Some aspects of matching priors

Setting: parametric inference, using a model (= likelihood), prior and posterior

Require: posterior probability statements to have sampling validity (Lindley, 1956)

Goals: default priors for 'routine' use

Bayesian/nonBayesian compromise

compare priors developed otherwise

Advantages:

Frequentist: marginalization for elimination of nuisance parameters

Bayesian: default prior 'should be' widely accepted

Overview

1. Edgeworth expansions for posterior quantiles, and probability matching

2. No general solution

3. Saddlepoint type/Strong matching

model	$f(y;\theta), \qquad \theta \in R^k$
data	$y = (y_1, \ldots, y_n)$
likelihood	$L(\theta) = L(\theta; y) = f(y;\theta)c(y)$
log-likelihood	$\ell(\theta) = \ell(\theta; y) = \log f(y;\theta) + a(y)$
prior, posterior	$\pi(\theta), \qquad \pi(\theta \vert y) \propto L(\theta; y)\pi(\theta)$
posterior quantile	$\theta^{(1-\alpha)}$ $\Pr_{\theta \vert Y}\{\theta \le \theta^{(1-\alpha)} \vert y\} = 1 - \alpha$
matching	$\Pr_{Y \vert \theta}\{\theta \le \theta^{(1-\alpha)}(\pi, Y)\} = 1 - \alpha +$
m.l.e.	$\hat{\theta} : \sup_\theta \ell(\theta) = \ell(\hat{\theta})$
obs. info.	$j(\hat{\theta})^{-1} =$ asy. var. $\hat{\theta}$
exp. info	$i(\theta) = (1/n)E\{\ell'(\theta)^2\}$

$$+o(n^{-1/2})$$

draw sketch of posterior and indicate quantile

note that equivalently we ask that posterior cdf for θ given Y be uniform under the sampling distribution, which is Welch and Peers approach

$$\pi(\theta|Y) \sim U(0,1)$$

under $f_Y(y;\theta)$

also there exist other approaches to 'default' or 'noninformative' priors

goal is typically [often] 'good' performance in repeated sampling

1. Edgeworth expansions ...

Want: $\text{Pr}_{Y|\theta}\{\theta \le \theta^{(1-\alpha)}(\pi, Y)\} = 1 - \alpha + o(n^{-1/2})$

– Step 1. Posterior density

$$\begin{aligned}\pi(\theta|y) &= \exp\{\ell(\theta)\}\pi(\theta)/\int \exp\{\ell(\theta)\}\pi(\theta)d\theta \\ &= \exp\{\ell(\hat{\theta}) + (\theta - \hat{\theta})\ell'(\hat{\theta}) + \ldots\}\{\pi(\hat{\theta}) + \ldots\}/ \\ &= \ldots = \phi(w)\{1 + \frac{1}{\sqrt{n}}(\mathrm{I}) + \frac{1}{n}(\mathrm{II}) + \ldots\}\end{aligned}$$

$w =$

$$\mathrm{I} = \left[\frac{\pi'(\hat{\theta})}{\pi(\hat{\theta})}\frac{1}{\{j(\hat{\theta})\}^{1/2}} + \frac{1}{3}\frac{\ell'''(\hat{\theta})}{\{j(\hat{\theta})\}^{3/2}}\right] w + (\cdots)w^3$$

– Step 2. Posterior cdf

$$\Pi(\theta|y) = \ldots = \Phi(w) + \phi(w)\{\frac{1}{\sqrt{n}}(\mathrm{I}') + \frac{1}{n}(\mathrm{II}') + \ldots\}$$

...1 Edgeworth expansions

– Step 3. Posterior quantile

$$\Pi(\theta^{(1-\alpha)}(y)|y) = \quad 1-\alpha \quad + \quad o(n^{-1/2})$$

$$\theta^{(1-\alpha)}(y) = \hat{\theta} + z_\alpha j(\hat{\theta})^{-1/2}\frac{1}{\sqrt{n}} +$$
$$j(\hat{\theta})^{-1/2}\frac{1}{n}\{(z_\alpha^2+2)A_3(y) + A_1(y)\} + .$$

– Step 4. Frequentist coverage

$$\Pr_{Y|\theta}\{\theta^{(1-\alpha)}(Y) \geq \theta\} = \Pr\{ \qquad \geq \theta\}$$
$$= \ldots = 1-\alpha + \frac{1}{\sqrt{n}}\phi(z_\alpha)T_1 + \frac{1}{n}z_\alpha\phi(z_\alpha)T_2 + \ldots$$

Mukerjee & Dey, 1993, Bka

$$T_1(\pi,\theta) = \left\{\frac{\pi'(\theta)}{\pi(\theta)} - \frac{i'(\theta)}{2i(\theta)}\right\} i^{-1/2}(\theta)$$

$$T_1(\pi,\theta) = 0 \iff \pi(\theta) \propto i^{1/2}(\theta)$$

Welch & Peers, 1963, Bka

Note matching is for all $\alpha \in (0, 1)$

n.b. $i(\theta)$ is expected Fisher information in one observation

This all assumes scalar parameter θ, but the same steps are followed for a scalar parameter of interest and vector of nuisance parameters; details somewhat messier; next slide

...1 Edgeworth expansions $\theta = (\theta_1, \ldots, \theta_k)$

– Want:

$$\text{Pr}_{Y|\theta}\{\theta_1 \leq \theta_1^{(1-\alpha)}(\pi, Y)\} = 1 - \alpha + o(n^{-1/2})$$

– Edgeworth expansion for *marginal* posterior

$$\pi_m(\theta_1|y) = \int \pi(\theta|y) d\theta_2 \ldots d\theta_k$$

– Cornish-Fisher inversion leads to

$$\theta_1^{(1-\alpha)} = \hat{\theta} + z_\alpha \hat{\sigma}_{11} + \ldots$$

– Frequentist coverage

$$= 1 - \alpha + \frac{1}{\sqrt{n}}\phi(z_\alpha)T_1(\pi, \theta) + \frac{1}{n} z_\alpha \phi(z_\alpha) T_2(\pi, \theta) + o(n^{-1})$$

2. No general solution

$T_1(\pi, \theta) = 0 \iff$

$$\sum_j \frac{\partial}{\partial \theta_j} \left[\{i^{11}(\theta)\}^{-1/2} i_{1j}(\theta) \pi(\theta) \right] = 0 \qquad (1)$$

Peers, 1965, Bka; Ghosh and Mukerjee, 1997

– in general has infinitely many solutions; e.g. suppose $i_{1j}(\theta) = 0, j = 2, \ldots, k$:

$$\frac{\partial}{\partial \theta_1} \{ i_{11}^{1/2}(\theta) \pi(\theta) \} = 0$$

$$\pi(\theta) \propto i_{11}^{1/2}(\theta) g(\theta_2, \ldots, \theta_k)$$

Tibshirani, 1989, Bka

– require (1) for each component in turn; leads to no solutions (in general)

...2 No general solution

What about matching to a higher order? T_2

Scalar parameter case: $\pi(\theta) \propto \{i(\theta)\}^{1/2}$:

$T_2(\pi, \theta) = 0 \iff$

$$\frac{d}{d\theta}\left[\frac{E\left(\frac{\partial \ell}{\partial \theta}\right)^3}{\{i(\theta)\}^{3/2}}\right] = 0$$

–In the orthogonal parameter case, the analogous condition is

$$\frac{1}{6} g(\theta_{(2)}) D_1 (i_{11}^{-3/2} i_{1,1,1})$$
$$+ \sum_{v=2}^{k} \sum_{s=2}^{k} D_v \{ i_{11}^{-1/2} i_{11s} i^{sv} g(\theta_{(2)}) \} = 0$$

where $i_{1,1,1} = E(\ell_1)^3$ and $i_{11s} = E(\ell_{11s})$

Mukerjee & Ghosh, 1997, Bka

...2 No general solution

Example: bivariate normal; $\theta_1 = \rho\mu_2/\mu_1$

– First order $\pi(\theta) \propto g(\theta_2, \theta_3, \theta_4, \theta_5) \left(\frac{\theta_3}{\theta_2}\right)^{1/2}$

– Second order $\pi(\theta) \propto g(\theta_3, \theta_4, \theta_5)\theta_2^{-1}$

Other matching criteria

– distribution function matching

– match under local alternatives

– match tolerance limits or other functions $h(\theta)$

– match distribution function for Wald or LR statistic

– match prediction limits

df matching

$$E \qquad \mathsf{Pr}_{\theta|Y}\{\sqrt{n}(\theta_1 - \widehat{\theta}_1)/\widehat{\sigma}_{11} \leq w|Y\}$$

$$= \mathsf{Pr}_{Y|\theta}\{\sqrt{n}(\theta_1 - \widehat{\theta}_1)/\widehat{\sigma}_{11} \leq w\} + O(n^{-j})$$

3 Saddlepoint-type expansions

1. Frequentist p-value for θ_1:

$$\Phi(r) + \phi(r)\left(\frac{1}{r} - \frac{1}{q}\right)$$

or

$$\Phi(r^*) = \Phi(r + \frac{1}{r}\log\frac{q}{r})$$

where

$$\begin{aligned} r &= \pm[2\{\ell_p(\widehat{\theta}_1) - \ell_p(\theta_1)\}]^{1/2} && \text{likelihood root} \\ q &= \{\chi(\widehat{\theta}) - \chi(\theta_1, \tilde{\theta}_{(2)})\}\sigma_{\chi\chi}^{-1/2} && \text{type of Wald stat.} \end{aligned}$$

– derived from p^* approximation

– accurate to $O(n^{-3/2})$

– approximates $\text{Pr}_{Y|\theta}\{R(\theta_1) \leq r(\theta_1)\}$

...3 Saddlepoint-type

2. Bayesian p-value for θ_1:

$$\Phi(r) + \phi(r)\left(\frac{1}{r} - \frac{1}{q}\right)$$

where

$$r = \pm[2\{\ell_p(\hat{\theta}_1) - \ell_p(\theta_1)\}]^{1/2} \qquad \text{likelihood root}$$

$$q = \ell_1(\theta_1, \tilde{\theta}_{(2)})\hat{\sigma}_{11}^{-1/2}\frac{\pi(\hat{\theta})}{\pi(\theta_1, \tilde{\theta}_{(2)})} \qquad \text{type of score stat.}$$

– derived from Laplace approximation to marginal posterior

– accurate to $O(n^{-3/2})$

– approximates $\text{Pr}_{\theta|Y}\{R \geq r(\theta_1)\}$

...3 Saddlepoint / strong matching

– Strong matching: $q_f = q_B$, i.e. $\Longleftrightarrow$

$$\frac{\pi(\theta_1, \tilde{\theta}_{(2)})}{\pi(\widehat{\theta})} = \ldots$$

– gives form of prior for θ_1, but not $\theta_{(2)}$

– depends on data

– not a very workable prescription as a 'default' prior

– but, does cast some light on the nature of matching priors:

...3 Strong matching

– frequentist p-value derived by finding an 'approximating exponential model' for $\ell(\theta)$, with canonical parameter $\varphi(\theta)$

– there is also an 'approximating location model', with location parameter $\beta(\theta)$

– the strong matching prior is flat in β: i.e. $\pi(\theta) \propto d\beta(\theta)$ in the scalar parameter case

– in the nuisance parameter case

$$\pi(\theta_1, \tilde{\theta}_{(2)}) \propto \left|\frac{\partial\theta_1(\theta)}{\partial\beta'(\theta)}\right|^{-1}_{(\theta_1,\tilde{\theta}_{(2)})} \cdot \text{info adjustment}$$

– If θ is a scalar then

$$\beta(\theta) = \int_{\hat{\theta}}^{\theta} -\frac{\ell_\theta(\theta)}{\varphi(\theta)} d\theta$$

... 3 Strong matching/data-dependent priors

– strong matching to 2nd order leads to $|j_{\varphi\varphi}(\varphi)|^{1/2}$, a data-dependent Jeffreys' prior

– data dependent priors may be inevitable

Pierce & Peters, 1994, Bka

– Example: Box-Cox model $y_i^{(\lambda)} = x_i'\beta + \sigma e_i$; $\theta = (\beta, \sigma, \lambda)$

$$\pi(\theta)d\theta \propto d\beta \frac{d\sigma}{\sigma} \frac{d\lambda}{(\dot{y}^{\lambda-1})^k}$$

Box & Cox, 1964, JRSSB

k is the dimension of β

$\dot{y}$ is the geometric mean

in mixture $c(\theta; y)$ deletes from the likelihood function the sample that comes entirely from the first component of the mixture

...3 Data dependent priors

– Example: mixture models

$$f(y;\theta) = \frac{1}{2}\phi(y) + \frac{1}{2}\phi(y-\theta)$$

– no fixed prior can match one-sided intervals to $O(n^{-1})$

$$\pi(\theta) \propto \{i(\theta)\}^{1/2} c(\theta; y)$$

where

$$c(\theta; y) = 1 - \prod \left\{ 1 + \frac{\phi(y_i - \theta)}{\phi(y_i)} \right\}$$

Wasserman, 2000, JRSSB

4 Conclusions

– no easy fix to the problem of nuisance parameters

– data-dependent priors may be necessary, even in a Bayesian context

– higher order asymptotics helps to understand problems in inference

– many other approaches to default priors, e.g. reference prior maximizes the Kullback-Liebler distance between the prior and the posterior

Berger & Bernardo, 1989, JASA ; Kass & Wasserman, 1996, JASA; Clarke & Yuan, 2001, preprint

– another approach: find 'the' likelihood for θ_1 (wip)