MRC Biostatistics Unit

UNIVERSITY OF CAMBRIDGE

# Markov melding
## A general method for integrating Bayesian models

Robert Goudie, MRC Biostatistics Unit, University of Cambridge

RSS Emerging Application Section workshop, 22 March 2019

# Motivation

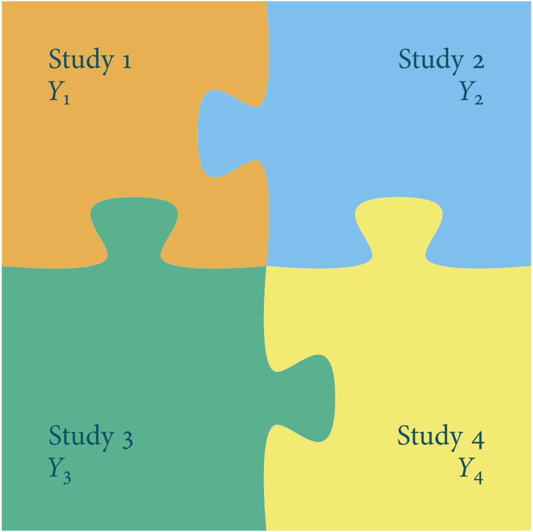
Study 1
$Y_1$

Study 2
$Y_2$

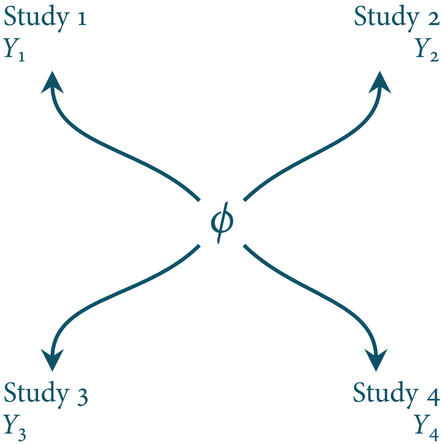
Study 3
$Y_3$

Study 4
$Y_4$

# Motivation



# Motivation



Study 1
$Y_1$

Study 2
$Y_2$

$\phi$

Study 3
$Y_3$

Study 4
$Y_4$

By using all available data we typically get

More precise estimates • More accurate reflection of true uncertainty • Minimisation of the risk of selection-type biases

# Motivation

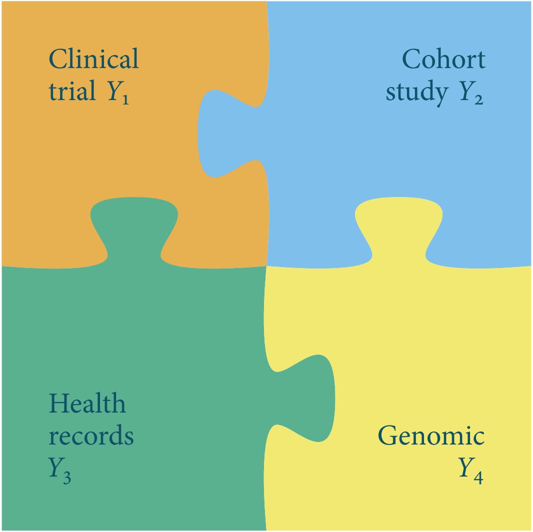
Clinical trial $Y_1$

Cohort study $Y_2$

Health records $Y_3$

Genomic $Y_4$

# Motivation



Will be hard to

formulate a suitable model • fit the resulting model • assess the resulting model

# Motivation

Clinical trial

$p_1(\phi, \psi_1, Y_1)$

# Motivation

Clinical trial

$p_1(\phi, \psi_1, Y_1)$

Cohort study

$p_2(\phi, \psi_2, Y_2)$

# Motivation

Clinical trial

$p_1(\phi, \psi_1, Y_1)$

Cohort study

$p_2(\phi, \psi_2, Y_2)$

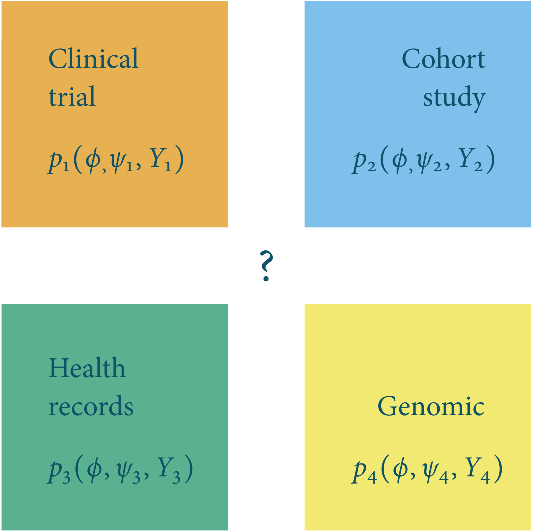
Health records

$p_3(\phi, \psi_3, Y_3)$

Genomic

$p_4(\phi, \psi_4, Y_4)$

## Motivation



Clinical trial

$p_1(\phi, \psi_1, Y_1)$

Cohort study

$p_2(\phi, \psi_2, Y_2)$

?

Health records

$p_3(\phi, \psi_3, Y_3)$

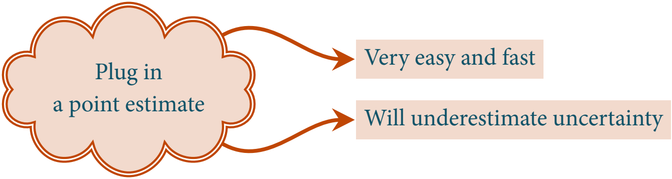
Genomic

$p_4(\phi, \psi_4, Y_4)$

Could discard the existing models and implementations, but this seems wasteful

# Modular approaches

# Modular approaches

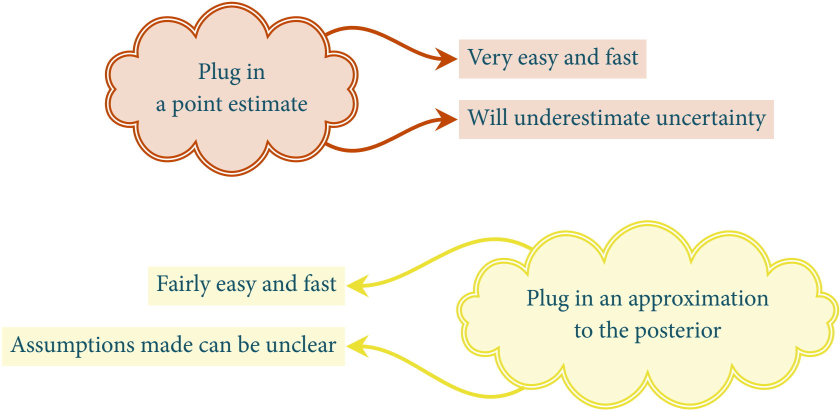


Plug in a point estimate

Very easy and fast

Will underestimate uncertainty

# Modular approaches



Plug in a point estimate

Very easy and fast

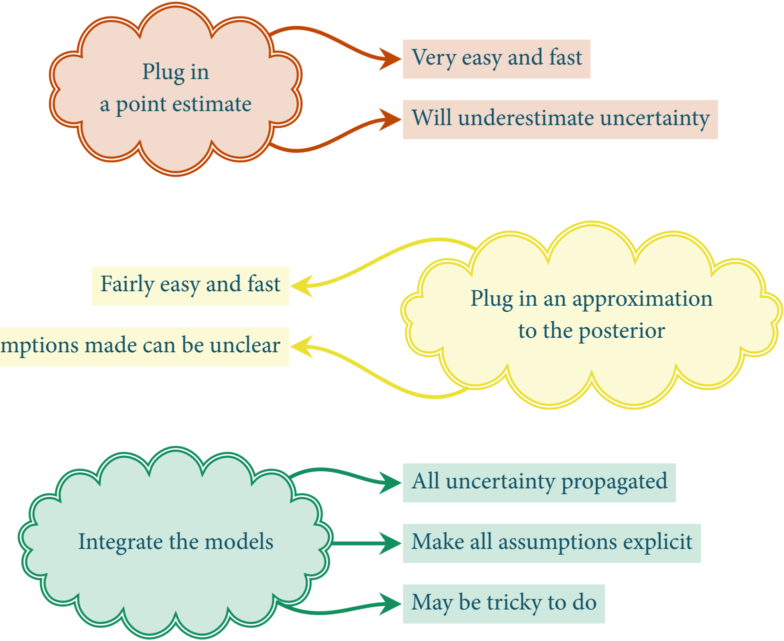
Will underestimate uncertainty

Plug in an approximation to the posterior

Fairly easy and fast

Assumptions made can be unclear

# Modular approaches



Plug in a point estimate

Very easy and fast

Will underestimate uncertainty

Plug in an approximation to the posterior

Fairly easy and fast

Assumptions made can be unclear

Integrate the models

All uncertainty propagated

Make all assumptions explicit

May be tricky to do

# A toy example Severity of flu

Quantity of interest: $p = \Pr(\text{being hospitalised} \mid \text{have influenza-like-illness})$

Observe:

- $y = 100$, the number of people in hospital with influenza-like-illness
- $n = 1000$, the number of people with influenza-like-illness

Model:

$$y \sim \text{Bin}(n, p) \qquad\qquad p \sim \text{Beta}(1, 9)$$

# A toy example Severity of flu

Quantity of interest: $p = \Pr(\text{being hospitalised} \mid \text{have influenza-like-illness})$

Observe:

- $y = 100$, the number of people in hospital with influenza-like-illness
- ~~$n = 1000$, the number of people with influenza-like-illness~~

Model:

$$y \sim \text{Bin}(n, p) \qquad p \sim \text{Beta}(1, 9)$$

$$n \sim \text{Po}(1000)$$

Quantity of interest: $p = \Pr(\text{being hospitalised} \mid \text{have influenza-like-illness})$

Observe:

- $y = 100$, the number of people in hospital with influenza-like-illness
- ~~$n = 1000$, the number of people with influenza-like-illness~~

Model:

$$y \sim \text{Bin}(n, p) \qquad p \sim \text{Beta}(1, 9)$$

$$n \sim \text{Po}(1000)$$

New data from a similar area: $x = 40$ out of $m = 500$ had influenza-like-illness.

Model for new data:

$$x \sim \text{Bin}(m, q) \qquad q \sim \text{Beta}(0.5, 5)$$

# A toy example Severity of flu

Quantity of interest: $p = \Pr(\text{being hospitalised} \mid \text{have influenza-like-illness})$

Observe:

- $y = 100$, the number of people in hospital with influenza-like-illness
- ~~$n = 1000$, the number of people with influenza-like-illness~~

Model:

$$y \sim \text{Bin}(n, p) \qquad\qquad p \sim \text{Beta}(1, 9)$$

$$n \sim \text{Po}(1000)$$

New data from a similar area: $x = 40$ out of $m = 500$ had influenza-like-illness.

Model for new data:

$$x \sim \text{Bin}(m, q) \qquad\qquad q \sim \text{Beta}(0.5, 5)$$

Given areas are similar, if we knew the total population size $N$ is in the original area, assume

$$n \sim \text{Bin}(N, q)$$

→ BUT now two models for $n$

# A toy example Severity of flu

Quantity of interest: $p = \Pr(\text{being hospitalised} \mid \text{have influenza-like-illness})$

Observe:

- $y = 100$, the number of people in hospital with influenza-like-illness
- ~~$n = 1000$, the number of people with influenza-like-illness~~

Model: $i = 1, 2$

$$y_i \sim \text{Bin}(n_i, p_i) \qquad\qquad p_i \sim \text{Beta}(1, 9)$$

$$n_i \sim \text{Po}(1000)$$

New data from a similar area: $x = 40$ out of $m = 500$ had influenza-like-illness.

Model for new data:

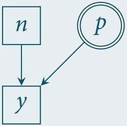
$$x \sim \text{Bin}(m, q) \qquad\qquad q \sim \text{Beta}(0.5, 5)$$

Given areas are similar, if we knew the total population size $N$ is in the original area, assume

$$n \sim \text{Bin}(N, q)$$ → BUT now two models for $n$

Now have a direct model for $n$ itself, **and** a model for $n_1$ and $n_2$ where $n = n_1 + n_2$

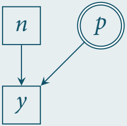
# A toy example Graphical representation



$$y \sim \text{Bin}(n, p)$$

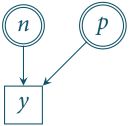
$$p \sim \text{Beta}(1, 9)$$

# A toy example Graphical representation



$$y \sim \text{Bin}(n, p)$$

$$p \sim \text{Beta}(1, 9)$$

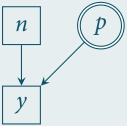


$$y \sim \text{Bin}(n, p)$$

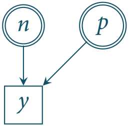
$$p \sim \text{Beta}(1, 9)$$

$$n \sim \text{Po}(1000)$$

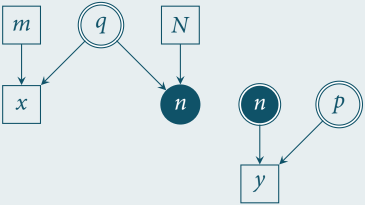
# A toy example Graphical representation



$$y \sim \mathrm{Bin}(n, p)$$
$$p \sim \mathrm{Beta}(1, 9)$$



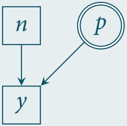
$$y \sim \mathrm{Bin}(n, p)$$
$$p \sim \mathrm{Beta}(1, 9)$$
$$n \sim \mathrm{Po}(1000)$$



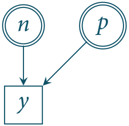
$$x \sim \mathrm{Bin}(m, q)$$
$$q \sim \mathrm{Beta}(0.5, 5)$$
$$n \sim \mathrm{Bin}(N, q)$$

$$y \sim \mathrm{Bin}(n, p)$$
$$p \sim \mathrm{Beta}(1, 9)$$
$$n \sim \mathrm{Po}(1000)$$

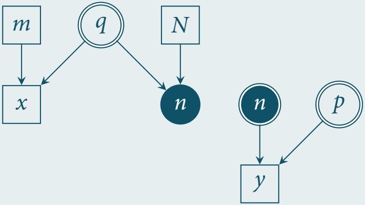
## A toy example Graphical representation



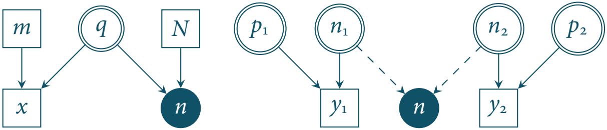
$$y \sim \mathrm{Bin}(n, p)$$
$$p \sim \mathrm{Beta}(1, 9)$$



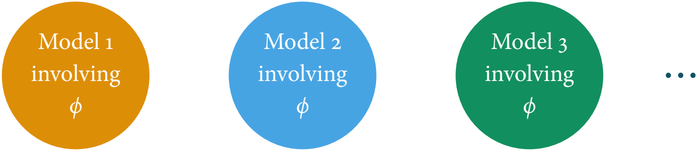
$$y \sim \mathrm{Bin}(n, p)$$
$$p \sim \mathrm{Beta}(1, 9)$$
$$n \sim \mathrm{Po}(1000)$$



$$x \sim \mathrm{Bin}(m, q) \qquad y \sim \mathrm{Bin}(n, p)$$
$$q \sim \mathrm{Beta}(0.5, 5) \qquad p \sim \mathrm{Beta}(1, 9)$$
$$n \sim \mathrm{Bin}(N, q) \qquad n \sim \mathrm{Po}(1000)$$



# Aims of this work



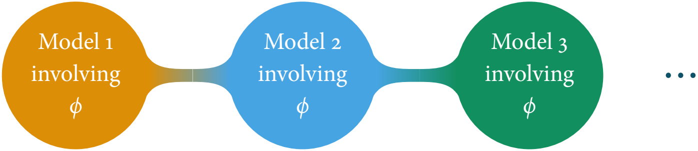
Model 1 involving $\phi$

Model 2 involving $\phi$

Model 3 involving $\phi$

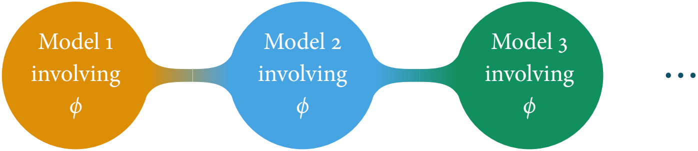
…

# Aims of this work



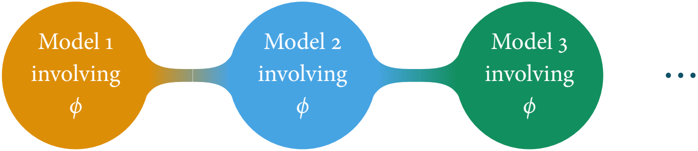
1. Create a generic method for joining submodels that share a common quantity $\phi$ into a single, joint model
   - Need to handle (implicitly) having two different priors for the same quantity
   - Need to handle models linked by non-invertible deterministic transformations

# Aims of this work



1. Create a generic method for joining submodels that share a common quantity $\phi$ into a single, joint model
    - Need to handle (implicitly) having two different priors for the same quantity
    - Need to handle models linked by non-invertible deterministic transformations
2. Fit the joint model in a staged/modular manner, one submodel at a time
    - Want the extra burden compared to plug-in approaches to be as small as possible

# Aims of this work

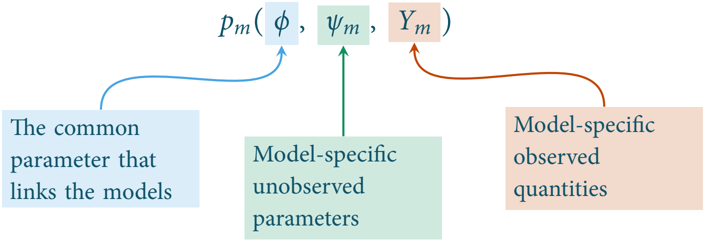


1. Create a generic method for joining submodels that share a common quantity $\phi$ into a single, joint model
   - Need to handle (implicitly) having two different priors for the same quantity
   - Need to handle models linked by non-invertible deterministic transformations

2. Fit the joint model in a staged/modular manner, one submodel at a time
   - Want the extra burden compared to plug-in approaches to be as small as possible

3. Understand the reverse operation: i.e. splitting large models into submodels to aid inference and understanding

Suppose we have models $m = 1, \dots, M$



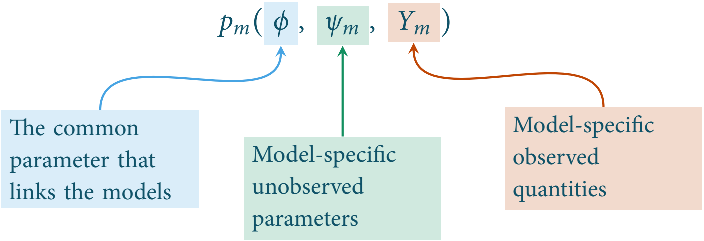
$$p_m(\phi, \psi_m, Y_m)$$

The common parameter that links the models

Model-specific unobserved parameters

Model-specific observed quantities

Suppose we have models $m = 1, \ldots, M$



$$p_m(\phi, \psi_m, Y_m)$$

The common parameter that links the models

Model-specific unobserved parameters

Model-specific observed quantities

↓

Want a generic method that integrates these models into a single joint model

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

# Joining models Markov combination

Suppose consistent marginals: i.e. $p_m(\phi) = p(\phi)$ is the same for all $m$

First isolate $\phi$ by conditioning:

$$p_m(\phi, \psi_m, Y_m) = p_m(\psi_m, Y_m \mid \phi)\, p_m(\phi)$$

This suggests the following joint model:

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

This is called Markov combination — Dawid and Lauritzen (1993), Massa and Lauritzen (2010)

Dawid, A. P. and Lauritzen, S. L. (1993). "Hyper Markov laws in the statistical analysis of decomposable graphical models". *Annals of Statistics* **21**, 1272–1317.

Massa, M. S. and Lauritzen, S. L. (2010). "Combining statistical models". In: *Contemporary Mathematics: Algebraic Methods in Statistics and Probability II*. ed. by Viana, M. A. G. and Wynn, H. P., pp. 239–260.

Suppose inconsistent marginals i.e. $p_1(\phi), \ldots, p_M(\phi)$ are not all equal

Instead choose a pooled density

$$p_{\text{pool}}(\phi) = g\Big(p_1(\phi), \ldots, p_M(\phi)\Big)$$

This suggests the following joint model:

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

We call this Markov melding (Goudie *et al.*, 2019)

Goudie, R. J. B. *et al.* (2019). "Joining and splitting models with Markov melding". *Bayesian Analysis* **14**, 81–109.

Similar problem to forming a single prior to use when several experts have been asked to supply their prior.

Several pooling functions have been suggested (O'Hagan *et al.*, 2006)

- Linear opinion pooling
  $$p_{\text{pool}}(\phi) = \sum_{m=1}^{M} w_m p_m(\phi)$$
- Logarithmic opinion pooling
  $$p_{\text{pool}}(\phi) \propto \prod_{m=1}^{M} p_m(\phi)^{w_m}$$
- Product of experts pooling (Hinton, 2002)
  $$p_{\text{pool}}(\phi) \propto \prod_{m=1}^{M} p_m(\phi)$$
- Dictatorial pooling
  $p_{\text{pool}}(\phi) = p_m(\phi)$ some $m \in \{1, \ldots, M\}$

O'Hagan, A. *et al.* (2006). *Uncertain Judgements: Eliciting Experts' Probabilities*. Chichester: John Wiley & Sons.
Hinton, G. E. (2002). "Training products of experts by minimizing contrastive divergence.". *Neural computation* **14**, 1771–1800.

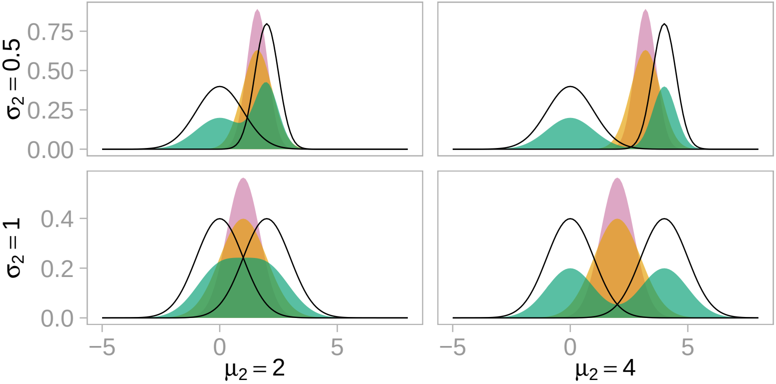


Figure: Pooling ($w_m = 0.5$) the black densities shown: $N(0, 1)$ and $N(\mu_2, \sigma_2^2)$; $\mu_2$ and $\sigma_2$ shown in row/column labels.

linear opinion pooling

$$p_{\text{pool}}(\phi) = \sum_{m=1}^{M} w_m p_m(\phi)$$

logarithmic opinion pooling

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

product of experts pooling

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

## Externally Bayesian pooling

A pooling function $g$ is called externally Bayesian if Bayesian updating and pooling are interchangeable.

$$g(\text{posterior}(\text{model}_1), \ldots, \text{posterior}(\text{model}_M)) \propto \text{posterior}(g(\text{model}_1, \ldots, \text{model}_M))$$

## Externally Bayesian pooling

A pooling function $g$ is called externally Bayesian if Bayesian updating and pooling are interchangeable.

$$g(\text{posterior}(\text{model}_1), \ldots, \text{posterior}(\text{model}_M)) \propto \text{posterior}(g(\text{model}_1, \ldots, \text{model}_M))$$

For $M$ models $p_i(\phi, Y) = p(Y \mid \phi) p_i(\phi) \qquad i = 1, \ldots, M$ with the same likelihood

$$g\Big(p_1(\phi \mid Y), \ldots, p_M(\phi \mid Y)\Big) \propto p(Y \mid \phi)\, g\Big(p_1(\phi), \ldots, p_M(\phi)\Big)$$

## Externally Bayesian pooling

A pooling function $g$ is called externally Bayesian if Bayesian updating and pooling are interchangeable.

$$g(\text{posterior}(\text{model}_1), \ldots, \text{posterior}(\text{model}_M)) \propto \text{posterior}(g(\text{model}_1, \ldots, \text{model}_M))$$

For $M$ models $p_i(\phi, Y) = p(Y \mid \phi) p_i(\phi) \qquad i = 1, \ldots, M$ with the same likelihood

$$g\Big(p_1(\phi \mid Y), \ldots, p_M(\phi \mid Y)\Big) \propto p(Y \mid \phi)\, g\Big(p_1(\phi), \ldots, p_M(\phi)\Big)$$

Logarithmic pooling is externally Bayesian when $\sum_{i=1}^{M} w_i = 1$. (Genest and Zidek, 1986)

Genest, C. and Zidek, J. V. (1986). “Combining probability distributions: A critique and an annotated bibliography”. *Statistical Science* **1**, 114–135.

A pooling function $g$ is called externally Bayesian if Bayesian updating and pooling are interchangeable.

$$g(\text{posterior}(\text{model}_1), \ldots, \text{posterior}(\text{model}_M)) \propto \text{posterior}(g(\text{model}_1, \ldots, \text{model}_M))$$

For $M$ models $p_i(\phi, Y) = p(Y \mid \phi)p_i(\phi) \qquad i = 1, \ldots, M$ with the same likelihood

$$g\Big(p_1(\phi \mid Y), \ldots, p_M(\phi \mid Y)\Big) \propto p(Y \mid \phi)\, g\Big(p_1(\phi), \ldots, p_M(\phi)\Big)$$

Logarithmic pooling is externally Bayesian when $\sum_{i=1}^{M} w_i = 1$. (Genest and Zidek, 1986)

However, this property is not applicable when combining several distinct likelihoods with distinct data, since

$$g\Big(p_1(\phi, \psi_1 \mid Y_1), \ldots, p_M(\phi, \psi_M \mid Y_M)\Big) \not\propto g\Big(p_1(\phi), \ldots, p_M(\phi)\Big) \prod_i p_i(Y_i, \psi_i \mid \phi)$$

Genest, C. and Zidek, J. V. (1986). "Combining probability distributions: A critique and an annotated bibliography". *Statistical Science* **1**, 114–135.

## Miscellaneous notes

Writing the model conditional on $\phi$ may not be analytically tractable

- e.g. suppose $\phi = f(\theta)$, where $f$ is a non-invertible deterministic function, like Bayesian melding (Poole and Raftery, 2000)
- Markov melding model turns out to be invariant to the choice of extension of $f$

Poole, D. and Raftery, A. E. (2000). “Inference for deterministic simulation models: The Bayesian melding approach”. *Journal of the American Statistical Association* **95**, 1244–1255.

Writing the model conditional on $\phi$ may not be analytically tractable

- e.g. suppose $\phi = f(\theta)$, where $f$ is a non-invertible deterministic function, like Bayesian melding (Poole and Raftery, 2000)
- Markov melding model turns out to be invariant to the choice of extension of $f$

A simple two-step approximate approach is sometimes used when joining two models

1. Approximate posterior marginal of $\phi$ under model 1 by $p_N(\phi \mid \widehat{\mu}, \widehat{\Sigma}) \approx p_1(\phi \mid y_1)$
2. Modify likelihood of the second model by a factor $p_N(\widehat{\mu} \mid \phi, \widehat{\Sigma})$

Turns out that this an approximation to Markov melding with PoE pooling.

Poole, D. and Raftery, A. E. (2000). "Inference for deterministic simulation models: The Bayesian melding approach". *Journal of the American Statistical Association* **95**, 1244–1255.

Writing the model conditional on $\phi$ may not be analytically tractable

- e.g. suppose $\phi = f(\theta)$, where $f$ is a non-invertible deterministic function, like Bayesian melding (Poole and Raftery, 2000)
- Markov melding model turns out to be invariant to the choice of extension of $f$

A simple two-step approximate approach is sometimes used when joining two models

1. Approximate posterior marginal of $\phi$ under model 1 by $p_N(\phi \mid \widehat{\mu}, \widehat{\Sigma}) \approx p_1(\phi \mid y_1)$
2. Modify likelihood of the second model by a factor $p_N(\widehat{\mu} \mid \phi, \widehat{\Sigma})$

Turns out that this an approximation to Markov melding with PoE pooling.

Markov melding is defined for any collection of submodels BUT that is not a guarantee that the joint model is appropriate

- If two submodels $p_m(\phi, \psi_m, Y_m)$ and $p_\ell(\phi, \psi_\ell, Y_\ell)$ strongly conflict, the posterior from the joint model will be misleading

Poole, D. and Raftery, A. E. (2000). “Inference for deterministic simulation models: The Bayesian melding approach”. *Journal of the American Statistical Association* **95**, 1244–1255.

# Inference

The joint posterior distribution $p_{\text{meld}}$ is

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

# Inference

The joint posterior distribution $p_{\text{meld}}$ is

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

Under product of experts pooling the posterior simplifies to

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

# Inference

The joint posterior distribution $p_{\text{meld}}$ is

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

Under product of experts pooling the posterior simplifies to

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

Under other pooling, the posterior distribution involves the prior marginals $p_m(\phi)$, $m = 1, \ldots, M$

1. These may be analytically tractable (or even directly available)
2. Otherwise, we can estimate them by sampling then using kernel density estimation

## Inference Two stage MCMC sampler

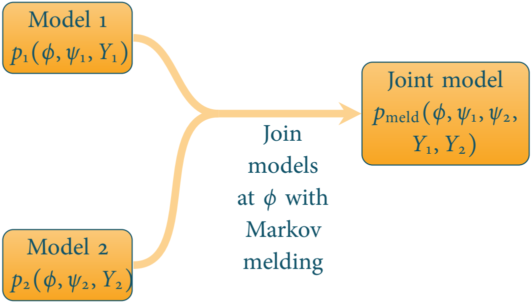
With $M = 2$ models:

Model 1
$p_1(\phi, \psi_1, Y_1)$

Model 2
$p_2(\phi, \psi_2, Y_2)$

## Inference | Two stage MCMC sampler

With $M = 2$ models:



Model 1

$p_1(\phi, \psi_1, Y_1)$

Model 2

$p_2(\phi, \psi_2, Y_2)$

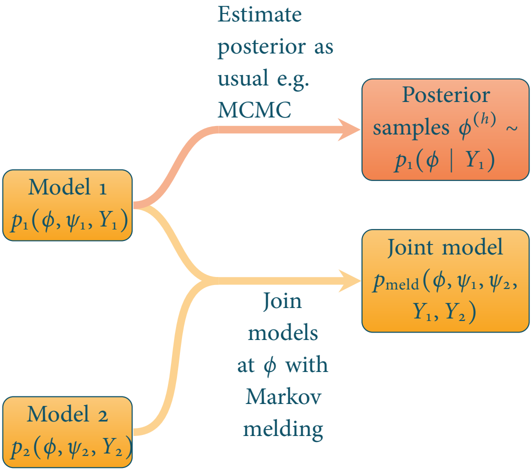
Join models at $\phi$ with Markov melding

Joint model

$p_{\text{meld}}(\phi, \psi_1, \psi_2, Y_1, Y_2)$

## Inference Two stage MCMC sampler

With $M = 2$ models:



Estimate posterior as usual e.g. MCMC

Posterior samples $\phi^{(h)} \sim p_1(\phi \mid Y_1)$

Model 1 $p_1(\phi, \psi_1, Y_1)$

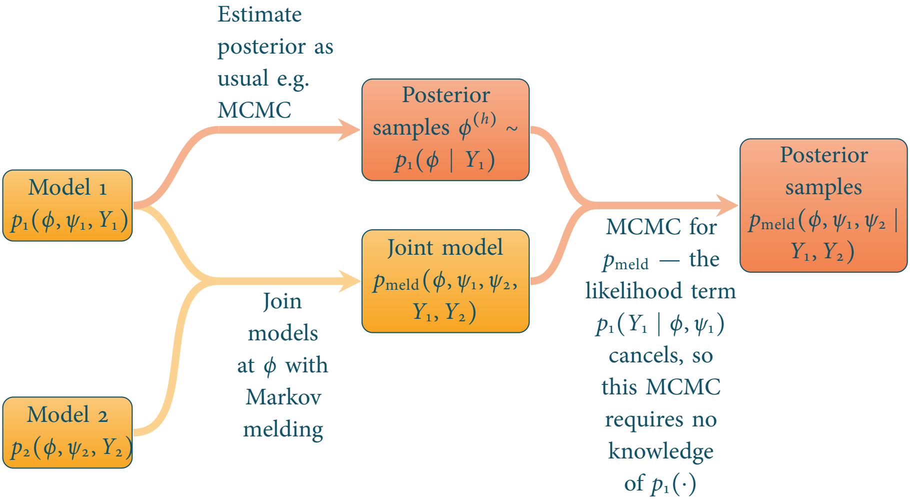
Joint model $p_{\text{meld}}(\phi, \psi_1, \psi_2, Y_1, Y_2)$

Join models at $\phi$ with Markov melding

Model 2 $p_2(\phi, \psi_2, Y_2)$

# Inference Two stage MCMC sampler

With $M = 2$ models:



Model 1 $p_1(\phi, \psi_1, Y_1)$

Model 2 $p_2(\phi, \psi_2, Y_2)$

Estimate posterior as usual e.g. MCMC

Posterior samples $\phi^{(h)} \sim p_1(\phi \mid Y_1)$

Join models at $\phi$ with Markov melding

Joint model $p_{\text{meld}}(\phi, \psi_1, \psi_2, Y_1, Y_2)$

MCMC for $p_{\text{meld}}$ — the likelihood term $p_1(Y_1 \mid \phi, \psi_1)$ cancels, so this MCMC requires no knowledge of $p_1(\cdot)$

Posterior samples $p_{\text{meld}}(\phi, \psi_1, \psi_2 \mid Y_1, Y_2)$

# Inference Two stage MCMC sampler

With $M = 2$ models:

1. Stage 1: Model 1 posterior
   Draw and retain samples $(\phi^{(h)}, \psi_1^{(h)}) \sim p_1(\phi, \psi_1 \mid y_1)$, $h = 1, \ldots, H$

# Inference Two stage MCMC sampler

With $M = 2$ models:

1. Stage 1: Model 1 posterior
   Draw and retain samples $(\phi^{(h)}, \psi_1^{(h)}) \sim p_1(\phi, \psi_1 \mid y_1), h = 1, \ldots, H$

2. Stage 2: Markov melded posterior

   To update $\psi_2 \mid (\phi, \psi_1)$ using the usual method for model 2

# Inference Two stage MCMC sampler

With $M = 2$ models:

1. Stage 1: Model 1 posterior
   Draw and retain samples $(\phi^{(h)}, \psi_1^{(h)}) \sim p_1(\phi, \psi_1 \mid y_1)$, $h = 1, \ldots, H$

2. Stage 2: Markov melded posterior

   To update $\psi_2 \mid (\phi, \psi_1)$ using the usual method for model 2

   To update $(\phi, \psi_1) \mid \psi_2$ draw an index $d$ uniformly at random from $\{1, \ldots, H\}$, so

   $$(\phi^\star, \psi_1^\star) = (\phi^{(d)}, \psi_1^{(d)})$$

   Accept with probability $\min(1, r)$ where

   $$r = \frac{p_{\text{pool}}(\phi^\star) \times p_1(\phi^\star)^{-1} \times p_2(\phi^\star, \psi_2, y_2) p_2(\phi^\star)^{-1}}{p_{\text{pool}}(\phi) \times p_1(\phi)^{-1} \times p_2(\phi, \psi_2, y_2) p_2(\phi)^{-1}}$$

# Inference Two stage MCMC sampler

With $M = 2$ models:

1. Stage 1: Model 1 posterior
   Draw and retain samples $(\phi^{(h)}, \psi_1^{(h)}) \sim p_1(\phi, \psi_1 \mid y_1)$, $h = 1, \dots, H$

2. Stage 2: Markov melded posterior

   To update $\psi_2 \mid (\phi, \psi_1)$ using the usual method for model 2

   To update $(\phi, \psi_1) \mid \psi_2$ draw an index $d$ uniformly at random from $\{1, \dots, H\}$, so

$$(\phi^\star, \psi_1^\star) = (\phi^{(d)}, \psi_1^{(d)})$$

   Accept with probability $\min(1, r)$ where

$$r = \frac{p_{\text{pool}}(\phi^\star) \times p_1(\phi^\star)^{-1} \times p_2(\phi^\star, \psi_2, y_2) p_2(\phi^\star)^{-1}}{p_{\text{pool}}(\phi) \times p_1(\phi)^{-1} \times p_2(\phi, \psi_2, y_2) p_2(\phi)^{-1}}$$

   Note the acceptance probability does not depend on the likelihood of model 1 at all

# Inference Two stage MCMC sampler

With $M = 2$ models:

1. Stage 1: Model 1 posterior
   Draw and retain samples $(\phi^{(h)}, \psi_1^{(h)}) \sim p_1(\phi, \psi_1 \mid y_1)$, $h = 1, \ldots, H$

2. Stage 2: Markov melded posterior

   To update $\psi_2 \mid (\phi, \psi_1)$ using the usual method for model 2

   To update $(\phi, \psi_1) \mid \psi_2$ draw an index $d$ uniformly at random from $\{1, \ldots, H\}$, so

$$(\phi^\star, \psi_1^\star) = (\phi^{(d)}, \psi_1^{(d)})$$

   Accept with probability $\min(1, r)$ where

$$r = \frac{p_{\text{pool}}(\phi^\star) \times p_1(\phi^\star)^{-1} \times p_2(\phi^\star, \psi_2, y_2) p_2(\phi^\star)^{-1}}{p_{\text{pool}}(\phi) \times p_1(\phi)^{-1} \times p_2(\phi, \psi_2, y_2) p_2(\phi)^{-1}}$$

Note the acceptance probability does not depend on the likelihood of model 1 at all
Extends naturally to an $M$-stage algorithm when $M > 2$

# Influenza example

## Background

Public health responses to influenza outbreaks rely on knowledge of severity: the probability that an infection results in a severe event such as hospitalisation or death

Presanis, A. M. *et al.* (2014). "Synthesising evidence to estimate pandemic (2009) A/H1N1 influenza severity in 2009–2011". *The Annals of Applied Statistics* **8**, 2378–2403.

# Influenza example

## Background

Public health responses to influenza outbreaks rely on knowledge of severity: the probability that an infection results in a severe event such as hospitalisation or death

## Aim

To estimate the total number $\chi$ of ICU admissions for the A/H1N1 strain during the 2010/2011 influenza season in England

Presanis, A. M. *et al.* (2014). "Synthesising evidence to estimate pandemic (2009) A/H1N1 influenza severity in 2009–2011". *The Annals of Applied Statistics* **8**, 2378–2403.

# Influenza example

## Background

Public health responses to influenza outbreaks rely on knowledge of severity: the probability that an infection results in a severe event such as hospitalisation or death

## Aim

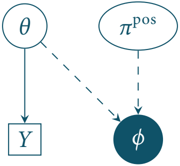
To estimate the total number $\chi$ of ICU admissions for the A/H1N1 strain during the 2010/2011 influenza season in England

## Data sources

- Observations of the (weekly) number of suspected prevalent cases of A/H1N1 in ICUs
- Weekly virological positivity data from the sentinel laboratory surveillance system
- Many other indirect data (number of GP consultations, suspected hospitalisations outside ICUs, deaths etc) — here simplified to an informative prior

See Presanis *et al.* (2014) for details

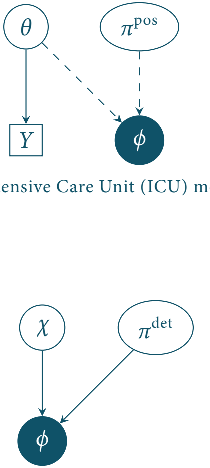
Presanis, A. M. *et al.* (2014). “Synthesising evidence to estimate pandemic (2009) A/H1N1 influenza severity in 2009–2011”. *The Annals of Applied Statistics* **8**, 2378–2403.



(a) Intensive Care Unit (ICU) model

Model (a)

- $Y$ is (weekly) data recording the number of suspected cases of A/H1N1 in ICUs
- $\pi^{\text{pos}}$ is probability of suspected A/H1N1 being real, based on virological data (not shown).
- Given $\theta$ and $\pi^{\text{pos}}$, we estimate the confirmed number $\phi$ of cases of A/H1N1 in ICUs



(a) Intensive Care Unit (ICU) model

(b) Severity model (simplified)

Figure: High-level DAGs of the A/H1N1 models

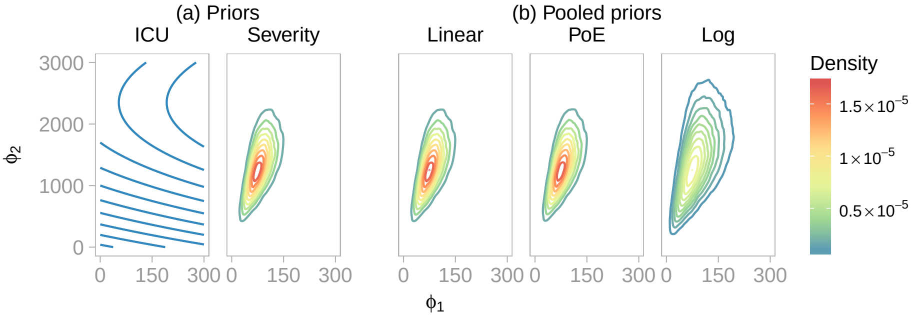
## Model (a)

- $Y$ is (weekly) data recording the number of suspected cases of A/H1N1 in ICUs
- $\pi^{\text{pos}}$ is probability of suspected A/H1N1 being real, based on virological data (not shown).
- Given $\theta$ and $\pi^{\text{pos}}$, we estimate the confirmed number $\phi$ of cases of A/H1N1 in ICUs

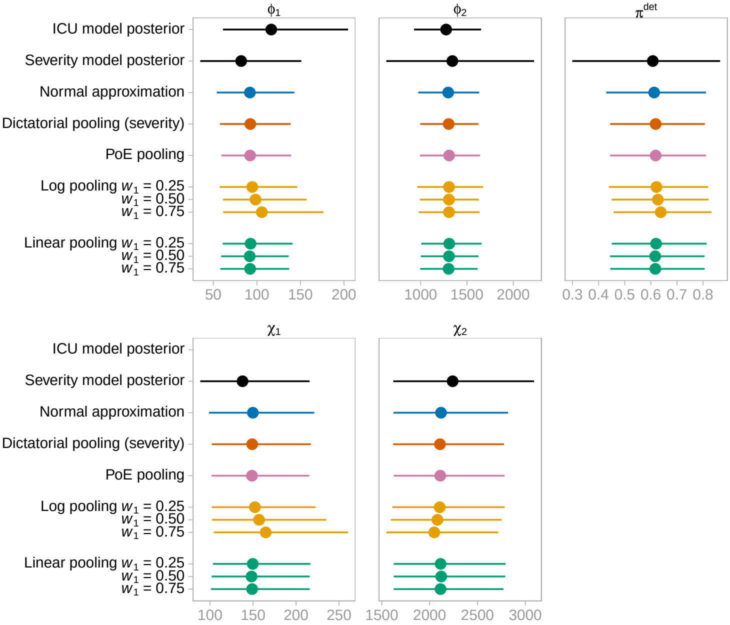
## Model (b)

- $\phi \sim \text{Bin}(\chi, \pi^{\text{det}})$, because data $Y$ is known to miss some cases in ICUs
- An informative prior is chosen for $\chi$ that represents the other data sources

## Influenza example Prior poolimg



# Influenza example Comparison of methods



Profile regression (Molitor et al, 2010) non-parameterically links a response vector $y$ with covariate data (e.g. biomarker data $b$) through cluster membership.

The likelihood, with parameters $\Omega$, for subject $i$ is an infinite mixture model, with latent cluster allocation $d_i$.

$$\Pr(y_i, b_i \mid \Omega) = \sum_d \underbrace{\Pr(y_i \mid d_i = d, \zeta_d)}_{\text{Response model}} \underbrace{\Pr(b_i \mid d_i = d, \chi_d)}_{\text{Biomarker model}}$$

This model is implemented in the R-package PReMiuM (Liverani *et al.*, 2015).

Liverani, S. *et al.* (2015). "PReMiuM: An R package for profile regression mixture models using Dirichlet processes". *Journal of Statistical Software* **64**.

# Joining models Longitudinal models with clustering

Profile regression (Molitor et al, 2010) non-parameterically links a response vector *y* with covariate data (e.g. biomarker data *b*) through cluster membership.

The likelihood, with parameters Ω, for subject *i* is an infinite mixture model, with latent cluster allocation $d_i$.

$$\Pr(y_i, b_i \mid \Omega) = \sum_d \underbrace{\Pr(y_i \mid d_i = d, \zeta_d)}_{\text{Response model}} \underbrace{\Pr(b_i \mid d_i = d, \chi_d)}_{\text{Biomarker model}}$$

This model is implemented in the R-package PReMiuM (Liverani *et al.*, 2015).

Now suppose you wish to use as the response vector *y* a quantity that is itself uncertain – that is we have another model for it.

Liverani, S. *et al.* (2015). "PReMiuM: An R package for profile regression mixture models using Dirichlet processes". *Journal of Statistical Software* **64**.

Profile regression (Molitor et al, 2010) non-parameterically links a response vector $y$ with covariate data (e.g. biomarker data $b$) through cluster membership.

The likelihood, with parameters $\Omega$, for subject $i$ is an infinite mixture model, with latent cluster allocation $d_i$.

$$\Pr(y_i, b_i \mid \Omega) = \sum_d \underbrace{\Pr(y_i \mid d_i = d, \zeta_d)}_{\text{Response model}} \underbrace{\Pr(b_i \mid d_i = d, \chi_d)}_{\text{Biomarker model}}$$

This model is implemented in the R-package PReMiuM (Liverani *et al.*, 2015).

Now suppose you wish to use as the response vector $y$ a quantity that is itself uncertain – that is we have another model for it.

No need to build a new implementation of profile regression *and* the other model – instead:

1. Save samples for $y$ from the extra model
2. Add a Metropolis-Hastings step for $y$ to PReMiuM as described above

Liverani, S. *et al.* (2015). “PReMiuM: An R package for profile regression mixture models using Dirichlet processes”. *Journal of Statistical Software* **64**.

# Splitting models

Suppose we have joint model

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

# Splitting models

Suppose we have joint model

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

$$\downarrow$$

We would like to split this joint into submodels $m = 1, \ldots, M$

$$p_m(\phi, \psi_m, Y_m)$$

such that the submodels are “faithful” to the original model — i.e. joining the submodels results in the original model.

# Splitting models

Suppose we have joint model

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

$$\downarrow$$

We would like to split this joint into submodels $m = 1, \ldots, M$

$$p_m(\phi, \psi_m, Y_m)$$

such that the submodels are “faithful” to the original model — i.e. joining the submodels results in the original model.

Potential uses for model splitting

- Dividing up computation
  The staged algorithm using the submodels $p_m(\phi, \psi_m, Y_m)$ will fit the original joint model
- Improving understanding
  Differences between submodels for $\phi$ under the joint model can be assessed.

# Splitting models

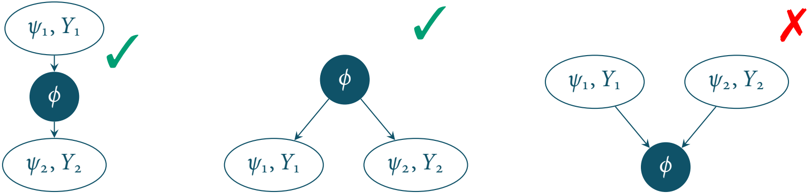
Splitting based on Markov melding is suitable only if

$$(\psi_m, Y_m) \perp\!\!\!\perp (\psi_\ell, Y_\ell) \mid \phi \qquad \text{for } m \neq \ell$$

# Splitting models

Splitting based on Markov melding is suitable only if

$$(\psi_m, Y_m) \perp\!\!\!\perp (\psi_\ell, Y_\ell) \mid \phi \qquad \text{for } m \neq \ell$$



(d) 'Chain' motif

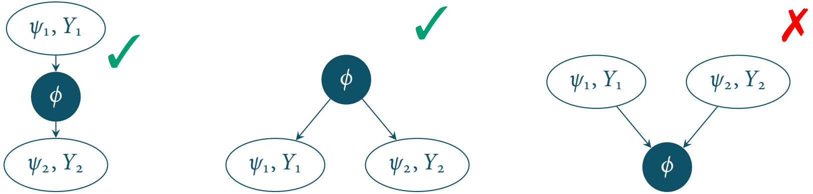
(e) 'Tail-to-tail' motif

(f) 'Head-to-head' motif

# Splitting models

Splitting based on Markov melding is suitable only if

$$(\psi_m, Y_m) \perp\!\!\!\perp (\psi_\ell, Y_\ell) \mid \phi \qquad \text{for } m \neq \ell$$



(g) ‘Chain’ motif

(h) ‘Tail-to-tail’ motif

(i) ‘Head-to-head’ motif

Suitable submodels are

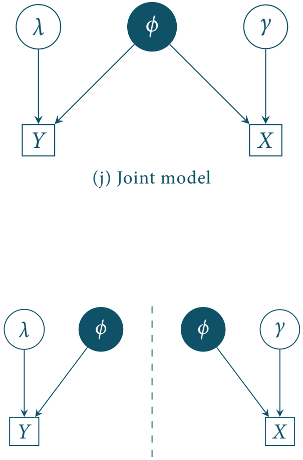
$$p_m(\phi, \psi_m, Y_m) = p(\psi_m, Y_m \mid \phi)\widetilde{p}_m(\phi), \qquad m = 1, \ldots, M$$

where $\widetilde{p}_1(\phi), \ldots, \widetilde{p}_M(\phi)$ are some new prior marginal distributions, which can be chosen freely provided that

$$p_{\text{pool}}(\phi) = g(\widetilde{p}_1(\phi), \ldots, \widetilde{p}_M(\phi)) = p(\phi)$$

e.g. with PoE pooling, either $\widetilde{p}_m(\phi) = p(\phi)^{1/M}$; or *any* factorisation of $p(\phi)$ into $M$ factors

# Ecology splitting example | Model structures



(j) Joint model

(k) The model split into components

Figure: High-level DAGs of the ecology models

Two distinct sources of data about the population of British Lapwing (a species of bird) in the UK

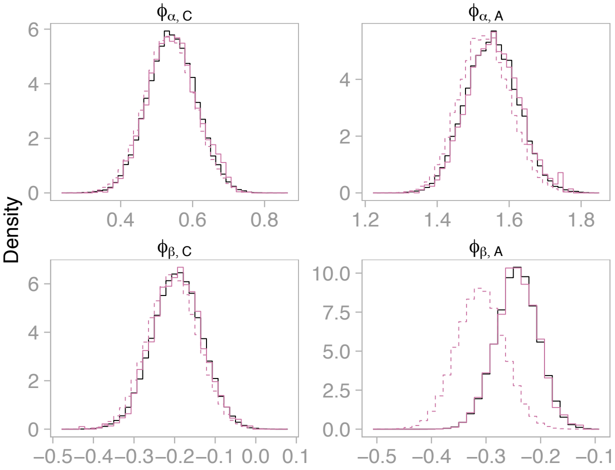
- Mark-recapture-recovery data $Y$ are modelled in terms of the recovery rate $\lambda$, and the survival rate $\phi$ for birds
- Index data $X$ are modelled in terms of the survival rate $\phi$, and the productivity rate $\gamma$ of adult female birds.

The joint model is complex and slow to fit.

**Can we split the joint model into components, then fit in stages, but still get inference for the original joint model?**

Brooks, S. P. *et al.* (2004). “A Bayesian approach to combining animal abundance and demographic data”. *Animal Biodiversity and Conservation* **27**, 515–529.

# Ecology splitting example Comparison of methods

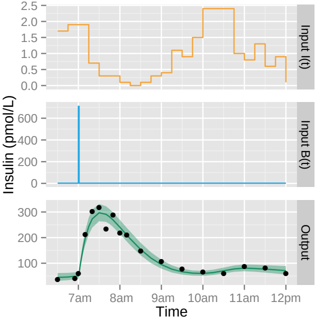


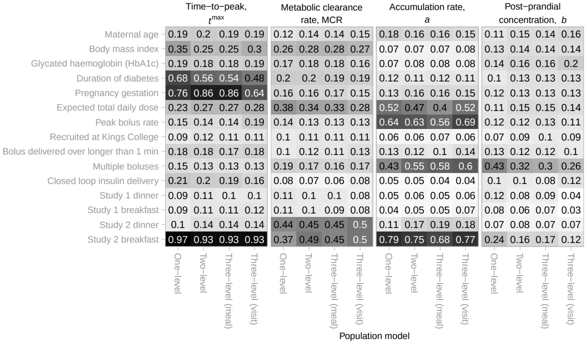
—— directly fitting full, joint model

- - - fitting only the recovery model ($(\lambda, \phi, y)$ part of model)

## Hierarchical model (Goudie *et al.*, 2015)

- Individual-level models are ODE models of insulin
- Want to perform Bayesian variable selection on models for ODE parameters
- Computation and inference much easier if we split the individual-level models from the variable selection model
- Also makes cross-validation easy to do





| | Time−to−peak, $t^{max}$ | | | | Metabolic clearance rate, MCR | | | | Accumulation rate, $a$ | | | | Post−prandial concentration, $b$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | One−level | Two−level | Three−level (meal) | Three−level (visit) | One−level | Two−level | Three−level (meal) | Three−level (visit) | One−level | Two−level | Three−level (meal) | Three−level (visit) | One−level | Two−level | Three−level (meal) | Three−level (visit) |
| Maternal age | 0.19 | 0.2 | 0.19 | 0.19 | 0.12 | 0.14 | 0.14 | 0.15 | 0.18 | 0.16 | 0.16 | 0.15 | 0.11 | 0.15 | 0.14 | 0.16 |
| Body mass index | 0.35 | 0.25 | 0.25 | 0.3 | 0.26 | 0.28 | 0.28 | 0.27 | 0.07 | 0.07 | 0.07 | 0.08 | 0.13 | 0.14 | 0.14 | 0.14 |
| Glycated haemoglobin (HbA1c) | 0.19 | 0.18 | 0.18 | 0.19 | 0.17 | 0.18 | 0.18 | 0.16 | 0.07 | 0.08 | 0.08 | 0.08 | 0.14 | 0.16 | 0.16 | 0.2 |
| Duration of diabetes | 0.68 | 0.56 | 0.54 | 0.48 | 0.2 | 0.2 | 0.19 | 0.19 | 0.12 | 0.11 | 0.12 | 0.11 | 0.1 | 0.13 | 0.13 | 0.13 |
| Pregnancy gestation | 0.76 | 0.86 | 0.86 | 0.64 | 0.16 | 0.16 | 0.17 | 0.15 | 0.13 | 0.16 | 0.16 | 0.15 | 0.12 | 0.13 | 0.13 | 0.13 |
| Expected total daily dose | 0.23 | 0.27 | 0.27 | 0.28 | 0.38 | 0.34 | 0.33 | 0.28 | 0.52 | 0.47 | 0.4 | 0.52 | 0.11 | 0.15 | 0.15 | 0.14 |
| Peak bolus rate | 0.15 | 0.14 | 0.14 | 0.19 | 0.14 | 0.13 | 0.13 | 0.13 | 0.64 | 0.63 | 0.56 | 0.69 | 0.12 | 0.12 | 0.13 | 0.11 |
| Recruited at Kings College | 0.09 | 0.12 | 0.11 | 0.11 | 0.1 | 0.11 | 0.11 | 0.11 | 0.06 | 0.06 | 0.07 | 0.06 | 0.07 | 0.09 | 0.1 | 0.09 |
| Bolus delivered over longer than 1 min | 0.18 | 0.18 | 0.17 | 0.18 | 0.1 | 0.12 | 0.11 | 0.13 | 0.13 | 0.12 | 0.1 | 0.14 | 0.13 | 0.12 | 0.12 | 0.1 |
| Multiple boluses | 0.15 | 0.13 | 0.13 | 0.13 | 0.19 | 0.17 | 0.16 | 0.17 | 0.43 | 0.55 | 0.58 | 0.6 | 0.43 | 0.32 | 0.3 | 0.26 |
| Closed loop insulin delivery | 0.21 | 0.2 | 0.19 | 0.16 | 0.08 | 0.07 | 0.06 | 0.08 | 0.05 | 0.05 | 0.04 | 0.04 | 0.1 | 0.1 | 0.08 | 0.12 |
| Study 1 dinner | 0.09 | 0.11 | 0.1 | 0.1 | 0.11 | 0.1 | 0.1 | 0.08 | 0.05 | 0.06 | 0.05 | 0.06 | 0.12 | 0.08 | 0.09 | 0.04 |
| Study 1 breakfast | 0.09 | 0.11 | 0.11 | 0.12 | 0.11 | 0.1 | 0.09 | 0.08 | 0.04 | 0.05 | 0.05 | 0.07 | 0.08 | 0.06 | 0.07 | 0.03 |
| Study 2 dinner | 0.1 | 0.14 | 0.14 | 0.14 | 0.44 | 0.45 | 0.45 | 0.5 | 0.11 | 0.17 | 0.19 | 0.18 | 0.07 | 0.08 | 0.07 | 0.07 |
| Study 2 breakfast | 0.97 | 0.93 | 0.93 | 0.93 | 0.37 | 0.49 | 0.45 | 0.5 | 0.79 | 0.75 | 0.68 | 0.77 | 0.24 | 0.16 | 0.17 | 0.12 |

Population model

Goudie, R. J. B. *et al.* (2015). "Rapid model exploration for complex hierarchical data: application to pharmacokinetics of insulin aspart.". *Statistics in Medicine* **34**, 3144–3158.

# Summary

Markov melding (Goudie *et al.*, 2019) — joining models

- Provides a generic method for joining submodels that share a common variable.
- Incorporates the idea of pooling of prior marginal distributions
- Requires that there is not strong conflict

Goudie, R. J. B. *et al.* (2019). "Joining and splitting models with Markov melding". *Bayesian Analysis* **14**, 81–109.

# Summary

Markov melding (Goudie *et al.*, 2019) — joining models

- Provides a generic method for joining submodels that share a common variable.
- Incorporates the idea of pooling of prior marginal distributions
- Requires that there is not strong conflict

Multi-stage algorithm

- Enables inference for the joint model to be conducted in submodel-specific stages
- May be easier/faster than fitting a monolithic joint model directly
- Only small changes required to existing software for a particular sub-model

Goudie, R. J. B. *et al.* (2019). “Joining and splitting models with Markov melding”. *Bayesian Analysis* **14**, 81–109.

# Summary

Markov melding (Goudie *et al*., 2019) — joining models

- Provides a generic method for joining submodels that share a common variable.
- Incorporates the idea of pooling of prior marginal distributions
- Requires that there is not strong conflict

Multi-stage algorithm

- Enables inference for the joint model to be conducted in submodel-specific stages
- May be easier/faster than fitting a monolithic joint model directly
- Only small changes required to existing software for a particular sub-model

Markov melding — splitting models

- Applying the method backwards enables splitting models
- Decomposition may aid understanding and computation for a large joint model

Goudie, R. J. B. *et al*. (2019). "Joining and splitting models with Markov melding". *Bayesian Analysis* **14**, 81–109.

# Thanks

MRC | Biostatistics Unit

UNIVERSITY OF CAMBRIDGE

Lorenz Wernisch

Anne Presanis

David Lunn

Daniela De Angelis

Brian Tom

Simon Schoenbuchner

Andrew Manderson