OBJECTIVE PRIORS: A SELECTIVE REVIEW

Malay Ghosh
University of Florida

May 1, 2008

OUTLINE

- INTRODUCTION
- TWO BASIC TOOLS
- DIVERGENCE PRIORS
- MATCHING PRIORS
- SUMMARY AND CONCLUSION

1 INTRODUCTION

- Bayes (1763), Laplace (1776): Uniform prior for the binomial proportion p in the absence of any other information.
- Criticism: Lack of invariance under one-to-one reparameterization.
- A uniform prior for p in the binomial case does not result in a uniform prior for p^2.
- A uniform prior for σ, the population standard deviation, does not result in a uniform prior for σ^2, and the converse is also true.

- Jeffreys' general rule prior, namely the positive square root of the determinant of the Fisher information matrix.

- Remains invariant under one-to-one reparameterization.

- One parameter case:
  $I(\theta) = E(d\text{log}f/d\theta)^2$.
  ϕ is 1:1 with θ. Then
  $I(\phi) = E(\frac{d\text{log}f}{d\phi})^2 =$
  $E(\frac{d\text{log}f}{d\theta})^2(\frac{d\theta}{d\phi})^2 = I(\theta)(\frac{d\theta}{d\phi})^2$.

- Hence, $I^{1/2}(\phi) = I^{1/2}(\theta)|\frac{d\theta}{d\phi}|$.

- The result generalizes to the multiparameter case.

- Jeffreys' general rule prior enjoys many optimality properties in the absence of nuisance parameters.
- Maximizes the distance between the prior and the posterior in a certain sense.
- Enjoys probability matching property, i.e. the coverage probability of the resulting Bayesian one-sided credible interval matches asymptotically the coverage probability of the corresponding frequentist confidence interval.
- Under a suitable topology, it is the unique invariant uniform prior (J.K. Ghosh *et al.*, 2006).

- Invariance as such may not always be a desirable criterion.
- For the N($0, \sigma^2$) distribution, Jeffreys' prior $\pi_J(\sigma^2) \propto (\sigma^2)^{-1}$ is an inverse J-shaped prior.
- Now if $r = (\sigma^2)^{-1}$, the precision parameter, $\pi_J(r) \propto r^{-1}$, also an inverse J-shaped prior!
- In the presence of nuisance parameters, Jeffreys' general rule prior does not enjoy the aforementioned optimality properties, and may have undesirable features.
- Interestingly, even in the one-parameter case, it is possible to find alternative priors motivated from one or other optimality criterion.

2 TWO BASIC TOOLS

- An asymptotic expansion of the posterior density (Johnson, 1970).
- Let $X_1, \ldots, X_n|\theta$ iid with common pdf $f(X|\theta)$, and let $\hat{\theta}_n$ denote the MLE of θ.
- $L_n(\theta) = \prod_1^n f(X_i|\theta)$ and $\ell_n(\theta) = \log L_n(\theta)$.
- $a_i = n^{-1}[d^i \ell_n(\theta)/d\theta^i]_{\theta=\hat{\theta}_n}$, $i = 1, 2, \ldots$.
- $\hat{I}_n == -a_2$, the observed per unit Fisher information number.
- Twice differentiable prior π.
- Let $T_n = \sqrt{n}(\theta - \hat{\theta}_n)\hat{I}_n^{1/2}$, and let $\pi_n^*(t)$ be the posterior pdf of T_n.

Theorem 1. $\pi_n^*(t) = \phi(t)[1 + n^{-1/2}\gamma_1(t; X_1, \ldots, X_n) + n^{-1}\gamma_2(t; X_1, \ldots, X_n)] + O_p(n^{-3/2})$, where $\phi(t)$ is the standard normal pdf,

$$\gamma_1(t; X_1, \ldots, X_n) = \frac{a_3 t^3}{6\hat{I}_n^{3/2}} + \frac{t}{\hat{I}_n^{1/2}} \frac{\pi'(\hat{\theta}_n)}{\pi(\hat{\theta}_n)};$$

$$\gamma_2(t; X_1, \ldots, X_n) = \frac{1}{24\hat{I}_n^2} a_4 t^4 + \frac{1}{72\hat{I}_n^3} a_3^2 t^6 + \frac{1}{2\hat{I}_n} t^2 \frac{\pi''(\hat{\theta}_n)}{\pi(\hat{\theta}_n)} + \frac{1}{6\hat{I}_n^2} a_3 t^4 \frac{\pi'(\hat{\theta}_n)}{\pi(\hat{\theta}_n)} - \frac{a_4}{8\hat{I}_n^2} - \frac{15a_3^2}{72\hat{I}_n^3} - \frac{1}{2\hat{I}_n} \frac{\pi''(\hat{\theta}_n)}{\pi(\hat{\theta}_n)} - \frac{a_3}{2\hat{I}_n^2} \frac{\pi'(\hat{\theta}_n)}{\pi(\hat{\theta}_n)}.$$

- Outline of Proof: Write
$\pi(\theta|X_1,\ldots,X_n) = \frac{\exp[\ell_n(\theta)]\pi(\theta)}{\int \exp[\ell_n(\theta)]\pi(\theta)d\theta}$
$= \frac{\exp[\ell_n(\theta)-\ell_n(\hat{\theta}_n)]\pi(\theta)}{\int \exp[\ell_n(\theta)-\ell_n(\hat{\theta}_n)]\pi(\theta)d\theta}$.

- Then $T_n = \sqrt{n}(\theta - \hat{\theta}_n)\hat{I}_n^{1/2}$ has pdf
$\pi_n^*(t) =$
$C_n^{-1}\exp[\ell_n\{\hat{\theta}_n + t(n\hat{I}_n)^{-1/2}\} -$
$\ell_n(\hat{\theta}_n)]\pi\{\hat{\theta}_n + t(n\hat{I}_n)^{-1/2}\}$,
where C_n^{-1} is the normalizing constant.

- Use the fourth order Taylor expansion of $\ell_n\{\hat{\theta}_n + t(n\hat{I}_n)^{-1/2}\}$ around $\hat{\theta}_n$ and note $\ell_n'(\hat{\theta}_n) = 0$.

- A similar fourth order Taylor expansion of $\pi\{\hat{\theta}_n + t(n\hat{I}_n)^{-1/2}\}$ around $\hat{\theta}_n$.

- Shrinkage argument of J.K. Ghosh and his colleagues:
  Let X be a real-or a vector-valued random variable with pdf $f(X|\theta)$, $\theta \in \Theta$, some open subset of R^p.

- The goal is to find an expression for $E[q(X,\theta)|\theta]$, where q is a real-valued function integrable in both X and θ.

- As a special case, one may consider $q(X,\theta) = I_{[\sqrt{n}(\hat{\theta}_n - \theta) \leq y]}$, where $\hat{\theta}_n$ is the MLE of θ.

- Then $E[q(X,\theta)|\theta]$ is the distribution function of the MLE of θ.

- It is also possible to find moments of the MLE of θ.

- Step 1: Let $\bar{\pi}(\cdot)$ be a proper prior on a compact rectangle as support and vanishing on the boundary of support, while remaining positive in the interior. Consider the posterior of θ under $\bar{\pi}(\cdot)$ and hence obtain $E^{\bar{\pi}}[q(X,\theta)|X]$.
- Step 2: Find $E[\{E^{\bar{\pi}}(q(X,\theta)|X)\}|\theta] = \lambda(\theta)$ for θ in the interior of the support of $\bar{\pi}(\cdot)$.
- Step 3: Integrate $\lambda(\cdot)$ with respect to $\bar{\pi}(\cdot)$, and then allow $\bar{\pi}(\cdot)$ to converge to the degenerate prior at the true value of θ, an interior point in the support of $\pi(\theta)$. This yields $E[q(X,\theta)|\theta]$.

- Explicit specification of the prior $\bar{\pi}(\theta)$ is not needed in Steps 1-3.
- Also, these steps can lead to significant reduction in the algebra underlying higher order frequentist asymptotics.
- The simplification arises from two counts. First, although the Bayesian approach to frequentist asymptotics requires Edgeworth and other assumptions, it avoids an explicit Edgeworth expansion, avoiding thereby the calculation of higher order cumulants. Second, it helps establish the results in an easily interpretable compact form.

- Example 1 (Bias of the MLE).
  Recall $T_n = \sqrt{n}(\theta - \hat{\theta}_n)\hat{I}_n^{1/2}$ and
  $a_3 = n^{-1}[d^3\ell_n(\theta)/d\theta^3]_{\theta=\hat{\theta}_n}$.

- The posterior of T_n under $\bar{\pi}$ is
  $\bar{\pi}_n^*(t) = \phi(t)[1 + n^{-1/2}(\frac{a_3 t^3}{6\hat{I}_n^{3/2}} + \frac{t}{\hat{I}_n^{1/2}}\bar{\pi}'(\hat{\theta}_n)\bar{\pi}(\hat{\theta}_n)) + \gamma_2/n + O_p(n^{-3/2})]$,
  where γ_2 involves only even powers of t and constants.

- $E(T_n|X_1, \ldots, X_n) = \frac{1}{\sqrt{n}}(\frac{a_3}{2\hat{I}_n^{3/2}} + \frac{1}{\hat{I}_n^{1/2}}\frac{\bar{\pi}'(\hat{\theta}_n)}{\bar{\pi}(\hat{\theta}_n)}) + O_p(n^{-3/2})$.

- $\lambda(\theta) = E(\hat{\theta}_n|\theta) = \theta - E(\frac{T_n}{n\hat{I}_n^{1/2}}|\theta)$.

- $\lambda(\theta) = \theta - \frac{1}{n}(\frac{g_3(\theta)}{2I^2(\theta)} + \frac{1}{I(\theta)}\frac{\bar{\pi}'(\theta)}{\bar{\pi}(\theta)}) + O(n^{-2})$, where $g_3(\theta) = E[d^3\text{log}(f(X_1|\theta)/d\theta^3]$.

- $\int \lambda(\theta)\bar{\pi}(\theta)d\theta = \int \theta\bar{\pi}(\theta)d\theta - \frac{1}{n}\int\{\frac{g_3(\theta)}{2I^2(\theta)}\bar{\pi}(\theta) + \frac{1}{I(\theta)}\bar{\pi}'(\theta)\}d\theta$.
- Integrating the last term by parts,
$\int \lambda(\theta)\bar{\pi}(\theta)d\theta = \int \theta\bar{\pi}(\theta)d\theta - \frac{1}{n}\int\{\frac{g_3(\theta)}{2I^2(\theta)} + \frac{I'(\theta)}{I^2(\theta)}\}\bar{\pi}(\theta)d\theta + O(n^{-2})$.
- $\lambda(\theta) = \theta - \frac{1}{nI^2(\theta)}(\frac{g_3(\theta)}{2} + I'(\theta)) + O(n^{-2})$.
- For the one-parameter exponential family,
$g_3(\theta) = -I'(\theta)$. Hence,
$\lambda(\theta) = \theta - \frac{I'(\theta)}{2nI^2(\theta)} + O(n^{-2})$.

- Example 2 (Asymptotic Expansion of the Distribution Function of the MLE).

  The goal is to find

  $\lambda(\theta) = P[\sqrt{n}(\hat{\theta}_n - \theta)\hat{I}_n^{1/2} \leq u|\theta] = P(T_n \geq -u|\theta)$.

- $P^{\bar{\pi}}[T_n \geq -u|X_1, \ldots, X_n]$
  $= 1 - \Phi(-u) + \frac{1}{\sqrt{n}}\{\frac{a_3 u^3}{6\hat{I}_n^{3/2}} + \frac{u}{\hat{I}_n^{1/2}}\frac{\bar{\pi}'(\hat{\theta}_n)}{\bar{\pi}(\hat{\theta}_n)}\} + O_p(n^{-1})$.

- $\lambda(\theta) = \Phi(u) + \frac{1}{\sqrt{n}I^{3/2}(\theta)}\{\frac{g_3(\theta)u^3}{6} + uI'(\theta)\} + O(n^{-1})$.

3 DIVERGENCE PRIORS

- Bernardo (1979): Maximize $E\left[\log \frac{\pi(\theta|X_1,\ldots,X_n)}{\pi(\theta)}\right]$, the Kullback-Leibler (K-L) divergence between the prior and the posterior, where expectation is taken over the joint distribution of $X_1, \ldots, X_n, \theta$.
- Berger and Bernardo (1989): If one does this maximization for a fixed n, this may lead to a discrete prior with finitely many jumps. Hence, one needs an asymptotic maximization.
- Bernardo (1979) showed that without any nuisance parameters, this leads to Jeffreys' general rule prior.

- More general distance
  $D_\pi(\beta) = \frac{1-\{\frac{\pi(\theta|X_1,\ldots,X_n)}{\pi(\theta)}\}^{-\beta}}{\beta(1-\beta)}, \beta < 1.$
- The limiting $\beta \to 0$ case yields the K-L distance.
- $\beta = 1/2$:Bhattacharyya-Hellinger;
  $\beta = -1$:chisquare.
- Rewrite with $X = (X_1, \ldots, X_n)$.
  Consider $\beta \neq 0$.
  $D^\pi(\beta) = \frac{1-\int\int \pi^{\beta+1}(\theta)\pi^{-\beta}(\theta|X)L_n(\theta)dXd\theta}{\beta(1-\beta)}$
  $= \frac{1-\int \pi^{\beta+1}(\theta)E[\{\pi^{-\beta}(\theta|X)\}|\theta]d\theta}{\beta(1-\beta)}$.
- Maximize $D^\pi(\beta)$: Minimize (maximize) according as $0 < \beta < 1$ and $-1 < \beta < 0$:
  $\int \pi^{\beta+1}(\theta)E[\{\pi^{-\beta}(\theta|X)\}|\theta]d\theta$.

- From Theorem 1. the posterior of θ is $\pi(\theta|X) = \frac{\sqrt{n}}{\sqrt{2\pi}\hat{I}_n^{1/2}}exp[-\frac{n}{2}(\theta - \hat{\theta}_n)^2\hat{I}_n][1+O_p(n^{-1/2})]$.
- $\pi^{-\beta}(\theta|X) = n^{-\beta/2}(2\pi)^{-\beta/2}\hat{I}_n^{-\beta/2}exp[\frac{n\beta}{2}(\theta - \hat{\theta}_n)^2\hat{I}_n][1+O_p(n^{-1/2})]$.
- $E[\{E^{\bar{\pi}}(\pi^{-\beta}(\theta)|X)\}|\theta] = n^{-\beta/2}(2\pi)^{-\beta/2}[I(\theta)]^{-\beta/2}exp(\beta/2)[1+O_p(n^{-1/2})]$.
- Hence, minimize ($0 < \beta < 1$) or maximize ($-1 < \beta < 0$) $\int[\pi(\theta)/I^{1/2}(\theta)]^\beta \pi(\theta)d\theta$ with respect to π subject to $\int \pi(\theta)d\theta = 1$.
- The solution is $\pi(\theta) = I^{1/2}(\theta)$.

- The solution is different when $\beta = -1$. Then $\pi^{\beta+1}(\theta) = 1$ and first order asymptotic expansion does not suffice.

- In this case the solution is (Ghosh, Mergel and Liu; 2007) $\pi(\theta) \propto \exp[-\int^{\theta} \frac{2g_3(u)+I'(u)}{4I(u)} du$, where $g_3(u) = E[\frac{d^3 \log f(X_1|u)}{du^3}]$.

- One-parameter Exponential Family: $g_3(u) = -I'(u)$.

- $\pi(\theta) \propto I^{1/4}(\theta)$.

- For the Binomial (n, θ) problem, one gets $\pi(\theta) \propto \theta^{-1/4}(1-\theta)^{-1/4}$ which is a Beta $\left(\frac{3}{4}, \frac{3}{4}\right)$ distribution.

- For the Poisson (θ) case, one gets $\pi(\theta) \propto \theta^{-1/4}$.

4 MATCHING PRIORS

- Let $X_1, \ldots, X_n|\theta$ be iid with common pdf $f(X|\theta)$. For $0 < \alpha < 1$, let $\theta^\pi_{1-\alpha}(X_1, \ldots, X_n) \equiv \theta^\pi_{1-\alpha}$ denote the $(1-\alpha)$th asymptotic posterior quantile of θ based on the prior π, i.e. $P^\pi[\theta \leq \theta^\pi_{1-\alpha}|X_1, \ldots, X_n] = 1 - \alpha + O_p(n^{-p})$, for some $p > 0$.
- If $P[\theta \leq \theta^\pi_{1-\alpha}|\theta] = 1 - \alpha + O_p(n^{-p})$, then some order of probability matching is achieved.
- If $p = 1$, then we call π a first order probability matching prior. If $p = 2$, then we call π a second order probability matching prior.

- An intuitive argument why Jeffrey's prior is a first order probability matching prior in the absence of nuisance parameters.

- If $X_1, \ldots, X_n | \theta$ iid $N(\theta, 1)$ and $\pi(\theta) = 1, -\infty < \theta < \infty$, then the posterior $\pi(\theta | X_1, \ldots, X_n)$ is $N(\bar{X}_n, n^{-1})$.

- Writing $z_{1-\alpha}$ as the $100(1-\alpha)\%$ quantile of the $N(0,1)$ distribution, one gets $P[\sqrt{n}(\theta - \bar{X}_n) \leq z_{1-\alpha} | X_1 \ldots, X_n) = 1 - \alpha = P[\sqrt{n}(\bar{X}_n - \theta) \geq -z_{1-\alpha} | \theta]$, so that the one sided credible interval $\bar{X}_n + z_{1-\alpha}/\sqrt{n}$ has exact frequentist coverage probability $1 - \alpha$.

- The above exact matching does not always hold. However, if $X_1, \ldots, X_n | \theta$ are iid, then $\hat{\theta}_n | \theta$ is asymptotically $N(0, (nI(\theta))^{-1})$.

- Then by the delta method $g(\hat{\theta}_n)|\theta \sim N[g(\theta), (g'(\theta))^2 (nI(\theta))^{-1}]$.

- So if $g'(\theta) = I^{1/2}(\theta)$ so that $g(\theta) = \int^\theta I^{1/2}(t)dt$, one gets $\sqrt{n}[g(\hat{\theta}_n) - g(\theta)]|\theta$ is asymptotically $N(0,1)$.

- Hence, with the uniform prior $\pi(\phi) = 1$ for $\phi = g(\theta)$, coverage matching is asymptotically achieved for ϕ.

- This leads to the prior $\pi(\theta) = \frac{d\phi}{d\theta} = g'(\theta) = I^{1/2}(\theta)$.

- First by asymptotic expansion,
  $P^\pi(T_n \le z_{1-\alpha}|X_1,\ldots,X_n)$
  $= \int_{-\infty}^{z_{1-\alpha}} \phi(t)[1 + \frac{1}{\sqrt{n}}\{\frac{a_3 t^3}{\hat{I}_n^{3/2}} + \frac{t}{\hat{I}_n^{1/2}} \frac{\pi'(\hat{\theta}_n)}{\pi(\hat{\theta}_n)}\} + O_p(n^{-1})]dt$
  $= 1 - \alpha - \frac{k_\alpha^\pi}{\sqrt{n}}\phi(z_{1-\alpha}) + O_p(n^{-1})$,
  where
  $k_\alpha^\pi = \frac{a_3}{\hat{I}_n^{3/2}}(z_{1-\alpha}^2 + 2) + \frac{1}{\hat{I}_n^{1/2}} \frac{\pi'(\hat{\theta}_n)}{\pi(\hat{\theta}_n)}$.
- Let $\theta_{1-\alpha}^\pi = \hat{\theta}_n + (n\hat{I}_n)^{-1/2}(z_{1-\alpha} + k_\alpha^\pi/\sqrt{n})$.
- $P^\pi(\theta \le \theta_{1-\alpha}^\pi|X_1,\ldots,X_n) = 1 - \alpha + O_p(n^{-1})$.

- Next find $P(\theta \leq \theta^{\pi}_{1-\alpha}|\theta)$.
- Step 1: $P^{\bar{\pi}}(\theta \leq \theta^{\pi_{1-\alpha}}|X)$ $= P^{\bar{\pi}}[T_n \leq \sqrt{n}(\theta^{\pi}_{1-\alpha} - \hat{\theta}_n)\hat{I}_n^{1/2}|X]$ $= 1 - \alpha + \frac{1}{\sqrt{n}\hat{I}_n^{1/2}}[\frac{\pi'(\hat{\theta}_n)}{\pi(\hat{\theta}_n)} - \frac{\bar{\pi}'(\hat{\theta}_n)}{\bar{\pi}(\hat{\theta}_n)}]\phi(z_{1-\alpha}) + O_p(n^{-1})$.
- Step 2: $\lambda(\theta) = 1 - \alpha + \frac{\phi(z_{1-\alpha})}{\sqrt{n}}(\frac{\pi'(\theta)}{\pi(\theta)} - \frac{\bar{\pi}'(\theta)}{\bar{\pi}(\theta)})\frac{1}{I^{1/2}(\theta)} + O(n^{-1})$.
- Step 3: $\int \lambda(\theta)\bar{\pi}(\theta)d\theta = 1 - \alpha + \frac{\phi(z_{1-\alpha})}{\sqrt{n}} \int \frac{\bar{\pi}(\theta)}{\pi(\theta)} \frac{d}{d\theta}(\pi(\theta)I^{-1/2}(\theta))d\theta + O(n^{-1})$.
- Need $\pi(\theta) \propto I^{1/2}(\theta)$ for the coefficient of $\frac{1}{\sqrt{n}}$ to disappear.

- Second Order Matching : An asymptotic expansion of the posterior distribution function up to $O(n^{-3/2})$ is needed.

- Jeffreys' prior needs to satisfy a second differential equation
$\frac{1}{3}\frac{d}{d\theta}[\pi(\theta)I^{-2}(\theta)g_3(\theta)] -$
$\frac{d^2}{d\theta^2}[\pi(\theta)I^{-1}(\theta)] = 0$.

- Jeffreys' prior is second order matching if and only $E[\frac{\{\frac{d\log f(X|\theta)}{d\theta}\}^3|\theta]}{I^{3/2}(\theta)}$ does not depend on θ.

- For the bivariate normal distribution with zero means, unit variances and correlation coefficient ρ, Jeffreys' prior is not a second order matching prior for ρ. So, there does not exist any second order matching prior.

- Nuisance Parameters: Let $\theta = (\theta_1, \ldots, \theta_p)$, where θ_1 is the parameter of interest, while $\theta_2, \ldots, \theta_p$ are the nuisance parameters.

- Welch and Peers (1963), Datta and J.K. Ghosh (1995): Let $I^{-1} = ((I^{jk}))$. The probability matching equation is $\sum_{j=1}^{p} \frac{\partial}{\partial \theta_j}\{\pi(\theta) I^{j1} (I^{11})^{-1/2}\} = 0$.

- Tibshirani (1989): Let θ_1 be orthogonal to $(\theta_2, \ldots, \theta_p)$ in the Fisherian sense, i.e. $I^{j1} = 0$ for $j = 2, 3, \ldots, p$.

- The equation simplifies to $\frac{\partial}{\partial \theta_1}\{\pi(\theta) I_{11}^{-1/2}\} = 0$.

- So $\pi(\theta) = I_{11}^{1/2} h(\theta_2, \ldots, \theta_p)$, where h is arbitrary.

- An intuitive explanation of the prior $\pi(\theta) = I_{11}^{1/2}$ under orthogonality is that $\sqrt{n}(\hat{\theta}_{1n} - \theta_1)|\theta \sim N(0, I_{11}^{-1})$. So, one may expect $I_{11}^{1/2}$ to be a first order matching prior.

- However, this prior is not a second order matching prior even when such a prior exists.

- An example where exact matching holds is the location-scale family of distributions $f(X|\mu, \sigma) = \sigma^{-1} f((X - \mu)/\sigma)$, μ(real), $\sigma(> 0)$.

- The prior $\pi(\mu, \sigma) = \sigma^{-1}$ is an exact matching prior when either μ or σ is the parameter of interest, and the other one is the nuisance parameter.

- Moment Matching Priors : Let $X_1, \ldots, X_n$ be iid with common pdf $f(X|\theta)$, and $T_n = \sqrt{n}(\theta - \hat{\theta}_n)$.
- The posterior $\pi_n^*(t)$ of T_n is $\pi_n^*(t) = \phi(t)[1 + \frac{1}{\sqrt{n}}(\frac{a_3 t^3}{6\hat{I}_n^{3/2}} + \frac{t}{\hat{I}_n^{1/2}} \frac{\pi'(\hat{\theta}_n)}{\pi(\hat{\theta}_n)}) + O_p(n^{-1})]$.
- $E(\theta|X_1, \ldots, X_n) = \hat{\theta}_n + \frac{1}{n}(\frac{a_3}{2\hat{I}_n^2} + \frac{1}{\hat{I}_n} \frac{\pi'(\hat{\theta}_n)}{\pi(\hat{\theta}_n)}) + O_p(n^{-2})$.
- $n[E(\theta|X_1, \ldots, X_n) - \hat{\theta}_n] \xrightarrow{\text{P}} \frac{g_3(\theta)}{2I^2(\theta)} + \frac{1}{I(\theta)} \frac{\pi'(\theta)}{\pi(\theta)}$.
- Moment matching prior: $\pi(\theta) = \text{exp}[-\frac{1}{2} \int^{\theta} \frac{g_3(v)}{I(v)} dv]$.

- One parameter exponential family: θ canonical parameter. $g_3(v) = -I'(v)$.
- $\pi(\theta) = \text{exp}[\frac{1}{2}\text{log}I(\theta)] = I^{1/2}(\theta)$.
- Not necessarily so for other parameters. For example,
  $f(X|\mu) = \mu^{-1}\text{exp}(-X/\mu)$.
  $\pi(\mu) = I(\mu)$.
- Bivariate normal with zero means, unit variances and correlation coefficient ρ:
  $\pi(\rho) = I(\rho)$.
- Inverse Gaussian:
  $f(X|\mu) =$
  $(2\pi X^3)^{-1/2}\text{exp}[-(X-\mu)^2/(2\mu^2 X)]$.
  $\pi(\mu) = I(\mu)$.

5 SUMMARY AND CONCLUSION

- For the one-parameter case, Jeffreys' general rule prior has many optimality properties, but is not necessarily always optimal according to every single meaningful criterion.
- In the presence of nuisance parameters, it is well-known that Jeffreys' general rule prior does not work well. For example, for the location-scale family with location parameter μ and scale parameter σ, Jeffreys himself recommended the prior $\pi(\mu, \sigma) = \sigma^{-1}$ instead of the general rule prior $\pi_J(\mu, \sigma) = \sigma^{-2}$.

- In this example, the former is the right-Haar prior, while the latter is the left-Haar prior.
- The right-Haar prior enjoys many exact coverage matching properties.
- While I have primarily focused on divergence priors and matching priors, right-Haar priors can be of interest in their own right (Severini *et al.* (2002), Eaton and Sudderth (2004)).