Objective Bayesian Analysis under Semi-invariance Structure

Xiaoyan(Iris) Lin, University of South Carolina

Jointly with Dongchu Sun, University of Missouri-Columbia

James O. Berger, SAMSI & Duke University

June, 2009

Outline

- Invariance (Left and Right Haar) Priors
- Semi-invariance Structures
  - Motivated Examples
  - Some Results
- Comments

Invariance Priors

Generalizes ‘invariance to parameterization’ to other transformations that seem to leave a problem unchanged. There are many illustrations. ‘Invariance to a group operation’ was systematically studied.

An example: location-scale group operation on a normal distribution:

- Suppose $X \sim N(\mu, \sigma^2)$.
- Then $X^* = aX + b \sim N(\mu^*, (\sigma^*)^2)$, where $\mu^* = a\mu + b$ and $\sigma^* = a\sigma$.

Desiderata:

- Final answers should be the same for the two problems.
- Since the X and X^* problems have identical ‘structure,’ $\pi(\mu, \sigma^2)$ should have the same form as $\pi(\mu^*, \sigma^*)$.

Mathematical consequence: use an *invariant measure* corresponding to the ‘group action’ of the problem, the *Haar measure* if unique, and the *right-Haar measure* otherwise (optimal from a frequentist perspective). For the example, $\pi^{RH}(\mu, \sigma) = 1/\sigma$ (independence-Jeffreys prior).

Invariance Priors (Cont.)

Suppose $p(\mathbf{x} \mid \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, has a group structure $\mathcal{G}$ with the induced group $\widetilde{\mathcal{G}}$. We still write $\mathcal{G}$ for $\widetilde{\mathcal{G}}$ and assume it is locally compact with the unit e. For any $g_0 \in \mathcal{G}$, the left and right haar measures:

$$\pi^l(g_0) = \frac{1}{|J(g \rightarrow g_0 \circ g)|_{g=e}|},$$
$$\pi^r(g_0) = \frac{1}{|J(g \rightarrow g \circ g_0)|_{g=e}|}.$$

- $\pi^l(g_0)$ is often the Jeffreys rule prior on $\boldsymbol{\Theta}$;
- $\pi^r(g_0)$ is often a reference prior (Chang and Eavesl, 1990).
- For a location-scale family $\sigma^{-1} f((x-\mu)/\sigma)$,
  $$\pi^l(\mu, \sigma) = 1/\sigma^2 \text{ and } \pi^r(\mu, \sigma) = 1/\sigma.$$
- The Bayesian credible sets of (μ, σ) based on π^r are exact!

Example: Two-parameter Weibull

If $X \sim \textit{Weibull}(\eta, \beta)$, $Y = \log(X)$ has the Type I Extreme Value Distribution, with the density,

$$f(y \mid \eta, \beta) = \beta \exp\Big[-\beta(y - \log\, \eta)\Big], \quad y \in \mathbb{R}.$$

- It is a location-scale family with location parameter $\mu = \log\, \eta$ and scale parameter $\sigma = 1/\beta$.
- The Jeffreys and the reference priors for (μ, σ) are

$$\pi^J(\mu, \sigma) \propto \frac{1}{\sigma^2}, \quad \pi^R(\mu, \sigma) \propto \frac{1}{\sigma}.$$

- $\pi^R(\mu, \sigma)$ is equivalent to $\pi^R(\eta, \beta) = 1/(\eta\beta)$.
- The corresponding Bayesian credible sets of (μ, σ) or (β, η) based on π^r are exact!

Motivated Example I: Three-parameter Weibull

Let $X_1, \cdots, X_n \sim \mathit{Weibull}(\psi, \eta, \beta)$, with pdf,

$$p(x \mid \psi, \eta, \beta) = \frac{\beta(x-\psi)^{\beta-1}}{\eta^{\beta}} \exp\left\{ -\left(\frac{x-\psi}{\eta}\right)^{\beta} \right\}, \quad x > \psi, \quad (1)$$

- Model response times in psychological research (with Jeff Rouder and Paul Speckman).
- Extreme value program at SAMSI in Spring of 2008.
- Reliability and survival analysis.
- Fisher Information of $\boldsymbol{\theta} = (\psi, \eta, \beta)$? No common support!
- An ad hoc prior for (ψ, η, β) could be used:

$$\pi^*(\psi, \eta, \beta) \propto \frac{1}{\eta\beta}.$$

- Does the posterior of (ψ, η, β) exist?

Motivated Example II: Three-parameter log-normal

Let $t_1, \cdots, t_n \sim log - normal(\psi, \eta, \beta)$, with pdf,

$$f(t \mid \gamma, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}(t-\gamma)} \exp\left\{ -\frac{1}{2\sigma^2}\left[\log(t-\gamma)-\mu\right]^2 \right\}, \ t > \gamma.$$

- in socialogical and biological processes such as income of a population (Altchison & Brown, 1957).
- in the dosage requisite to produce a given response in biological assay (Finney, 1952).
- Hill (1963) considered the "independent" Jeffreys prior:

$$\pi^{IJ}(\gamma, \mu, \sigma) = \frac{1}{\sigma}. \tag{2}$$

- From Hill (1963),
  - if $\gamma \in \mathbb{R}$, the joint posterior is always improper!
  - if $\gamma \in (\gamma_0, \infty)$, the joint posterior is proper if $n \geq 3$.

A Statistical Models with Semi-invariance Structure

Consider a statistical model $p(\mathbf{x} \mid \boldsymbol{\theta}, \boldsymbol{\xi})$, where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $\boldsymbol{\xi} \in \Xi$. Suppose for each fixed $\boldsymbol{\theta}$, there is the (same) group invariance structure on Ξ, namely $\mathcal{G}$ with unit element $\mathbf{e}$.

- **Case A.** $\boldsymbol{\xi}$ is of interest and $\boldsymbol{\theta}$ is nuisance parameter(s).
- **Case B.** $\boldsymbol{\theta}$ is of interest and $\boldsymbol{\xi}$ is nuisance parameter(s);

Theorem 1

Case A. If $\boldsymbol{\xi}$ is of interest and $\boldsymbol{\theta}$ is nuisance parameter, the reference prior of $(\boldsymbol{\theta}, \boldsymbol{\xi})$ is

$$\pi(\boldsymbol{\theta}, \boldsymbol{\xi}) = \pi^R(\boldsymbol{\theta})\pi^r(\boldsymbol{\xi}),$$

where $\pi^R(\boldsymbol{\theta})$ is the reference or Jeffreys prior for $p(\mathbf{x} \mid \boldsymbol{\theta}, \boldsymbol{\xi} = \mathbf{e})$ and $\pi^r(\boldsymbol{\xi})$ is the right haar prior of $\boldsymbol{\xi}$ for the group $\mathcal{G}$ on Ξ.

Semi-location Family

Fact 1 Consider

$$p(x \mid \mu, \boldsymbol{\theta}) = g\left(x - \mu \mid \boldsymbol{\theta}\right), \tag{3}$$

where $g(\cdot \mid \boldsymbol{\theta})$ is a known density depending on $\boldsymbol{\theta}$ only, and $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^k$. Clearly, the right haar prior for μ is a constant prior,

$$\pi^r(\mu) \propto 1.$$

The reference prior when $\boldsymbol{\theta}$ is of interest is

$$\pi(\mu, \boldsymbol{\theta}) = \pi^R(\boldsymbol{\theta}),$$

$\pi^R(\boldsymbol{\theta})$ is the reference prior for $\boldsymbol{\theta}$ under the model $g(\cdot \mid \boldsymbol{\theta})$.

Motivated Example II: Three-parameter log-normal–Continued

$$f(t \mid \gamma, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}(t-\gamma)} \exp\left\{ -\frac{1}{2\sigma^2}\Big[\log(t-\gamma)-\mu\Big]^2 \right\}, \ t > \gamma.$$

Let $\xi = \gamma$ and $\boldsymbol{\theta} = (\mu, \sigma)$. We apply the one-at-a-time reference priors $\pi^R(\mu, \sigma) = 1/\sigma$ and

$$\pi^R(\gamma, \mu, \sigma) = \frac{1}{\sigma}. \tag{4}$$

Semi-location-scale Family

Fact 2 Consider

$$p(x \mid \mu, \sigma, \boldsymbol{\theta}) = \frac{1}{\sigma} g\left(\frac{x-\mu}{\sigma} \mid \boldsymbol{\theta}\right), \tag{5}$$

where $\sigma > 0$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^k$. Here μ is a location parameter, σ is a scale parameter, and $g(\cdot, \boldsymbol{\theta})$ is a known density depending on $\boldsymbol{\theta}$ only. We let $\boldsymbol{\xi} = (\mu, \sigma)$. Clearly,

$$\pi^r(\mu, \sigma) \propto \frac{1}{\sigma}.$$

Then

$$\pi^R(\mu, \sigma, \boldsymbol{\theta}) = \frac{1}{\sigma} \pi^R(\boldsymbol{\theta}),$$

where $\pi^R(\boldsymbol{\theta})$ is the reference prior for $\boldsymbol{\theta}$ under the model $g(\cdot \mid \boldsymbol{\theta})$.

Motivated Example I: Three-parameter Weibull–Continued

$$p(x \mid \psi, \eta, \beta) = \frac{\beta(x-\psi)^{\beta-1}}{\eta^\beta} \exp\left\{-\left(\frac{x-\psi}{\eta}\right)^\beta\right\}, \quad x > \psi,$$

It is a semi-location-scale family, with $\theta = \beta$,

$$p(x \mid \mu, \eta, \beta) = \frac{1}{\eta} g\left(\frac{x-\mu}{\eta} \;\middle|\; \beta\right), \tag{6}$$

where $g(y \mid \beta) = \beta y^{\beta-1} e^{-y^\beta} 1_{(0,\infty)}(y)$.
If $Y \sim g(y \mid \beta)$, $\log(Y)$ is a scale family with scale parameter $1/\beta$.
The reference prior for β is of the form, $\pi^R(\beta) = 1/\beta,\ \beta > 0$.
Consequently, the reference prior for (μ, η, β) is

$$\pi(\mu, \eta, \beta) = \frac{1}{\eta\beta}, \quad \mu \in \mathbb{R}, \eta > 0, \beta > 0. \tag{7}$$

- if $\mu \in \mathbb{R}$, the joint posterior is always improper!
- if $\mu \in (\mu_0, \infty)$, the joint posterior is proper if $n \geq 3$.

Simulations about the frequentist coverage of three parameter Weibull distribution are not finished yet!

Example III: Three-parameter Gamma

Consider one kind of the three parameter Gamma density,

$$f(x \mid \mu, \lambda, \alpha) = \frac{1}{\lambda\Gamma(\alpha)}\left(\frac{x-\mu}{\lambda}\right)^{\alpha-1}\exp\left(-\frac{x-\mu}{\lambda}\right), \quad x > \mu. \tag{8}$$

Let $y = (x-\mu)/\lambda$. Then y depends on α only with the pdf,

$$g(y \mid \alpha) = \frac{1}{\Gamma(\alpha)} y^{\alpha-1}\exp(-y), \quad y > 0.$$

The fisher information for α is $-E\left(\frac{\partial^2 l}{\partial\alpha^2}\right) = \frac{\partial^2 \log\Gamma(\alpha)}{\partial\alpha^2}$. Then $\pi^R(\alpha) = \sqrt{\frac{\partial^2}{\partial\alpha^2}\log\Gamma(\alpha)}$ and

$$\pi^R(\mu, \lambda, \alpha) = \frac{1}{\lambda}\sqrt{\xi(\alpha)}, \quad \mu \in \mathbb{R}, \lambda > 0, \alpha > 0. \tag{9}$$

Fact

Under $\pi^R(\mu, \lambda, \alpha)$, the joint posterior exists iff $n \geq 3$.

Simulations about the frequentist coverage of three parameter Gamma distribution are not finished yet!

Case B. $\boldsymbol{\theta}$ is of Interest

Suppose θ is the parameter of interest and $\boldsymbol{\xi}$ is nuisance parameter. There is a group structure $\mathcal{G}$ for Ξ with the right haar prior $\pi^r(\Xi)$. Assume that there is a maximum invariance statistic $\mathbf{T}$ for Ξ. Since

$$f(\mathbf{x} \mid \boldsymbol{\theta}, \boldsymbol{\xi}) = f(\mathbf{T} \mid \theta) f(\mathbf{x} \mid \mathbf{T}, \boldsymbol{\theta}, \boldsymbol{\xi}),$$

the marginal likelihood funciton of $\boldsymbol{\theta}$ is

$$\int_{\Xi} f(\mathbf{x} \mid \boldsymbol{\theta}, \boldsymbol{\xi}) \pi^r(\boldsymbol{\xi}) d\boldsymbol{\xi} = f(\mathbf{T} \mid \theta)\ H,$$

where

$$H = \int_{\Xi} f(\mathbf{x} \mid \mathbf{T}, \boldsymbol{\theta}, \boldsymbol{\xi}) \pi^r(\boldsymbol{\xi}) d\boldsymbol{\xi}. \tag{10}$$

Theorem 2

(a) H in (10) does not depend on $\boldsymbol{\theta}$.

(b) The reference prior for $\boldsymbol{\theta}$ is based on the model $f(\mathbf{T} \mid \theta)$ only.

Example IV: Bivariate Normal

The bivariate normal distribution of (x_1, x_2) has mean parameters (μ_1, μ_2) and variance matrix $\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, where ρ is the correlation between x_1 and x_2.

The priors when ρ is of interest have been facinating since 1930. See Berger and Sun (2008). Let $\boldsymbol{\xi} = (\mu_1, \mu_2, \sigma_1, \sigma_2)$ and consider a group of transformations

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \to \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} b_1 & 0 \\ 0 & b_2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

where $a_1, a_2 \in \mathbb{R}$, $b_1, b_2 > 0$. The right haar is $\pi^r(\boldsymbol{\xi}) = 1/(\sigma_1\sigma_2)$.
One can show that $\pi^R(\rho) = 1/(1-\rho^2)$.
The reference prior for $\{\rho, \boldsymbol{\xi}\}$ is the Lindley-Bayarri prior

$$\pi^R(\rho, \mu_1, \mu_2, \sigma_1, \sigma_2) = \frac{1}{\sigma_1\sigma_2(1-\rho^2)}.$$

Comments

- Invariance priors are reference priors are quite successful in practice.
- For a model with semi-invariance structure, we would take advantages of the invariance component.
- It can be applied to many examples.
- *Weakness?*
- Frequentist matching under a semi-invariance structure?