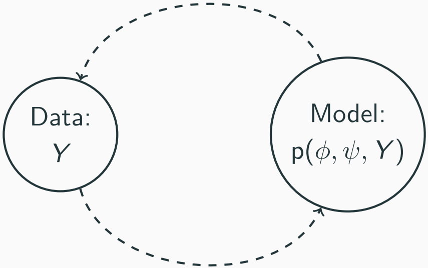
# Markov melding, multi-stage sampling, and self-density ratios

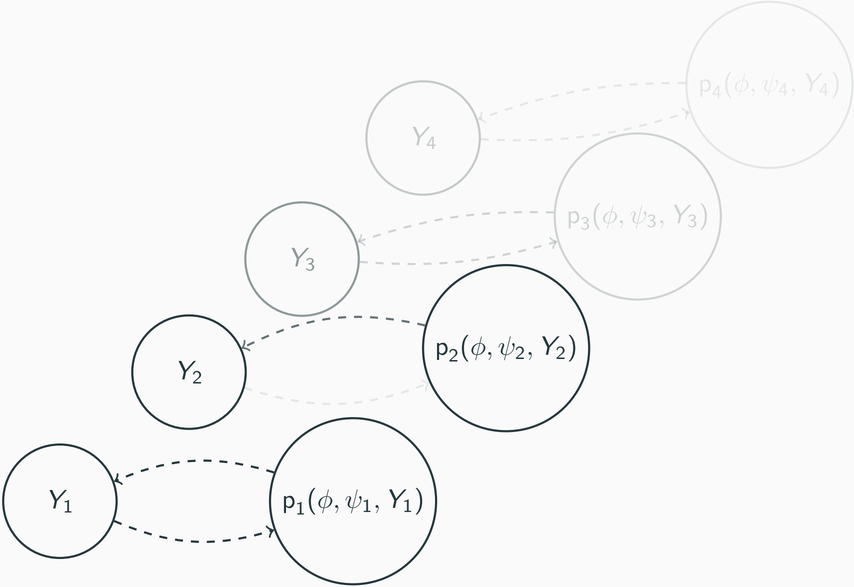
Andrew Manderson

2019-10-07

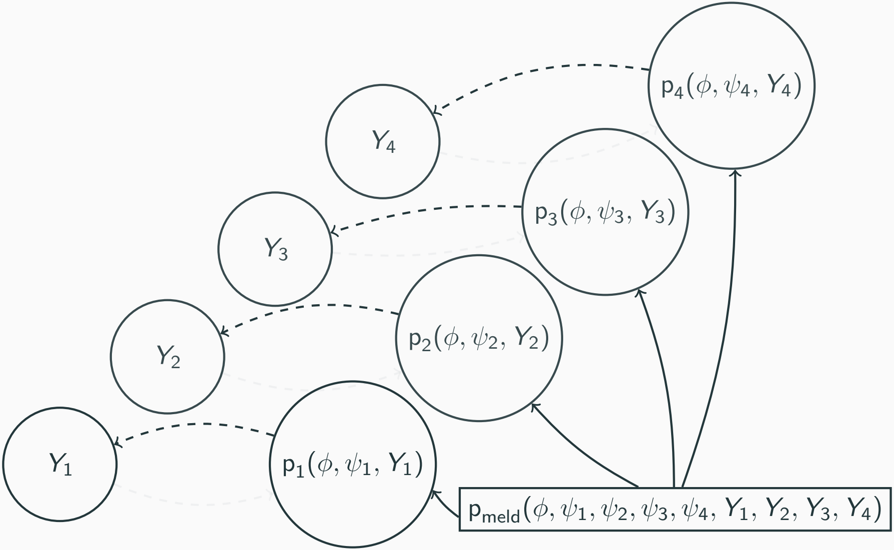
# Statistical modelling



# Modern statistical modelling



# Melding based statistical modelling?



# Markov melding

- Markov melding (Goudie et al. 2019) is a method for joining models with common component: $\phi$
- Ideally we would specify the generative model (For $M = 2$ submodels):

$$\mathrm{p}_{\mathrm{meld}}(\phi, \psi_1, \psi_2, Y_1, Y_2) = \mathrm{p}_{\mathrm{pool}}(\phi)\ \mathrm{p}_1(\psi_1, Y_1 \mid \phi)\ \mathrm{p}_2(\psi_2, Y_2 \mid \phi)$$

- Models developed in isolation, re-express as:

$$\mathrm{p}_{\mathrm{meld}}(\phi, \psi_1, \psi_2, Y_1, Y_2) = \mathrm{p}_{\mathrm{pool}}(\phi)\ \frac{\mathrm{p}_1(\phi, \psi_1, Y_1)}{\mathrm{p}_1(\phi)}\ \frac{\mathrm{p}_2(\phi, \psi_2, Y_2)}{\mathrm{p}_2(\phi)}$$

- Analytic form of $\mathrm{p}_m(\phi)$ often unknown
  - Can cause numerical issues in the model fitting process
  - Requires accurate estimation in low probability regions
  - We will revisit

# Multi-stage sampler

- Need to sample the melded posterior:

$$\begin{aligned} \mathrm{p}_{\mathrm{meld}}(\phi, \psi_1, \psi_2 \mid Y_1, Y_2) &\propto \mathrm{p}_{\mathrm{meld}}(\phi, \psi_1, \psi_2, Y_1, Y_2) \\ &= \mathrm{p}_{\mathrm{pool}}(\phi) \, \frac{\mathrm{p}_1(\phi, \psi_1, Y_1)}{\mathrm{p}_1(\phi)} \, \frac{\mathrm{p}_2(\phi, \psi_2, Y_2)}{\mathrm{p}_2(\phi)} \end{aligned}$$

- Submodels can be complex $\rightarrow$ sample in stages
  - Multi-stage sampling & approximations are common in practice
  - Staged sampling targets components of $\mathrm{p}_{\mathrm{meld}}(\phi, \psi_1, \psi_2 \mid Y_1, Y_2)$ in a cumulative manner
  - Can reuse other implementations of $\mathrm{p}_m(\phi, \psi_m, Y_m)$
  - Potentially more efficient

# Stage one acceptance probability

- Say we choose to target, $\mathrm{p}_{\mathrm{meld},1}(\phi, \psi_1 \mid Y_1)$:

$$\mathrm{p}_{\mathrm{meld}}(\phi, \psi_1, \psi_2 \mid Y_1, Y_2) \propto \mathrm{p}_{\mathrm{pool}}(\phi) \frac{\mathrm{p}_1(\phi, \psi_1, Y_1)}{\mathrm{p}_1(\phi)} \frac{\mathrm{p}_2(\phi, \psi_2, Y_2)}{\mathrm{p}_2(\phi)}$$

$$\rightarrow \qquad \mathrm{p}_{\mathrm{meld},1}(\phi, \psi_1 \mid Y_1)$$

- Given our stage one target, the stage one sampler has the following acceptance probability:

$$\alpha((\phi^*, \psi_1^*), (\phi, \psi_1)) = \frac{\mathrm{p}_1(\phi^*, \psi_1^*, Y_1)\mathrm{p}_1(\phi)\mathcal{Q}(\phi, \psi_1 \mid \phi^*, \psi_1^*)}{\mathrm{p}_1(\phi, \psi_1, Y_1)\mathrm{p}_1(\phi^*)\mathcal{Q}(\phi^*, \psi_1^* \mid \phi, \psi_1)}$$

- (asymptotically) Produces samples from the stage one target

## Stage two acceptance probability

- Stage two target, the melded posterior:

$$\mathrm{p}_{\mathrm{meld}}(\phi, \psi_1, \psi_2 \mid Y_1, Y_2) \;\propto\; \mathrm{p}_{\mathrm{pool}}(\phi) \, \frac{\mathrm{p}_1(\phi, \psi_1, Y_1)}{\mathrm{p}_1(\phi)} \, \frac{\mathrm{p}_2(\phi, \psi_2, Y_2)}{\mathrm{p}_2(\phi)}$$

- Use stage one samples of $\phi$ as the proposal distribution in a Gibbs update for $\phi \mid \psi$:

$$\alpha(\phi^*, \phi) = \frac{\mathrm{p}_{\mathrm{pool}}(\phi^*)\mathrm{p}_1(\phi^*, \psi_1, Y_1)\mathrm{p}_2(\phi^*, \psi_2, Y_2)\mathrm{p}_1(\phi)\mathrm{p}_2(\phi)}{\mathrm{p}_{\mathrm{pool}}(\phi)\mathrm{p}_1(\phi, \psi_1, Y_1)\mathrm{p}_2(\phi, \psi_2, Y_2)\mathrm{p}_1(\phi^*)\mathrm{p}_2(\phi^*)} \, \frac{\mathrm{p}_1(\phi, \psi_1, Y_1)\mathrm{p}_1(\phi^*)}{\mathrm{p}_1(\phi^*, \psi_1, Y_1)\mathrm{p}_1(\phi)}$$

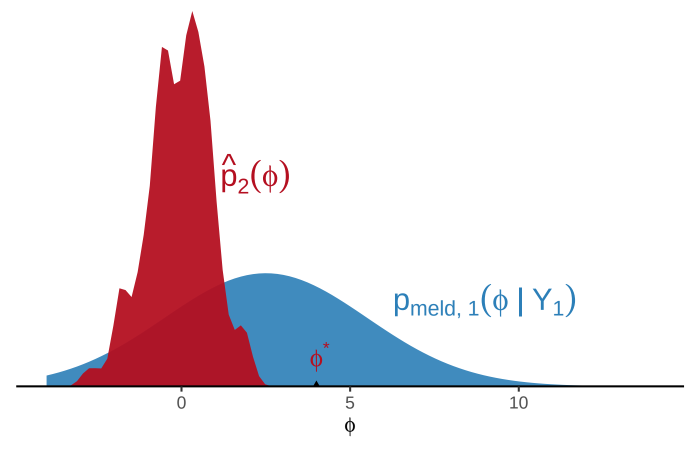
- Stage one terms cancel:

$$\alpha(\phi^*, \phi) = \frac{\mathrm{p}_{\mathrm{pool}}(\phi^*)\mathrm{p}_2(\phi^*, \psi_2, Y_2)\mathrm{p}_2(\phi)}{\mathrm{p}_{\mathrm{pool}}(\phi)\mathrm{p}_2(\phi, \psi_2, Y_2)\mathrm{p}_2(\phi^*)}$$

- $\mathrm{p}_2(\phi)$ not known analytically $\rightarrow$ substitute KDE: $\hat{\mathrm{p}}_2(\phi)$

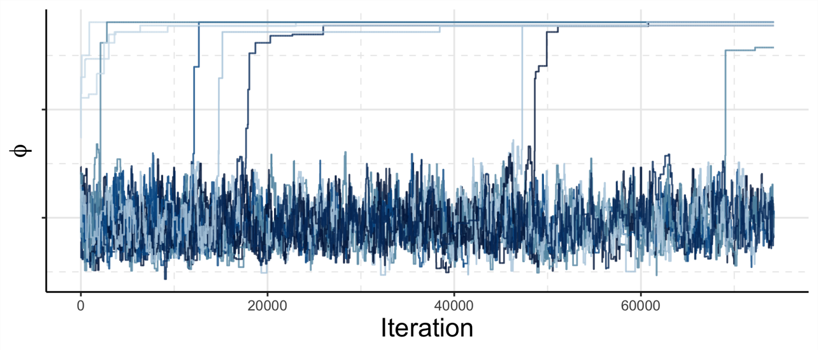
# Conflict

$$\alpha(\phi^*, \phi) = \frac{\mathrm{p}_{\mathrm{pool}}(\phi^*)\mathrm{p}_2(\phi^*, \psi_2, Y_2)}{\mathrm{p}_{\mathrm{pool}}(\phi)\mathrm{p}_2(\phi, \psi_2, Y_2)} \cdot \frac{\hat{\mathrm{p}}_2(\phi)}{\hat{\mathrm{p}}_2(\phi^*)}$$



# Conflict in action

$$\alpha(\phi^*, \phi) = \frac{\mathrm{p}_{\text{pool}}(\phi^*)\mathrm{p}_2(\phi^*, \psi_2, Y_2)}{\mathrm{p}_{\text{pool}}(\phi)\mathrm{p}_2(\phi, \psi_2, Y_2)} \cdot \frac{\hat{\mathrm{p}}_2(\phi)}{\hat{\mathrm{p}}_2(\phi^*)}$$




- H1N1 example from Goudie et al. (2019)

# Self density ratios

We only interact with unknown marginal via self-density ratio (Hiraoka, Hamada, and Hori 2018):

$$\mathrm{r}(\phi_{\mathrm{nu}}, \phi_{\mathrm{de}}) = \frac{\mathrm{p}(\phi_{\mathrm{nu}})}{\mathrm{p}(\phi_{\mathrm{de}})}$$

*Weighted-sample self-density ratio estimation* (WSRE) Intuition:

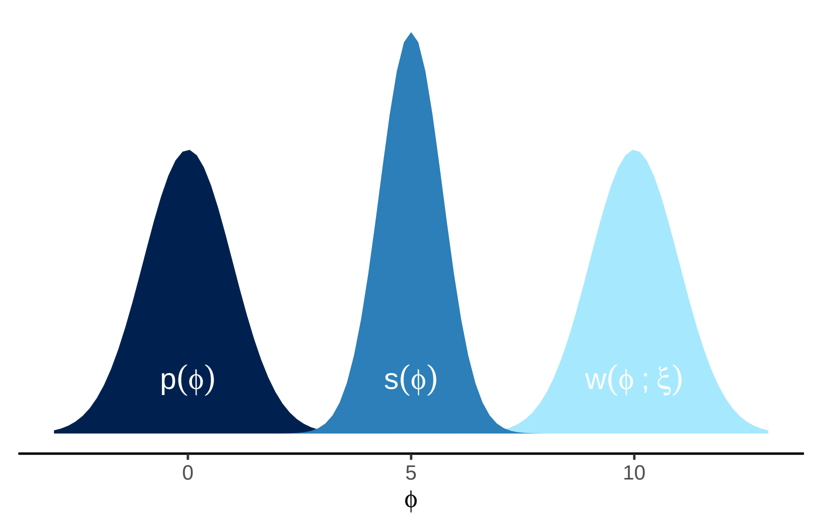
- Sampling $\phi \sim \mathrm{p}(\phi, \psi, Y)$ admits $\phi \sim \mathrm{p}(\phi)$
- Instead, sample:
  $\phi \sim \mathrm{p}(\phi, \psi, Y)\mathrm{w}(\phi; \xi) \quad \rightarrow \quad \phi \sim \frac{1}{Z}\mathrm{p}(\phi)\mathrm{w}(\phi; \xi) = \mathrm{s}(\phi)$
- Use the weighted sample density estimator of Jones (1991):

$$\hat{\mathrm{p}}(\phi) = \frac{1}{ZNh} \sum_{i=1}^{N} \mathrm{w}(\phi; \xi)^{-1} \mathrm{K}_h(\phi - \phi_i)$$

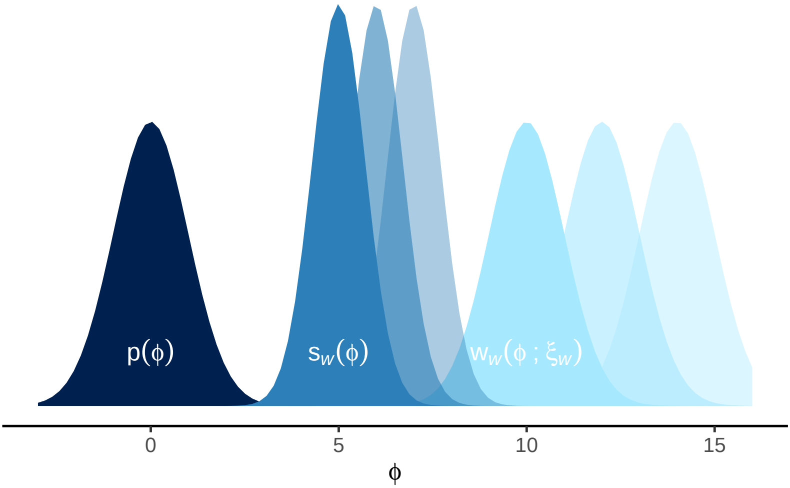
- Ratio estimator is then:

$$\hat{\mathrm{r}}(\phi_{\mathrm{nu}}, \phi_{\mathrm{de}}) = \frac{\hat{\mathrm{p}}(\phi_{\mathrm{nu}})}{\hat{\mathrm{p}}(\phi_{\mathrm{de}})}$$

# WSRE intuition - 1

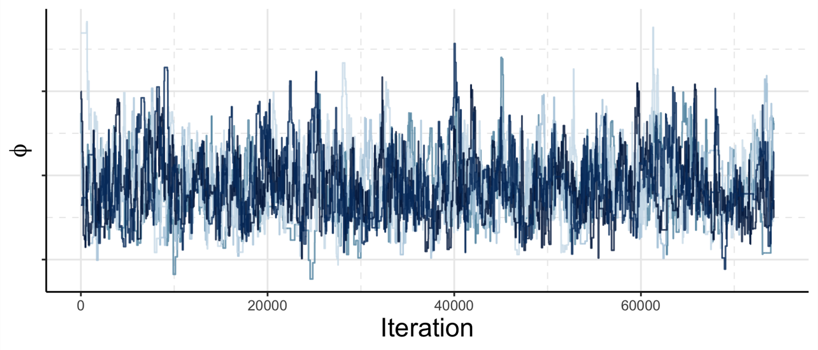


# WSRE intuition - $W$



# H1N1 example - 2

$$\alpha(\phi^*, \phi) = \frac{\mathrm{p}_{\mathrm{pool}}(\phi^*)\mathrm{p}_2(\phi^*, \psi_2, Y_2)}{\mathrm{p}_{\mathrm{pool}}(\phi)\mathrm{p}_2(\phi, \psi_2, Y_2)} \cdot \hat{\mathrm{r}}(\phi, \phi^*)$$



# Future plans

- Markov melding for ICU delirium, using electronic health record data
  - Observational data (Temperature, respiratory rate)
  - Blood test data
  - Daily ICU drug dose data
  - Each source is idiosyncratic $\rightarrow$ requires substantial modelling effort (and domain expertise)
- Melding for multiple $\phi$: $\mathrm{p}_1(\phi_1, Y_1)$ $\mathrm{p}_2(\phi_1, \phi_2, Y_2)$ $\mathrm{p}_3(\phi_2, Y_3)$
- Addressing other kinds of internal difference in scale and location (conflict)

# Links

Should you also wish to estimate self-density ratios:

- https://github.com/hhau/wsre

The traceplots are from this example:

- https://github.com/hhau/full-melding-example

This talk:

- https://github.com/hhau/bsu-together-20mins

# References

Goudie, Robert J. B., Anne M. Presanis, David Lunn, Daniela De Angelis, and Lorenz Wernisch. 2019. "Joining and Splitting Models with Markov Melding." *Bayesian Anal.* 14 (1). International Society for Bayesian Analysis: 81–109. doi:10.1214/18-BA1104.

Hiraoka, Kazuyuki, Toshihiko Hamada, and Gen Hori. 2018. "Necessary and Sufficient Conditions of Proper Estimators based on Self Density Ratio for Unnormalized Statistical Models." *Neural Networks* 98: 263–70. doi:10.1016/j.neunet.2017.11.018.

Jones, M. C. 1991. "Kernel Density Estimation for Length Biased Data." *Biometrika* 78 (3). Oxford University Press, Biometrika Trust: 511–19. http://www.jstor.org/stable/2337020.