Joining Bayesian submodels with Markov melding

Robert Goudie, MRC Biostatistics Unit, University of Cambridge

CRiSM Seminar, University of Warwick, 2nd November 2022

Motivation

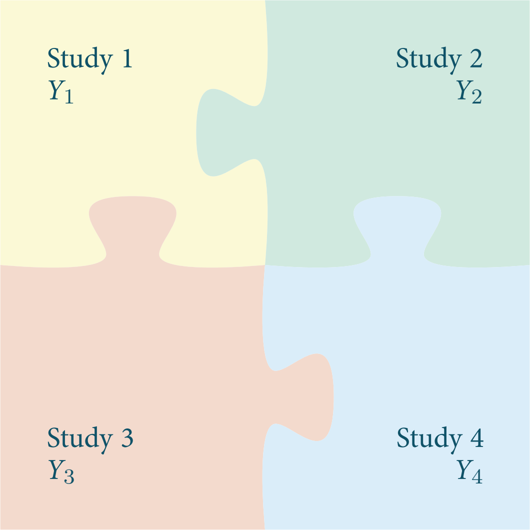
Study 1
Y_1

Study 2
Y_2

Study 3
Y_3

Study 4
Y_4

Motivation



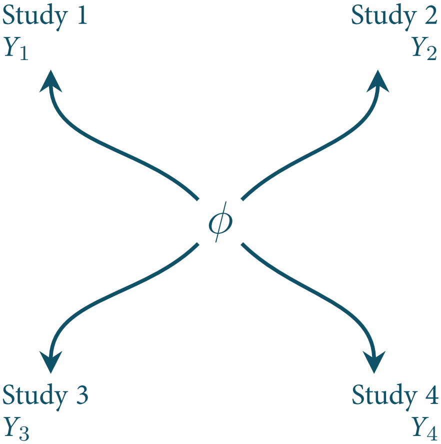
Study 1
Y_1

Study 2
Y_2

Study 3
Y_3

Study 4
Y_4

Motivation



Study 1 Y_1

Study 2 Y_2

ϕ

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

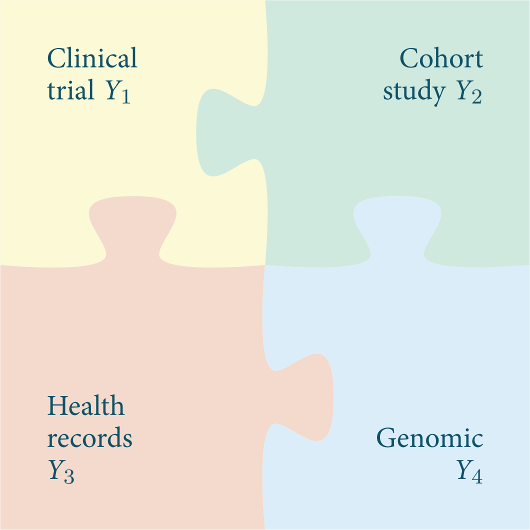
Clinical trial Y_1

Cohort study Y_2

Health records Y_3

Genomic Y_4

Motivation



Clinical trial Y_1

Cohort study Y_2

Health records Y_3

Genomic Y_4

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

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

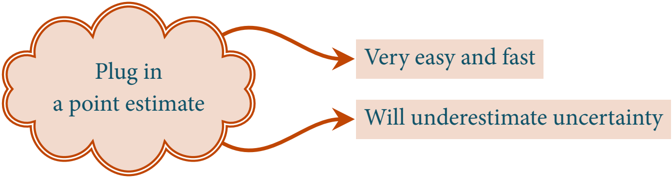
Genomic

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

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

Modular approaches

Modular approaches

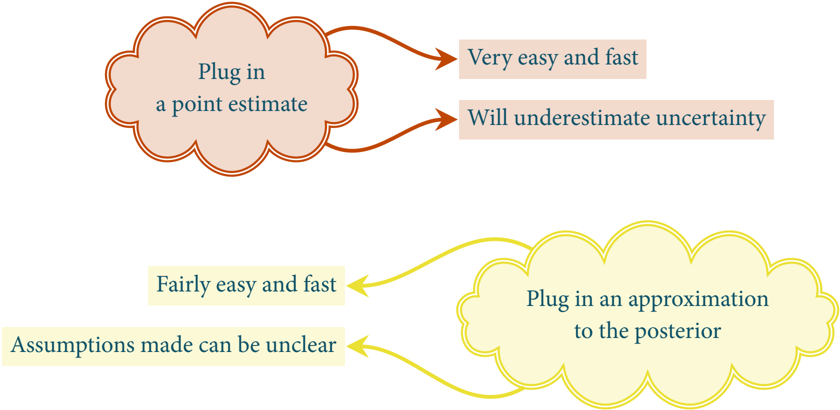


Plug in
a point estimate

Very easy and fast

Will underestimate uncertainty

Modular approaches



Plug in
a point estimate

Very easy and fast

Will underestimate uncertainty

Fairly easy and fast

Assumptions made can be unclear

Plug in an approximation
to the posterior

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

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

Observe:

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

Model:

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

A toy example Severity of influenza

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

Observe:

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

Model:

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

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

A toy example Severity of influenza

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

Observe:

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

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 symptoms.

Model for new data:

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

A toy example | Severity of influenza

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

Observe:

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

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 symptoms.

Model for new data:

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

If we assume q is the same in the original area, and we knew the total population size N,

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

A toy example Severity of influenza

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

Observe:

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

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 symptoms.

Model for new data:

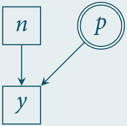
$$x \sim \text{Bin}(m, q) \qquad\qquad q \sim \text{Beta}(0.5, 5)$$

If we assume q is the same in the original area, and we knew the total population size N,

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

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

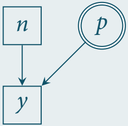
A toy example Graphical representation



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

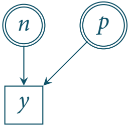
$$p \sim \mathrm{Beta}(1, 9)$$

A toy example Graphical representation



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

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

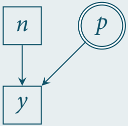


$$y \sim \text{Bin}(n, p)$$

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

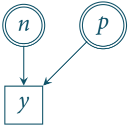
$$n \sim \text{Po}(1000)$$

A toy example | Graphical representation



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

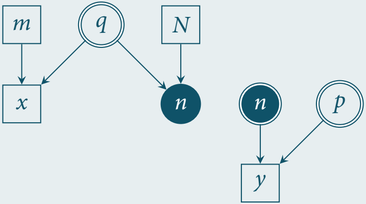
$$p \sim \mathrm{Beta}(1, 9)$$



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

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

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



$$x \sim \mathrm{Bin}(m, q)$$

$$q \sim \mathrm{Beta}(0.5, 5)$$

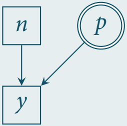
$$n \sim \mathrm{Bin}(N, q)$$

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

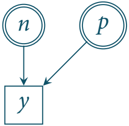
$$p \sim \mathrm{Beta}(1, 9)$$

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

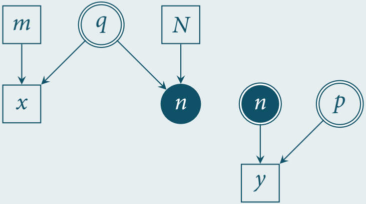
A toy example Graphical representation



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

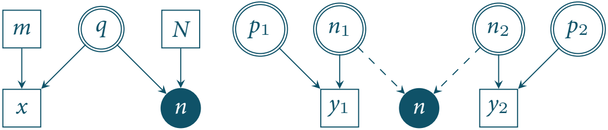


$$y \sim \text{Bin}(n, p)$$
$$p \sim \text{Beta}(1, 9)$$
$$n \sim \text{Po}(1000)$$



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

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



Aims of this work

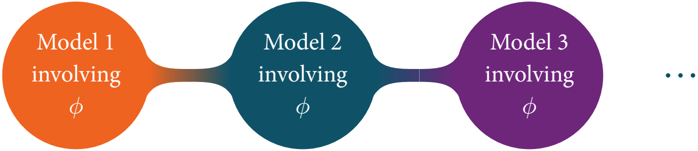
Model 1
involving
ϕ

Model 2
involving
ϕ

Model 3
involving
ϕ

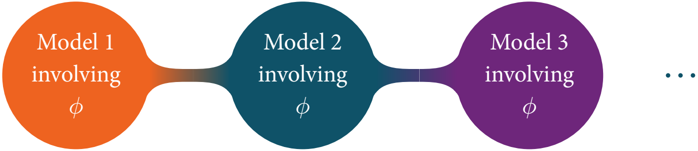
$\cdots$

Aims of this work



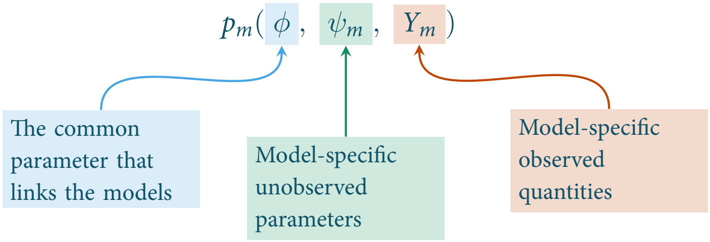
1. Create a generic method for joining submodels that share a common quantity ϕ into a single, joint model
    - Need to handle (implicitly) having two different priors for the same quantity
    - Need to handle models linked by non-invertible deterministic transformations

Aims of this work



1. Create a generic method for joining submodels that share a common quantity ϕ 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

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



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

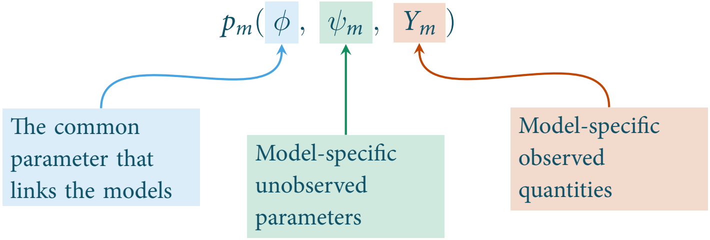
The common parameter that links the models

Model-specific unobserved parameters

Model-specific observed quantities

Notation

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

$$\downarrow$$

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 ϕ 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:

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

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

- $(\psi_m, Y_m) \perp\!\!\!\perp (\psi_\ell, Y_\ell) \mid \phi$ for $m \neq \ell$
- $p_{\text{comb}}(\psi_m, Y_m \mid \phi) = p_m(\psi_m, Y_m \mid \phi)$ for all m
- $p_{\text{comb}}(\phi, \psi_m, Y_m) = p_m(\phi, \psi_m, Y_m)$ for all m

---

Dawid and Lauritzen (1993). “Hyper Markov Laws in the Statistical Analysis of Decomposable Graphical Models”. *Annals of Statistics* **21**, 1272–1317.
Massa and Lauritzen (2010). “Combining statistical models”. In: *Contemporary Mathematics: Algebraic Methods in Statistics and Probability II*. ed. by Viana and Wynn, 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[1]

- $(\psi_m, Y_m) \perp\!\!\!\perp (\psi_\ell, Y_\ell) \mid \phi$ for $m \neq \ell$
- $p_{\text{meld}}(\psi_m, Y_m \mid \phi) = p_m(\psi_m, Y_m \mid \phi)$ for all m
- But $p_{\text{meld}}(\phi, \psi_m, Y_m) \neq p_m(\phi, \psi_m, Y_m)$ in general

---

[1] Goudie *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 *et al.* (2006). *Uncertain Judgements: Eliciting Experts' Probabilities*. Chichester: John Wiley & Sons.
Hinton (2002). "Training Products of Experts by Minimizing Contrastive Divergence.". *Neural computation* **14**, 1771–1800.

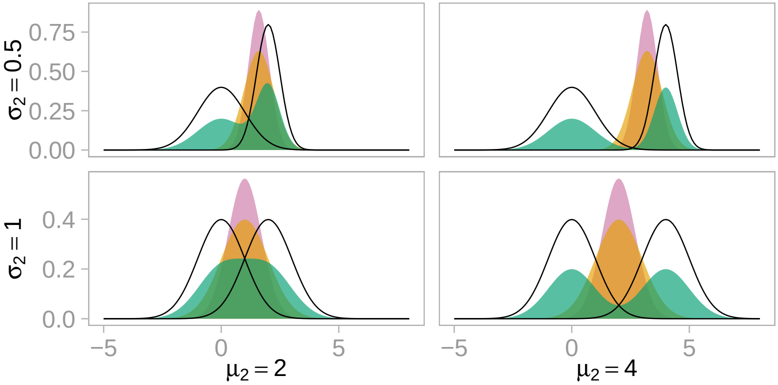


Figure: Pooling ($w_m = 0.5$) the black densities shown: $\mathrm{N}(0, 1)$ and $\mathrm{N}(\mu_2, \sigma_2^2)$; μ_2 and σ_2 shown in row/column labels.

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

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

Joining models 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

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

Joining models 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 and Zidek (1986). “Combining Probability Distributions: A Critique and an Annotated Bibliography”. *Statistical Science* **1**, 114–135.

Joining models Externally Bayesian pooling

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

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

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

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

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

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

Genest and Zidek (1986). “Combining Probability Distributions: A Critique and an Annotated Bibliography”. *Statistical Science* **1**, 114–135.

Joining models Bayesian melding

Poole and Raftery (2000) consider a deterministic model of the form:

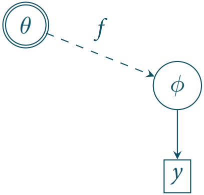
$$\phi = f(\theta)$$

Poole and Raftery (2000). "Inference for Deterministic Simulation Models: The Bayesian Melding Approach". *Journal of the American Statistical Association* **95**, 1244–1255.

Poole and Raftery (2000) consider a deterministic model of the form:

$$\phi = f(\theta)$$

Standard Bayesian

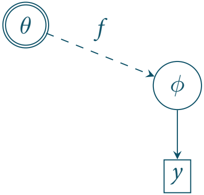


Poole and Raftery (2000). “Inference for Deterministic Simulation Models: The Bayesian Melding Approach”. *Journal of the American Statistical Association* **95**, 1244–1255.

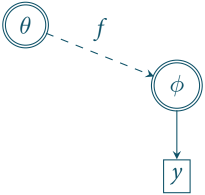
Poole and Raftery (2000) consider a deterministic model of the form:

$$\phi = f(\theta)$$

Standard Bayesian



Bayesian melding



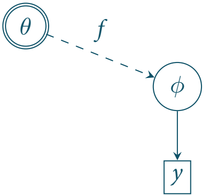
Poole and Raftery (2000). “Inference for Deterministic Simulation Models: The Bayesian Melding Approach”. *Journal of the American Statistical Association* **95**, 1244–1255.

Joining models Bayesian melding

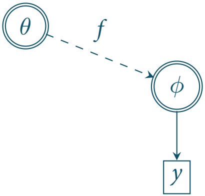
Poole and Raftery (2000) consider a deterministic model of the form:

$$\phi = f(\theta)$$

Standard Bayesian



Bayesian melding

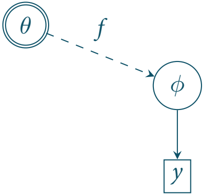


Poole and Raftery (2000). “Inference for Deterministic Simulation Models: The Bayesian Melding Approach”. *Journal of the American Statistical Association* **95**, 1244–1255.

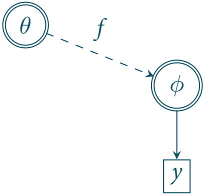
Poole and Raftery (2000) consider a deterministic model of the form:

$$\phi = f(\theta)$$

Standard Bayesian



Bayesian melding



- Extra information about the *output* of the deterministic function
- ϕ now has two different prior distributions
- Extend f to an invertible function, back-transform prior, then pool the two priors for θ

Poole and Raftery (2000). “Inference for Deterministic Simulation Models: The Bayesian Melding Approach”. *Journal of the American Statistical Association* **95**, 1244–1255.

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

---

Eddy *et al.* (1992). *Meta-Analysis by the Confidence Profile Method*. London: Academic Press.
Jackson *et al.* (2009). “Bayesian Graphical Models for Regression on Multiple Data Sets with Different Variables.”. *Biostatistics* **10**, 335–351.
Albert *et al.* (2011). “A Bayesian Evidence Synthesis for Estimating Campylobacteriosis Prevalence”. *Risk Analysis* **31**, 1141–1155.
Commenges and Hejblum (2012). “Evidence Synthesis through a Degradation Model Applied to Myocardial Infarction”. *Lifetime Data Analysis* **19**, 1–18.

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

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

e.g. Eddy *et al.* (1992), Jackson *et al.* (2009), Albert *et al.* (2011), Commenges and Hejblum (2012)

1. Obtain the posterior distribution $p_1(\phi, \psi_1 \mid y_1)$ under the first model
2. Approximate posterior marginal of ϕ under model 1 by $p_N(\phi \mid \widehat{\mu}, \widehat{\Sigma}) \approx p_1(\phi \mid y_1)$
3. 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 Products of Expert pooling.

Eddy *et al.* (1992). *Meta-Analysis by the Confidence Profile Method*. London: Academic Press.

Jackson *et al.* (2009). "Bayesian Graphical Models for Regression on Multiple Data Sets with Different Variables.". *Biostatistics* **10**, 335–351.

Albert *et al.* (2011). "A Bayesian Evidence Synthesis for Estimating Campylobacteriosis Prevalence". *Risk Analysis* **31**, 1141–1155.

Commenges and Hejblum (2012). "Evidence Synthesis through a Degradation Model Applied to Myocardial Infarction". *Lifetime Data Analysis* **19**, 1–18.

- Markov melding: aims to construct a suitable joint model

  Can be viewed as replacing prior marginal distributions

$$p_{\text{original}}(\phi, \psi, Y) \quad \rightarrow \quad p_{\text{meld}}(\phi, \psi, Y) = \frac{p_{\text{original}}(\phi, \psi, Y)}{p_{\text{original}}(\phi)} p_{\text{pool}}(\phi)$$

  then use the standard Bayesian posterior as our result

- Cut distributions: aims to avoid (some of) the consequences of a joint model

  Can be viewed as replacing posterior marginal distributions

$$\begin{aligned} & p_{\text{original}}(\phi, \psi \mid Y, Z) \\ & = p_{\text{original}}(\psi \mid Y, \phi) p_{\text{original}}(\phi \mid Y, Z) \quad \rightarrow \quad p_{\text{cut}}(\phi, \psi) = \frac{p_{\text{original}}(\phi, \psi \mid Y, Z)}{p_{\text{original}}(\phi \mid Y, Z)} p(\phi \mid Z) \end{aligned}$$

  and use that as our result.

Inference

Recall that the Markov melding model is

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

The joint posterior distribution p_{meld} is

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

p_{meld} is just a posterior distribution, so can target it directly with any standard method

Inference

But this is unappealing for the sorts of applications we consider because...

1. Pre-existing implementations of p_m, $m = 1, \ldots, M$
   - Often each implemented in a different language (C++, R, Python...) and/or a probabilistic programming language (Stan, JAGS, BUGS, Nimble, Turing, ...)
2. Fragility: each p_m may require "hand-holding"
   - Tuning parameters, initial conditions...
   - Cleaning erroneous data etc
3. Interest goes beyond p_{meld}
   - Sub/intermediate posteriors are of interest themselves: which submodel leads the overall posterior distribution to have some particular feature?
4. Submodels may not be fixed
   - May have several alternatives for p_ℓ some $\ell \in \{1, \ldots, M\}$
5. Also (ideally)...
   - data too large to fit on a single computer ('omics type data)
   - model too slow to fit on a single computer
   - siloed private data

Inference

We adopt the modular/two-stage/recursive computation scheme, proposed by several authors, which is more convenient

- Liang and Weiss (2007)
- Tom *et al.* (2010)
- Lunn *et al.* (2013)
- Hooten *et al.* (2019)

Liang and Weiss (2007). “A Hierarchical Semiparametric Regression Model for Combining HIV-1 Phylogenetic Analyses Using Iterative Reweighting Algorithms”. *Biometrics* **63**, 733–741.
Tom *et al.* (2010). “Reuse, Recycle, Reweigh: Combating Influenza through Efficient Sequential Bayesian Computation for Massive Data”. *The Annals of Applied Statistics* **4**, 1722–1748.
Lunn *et al.* (2013). “Fully Bayesian Hierarchical Modelling in Two Stages, with Application to Meta-Analysis.”. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **62**, 551–572.
Hooten *et al.* (2019). “Making Recursive Bayesian Inference Accessible”. *The American Statistician* **186**, 1–10.

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

$$\begin{aligned} r &= \frac{p_{\text{meld}}(\phi^\star, \psi_1^\star, \psi_2 \mid y_1, y_2)}{p_{\text{meld}}(\phi, \psi_1, \psi_2 \mid y_1, y_2)} \times \frac{q(\phi \mid \phi^\star)}{q(\phi^\star \mid \phi)} \\ &= \frac{p_{\text{pool}}(\phi^\star) \times \cancel{p_1(\phi^\star, \psi_1^\star, y_1)} p_1(\phi^\star)^{-1} \times p_2(\phi^\star, \psi_2, y_2) p_2(\phi^\star)^{-1}}{p_{\text{pool}}(\phi) \times \cancel{p_1(\phi, \psi_1, y_1)} p_1(\phi)^{-1} \times p_2(\phi, \psi_2, y_2) p_2(\phi)^{-1}} \times \frac{\cancel{p_1(\phi, \psi_1 \mid y_1)}}{\cancel{p_1(\phi^\star, \psi_1^\star \mid y_1)}} \\ &= \frac{p_{\text{pool}}(\phi^\star) S \times p_1(\phi^\star)^{-1} \times p_2(\phi^\star, \psi_2, y_2) p_2(\phi^\star)^{-1}}{p_{\text{pool}}(\phi) \times p_1(\phi)^{-1} \times p_2(\phi, \psi_2, y_2) p_2(\phi)^{-1}} \end{aligned}$$

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

   $$\begin{aligned} r &= \frac{p_{\text{meld}}(\phi^\star, \psi_1^\star, \psi_2 \mid y_1, y_2)}{p_{\text{meld}}(\phi, \psi_1, \psi_2 \mid y_1, y_2)} \times \frac{q(\phi \mid \phi^\star)}{q(\phi^\star \mid \phi)} \\ &= \frac{p_{\text{pool}}(\phi^\star) \times \cancel{p_1(\phi^\star, \psi_1^\star, y_1)} p_1(\phi^\star)^{-1} \times p_2(\phi^\star, \psi_2, y_2) p_2(\phi^\star)^{-1}}{p_{\text{pool}}(\phi) \times \cancel{p_1(\phi, \psi_1, y_1)} p_1(\phi)^{-1} \times p_2(\phi, \psi_2, y_2) p_2(\phi)^{-1}} \times \frac{\cancel{p_1(\phi, \psi_1 \mid y_1)}}{\cancel{p_1(\phi^\star, \psi_1^\star \mid y_1)}} \\ &= \frac{p_{\text{pool}}(\phi^\star) S \times p_1(\phi^\star)^{-1} \times p_2(\phi^\star, \psi_2, y_2) p_2(\phi^\star)^{-1}}{p_{\text{pool}}(\phi) \times p_1(\phi)^{-1} \times p_2(\phi, \psi_2, y_2) p_2(\phi)^{-1}} \end{aligned}$$

Extends naturally to an M-stage algorithm when $M > 2$

Inference

What is appealing about this approach?

1. Second stage requires only posterior samples from the first $p_1(\phi, \psi_1 \mid y_1)$
    - Any implementation of a Monte Carlo algorithm for p_1 will output this
2. Second stage acceptance probability does not involve $p_1(\phi, \psi_1, y_1)$
    - No need to code this in the second stage
3. Second stage can be implemented completely separately from stage one
    - No need to use the same programming language etc

What is not appealing....

- Samples from stage one should be independent
- If the region of posterior mass of p_{meld} is not a subregion of the high posterior support of p_1, then degeneracy-type problems
- etc...

Influenza A/H1N1 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 *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 A/H1N1 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 χ of intensive care unit (ICU), admissions for the A/H1N1 strain during the 2010/2011 influenza season in England

---

Presanis *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 A/H1N1 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

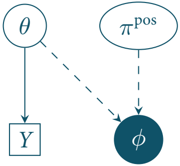
To estimate the total number χ of intensive care unit (ICU), admissions for the A/H1N1 strain during the 2010/2011 influenza season in England

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 *et al*. (2014). "Synthesising Evidence to Estimate Pandemic (2009) A/H1N1 Influenza Severity in 2009–2011". *The Annals of Applied Statistics* **8**, 2378–2403.

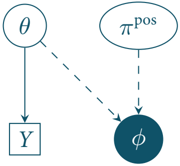


(a) Intensive Care Unit (ICU) model

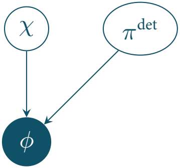
Model (a)

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

Influenza A/H1N1 example | Model structures



(a) Intensive Care Unit (ICU) model



(b) Severity model (simplified)

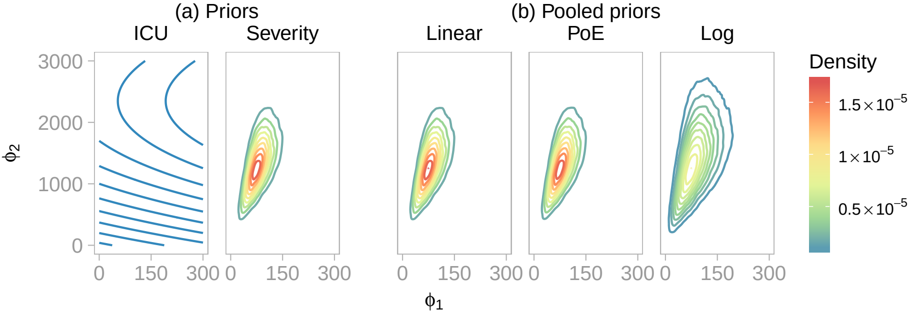
Figure: High-level DAGs

Model (a)

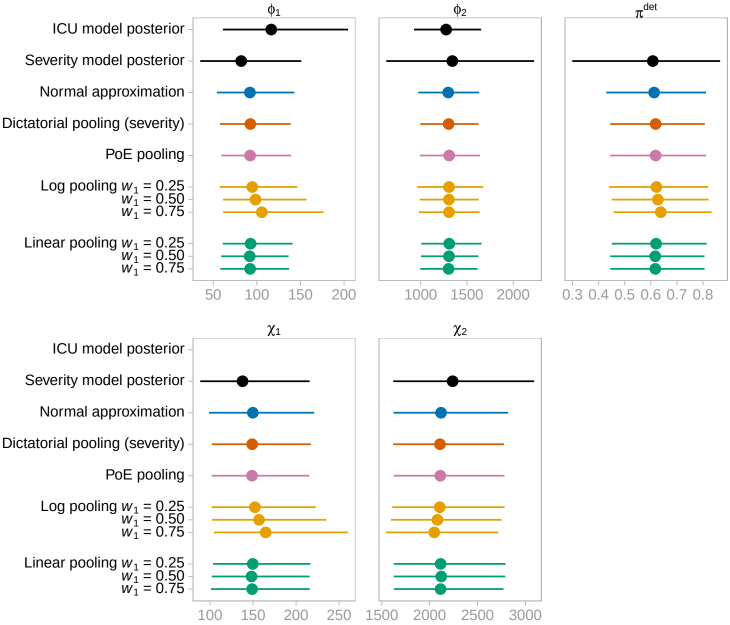
- Y is (weekly) data recording the number of suspected cases of A/H1N1 in ICUs
- π^{pos} is probability of suspected A/H1N1 being real, based on virological data (not shown).
- Given θ and π^{pos}, we estimate the confirmed number ϕ of cases of A/H1N1 in ICUs

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 χ that represents the other data sources


(a) Priors
ICU
Severity
(b) Pooled priors
Linear
PoE
Log
Density
1.5×10^{-5}
1×10^{-5}
0.5×10^{-5}
3000
2000
1000
0
ϕ_2
0
150
300
ϕ_1

Influenza A/H1N1 example Comparison of methods



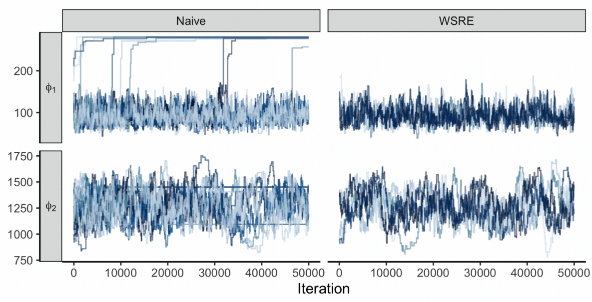
Inference Handling prior marginal distributions

The acceptance probability in the multi-stage algorithm for a proposal $\phi \rightarrow \phi^\star$ involves the self-density ratio of prior marginals $p_1(\phi)/p_1(\phi^\star)$

- The prior marginals $p_1(\phi) = \int p_1(\phi, \psi_1, Y_1) d\psi_1 \, dY_1$ are not usually tractable
- Originally we plugged in kernel density estimates from Monte Carlo samples

But can be unstable – underestimation in the tails of the denominator leads to an exploding self-density ratio estimate

- Sampling from (multiple) weighted marginals then using weighted KDE can help[2]



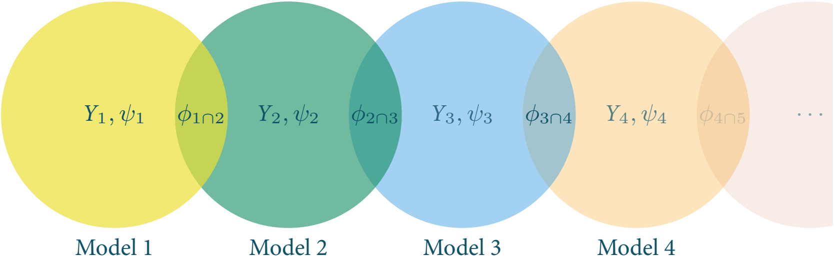
---

[2] Manderson and Goudie (2022). "A Numerically Stable Algorithm for Integrating Bayesian Models Using Markov Melding". *Statistics and Computing* **32**, 24.

Chained Markov melding

There is not always a single quantity ϕ shared between all submodels

For example, the submodels may form a “chain”, in which adjacent submodels have common parameters



Chained Markov melding Notation

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

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

Common parameter $\phi_m = (\phi_{m-1 \cap m}, \phi_{m \cap m+1})$ for $m = 2, \dots M - 1$

with $\phi_1 = \phi_{1 \cap 2}$ and $\phi_M = \phi_{M-1 \cap M}$

Model-specific unobserved parameters

Model-specific observed quantities

Chained Markov melding
Notation

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

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

Common parameter $\phi_m = (\phi_{m-1 \cap m}, \phi_{m \cap m+1})$ for $m = 2, \dots M - 1$

with $\phi_1 = \phi_{1 \cap 2}$ and $\phi_M = \phi_{M-1 \cap M}$

Model-specific unobserved parameters

Model-specific observed quantities

$$\downarrow$$

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

$$p(\phi_{1 \cap 2}, \dots, \phi_{M-1 \cap M}, \psi_1, \dots, \psi_M, Y_1, \dots, Y_M)$$

Chained Markov melding Definition

In the $M = 3$ case, we propose[3]:

$$p_{\text{meld}}(\phi_{1\cap 2}, \phi_{2\cap 3}, \psi_1, \psi_2, \psi_3, Y_1, Y_2, Y_3) =$$
$$p_{\text{pool}}(\phi_{1\cap 2}, \phi_{2\cap 3}) \frac{p_1(\phi_{1\cap 2}, \psi_1, Y_1)}{p_1(\phi_{1\cap 2})} \frac{p_2(\phi_{1\cap 2}, \phi_{2\cap 3}, \psi_2, Y_2)}{p_2(\phi_{1\cap 2}, \phi_{2\cap 3})} \frac{p_3(\phi_{2\cap 3}, \psi_3, Y_M)}{p_3(\phi_{2\cap 3})}$$

---

[3] Manderson and Goudie (in press). “Combining Chains of Bayesian Models with Markov Melding”. *Bayesian Analysis*.

Chained Markov melding Definition

In the $M = 3$ case, we propose[3]:

$$p_{\text{meld}}(\phi_{1\cap 2}, \phi_{2\cap 3}, \psi_1, \psi_2, \psi_3, Y_1, Y_2, Y_3) = \\ p_{\text{pool}}(\phi_{1\cap 2}, \phi_{2\cap 3}) \frac{p_1(\phi_{1\cap 2}, \psi_1, Y_1)}{p_1(\phi_{1\cap 2})} \frac{p_2(\phi_{1\cap 2}, \phi_{2\cap 3}, \psi_2, Y_2)}{p_2(\phi_{1\cap 2}, \phi_{2\cap 3})} \frac{p_3(\phi_{2\cap 3}, \psi_3, Y_M)}{p_3(\phi_{2\cap 3})}$$

Unless $p_2(\phi_{1\cap 2}, \phi_{2\cap 3}) = p_2(\phi_{1\cap 2})p_2(\phi_{2\cap 3})$, this is usually a different model to the model given by applying “common ϕ melding” twice i.e.

$$\text{meld}(\text{meld}(p_1, p_2), p_3)$$

---

[3] Manderson and Goudie (in press). “Combining Chains of Bayesian Models with Markov Melding”. *Bayesian Analysis*.

Chained Markov melding Definition

In the $M = 3$ case, we propose[3]:

$$p_{\text{meld}}(\phi_{1\cap 2}, \phi_{2\cap 3}, \psi_1, \psi_2, \psi_3, Y_1, Y_2, Y_3) =$$
$$p_{\text{pool}}(\phi_{1\cap 2}, \phi_{2\cap 3}) \frac{p_1(\phi_{1\cap 2}, \psi_1, Y_1)}{p_1(\phi_{1\cap 2})} \frac{p_2(\phi_{1\cap 2}, \phi_{2\cap 3}, \psi_2, Y_2)}{p_2(\phi_{1\cap 2}, \phi_{2\cap 3})} \frac{p_3(\phi_{2\cap 3}, \psi_3, Y_M)}{p_3(\phi_{2\cap 3})}$$

Unless $p_2(\phi_{1\cap 2}, \phi_{2\cap 3}) = p_2(\phi_{1\cap 2}) p_2(\phi_{2\cap 3})$, this is usually a different model to the model given by applying "common ϕ melding" twice i.e.

$$\text{meld}(\text{meld}(p_1, p_2), p_3)$$

Generalisation to a chain of M models is natural:

$$p_{\text{meld}}(\boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{Y}) = p_{\text{pool}}(\boldsymbol{\phi}) \frac{p_1(\phi_{1\cap 2}, \psi_1, Y_1)}{p_1(\phi_{1\cap 2})} \frac{p_M(\phi_{M-1\cap M}, \psi_M, Y_M)}{p_M(\phi_{M-1\cap M})}$$
$$\times \prod_{m=2}^{M-1} \left(\frac{p_m(\phi_{m-1\cap m}, \phi_{m\cap m+1}, \psi_m, Y_m)}{p_m(\phi_{m-1\cap m}, \phi_{m\cap m+1})} \right)$$

---

[3] Manderson and Goudie (in press). "Combining Chains of Bayesian Models with Markov Melding". *Bayesian Analysis*.

We form the pooled prior by combining the prior marginals of all common parameters

$$
\begin{aligned}
p_{\text{pool}}(\boldsymbol{\phi}) &= g(p_1(\phi_1), p_2(\phi_2), \ldots, p_M(\phi_M)) \\
&= g(p_1(\phi_{1\cap 2}),\ p_2(\phi_{1\cap 2}, \phi_{2\cap 3}),\ \ldots,\ p_2(\phi_{M-2\cap M-1}, \phi_{M-1\cap M}),\ p_M(\phi_{M-1\cap M})),
\end{aligned}
$$

- Logarithmic opinion pooling

$$
p_{\text{pool, log}}(\boldsymbol{\phi}) \propto \prod_{m=1}^{M} p_m(\phi_m)^{\lambda_m}
$$

  where $\lambda_1 = \cdots = \lambda_M = 1$ is a special case we call products-of-experts (PoE) pooling

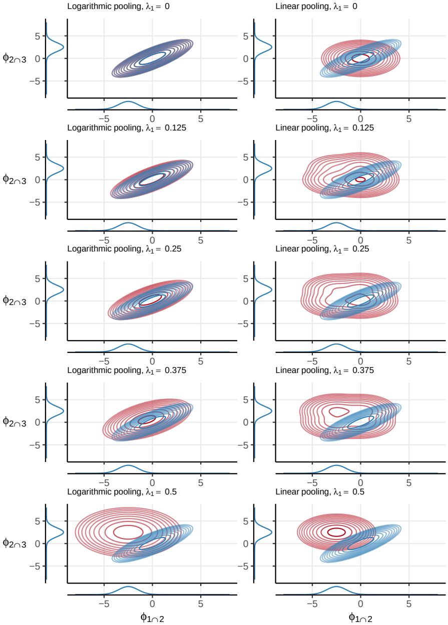
- Linear opinion pooling less obvious: $p_1(\phi_{1\cap 2}) + p_2(\phi_{1\cap 2}, \phi_{2\cap 3}) = ?$.
  Nearest analogue: pool marginals

$$
p_{\text{pool},m}(\phi_{m\cap m+1}) \ \propto\ \lambda_{m,1} p_m(\phi_{m\cap m+1}) + \lambda_{m,2} p_{m+1}(\phi_{m\cap m+1})
$$

  Then take product of marginals — which obviously induces prior independence.

- Dictatorial pooling
  Essentially choose one (of the two possible) priors for each common parameter $\phi_{m-1\cap m}$

Chained Markov melding Pooling examples



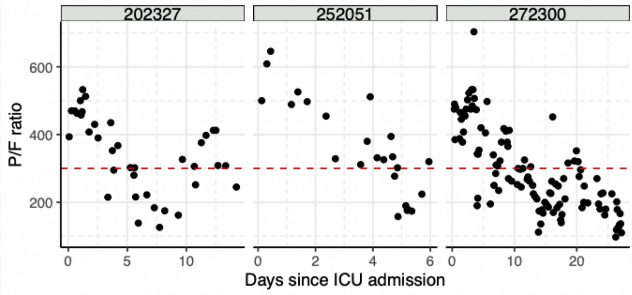
Interested in the time to onset of respiratory failure amongst patients in intensive care units (ICU), and factors that are associated with its onset

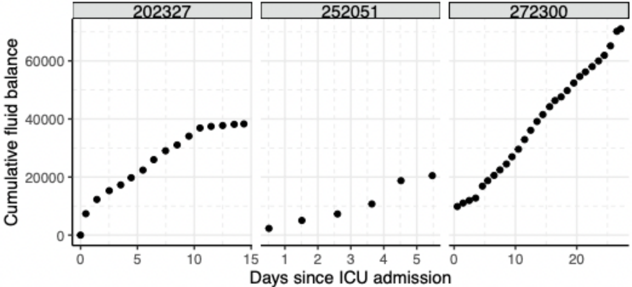
- Respiratory failure defined as the P/F ratio $\leq$ 300mmHg
  - P/F ratio is measured regularly, but is highly variable
  - The time to P/F ratio $\leq$ 300mmHg is thus quite uncertain
- Time to respiratory failure likely depends on various baseline covariates
  - Demographics, laboratory test results
- There is some also evidence for a relationship with rate of fluid intake
  - The rate can be estimated from cumulative fluid balance records

Aim: to integrate together 3 separate models that relate:

1. B-spline: model for P/F ratio data
2. Time-to-event: model for respiratory failure dependence on baseline covariates and fluid intake rate
3. Piecewise linear model: model for cumulative fluid balance

Uncertain-time-to-event example Raw data





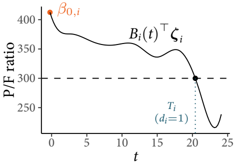
Uncertain-time-to-event example Submodels

Model 1: P/F ratio model

$$Y_1 = \beta_{0,i} + B_i(t_i)^\top \boldsymbol{\zeta}_i + \varepsilon_i.$$

with estimated time of respiratory failure

$$\boxed{\phi_{1\cap 2}} = T_i = \min_t \{\beta_{0,i} + B_i(t)\boldsymbol{\zeta}_i = 300\}$$

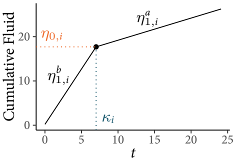


Model 2: Weibull time-to-event model for time $\boxed{\phi_{1\cap 2}}$, with hazard

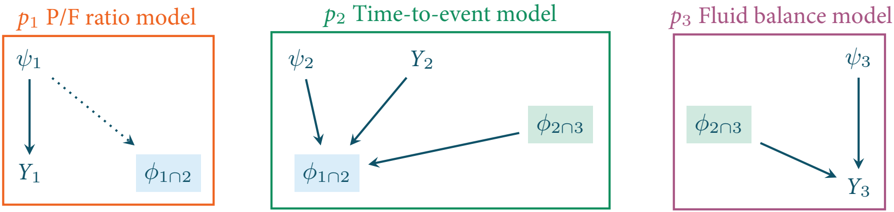
$$h_i(t) = \gamma t^{\gamma-1} \exp\left\{ \boldsymbol{w}_i^\top \boldsymbol{\theta} + \alpha \frac{\partial}{\partial t} \boxed{\phi_{2\cap 3}(t)} \right\}$$

Model 3: Cumulative fluid balance model, with mean:

$$\boxed{\phi_{2\cap 3}(t)} = m_i(t) = \eta_{0,i} + \eta^b_{1,i}(t - \kappa_i)\mathbf{1}_{\{t<\kappa_i\}} + \eta^a_{1,i}(t - \kappa_i)\mathbf{1}_{\{t\geq\kappa_i\}}$$



Uncertain-time-to-event example Model structure



p_1 P/F ratio model

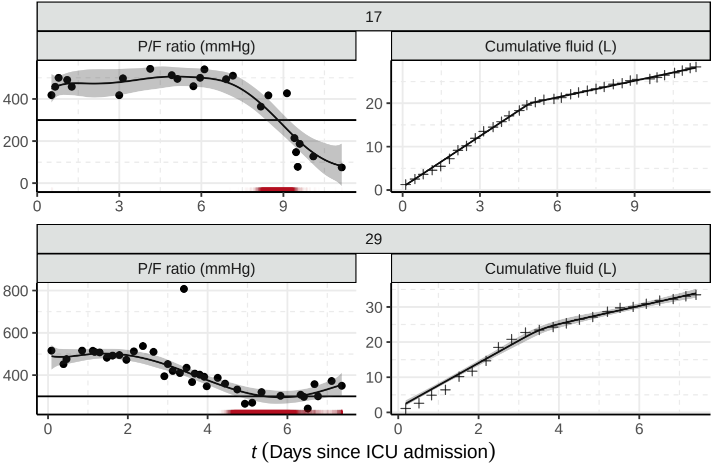
p_2 Time-to-event model

p_3 Fluid balance model

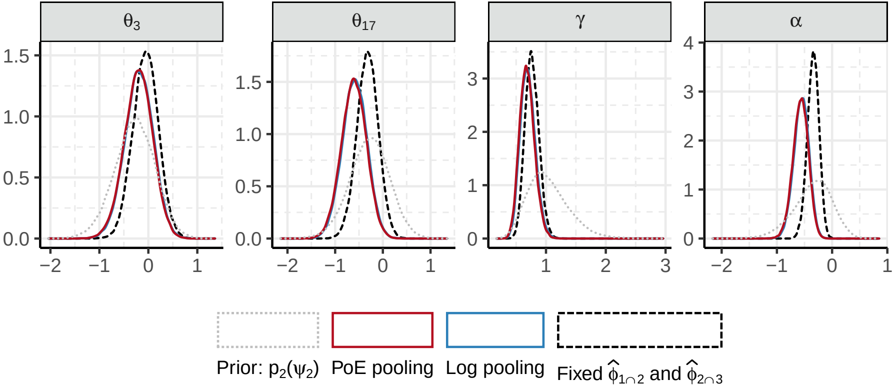
Will compare the results from

- Chained Markov melding (with different pooling functions)
- Plugging in point-estimates from models 1 and 3 into model 2

Uncertain-time-to-event example Model fit



Posterior density estimates for baseline coefficients (θ_3, θ_{17}), Weibull hazard rate γ and hazard-longitudinal association α



- Posterior insensitive to Logarithmic pooling vs Products of Experts (PoE) pooling here
- Disparity between chained Markov melding posterior and plugging-in fixed point estimates $\widehat{\phi}_{1\cap 2}$ and $\widehat{\phi}_{2\cap 3}$

With the Turing-RSS Health Data Lab in the UK, exploited these ideas to integrate several data sources relating to COVID-19

1. Randomised surveillance data (REACT study)
2. Targeted surveillance data from hospitals and (self-selected) wider community
3. Population meta-data (ethnicity, deprivation)
4. Commuter travel data

See Nicholson *et al*. (2022) for details. Argues that the pandemic...

> *...brought into focus a number of interesting challenges to conventional statistical practice arising, in particular, from the need to model real-time, messy data from diverse sources, in order to efficiently address rapidly evolving [...] demands. The dynamic nature [...] led to frequent changes in the specific questions being asked of the data, with focus often shifting unpredictably and suddenly.*

Argues that interoperable approaches → agility: ability to rapidly interlink and recycle statistical modelling outputs across analyses, with transferable components across problems

---

Nicholson *et al*. (2022). "Interoperability of Statistical Models in Pandemic Preparedness: Principles and Reality". *Statistical Science* **37**, 183–206.

Summary

Markov melding — joining models

- Provides a generic method for joining submodels
    - that share a common variable[4]
    - or that are linked in a chain-like structure[5]
- Incorporates the idea of pooling of prior marginal distributions
- Requires that there is not strong conflict

---

[4] Goudie *et al.* (2019). "Joining and Splitting Models with Markov Melding". *Bayesian Analysis* **14**, 81–109.

[5] Manderson and Goudie (in press). "Combining Chains of Bayesian Models with Markov Melding". *Bayesian Analysis*.

[6] Manderson and Goudie (2022). "A Numerically Stable Algorithm for Integrating Bayesian Models Using Markov Melding". *Statistics and Computing* **32**, 24.

Summary

Markov melding — joining models

- Provides a generic method for joining submodels
  - that share a common variable[4]
  - or that are linked in a chain-like structure[5]
- 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
- But can be unstable: weighted KDE can help[6]

---

[4] Goudie *et al.* (2019). "Joining and Splitting Models with Markov Melding". *Bayesian Analysis* **14**, 81–109.

[5] Manderson and Goudie (in press). "Combining Chains of Bayesian Models with Markov Melding". *Bayesian Analysis*.

[6] Manderson and Goudie (2022). "A Numerically Stable Algorithm for Integrating Bayesian Models Using Markov Melding". *Statistics and Computing* **32**, 24.

Future work

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

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

Feasible with low-dimensional common parameters + little conflict between models
Still a long way off this being straightforward in general

Funded PhD Opportunities

MRC
Biostatistics
Unit

UNIVERSITY OF
CAMBRIDGE

Interested in doing a PhD in Biostatistics, at the interface between methodological development and biomedical applications?

- Efficient Study Design
- Precision Medicine
- Causal Mechanisms
- Population Health
- Biostatistical Machine Learning

Multiple funded PhD studentships for available for October 2023

Applications for funding close on January 5th 2023. For more details see

`https://www.mrc-bsu.cam.ac.uk/training/phd/` • `phdstudy@mrc-bsu.cam.ac.uk`

Thanks

MRC Biostatistics Unit

UNIVERSITY OF CAMBRIDGE

Lorenz Wernisch
Anne Presanis
David Lunn
Daniela De Angelis

Andrew Manderson

Turing-RSS Health Data Lab
George Nicholson (Oxford)
Sylvia Richardson

Goudie et al (2019)

Manderson and Goudie (2022)
Manderson and Goudie (in press)

Nicholson et al (2022)

UKRI Medical Research Council (MRC) • The Alan Turing Institute under the UK EPSRC • NIHR Cambridge Biomedical Research Centre

---

Goudie *et al*. (2019). "Joining and Splitting Models with Markov Melding". *Bayesian Analysis* **14**, 81–109.
Manderson and Goudie (2022). "A Numerically Stable Algorithm for Integrating Bayesian Models Using Markov Melding". *Statistics and Computing* **32**, 24.
Manderson and Goudie (in press). "Combining Chains of Bayesian Models with Markov Melding". *Bayesian Analysis*.
Nicholson *et al*. (2022). "Interoperability of Statistical Models in Pandemic Preparedness: Principles and Reality". *Statistical Science* **37**, 183–206.