Discussion of "Objective data-dependent distributions" by Ryan Martin and "Posterior distributions with implicit objective priors" by Erlis Ruli, L. Ventura, N. Sartori and P. Jacob

Dongchu Sun

University of Missouri

O'Bayes 2019

Warwick, June 30, 2019

Thank Ryan and Erlis for sending me the slides at least one week earlier!

Posterior distributions with implicit objective priors

- An excellent review of the existing objective priors.
- Sometime accurate estimation of unknown parameters is the main interest.
- For a regular model, the asymptotic bias of posterior mode can be written as

$$b(\theta) = \frac{b_1(\theta)}{n} + \frac{b_2(\theta)}{n^2} + \ldots, \quad \theta \in \mathbb{R}^d.$$

- From Firth (1993), for the solution $\hat{\theta}^*$ of a modified score function,

$$\tilde{l}_r(\theta) = l_r(\theta) + a_r(\theta), r = 1, \cdots, d,$$

  the term $b_1(\theta)$ vanished. Here $a_r(\theta)$ is a suitable $O_p(1)$.
- Under the canonical multiparamter exponential family, $\hat{\theta}^*$ is the MAP under the Jeffreys prior! In this case, the Jeffreys rule prior would give higher order unbiased estimators.

Review of "Implicit Priors"

- For any parametric model with one parameter θ, the posterior one-side crdible interval of θ would be a first order matching prior for θ.
- For the canonical multiparamter exponential family, the posterior mode under Jeffreys rule prior would give higher order unbiased estimators.
- For non-canonical exponential family, no-close form expression are avaible for a Bias-Reduction prior density. So it is implicit.
- Three methods for approximating such "implicit" posteriors.
  - (a) Global Approximation via The Rao score function –a bimodal posterior.
  - (b) Local approximation based on Taylor expansion —better!
  - (c) Metropolis Hasting—slow convergence!

Questions

- How about the posterior means?
- Nonregular cases? e.g., there is no common support.
- What if we want to estimate several functions of parameters?
- What if we consider estimation under other losses? e.g. entropy loss.

Ryan's Motivation of "Objective data-dependent distributions"

- simpler prior specification
- faster computation
- robustness to model misspecification
- fast asymptotic concentration rates
- credible sets achieve nominal coverage
- identification of the quantity of interest:
  - through a statistical model
  - through some kind of loss function
  - Describe three different DDM constructions through various examples

DDM1: Bayes with empirical priors

- Example: The sparse normal means
- Using power prior distribution for regression models, Chen and Ibrahm (2000).
- Reach a faster convergence rate with correct marginal credible interval
- Based on at least five manuscripts....

DDM2: Gibbs posteriors

- Example. Minimum Clinically Important Difference (MCID):

$$\theta = \theta(P) = arg\min_{\varphi} P(Z \neq sign(X - \varphi))$$

- Parameters aren't always defined through a statistical model
  - quantiles, quantile regression, VaR, etc
  - area under the ROC curve
- Benefits: to avoid
  - nuisance parameters, hence marginalization
  - extra priors and extra computation
  - risk of misspecification bias
- Based on at least three manuscripts....

DDM3: permutations

- To estimate the density f based on iid data Y^n, gave a version of Newton's recursive algorithm for estimating a (mixture) density.
- Start with an initial guess $\hat{f}_0$, a weight sequence $w_i \in (0,1)$, and a Gaussian copula density c_p. For $i \geq 1$, update

$$\hat{f}_i(y) = \hat{f}_{i-1}(y)[1 + w_i\{c_p(\hat{F}_{i-1}(y), \hat{F}_{i-1}(Y_i) - 1\}]$$

- Take $\hat{f}_n$ as the final estimate based on Y^n.
- Properties:
  - very fast to compute
  - asymptotic L_1 consistency

DDM3: permutations—

- Create a DDM measure Π^n for f by randomly permuting the data sequence, i.e.,
- $S \sim Unif(permutations)$
- $\hat{f}_n^S$ = estimate based on data $Y_{S(1)}, \cdots, Y_{S(n)}$.
- Key: $\hat{f}_n$ depends on the data ordering. Could average over all permutations to reduce/eliminate this dependence.
- Based on at least two manuscripts

Questions

- DDW1: What is the noninformative prior for $\pi(\theta_S \mid S) \sim N_{|S|}(\mathbf{Y}_S, \gamma^{-1}\mathbf{I}_{|S|})$? Centered at the data? small γ?
- DDM2: The Gibbs posterior is

  $$\Pi_\omega^n(d\theta) \propto e^{-wnR_n(\theta)}\Pi(\theta),$$

  where Π is a prior for θ and $w > 0$ is the learning rate. Questions: How to choose Π and w?
- DDM3: permutations: What is the convergence rate comparing to usual optimal rate for nonparametric density estimates? There are $n!$ permutations, what is the computational time? Would nonGaussian copula c_p work better if you want to molti-mode density, such as mixtures?

General Questions

- In general how to construct a data dependent distribution?
- Bayesian, Fiducial, Frequentist methods? A confidence distribution and beyond?
- A practical data-dependent distribution of parameters? What about model selection, or prediction?
- "model-free" posterior? A Psuedo-likelihood?
- Such a data dependent distribution is often not unique, especially at multiparameter case. How do you choose?