Markov Melding Notes

Adam Howes

3 May 2019

Logistic regression

Consider response $Y \in \{0, 1\}$ modelled as $Y \sim \text{Bern}(q)$ and covariates $x \in \mathbb{R}^p$ with

$$\log \left(\frac{q(x)}{1 - q(x)} \right) = \beta^T x,$$

where $\beta \in \mathbb{R}^p$. Then

$$q(x) = \frac{\exp\left(\beta^T x\right)}{1 + \exp\left(\beta^T x\right)}$$

Observe labelled data $\{(x_1, y_1), \ldots, (x_n, y_n)\}$. The likelihood function is

$$\mathcal{L}(\beta_0, \beta) = \prod_{i=1}^{n} q\left(x_i\right)^{y_i} \left(1 - q(x_i)\right)^{1 - y_i}$$

Place Gaussian prior on β such that

$$\beta \sim \mathcal{N}_p(\mu, \text{diag}(\sigma_1^2, \ldots, \sigma_p^2))$$

Then the posterior is proportional to

$$p(\beta_0, \beta | y_1, \ldots, y_n) \propto \prod_{j=1}^{p} \exp\left(\frac{1}{2\sigma_j^2} (\beta_j - \mu_j)^2 \right) \prod_{i=1}^{n} q\left(x_i\right)^{y_i} \left(1 - q\left(x_i\right)\right)^{1 - y_i}.$$

Taking the logarithm gives

$$\log p(\beta_0, \beta | y_1, \ldots, y_n) \propto \sum_{j=0}^{p} \frac{1}{2\sigma_j^2} (\beta_j - \mu_j)^2 + \sum_{i=1}^{n} y_i \log q\left(x_i\right) + (1 - y_i) \log\left(1 - q\left(x_i\right)\right)$$

The log-likelihood can be rewritten as

$$\begin{aligned}
\sum_{i=1}^{n} y_i \log q\left(x_i\right) + (1 - y_i) \log\left(1 - q\left(x_i\right)\right) &= \sum_{i=1}^{n} y_i \log \left(\frac{q\left(x_i\right)}{1 - q\left(x_i\right)} \right) + \log\left(1 - q\left(x_i\right)\right) \\
&= \sum_{i=1}^{n} y_i \log \left(\frac{q\left(x_i\right)}{1 - q\left(x_i\right)} \right) + \log\left(1 - q\left(x_i\right)\right) \\
&= \sum_{i=1}^{n} y_i \left(\beta^T x\right) - \log\left(1 + \exp(\beta^T x)\right)
\end{aligned}$$

Monte Carlo

Johansen (2018)

Metropolis-Hastings sampler

Target density f and starting with $\mathbf{X}^{(0)} := \left(X_1^{(0)}, \dots, X_p^{(0)}\right)$, iterate for $t = 1, 2, \dots$

1. Draw $\mathbf{X} \sim q\left(\cdot|\mathbf{X}^{(t-1)}\right)$
2. With probability $\min\left\{1, \frac{f(\mathbf{X})\cdot q\left(\mathbf{X}^{(t-1)}|\mathbf{X}\right)}{f\left(\mathbf{X}^{(t-1)}\right)\cdot q\left(\mathbf{X}|\mathbf{X}^{(t-1)}\right)}\right\}$ set $\mathbf{X}^{(t)} = \mathbf{X}$, else set $\mathbf{X}^{(t)} = \mathbf{X}^{(\mathbf{t-1})}$

Note that if the proposal q is symmetric (random walk metropolis-hastings) then the acceptance probability simplifies to $\min\left\{1, \frac{f(\mathbf{X})}{f\left(\mathbf{X}^{(t-1)}\right)}\right\}$

(Random scan) Gibbs sampler

Target density f and starting with $\mathbf{X}^{(0)} := \left(X_1^{(0)}, \dots, X_p^{(0)}\right)$, iterate for $t = 1, 2, \dots$

1. Draw $j \sim \text{Unif}\{1, \dots, p\}$
2. (To-do when I'm at better computer)

Markov Melding

Goudie (2018)

Introduction to Markov melding

- Aims of work:

1. Join submodels p_m into a single joint model
    - Must implicitly handle two different priors for same quantity
    - Must handle non-invertible deterministic transformations
2. Fit the submodels one at a time
    - Minimize burden on practitioners
3. Understanding of reverse operation to joining - splitting

- Models $m = 1, \dots, M$ each with joint density $p_m(\phi, \psi_m, Y_m)$ where:
    - ϕ is the common parameter linking the models
    - ψ_m are model specific unobserved parameters
    - Y_m are model specific observed quantities
- Join together to create $p(\phi, \psi_1, \dots, \psi_M, Y_1, \dots, Y_M)$

Pooling marginal distributions

- Linear pooling
- Logarithmic pooling
- Product of Experts (special case of logarithmic pooling)
- Dictatorial pooling

Inference and computation

Joint posterior, given data $Y_m = y_m$ for $m = 1, \dots, M$, under Melded model is

$$p_{\text{meld}}(\phi, \psi_1, \dots, \psi_M | y_1, \dots, y_M) \propto p_{\text{pool}}(\phi) \prod_{m=1}^{M} \frac{p_m(\phi, \psi_m, y_m)}{p_m(\phi)}$$

Metropolis-Hastings candidate values $(\phi^\star, \psi_1^\star, \dots, \psi_M^\star)$ drawn from a proposal $q(\phi^\star, \psi_1^\star, \dots, \psi_M^\star | \phi, \psi_1, \dots, \psi_M)$ and accepted with probability $\min(1, r)$ where

$$r = \frac{R(\phi^\star, \psi_1^\star, \dots, \psi_M^\star, \phi, \psi_1, \dots, \psi_M)}{R(\phi, \psi_1, \dots, \psi_M, \phi^\star, \psi_1^\star, \dots, \psi_M^\star)}$$

where $R(\phi^\star, \psi_1^\star, \dots, \psi_M^\star, \phi, \psi_1, \dots, \psi_M)$ is the target-to-proposal density ratio

$$R(\phi^\star, \psi_1^\star, \dots, \psi_M^\star, \phi, \psi_1, \dots, \psi_M) = p_{\text{pool}}(\phi^\star) \prod_{m=1}^{M} \frac{p_m(\phi^\star, \psi_m^\star, y_m)}{p_m(\phi^\star)} \times \frac{1}{q(\phi^\star, \psi_1^\star, \dots, \psi_M^\star | \phi, \psi_1, \dots, \psi_M)}$$

Metropolis-within-Gibbs

Sample from the full conditionals using Metropolis-Hastings.

For each of the **latent parameter updates** (ψ_m for $m = 1, \dots, M$) we have

$$R(\phi, \psi_1, \dots, \psi_m^\star, \dots, \psi_M, \phi, \psi_1, \dots, \psi_M) = p_{\text{pool}}(\phi) \prod_{j \neq m} \frac{p_j(\phi, \psi_j, y_j)}{p_j(\phi)} \times \frac{p_m(\phi, \psi_m^\star, y_m)}{p_m(\phi)} \frac{1}{q(\psi_m^\star | \psi_m)}$$

so that

$$r = \frac{p_m(\phi, \psi_m^\star, y_m) \times \frac{1}{q(\psi_m^\star | \psi_m)}}{p_m(\phi, \psi_m, y_m) \times \frac{1}{q(\psi_m | \psi_m^\star)}}$$

and for the **link parameter update**

$$R(\phi, \psi_1, \dots, \psi_m^\star, \dots, \psi_M, \phi, \psi_1, \dots, \psi_M) = p_{\text{pool}}(\phi^\star) \prod_{m=1}^{M} \frac{p_m(\phi^\star, \psi_m, y_m)}{p_m(\phi^\star)} \times \frac{1}{q(\phi^\star | \phi)}$$

Multi-stage Metropolis-within-Gibbs

Factorise the pooled prior (can be done in many ways)

$$p_{\text{pool}}(\phi) = \prod_{m=1}^{M} p_{\text{pool},m}(\phi)$$

Define lth stage posterior as

$$p_{\text{meld},l}(\phi, \psi_1, \dots, \psi_\ell | y_1, \dots, y_\ell) \propto \prod_{m=1}^{\ell} \left(\frac{p_m(\phi, \psi_m, y_m)}{p_m(\phi)} p_{\text{pool},m}(\phi) \right)$$

Basis obtain samples $\left(\phi^{(h,1)}, \psi_1^{(h,1)}\right)$ for $h = 1, \dots, H_1$ from $p_{\text{meld},1}(\phi, \psi_1 | y_1)$ (by MCMC typically)

Inductive construct a Metropolis-within-Gibbs sampler for $(\phi, \psi_1, \dots, \psi_\ell)$ given the data $(y_1, \dots, y_\ell)$

References

- Johansen, A. (2018). *ST407 Monte Carlo Methods*. University of Warwick course notes
- Goudie, R. J. B., A. M. Presanis, D. Lunn, D. De Angelis, and L. Wernisch (2018). *Joining and splitting models with Markov melding. Bayesian Analysis* (to appear)
- Goudie R. J. B. (2019). *Markov melding: A general method for integrating Bayesian models* RSS Emerging Application Section workshop