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

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

Taking the logarithm gives

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

The log-likelihood can be rewritten as

$$\sum_{i=1}^{n} y_i \log q\left(x_i\right) + (1 - y_i) \log\left(1 - q\left(x_i\right)\right) = \sum_{i=1}^{n} y_i \log\left(\frac{q\left(x_i\right)}{1 - q\left(x_i\right)}\right) + \log\left(1 - q\left(x_i\right)\right)$$

$$= \sum_{i=1}^{n} y_i \log\left(\frac{q\left(x_i\right)}{1 - q\left(x_i\right)}\right) + \log\left(1 - q\left(x_i\right)\right)$$

$$= \sum_{i=1}^{n} y_i \left(\beta^T x\right) - \log\left(1 + \exp(\beta^T x)\right)$$

1

Monte Carlo

(Johansen 2018)

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

Metropolis-Hastings sampler

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 $\mathbf{X}^{(t)} = \mathbf{X}^{(t-1)}$

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

(Random scan) Gibbs sampler

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

(Random scan) Metropolis-within-Gibbs

1. Draw $j \sim \text{Unif}\{1, \ldots, p\}$
2.
   a) Draw $X_j \sim q_j \left(\cdot | \mathbf{X}^{(t-1)} \right)$ and set $\mathbf{X} = \left(X_1^{(t-1)}, \ldots, X_j, \ldots, X_p^{(t-1)} \right)$
   b) 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 $\mathbf{X}^{(t)} = \mathbf{X}^{(t-1)}$

(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 $m = 1, \ldots, M$ 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 $p(\phi, \psi_1, \ldots, \psi_M, Y_1, \ldots, Y_M)$

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 $m = 1, \ldots, M$, under Melded model is

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

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

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

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

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

Metropolis-within-Gibbs

Sample from the full conditionals using Metropolis-Hastings.

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

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

so that

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

and for the **link parameter update**

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

3

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

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

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

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

Example: Gambia Malaria Data

The `gambia` dataset from the R package geoR contains observations of 2035 Gambian children. The eight variables measured are:

- `x` the x-coordinate of the village (Universal Transverse Mercator)
- `y` the y-coordinate of the village (Universal Transverse Mercator)
- `pos` presence (1) or absence (0) of malaria in a blood sample taken from the child
- `age` age of the child, in days
- `netuse` indicator variable denoting whether (1) or not (0) the child regularly sleeps under a bed-net
- `treated` indicator variable denoting whether (1) or not (0) the bed-net is treated (coded 0 if `netuse` = 0)
- `green` satellite-derived measure of the green-ness of vegetation in the immediate vicinity of the village (arbitrary units)
- `phc` indicator variable denoting the presence (1) or absence (0) of a health center in the village

Full model

Logistic regression of response `pos` on the other variables including an intercept term but excluding the co-ordinates `x` and `y`.

$$\log\left(\frac{q(x)}{1 - q(x)}\right) = \eta$$

$$\eta = \beta_1 + \beta_2 \cdot \text{age} + \beta_3 \cdot \text{netuse} + \beta_4 \cdot \text{treated} + \beta_5 \cdot \text{green} + \beta_6 \cdot \text{phc}$$

Split models

	β_1	β_2	β_3	β_4	β_5	β_6
Submodel 1	✓	✓			✓	✓
Submodel 2	✓		✓	✓		✓

The link parameter is $\phi = (\beta_1, \beta_6)$ and model specific parameters are $\psi_1 = (\beta_1, \beta_6)$ and $\psi_2 = (\beta_3, \beta_4)$.

$$\eta = \beta_1 + \beta_2 \cdot \text{age} + \beta_5 \cdot \text{green} + \beta_6 \cdot \text{phc}$$

$$\eta = \beta_1 + \beta_3 \cdot \text{netuse} + \beta_4 \cdot \text{treated} + \beta_6 \cdot \text{phc}$$

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
- Ribeiro Jr, P. J., & Diggle, P. J. (2001). *geoR: a package for geostatistical analysis.* R news, 1(2), 14-18.