Evaluating Improper Priors via Markov Chain Arguments

Galin L. Jones

School of Statistics
University of Minnesota

January 9, 2007

Introduction

General Setting

Suppose $X \sim f(\cdot \mid \theta)$ for $\theta \in \Theta$ and let $\nu(\theta)$ be a prior such that

- $\int_\Theta \nu(\theta)\, d\theta = \infty$, so ν is *improper*
- The formal posterior $\pi(\theta \mid x) \propto f(x|\theta)\nu(\theta)$ is proper, so for all $x \in \mathsf{X}$

$$m(x) = \int_\Theta f(x|\theta)\nu(\theta)\, d\theta < \infty$$

When is $\nu(\theta)$ a reasonable prior?

Strong Admissibility

$X \sim f(\cdot \mid \theta)$ for $\theta \in \Theta$ and $\pi(\theta \mid x) \propto f(x|\theta)\nu(\theta)$

Assuming squared error loss the formal Bayes estimator of $g(\theta)$ is $\delta_g(X)$ where

$$\delta_g(x) = \int_\Theta g(\theta)\pi(\theta \mid x)\, d\theta$$

and the risk function is

$$R(\delta_g, \theta) = \int_{\mathsf{X}} [\delta_g(x) - g(\theta)]^2 f(x|\theta)\, dx$$

Strong Admissibility

$R(\delta_g, \theta) = \int_{\mathsf{X}} [\delta_g(x) - g(\theta)]^2 f(x|\theta)\, dx$

The estimator $\delta_g(X)$ is *admissible* if there does not exist another estimator $\delta(X)$ such that $R(\delta_g, \theta) \geq R(\delta, \theta)$ for all $\theta \in \Theta$ and $R(\delta_g, \theta) > R(\delta, \theta)$ for some θ.

The prior (posterior) is *strongly admissible* if the formal Bayes estimator of every bounded function of θ is admissible.

Strong admissibility ensures a certain robustness of the formal Bayes method.

Where are we?

$X \sim f(\cdot \,|\, \theta)$ for $\theta \in \Theta$ and an improper prior $\nu(\theta)$ yields $\pi(\theta \,|\, x) \propto f(x|\theta)\nu(\theta)$

Is ν (or equivalently π) reasonable?

That is, can we say anything about the robustness of the formal Bayes method?

The goal is to establish conditions for the prior to be strongly admissible; i.e., conditions that ensure the formal Bayes estimator of every bounded function of θ is admissible.

Strong admissibility seems difficult to verify. So we might look for sufficient conditions.

Markov Chain Connection

Fix $\theta \in \Theta$. Then $\pi(\eta|x)f(x|\theta)$ is a density in (η, x) which implies that

$$\int_{\mathsf{X}} \pi(\eta|x)f(x|\theta)\, dx$$

is a density in η.

Let $W = \{W_0, W_1, W_2, \dots\}$ be a Markov chain on Θ with transition density $\theta \to \eta$

$$r(\eta|\theta) = \int_{\mathsf{X}} \pi(\eta|x)f(x|\theta)\, dx$$

so that the transition kernel is

$$\Pr(W_{n+1} \in A \mid W_n = \theta) = P(A|\theta) = \int_A r(\eta|\theta)\, d\eta$$

Markov Chain Connection

$$r(\eta|\theta) = \int_{\mathsf{X}} \pi(\eta|x) f(x|\theta)\, dx$$

It is easy to see that

- W is ν–symmetric: $r(\eta|\theta)\nu(\theta) = r(\theta|\eta)\nu(\eta)$
- ν is the invariant "density" for W since

$$\int_{\Theta} r(\eta|\theta)\nu(\theta)\, d\theta = \nu(\eta)$$

Markov Chain Connection

$W = \{W_0, W_1, W_2, \ldots\}$ is a Markov chain on Θ with transition density $\theta \to \eta$

$$r(\eta|\theta) = \int_{\mathsf{X}} \pi(\eta|x) f(x|\theta)\, dx$$

Define $\tau_A = \inf\{n \geq 1 \;:\; W_n \in A\}$

W is *locally-ν-recurrent* if
$\nu(A) > 0 \;\Rightarrow\; \nu\{\theta \in A \;:\; \Pr_\theta(\tau_A < \infty) < 1\} = 0$

Except for a set of ν-measure 0 the chain returns to A with probability 1.

Eaton (1992, Annals): If W is locally-ν-recurrent then ν is strongly admissible.

Markov Chain Connection

Irreducibility

A Markov chain is *irreducible* if for each "big" set A and for every w there exists an $n > 0$ such that

$$P^n(A\,|\,w) > 0\,.$$

Recurrence

An irreducible Markov chain is *recurrent* if for each w and each "big" set A

$$\sum_{n=1}^{\infty} P^n(A|w) = \infty$$

Theorem

A symmetric, irreducible Markov chain is recurrent if and only if it is locally recurrent.

Multivariate Normal

$$X|\theta \sim \mathrm{N}(\theta, I_p)$$
$$\nu(\theta) \propto 1$$

Then the formal posterior is $\theta \mid x \sim \mathrm{N}(x, I_p)$ and the one step transition of the Markov chain can be described as follows:

Given the chain is at $\theta \in \mathbb{R}^p$, the next state of the chain is $\theta + V$ where $V \sim N(0, 2I_p)$.

The chain is thus a random walk on $\mathbb{R}^p$ and is recurrent for $p = 1, 2$. Hence ν is strongly admissible.
This can be extended to general location families.

Example

Eaton's chain is often irreducible, but not always.

$$f(x|\theta) = e^{\theta - x} I(\theta < x < \infty) I(0 \leq \theta < \infty) + e^{x-\theta} I(-\infty < x < \theta) I(-\infty < \theta < 0)$$

and take the prior to be Lebesgue measure on $\mathbb{R}$.
Then Eaton's chain W lives on $\mathbb{R}$ but it is not too difficult to show that if $\theta \geq 0$

$$\int_{-\infty}^{0} r(\eta|\theta)\, d\eta = 0$$

From which one can deduce that if the chain is started in $[0, \infty)$ it stays in $[0, \infty)$ forever and if it starts in $(-\infty, 0)$ it stays in $(-\infty, 0)$ forever.

Another Markov chain

Eaton's chain $W = \{W_0, W_1, W_2, \ldots\}$ lives on Θ with transition density $\theta \to \eta$

$$r(\eta|\theta) = \int_{\mathsf{X}} \pi(\eta|x) f(x|\theta) \, dx$$

Fix $x \in \mathsf{X}$. Then $f(y|\theta)\pi(\theta|x)$ is a density in (y, θ) which implies that

$$\int_{\Theta} f(y|\theta)\pi(\theta|x) \, d\theta$$

is a density in y.

Another Markov chain

Let $\tilde{W} = \{\tilde{W}_0, \tilde{W}_1, \tilde{W}_2, \ldots\}$ be a Markov chain on X with transition density $x \to y$

$$\tilde{r}(y|x) = \int_{\Theta} f(y|\theta)\pi(\theta|x)\, d\theta$$

and $\tilde{r}(y|x)m(x) = \tilde{r}(x|y)m(y)$.

Theorem

Assume either W or $\tilde{W}$ is irreducible. Then W is recurrent if and only if $\tilde{W}$ is recurrent.

Brown (1971, Annals) and Johnstone (1984, Annals)

Where are we?

$X \sim f(\cdot \,|\, \theta)$ for $\theta \in \Theta$ and an improper prior $\nu(\theta)$ yields $\pi(\theta \,|\, x) \propto f(x|\theta)\nu(\theta)$

The goal is to establish conditions for the prior to be strongly admissible; i.e., conditions that ensure the formal Bayes estimator of every bounded function of θ is admissible.

Strong admissibility seems difficult to verify but we can instead verify that W (or $\tilde{W}$) is locally recurrent...which also seems hard to verify directly. So we might look for sufficient conditions.

Establishing Local Recurrence

W is *locally-ν-recurrent* if
$\nu(A) > 0 \Rightarrow \nu\{\theta \in A : \Pr_\theta(\tau_A < \infty) < 1\} = 0$

Drift: $\Delta V(w) := \mathrm{E}[V(W_{n+1}) | W_n = w] - V(w)$

<u>Criterion for Recurrence</u>

Suppose W is a Markov chain on Θ with invariant measure ν and $V : \Theta \to \mathbb{R}^+$ such that

$$P(\limsup_{n\to\infty} V(W_n) = \infty \,|\, w_0) = 1$$

and there exists a set $C \subseteq \Theta$ such that

$$\Delta V(w) \leq 0 \;\text{ for all }\; w \in \bar{C}$$

then W is locally-ν-recurrent.

The condition

$$P(\limsup_{n\to\infty} V(W_n) = \infty \mid w_0) = 1 \tag{1}$$

is different than irreducibility.

Example

Let W be a Markov chain on $\mathbb{R}$ that evolves as follows: given $W_n = w \in \mathbb{R}$ obtain

$$W_{n+1} \sim \text{Uniform}(w, w+1)$$

Need to pick an appropriate V. Let $V : \mathbb{R}^+ \to \mathbb{R}^+$ be *unbounded off compact sets*; i.e., for any $0 < k < \infty$ the set

$$\{w \; : \; V(w) \le k\}$$

is compact.

Then (1) is satisfied but the chain is clearly reducible.

Drift on $\mathbb{R}^+$

$$\Delta V(w) \le 0 \quad \text{for all} \quad w \in \bar{C}$$

$C = [0, m]$ for some “large enough” m
$V(w) = \log(\log(w + e))$
$\mu_i(w) = \text{E}[(W_1 - W_0)^i | W_0 = w]$ for $i = 1, 2, 3$

Theorem

Let ψ_1 be a real-valued function such that

$$\frac{2w\mu_1(w)}{\mu_2(w)} \le 1 + \psi_1(w) \quad \text{and} \quad \lim_{w \to \infty} \log(w)\psi_1(w) = 0 \; .$$

Also assume

$$\lim_{w \to \infty} \frac{\mu_3(w)}{\mu_2(w)} \frac{\log(w)}{w} = 0 \; .$$

Then there exists $m > 0$ such that $\Delta V(w) \le 0$ for all $w > m$.

Multivariate Normal

$$X|\theta \sim \mathrm{N}(\theta, I_p)$$
$$\nu(\theta) = [a + \|\theta\|^2]^{-b}$$

Question

Does the posterior exist?

Answer

If either (i) $a > 0$ and $0 < b \leq p/2$ or (ii) $a = 0$ and $0 < b < p/2$ a formal posterior distribution exists.

Question

When is ν strongly admissible?

Answer If $p \geq 3$ and either (i) $a > 0$ and $p/2 - 1 \leq b \leq p/2$ or (ii) $a = 0$ and $p/2 - 1 < b < p/2$

Multivariate Normal

$$X|\theta \sim \mathrm{N}(\theta, I_p)$$
$$\nu(\theta) = [a + \|\theta\|^2]^{-b}$$

Question If $h : \Theta \to [0, \infty)$, when is the "perturbed" prior $\nu_h(d\theta) = h(\theta)\nu(d\theta)$ strongly admissible?

Answer If h is bounded, $p \geq 3$ and either (i) $a > 0$ and $p/2 - 1 \leq b \leq p/2$ or (ii) $a = 0$ and $p/2 - 1 < b < p/2$

New Question Does ν fail to be strongly admissible when $a > 0$ and $b < p/2 - 1$ or when $a = 0$ and $b \leq p/2 - 1$?

Remarks

- Strong admissibility attempts to capture a robustness property of the formal Bayes method.
- Markov chain stability properties are closely connected to strong admissibility.
- I have described general conditions for recurrence of Markov chains...
- ... and applied them to a specific example.
- While I have focused on the multivariate Normal distribution there have been a series of papers applying the methods to other situations.
- These results can also be extended to admissibility of estimators of unbounded functions.

Truncated Univariate Normal

$$X|\theta \sim \mathrm{N}(\theta, 1)$$
$$\nu(\theta) = I(\theta \in \mathbb{R}^+)$$

Then $m(x) = \Phi(x)$ and

$$\pi(\theta|x) = \frac{1}{\Phi(x)\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2} I(\theta \in \mathbb{R}^+)$$

Then the formal Bayes estimator of θ is

$$\delta_\theta(X) = X + \phi(x)/\Phi(x)$$

and

$$r^*(\eta|\theta) = \frac{(\eta-\theta)^2}{\mu_2(\theta)} r(\eta|\theta)$$

is a transition density that defines a Markov chain W^* on $\mathbb{R}^+$. I can use a previous theorem to establish local recurrence of W^* which in turn will guarantee that $\delta_\theta(X)$ is an admissible estimator.