# Markov Melding Notes

*Adam Howes*

*08 July, 2019*

## 0. Notation

- Captial $X$ is a random variable, $\mathbf{X}$ is a vector of random variables and $x$ and $\mathbf{x}$ are their realisations

## 1. Background

### 1.1 Monte Carlo

(Johansen, 2018)

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

#### 1.1.1. 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 $\min\left\{1, \frac{f(x)}{f\left(x^{(t-1)}\right)}\right\}$.

#### 1.1.2. (Random scan) Gibbs sampler

1. Draw $j \sim \text{Unif}\{1, \dots, p\}$
2. Draw $x_j^{(t)} \sim f_{x_j | x_{-j}}\left(\cdot | x_1^{(t-1)}, \dots, x_{j-1}^{(t-1)}, x_{j+1}^{(t-1)}, \dots, x_p^{(t-1)}\right)$, and set $x_i^{(t)} := x_i^{(t-1)}$ for all $i \neq j$

#### 1.1.3. (Random scan) Metropolis-within-Gibbs

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

## 1.2. Meta-analysis, evidence synthesis, combining expert opinion etc.

### 1.2.1 (O'Hagan 2006) Chapter 9: Multiple Experts

- Want to obtain a single distribution which encapsulates the beliefs of several experts
- Two approaches
  - Mathematical aggregation: elicit distribution from each expert individually and independently then mathematically combine
  - Behavioural aggregation: Create an interaction beween the group of experts through which a single distribution is elicited from the group as a whole
- In reference to Markov melding, it seems as though Behavioural approaches are not possible
- Each of a group of $n$ experts asked individually for her beliefs about some unknown quantity $\theta$, eliciting distributions $f_i(\theta)$ for $i = 1, \ldots, n$
- Formal Bayesian perspective (DM is a supra-Bayesian): DM begins with his own prior $f(\theta)$ for $\theta$ and has posterior $f(\theta|D)$ after incorporating the experts opinions $D = \{f_1(\theta), \ldots, f_n(\theta)\}$. This is difficult as DM must construct likelihood $f(D|\theta)$
- Simpler and widely used technique is opinion pooling where a consensus distribution $f(\theta)$ is obtained as some function of the individual distributions $\{f_1(\theta), \ldots, f_n(\theta)\}$
- Linear opinion pool $f(\theta) \propto \sum_{i=1}^{n} w_i f_i(\theta)$ where the weights $w_i$ sum to one
  - Could weight all experts equally $w_i = 1/n$ for all $i$
  - Alternatively, give more weight to some expert
  - Coherent marginalisation
  - This approach is not externally Bayesian: after recieving new information, updating the priors then pooling the result is not the same as updating the pooled prior
  - Not consistent with regard to judgements of independence
- Logarithmic opinion pool $f(\theta) \propto \prod_{i=1}^{n} f_i(\theta)^{w_i}$
  - Again can weight opinions as wishes
  - Externally Bayesian and consistent about independence
  - However, unlike linear, no coherent marginalisation (no pooling satisfies both externally Bayesian and coherent marginalisation)
- Product of Experts (special case of logarithmic pooling) $f(\theta) \propto \prod_{i=1}^{n} f_i(\theta)$ *To-do: look at Hinton 2002*
- Unlike supra-Bayesian approach, the result of opinion pooling does not represent the actual beliefs of any individual as so may not behave as one would expect a probability distribution to
- "In general, while the linear opinion pool has been quite widely used in practice, the logarithmic opinion pool has been largely ignored, perhaps because it is perceived to lead to unrealistically strong aggregated beliefs"
- Dictatorial pooling $f(\theta) = f_i(\theta)$ for some $i = 1, \ldots, n$ (not mentioned in book)
- Cooke's method
- For more about opinion pooling, could look at (Clemen, 1999

## 1.3. Graphical models

### 1.3.1 (Smith, 2010) Chapter 7: Bayesian networks

#### 1.3.1.1 Relevance, informativeness and independence

- Client believes that measurement $X$ is irrelevant for predicting $Y$ given the measurement $Z$, written $Y \perp X|Z$ if she believes now that once she learns the value of $Z$ then then measurement of $X$ will provide her with no extra useful information with which to predict the value of $Y$.
- In this case, if she is a Bayesian, then she could write her conditional density $p(y|x, z) = p(y|z)$ so that it did not depend on the value of $x$ for all possible values of $(x, y, z)$. Equivalently the joint mass function can be factorised as $p(x, y, z) = p(y|z)p(x|z)p(z)$
- Two most important and universally applicable rules:
  - Symmetry $Y \perp X|Z \iff X \perp Y|Z$
  - Perfect composition $X \perp (Y, Z)|W \iff X \perp Y|(W, Z) \iff X \perp Z|W$
- *To-do: look at some work of Pearl?*

#### 1.3.1.2. Bayesian networks and DAGs

- Bayesian network is a simple and convenient way of representing a factorisation of a joint pdf of a vector of random variables $\mathbf{X} = (X_1, X_2, \ldots, X_n)$.
- Always the case that $p(\mathbf{x}) = p(x_1)p(x_2|x_1)p(x_3|x_2, x_1) \times \cdots \times p(x_n|x_1, x_2, \ldots, x_{n-1})$
- Often many of the functions $p(x_i|x_1, x_2, \ldots, x_{i-1})$ are explicit functions of components of $X$ whose indices lie in a proper subset $Q_i \subset \{1, 2, \ldots, i-1\}$ so that $p(x_i|x_1, x_2, \ldots, x_{i-1}) = p(x_i|\mathbf{x}_{Q_i})$
- Now $p(\mathbf{x}) = p(x_1) \prod_{i=2}^{n} p(x_i|\mathbf{x}_{Q_i})$
- Let the remainder set $R_i = \{1, 2, \ldots, i-1\} \; Q_i$ then the above bullet point is equivalent to the set of $n-1$ irrelevance statements $X_i \perp \mathbf{X}_{R_i}|\mathbf{X}_{Q_i}$, $2 \leq i \leq n$
- Definition: A directed acyclic graph (DAG) $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with set of vertices $\mathcal{V}$ and set of directed edges $\mathcal{E}$ is a directed graph having no directed cycles
- Definition: A Bayesian network (BN) on the set of measurements $\{X_1, X_2, \ldots, X_n\}$ is a set of the $n-1$ conditional irrelevance statements together with a DAG $\mathcal{G}$. The set of vertices $\mathcal{G} = \{X_1, X_2, \ldots, X_n\}$ and a directed edge from $X_i$ to $X_j$ is in $\mathcal{E}$ if and only if $i \in Q_j$

### 1.3.2 Factor graphs

- 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."

## 1.4. Misc. reading

### 1.4.1. (Spiegelhalter, 1995) Bayesian approaches to random-effects meta-analysis

- Meta analysis, also known as systematic overview, is a statistical procedure in which the results of several independent studies are integrated. Aim is to resolve issues that cannot be concluded from a single study alone
- *Note: Gelman blog post about why fixed and random effects terminology is confusing*
- Fixed-effect analysis: a common effect across studies is estimated
- Random-effects model: a probability model for individual study effects is assumed
- "In contrast to the pooled estimate arising from a fixed-effect model, a random-effects meta-analysis assumes that the true effects in each trial are not necessarily equal, but are random observations drawn from some common population distribution"

Let $r_i^C$ denote the number of infections in the control group in trial $i$, arising from $n_i^C$ cases each assumed to have probability $p_i^C$ of developing an infection. Adopt equivalent notation for the treatment group and assume that $\delta_i$ is the true treatment effect on a log-odds scale such that

$$\delta_i = \text{logit}(p_i^T) - \text{logit}(p_i^C).$$

The "average" infection rate in the $i$th trial is

$$\mu_i = \frac{1}{2}(\text{logit}(p_i^T) + \text{logit}(p_i^C)).$$

Individual trial effects are drawn from some Gaussian population with mean $d$ and variance $\sigma^2$, giving the full model as

$$\begin{aligned}
r_i^C &\sim \text{Binomial}\left(p_i^C, n_i^C\right) \\
r_i^T &\sim \text{Binomial}\left(p_i^T, n_i^T\right) \\
\text{logit}(p_i^C) &= \mu_i - \delta_i/2 \\
\text{logit}(p_i^T) &= \mu_i + \delta_i/2 \\
\delta_i &\sim \text{Normal}(d, \sigma^2)
\end{aligned}$$

- Empirical Bayes' methods estimate $d$ and $\sigma^2$ from the data by moment-matching, then make inferences conditional on these estimates $\hat{d}$, $\hat{\sigma}^2$ but this does not propigate uncertainty about the estimates. Instead full Bayes puts priors on the unknown parameters
- Based on suitable DAG, see Figure 1., the joint distribution is $p(V) = \prod_{v \in V} p(v|\text{pa}(v))$
- Use Gibbs sampling (BUGS) to perform inference

#### 1.4.2. (Fleiss, 1993) The statistical basis of meta-analysis

- $C$ denote the total number of studies to be analyzed, $c = 1, \ldots, C$
  - First approach: take these $C$ studies as the only ones of interest
  - Second approach: take the $C$ studies as a sample from a larger population of studies (fixed vs random set of studies)
-

#### 1.4.3. (Kennedy, 2001) Bayesian calibration of computer models

- Calibration: the activity of adjusting the unknown rate parameters until the outputs of the model fit the observed data
- Bayesian approach with unknown inputs as parameter vector $\theta$
- Uncertainties in computer models
  - Parameter uncertainty: uncertainty about the values of some of the computer code inputs

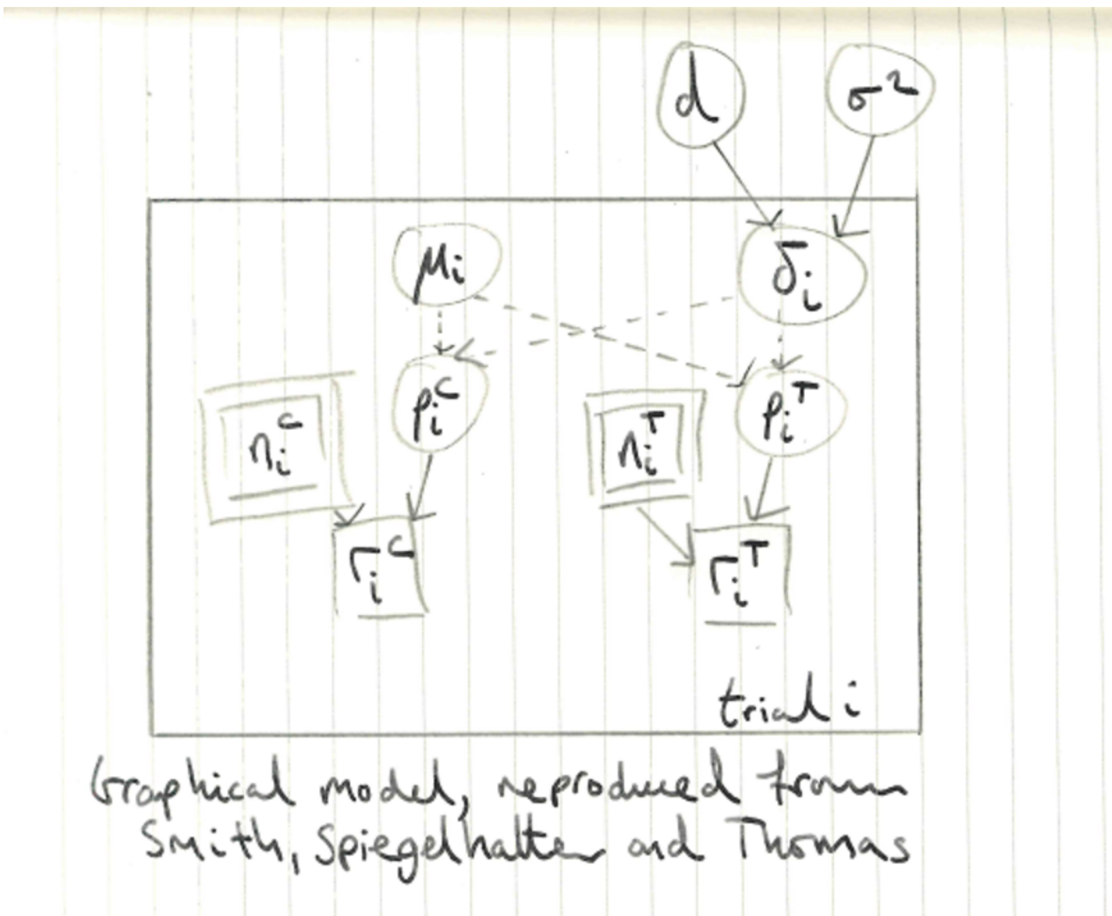


Figure 1: *To-do: reproduce this and similar models in Tikz* Solid line is stochastic dependence, dashed line is logical dependence

- Model inadequacy: no model is perfect. Can't predict the true value of the process (Note that this is not about stochasticity of the process, define model inadequacy to be difference between true mean value of the real world process and the code output at the true values of inputs)
- Residual variability: the real process may not always take the same value for the same repeated inputs. This incompases potential inherent unpredictability, but it may also be that this variation would be eliminated or reduced if only more input conditions were recognised and specified within the model.
- Parametric variability: prediction with random parametric inputs (my interpretation)
- Observation error: uncertainty about the observations used for calibration
- Code uncertainty

- Objective of uncertainty analysis is to study the distribution of the code output that is induced by probability distributions on inputs. Simple Monte Carlo approach: draw configurations of inputs at random from their distribution and run code for each sample input. LHS is better than random sample
- Sensitivity analysis aims to characterize how the code output responds to changes in the inputs
- Generalised likelihood uncertainty estimation: MC sample from prior on unknown inputs, predictions made using this sample weighted by likelihood
- Craig et al. (1992) calibration adopting Bayes linear philosophy (Goldstein)
- Raftery et al. (1995) an attempt to combine prior expert opinion on both the calibration parameters and the model output: Bayesian synthesis. Critizised by Wolpert (1995) and Schweder and Hjort (1996), and in a follow-up paper Poole and Raftery (1998) propose Bayesian melding. Neither method explicitly recognises model inadequacy (!!) - related to link parameter discussions?
- You can use $f(\cdot) \sim \mathcal{GP}(m(\cdot), k(\cdot,\cdot))$ to model unknown functions as random

#### 1.4.4. (Poole, 2000) Inference for deterministic simulation models: the Bayesian melding approach

# 2. (Goudie, 2018) Markov Melding

## 2.1. Introduction to Markov melding

- Motivation: by using all available data typically get:
  - More precise estimates
  - More accurate reflection of true uncertainty
  - Minimise risk of selection-type biases
- Modular approaches
  1. Plug in a point estimate
     - Very easy and fast
     - Underestimates uncertainty
  2. Plug in an approximation to the posterior
     - Quite easy and fast
     - Assumptions made can be unclear
  3. Integrate the models

    - All uncertainty propagated
    - All assumptions explicit
    - Probably tricky to do
- 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:
  - $\phi$ is the common parameter linking the models
  - $\psi_m$ are model specific unobserved parameters
  - $Y_m$ are model specific observed quantities
- Join together to create $p(\phi, \psi_1, \ldots, \psi_M, Y_1, \ldots, Y_M)$

### 2.2. Markov combination

*To-do: add references to Dawid, Massa*

Suppose that the marginals are consistent i.e. $p_m(\phi) = p(\phi)$ for all $m$ then one can define the Markov combination $p_{\text{comb}}$ of submodels $p_1, \ldots, p_m$ as

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

### 2.3. Pooling marginal distributions

If the marginals are inconsistent then instead a pooled density $p_{\text{pool}}(\phi) = g(p_1(\phi), \ldots, p_M(\phi))$ can be used instead. Similar idea to combining expert opinions. This suggests the joint model

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

Goudie calls this Markov melding.

## 2.4. 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

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

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

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

### 2.4.1. Metropolis-within-Gibbs

Sample from the full conditionals using Metropolis-Hastings.

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

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

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

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

### 2.4.2. 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 $l$th stage posterior as

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

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

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

## 3. (Lindsten, 2017) D&C-SMC

*To-do*

## 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 - similar to latitude and longitude)
- `y` the y-coordinate of the village (UTM)
- `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 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

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

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

Observe labelled data $\mathcal{D} = \{(x_1, y_1), \ldots, (x_n, y_n)\}$. 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 $\beta$ and $\beta_0$ such that

$$\beta_0 \sim \mathcal{N}(\mu_0, \sigma_0^2),\ \beta \sim \mathcal{N}_p(\mu, \mathrm{diag}(\sigma_1^2, \ldots, \sigma_p^2))$$

Then the posterior is proportional to

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

Taking the logarithm gives

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

The log-likelihood can be rewritten as

$$\begin{aligned}
\sum_{i=1}^{n} \{y_i \log q\left(x_i\right) + (1 - y_i) \log\left(1 - q\left(x_i\right)\right)\} &= \sum_{i=1}^{n} \{y_i \log\left(\frac{q\left(x_i\right)}{1 - q\left(x_i\right)}\right) + \log\left(1 - q\left(x_i\right)\right)\} \\
&= \sum_{i=1}^{n} \{y_i \log\left(\frac{q\left(x_i\right)}{1 - q\left(x_i\right)}\right) + \log\left(1 - q\left(x_i\right)\right)\} \\
&= \sum_{i=1}^{n} \{y_i \left(\beta_0 + \beta^T x\right) - \log\left(1 + \exp(\beta_0 + \beta^T x)\right)\},
\end{aligned}$$

so that the log-posterior is

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

```
# (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
}
```

## 4.2. Full and 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 | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\beta_5$ |
|---|---|---|---|---|---|---|
| Submodel 1 | ✓ | ✓ | | | ✓ | ✓ |
| Submodel 2 | ✓ | | ✓ | ✓ | | ✓ |

The link parameter is $\phi = \beta_5$ and model specific parameters are $\psi_1 = (\beta_{0,1}, \beta_1, \beta_4)$ and $\psi_2 = (\beta_{0,2}, \beta_2, \beta_3)$. 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.

The joint distribution corresponding to submodel $\mathcal{M}_1$, as a function of the parameters, is proportional to the posterior, which itself is proportional to the prior times the likelihood

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

Similarly for $\mathcal{M}_2$

$$\begin{aligned}
p_2(\phi, \psi_2, \mathbf{y}_2) &\propto p_2(\phi, \psi_2 | \mathbf{y}_2) \\
&= p_2(\beta_{0,2}, \beta_2, \beta_3, \beta_5 | \mathbf{y}) \\
&\propto \underbrace{\exp\left(\frac{1}{2\sigma^2_{0,2}}(\beta_{0,2} - \mu_{0,2})^2\right) \prod_{j=2,3,5} \exp\left(\frac{1}{2\sigma^2_j}(\beta_j - \mu_j)^2\right)}_{\text{Prior on } (\phi,\psi_1)=(\beta_{0,2},\beta_2,\beta_3,\beta_5)} \times \underbrace{\prod_{i=1}^{2035} q_2(x_i)^{y_i} (1 - q_2(x_i))^{1-y_i}}_{\text{Likelihood}}.
\end{aligned}$$

Therefore the Markov combination in this case is

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

Informally $p_{\text{comb}}$ contains the product of the priors on the full set of parameters $(\phi, \psi_1, \psi_2)$ with both likelihoods from $\mathcal{M}_1$ and $\mathcal{M}_2$. So this is almost like Bayesian inference for full model but with a different likelihood. It seems as if the information contained by the data $\mathcal{D}$ is being used more than once.

### 4.3. Monte Carlo schemes

Want to sample from the target (*) using the methods described in (Goudie 2018). Continue to use (symmetric) normal proposals throughout.

#### 4.3.1 Metropolis-within-Gibbs

- Update first latent parameter $\psi_1$ by systematic scan Metropolis-within-Gibbs
- Update second latent parameter $\psi_2$ by systematic scan Metropolis-within-Gibbs
- Update link parameter $\phi$ by Metropolis-within-Gibbs

#### 4.3.2 Multi-stage Metropolis-within-Gibbs

*To-do*

## References


- Johansen, A. (2018). ST407 Monte Carlo Methods. University of Warwick course notes
- O'Hagan, A. et al. (2006). Uncertain Judgements: Eliciting Experts' Probabilities. Chichester: John Wiley & Sons.
- Clemen, R. T., & Winkler, R. L. (1999). Combining probability distributions from experts in risk analysis. Risk analysis, 19(2), 187-203.
- Smith, J. Q. (2010). Bayesian decision analysis: principles and practice. Cambridge University Press.
- Smith, T. C., Spiegelhalter, D. J., & Thomas, A. (1995). Bayesian approaches to random-effects meta-analysis: a comparative study. Statistics in medicine, 14(24), 2685-2699.

- Fleiss, J. L. (1993). Review papers: The statistical basis of meta-analysis. Statistical methods in medical research, 2(2), 121-145.
- Kennedy, M. C., & O'Hagan, A. (2001). Bayesian calibration of computer models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63(3), 425-464.
- Poole, D., & Raftery, A. E. (2000). Inference for deterministic simulation models: the Bayesian melding approach. Journal of the American Statistical Association, 95(452), 1244-1255.
- Lindsten, F., Johansen, A. M., Naesseth, C. A., Kirkpatrick, B., Schon, T. B., Aston, J. A. D., & Bouchard-Cote, A. (2017). Divide-and-conquer with sequential Monte Carlo. Journal of Computational and Graphical Statistics, 26(2), 445-458.
- 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.