Combining multiple sources of information with Markov melding

Examples, challenges, and (partial) solutions

Andrew Manderson

2020-01-28

MRC Biostatistics Unit, University of Cambridge & The Alan Turing Institute.

Model criticism - 1

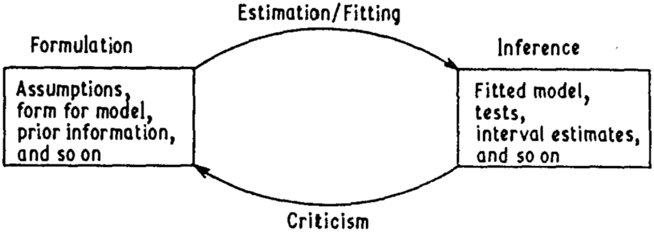


Figure 1.2.1 Schematic outline of model fitting. Adapted from Box (1979, 1980)

- From Cook and Weisberg (1982), but really from Box (1979; 1980)

Model criticism - 2

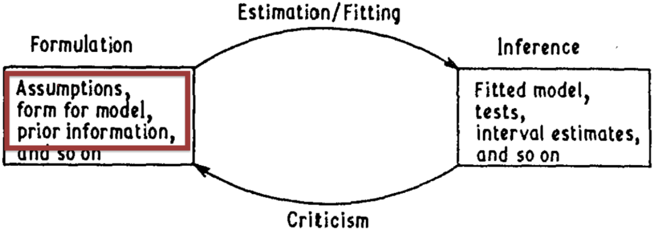


Figure 1.2.1 Schematic outline of model fitting. Adapted from Box (1979, 1980)

- From Cook and Weisberg (1982), but really from Box (1979; 1980)
- Non-linear increase in difficulty when $n, p \to \infty$

A complex applied analysis

The Annals of Applied Statistics
2014, Vol. 8, No. 4, 2378–2403
DOI: 10.1214/14-AOAS775

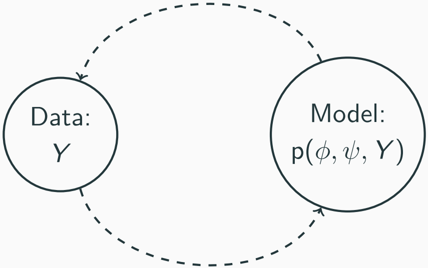
SYNTHESISING EVIDENCE TO ESTIMATE PANDEMIC (2009) A/H1N1 INFLUENZA SEVERITY IN 2009–2011[1]

BY ANNE M. PRESANIS*, RICHARD G. PEBODY†, PAUL J. BIRRELL*, BRIAN D. M. TOM*, HELEN K. GREEN†, HAYLEY DURNALL‡, DOUGLAS FLEMING‡ AND DANIELA DE ANGELIS*

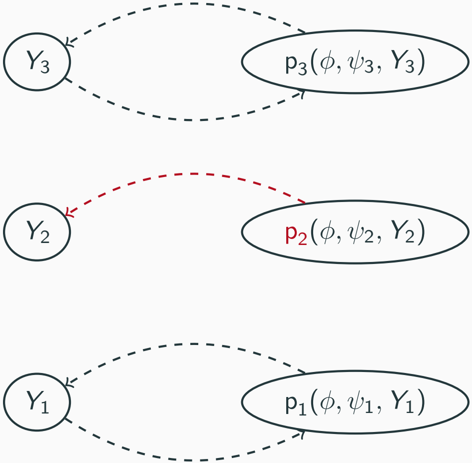
Medical Research Council Biostatistics Unit, Public Health England† and Royal College of General Practitioners‡*

- 5 sources of data, each idiosyncratic.
  - Some known to be biased in particular directions
  - Different levels of aggregation, national / hospital specific data
- Specifying and critiquing a joint model for all of this?
- Will return to a simplified version

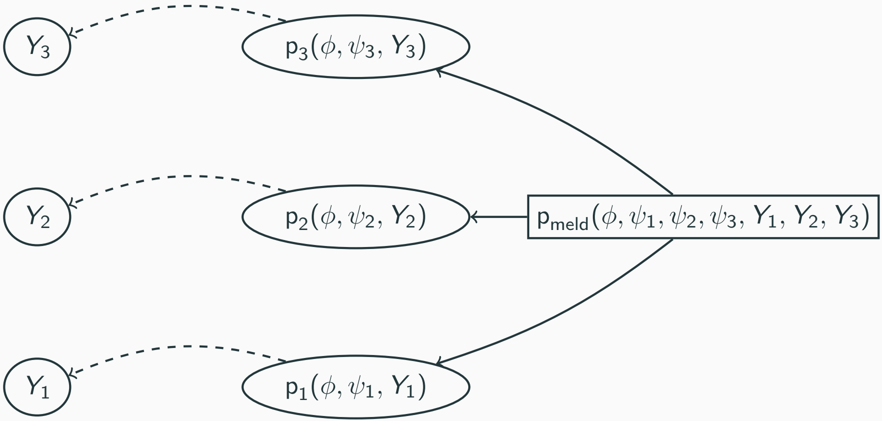
Statistical modelling



Statistical submodelling



Melding based statistical modelling?



Specification: Markov melding

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

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

- Submodels 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

Specification: Notes on the pooled prior

$\text{p}_{\text{pool}}(\phi)$ addresses two issues:

1. Necessary because we have two priors for ϕ (Poole and Raftery 2000)

2. $\text{p}_{\text{pool}}(\phi)$ should be appropriate for ϕ under the melded model *and* both submodels
   - Suggests defining $\text{p}_{\text{pool}}(\phi) = g(\text{p}_1(\phi), \text{p}_2(\phi))$
   - Similar problem encountered when experts encode their opinions as priors (O'Hagan et al. 2006)
     - **Log**: $\text{p}_{\text{pool}}(\phi) = \text{p}_1(\phi)^{w_1} \text{p}_2(\phi)^{w_2}$
     - Linear: $\text{p}_{\text{pool}}(\phi) = w_1 \text{p}_1(\phi) + w_2 \text{p}_2(\phi)$
     - Product-of-experts: $\text{p}_{\text{pool}}(\phi) = \text{p}_1(\phi)\text{p}_2(\phi)$ (Hinton 2002)
     - Dictatorial: $\text{p}_{\text{pool}}(\phi) = \text{p}_m(\phi)$ for an $m \in \{1, 2\}$

Estimation: 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

Estimation: Stage one acceptance probability

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

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

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

- 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

Estimation: Stage two acceptance probability

- Stage two target, the melded posterior:

$$\mathrm{p_{meld}}(\phi, \psi_1, \psi_2 \mid Y_1, Y_2) \propto \mathrm{p_{pool}}(\phi) \frac{\mathrm{p_1}(\phi, \psi_1, Y_1)}{\mathrm{p_1}(\phi)} \frac{\mathrm{p_2}(\phi, \psi_2, Y_2)}{\mathrm{p_2}(\phi)}$$

- Use stage one samples of ϕ, ψ_1 as the proposal distribution in a Metropolis-within-Gibbs update for $\phi, \psi_1 \mid \psi_2$:

$$\alpha((\phi^*, \psi_1^*), (\phi, \psi_1)) =$$

$$\frac{\mathrm{p_{pool}}(\phi^*)\mathrm{p_1}(\phi^*, \psi_1^*, Y_1)\mathrm{p_2}(\phi^*, \psi_2, Y_2)\mathrm{p_1}(\phi)\mathrm{p_2}(\phi)}{\mathrm{p_{pool}}(\phi)\mathrm{p_1}(\phi, \psi_1, Y_1)\mathrm{p_2}(\phi, \psi_2, Y_2)\mathrm{p_1}(\phi^*)\mathrm{p_2}(\phi^*)} \frac{\mathrm{p_1}(\phi, \psi_1, Y_1)\mathrm{p_1}(\phi^*)}{\mathrm{p_1}(\phi^*, \psi_1^*, Y_1)\mathrm{p_1}(\phi)}$$

- Stage one terms cancel:

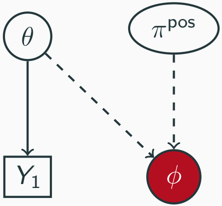
$$\alpha(\phi^*, \phi) = \frac{\mathrm{p_{pool}}(\phi^*)\mathrm{p_2}(\phi^*, \psi_2, Y_2)\mathrm{p_2}(\phi)}{\mathrm{p_{pool}}(\phi)\mathrm{p_2}(\phi, \psi_2, Y_2)\mathrm{p_2}(\phi^*)}$$

- Generic Metropolis-Hastings update for $\psi_2 \mid \phi, \psi_1$

Example: Simplified version of Presanis et. al. (2014)

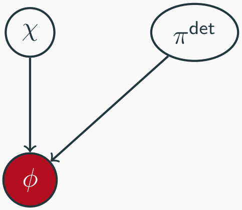
- We will now consider a simplified version of Presanis et. al. (2014)
- Submodel details are not overly important for our purpose
- Consider 2 submodels:
  1. Models number of patients in the ICU with H1N1
  2. Collapses the other 4 data sources into an informative prior structure

Example: Submodel 1 - ICU submodel



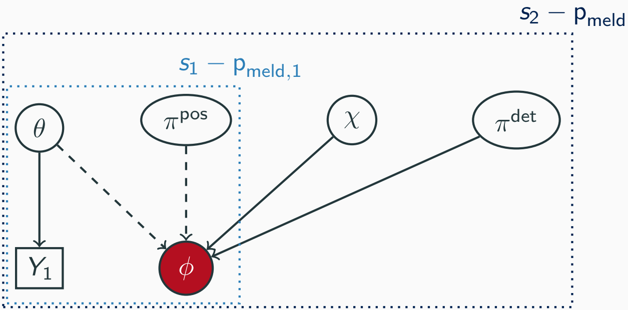
- Notation: $\psi_1 = (\theta, \pi^{\text{pos}})$
- $\text{p}_1(\theta, Y)$ is a thinned Poisson process as an observational model for Y
  - Can derive total number of people in ICU with influenza
- π^{pos} derived from virological data
  - Informs the proportion of people with influenza who are positive for H1N1
- Submodel deterministically defines $\phi = f(\theta, \pi^{\text{pos}})$.

Example: Submodel 2 - Simplified severity model



- Notation: $\psi_2 = (\chi, \pi^{\text{det}})$
- $\phi \sim \text{Bin}(\chi, \pi_{\text{det}})$ to account for known underestimation of ϕ
- Informative prior for χ summarises other components of severity model

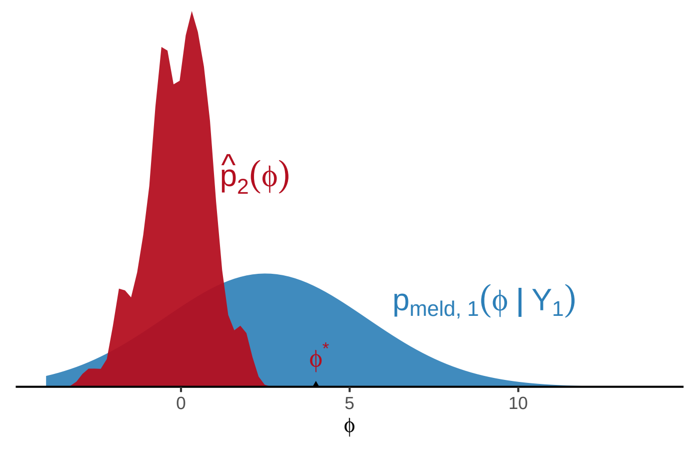
Example: Multi-stage sampler - in DAG form



- After stage 2 (s_2), every node's distribution incorporates information from Y_1 and χ.

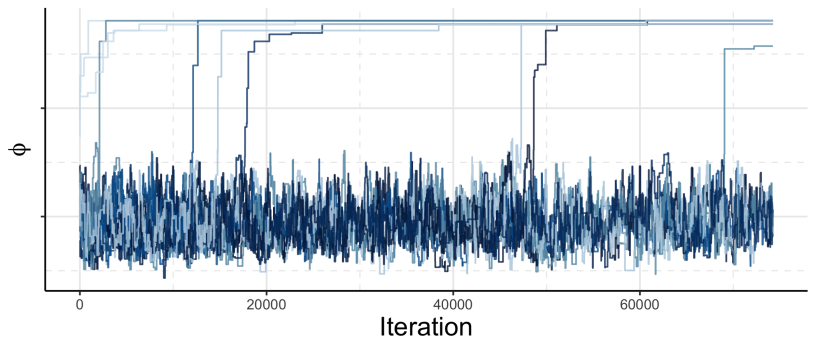
Issue 1: Things don't always work!

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



Issue 1: KDE error in action

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



Method 1: 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) (Manderson and Goudie 2020) Intuition:

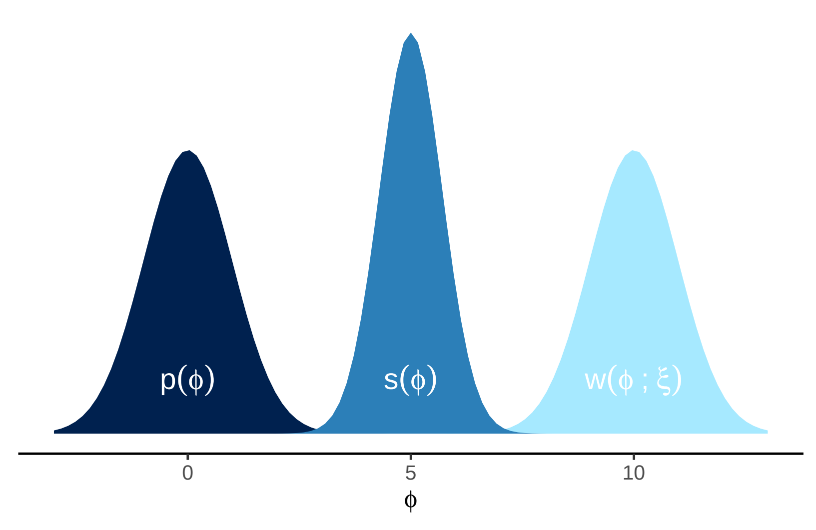
- Sampling $\phi \sim \mathrm{p}(\phi, \psi, Y)$ admits $\phi \sim \mathrm{p}(\phi)$
- Instead, sample:
  $\phi \sim \mathrm{p}(\phi, \psi, Y)\mathrm{w}(\phi; \xi) \quad \rightarrow \quad \phi \sim \frac{1}{Z}\mathrm{p}(\phi)\mathrm{w}(\phi; \xi) = \mathrm{s}(\phi)$
- Use the weighted sample density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

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

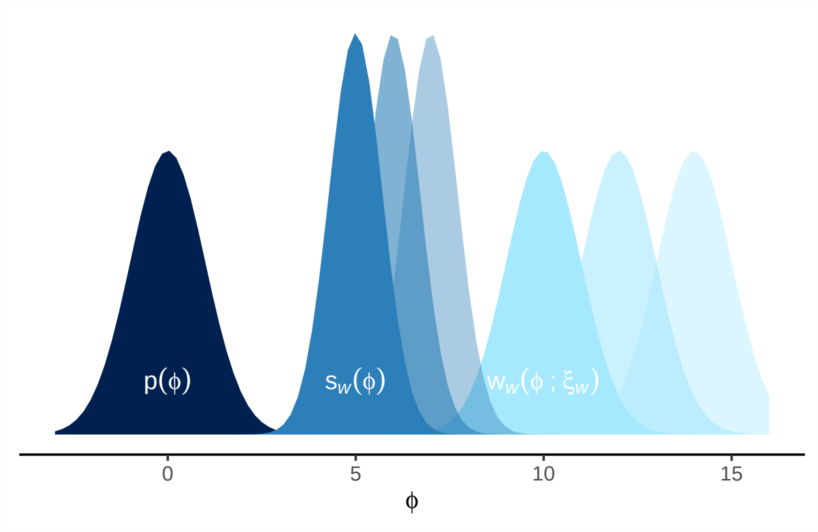
- Ratio estimator is then:

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

Method 1: WSRE intuition - 1

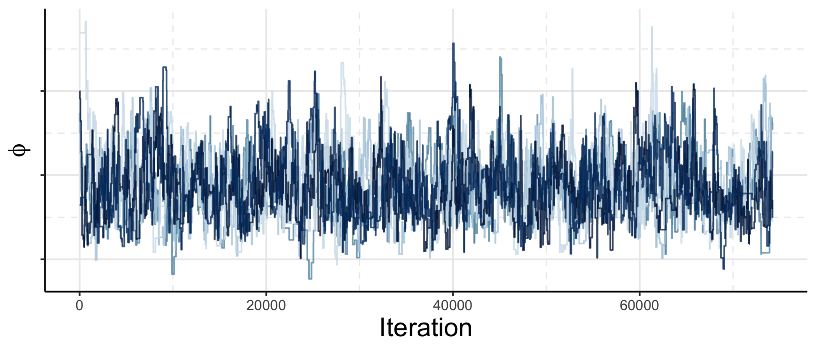


Method 1: WSRE intuition - W

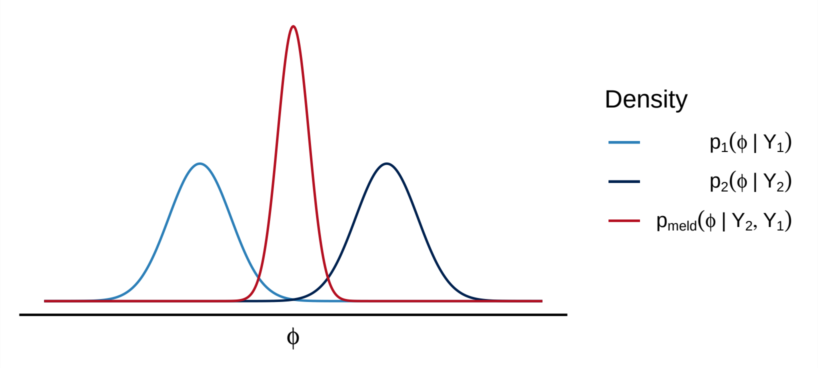


Method 1: H1N1 example - 2

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

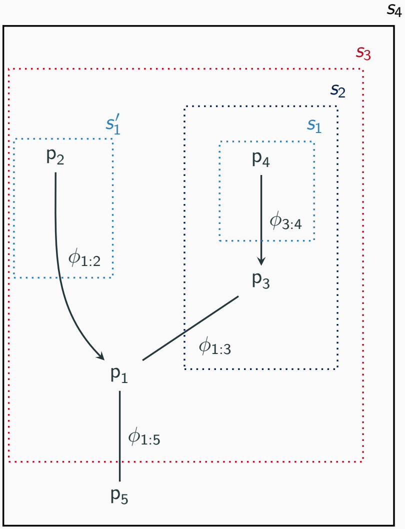


Issue 2: Submodel conflict



- Last issue arose (primarily) due to differences in scale
- What about more general model conflict?
  - There is a philosophical issue here
- *Partial solution*: Automatic, multiple importance sampling (Paananen et al. 2019)

***Issue 3*: Multiple ϕ?**



Conclusion

45

A trio of inference problems that could win you a Nobel Prize in statistics (if you help fund it)

Xiao-Li Meng
Department of Statistics
Harvard University, Cambridge, MA

- Multi-source inference is extremely difficult in practice (Meng 2014) (along with multi-resolution and multi-phase)

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:

- **TODO: Push this branch to github / link to it**

Box, George E. P. 1979. "Robustness in the Strategy of Scientific Model Building." In *Robustness in Statistics*, edited by Robert L. Launer and Graham N. Wilkinson, 201–36. Academic Press. doi:https://doi.org/10.1016/B978-0-12-438150-6.50018-2.

———. 1980. "Sampling and Bayes' Inference in Scientific Modelling and Robustness." *Journal of the Royal Statistical Society. Series A (General)* 143 (4). Wiley: 383–430. http://www.jstor.org/stable/2982063.

Cook, R.D., and S. Weisberg. 1982. *Residuals and Influence in Regression*. Monographs on Statistics and Applied Probability. Chapman & Hall.

Goudie, Robert J. B., Anne M. Presanis, David Lunn, Daniela De Angelis, and Lorenz Wernisch. 2019. "Joining and Splitting Models with Markov Melding." *Bayesian Analysis* 14 (1). International Society for Bayesian

Analysis: 81–109. doi:10.1214/18-BA1104.

Hinton, Geoffrey E. 2002. "Training Products of Experts by Minimizing Contrastive Divergence." *Neural Computation* 14 (8): 1771–1800. doi:10.1162/089976602760128018.

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.

Manderson, Andrew A., and Robert J. B. Goudie. 2020. "A Numerically Stable Algorithm for Integrating Bayesian Models Using Markov Melding."

arXiv E-Prints, arXiv:2001.08038.

Meng, Xiao-Li. 2014. "A Trio of Inference Problems That Could Win You a Nobel Prize in Statistics (If You Help Fund It)." CRC Press. http://nrs.harvard.edu/urn-3:HUL.InstRepos:11718181.

O'Hagan, A., C.E. Buck, A. Daneshkhah, J.R. Eiser, P.H. Garthwaite, D.J. Jenkinson, J.E. Oakley, and T. Rakow. 2006. *Uncertain Judgements: Eliciting Experts' Probabilities*. Statistics in Practice. Wiley. doi:10.1002/0470033312.

Paananen, Topi, Juho Piironen, Paul-Christian Bürkner, and Aki Vehtari. 2019. "Pushing the Limits of Importance Sampling Through Iterative Moment Matching." *arXiv E-Prints*, arXiv:1906.08850.

Poole, David, and Adrian E. Raftery. 2000. "Inference for Deterministic Simulation Models: The Bayesian Melding Approach." *Journal of the*

References iv

American Statistical Association 95 (452). Taylor & Francis: 1244–55. doi:10.1080/01621459.2000.10474324.