquad w_i = \Phi(x_i^T \beta)$$

$$\Longleftrightarrow$$

$$y_i = \mathbb{I}_{z_i > 0}$$

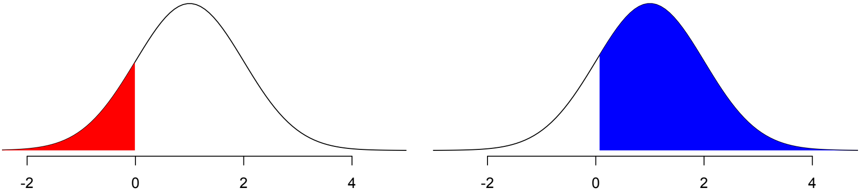
$$z_i = x_i^T \beta + \epsilon_i$$

$$\epsilon_i \sim N(0, 1)$$

$$L(\beta) = \prod_{i=1}^{N} L_i(\beta) = \prod_{i=1}^{N} \{\Phi(x_i^T \beta)\}^{1-y_i} \cdot \{1 - \Phi(x_i^T \beta)\}^{y_i}$$

The key trick: write each term as the integral of a simpler quantity:

$$\begin{aligned} L_i(\beta) &= \frac{1}{\sqrt{2\pi}} \int_{A_i} \exp\{-(z_i - x_i^T \beta)^2 / 2\} \, dz_i \\ A_i &= \begin{cases} (-\infty, 0)\,, & y_i = 0 \\ (0, \infty)\,, & y_i = 1\,. \end{cases} \end{aligned}$$



Auxiliary variables

$$\begin{aligned} p(\beta \mid Y) &\propto p(\beta)L(\beta) \\ &= p(\beta) \prod_{i=1}^{N} \int_{A_i} \phi(z_i; x_i^T \beta, 1)\, dz_i \\ &= \int_{\mathbb{R}^n} \mathbb{I}\left\{ z \in \prod_{i=1}^{n} A_i \right\} p(\beta) \prod_{i=1}^{N} \phi(z_i; x_i^T \beta, 1)\, dz \\ &\propto \int_{R^n} p(\beta, z \mid Y) dz \end{aligned}$$

Auxiliary variables

$$
\begin{aligned}
p(\beta \mid Y) &\propto p(\beta)L(\beta) \\
&= p(\beta) \prod_{i=1}^{N} \int_{A_i} \phi(z_i; x_i^T \beta, 1)\, dz_i \\
&= \int_{\mathbb{R}^n} \mathbb{I}\left\{ z \in \prod_{i=1}^{n} A_i \right\} p(\beta) \prod_{i=1}^{N} \phi(z_i; x_i^T \beta, 1)\, dz \\
&\propto \int_{R^n} p(\beta, z \mid Y) dz
\end{aligned}
$$

Gibbs sampling: alternate between $p(\beta|z, Y)$ (Gaussian) and $p(z|\beta, Y)$ (Truncated Gaussian)

Auxiliary variables

$$\begin{aligned} p(\beta \mid Y) &\propto p(\beta)L(\beta) \\ &= p(\beta)\prod_{i=1}^{N}\int_{A_i}\phi(z_i; x_i^T\beta, 1)\, dz_i \\ &= \int_{\mathbb{R}^n}\mathbb{I}\left\{z \in \prod_{i=1}^{n} A_i\right\} p(\beta)\prod_{i=1}^{N}\phi(z_i; x_i^T\beta, 1)\, dz \\ &\propto \int_{R^n} p(\beta, z \mid Y) dz \end{aligned}$$

Gibbs sampling: alternate between $p(\beta|z, Y)$ (Gaussian) and $p(z|\beta, Y)$ (Truncated Gaussian)

Similar ideas for models involving fancier likelihoods (binomial, negative binomial etc.)

Auxiliary variable representation for logistic likelihood?

$$p(\beta \mid Y) \propto p(\beta) \cdot \prod_{i=1}^{N} \frac{\{\exp(x_i^T \beta)\}^{y_i}}{1 + \exp(x_i^T \beta)}$$

$$\stackrel{?}{=} \int p(\beta, z \mid Y) dz \; .$$

Polya Gamma distribution

The approach is based on a new distribution that we call the Polya-Gamma class: $X \sim \text{PG}(b, c)$ if

$$
\begin{aligned}
X &\overset{D}{=} \frac{1}{2\pi^2} \sum_{k=1}^{\infty} \frac{g_k}{(k - 1/2)^2 + c^2/(4\pi^2)} \\
g_k &\overset{iid}{\sim} \text{Ga}(b, 1)
\end{aligned}
$$

Polya Gamma distribution

The approach is based on a new distribution that we call the Polya-Gamma class: $X \sim \text{PG}(b, c)$ if

$$
\begin{aligned}
X &\stackrel{D}{=} \frac{1}{2\pi^2} \sum_{k=1}^{\infty} \frac{g_k}{(k - 1/2)^2 + c^2/(4\pi^2)} \\
g_k &\stackrel{iid}{\sim} \text{Ga}(b, 1)
\end{aligned}
$$

Our basic result is that

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

where $\kappa = a - b/2$ and $\omega \sim \text{PG}(b, 0)$.

Polya Gamma distribution

- $PG(1, 0)$ has laplace transform $cosh^{-1}(\sqrt{t/2})$
- $PG(1, 0)$ is infinite sum of exponentials

Polya Gamma distribution

- $PG(1, 0)$ has laplace transform $cosh^{-1}(\sqrt{t/2})$
- $PG(1, 0)$ is infinite sum of exponentials
- $PG(b, 0)$ has laplace transform $cosh^{-b}(\sqrt{t/2})$
- $PG(b, c)$: $p(w|b, c) \propto \exp(-\frac{c^2 w}{2})p(w|b, 0)$
- Laplace transform useful to derive moments
- Can also be represented as an alternating sign sum of inverse gaussian densities (later)

Model and Inference

- y_i: the number of successes
- n_i: the number of trials
- $x_i = (x_{i1}, \ldots, x_{ip})$: the vector of regressors.
- $\kappa = (y_1 - n_1/2, \ldots, y_N - n_N/2)$
- Prior: $\beta \sim \mathrm{N}(b, B)$

Model and Inference

- y_i: the number of successes
- n_i: the number of trials
- $x_i = (x_{i1}, \ldots, x_{ip})$: the vector of regressors.
- $\kappa = (y_1 - n_1/2, \ldots, y_N - n_N/2)$
- Prior: $\beta \sim \mathsf{N}(b, B)$

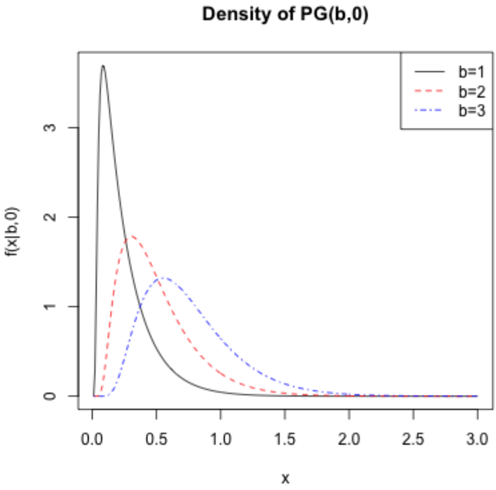
The Polya-Gamma method has only two steps:

$$\begin{aligned}(\omega_i \mid \beta) &\sim \mathsf{PG}(n_i, x_i^T \beta) \\ (\beta \mid y, \omega) &\sim \mathsf{N}(m_\omega, V_\omega)\,,\end{aligned}$$

where

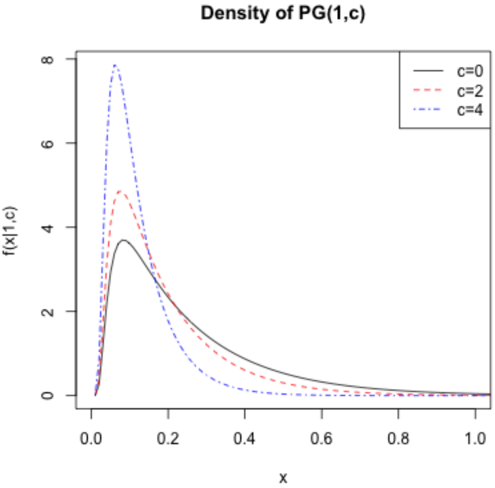
$$\begin{aligned}V_\omega &= (X^T \Omega X + B^{-1})^{-1} \\ m_\omega &= V_\omega (X^T \kappa + B^{-1} b) \\ \Omega &= \mathsf{diag}(\omega_1, \ldots, \omega_N)\,.\end{aligned}$$

PG(b,0)



More Gaussian-ish as *b* increases

PG(1,c)



More like a point mass as *c* increases

How to sample from PG?

It turns out we can simulate PG random variates exactly, without truncating the infinite sum:

$$X \overset{D}{=} \frac{1}{2\pi^2} \sum_{k=1}^{\infty} \frac{g_k}{(k - 1/2)^2 + c^2/(4\pi^2)} \,.$$

We do this using a simple, efficient rejection sampler.

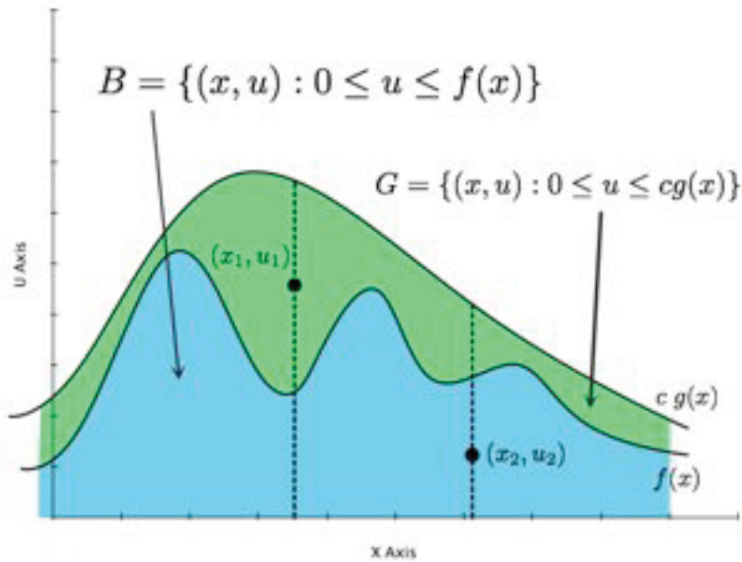
The proposal: exponential, uniform, and normal draws.

Checking for acceptance: roughly like one IG density evaluation.

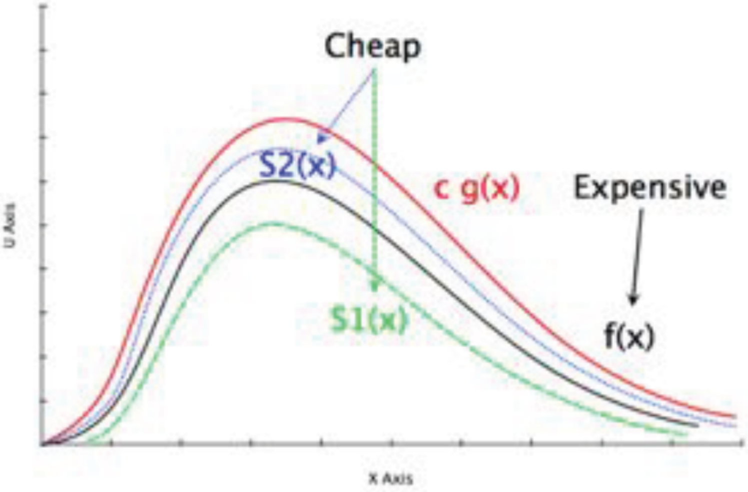
Acceptance probability: in practice, usually better than 0.9998 …

and uniformly bounded below at 0.9992.

Rejection sampler



Squeeze principle



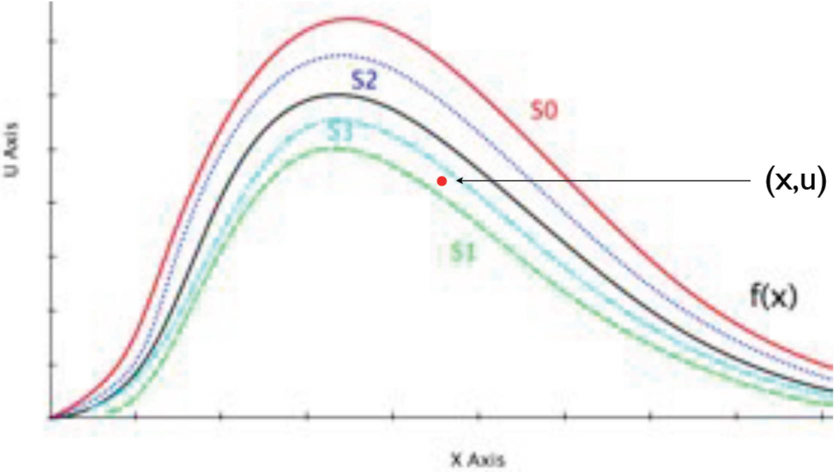
Squeeze principle

Given bounds $S_1(x) \leq f(x)$ and $S_2(x) > f(x)$:

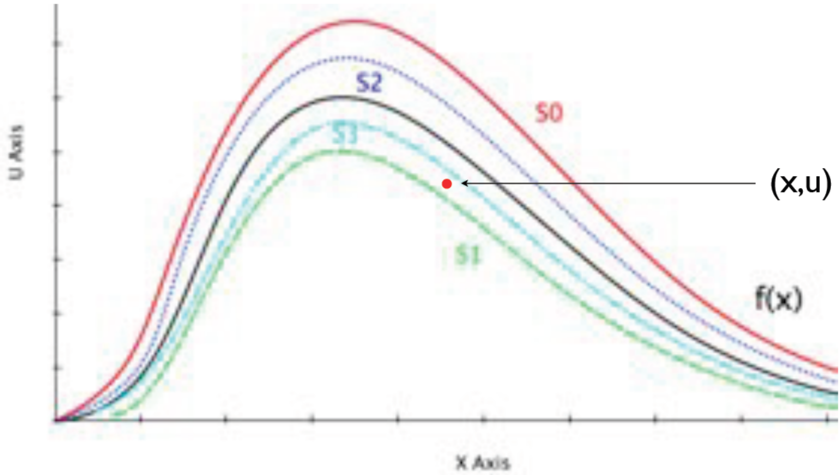
1. $x \sim g$ and $u \sim U(0, cg(x))$ as before.
2. $u \leq S_1(x)$: accept.
3. $u > S_2(x)$: reject.
4. $u \leq f(x)$: accept. If not, reject.

If the bounds are good, we rarely have to evaluate $f(x)$.

Squeeze$^{\infty}$



Squeeze$^{\infty}$



Implemented in BayesLogit R package

Some interesting extensions

- ”Efficient Data Augmentation in Dynamic Models for Binary and Count Data” by Windle et al.
  - Extension of this idea to sequences of binary/count data, where β_t (parameter) dynamics are Gaussian
  - Conditioned on ψ_i, can run exact sampler for β’s (forward filter backward sampling)
- Neuroscience (e.g. spike counts)
- Network data where connections vary with time
- Sports data where probability of win drifts over time

Thanks!