

Lecture XI: Robustness and Asymptotics

Charles B. Moss

May 28, 2013

I. Robustness

- A. The development of priors is one of the most vexing areas for Bayesian analysis. One concept is to create bounds for the kernel

$$k(\theta_A | A) \ni a(\theta_A) < k(\theta_A | A) < b(\theta_A) \quad (1)$$

where $a(\theta_A)$ and $b(\theta_A)$ are kernels of prior distributions that yield proper densities.

- B. As a special case we consider

$$b(\theta_A) = r \cdot a(\theta_A). \quad (2)$$

"The density ratio class then permits ratios of prior probabilities of subsets Θ_A to vary by a factor of up to r from the corresponding ratios implied by the prior density kernel $k(\theta_A | A)$ (Geweke)."

- C. In this section we are interested in the estimator

$$g(\theta_A) = \int_{\Omega} h(\omega) p(\omega | y^0, \theta_A, A) d\nu(\omega). \quad (3)$$

For any kernel $k(\theta_A | A)$ we obtain

$$E[g(\theta_A) | y^0, A] = \frac{\int_{\Theta_A} k(\theta_A | A) p(y^0 | \theta_A, A) g(\theta_A) d\nu(\theta_A)}{\int_{\Theta_A} k(\theta_A | A) p(y^0 | \theta_A, A) d\nu(\theta_A)} \quad (4)$$

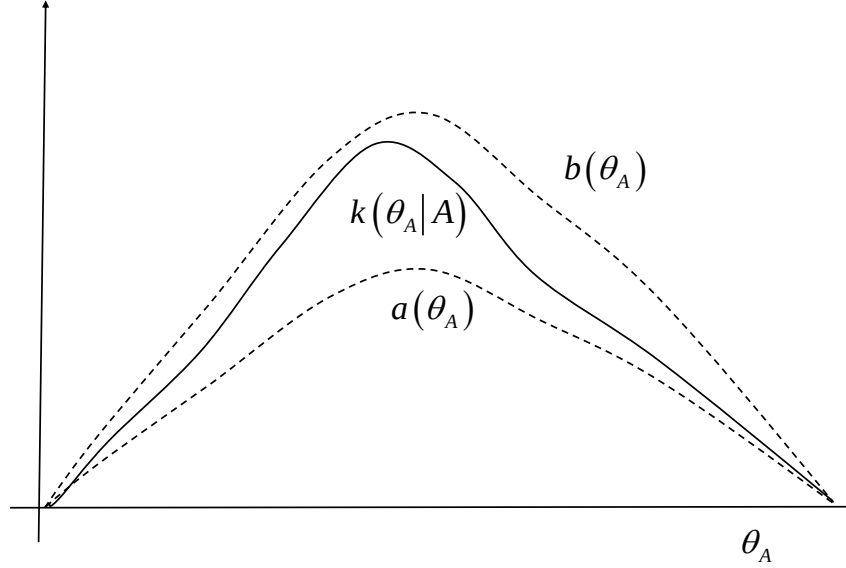


Figure 1: Bounded Kernel

Note that the denominator is

$$c = \int_{\Theta_A} k(\theta_A | A) p(y^0 | \theta_A, A) d\nu(\theta_A) \quad (5)$$

following our discussion in Lecture 10.

1. First let $a(\theta_A)$ and $b(\theta_A)$ be functions such that $0 \leq a(\theta_A) \leq k(\theta_A | A) \leq b(\theta_A) \forall \theta_A \in \Theta_A$ as depicted graphically in Figure 1 and $b(\theta_A) p(y^0 | \theta_A, A)$ is finitely integrable on Θ_A .
2. Next, we determine the range of values for Equation 4 over all the set of possible kernels $k(\theta_A | A) \subset S$ such that $0 \leq a(\theta_A) \leq k(\theta_A | A) \leq b(\theta_A)$

$$\begin{aligned} \underline{E}[g(\theta_A) | y^0, A] &= \inf_{k \in S} E[g(\theta_A) | y^0, A] \\ \overline{E}[g(\theta_A) | y^0, A] &= \sup_{k \in S} E[g(\theta_A) | y^0, A] \end{aligned} \quad (6)$$

3. **Theorem 3.3.1 Bounding a Posterior Moment over the Density Ratio Class of Priors** Let $b(\theta_A)$ be a prior density kernel for which the posterior density kernel $b(\theta_A)p(y^0|\theta_A, A)$ is finitely integrable on Θ_A , and let $a(\theta_A) \leq b(\theta_A)$ be a second prior density kernel. Let S be the set of all prior density kernels k satisfying $a(\theta_A) \leq k(\theta_A) \leq b(\theta_A) \forall \theta_A \in \Theta_A$, and suppose that Equation 4 is bounded above for $k \in S$. Let

$$k(\theta_A; c) = \begin{cases} a(\theta_A) & \text{if } g(\theta_A) \leq c \\ b(\theta_A) & \text{if } g(\theta_A) \geq c \end{cases} \quad (7)$$

Then the unique solution of

$$f(c) = \int_{\Theta_A} [g(\theta_A) - c] p(y^0|\theta_A, A) k(\theta_A; c) d\nu(\theta_A) = 0 \quad (8)$$

is $\hat{c} = E[g(\theta_A)|y^0, A] = \sup_{k \in S} E[g(\theta_A)|y^0, A]$.

4. "Because Theorem 3.3.1 remains valid with the formal substitution $p(y^0|\theta_A, A) = 1$, it provides bounds on prior expectations in a density ratio class of prior distributions as well (Geweke)."

II. Asymptotic Analysis

- A. Asymptotic analysis examines the properties of the posterior distribution as the sample size approaches infinity (i.e., the behavior of $p(\theta_A|Y_T^0, A)$ as $T \rightarrow \infty$).
1. These properties depend on the sequence $Y_T^0 = \{y_1^0, y_2^0, \dots, y_T^0\}$.
 - a. Here Geweke changes his notation by denoting the probability of the sequence Y_T as $p(Y_T|D)$ where D is the *data-generating process*.
 - b. Hence, the joint distribution for Y_T and θ_A is denoted $p(Y_T|D)p(\theta_A|Y_T, A)$.
 2. **Theorem 3.4.1 Asymptotic Concentration of the Posterior Distribution for a Discrete Parameter Vector** Suppose that
 - a. The vector of unobservables $\tilde{\theta}_A$ has a discrete prior distribution over a finite set of n points θ_{Ai} and $p(\theta_{Ai}|A) > 0$ ($i = 1, \dots, n$).

- b. $T^{-1} \ln p(Y_T | \theta_{Ai}, A) \xrightarrow{\text{a.s.}} l(\theta_{Ai}; A)$ ($i = 1, \dots, n$).
- c. For one $j \in \{1, \dots, n\}$, $l(\theta_{Aj}; A) > l(\theta_{Ai}; A)$ for all $i = 1, \dots, n$, $i \neq j$.

Then

$$\lim_{T \rightarrow \infty} P\left(\tilde{\theta}_A = \theta_{Aj} \mid \tilde{Y}_T, A\right) = 1. \quad (9)$$

3. **Definition 3.4.1** Given two densities $p(y|A)$ and $p(y|B)$ for the same observable $\tilde{y}$ and defined with respect to the same measure ν , the *Kullback-Leibler information criteria* (KLIC) distance from $p(y|A)$ to $p(y|B)$ is

$$\begin{aligned} K[p(y|A), p(y|B)] &= \int_{\Psi} p(y|A) \ln \left[\frac{p(y|A)}{p(y|B)} \right] d\nu(y) \\ &= E \left\{ \ln \left[\frac{p(\tilde{y}|A)}{p(\tilde{y}|B)} \right] \mid A \right\} \end{aligned} \quad (10)$$

Note that the KLIC distance is directed

$$K[p(y|A), p(y|B)] \neq K[p(y|B), p(y|A)], \quad (11)$$

and one can be finite while the other is infinite.

4. "Condition 3 of Theorem 3.4.1 may now be restated, for the case of i.i.d. distributions, as

$$\begin{aligned} &E[\ln(p(\tilde{y}|\theta_{Aj}, A)) | D] > E[\ln(p(\tilde{y}|\theta_{Ai}, A)) | D] \\ \Leftrightarrow &\int_{\Psi} \ln(p(y|\theta_{Aj}, A) p(y|D)) d\nu(y) > \int_{\Psi} \ln(p(y|\theta_{Ai}, A) p(y|D)) d\nu(y) \\ \Leftrightarrow &K[p(y|D), p(y|\theta_{Aj}, A)] < K[p(y|D), p(y|\theta_{Ai}, A)] \end{aligned} \quad (12)$$

for each i, j (Geweke)." As the sample size increases the posterior distribution concentrates to the parameter vector θ_{Aj} that provides the smallest KLIC for the data-generating function.

5. When the parameters are continuously distributed, the probability of any one vector of parameters approaches zero. Hence, we have to modify this intuition slightly.
6. **Theorem 3.4.2 Asymptotic Concentration of the Posterior Distribution for a Continuous Parameter Vector** Suppose that
 - a. The prior distribution of θ_A is absolutely continuous and $P(\theta_A \in C | A) > 0$ for all $C \subset \Theta_A$ for which $\int_C d\theta_A > 0$.
 - b. $T^{-1} \ln(p(Y_T | \theta_A, A) l(\theta_A; A))$ is uniformly distributed for all $\theta_A \in \Theta_A$.
 - c. $l(\theta_A; A)$ is a continuous function of θ_A with a unique global mode at $\theta_A = \theta_A^*$, and there exist l_L and l_U for which $M(\theta_A^*) = \{\theta_A : l_L < l(\theta_A) < l_U\}$ is a bounded open neighborhood of θ_A^* .

Then for any open neighborhood $N(\theta_A^*)$ of θ_A^* ,

$$\lim_{T \rightarrow \infty} P[\tilde{\theta}_A \in N(\theta_A^*) | \tilde{Y}_T, A] = 1. \quad (13)$$

7. Basically **Theorem 3.4.2** demonstrates that the likelihood will concentrate into a bounded neighborhood around θ_A^* .

B. **Theorem 3.4.3 Asymptotic Posterior Distribution for a Continuous Parameter Vector** Suppose that for all T , $\ln(p(\theta_A | Y_T, A))$ is twice differentiable for all $\theta_A \in \Theta_A$ and $Y_T \in \Psi_T$. Denote $L_T(\theta_A) = \ln(p(\theta_A | Y_T, A))$, $L'_T(\theta_A) = \partial L_T(\theta_A) / \partial \theta_A$ and $L''_T(\theta_A) = \partial^2 L_T(\theta_A) / \partial \theta_A \partial \theta'_A$. Suppose that with probability 1 there exists T^* such that

1. For all $T > T^*$, $L_T(\theta)$ has a strict local minimum at $\theta_{AT} = \theta_{AT}^*$, at which point $L'_T(\theta_{AT}^*) = 0$ and $L''_T(\theta_{AT}^*)$ is a negative definite matrix.
2. For all $T > T^*$, the largest eigenvalue σ_T^2 of the positive definite matrix $\Sigma_T = -[L''_T(\theta_{AT}^*)]^{-1}$ satisfies the condition $\lim_{T \rightarrow \infty} \sigma_T^2 = 0$.
3. Given any $\epsilon > 0$ there exists $T(\epsilon)$ and $\delta(\epsilon) > 0$ such that for all $T > T(\epsilon)$ and $\theta_A \in \{\theta_A : (\theta_A - \theta_{AT}^*)'(\theta_A - \theta_{AT}^*) < \delta(\epsilon)\}$.

$$I_k - B(\epsilon) \leq L''_T(\theta_A) [L''_T(\theta_{AT}^*)]^{-1} \leq I_k + B(\epsilon) \quad (14)$$

where $B(\epsilon)$ is a positive semidefinite matrix whose largest eigenvalue tends to zero as $\epsilon \rightarrow 0$.

4. There exists a point $\theta_A^* \in \Theta_A$ such that for any open neighborhood $N(\theta_A^*)$ of θ_A^*

$$\lim_{T \rightarrow \infty} P[\theta_A \in N(\theta_A^*) | Y_T, A] = 1. \quad (15)$$

For all $T > T^*$, let $\Sigma_T^{1/2}$ be any $k_A \times k_A$ matrix such that $(\Sigma_T^{-1/2})' \Sigma_T \Sigma_T^{1/2} = I$, let θ_{AT} be the random vector corresponding to the pdf $p(\theta_A | Y_A, A)$, and let $\tilde{z}_T = \Sigma_T^{1/2} (\theta_{AT} - \theta_A^*)$. Then $\tilde{z}_T \rightarrow N(0, I_{k_A})$.