"The Formal Definition of Reference Priors"

by Berger, Bernardo and Sun (2009)

Hans Niederhausen

Dept. of Physics and Astronomy
Stony Brook University

December 3, 2015

Motivation

objectivity is a fundamental scientific desideratum

"Bayesians address the question everyone is interested in by using assumptions no-one believes" - Louis Lyons (2006)

- Reference priors offer objective Bayesian inference (Bernardo, 1979)

- Bernardo, Berger and Sun (2009) give rigorous definition for single parameter models (this talk!)

- Why do I care about Bayes?
  → include scientific information/knowledge
  (e.g. about nuisance parameters - ubiquitous in physics)

(particle physicists mostly use classical inference: "profile likelihood" cf. presentation by Hechuan Wang)

Definitions

DEFINITION 8 (Reference Prior):
A function $\pi(\theta) = \pi(\theta|M, P)$ is a reference prior for model M given a class of prior functions P if it is __permissible__ and has the __MMI property__

some properties

- independent of sample size
- compatible with sufficient statistics
- consistent under re-parametrization (invertible, one-to-one)
- reduce to Jeffrey's priors (if regularity conditions are met)
  $\pi(\theta) = \sqrt{I(\theta)}$ with $I(\theta) = -E_x\left\{\frac{\partial^2}{\partial\theta^2}\log[f(x|\theta)]\right\}$

Permissibiliy

DEFINITION 5 (Permissible Prior):

A strictly positive continuous function $\pi(\theta)$ is a permissible prior if 1.) $\int_{\Theta} f(x|\theta)\pi(\theta)\,d\theta < \infty$ for all $x \in X$ and 2.) for some approximating compact sequence $\{\Theta\}_{i=1}^{\infty}$ converging to Θ, the corresponding posterior sequence is <u>expected logarithmically convergent</u> to $\pi(\theta|x) \propto f(x|\theta)\pi(\theta)$

$\rightarrow$ view improper prior (defined on Θ) as approximation to the analysis based on the "true but difficult to specify" large bounded space

(aside: physics problems typically do NOT live in non-bounded parameter spaces)

$\rightarrow$ reference prior defined on Θ, when restricted to any large compact Θ_0, yields reference prior on Θ_0

(by logarithmic convergence)

Permissibiliy

DEFINITION 4 (Expected Logarithmic Convergence):

[...] The corresponding sequence of posteriors $\{\pi_i(\theta|x)\}_{i=1}^{\infty}$ is expected logarithmically convergent to $\pi(\theta|x)$ if

$$\lim_{i \to \infty} \int_X \kappa\left\{\pi(\theta|x)|\pi_i(\theta|x)\right\} p_i(x)\, dx = 0 \qquad (1)$$

where $p_i(x) = \int_{\Theta_i} f(x|\theta)\pi_i(\theta)\, d\theta$ (prior predictive distribution) and

$\kappa\{p|\tilde{p}\} = \int_{\Theta} p(\theta)\log\frac{p(\theta)}{\tilde{p}(\theta)}d\theta$ (Kullback-Leibler divergence)

The expectation in eq. (1) is needed to ensure good coverage of the corresponding interval estimates

(K-L divergence was also covered in Chen Mingshen's presentation)

MMI property

"MMI property" uses concept of missing information

DEFINITION 6 (Expected Information):

The information to be expected from one observation from Model
$M \equiv \{f(x|\theta), \theta \in \Theta, x \in X\}$ *with prior* $q(\theta)$ *is*

$$I(q|M) = \int \int_{X \times \Theta} f(x|\theta) q(\theta) \log \frac{p(\theta|x)}{q(\theta)} dx d\theta \qquad (2)$$

$\rightarrow$ interpret as expected K-L divergence between prior and posterior distributions from one observation

$\rightarrow$ measure of information gain from one observation

MMI property

Infinitely many observations $\{x_1, ..., x_k\}$ ($k \to \infty$) provide any missing information about the value of θ

$\to I\left(q|M^k\right)$ is measure of missing information (as $k \to \infty$).

DEFINITION 7 (Maximizing Missing Information):

The function $\pi\left(\theta\right)$ is said to have the MMI property for model M given P if, for any compact set $\Theta_0 \in \Theta$ and any $p \in P$,

$$\lim_{k \to \infty} \left\{ I\left(\pi_0|M^k\right) - I\left(p_0|M^k\right) \right\} \geq 0 \qquad (3)$$

where π_0 and p_0 are the normalized restrictions of π and p to Θ_0.

Reference Prior

THEOREM 7 (Explicit Form):

Let $\pi^*(\theta)$ be strictly positive, continuous, s.t. posterior distribution $\pi^*(\theta|t_k) \propto f(t_k|\theta)\pi^*(\theta)$ is proper and consistent. For any interior point $\theta_0 \in \Theta$ define

$$f_k(\theta) = \exp\left\{ \int_{T_k} p(t_k|\theta) \log\left[\pi^*(\theta|t_k)\right] dt_k \right\} \tag{4}$$

$$f(\theta) = \lim_{k \to \infty} \frac{f_k(\theta)}{f_k(\theta_0)} \tag{5}$$

$\pi(\theta) = f(\theta)$ is the <u>reference prior</u> for model M, if

1) $f_k(\theta)$ is continuous
2) $\left\{ f_k^0(\theta)/f_k^0(\theta_0) \right\}$ monotonic in k or bounded above (in θ)
3) $f(\theta)$ is permissible

+ authors describe routine for numerical computation of $\pi(\theta)$ on discrete points in parameter space
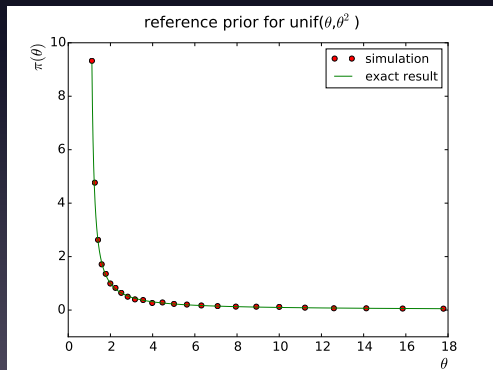
Numerical Calculations

A C++ implementation (requires BOOST) of the described algorithm available from:

‣ https://github.com/HansN87/ReferencePriors

applied to example 10 - unif $\left(\theta, \theta^2\right)$:

$$f\left(x|\theta\right) = \frac{1}{\theta^2 - \theta}, \ \ \theta < x < \theta^2, \ \ \theta > 1 \tag{6}$$

$$\pi\left(\theta\right) \propto \frac{2\theta - 1}{\theta\left(\theta - 1\right)} \exp\left\{\psi\left(\frac{2\theta}{2\theta - 1}\right)\right\} \tag{7}$$
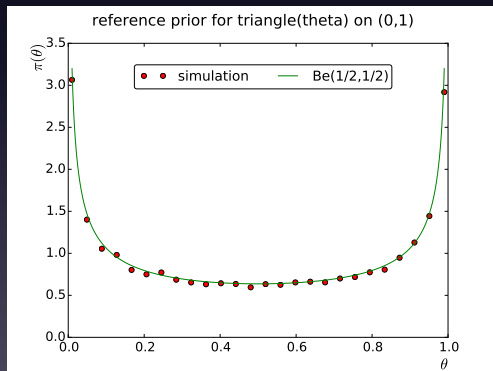

reference prior for unif(θ, θ^2)

Numerical Calculations

applied to example 11 - triangular distribution (no exact solution known):

$$f(x|\theta) = \begin{cases} \frac{2x}{\theta} & \text{for } 0 < x \le \theta \\ \frac{2(1-x)}{1-\theta} & \text{for } \theta < x < 1 \end{cases} \quad \text{and } 0 < \theta < 1 \quad (8)$$

reference prior for triangle(theta) on (0,1)

Conclusion

- Reference priors offer objective Bayesian inference

- Based on concept of maximizing missing information

- Presented analytic formula + numerical algorithm ...

- ... that provide solutions even in non-regular cases

- Multivariate generalizations exist but appear more involved

- Check out [1] for more details

Thank you!

[1] Berger, Bernardo and Sun, Ann. Statist. Volume 37, Number 2 (2009) + references therein