

Lecture X: Improper Prior Distributions

Charles B. Moss

May 21, 2013

I. Improper Prior Distributions

- A. One suggestion is for researchers to use fairly flat or dispersed prior densities (especially in the relevant parameter space or Θ_A or the area in which we believe the likelihood function is concentrated).
 - 1. This procedure is appealing because a "flat" prior density function seems to convey the idea that we have little prior information.
 - 2. Thus, this flat distribution would be acceptable to a "diverse group of people who may have very different priors."
- B. "This rationale is misleading, and this can be seen by considering the effects of reparameterization of a model (Geweke)."
 - 1. Suppose that we can derive model B from model A such that $\theta_B = f(\theta_A)$ with $f(\cdot)$ being a one-to-one transformation.
 - 2. We can write $\theta_A = h(\theta_B)$ with $h = f^{-1}$ and $\theta_B \in \Theta_B$, with Θ_B being the image of Θ_A under $f(\cdot)$.
 - 3. The new observables density is

$$p(y|\theta_B, B) = p[y|\theta_A = h(\theta_B), A]. \quad (1)$$

The new prior density is

$$p(\theta_B|B) = p[\theta_A = h(\theta_B)|A] \left[\frac{\partial h(\theta_B)}{\partial \theta'_B} \right] \quad (2)$$

where $\partial h(\theta_B)/\partial \theta'_B$ is the Jacobian term. Note that we can make the distribution $p(\theta_B|B)$ nearly flat by choosing the

appropriate Jacobian term while the original prior $p(\theta_A | A)$ is not flat.

4. For purposes of discussion, consider a vector of interest ω

$$p(\omega | y, \theta_B, B) = p[\omega | y, \theta_A = h(\theta_B), A] \quad (3)$$

5. Working through the model

$$\begin{aligned} p(\omega | y, B) &\propto \int_{\Theta_B} p(\omega | y, \theta_B, B) p(\theta_B | B) d\nu(\theta_B) \\ &= \int_{\Theta_B} p[\omega | y, \theta_A = h(\theta_B), A] p[y | \theta_A = h(\theta_B), A] \\ &\quad \cdot p[\theta_A = h(\theta_B) | A] \left[\frac{\partial h(\theta_B)}{\partial \theta'_B} \right] d\nu(\theta_B) \\ &= \int_{\Theta_A} p(\omega | y, \theta_A, A) p(y | \theta_A, A) p(\theta_A | A) \\ &\quad \cdot \left| \frac{\partial f(\theta_A)}{\partial \theta'_A} \right|^{-1} \left| \frac{\partial f(\theta_A)}{\partial \theta'_A} \right| d\nu(\theta_A) \\ &= \int_{\Theta_A} p(\omega | y, \theta_A, A) p(y | \theta_A, A) p(\theta_A | A) d\nu(\theta_A) \\ &\propto p(\omega | y, A) \end{aligned} \quad (4)$$

The shape of the prior alone is no indication of how much information it conveys.

- B. Next, consider a sequence of models $A = \{A_j\}_{j=1}^{\infty}$ each with the same parameter vector θ_A (possibly different different normal distributions), data density $p(y | \theta_A, A)$, and vector of interest ω , but different priors $p(\theta_A | A_j)$.

1. Next, we conceptualize the sequence of priors using a sequence of kernels $k(\theta_A | A_j) \propto p(\theta_A | A_j)$. The corresponding sequence of posterior density functions is then

$$p(\theta_A | y^0, A_j) \propto p(y^0 | \theta_A, A) k(\theta_A | A_j) \quad (5)$$

2. **Definition 3.2.1** Let $k(\theta_A | A_j)$ be a sequence of prior density kernels for which $k(\theta_A | A) = \lim_{j \rightarrow \infty} k(\theta_A | \theta_A, A_j) \forall \theta_A \in \Theta_A$ exists but is not finitely integrable. If $\lim_{j \rightarrow \infty} p(\theta_A | y^0, A_j) \propto k(\theta_A | A) p(y^0 | \theta_A, A)$ exists and is finitely integrable, then $k(\theta_A | A)$ is an *improper prior density kernel* for θ_A in the model A with data y^0 .
3. "An attraction of using an improper prior distribution is that it can reflect some limiting properties of the sequence of distributions $\omega | (Y_T^0, A_j)$ and moments $E[h(\omega) | (y^0, A_j)]$ (Geweke)."
4. I have always been drawn to the improper prior because it focuses on the important portion - we can always integrate out the unknown constant as depicted in **Theorem 3.2.1**.
5. **Theorem 3.2.1 Convergence of Posterior Densities Given a Sequence of Prior Densities** Let $p(\theta_A, A_j)$ be a sequence of prior densities with corresponding kernels $k(\theta_A | A_j)$. Suppose that for all $\theta_A \in \Theta_A$, $k(\theta_A | A_j)$ is monotone non-decreasing with

$$\lim_{j \rightarrow \infty} k(\theta_A | A_j) = k(\theta_A | A) \quad (6)$$

where $k(\theta_A | A)$ is an improper prior density kernel of θ_A . Suppose further that

$$c_j = \int_{\Theta_A} p(y^0 | \theta_A, A) k(\theta_A | A_j) d\nu(\theta_A) \quad (7)$$

has finite limit c . Then $\theta_A | (y^0, A_j) \xrightarrow{d} \theta_A | (y^0, A)$ and a kernel of the limiting posterior $\lim_{j \rightarrow \infty} p(\theta_A | y^0, A_j)$ is $p(y^0 | \theta_A, A) k(\theta_A | A)$.

6. The implications of this theorem are important: "When the conditions of Theorem 3.2.1 are satisfied..."
 - a. "reports of posterior densities of parameters using the improper prior can be interpreted as limits of sequences of posterior densities employing priors whose kernels converge to the improper prior kernel."
 - b. "These conditions imply that there exist convergent sequences of credible sets, as well – that is

$$P(\theta_A \in C | y^0, A) 1 - \alpha \Rightarrow \lim_{j \rightarrow \infty} P(\theta_A \in C | y^0, A_j) = 1 - \alpha. \quad (8)$$

- c. **Corollary 3.2.1 Convergence of Posterior Moments Given a Sequence of Prior Densities** Suppose $\theta_A | (y^0, A_j) \xrightarrow{d} \theta_A | (y^0, A)$ and $g : \Theta_A \rightarrow R$ is continuous. If $\lim_{j \rightarrow \infty} E[g(\theta_A) | y^0, A_j] = g^*$, then $g^* = E[g(\theta_A) | y^0, A]$.
- d. **Corollary 3.2.2 Convergence of the Posterior Distribution of a Vector of Interest Given a Sequence of Prior Densities** Suppose $\theta_A | (y^0, A_j)$, $\theta_A | (y^0, A)$, and for all $\omega \in \Omega$, $g(\theta_A; \omega) = p(\omega | y^0, \theta_A, A_j)$ is a continuous function for θ_A . Then $\omega | (y^0, A_j) \xrightarrow{d} \omega | (y^0, A)$.
- e. **Corollary 3.2.3 Convergence of Posterior Moments of a Vector of Interest Given a Sequence of Prior Densities** Suppose $\omega | (y^0, A_j) \xrightarrow{d} \omega | (y^0, A)$ and $h : \Omega \rightarrow R$ is continuous. If $E[h(\omega) | y^0, A_j] = h^*$, then $h^* = E[h(\omega) | y^0, A]$.
- f. Note that these corollaries demonstrate that the improper distributions can be used to take the expectations of $g(\theta_A)$ and $h(\omega)$.

7. **Theorem 3.2.2 Marginal Likelihood Given an Improper Prior** The conditions of Theorem 3.2.1 imply

$$\lim_{J \rightarrow \infty} \int_{\Theta_A} p(\theta_A | A_j) p(y^0 | \theta_A, A) d\nu(\theta_A) = 0. \quad (9)$$

C. **Definition 3.2.2** If $\{y_t\}$ is i.i.d. with pdf $p(y_t | \theta_A, A)$, and $p(y_t | \theta_A, A)$ is differentiable with respect to $\theta_A \forall \theta_A \in \Theta_A$, then the *Jeffreys invariant prior density kernel* is

$$\begin{aligned} k(\theta_A | A) &\propto \left| E \left[\frac{\partial \ln p(y_t | \theta_A, A)}{\partial \theta_A} \cdot \frac{\partial \ln p(y_t | \theta_A, A)}{\partial \theta'_A} | \theta_A, A \right] \right|^{1/2} \\ &= \left| E \left[\frac{\partial^2 \ln p(y_t | \theta_A, A)}{\partial \theta_A \partial \theta'_A} \right] | \theta_A, A \right|^{1/2}. \end{aligned} \quad (10)$$

- D. **Theorem 3.2.3** Invariance of the Jeffreys Prior The Jeffreys invariant prior density is invariant under one-to-one reparameterization.