David Giles

Bayesian Econometrics

3. Properties of Bayes Estimators & Tests

(i) Some general results for Bayes Decision Rules.

(ii) Bayes estimators - some exact results.

(iii) Bayes estimators - some asymptotic (large-*n*) results.

(iv) Bayesian interval estimation.

General Results for Bayes Decision Rules

- Recall - Bayes rules and MEL rules (*generally equivalent*)
- Minimum Expected Loss (MEL) Rule:

  "Act so as to **M**inimize (posterior) **E**xpected **L**oss"

$$\int_{\Omega} L(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}) p(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta}$$

- Bayes' Rule:

  "Act so as to Minimize Average Risk."

  (Often called the "Bayes' Risk"):

$$r(\widehat{\boldsymbol{\theta}}) = \int R\,(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

- The action will involve selecting an Estimator, or rejecting some Hypothesis, for instance.

- *We have the following general results –*

(i) A Mini-max Rule always exists.

(ii) Any decision rule that is Admissible must be a Bayes' Rule with respect to some prior distribution.

(iii) If the prior distribution is "proper", every Bayes' Rule is Admissible.

(iv) If a Bayes' Rule has constant risk, then it is Mini-max.

(v) If a Mini-max Rule corresponds to a *unique* Bayes' Rule, then this Mini-max Rule is also *unique*.

Bayes Estimators – Some Exact Results

- As long as the prior p.d.f., $p(\boldsymbol{\theta})$ is "proper", the corresponding Bayes' estimator is Admissible.
- Bayesians are not really interested in the properties of their estimators in a "repeated sampling" situation.
- Interested in behaviour that's conditional on the data in the *current sample*.
- Contrast "Bayesian Posterior Probability Intervals" (or, "Bayesian Credible Intervals") with "Confidence Intervals".
- However, note that Bayes' estimators may be biased or unbiased in finite samples.

Bayes Estimators – Some Asymptotic Results

- Intuitively, we'd expect that as the sample size grows, the Likelihood Function will dominate the Prior.
- In the limit, we might expect that Bayes' estimators will converge to MLE's.
- An exception will be if the prior is "totally dogmatic" (degenerate).
- So, not surprisingly, Bayes estimators are *weakly consistent*.
- The principal asymptotic result associated with Bayes' estimators is the so-called "Bernstein-von Mises Theorem".

Sergei Bernstein

(1880 – 1968)

Richard von Mises

(1883 – 1953)

Theorem:

Unless $p(\boldsymbol{\theta})$ is degenerate, $\lim_{n\to\infty}\{p(\boldsymbol{\theta}|\boldsymbol{y})\} = N[\widetilde{\boldsymbol{\theta}}, I(\widetilde{\boldsymbol{\theta}})^{-1}]$; where $\widetilde{\boldsymbol{\theta}}$ is the MLE of $\boldsymbol{\theta}$.

Proof: (*for the scalar case*)

From *Bayes' Theorem* –

$$p(\theta|\boldsymbol{y}) \propto p(\theta)L(\theta|\boldsymbol{y}) = p(\theta)exp\{logL(\theta|\boldsymbol{y})\}$$

- Take a Taylor's series expansion for $p(\theta)$ about the MLE, $\tilde{\theta}$ –

$$p(\theta) = p(\tilde{\theta}) + (\theta-\tilde{\theta})p'(\tilde{\theta}) + \frac{1}{2}(\theta-\tilde{\theta})^{2}p''(\tilde{\theta}) + \cdots$$

$$= p(\tilde{\theta})[1 + (\theta-\tilde{\theta})\frac{p'(\tilde{\theta})}{p(\tilde{\theta})} + \frac{1}{2}(\theta-\tilde{\theta})^{2}\frac{p''(\tilde{\theta})}{p(\tilde{\theta})} + \cdots]$$

$$\propto [1 + (\theta - \tilde{\theta})\frac{p'(\tilde{\theta})}{p(\tilde{\theta})} + \frac{1}{2}(\theta - \tilde{\theta})^{\mathbf{2}}\frac{p''(\tilde{\theta})}{p(\tilde{\theta})} + \cdots]$$

- Let $l(\theta) = logL(\theta)$.

- $exp\{logL(\theta|\boldsymbol{y})\} = exp\{l(\theta)\}$

$$= exp\{l(\tilde{\theta}) + (\theta - \tilde{\theta})l'(\tilde{\theta}) + \frac{1}{2}(\theta - \tilde{\theta})^{\mathbf{2}}l''(\tilde{\theta}) + \cdots\}$$

- Note that $l'(\tilde{\theta}) = 0$.

- So,

$$exp\{l(\theta)\} = \color{red}{exp\{l(\tilde{\theta})\}}\, exp\{\frac{1}{2}(\theta - \tilde{\theta})^{\mathbf{2}}l''(\tilde{\theta})\}\, exp\{\frac{1}{6}(\theta - \tilde{\theta})^{\mathbf{3}}l'''(\tilde{\theta})\} \dots.$$

Or,

$$exp\{l(\theta)\} \propto exp\{\frac{1}{2}(\theta-\tilde{\theta})^{\mathbf{2}} l''(\tilde{\theta})\}\ exp\{\frac{1}{6}(\theta-\tilde{\theta})^{\mathbf{3}} l'''(\tilde{\theta})\} \ldots.$$

- Expand this exponential:

$$exp\{l(\theta)\} \propto exp\{\frac{1}{2}(\theta-\tilde{\theta})^{\mathbf{2}} l''(\tilde{\theta})\}\ [1+\{\frac{1}{6}(\theta-\tilde{\theta})^{\mathbf{3}} l'''(\tilde{\theta})\}+\cdots].$$

- The leading term is

$$exp\{l(\theta)\} \propto \mathrm{e}xp\{\frac{1}{2}(\theta-\tilde{\theta})^{\mathbf{2}} l''(\tilde{\theta})\}$$

- This will apply for large *n*

In this case,

$$p(\theta|\boldsymbol{y}) \propto exp\{l(\theta)\}p(\theta)$$

$$\propto exp\left\{\frac{1}{2}\left(\theta-\tilde{\theta}\right)^{\mathbf{2}} l"\left(\tilde{\theta}\right)\right\}\left[1+\left(\theta-\tilde{\theta}\right)\frac{p'(\tilde{\theta})}{p(\tilde{\theta})}+\frac{1}{2}\left(\theta-\tilde{\theta}\right)^{\mathbf{2}}\frac{p''(\tilde{\theta})}{p(\tilde{\theta})}+\cdots\right]$$

If *n* is large enough –

$$p(\theta|\boldsymbol{y}) \propto exp\left\{\frac{1}{2}\left(\theta-\tilde{\theta}\right)^{\mathbf{2}} l"\left(\tilde{\theta}\right)\right\}.$$

This is the kernel of a Normal density, centered at $\tilde{\theta}$, and with a variance of $-1/l"\left(\tilde{\theta}\right)$.

- Asymptotically, our Bayes estimator under a zero-one Loss Function will coincide with the MLE.

Bayesian Interval Estimation

- Consider the Bayesian *counterpart* to a Confidence Interval.
- Totally different interpretation – nothing to do with repeated sampling.
- An interval of the form, (θ_L, θ_U), such that $Pr.[(\theta_L < \theta < \theta_U)|\boldsymbol{y}] = p\%$, is a $p\%$ Bayesian Posterior Probability Interval, or a $p\%$ Bayesian Credible Interval for θ.
- Obvious extension to the case where $\boldsymbol{\theta}$ is a *vector*: a $p\%$ Bayesian Posterior Probability *Region*, or a $p\%$ Bayesian Credible *Region* for $\boldsymbol{\theta}$.

- When we construct a (frequentist) Confidence Interval or Region, we usually try to make the interval as short (small) as possible for the desired confidence level.

- We can also consider constructing an "optimal" Bayesian Posterior Probability Region, as follows:

  Given a posterior density function, $p(\boldsymbol{\theta} \mid \boldsymbol{y})$, let A be a subset of the parameter space such that:

  (i) $Pr.[\boldsymbol{\theta} \in A \mid \boldsymbol{y}] = (1 - \alpha)$

  (ii) For all $\boldsymbol{\theta_1} \in A$ and $\boldsymbol{\theta_2} \notin A$, $p(\boldsymbol{\theta_1} \mid \boldsymbol{y}) \geq p(\boldsymbol{\theta_2} \mid \boldsymbol{y})$

then A is a **Highest Posterior Density (HPD) Region** of content $(1 - \alpha)$ for $\boldsymbol{\theta}$.

- For a given α, the HPD has the smallest possible volume.
- If $p(\boldsymbol{\theta} \mid \boldsymbol{y})$ is not uniform over every region of the parameter space, then the HPD is ***unique***.
- A BPI (BCI) can be used in obvious way to test simple null hypotheses, just as a C.I. is used for this purpose by non-Bayesians.
- See more of this later in the course.

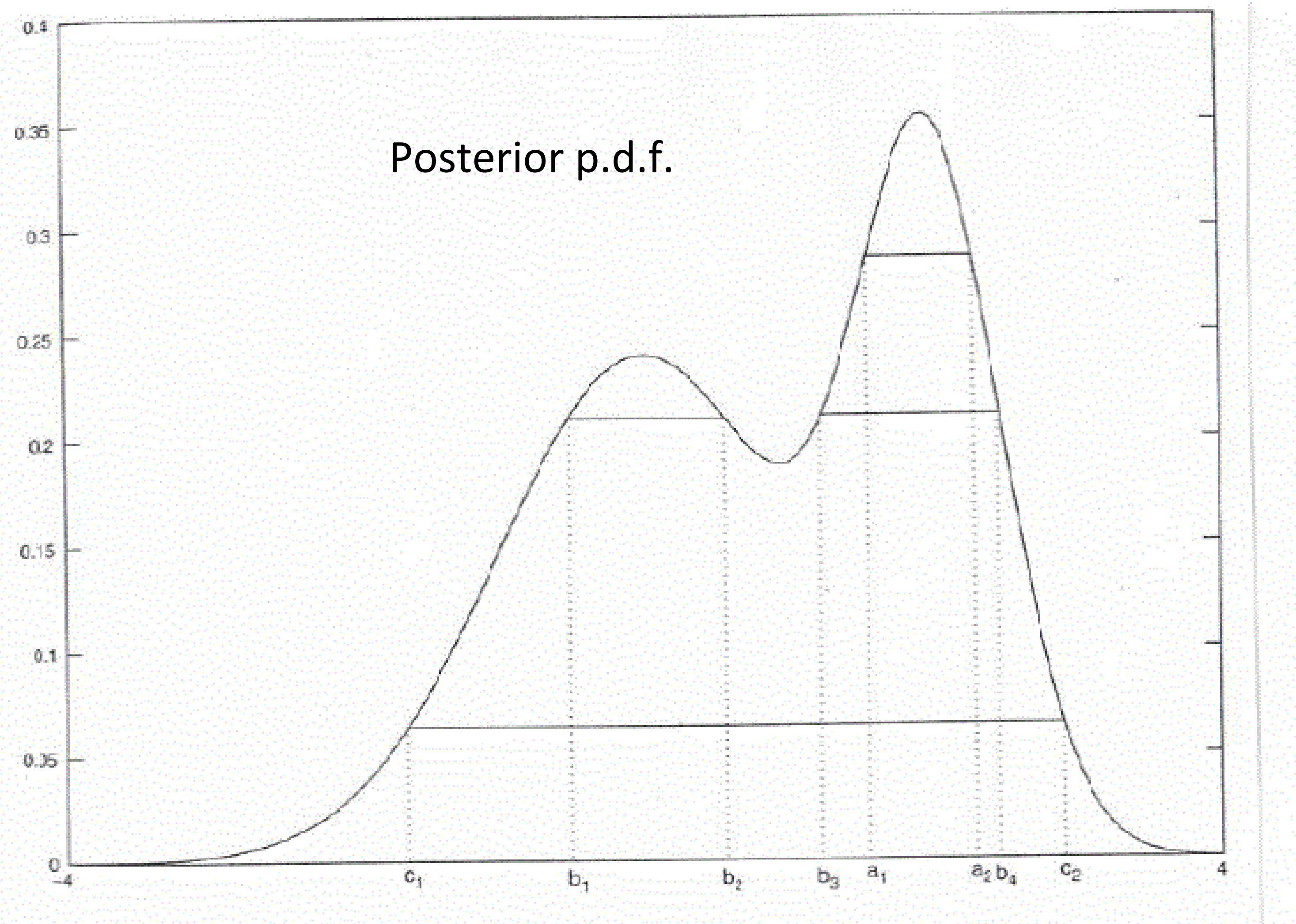

Figure 5.2 — Three different HPD intervals.

$[a_1 \ , \ a_2]$ $[b_1 \ , \ b_2] \cup [b_3 \ , \ b_4]$ $[c_1 \ , \ c_2]$

Example 1

$$y_i \sim N[\mu, \sigma_0^2] \quad ; \quad \sigma_0^2 \quad \text{is } \textit{known}$$

- Before we see the sample of data, we have prior beliefs about value of μ:

$$p(\mu) = p(\mu|\sigma_0^2) \sim N[\bar{\mu}, \bar{v}]$$

That is,

$$p(\mu) \propto exp\left\{-\frac{1}{2\bar{v}}(\mu - \bar{\mu})^2\right\}$$

- Now we take a random sample of data:

$$\boldsymbol{y} = (y_1, y_2, \dots \dots, y_n)$$

- The joint data density (*i.e.*, the Likelihood Function) is:

$$p(\boldsymbol{y}|\mu, \sigma_0^2) = L(\mu|\boldsymbol{y}, \sigma_0^2) \propto \ exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(y_i - \mu)^2\right\}$$

- Bayes' Theorem:

$$p(\mu|\boldsymbol{y}, \sigma_0^2) \propto p(\mu|\sigma_0^2)p(\boldsymbol{y}|\mu, \sigma_0^2)$$

- So,

$$p(\mu|\boldsymbol{y}, \sigma_0^2) \propto exp\left\{-\frac{1}{2}\left[\left(\frac{1}{\bar{v}} + \frac{n}{\sigma_0^2}\right)\left(\mu - \frac{\left(\frac{\bar{\mu}}{\bar{v}} + \frac{n\bar{y}}{\sigma_0^2}\right)}{\left(\frac{1}{\bar{v}} + \frac{n}{\sigma_0^2}\right)}\right)^2\right]\right\}$$

- The Posterior distribution for μ is $N[\bar{\bar{\mu}}, \bar{\bar{v}}]$, where

$$\bar{\bar{\mu}} = \frac{\left(\left(\frac{1}{\bar{v}}\right)\bar{\mu} + \left(\frac{n}{\sigma_0^2}\right)\bar{y}\right)}{\left(\frac{1}{\bar{v}}+\frac{n}{\sigma_0^2}\right)}$$

$$\frac{1}{\bar{\bar{v}}} = \left(\frac{1}{\bar{v}} + \frac{n}{\sigma_0^2}\right)$$

- So, a 95% HPD interval for μ is obtained as follows –

$$Pr.\left[-1.96 < Z < 1.96\right] = 95\%$$

$$Pr.\left[-1.96 < \frac{(\mu - \bar{\bar{\mu}})}{\sqrt{\bar{\bar{v}}}} < 1.96 \mid \boldsymbol{y}\right] = 95\%$$

$$Pr.\left[\bar{\bar{\mu}} - 1.96\sqrt{\bar{\bar{v}}} < \mu < \bar{\bar{\mu}} + 1.96\sqrt{\bar{\bar{v}}} \mid \boldsymbol{y}\right] = 95\%$$

- The 95% HPD interval for μ is the interval

$$\{\, \bar{\bar{\mu}} - 1.96\sqrt{\bar{\bar{\bar{v}}}} \;\; ; \;\; \bar{\bar{\mu}} + 1.96\sqrt{\bar{\bar{\bar{v}}}} \,\}$$

- The (posterior) probability that μ lies in this interval is 95%.

Example 2

- Random sample of *n* observations from an Exponential distribution , so

$$p(y_i \,|\theta) = \theta \exp(-\theta y_i) \quad ; \quad \theta > 0$$

- Prior density for θ is chosen to be Gamma (α , β), where $\alpha, \beta > 0$:

$$p(\theta) \propto \theta^{\alpha-1} \exp(-\frac{\theta}{\beta})$$

Shape Scale

- So, the likelihood function is

$$p(\boldsymbol{y} \,|\theta) = \theta^n \exp(-n\bar{y}\theta)$$

- Bayes' Theorem:

$$p(\theta \mid \boldsymbol{y}) \propto \theta^{n+\alpha-1} \exp\{-(n\bar{y} + \beta^{-1})\theta\}$$

- This is posterior is Gamma, with parameters $(n + \alpha)$ and $1/(n\bar{y} + \beta^{-1})$.

- Another example of *Natural Conjugacy*.

- Bayes point estimators of θ

(i) Quadratic Loss: $\hat{\theta} = (n + \alpha)/\ (n\bar{y} + \beta^{-1})$.

(ii) Zero-One Loss: $\hat{\theta} = (n + \alpha - 1)/\ (n\bar{y} + \beta^{-1})$.

- Suppose that $\alpha = 2$ and $\beta = 4$.
- If n = 2, and $\bar{y} = 0.125$, what is the posterior probability of the BCI , [3.49 , 15.5]?
- In this case the posterior density is Gamma [4 , 2].
- This is the same as a Chi-Square density with 8 degrees of freedom.
- Because Gamma [(v / 2) , 2] = Chi-Square (v)
- $Pr.\left[\chi^2_{(8)} < 3.49\right] = 0.10$, and $.\left[\chi^2_{(8)} < 15.5\right] = 0.95$.
- So, the posterior probability for this BCI is (0.95 – 0.10) = 0.85.
- What are the Bayes point estimates of θ under quadratic and 0-1 losses?