Markov Melding Notes

Adam Howes

1. Logistic regression

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

$$\log\left(\frac{q(x)}{1 - q(x)}\right) = \beta_0 + \beta^T x,$$

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

$$q(x) = \frac{\exp\left(\beta_0 + \beta^T x\right)}{1 + \exp\left(\beta_0 + \beta^T x\right)}$$

```r
# Classify to 1 with probability
q <- function(x, b) {
  exp(b %*% x) / (1 + exp(b %*% x))
}
```

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 priors on β and β_0 such that

$$\beta_0 \sim \mathcal{N}(\mu_0, \sigma_0^2)$$

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

Then the posterior is proportional to

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

Taking the logarithm gives

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

The log-likelihood can be rewritten as

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

```r
# (proportional to) log posterior in the indep normals prior case
logpost <- function(b, X, mu, sigma) {
  logprior <- sum((b - mu)^2 / 2*sigma)
  nu <- apply(X, 1, function(x) b %*% x) # Vector of linear predictors
  loglike <- sum(nu[Y == 1]) + sum(-log(1 + exp(nu)))
  logprior + loglike
}
```

2. Monte Carlo

(Johansen 2018)

For the following samplers targeting density f and starting with $x^{(0)} := \left(x_1^{(0)}, \ldots, x_p^{(0)}\right)$, iterate for $t = 1, 2, \ldots$

2.1. Metropolis-Hastings sampler

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

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

2.2. (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 $i \neq j$

2.3. (Random scan) Metropolis-within-Gibbs

1. Draw $j \sim \text{Unif}\{1, \ldots, p\}$
2. a) Draw $x_j \sim q_j\left(\cdot|x^{(t-1)}\right)$ and set $x = \left(x_1^{(t-1)}, \ldots, x_j, \ldots, x_p^{(t-1)}\right)$

b) With probability $\min\left\{1, \frac{f(x)\cdot q\left(x^{(t-1)}|x\right)}{f\left(x^{(t-1)}\right)\cdot q\left(x|x^{(t-1)}\right)}\right\}$ set $x^{(t)}=x$, else set $x^{(t)}=x^{(t-1)}$

2.4. (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."

3. Markov Melding

(Goudie 2018)

3.1. 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 $p(\phi, \psi_1, \ldots, \psi_M, Y_1, \ldots, Y_M)$

One can define the Markov combination p_{comb} of submodels $p_1, \ldots, p_m$ as

$$
\begin{aligned}
p_{\text{comb}}\left(\phi, \psi_1, \ldots, \psi_M, Y_1, \ldots, Y_M\right) &= p(\phi) \prod_{m=1}^{M} p_m\left(\psi_m, Y_m|\phi\right) \\
&= p(\phi) \prod_{m=1}^{M} \frac{p_m\left(\phi, \psi_m, Y_m\right)}{p(\phi)} \\
&= \frac{\prod_{m=1}^{M} p_m\left(\phi, \psi_m, Y_m\right)}{p(\phi)^{M-1}}
\end{aligned}
$$

3.2. Pooling marginal distributions

- $p_{\text{pool}}(\phi) = g(p_1(\phi), \ldots, p_M(\phi))$
- Types of pooling include
    - Linear pooling $p_{\text{pool}}(\phi) \propto \sum_{m=1}^{M} w_m p_m(\phi)$

- Logarithmic pooling $p_{\text{pool}}(\phi) \propto \prod_{m=1}^{M} p_m(\phi)^{w_m}$
- Product of Experts (special case of logarithmic pooling) $p_{\text{pool}}(\phi) \propto \prod_{m=1}^{M} p_m(\phi)$
- Dictatorial pooling $p_{\text{pool}}(\phi) = p_m(\phi)$ for some $m = 1, \dots, M$

3.3. 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

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

3.3.1. 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

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

3.3.2. 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 $(y_1, \ldots, y_\ell)$

4. Example: Gambia Malaria Data

The `gambia` dataset from the R package geoR contains observations of $n = 2035$ 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

4.1. Submodels

Firstly, the full model $\mathcal{M}$ is the logistic regression of response `pos` on the other variables including an intercept term but excluding the co-ordinates `x` and `y`.

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

$$\eta = \beta_0 + \beta_1 \cdot \texttt{age} + \beta_2 \cdot \texttt{netuse} + \beta_3 \cdot \texttt{treated} + \beta_4 \cdot \texttt{green} + \beta_5 \cdot \texttt{phc}$$

Define submodels $\mathcal{M}_1$ and $\mathcal{M}_2$ with linear predictors

$$\eta_1 = \beta_{0,1} + \beta_1 \cdot \texttt{age} + \beta_4 \cdot \texttt{green} + \beta_5 \cdot \texttt{phc},$$

and

$$\eta_2 = \beta_{0,2} + \beta_2 \cdot \texttt{netuse} + \beta_3 \cdot \texttt{treated} + \beta_5 \cdot \texttt{phc}.$$

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

The link parameter is $\phi = \beta_5$ and model specific parameters are $\psi_1 = (\beta_{0,1}, \beta_1, \beta_6)$ and $\psi_2 = (\beta_{0,2}, \beta_3, \beta_4)$. Both submodels have the same observable random variables $Y_1 = Y_2 = Y$, the response variable `pos`.

Define q_k for $k = 1, 2$ by

$$q_k(x) = \frac{\exp(\eta_k)}{1 + \exp(\eta_k)}$$

Take a normal prior as in Section 1 for $\beta_{1:5}$, and similarly take a normal prior for the intercepts

$$\beta_{0,k} \sim \mathcal{N}(\mu_{0,k}, \sigma^2_{0,k}).$$

Then the submodels have consistent prior marginals in the link parameter and Markov combination can be applied.

$$\begin{aligned}
p_1(\phi, \psi_1, \mathbf{y}_1) &\propto p_1(\phi, \psi_1 | \mathbf{y}_1) \\
&= p_1(\beta_{0,1}, \beta_1, \beta_4, \beta_5 | \mathbf{y}) \\
&\propto \exp\left(\frac{1}{2\sigma^2_{0,1}}(\beta_{0,1} - \mu_{0,1})^2\right) \prod_{j=1,4,5} \exp\left(\frac{1}{2\sigma^2_j}(\beta_j - \mu_j)^2\right) \prod_{i=1}^{2035} q_1(x_i)^{y_i}(1 - q_1(x_i))^{1-y_i}
\end{aligned}$$

and similarly

$$\begin{aligned}
p_2(\phi, \psi_2, \mathbf{y}_2) &\propto p_2(\phi, \psi_2 | \mathbf{y}_2) \\
&= p_2(\beta_{0,2}, \beta_2, \beta_3, \beta_5 | \mathbf{y}) \\
&\propto \exp\left(\frac{1}{2\sigma^2_{0,2}}(\beta_{0,2} - \mu_{0,2})^2\right) \prod_{j=2,3,5} \exp\left(\frac{1}{2\sigma^2_j}(\beta_j - \mu_j)^2\right) \prod_{i=1}^{2035} q_2(x_i)^{y_i}(1 - q_2(x_i))^{1-y_i}
\end{aligned}$$

Therefore the Markov combination in this case is

$$\begin{aligned}
p_{\text{comb}}(\phi, \psi_1, \psi_2, \mathbf{y}_1, \mathbf{y}_2) &= \frac{p_1(\phi, \psi_1, \mathbf{y}_1)\, p_2(\phi, \psi_2, \mathbf{y}_2)}{p(\phi)} \\
&\propto \frac{p_1(\beta_{0,1}, \beta_1, \beta_4, \beta_5 | \mathbf{y}) p_2(\beta_{0,2}, \beta_2, \beta_3, \beta_5 | \mathbf{y})}{p(\beta_5)} \\
&\propto \prod_{k=1}^{2} \exp\left(\frac{1}{2\sigma^2_{0,k}}(\beta_{0,k} - \mu_{0,k})^2\right) \prod_{j=1}^{5} \exp\left(\frac{1}{2\sigma^2_j}(\beta_j - \mu_j)^2\right) \\
&\times \prod_{i=1}^{2035} q_1(x_i)^{y_i}(1 - q_1(x_i))^{1-y_i}\, q_2(x_i)^{y_i}(1 - q_2(x_i))^{1-y_i}
\end{aligned}$$

Informal: almost like Bayesian inference for full model with likelihood squared (not quite)

4.2. Monte Carlo scheme

Continue to use symmetric normal proposals.

For $j \in \{1, 4\}$ such that β_j is one of the latent parameters in ψ_1

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.