Some aspects of matching priors

Overview

1. Matching priors via Edgeworth type expansions

2. Matching priors via saddlepoint-type expansions (strong matching)

3. Example

with Rahul Mukerjee, Don Fraser

model	$f(y;\theta), \qquad \theta = (\psi, \lambda)$
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	$\pi(\theta), \qquad \pi(\theta \mid y) \propto L(\theta; y)\pi(\theta)$
	$\pi_m(\psi \mid y) = \int \pi(\theta \mid y) d\lambda$
m.l.e.	$\hat{\theta} : \sup_\theta \ell(\theta) = \ell(\hat{\theta})$
restricted m.l.e.	$\hat{\theta}_\psi : \sup_\lambda \ell(\psi, \lambda) = \ell(\hat{\theta}_\psi)$
obs. info.	$j(\hat{\theta})^{-1} = \text{asy. var. } \hat{\theta}$
exp. info	$i(\theta) = (1/n)E\{\ell'(\theta)^2\}$
partitioned info	$i(\theta) = \begin{pmatrix} i_{\psi\psi} & i_{\psi\lambda} \\ i_{\lambda\psi} & i_{\lambda\lambda} \end{pmatrix}$
	$j(\theta) = \begin{pmatrix} j^{\psi\psi} & j^{\psi\lambda} \\ j^{\lambda\psi} & j^{\lambda\lambda} \end{pmatrix}$

1. Matching priors: Edgeworth-type

a. Quantile matching - posterior quantile is also a confidence bound;

Posterior density

$$\pi_m(\psi|y) \;=\; \phi(w) + \frac{1}{\sqrt{n}}\phi(w)(\text{poly in } w) + \frac{1}{n}\phi(w)(\mathrm{II}) +$$

$$w = (\psi - \widehat{\psi})\{j^{\psi\psi}(\widehat{\theta})\}^{-1/2}$$

Posterior cdf

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

Quantile

$$\psi^{(1-\alpha)}(y) = \widehat{\psi} + z_\alpha\{j^{\psi\psi}(\widehat{\theta})\}^{1/2} + \ldots$$

Coverage

$$\Pr_{Y|\theta}\{\psi^{(1-\alpha)}(y) \geq \psi\} = \alpha + \frac{1}{\sqrt{n}}T_1 + \frac{1}{n}T_2 + \ldots$$

poly in w has coefficients that are messy functions of data, e.g. $\{\widehat{j}^{\psi\psi}\}^{-1/2}(\widehat{\pi}_j/\widehat{\pi})\widehat{j}^{\psi\lambda_j}$

shrinkage argument used to get $P_{Y|\theta}$

...1 Edgeworth expansions

$$T_1 = 0 \iff \pi(\psi, \lambda) \propto \{i_{\psi\psi}(\theta)\}^{1/2} g(\lambda)$$

Peers 1965

$$T_2 = 0 \iff \quad \frac{1}{6} g(\lambda) D_1(i_{11}^{-3/2} i_{1,1,1}) + \sum_{v=2}^{k} \sum_{s=2}^{k} D_v\{i_{11}^{-1/2} i_{11s} i^{sv} g(\lambda)\} = 0$$

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

Mukerjee & Ghosh, 1997, Bka

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

– First order $\pi(\lambda) \propto g(\lambda_1, \lambda_2, \lambda_3, \lambda_4) \left(\frac{\lambda_2}{\lambda_1}\right)^{1/2}$

– Second order $\pi(\lambda) \propto g(\lambda_2, \lambda_3, \lambda_4)\lambda_1^{-1}$

...1 Edgeworth type

b Other matching criteria

– distribution function matching

$$E \quad \text{Pr}_{\theta|Y}\{\sqrt{n}(\psi - \widehat{\psi})\{j^{\psi\psi}\}^{1/2} \leq w|Y\}$$

$$= \text{Pr}_{Y|\theta}\{\sqrt{n}(\psi - \widehat{\psi})/\{j^{\psi\psi}\}^{1/2} \leq w\} + O(n^{-j})$$

– from quantile argument, have expansion for $\text{Pr}_{\theta|Y}\{\sqrt{n}(\psi - \widehat{\psi})\{j^{\psi\psi}\}^{1/2} \leq w|Y\}$

– finding $E_{Y|\theta}$ means getting expected values of numerous coefficients

– use shrinkage argument on this result, to get $\text{Pr}_{Y|\theta}\{\sqrt{n}(\psi - \widehat{\psi})/\{j^{\psi\psi}\}^{1/2} \leq w\}$

then compare for each w (in bivariate normal gives same result, although not always)

...1 Edgeworth type

– match coverage under local alternatives

Mukerjee and Reid, 1999

in bivariate normal gives $g(\lambda_3, \lambda_4)/(\lambda_1\lambda_2)$

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

– match distribution function for Wald or LR statistic

– match prediction limits (very different expansion)

Datta et al., 2000, AnnStat; Sweeting 2001 Bka alternatives:

$\mathsf{Pr}_{Y|\theta}\{\psi \le \psi^{(1-\alpha)}(y)\} = 1 - \alpha + O(n^{-1})$ and

$\mathsf{Pr}_{Y|\theta}\{\psi + \delta(i^{\psi\psi})^{1/2} \le \psi^{(1-\alpha)}(Y)\} =$

$E_{Y|\theta}[\mathsf{Pr}_{\theta|Y}\{\psi+\delta(i^{\psi\psi})^{1/2} \le \psi^{(1-\alpha)}(Y)|Y\}]+O(n^{-1})$

(or $O(n^{-3/2})$)

2. Saddlepoint-type expansions

1. Bayesian p-value for ψ:

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

or

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

where

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

$$q = \ell_\psi(\psi, \hat\lambda_\psi)\left\{\frac{|j(\hat\theta)|}{|j_{\lambda\lambda}(\psi, \hat\lambda_\psi)|}\right\}^{-1/2}\frac{\pi(\hat\theta)}{\pi(\psi, \hat\lambda_\psi)} \qquad \text{type of s}$$

– derived from Laplace approximation to marginal posterior

– accurate to $O(n^{-3/2})$

– approximates $\Pr_{\theta|Y}\{R \geq r(\psi)|y\}$

– $R = r(\Psi, y)$

...2 Saddlepoint-type expansions

1. Frequentist p-value for θ_1:

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

where

$$r = \pm[2\{\ell_p(\widehat{\psi}) - \ell_p(\psi)\}]^{1/2} \qquad \text{likelihood root}$$

$$q = \{\chi(\widehat{\theta}) - \chi(\widehat{\theta}_\psi)\}\left\{\frac{|j(\widehat{\varphi})|}{|j_{(\lambda\lambda)}(\widehat{\theta}_\psi)|}\right\}^{1/2} \qquad \text{type of Wald}$$

– derived from p^* approximation

– accurate to $O(n^{-3/2})$

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

– $R = r(\psi, Y)$

$$\varphi(\theta) = \ell_{;V}(\theta; y^0) \qquad \chi(\theta) = \psi_\varphi(\widehat{\theta}_\psi)/| \qquad | \cdot \varphi(\theta)$$

$$|j_{(\lambda\lambda)}| = |j_{\lambda\lambda}||\varphi_\lambda|^{-2}$$

...3 Saddlepoint / strong matching

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

$$\frac{\pi(\psi, \hat{\lambda}_\psi)}{\pi(\hat{\theta})} = \frac{\ell_\psi(\hat{\theta}_\psi)}{\chi(\hat{\theta}\,) - \chi(\hat{\theta}_\psi)} \frac{|\varphi_\theta(\hat{\theta}\,)|}{|j(\hat{\theta}\,)|} \frac{|j_{\lambda\lambda}(\hat{\theta}_\psi)|}{|\varphi_\lambda(\hat{\theta}_\psi)|}$$

– gives form of prior for ψ, but not λ

– depends on data

– not a very workable prescription as a 'default' prior (e.g. in bivariate normal prescribes flat prior for regression coefficient ψ)

– 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 $\ell(\theta; y)$, with canonical parameter $\varphi(\theta)$

– i.e. has the properties of an exponential model, e.g. $\partial^2\ell/\partial\varphi^2$ depends only on φ, $\partial^2\ell/\partial\varphi\partial s = 1$, etc.

– Except for a term of $O(n^{-1})$, of the form $c\varphi^2 y^2$ (and we are ignoring terms of $O(n^{-3/2})$ and higher

– there is also an 'approximating location model', with location parameter $\beta(\theta)$, except for a single $O(n^{-1})$ term

– to $O(n^{-1})$ the strong matching prior is flat in β: i.e. $\pi(\theta) \propto d\beta(\theta)$

– if the model is location to $O(n^{-3/2})$, i.e. the single term is missing, then $d\beta$ is strong matching to $O(n^{-3/2})$

– in the nuisance parameter case, if ψ is linear in β, then we have strong matching for ψ

– otherwise we need to target ψ, using

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

– If θ is a scalar then

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

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

3 Example: Location model with curved parameter of interest

$Y_1 \sim N(\theta_1, 1)$, $Y_2 \sim N(\theta_2, 1)$ independent

$\psi^2 = (R + \theta_1)^2 + \theta_2^2$; $R known$

$r^2 = \surd\{(R + y_1)^2 + y_2^2)\}$

Bayesian posterior $(\theta_1, \theta_2 | y) \sim N(y_1, y_2)$

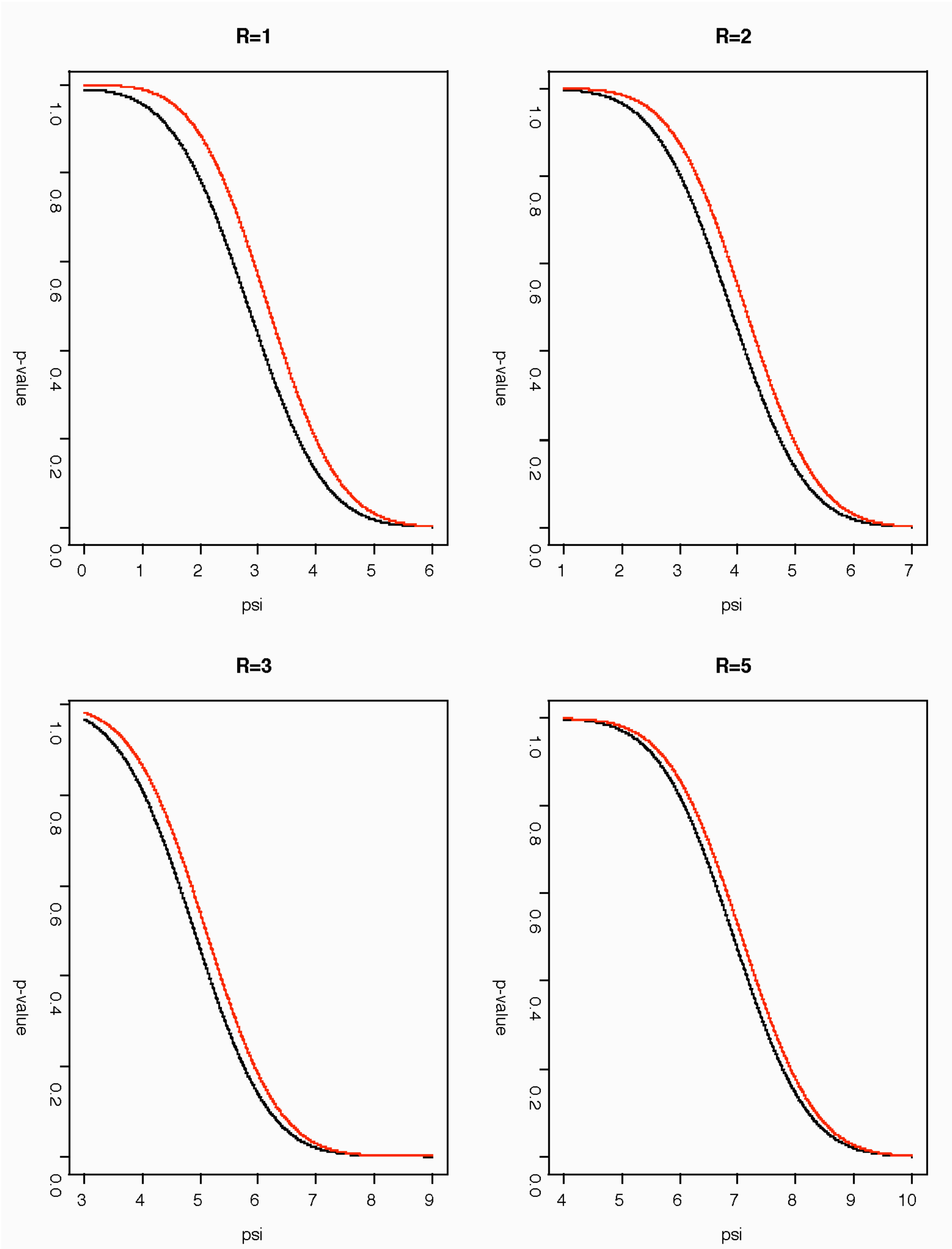
frequentist p-value (marginal) Pr$\{r \leq r^0; \psi^0)$

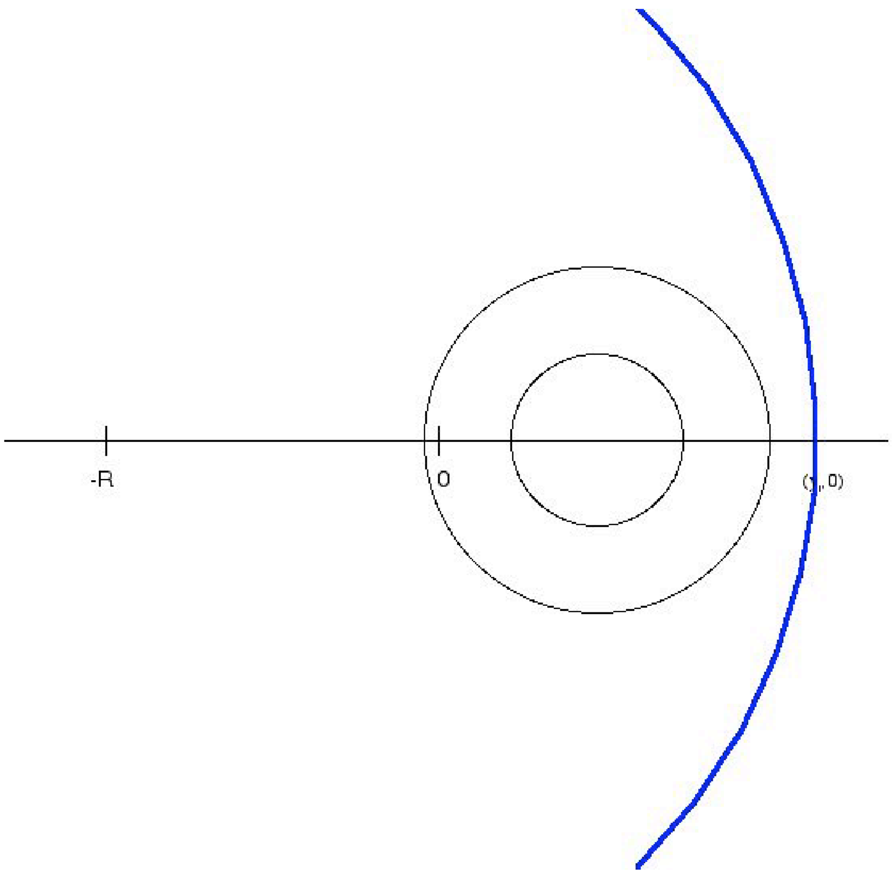
Bayesian p-value Pr$\{\psi \geq \psi^0 | y)$

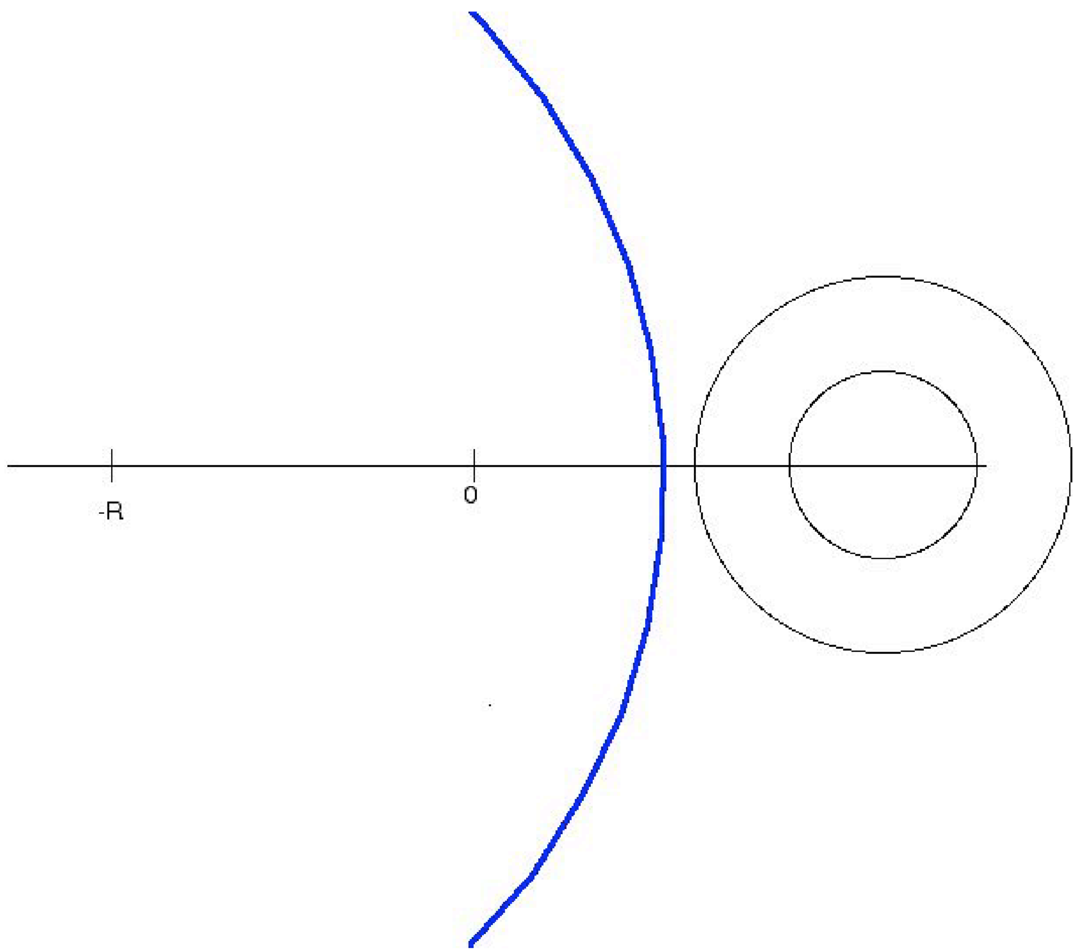
frequentist $=$ Pr$\{\chi_2^{2'}(\psi^0)^2 \leq (y_1^0 + R)^2 + y_2^{02}\}$

Bayesian $=$ Pr$\{\chi_2^{2'}((y_1^0 + R)^2 + y_2^{02}) \geq \psi^{02}\}$

Bayesian-frequentist $=$ Pr$\{X_1 - X_2 = 0\}$: $X_1 \sim Po((y_1^0 + R)^2 + y_2^{02}), X_2 \sim Po(\psi^{02})$


R=1
R=2
R=3
R=5
p-value
psi
0.0
0.2
0.4
0.6
0.8
1.0


-R
0


-R
0

4 Conclusions

– no easy fix to the problem of nuisance parameters

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

– 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 (Fraser, 2002)