Partial Bayes: Exact Inference with Partially Specified Bayesian Models

June 28, 2017 @ ICSA Symposium

Yixuan Qiu
Purdue University

Joint work with Prof. Lingsong Zhang and Chuanhai Liu

Outline

Motivation

Our Approach

Models and Simulations

Summary and Discussion

Outline

Motivation

Our Approach

Models and Simulations

Summary and Discussion

Partial Prior Problems

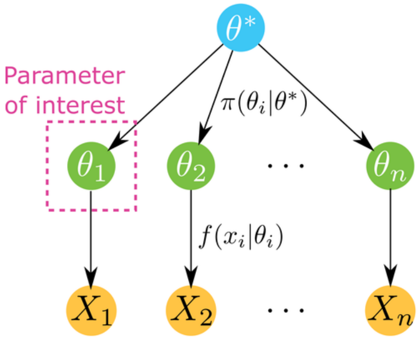
- Partial prior problems can be formulated as a partially specified Bayesian model
- One typical case is the Bayesian hierarchical model with unknown hyper-parameters

Example: Bayesian Hierarchical Model (Efron, 2010)

- $X_i|\mu_i \sim N(\mu_i, 1), i = 1, \ldots, K$ where X_i's are independent observed data, and μ_i's are unknown individual means
- $\mu_i \sim N(0, \tau^2)$ is the common prior, where τ^2 is unknown
- Want to make inference about μ_i

PURDUE UNIVERSITY

Bayesian Hierarchical Model



Unknown hyper-parameters

Parameters with prior distributions depending on θ^*

Observed data whose distributions depend on θ_i

Partial Prior Problems

- Another situation is when the prior is only available for a low-dimensional function of the parameters

Example: Meta-Analysis (Xie et al, 2013)

- $X \sim Bin(m, p_1)$ and $Y \sim Bin(n, p_2)$ are two independent observations
- Prior is on $\delta = p_1 - p_2 \sim \pi$
- Want to make inference about δ

Model Specification

Component	Formula
Sampling model	$X\|\theta \sim f(x\|\theta)$
Parameter partition	$\theta = (\tilde{\theta}, \theta^*)$
Partial prior	$\tilde{\theta}\|\theta^* \sim \pi(\tilde{\theta}\|\theta^*)$
Component without prior	θ^*
Parameter of interest	$\eta = h(\tilde{\theta})$

PURDUE UNIVERSITY

Inference Objective

- To construct interval estimators for the parameter of interest such that different approaches can be compared
- Definition follows Morris (1983):

$$P(\eta \in C(X)) \geq 1 - \alpha, \text{ for all } \theta^*$$

where P is computed over the joint distribution of (θ, X)

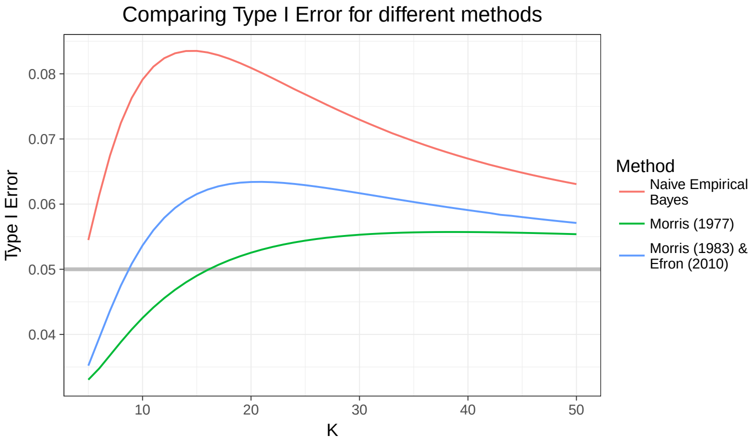
- Validity is our primary concern

Prior Work

- Full Bayesian Approach (Lindley and Smith, 1972)
  - Put prior on the unknown hyper-parameters
- Empirical Bayes (Robbins, 1956; Efron and Morris, 1971, 1972, 1973, 1975)
  - Use data to estimate unknown hyper-parameters, and plug the estimators into the prior
- Empirical Bayes confidence interval (Morris, 1983; Carlin and Gelfand, 1990)
  - Correct the bias in the naive Empirical Bayes
- Confidence Distribution (Xie et al, 2011; Xie et al, 2013)
  - Combine evidence from independent sources

Preserving Validity Is Non-trivial

- Bayesian hierarchical model example (Efron, 2010)



Outline

Tool – Inferential Models (IMs)

- A framework for prior-free and exact statistical inference
- Originally developed in Martin and Liu (2013), with extensions (Martin and Liu, 2015a, b) in a sequence
- We find the IMs very useful in providing solutions to partially specified Bayesian models
  - Dissolves boundaries between Frequentist and Bayesian, utilizing knowledge from both
  - Guarantees validity

Inferential Models - A Brief Introduction

A minimal example: $X \sim N(\theta, 1)$. Want to make inference on θ

Association step

Connect data, parameter, and unobserved auxiliary random variable through an association function

- $X = \theta + Z,\ Z \sim N(0, 1)$
- $\Theta_x(z) = x - z$

Prediction step

Predict the unobserved auxiliary random variable with a nested random set

- Define $\mathcal{S} = (-|V|, |V|), V \sim N(0, 1)$ to be a random interval
- $\mathcal{S}$ will cover the true value of Z with high probability

Inferential Models - A Brief Introduction cont.

Combination step

Transform the uncertainty from the auxiliary space to the parameter space

- Given observed data x, we expect

$$\Theta_x(\mathcal{S}) = \bigcup_{z \in \mathcal{S}} \Theta_x(z) = (x - |V|, x + |V|)$$

  to cover θ with high probability

PURDUE UNIVERSITY

Inferential Models - A Brief Introduction cont.

For any assertion on θ, for example $A = \{\theta : 1 < \theta < 2\}$, we compute two quantities:

Belief function

How much evidence supports that A is true (direct evidence)

$$\begin{aligned} \text{bel}_x(A) &= P\{\Theta_x(\mathcal{S}) \subseteq A | \Theta_x(\mathcal{S}) \neq \varnothing\} \\ &= P\{x - |V| > 1, x + |V| < 2\} \end{aligned}$$

Plausibility function

How much evidence does not support that A is false (indirect evidence)

$$\text{pl}_x(A) = 1 - \text{bel}_x(A^c)$$

PURDUE UNIVERSITY

Inferential Models - A Brief Introduction cont.

- Plausibility function can be used to construct $100(1-\alpha)\%$ interval estimator (plausibility region)

$$\mathrm{PR}_x(\alpha) = \{\theta : \mathrm{pl}_x(\{\theta\}) > \alpha\}$$

or hypothesis testing

$$\text{Reject } A \text{ if } \mathrm{pl}_x(A) \leq \alpha$$

- IMs guarantee the validity of the interval estimator

$$P_{X|\theta}\{\mathrm{PR}_x(\alpha) \ni \theta\} \geq 1 - \alpha$$

IMs on Partially Specified Bayesian Models

We use IMs to work out a simple but illustrative partially specified Bayesian model

Two-observation Normal-Normal Model

- $X_i|\mu_i \sim N(\mu_i, 1), i = 1, 2.$ X_1, X_2 are independent. μ_1, μ_2 are unknown means
- $\mu_i \sim N(\mu, 1)$ is the independent common prior, where μ is unknown
- Want to make inference about μ_1 (or μ_2)

PURDUE UNIVERSITY

Association Step

- Writing down association is typically trivial

$$\begin{cases} X_1 = \mu_1 + e_1 = (\mu + \varepsilon_1) + e_1 \\ X_2 = \mu_2 + e_2 = (\mu + \varepsilon_2) + e_2 \end{cases}, e_i, \varepsilon_i \overset{indep}{\sim} N(0,1)$$

Prediction Step

- An equivalent formulation of association is

$$\begin{cases} X_1 & = \mu_1 + e_1 \\ X_2 - X_1 & = \varepsilon_2 - \varepsilon_1 + e_2 - e_1 \end{cases}$$

- We can directly predict e_1 in the first equation, but
- The second equation tells that given the data, $\varepsilon_2 - \varepsilon_1 + e_2 - e_1$ is fully observed to be $x_2 - x_1$, which provides extra information about e_1
  - Combine information using Conditional IMs (Martin and Liu, 2015a)
  - Construct a predictive random set $\mathcal{S}$ for e_1 according to $e_1 | \{\varepsilon_2 - \varepsilon_1 + e_2 - e_1 = x_2 - x_1\} \sim N\left(\frac{1}{4}(x_1 - x_2), \frac{3}{4}\right)$

PURDUE UNIVERSITY

Combination Step and Plausibility Calculation

- Random set for parameter

$$\Theta_x(\mathcal{S}) = \bigcup_{e \in \mathcal{S}} \Theta_x(e) = \bigcup_{e \in \mathcal{S}} \{x_1 - e\}$$

- For assertion $A = \{\theta\}$, compute plausibility function

$$\text{pl}_x(\{\theta\}) = 2\Phi\left(-\frac{2}{\sqrt{3}}\left|\frac{3}{4}x_1 + \frac{1}{4}x_2 - \theta\right|\right)$$

- A $100(1-\alpha)$% interval estimator for μ_1 is given by

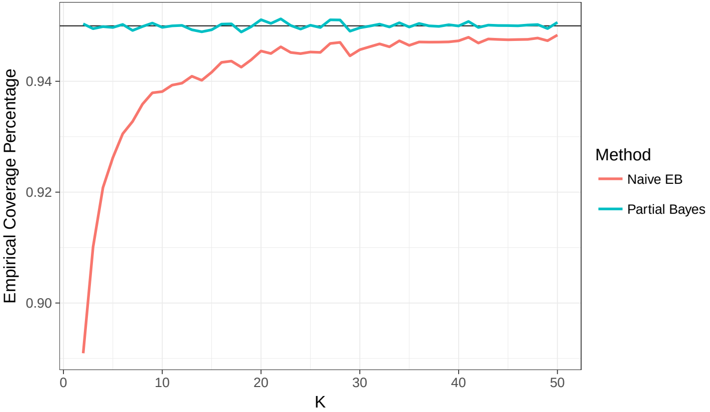
$$\text{PR}_X(\alpha) = \{\theta : \text{pl}_X(\{\theta\}) > \alpha\} = \left(\frac{3}{4}X_1 + \frac{1}{4}X_2\right) \pm \frac{\sqrt{3}}{2} z_{\alpha/2}$$

- Naive Empirical Bayes provides the interval estimator as

$$\left(\frac{3}{4}X_1 + \frac{1}{4}X_2\right) \pm \frac{\sqrt{2}}{2} z_{\alpha/2}$$

Comparing Two Solutions

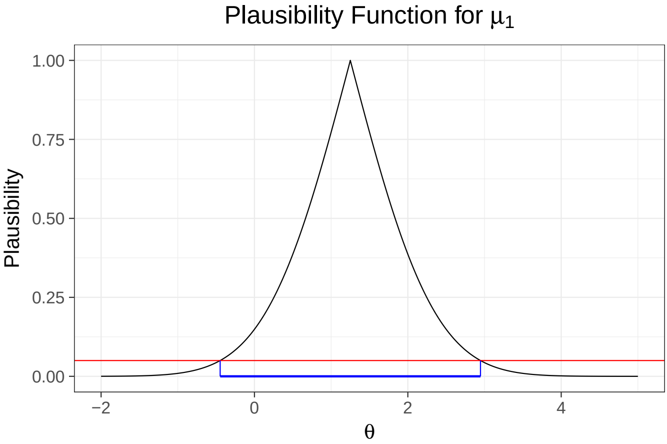
- The plot below shows the performance of the Partial Bayes solution



PURDUE UNIVERSITY

Plausibility Plot

- Assume that we have observed $x_1 = 1, x_2 = 2$



PURDUE UNIVERSITY

Outline

Other Popular Models

Normal-Normal Hierarchical Model

- $X_i|\mu_i \sim N(\mu_i, 1), i = 1, \ldots, K$ where X_i's are independent observed data, and μ_i's are unknown individual means
- Common prior $\mu_i \sim N(\mu, \tau^2)$. Both μ and τ^2 are unknown
- Want to make inference about μ_i

Gamma-Poisson Hierarchical Model

- $X_i|\lambda_i \sim Pois(\lambda_i), i = 1, \ldots, K$ where X_i's are independent observed data, and λ_i's are unknown individual means
- Common prior $\lambda_i \sim \gamma\text{G}(s)$. s is known and γ is unknown
- Want to make inference about λ_i

Normal-Normal Hierarchical Model

Component	Formula
Sampling model	$X \| (\tilde{\theta}, \theta^*) \sim \prod_i N(\mu_i, 1)$
Partial prior	$\tilde{\theta} = (\mu_1, \ldots, \mu_n), \tilde{\theta} \| \theta^* \sim \prod_i N(\mu, \tau^2)$
Component w/o prior	$\theta^* = (\mu, \tau^2)$
Parameter of interest	$\eta = h(\tilde{\theta}) = \mu_1$

Simulation Result

		Coverage		Median interval width	
τ^2	n	PB	Naive EB	PB	Naive EB
1	5	0.9526	0.8547	3.650	2.862
	10	0.9492	0.8684	3.304	2.779
	100	0.9529	0.9399	2.922	2.762
	1000	0.9567	0.9496	2.857	2.771
10	5	0.9501	0.9200	3.926	3.705
	10	0.9553	0.9361	3.907	3.722
	100	0.9589	0.9487	3.874	3.736
	1000	0.9585	0.9525	3.844	3.737

Gamma-Poisson Hierarchical Model

Component	Formula
Sampling model	$X\|(\tilde{\theta}, \theta^*) \sim \prod_i Pois(\lambda_i)$
Partial prior	$\tilde{\theta} = (\lambda_1, \ldots, \lambda_n), \tilde{\theta}\|\theta^* \sim \prod_i \gamma \mathsf{G}(s)$
Component w/o prior	$\theta^* = \gamma$
Parameter of interest	$\eta = h(\tilde{\theta}) = \lambda_1$

Simulation Result

		Coverage		Median interval width	
s	n	PB	Naive EB	PB	Naive EB
1	5	0.9550	0.9225	4.666	3.553
	10	0.9695	0.9398	4.444	3.610
	50	0.9560	0.9480	4.021	3.627
10	5	0.9547	0.9392	15.303	13.983
	10	0.9505	0.9434	14.592	14.008
	50	0.9534	0.9551	14.125	14.047

PURDUE UNIVERSITY

Outline

Motivation

Our Approach

Models and Simulations

Summary and Discussion

Summary

- Partially specified Bayesian models have a long history and are useful for data analysis
- We provided solutions to some popular models using the Inferential Models
- Simulation results demonstrate the validity of Partial Bayes solutions

Discussion

- Solving partially specified Bayesian models with validity guarantee is a challenging task, and is usually non-trivial
- Dealing with discrete random variables is even harder
- Computation can be intensive
- To explore the possibility of solving such problems using other successful inference frameworks
  - Confidence distribution
  - Generalized fiducial inference
  - Dempster-Shafer theory
  - ...

THANK YOU!