6.7 Bayesian Point Estimation

In their comprehensive treatise, Bernardo and Smith (1994, *Bayesian Theory*, p. 4) offer the following summarization of *Bayesian statistics*:

Bayesian Statistics offers a rationalist theory of *personalistic* beliefs in contexts of uncertainty, with the central aim of characterizing how an individual *should* act in order to avoid certain kinds of undesirable behavioral inconsistencies.
• The theory establishes *expected utility maximization* as the basis for rational decision making and *Bayes’ Theorem* as the key to the way beliefs should fit together in the light of changing evidence.

• The goal is to establish rules and procedures for individuals concerned with *disciplined uncertainty accounting*.

  ◦ The theory is *not* descriptive in the sense of claiming to model actual behavior.

  ◦ Rather, it is *prescriptive*, in the sense of saying “if you wish to avoid the possibility of these undesirable consequences you must act in the following way.”
Table 1: Earliest References in JSTOR to “Bayes” or “Bayesian”

<table>
<thead>
<tr>
<th>Discipline</th>
<th>First</th>
<th>Fifth</th>
<th>Tenth</th>
<th>Hundredth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economics</td>
<td>1925</td>
<td>1941</td>
<td>1949</td>
<td>1970</td>
</tr>
<tr>
<td>General Science</td>
<td>1763</td>
<td>1922</td>
<td>1934</td>
<td>1982</td>
</tr>
<tr>
<td>Philosophy</td>
<td>1884</td>
<td>1937</td>
<td>1940</td>
<td>1971</td>
</tr>
<tr>
<td>Statistics</td>
<td>1907</td>
<td>1918</td>
<td>1921</td>
<td>1951</td>
</tr>
</tbody>
</table>

Note: Use of “Bayes” must refer to Thomas Bayes.
### Table 2: Earliest References in JSTOR to “MCMC”

<table>
<thead>
<tr>
<th>Discipline</th>
<th>First</th>
<th>Fifth</th>
<th>Tenth</th>
<th>Hundredth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economics</td>
<td>1993</td>
<td>1994</td>
<td>1996</td>
<td>2005</td>
</tr>
<tr>
<td>Finance</td>
<td>1998</td>
<td>2001</td>
<td>2003</td>
<td></td>
</tr>
<tr>
<td>Philosophy</td>
<td>2010</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Political Science</td>
<td>1998</td>
<td>2000</td>
<td>2001</td>
<td>2010</td>
</tr>
<tr>
<td>Psychology</td>
<td>2005</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sociology</td>
<td>1997</td>
<td>2002</td>
<td>2004</td>
<td></td>
</tr>
<tr>
<td>Zoology</td>
<td>2000</td>
<td>2002</td>
<td>2004</td>
<td>2010</td>
</tr>
</tbody>
</table>

**Note:** Use of “MCMC” must refer to Markov Chain Monte Carlo.
Alan Greenspan, AEA Meetings, January 3, 2004:

“... As a consequence, the conduct of monetary policy in the United States has come to involve, at its core, crucial elements of risk management. This conceptual framework emphasizes understanding as much as possible the many sources of risk and uncertainty that policymakers face, quantifying those risks when possible, and assessing the costs associated with each of the risks. In essence, the risk management approach to monetary policymaking is an application of Bayesian decision-making. ... Our problem is not, as is sometimes alleged, the complexity of our policymaking process, but the far greater complexity of a world economy whose underlying linkages appear to be continuously evolving. Our response to that continuous evolution has been disciplined by the Bayesian type of decision-making in which we have engaged. ...”
Bayesian Impact

Poirier (2006, *Bayesian Analysis* 1, 969-980)

http://ba.stat.cmu.edu/journal/2006/vol01/issue04/poirier.pdf
Bayesian Analysis (2006) 1, Number 4, pp. 969–980

The Growth of Bayesian Methods in Statistics and Economics Since 1970

Dale J. Poirier

Abstract. To measure the impact of Bayesian reasoning, this paper investigates the occurrence of two words, “Bayes” and “Bayesian,” since 1970 in journal articles in a variety of disciplines, with a focus on economics and statistics. The growth in statistics is documented, but the growth in economics is largely confined to economic theory/mathematical economics rather than econometrics.

Keywords: Bayesian impact, Journals
Figure 1: Various JSTOR Disciplines
Figure 2: Statistics Journals Containing “Bayer” or “Bayesian”: More Theoretical
(b) Less Theoretical Statistics Journals

American Statistician

Applied Statistician

Biometrics

International Statistics Review

JRSSA

Statistician
Figure 3: Economics Journals Containing “Bayes” or “Bayesian”

(a) All-Purpose Economics Journals

Econometrica

International Economic Review

Year

Year

Proportion

Proportion

Review of Economic Statistics

Review of Economic Studies

Year

Year

Proportion

Proportion

(b) Econometrica

Econometrics

Economic Theory / Math Econ

Year

Year

Proportion

Proportion
(c) Econometrics Journals

Journal of Applied Econometrics

Proportion

Year

JBES

Proportion

Year

Journal of Econometrics

Proportion

Year

Econometric Theory

Proportion

Year
(d) Non-Econometrics Journals

**American Economic Review**

**Economic Journal**

**Economics Letters**

**Journal of Economic Literature**

**Journal of Economic Perspectives**

**Journal of Human Resources**
(d) Non-Econometrics Journals (continued)

Journal of Labor Economics

Quarterly Journal of Economics

Journal of Political Economy

RAND Journal of Economics
http://www.youtube.com/watch?v=TbV15eUvhM4
In 1891 a Scottish mathematician, George Chrystal, commented on Laplace’s methods as follows:

“The laws of ... Inverse Probability being dead, they should be decently buried out of sight, and not embalmed in text books and examination papers ... The indiscretions of great men should be quietly allowed to be forgotten.”
“... If Bayes’ story were a TV melodrama, it would need a clear-cut villain, and Fisher would probably be the audience’s choice by acclamation. ... Even with thick glasses he could barely see three feet and had to be rescued from oncoming buses. His clothes were so rumpled that his family thought he looked like a tramp; he smoked a pipe even while swimming; and if a conversation bored him, he sometimes removed his false teeth and cleaned them in public.”
“It is difficult for me to tone down the missionary zeal acquired in youth, but perhaps the good battle is justified since there are still many heathens.”

I. J. Good (1976)
Recall:

- According to the subjective interpretation, probability is a property of an individual’s *perception of reality*, rather than a property of reality itself.

- There are no “true unknown probabilities” in the world to be discovered. Probability is in the eye of the beholder. *“Probability does not exist.”*

- Supporters of subjective probability and decision-making based on maximization of expected utility, see the matter as a *normative* proposition, i.e., “rational” people *ought* to behave in this manner.
The probabilistic basis in Section 6.3 for what constituted “good” estimates was couched in terms of the *procedures* (estimators) used to generate the estimates, and how these procedures perform on average in repeated samples.

- This “averaging” involved expectation over the *entire* sample space.

- According to frequentists, the criteria for deciding what constitutes a desirable estimate, involves not just the observed data, but also all the data that could have been observed.

- This viewpoint is in direct conflict with the Likelihood Principal.
Lancaster (2004, p. 8): Bayesian inference is not “objective.” ... The typical seminar in our subject is an exercise in persuasion in which the speaker announces her beliefs in the form of a model containing and accompanied by a set of assumptions, these being additional (tentative) beliefs. She attempts to persuade her audience of the reasonableness of these beliefs by showing that some embody “rational” behavior by the agents she is discussing and promising that other beliefs are shown by the evidence to be not inconsistent with the data. She presents her results and shows how some of her beliefs seem to be true and others false and in need of change. The entire process is subjective and personal. All that a Bayesian can contribute to this is to ensure that the way in which she revises her beliefs conforms to the laws of probability, in particular, Bayes’ Theorem.
• From the *subjective Bayesian* viewpoint both the likelihood and the prior are *subjective concepts* and a *subjective interpretation of probability* is adopted.

  ◦ It is possible to adopt the so-called *objective Bayesian* viewpoint in which the prior is chosen by some “objective” means.

  ◦ For a comparison of the two viewpoints, see Berger (2006), Goldstein (2006), and the comments by eight discussants.

  
  http://ba.stat.cmu.edu/vol01is03.php
• The meaningful distinction between the “prior” and the likelihood is that the latter represents a *window* for viewing the observable world shared by a group of researchers who agree to disagree in terms of possibly different priors.

• Poirier (1988, *JEP*) introduced the metaphor “window” for a likelihood function because *a window captures the essential role played by the likelihood, namely, as a parametric medium for viewing the observable world.*
• In the context of de Finetti’s Representation Theorem 5.8.1, such *intersubjective agreement* (objectivity?) over the window (likelihood) may result from agreed upon symmetries in probabilistic beliefs concerning observables.

• Disagreements over priors are addressed by sensitivity analysis evaluating the robustness of posterior quantities of interest to changes in the prior.

• Bayesian treat all quantities, both observable and unobservable, as random variables.
According to both the logical and the subjective interpretations of probability, it is meaningful to use probability to describe degrees of belief about unknown entities such as a K-dimensional parameter vector \( \theta \in \Theta \).

Before observing the data \( y \), these beliefs are described by a pdf \( f(\theta) \) known as the **prior density**.

For any value of \( \theta \), the data are viewed as arising from the pdf \( f(y|\theta) \), which when viewed as a function of \( \theta \) given \( y \), is known as the **likelihood function** \( L(\theta; y) \).
Components of Bayesian Analysis

- **Prior:** \( f(\theta) \)

- **Likelihood (window):** \( \mathcal{L}(\theta; y) \)

- **Bayes Theorem:** \( f(\theta|y) \propto f(\theta) \mathcal{L}(\theta; y) \)
  Interval estimation follows directly from \( f(\theta|y) \).

- **Loss functions:** \( C(\hat{\theta}, \theta), C(d, \theta), \) or \( C(\hat{y}_*, y_*) \)

- **Commandment:** minimize expected posterior loss
  \[
  \min \hat{\theta} \quad E_{\theta|y}[C(\hat{\theta}, \theta)] \\
  \min d \quad E_{\theta|y}[C(d, \theta)] \\
  \min \hat{y}_* \quad E_{y_*|y}[C(\hat{y}_*, y_*)]
  \]

- **Prediction of future** \( y_* \):
  \[
  f(y_*|y) = E_{\theta|y}[f(y_*|y, \theta)] = \int f(y_*|y, \theta) f(\theta|y) \, d\theta
  \]

Sensitivity analysis w.r.t. prior, loss, likelihood.
• The major disagreement between Bayesians and frequentists is whether to minimize $E_{\theta|y}[C(\hat{\theta}, \theta)]$ or $E_{Y|\theta}[C(\bar{\theta}, \theta)]$, i.e.,

$$\theta \mid Y = y \text{ versus } Y \mid \theta.$$  

*This choice underlies most statistical debates.*
Manipulating density functions according to the rules of probability yields \textit{Bayes Theorem} for densities:

\[
f(\theta|y) = \frac{f(\theta, y)}{f(y)} = \frac{f(\theta) \mathcal{L}(\theta; y)}{f(y)} \quad (6.7.1)
\]

\[\propto f(\theta) \mathcal{L}(\theta; y),\]

where \(f(\theta|y)\) is the \textit{posterior} density, and the denominator in (6.7.1), the \textit{marginal density of the data} or the \textit{marginal likelihood}, is

\[
f(y) = E_\theta[\mathcal{L}(\theta; y)] = \int \mathcal{L}(\theta; y) f(\theta) \, d\theta, \quad (6.7.2)
\]

and does not depend on \(\theta\). The corresponding prior and posterior cdfs are denoted \(F(\theta)\) and \(F(\theta|y)\).
Bayesian analysis is a learning process in that prior beliefs $f(\theta)$ are updated by (6.7.1) in light of data $y$ to obtain the posterior beliefs summarized by $f(\theta | y)$.

- The posterior distribution summarizes beliefs about $\theta$ given the observed data $Y = y$.

- The distribution $\theta | y$ is in sharp contrast to the sampling distribution $Y | \theta = \theta_0$.

- Use of posterior density (6.7.1) as a basis for estimation and inference is in agreement with the LP.
• **Bayesian inference** involves updating prior beliefs into posterior beliefs conditional on observed data.

  ○ It requires only a few general principles that are applied over and over again in different settings.

  ○ Bayesians begin by specifying a joint distribution of all quantities under consideration (both observable and unobservable) except known constants.

  ○ **Bayesianism** reduces statistical inference to applied probability.
Before discussing *Bayesian point estimation*, it is useful to consider some examples of posterior density (6.7.1). The ease with which Bayesian analysis can be applied is greatly enhanced when there exists *conjugate prior distributions* as now loosely defined.

**Definition 6.7.1:** Given the likelihood $\mathcal{L}(\theta; y)$, if there exists a family of prior distributions with the property that no matter the data actually observed, the posterior distribution $f(\theta|y)$ is also member of this family, then this family of prior distributions is called a *conjugate family*, and any member of the family is said to be a *conjugate prior*. A conjugate prior that has the same functional form as the likelihood function is called a *natural conjugate prior*. 
• The class of all distributions is trivially a conjugate family for all likelihood functions.

  ○ For a given likelihood there may not exist a nontrivial conjugate family.

  ○ The existence of a sufficient statistic of fixed dimension independent of sample size, however, insures the existence of a non-trivial conjugate family.

• For example, natural conjugate priors are available whenever the likelihood function belongs to the exponential family, as well as in other cases such as $U[0, \theta]$ (see Exercise 6.7.6).
Conjugate priors are computationally attractive because usually it is not necessary to compute the K-dimensional integral

\[ f(y) = E_\theta[\mathcal{L}(\theta; y)] = \int_{\mathbb{R}^K} \mathcal{L}(\theta; y) f(\theta) \, d\theta, \quad (6.7.2) \]

which serves as an \textit{integrating constant} to insure that

\[ f(\theta|y) = \frac{f(\theta, y)}{f(y)} = \frac{f(\theta) \mathcal{L}(\theta; y)}{f(y)} \quad (6.7.1) \]

integrates to unity. Since the posterior belongs to the same family as the prior, such knowledge can usually be used to appropriately scale the numerator in (6.7.1) without any explicit integration.
• Analytical expressions for moments are often readily available.

• When conjugate families exist they are quite large.

  ◦ Mixtures of conjugate priors are also conjugate priors because the corresponding posterior distributions is a mixture of the component posteriors.

  ◦ Although mixture distributions tend to be “parameter-rich,” they are capable of representing a wide range of prior beliefs (e.g., multimodal priors) while still enjoying the computational advantages of the component conjugate priors.
Diaconis and Ylvisaker (1979) show that the characteristic feature of conjugate priors for exponential family likelihoods is that the resulting posterior means are linear in the observations, as will become clear in the examples that follow.

Natural conjugate priors are especially attractive in situations where it is desired to interpret the prior information as arising from a fictitious sample from the same underlying population that gave rise to the likelihood.
Example 6.7.1: Given $\theta$, where $0 < \theta < 1$, consider $T$ i.i.d. Bernoulli variables $Y_t \ (t = 1, 2, \ldots, T)$ with p.m.f.:

$$f(y_t | \theta) = \begin{cases} 
\theta, & \text{if } y_t = 1 \\
1 - \theta, & \text{if } y_t = 0
\end{cases}.$$  

(6.7.3)

- **de Finetti’s Representation Theorem** implies the problem has the alternative interpretation in which subjective beliefs over arbitrarily long sequences of the $Y_t$'s are assumed exchangeable and coherent.

- The likelihood function (i.e., parametric window) through which the observables are viewed is

$$\mathcal{L}(\theta; y) = \theta^m (1 - \theta)^{T-m},$$  

(6.7.4)

where $m = T \overline{y}$ is the number of successes (i.e., $y_t = 1$) in $T$ trials.
Suppose that prior beliefs concerning $\theta$ are represented by a beta distribution (Definition 3.3.8) with density

$$f(\theta | \alpha, \delta) = \frac{1}{B(\alpha, \delta)} \theta^{\alpha-1} (1 - \theta)^{\delta-1}, \quad 0 < \theta < 1,$$

where $\alpha > 0$ and $\delta > 0$ are known constants and

$$B(\alpha, \delta) \equiv \frac{\Gamma(\alpha) \Gamma(\delta)}{\Gamma(\alpha + \delta)} \quad (3.3.44)$$

Given the variety of shapes the beta p.d.f. can assume (see Figure 3.3.4), this class of priors can represent a wide range of prior opinions.
The denominator (6.7.2) of posterior density (6.7.1) is easy to compute in this instance. Define

\[ \bar{\alpha} = \alpha + m, \quad (6.7.6) \]

\[ \bar{\delta} = \delta + T - m, \quad (6.7.7) \]

and consider
\[ f(y) = \int_{0}^{1} \left[ B(\alpha, \delta) \right]^{-1} \theta^{\alpha-1} (1 - \theta)^{\delta-1} \theta^{m} (1 - \theta)^{T-m} \, d\theta \]

\[ = \left[ \frac{B(\bar{\alpha}, \delta)}{B(\alpha, \delta)} \right] \int_{0}^{1} [B(\bar{\alpha}, \delta)]^{-1} \theta^{\bar{\alpha}-1} (1 - \theta)^{\delta-1} \, d\theta \]

\[ = \frac{B(\bar{\alpha}, \delta)}{B(\alpha, \delta)}, \quad (6.7.8) \]

where the integral in the middle line of (6.7.8) equals unity because the integrand is a beta p.d.f.
• From (6.7.1), (6.7.5) and (6.7.8) it follows that

\[
f(\theta|y) = \left[ \frac{[B(\alpha, \delta)]^{-1}}{B(\overline{\alpha}, \overline{\delta}) / B(\alpha, \delta)} \right] \theta^{\alpha-1} (1 - \theta)^{\delta-1} \theta^{m} (1 - \theta)^{T-m}
\]

\[= [B(\overline{\alpha}, \overline{\delta})]^{-1} \theta^{\overline{\alpha}-1} (1 - \theta)^{\overline{\delta}-1}, \quad 0 < \theta < 1. \tag{6.7.9}
\]

• Therefore, since posterior density (6.7.9) is itself a beta p.d.f. with parameters \( \overline{\alpha} \) and \( \overline{\delta} \) given by (6.7.6) and (6.7.7), it follows that the conjugate family of prior distributions for a Bernoulli likelihood is the beta family of p.d.f.s.
Example 6.7.2: Given $\theta = [\theta_1, \theta_2]' \in \mathbb{R} \times \mathbb{R}^+$, consider a random sample $y_t$ \((t = 1, 2, ..., T)\) from a $N(\theta_1, \theta_2^{-1})$ population. It is convenient to work in terms of $\theta_2$, the reciprocal of the variance (known as the precision). For now assume $\theta_2$ is known; this assumption is dropped in Example 6.7.4. Suppose prior beliefs concerning the unknown mean $\theta_1$ are represented by

$$\theta_1 | \theta_2 \sim N(\mu, h^{-1}), \quad (6.7.10)$$

where $\mu$ and $h > 0$ are given.
Let
\[ h = \left[ \theta_2^{-1}/T \right]^{-1} = T\theta_2, \]  
\[ \bar{h} = h + h, \]  
\[ \bar{\mu} = \bar{h}^{-1}(h\mu + h\bar{y}). \]
It is useful to employ two identities.

- The first identity is Exercise 5.2.2:

\[
\sum_{t=1}^{T} (y_t - \theta_1)^2 = \sum_{t=1}^{T} (y_t - \bar{y})^2 + T(\bar{y} - \theta_1)^2
\]

\[
= \nu s^2 + T(\bar{y} - \theta_1)^2
\]

for all \( \theta_1 \), where

\[
\nu \equiv T - 1.
\]

\[
s^2 = \nu^{-1} \sum_{t=1}^{T} (y_t - \bar{y})^2.
\]
The second identity is left as an exercise for the reader:

\[ h(\theta_1 - \mu)^2 + h(\bar{y} - \theta_1)^2 = \bar{h}(\theta_1 - \overline{\mu})^2 + (h^{-1} + h^{-1})^{-1}(\overline{y} - \mu)^2, \quad (6.7.17) \]

for all \( \theta_1, \ h, \) and \( h. \)
Using identity (6.7.14), the likelihood function is

\[ L(\theta_1; y | \theta_2) = \prod_{t=1}^{T} \varphi(y_t | \theta_1, \theta_2^{-1}) \]

\[ = (2 \pi \theta_2^{-1})^{-T/2} \exp \left[ -\frac{\theta_2}{2} \sum_{t=1}^{T} (y_t - \theta_1)^2 \right] \]

\[ = (2 \pi \theta_2^{-1})^{-T/2} \exp \left[ -\frac{h}{2T} \left[ \nu s^2 + T(\bar{y} - \theta_1)^2 \right] \right] \]

\[ = c_1(\theta_2) \varphi(\bar{y} | \theta_1, h^{-1}), \]

where

\[ c_1(\theta_2) = (2\pi)^{-\nu/2} T^{-\nu/2} \theta_2^{\nu/2} \exp(-\frac{1}{2} \theta_2 \nu s^2) \]

\[ (6.7.18) \]

\[ (6.7.19) \]

does not depend on \( \theta_1 \).
Note:

- The factorization in (6.7.18) demonstrates that $\bar{Y}$ is a sufficient statistic for $\theta_1$.

- Density $\varphi(\bar{y} | \theta_1, h^{-1})$ corresponds to the sampling distribution of the sample mean [see Theorem 5.3.5(a)], given $\theta$.

- Example 6.7.2 involves “completing the square.” See Jackman (2009, Propositions C.1 and C.2).
Using identity (6.7.17) and factorization (6.7.18), the numerator of (6.7.1) is

\[
f(\theta_1 | \theta_2) \equiv (\theta_1 ; y | \theta_2)
\]

\[
= \varphi(\theta_1 | \mu, h^{-1}) \; c_1(\theta_2) \; \varphi(y | \theta_1, h^{-1})
\]

\[
= c_1(\theta_2) \; (2\pi h^{-1})^{-\frac{1}{2}} \; (2\pi h^{-1})^{-\frac{1}{2}} \; \exp\left[-\frac{1}{2} \left\{ h(\theta_1 - \mu)^2 + h(\bar{y} - \theta_1)^2 \right\} \right]
\]

\[
= c_1(\theta_2) \; (2\pi h^{-1})^{-\frac{1}{2}} \; (2\pi h^{-1})^{-\frac{1}{2}} \exp\left[-\frac{1}{2} \left\{ \bar{h}(\theta_1 - \bar{\mu})^2 + (h^{-1} - h^{-1})^{-1}(\bar{y} - \mu)^2 \right\} \right]
\]

\[
= \left[ c_1(\theta_2) \; \varphi(\bar{y} | \mu, h^{-1} + h^{-1}) \right] \varphi(\theta_1 | \bar{\mu}, h^{-1}^{-1}). \tag{6.7.20}
\]
The denominator of (6.7.1), the density of the data given $\theta_2$, can then be easily obtained from (6.7.20):

$$
\int_{-\infty}^{\infty} f(y|\theta_2) = \int f(\theta_1|\theta_2) \mathcal{L}(\theta_1; y|\theta_2) \, d\theta_1
$$

$$
= c_1(\theta_2) \varphi(\bar{y}|\mu, h^{-1} + h^{-1}) \int_{-\infty}^{\infty} \varphi(\theta_1|\bar{\mu}, \bar{h}^{-1}) \, d\theta_1
$$

$$
= c_1(\theta_2) \varphi(\bar{y}|\mu, h^{-1} + h^{-1}). \tag{6.7.21a}
$$

The density $\varphi(\bar{y}|\mu, h^{-1} + h^{-1})$ in (6.7.21a) corresponds to the sampling distribution of the sample mean given $\theta_2$. 
Using Exercise 6.7.50, the marginal density of the data (remember $\theta_2$ is assumed known) in (6.7.21a) can alternatively be expressed as

$$f(y|\theta_2) = \varphi\left(y|\mu, \theta_2^{-1} \left[I_T + h^{-1} \imath_T \imath_T'\right]\right).$$  \hspace{1cm} (6.7.21b)
Dividing (6.7.20) by (6.7.21a) yields the posterior density of $\theta_1$, given $\theta_2$:

$$f(\theta_1 | y, \theta_2) = \varphi(\theta_1 | \bar{\mu}, \bar{h}^{-1}).$$  \hspace{1cm} (6.7.22)

where

$$\bar{h} = h + h,$$  \hspace{1cm} (6.7.12)

$$\bar{\mu} = \bar{h}^{-1}(h\mu + h\bar{y}).$$  \hspace{1cm} (6.7.13)

The interpretation of quantities (6.7.12) and (6.7.13) is now clear from (6.7.22): they are the posterior precision and posterior mean, respectively.
• Note that it is the additivity of precisions in (6.7.12) that motivates working with precisions rather than variances.

• Since posterior density (6.7.22) and prior density (6.7.10) are both members of the normal family, it follows that the natural conjugate prior for the case of random sampling from a normal population with known variance is itself a normal density.
Example 6.7.3 [Box and Tiao (1973, pp. 15-18)]: Consider two researchers, A and B, concerned with obtaining accurate estimates of a parameter $\theta$. Suppose that the prior beliefs of A are $\theta \sim N(900, 20^2)$ and the prior beliefs of B are $\theta \sim N(800, 80^2)$. Note that researcher A is far more certain, a priori, than researcher B. Figure 6.7.1(a) depicts the respective prior pdfs.

Suppose that given $\theta$ an unbiased method of experimental measurement is available and one observation is taken from a $N(\theta, 40^2)$ distribution and it turns out to be $y_1 = 850$. Then based on this observation both researchers will update their prior beliefs as in Example 6.7.2 to obtain the posterior distributions $\theta \mid y_1 \sim N(890, [17.9]^2)$ and $\theta \mid y_1 \sim N(840, [35.7]^2)$ for researchers A and B, respectively. These posterior p.d.f.s are depicted in Figure 6.7.1(c) which suggests that the beliefs of the two researchers have become more similar, but remain distinct.
If 99 additional observations are gathered so that the sample mean based on all $T = 100$ observations is $\bar{y} = 870$, then the posterior distributions for researchers A and B become $\theta \mid \bar{y} \sim N(871.2, [3.9]^2)$ and $\theta \mid \bar{y} \sim N(869.8, [3.995]^2)$, respectively. These posterior p.d.f.s are shown in Figure 6.7.1(e), and near total agreement is now evident. Moral of the story: a sufficiently large sample will “swamp” the non-dogmatic prior.
Example (Final, 2006): Consider Exercise 6.7.2. Let $a \in \mathbb{R}$ and $b \in \mathbb{R}^+$. Suppose the researcher reports the Bayesian posterior $N(a, b)$ with the same likelihood given in the exercise. Will there always exist a prior $N(\mu, h^{-1})$ to rationalize these posterior beliefs?
Solution: Proceeding as in Exercise 6.7.2 with $h = 4/100 = .25$ and $\bar{y} = 56$, we need to solve the equations

$$\bar{\mu} = \bar{h}^{-1}(h\mu + h\bar{y}),$$

$$\bar{h} = h + h,$$

for $\mu$ and $h^{-1}$. This yields

$$\mu = \frac{\bar{\mu}\bar{h} - h\bar{y}}{h} = \frac{b^{-1}a - .25\bar{y}}{b^{-1} - .25},$$

$$h = \bar{h} - h = b^{-1} - .25 > 0,$$

provided $b^{-1} > .25$. If this condition is not met, then there will not be a rationalizing prior.