David Giles

Bayesian Econometrics

1. General Background

2. Constructing Prior Distributions

3. Properties of Bayes Estimators and Tests

4. Bayesian Analysis of the Multiple Regression Model

5. Bayesian Model Selection / Averaging

6. Bayesian Computation – Monte Carlo Markov Chain (MCMC)

1. General Background

Major Themes

(i) Flexible & explicit use of Prior Information, together with data, when drawing inferences.

(ii) Use of an explicit Decision-Theoretic basis for inferences.

(iii) Use of Subjective Probability to draw inferences about once-and-for-all events.

(iv) Unified set of principles applied to a wide range of inferential problems.

(v) No reliance on Repeated Sampling.

(vi) No reliance on large-n asymptotics - exact Finite-Sample Results.

(vii) Construction of Estimators, Tests, and Predictions that are "Optimal".

Probability

Definitions

1. Classical, "*a priori*" definition.

2. Long-run relative frequency definition.

3. Subjective, "personalistic" definition:

   (i) Probability is a personal "degree of belief".

   (ii) Formulate using subjective "betting odds".

   (iii) Probability is dependent on the available Information Set.

   (iv) Probabilities are revised (updated) as new information arises.

   (v) Once-and-for-all events can be handled formally.

   (vi) This is what most Bayesians use.

Rev. Thomas Bayes (1702?-1761)

Bayes' Theorem

(i) *Conditional Probability*:

$$p(A|B) = p(A \cap B)/p(B) \qquad p(B|A) = p(A \cap B)/p(A) \qquad (1)$$

$$p(A \cap B) = p(A|B)p(B) = p(B|A)p(A) \qquad (2)$$

(ii) *Bayes' Theorem*:

From (1) and (2): $p(A|B) = p(B|A)p(A)/p(B)$ (3)

(iii) *Law of Total Probability*:

Divide event A into k mutually exclusive & exhaustive events, $\{B_1, B_2, ..., B_k\}$.

Then, $$A = (A \cap B_1) \cup (A \cap B_2) \cup \ldots\ldots\ldots \cup (A \cap B_k) \qquad (4)$$

From (2): $p(A \cap B) = p(A|B)p(B)$

So, from (4):

$$p(A) = p(A|B_1)p(B_1) + p(A|B_2)p(B_2) + \cdots + p(A|B_k)p(B_k)$$

(iv) *Bayes' Theorem Re-stated*:

$$\boldsymbol{p(B_j|A) = p(A|B_j)p(B_j)/\left[\sum_{i=1}^{k}(p(A|B_i)p(B_i))\right]}$$

(v) *Translate This to Inferential Problem*:

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})/p(\boldsymbol{y})$$

or,

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})/\int p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})\, d\boldsymbol{\theta}$$

(**Note:** *multi-dimensional multiple integral*.)

We can write this as:

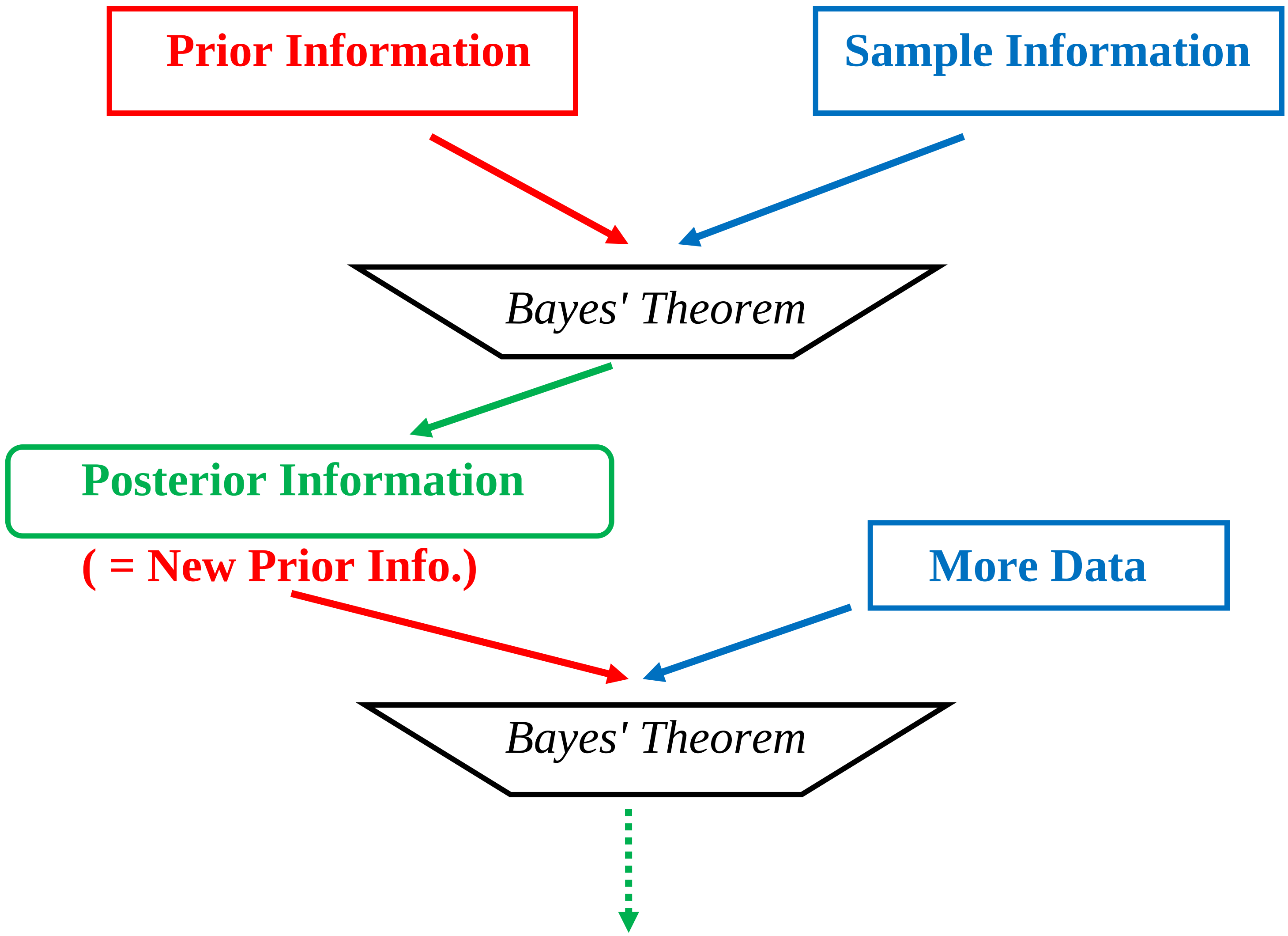
$$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta})$$

or,

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

That is: Posterior Density for $\boldsymbol{\theta}$ $\propto$ Prior Density for $\boldsymbol{\theta}$ $\times$ Likelihood Function

- *Bayes' Theorem* provides us with a way of updating our prior information about the parameters when (new) data information is obtained:



- As we'll see later, the Bayesian updating procedure is strictly "additive".
- If $\int p(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta} = 1,$

  then we can always "recover" the proportionality constant later:

$$\int p(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta} = \int \frac{p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}{\int p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}} = 1$$

- So, the proportionality constant is

$$c = 1/\int p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta} \quad .$$

- This is just a number, but obtaining it requires *multi-dimensional integration*.
- So, even "normalizing" the posterior density may be computationally burdensome.

Example 1

We have data generated by a Bernoulli distribution, and we want to estimate the probability of a "success":

$$p(y) = \theta^{y}(1-\theta)^{1-y} \quad ; \quad y = 0, 1 \quad ; \quad 0 \leq \theta \leq 1$$

Uniform prior for $\theta \in [0\,,1]$:

$$p(\theta) = 1 \quad ; \quad 0 \leq \theta \leq 1$$

(mean & variance?)

$$= 0 \quad ; \quad \text{otherwise}$$

Random sample of *n* observations, so the Likelihood Function is

$$L(\theta|\boldsymbol{y}) = p(\boldsymbol{y}|\theta) = \prod_{i=1}^{n}\left[\theta^{y_i}(1-\theta)^{(1-y_i)}\right] = \theta^{\sum y_i}(1-\theta)^{n-\sum(y_i)}$$

Bayes' Theorem:

$$p(\theta|\boldsymbol{y}) \propto p(\theta)p(\boldsymbol{y}|\theta) \propto \theta^{\sum y_i}(1-\theta)^{n-\sum(y_i)}$$

How would we normalize this posterior density function?

Statistical Decision Theory

- Loss Function - personal in nature.
- $L(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}) \geq 0$, for all $\boldsymbol{\theta}$; $L(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}) = 0$ iff $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}$.
- Note that $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}(\boldsymbol{y})$.
- Examples (*scalar case*; each loss is symmetric)

(i) $L_1(\theta, \hat{\theta}) = c_1(\theta - \hat{\theta})^2$; $c_1 > 0$

(ii) $L_2(\theta, \hat{\theta}) = c_2|\theta - \hat{\theta}|$; $c_2 > 0$

(iii) $L_3(\theta, \hat{\theta}) = c_3$; if $|\theta - \hat{\theta}| > \varepsilon$ $c_3 > 0$; $\varepsilon > 0$

$= 0$; if $|\theta - \hat{\theta}| \leq \varepsilon$ (Typically, $c_3 = 1$)

- Risk Function: Risk = Expected Loss
- $R(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}) = \int L\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}(\boldsymbol{y})\right) p(\boldsymbol{y}|\boldsymbol{\theta}) d\boldsymbol{y}$
- Average is taken over the *sample space*.

- Minimum Expected Loss (MEL) Rule:

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

  (*e.g.*, choose an estimator, select a model/hypothesis)

- Bayes' Rule:

  "Act so as to Minimize Average Risk."

  (Often called the "Bayes' Risk".)

- $r(\hat{\boldsymbol{\theta}}) = \int R\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}\right) p(\boldsymbol{\theta}) d\boldsymbol{\theta}$ (Averaging over the parameter space.)

- Typically, this will be a *multi-dimensional integral.*

- So, the Bayes' Rule implies

$$\hat{\boldsymbol{\theta}} = argmin.\left[r(\hat{\boldsymbol{\theta}})\right] = argmin. \int_{\Omega} R\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}\right) p(\theta) d\boldsymbol{\theta}$$

- Sometimes, the Bayes' Rule is not defined. (If double integral diverges.)

- The MEL Rule is almost always defined.

- Often, the Bayes' Rule and the MEL Rule will be equivalent.

- To see this last result:

$$\hat{\boldsymbol{\theta}} = argmin. \int_{\Omega} [\int_{Y} L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) p(\boldsymbol{y}|\boldsymbol{\theta}) d\boldsymbol{y}] p(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

$$= argmin. \int_{\Omega} [\int_{Y} L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) p(\boldsymbol{\theta}|\boldsymbol{y}) p(\boldsymbol{y}) d\boldsymbol{y}] d\boldsymbol{\theta}$$

$$= argmin. \int_{Y} [\int_{\Omega} L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) p(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta}] p(\boldsymbol{y}) d\boldsymbol{y}$$

- Have to be able to interchange order of integration - Fubini's Theorem.
- Note that $p(\boldsymbol{y}) \geq 0$. So,

$$\hat{\boldsymbol{\theta}} = argmin. \int_{\Omega} L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) p(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta} \; .$$

- That is, $\hat{\boldsymbol{\theta}}$ satisfies the MEL Rule.
- May as well use the MEL Rule in practice.
- What is the MEL (Bayes') estimator under particular Loss Functions?

- Quadratic Loss (L_1): $\hat{\boldsymbol{\theta}} = Posterior\ Mean = E(\boldsymbol{\theta}|\boldsymbol{y})$
- Absolute Error Loss (L_2): $\hat{\boldsymbol{\theta}} = Posterior\ Median$
- Zero-One Loss (L_3): $\hat{\boldsymbol{\theta}} = Posterior\ Mode$
- Consider L_1 as example:
- $min._{\hat{\theta}} \int c_1(\hat{\theta} - \theta)^2 p(\theta|y)d\theta$

$$\Longrightarrow 2c_1 \int (\hat{\theta} - \theta)p(\theta|y)d\theta = 0$$

$$\Longrightarrow \hat{\theta} \int p(\theta|y)d\theta = \int \theta p(\theta|y)d\theta$$

$$\Longrightarrow \hat{\theta} = \int \theta p(\theta|y)d\theta = E(\theta|y)$$

(*Check the 2nd-order condition*)

See CourseSpaces handouts for the cases of L_2 and L_3 Loss Functions.

Example 2

We have data generated by a Bernoulli distribution, and we want to estimate the probability of a "success":

$$p(y) = \theta^{y}(1-\theta)^{1-y} \quad ; \quad y = 0, 1 \quad ; \quad 0 \leq \theta \leq 1$$

(a) *Uniform* prior for ∈ [0 , 1] :

$$p(\theta) = 1 \quad ; \quad 0 \leq \theta \leq 1 \qquad \text{(mean = 1/2; variance = 1/12)}$$

$$= 0 \quad ; \quad \text{otherwise}$$

Random sample of *n* observations, so the Likelihood Function is

$$L(\theta|\boldsymbol{y}) = p(\boldsymbol{y}|\theta) = \prod_{i=1}^{n}\left[\theta^{y_i}(1-\theta)^{(1-y_i)}\right] = \theta^{\sum y_i}(1-\theta)^{n-\sum(y_i)}$$

Bayes' Theorem:

$$p(\theta|\boldsymbol{y}) \propto p(\theta)p(\boldsymbol{y}|\theta) \propto \theta^{\sum y_i}(1-\theta)^{n-\sum(y_i)} \qquad \textbf{(*)}$$

Apart from the normalizing constant, this is a Beta p.d.f.:

- A continuous random variable, X, has a **Beta distribution** if its density function is

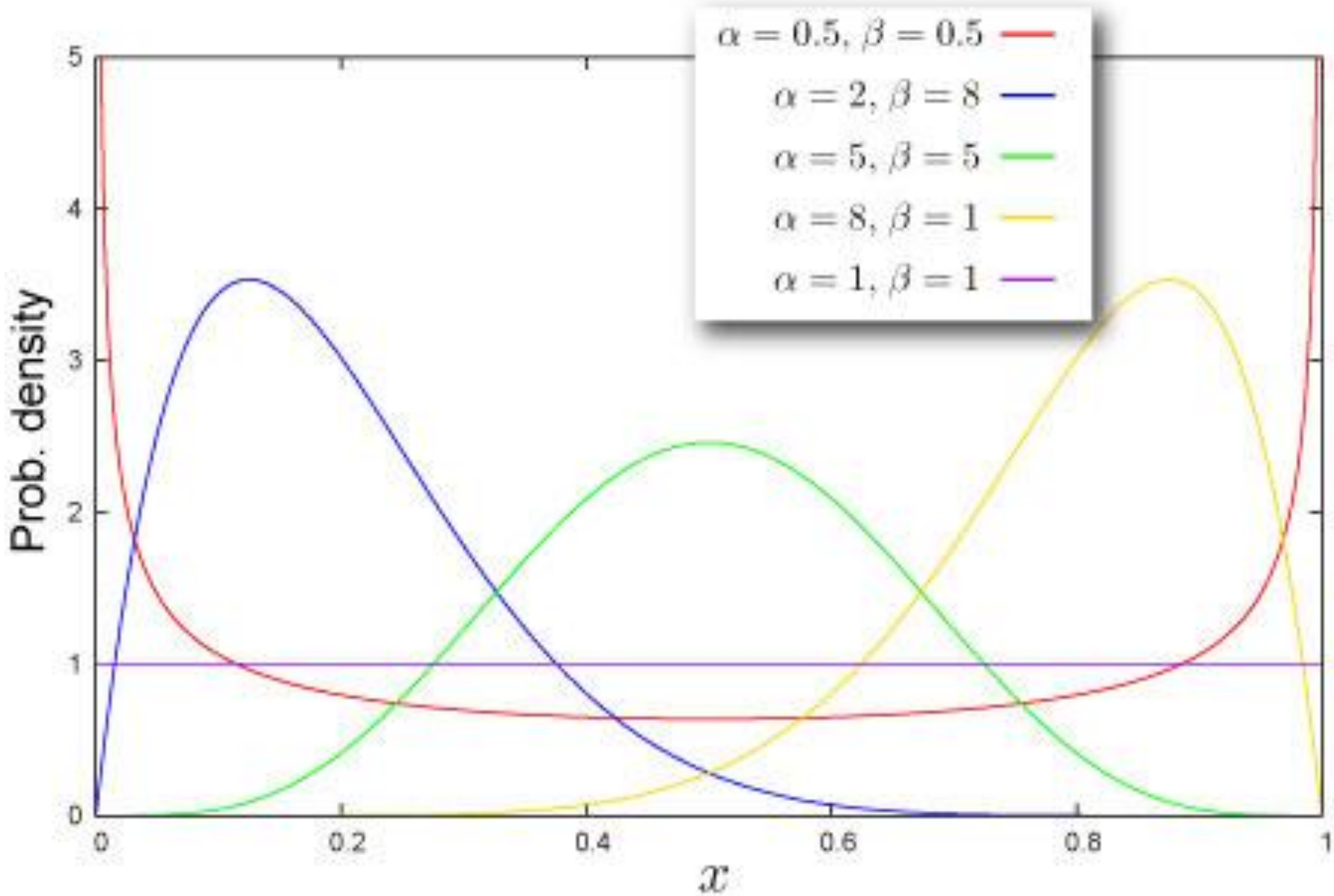
$$p(x|\alpha,\beta) = \frac{1}{B(\alpha,\beta)} x^{(\alpha-1)}(1-x)^{(\beta-1)} \quad ; \; 0 < x < 1 \; ; \quad \alpha,\beta > 0$$

- Here, $B(\alpha,\beta) = \int_0^1 t^{\alpha}(1-t)^{\beta-1}dt$ is the "Beta Function".
- We can write it as: $B(\alpha,\beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$

  where $\Gamma(t) = \int_0^{\infty} x^{t-1}e^{-x}dx$ is the "Gamma Function"
- See the handout, "Gamma and Beta Functions".
- $E[X] = \alpha/(\alpha+\beta)$; $V[X] = \alpha\beta/[(\alpha+\beta)^2(\alpha+\beta+1)]$

  Mode $[X] = (\alpha-1)/(\alpha+\beta-2)$; if $\alpha,\beta > 1$

  Median $[X] \cong (\alpha-1/3)/(\alpha+\beta-2/3)$; if $\alpha,\beta > 1$
- Can extend to have support [a , b], and by adding other parameters.
- It's a very flexible and useful distribution, with a *finite support*.


$\alpha = 0.5, \beta = 0.5$
$\alpha = 2, \beta = 8$
$\alpha = 5, \beta = 5$
$\alpha = 8, \beta = 1$
$\alpha = 1, \beta = 1$
Prob. density
x
0
0.2
0.4
0.6
0.8
1
1
2
3
4
5

- Going back to Slide 14:

$$p(\theta|\boldsymbol{y}) \propto p(\theta)p(\boldsymbol{y}|\theta) \propto \theta^{\sum y_i}(1-\theta)^{n-\sum(y_i)} \quad ; \; 0 < \theta < 1 \qquad \textbf{(*)}$$

Or,

$$p(\theta|\boldsymbol{y}) = \frac{1}{B(\alpha,\beta)}p(\theta)p(\boldsymbol{y}|\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}; \; 0 < \theta < 1$$

where $\alpha = n\bar{y} + 1 \;\; ; \;\; \beta = n(1-\bar{y}) + 1$

- So, under various loss functions, we have the following Bayes' estimators:

(i) L_1: $\hat{\theta}_1 = E[\theta|\boldsymbol{y}] = \frac{\alpha}{\alpha+\beta} = (n\bar{y}+1)/(n+2)$

(ii) L_3: $\hat{\theta}_3 = Mode[\theta|\boldsymbol{y}] = \frac{(\alpha-1)}{(\alpha+\beta-2)} = \bar{y}$

- Note that the MLE is $\tilde{\theta} = \bar{y}$.   (The Bayes' estimator under zero-one loss.)

- The MLE ($\hat{\theta}_3$) is *unbiased* (because $E[y] = \theta$); consistent, and BAN.

- We can write: $\hat{\theta}_1 = (\bar{y} + 1/n)/(1 + 2/n)$

  So, $\lim_{(n\to\infty)} \left(\hat{\theta}_1\right) = \bar{y}$; and $\hat{\theta}_1$ is also consistent and BAN.

(b) *Beta* prior for ∈ [0 , 1] :

$$p(\theta) \propto \theta^{a-1}(1-\theta)^{b-1} \quad ; \quad a, b > 0$$

- Recall that Uniform [0 , 1] is a special case of this.

- We can assign values to *a* and *b* to reflect prior beliefs about θ .

- *e.g.*, $E[\theta] = a/(a + b)$; $V[\theta] = ab/[(a + b)^2(a + b + 1)]$

- Set values for $E[\theta]$ and $V[\theta]$, and then solve for implied values of a and b.

- Recall that the Likelihood Function is:

$$L(\theta|\boldsymbol{y}) = p(\boldsymbol{y}|\theta) = \theta^{n\bar{y}}(1-\theta)^{n(1-\bar{y})}$$

- So, applying Bayes' Theorem:

"kernel" of the density

$$p(\theta|\boldsymbol{y}) \propto p(\theta)p(\boldsymbol{y}|\theta) \propto \theta^{n\bar{y}+a-1}(1-\theta)^{n(1-\bar{y})+b-1}$$

- This is a Beta density, with $\alpha = n\bar{y} + a \;\; ; \;\; \beta = n(1-\bar{y}) + b$

- An example of "Natural Conjugacy"

- So, under various loss functions, we have the following Bayes' estimators:

(i) L_1: $\hat{\theta}_1 = E[\theta|\boldsymbol{y}] = \frac{\alpha}{\alpha+\beta} = (n\bar{y}+a)/(n+a+b)$

(ii) L_3: $\hat{\theta}_3 = Mode[\theta|\boldsymbol{y}] = \frac{(\alpha-1)}{(\alpha+\beta-2)} = (n\bar{y} + a - 1)/(a + b + n - 2)$

- Recall that the MLE is $\tilde{\theta} = \bar{y}$.

- We can write: $\hat{\theta}_1 = (\bar{y} + a/n)/(1 + a/n + b/n)$

  So, $\lim_{(n\to\infty)}(\hat{\theta}_1) = \bar{y}$; and $\hat{\theta}_1$ is consistent and BAN.

- We can write: $\hat{\theta}_3 = (\bar{y} + a/n - 1/n)/(1 + a/n + b/n - 2/n)$

  So, $\lim_{(n\to\infty)}(\hat{\theta}_3) = \bar{y}$; and $\hat{\theta}_3$ is consistent and BAN.

- In the various examples so far, the Bayes' estimators have had the same asymptotic properties as the MLE.

- We'll see later that *this result holds in general* for Bayes' estimators.

Example 3

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

- 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,

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

"*kernel*" *of the density*

- Note: *we are not saying that μ is random*. Just uncertain about its value.

- 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:

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

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

"kernel" of the density

- Now we'll apply Bayes' Theorem:

Likelihood function

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

- So,

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

We can write:

$$\sum_{i=1}^{n}(y_i-\mu)^2=\sum_{i=1}^{n}(y_i-\bar{y})^2+n(\bar{y}-\mu)^2$$

So,

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

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

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

- "Complete the square":

$$ax^2 + bx + c = a(x + \frac{b}{2a})^2 + (c - \frac{b^2}{4a})$$

- In our case,

$x = \mu$

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

$b = -2\left(\frac{\bar{\mu}}{\bar{v}} + \frac{n\bar{x}}{\sigma_0^2}\right)$

$c = 0$

- So, we have:

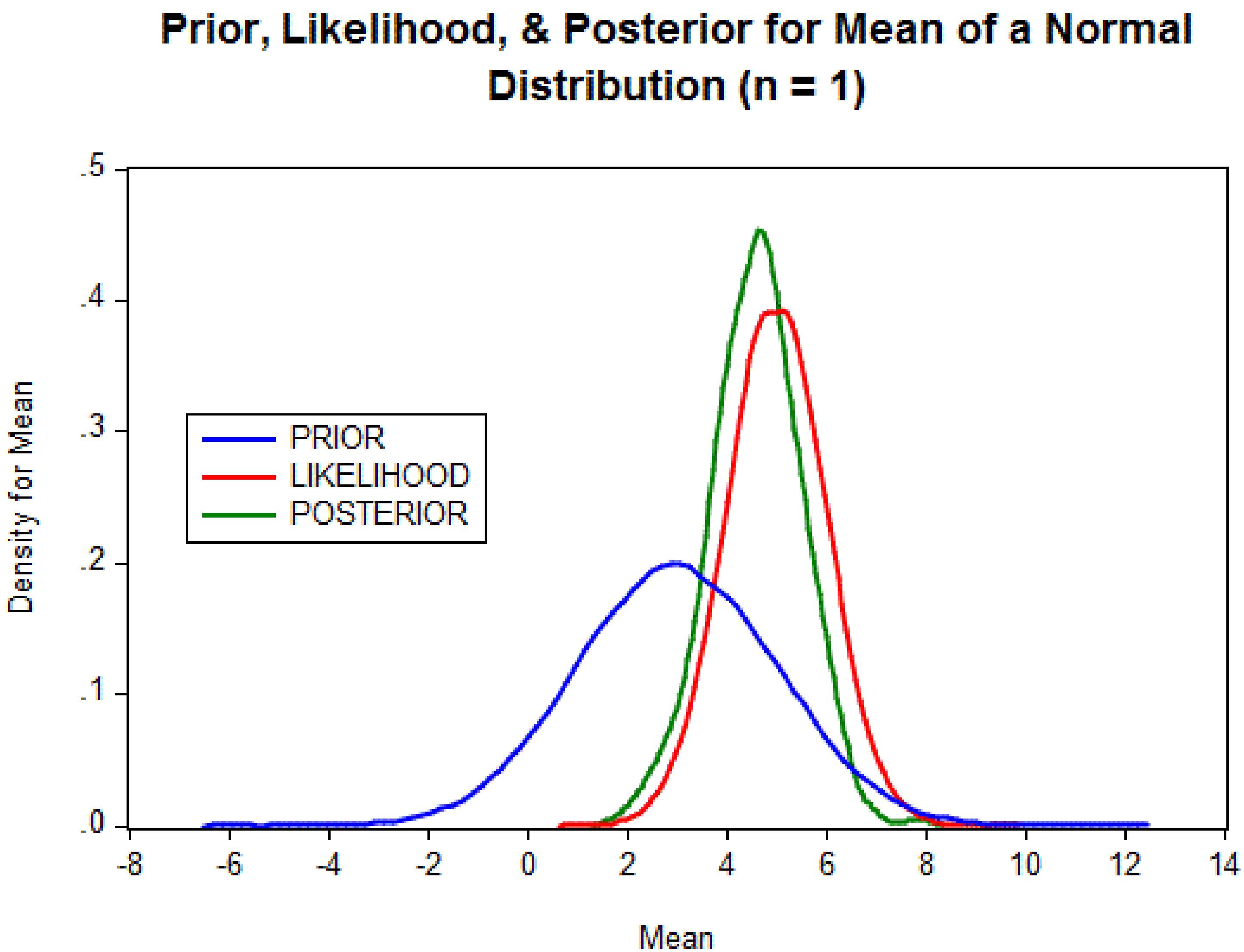
$$p(\mu|\boldsymbol{y},\sigma_0^2) \propto exp\left\{-\frac{1}{2}\left[\left(\frac{1}{\bar{v}}+\frac{n}{\sigma_0^2}\right)\left(\mu-\frac{\left(\frac{\bar{\mu}}{\bar{v}}+\frac{n\bar{y}^2}{\sigma_0^2}\right)}{\left(\frac{1}{\bar{v}}+\frac{n}{\sigma_0^2}\right)}\right)^2\right]\right\}$$

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

$$\bar{\bar{\mu}} = \frac{\left(\left(\frac{1}{\bar{v}}\right)\bar{\mu} + \left(\frac{n}{\sigma_0^2}\right)\bar{y}\right)}{\left(\frac{1}{\bar{v}}+\frac{n}{\sigma_0^2}\right)} \quad ; \qquad \textit{weighted average interpretation?}$$

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

- $\bar{\bar{\mu}}$ is the **MEL (Bayes) estimator** of μ under loss functions L_1, L_2, L_3.
- Another example of "Natural Conjugacy".
- Show that $\lim_{(n\to\infty)}(\bar{\bar{\mu}}) = \bar{y}$ (= MLE: *consistent, BAN*)


Prior, Likelihood, & Posterior for Mean of a Normal Distribution (n = 1)
.5
.4
.3
.2
.1
.0
-8
-6
-4
-2
0
2
4
6
8
10
12
14
Density for Mean
Mean
PRIOR
LIKELIHOOD
POSTERIOR

Dealing With "Nuisance Parameters"

- Typically, $\boldsymbol{\theta}$ is a vector: $\boldsymbol{\theta}' = (\theta_1, \theta_2, \dots \dots, \theta_p)$

  e.g., $\boldsymbol{\theta}' = (\beta_1, \beta_2, \dots \dots, \beta_k;\ \sigma)$

- Often interested primarily in only some (one?) of the parameters. The rest are "nuisance parameters".

- However, they can't be ignored when we draw inferences about parameters of interest.

- Bayesians deal with this in a very flexible way - they construct the **Joint Posterior density** for the full $\boldsymbol{\theta}$ vector; and then they *marginalize* this density by integrating out the nuisance parameters.

- For instance:

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = p(\boldsymbol{\beta}, \sigma|\boldsymbol{y}) \propto p(\boldsymbol{\beta}, \sigma) L(\boldsymbol{\beta}, \sigma|\boldsymbol{y})$$

  Then,

$$p(\boldsymbol{\beta}|\boldsymbol{y}) = \int_0^\infty p(\boldsymbol{\beta},\sigma|\boldsymbol{y})d\sigma \qquad \text{(\textit{marginal posterior for} } \boldsymbol{\beta}\text{)}$$

$$p(\beta_i|\boldsymbol{y}) = \int\int \dots . \int p(\boldsymbol{\beta}|\boldsymbol{y})d\beta_1 \dots . d\beta_{i-1}d\beta_{i+1} \dots . d\beta_k$$

(*marginal posterior for* β_i)

$$p(\sigma|y) = \int p(\boldsymbol{\beta},\sigma|\boldsymbol{y})d\boldsymbol{\beta} = \int\int \dots . \int p(\boldsymbol{\beta},\sigma|\boldsymbol{y})d\beta_1 \dots . d\beta_k$$

(*marginal posterior for* σ)

- Computational issue - can the integrals be evaluated analytically?
- If not, we'll need to use numerical procedures.
- Standard numerical "quadrature" infeasible if $p > 3, 4$.
- Need to use alternative methods for obtaining the Marginal Posterior densities from the Joint Posterior density.

- Once we have a Marginal Posterior density, such as $p(\beta_i|\boldsymbol{y})$, we can obtain a Bayes point estimator for β_i.
- For example:

$$\widehat{\beta_i} = E[\beta_i|\boldsymbol{y}] = \int_{-\infty}^{\infty} \beta_i p(\beta_i|\boldsymbol{y}) d\beta_i$$

$$\widetilde{\beta_i} = argmax.\{p(\widetilde{\beta_i}|\boldsymbol{y})\}$$

$$\breve{\beta_i} = M, \text{ such that } \int_{-\infty}^{M} p(\beta_i|\boldsymbol{y}) d\beta_i = 0.5$$

- The "posterior uncertainty" about β_i can be measured, for instance, by computing

$$var.(\beta_i|\boldsymbol{y}) = \int_{-\infty}^{\infty} \left(\beta_i - \widehat{\beta_i}\right)^2 p(\beta_i|\boldsymbol{y}) d\beta_i$$

Bayesian Updating

- Suppose that we apply our Bayesian analysis using a first sample of data, $\boldsymbol{y}_1$:

$$p(\boldsymbol{\theta}|\boldsymbol{y}_1) \propto p(\boldsymbol{\theta})p(\boldsymbol{y}_1|\boldsymbol{\theta})$$

- Then we obtain a new *independent* sample of data, $\boldsymbol{y}_2$.
- The previous Posterior becomes our new Prior for $\boldsymbol{\theta}$ and we apply Bayes' Theorem again:

$$\begin{aligned} p(\boldsymbol{\theta}|\boldsymbol{y}_1, \boldsymbol{y}_2) &\propto p(\boldsymbol{\theta}|\boldsymbol{y}_1)p(\boldsymbol{y}_2|\boldsymbol{\theta}, \boldsymbol{y}_1) \\ &\propto p(\boldsymbol{\theta})p(\boldsymbol{y}_1|\boldsymbol{\theta})p(\boldsymbol{y}_2|\boldsymbol{\theta}, \boldsymbol{y}_1) \\ &\propto p(\boldsymbol{\theta})p(\boldsymbol{y}_1, \boldsymbol{y}_2|\boldsymbol{\theta}) \end{aligned}$$

- The Posterior is the same as if we got both samples of data at once.
- The Bayes updating procedure is "additive" with respect to the available information set.

Advantages & Disadvantages of Bayesian Approach

1. *Unity of approach* regardless of inference problem: Prior, Likelihood function, Bayes' Theorem, Loss function, MEL rule.
2. Prior information (if any) incorporated *explicitly* and *flexibly*.
3. Explicit use of a *Loss function* and *Decision Theory*.
4. *Nuisance parameters* can be handled easily - not ignored.
5. Decision rules (estimators, tests) are ***Admissible***.
6. *Small samples* are alright (even $n = 1$).
7. *Good asymptotic properties* - same as MLE.

<u>However</u>:

1. Difficulty / cost of *specifying the Prior* for the parameters.
2. Results may be *sensitive to choice of Prior* - need robustness checks.
3. Possible *computational issues*.