

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

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

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

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

Place Gaussian prior on β such that

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

Then the posterior is proportional to

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

Taking the logarithm gives

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

The log-likelihood can be rewritten as

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

Monte Carlo

(Johansen 2018)

For the following samplers targeting density f and starting with $\mathbf{X}^{(0)} := (X_1^{(0)}, \dots, X_p^{(0)})$, iterate for $t = 1, 2, \dots$

Metropolis-Hastings sampler

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

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

(Random scan) Gibbs sampler

1. Draw $j \sim \text{Unif}\{1, \dots, p\}$
2. Draw $X_j^{(t)} \sim f_{X_j | X_{-j}}(\cdot | X_1^{(t-1)}, \dots, X_{j-1}^{(t-1)}, X_{j+1}^{(t-1)}, \dots, X_p^{(t-1)})$, and set $X_i^{(t)} := X_i^{(t-1)}$ for all $i \neq j$

(Random scan) Metropolis-within-Gibbs

1. Draw $j \sim \text{Unif}\{1, \dots, p\}$
2. a) Draw $X_j \sim q_j(\cdot | \mathbf{X}^{(t-1)})$ and set $\mathbf{X} = (X_1^{(t-1)}, \dots, X_j, \dots, X_p^{(t-1)})$
b) With probability $\min \left\{ 1, \frac{f(\mathbf{X}) \cdot q(\mathbf{X}^{(t-1)} | \mathbf{X})}{f(\mathbf{X}^{(t-1)}) \cdot q(\mathbf{X} | \mathbf{X}^{(t-1)})} \right\}$ set $\mathbf{X}^{(t)} = \mathbf{X}$, else set $\mathbf{X}^{(t)} = \mathbf{X}^{(t-1)}$

(Divide-and-Conquer with) Sequential Monte Carlo

- Bayesian network is a directed acyclic graph
- Factor graphs (Bishop 2016): “Factor graphs make this decomposition explicit by introducing additional nodes for the factors themselves in addition to the nodes representing the variables.”

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 $(\phi^*, \psi_1^*, \dots, \psi_M^*)$ drawn from a proposal $q(\phi^*, \psi_1^*, \dots, \psi_M^* | \phi, \psi_1, \dots, \psi_M)$ and accepted with probability $\min(1, r)$ where

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

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

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

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

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

so that

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

and for the **link parameter update**

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

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 l th 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 $(\phi^{(h,1)}, \psi_1^{(h,1)})$ 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_ℓ)

References

- Johansen, A. (2018). *ST407 Monte Carlo Methods*. University of Warwick course notes
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- 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