

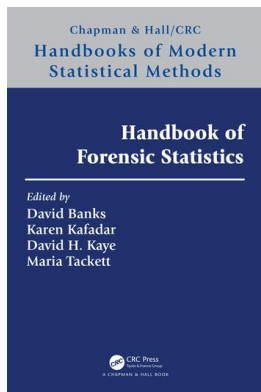
This article was downloaded by: 10.3.98.104

On: 28 Oct 2021

Access details: *subscription number*

Publisher: *CRC Press*

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: 5 Howick Place, London SW1P 1WG, UK



Handbook of Forensic Statistics

David Banks, Karen Kafadar, David H. Kaye, Maria Tackett

Bayesian Methods and Forensic Inference

Publication details

<https://www.routledgehandbooks.com/doi/10.1201/9780367527709-3>

David Banks, Maria Tackett

Published online on: 06 Nov 2020

How to cite :- David Banks, Maria Tackett. 06 Nov 2020, *Bayesian Methods and Forensic Inference* from: Handbook of Forensic Statistics CRC Press

Accessed on: 28 Oct 2021

<https://www.routledgehandbooks.com/doi/10.1201/9780367527709-3>

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: <https://www.routledgehandbooks.com/legal-notices/terms>

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

3

Bayesian Methods and Forensic Inference

David Banks and Maria Tackett

CONTENTS

3.1	Introduction	73
3.2	The Basics	74
3.2.1	A Beta-Binomial Mock Example	76
3.2.2	A Gamma-Poisson Mock Example	79
3.3	Markov Chain Monte Carlo	85
3.4	Broad Applications	86
3.5	Summary	89
	Acknowledgments	89
	References	89

3.1 Introduction

Bayesian methods are well-established analytical tools. Along with frequentist methods (see Chapter 2), these comprise the two dominant approaches to statistical inference. (A third group of methods, called fiducial or likelihood inference has found less favor.) Reviews of all three perspectives can be found in Young et al. (2005).

The Bayesian approach differs from the frequentist approach in several important ways. First, a Bayesian expresses uncertainty as a probability, which a frequentist is not allowed to do. For example, when a frequentist sets a confidence interval, he may not say something such as “the probability that the interval contains the true mean of the population is 0.95.” Instead, he must say that “95% of similarly constructed intervals will contain the true mean.” This is because the interval either contains the true mean or it does not. The frequentist does not know which of these is correct, but he may not express his uncertainty as a probability.

In contrast, a Bayesian seeing the same data will construct a *credible interval*, a region which, to a Bayesian, has some pre-specified probability of containing the unknown parameter value. In some cases a credible interval with probability 0.95 of containing the parameter value is in exact numerical agreement with the 95% confidence interval. But the philosophy and interpretation of a credible interval is very different from that of a confidence interval.

A second difference is that the Bayesian may integrate over the parameter space. This is important in several ways, but one application is in hypothesis testing. A frequentist will test the null hypothesis against the alternative, and he will either reject or fail to reject the null hypothesis at some pre-specified alpha level (often 0.05 or 0.01). In contrast, the Bayesian will calculate her *posterior distribution*, a probability distribution that describes

her belief about the value of the unknown parameter after observing the data. Then she can integrate her posterior distribution over the region corresponding to the null hypothesis, and thus calculate her probability that the null hypothesis is true.

The third difference is the most controversial. Bayesians often start with subjective beliefs about the probabilities of events. For example, based on life experience or just plain prejudice, a Bayesian juror might have an initial belief that the chance that the defendant is guilty is 0.9, which should then be modified through the use of a mathematical tool called "Bayes' Rule" as evidence is presented. This initial belief is difficult to square with the legal tenet that a defendant should be presumed innocent until guilt is proven, but it may be a more realistic representation of juror thinking than to assume each juror is a blank slate upon which the lawyers paint their narratives.

Savage (1954) proved that under a very general and reasonable definition of rationality, any rational person must act as a Bayesian. Specifically, if agents have beliefs that can be characterized through a probability function on some space, then rational agents should make decisions that maximize their expected utility function (i.e., decisions that, on average, provide the largest benefit or least loss where the benefit or loss need not be monetary). And this compels them to act as Bayesians. On the other hand, another theorem shows that a committee (such as a jury) cannot act collectively as a Bayesian unless they initially hold identical opinions (*prior beliefs* or simply *priors*) about the probability distributions for all relevant unknowns in the problem and also agree on the mechanism for generating the observations (cf. Kadane et al., 1993). The usual way of addressing this problem is to invoke another theorem, which states that under general conditions, two Bayesians who observe the same stream of data will converge in their beliefs. Thus, in the context of a jury trial, each will juror have separate prior beliefs, and the lawyers would attempt to present enough evidence that all members of the jury would converge towards unanimity (as required in many states), so that all would vote for acquittal or conviction. Some jurors might need little evidence, others a great deal, but all opinions should be changing in the same direction.

3.2 The Basics

Bayes' Theorem (or Bayes' Rule) is a completely standard and universally accepted consequence of elementary probability. Recall that the conditional probability of an event A given that event B is observed is defined as

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \text{ and } B]}{\mathbb{P}[B]}$$

where $\mathbb{P}[A \text{ and } B]$ is the probability that both events A and B occur, and $\mathbb{P}[B]$ is the probability that event B occurs.

Let A_1, \dots, A_k be a finite partition of the set of possible outcomes, where a finite partition means that one of A_1, \dots, A_k must happen, but the events are mutually exclusive (disjoint), meaning that it is impossible for two or more of them to happen at the same time. In drawing a card from a standard deck, one possible finite partition is Clubs, Diamonds, Hearts and Spades; another finite partition is Red and Black; a third is aces, twos, ..., kings.

Bayes' Rule says that if A_1, \dots, A_k is a finite partition and one observes the event B , then the probability of A_i given that B is

$$\mathbb{P}[A_i|B] = \frac{P[B|A_i] \times \mathbb{P}[A_i]}{\sum_{j=1}^k P[B|A_j] \times \mathbb{P}[A_j]} \tag{3.1}$$

One can show that the numerator is just $\mathbb{P}[A_i \text{ and } B]$ and, for a finite partition, the denominator is just $\mathbb{P}[B]$, so this is just a rewriting of the definition of conditional probability. Richard Price published this theorem in 1763, on behalf of the Reverend Thomas Bayes, who had discovered it, in the *Philosophical Transactions of the Royal Society* (Bayes and Price, 1763). To illustrate an application of Bayes's Theorem, consider the enzyme-linked immunosorbent assay (ELISA) test for HIV. Chou et al. (2005) of the U.S. Preventive Services Task Force studied the ELISA test carefully, and found that if one has HIV, ELISA signals with probability 0.997 (the sensitivity), and if one does not have HIV, then it does not signal with probability 0.985 (the specificity). And it is known that about 0.34% of the U.S. population is HIV positive (CDC, 2017).

Suppose a random person is given a blood test (e.g., the person wants to join the army, and the blood test is a routine part of the physical examination). And suppose the ELISA test signals that the person has HIV. What is the chance that the ELISA test is correct and the subject is truly HIV positive? Note that having HIV or not having HIV constitutes a finite partition with just two events: one of the two must be true, but it is not possible for both to be true. Using Bayes's Rule, one finds that the probability of being HIV positive given a positive ELISA test is

$$\begin{aligned} &\mathbb{P}[\text{HIV} \mid \text{pos test}] \\ &= \frac{\mathbb{P}[\text{pos test} \mid \text{HIV}] * \mathbb{P}[\text{HIV}]}{\mathbb{P}[\text{pos test} \mid \text{HIV}] * \mathbb{P}[\text{HIV}] + \mathbb{P}[\text{pos test} \mid \text{no HIV}] * \mathbb{P}[\text{no HIV}]} \\ &= (0.997) * (0.0034) / [(0.997) * (0.0034) + (1 - 0.985) * (1 - 0.0034)] \\ &= 0.1848. \end{aligned}$$

So even though the test signals HIV, because the prevalence of HIV in the population is so low, the actual chance of being HIV positive is fairly small.

This calculation requires that the person being tested is chosen at random (i.e., in a manner that is unrelated to HIV status). It would not apply to someone who came to the physician because he felt ill. Also note that this example can be generalized to breathalyzer tests, fingerprint matches, and other forensic-science tests.

In this example, the Bayesian analyst did not need to use subjective probability, because she knew the baseline prevalence of HIV in the U.S. population and the problem assumed that person being tested was a random draw from that population, allowing her to use an objective probability. However, if the person were being tested because he felt ill or had engaged in risky behavior, then a Bayesian physician would have a different, probably subjective, probability of HIV infection, and that probability would be larger than 0.0034.

Although Bayesian inference has compelling intellectual properties, in its mathematical form, it generally requires the evaluation of complex integrals. Specifically, the integral form of the Bayes' Rule, which is the analogue of the discrete form given in (3.1), is

$$\pi^*(\theta|x_1, \dots, x_n) = \frac{f(x_1, \dots, x_n|\theta)\pi(\theta)}{\int_{-\infty}^{\infty} f(x_1, \dots, x_n|\theta)\pi(\theta) d\theta} \tag{3.2}$$

where $\pi(\theta)$ is the prior density for the unknown parameter θ , $\pi^*(\theta|x_1, \dots, x_n)$ is the posterior density for θ after observing the data x_1, \dots, x_n , and $f(x_1, \dots, x_n|\theta)$ is the density for the distribution of the data given the value of θ . A few examples will clarify all this, but the main point is that the analyst had prior beliefs that were updated after seeing data to form posterior beliefs, and that this updating was done according to Bayes' Rule.

The integral in the denominator of (3.2) was a stumbling block to the use of Bayesian inference for several centuries. In most cases, the integral could not be evaluated, and thus the updating calculation could not be completed. Essentially, the only circumstances in which tractable calculation was possible were with conjugate families. In conjugate families, the prior belief on the value of the unknown parameter is assumed to be a member of a convenient family, and the distribution that generates the data, conditional on the value of the parameter, is also assumed to be a member of a convenient family. The convenience of these two families derives from the fact that when the prior belief about the parameter is updated by the data to obtain the new (posterior) belief about the parameter, it turns out that the new belief is in the same distributional family as the prior. The three most prominent of these conjugate families are the beta-binomial, the normal-normal, and the gamma-Poisson families.

Two hypothetical applications illustrate different facets of Bayesian reasoning. These examples will use the beta-binomial and gamma-Poisson families.

3.2.1 A Beta-Binomial Mock Example

The beta-binomial is used when making inferences about a proportion or a probability. For example, suppose a forensic accountant wants to estimate the proportion of fraudulent claims filed with an insurance company by an automobile repair chain. Formally, one would have to assume that an infinite number of claims have been filed, and that the forensic accountant is sampling this population of claims at random. In practice, it is sufficient to simply assume that a very large number of claims have been filed, but it is still essential that the sampling be random.

The binomial distribution describes the number of "successes" in a fixed number of trials, where the chance of success on each trial is independent with constant probability θ . For n trials, the probability of exactly k successes is

$$\mathbb{P}[k \text{ successes}] = \binom{n}{k} \theta^k (1 - \theta)^{n-k}$$

where

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1) \cdots 1}{[k(k-1) \cdots 1][(n-k)(n-k-1) \cdots 1]}$$

(Deeper mathematics shows that $0! = 1$, which is a case that can arise in practice; for example, if none of the trials is a success, $k = 0$.) The binomial distribution is iconically used to describe the number of Heads obtained in n tosses of a coin that has probability θ of coming up Heads.

In our example, the forensic accountant decided how many claims to sample; this is her n . She then draws a random sample of n claims and audits each. A success is finding a fraudulent claim. If she records a 0 for a failure and a 1 for a success, then, in the notation of (3.2), she observes x_1, \dots, x_n , where each observation x_i is either a one or a zero. Based

on this sample, she now wants to make an inference about θ , the proportion of fraudulent claims.

As a Bayesian, she has a prior opinion about the value of θ . If that opinion happens to be expressible as a beta distribution, then she is in a conjugate family and can calculate the awkward integral in the denominator of (3.2).

A beta random variable may take any value between 0 and 1, inclusive. Since the beta distribution describes a continuous random variable, the probability that it takes any specific value is zero, but if one integrates the density function of a beta distribution between, say, 0 and 0.5, one obtains the probability the value p is less than or equal to 0.5.

The beta distribution is indexed by two parameters, traditionally represented as α and β . Both α and β must be greater than zero. The density function of a random variable θ with Beta(α, β) distribution is

$$\pi(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1} \tag{3.3}$$

where $\Gamma(z) = (z - 1)!$ for all positive z (the definition in the case of non-integer z values derives from an integral representation of this function).

The beta family is quite flexible, and can represent many possible beliefs about θ . Figure 3.1 shows the shape of the density function for three possible pairs of indices (α, β).

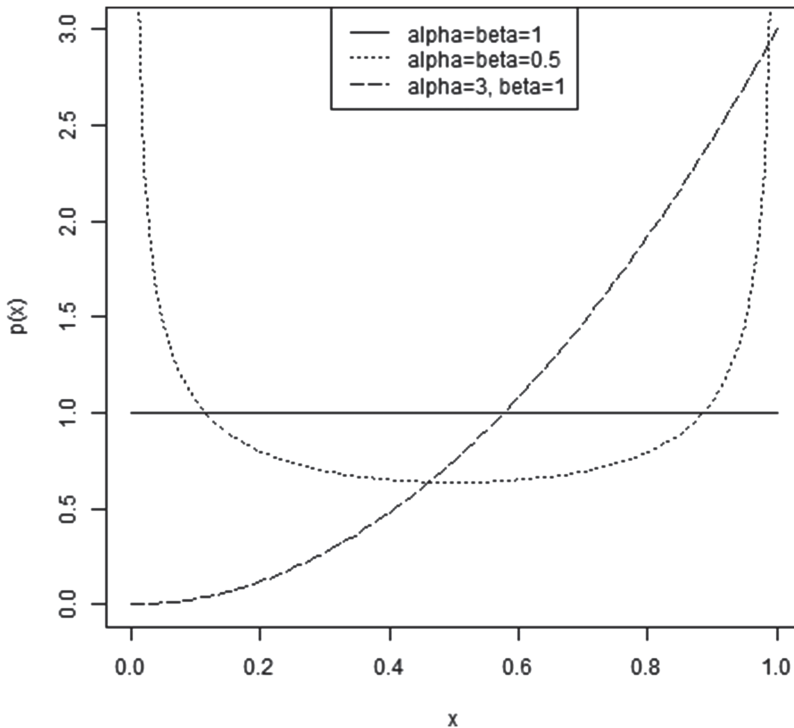


FIGURE 3.1

This figure shows three different densities for the beta distribution, which illustrate the wide range of prior beliefs that can be incorporated. The case $\alpha = \beta = 1$ is uninformative about the probability of success θ ; the case $\alpha = \beta = 0.5$ represents the belief that θ is probably close to zero or one; and the case $\alpha = 3, \beta = 1$ puts more weight on larger values of θ .

Note that when $\alpha = \beta = 1$, the density is flat. The forensic accountant might use this density as her prior if she were completely agnostic about θ and thought that all values between 0 and 1 were equally likely. This would be a noninformative prior, since no value of θ is favored. (For a discussion of using a flat prior with data from a Supreme Court case, see Kaye, 1982, p. 779.) Alternatively, it might be reasonable to think that the automobile repair chain is either mostly honest or mostly dishonest, in which case she could use the informative beta prior whose density has $\alpha = \beta = 0.5$. Or, based on other evidence, she might believe that the chain is a criminal enterprise, and then she might pick the density corresponding to $\alpha = 3$ and $\beta = 1$.

At this point the Bayesian forensic accountant has collected the sample x_1, \dots, x_n and selected the α and β that specify her beta prior. She now wants to solve (3.2) in order to determine her new belief about the density of θ , the probability of a fraudulent claim given what she has learned from her sample. She can replace x_1, \dots, x_n by the number of successes, i.e., the number of fraudulent claims. (In statistics, there is often a *sufficient statistic*, one that summarizes all the relevant information in the data; in this case one can show that $k = \sum x_i$ is sufficient for inference on θ .) Then the posterior density is

$$\begin{aligned} \pi^*(\theta | k) &= \frac{f(k | \theta)\pi(\theta)}{\int_{-\infty}^{\infty} f(k | \theta)\pi(\theta) d\theta} \\ &= \frac{\left[\binom{n}{k} \theta^k (1 - \theta)^{n-k} \right] * \left[\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1} \right]}{\int_0^1 \left[\binom{n}{k} \theta^k (1 - \theta)^{n-k} \right] * \left[\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1} \right] d\theta} \\ &= \frac{\Gamma(n + \alpha + \beta)}{\Gamma(k + \alpha)\Gamma(n - k + \beta)} \theta^{k + \alpha - 1} (1 - \theta)^{n - k + \beta - 1} \end{aligned}$$

which is the Beta($\alpha + k, \beta + n - k$) density.

The trick to making this conjugate family work is to recognize that the denominator has integrated out θ , so it is some constant. And the numerator is a member of the beta family, since its kernel (the terms in the integrand that depend upon θ) is the product of θ to a power times $1 - \theta$ to a power. So the constant in the denominator combined with the constants in the numerator have to be the value that forces the integral of the density to equal 1. From (3.3), it is clear that the constant must be $\Gamma(n + \alpha + \beta) / \Gamma(k + \alpha)\Gamma(n - k + \beta)$.

The posterior density is the complete expression of the forensic accountant's new belief. She can use it in several ways. For example, the district attorney may have an informal rule of thumb that if the proportion of invalid claims is less than 3%, then it is probably error rather than fraud, and not worth prosecutorial effort. In that case, the forensic accountant would integrate her posterior for θ over the region from 0 to 0.03, to calculate her probability that the district attorney should decline to prosecute.

But it is instructive to look at the mean and variance of her posterior distribution, to see how the new data have affected her previous opinion. One can show that the mean μ and variance σ^2 of a beta distribution with parameters α and β are

$$\mu = \frac{\alpha}{\alpha + \beta} \quad \sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

Looking first at the posterior mean μ^* , which is a beta distribution with parameters $\alpha^* = \alpha + k$ and $\beta^* = \beta + n - k$, we can write

$$\mu^* = \frac{\alpha + k}{\alpha + \beta + n} = c \frac{\alpha}{\alpha + \beta} + (1 - c) \frac{k}{n}$$

where $c = (\alpha + \beta)/(n + \alpha + \beta)$. This representation of μ^* is a weighted average of the mean of the prior and the mean of the data (k out of n successes implies the average of the data is k/n). As the sample size increases, more weight is placed upon upon the empirical average, and less upon the prior mean, which makes sense. The forensic accountant might have silly prior beliefs, but if she observes lots of data and follows Bayes' Rule, her opinions will shift toward what the data are revealing. This type of shift, to take more account of the data as the sample size grows, is standard in Bayesian analysis.

Looking at the posterior variance, one sees that

$$\sigma^{*2} = \frac{(\alpha + k)(\beta + n - k)}{(\alpha + \beta + n)^2(\alpha + \beta + n + 1)}.$$

This formula shows that as n increases, the posterior variance shrinks, so the forensic accountant's uncertainty becomes smaller and her probability concentrates upon a shrinking region within the interval $[0, 1]$ in which θ , the probability of success, must lie.

Before leaving this mock example, there are a few points to underscore regarding the use of subjective priors. First, Savage's results on the rationality of Bayesian decision makers requires the analyst to use her true subjective prior, rather than a noninformative (or objective) prior. Second, noninformative priors are not simple; they may not exist, they may not be unique, and they may be improper (meaning that their integral is infinite, rather than equal to 1). For the beta-binomial example, the Haldane prior, the Jeffreys prior, and the maximum entropy prior are all, in different technical senses, objective priors (cf. Kass and Wasserman, 1996). Third, as the sample size grows, all Bayesians who put non-zero probability on the entire domain of θ will converge to the same opinion, no matter which starting point prior each chose.

3.2.2 A Gamma-Poisson Mock Example

For another mock case study that illustrates several facets of Bayesian reasoning, suppose that a hospital review system flags a nurse who has had an unusually large number of his patients die during his shift. The police are consulted, and the district attorney wants to determine how improbable that observed number of deaths might be. The matter is complicated since well-known cases, such as that of Lucia de Berk (cf. Derksen and Meising, 2009), have highlighted the fact that, just by chance, an innocent nurse may have an unusual number of fatalities on his shifts. And if hospitals monitor many, many nurses, then it is certain that some innocent nurses will seem suspicious.

The analysis in this situation could entail the use of the gamma-Poisson conjugate family. The Poisson distribution describes the number of independent events that occur in a fixed amount of time, a fixed area, or a fixed volume. Assuming that an innocent nurse has independent deaths on his watch, is reasonable to model the number of deaths as a Poisson random variable with unknown mean θ . For expository simplicity, it is helpful to assume that the period of time in this application is one year.

A Poisson random variable can only take the values $0, 1, \dots$, without any upper limit. (Obviously, in applications, there is generally some upper bound, so this is technically only

an approximation, but it is widely used and broadly accurate.) The probability of observing exactly k deaths is

$$\mathbb{P}[X = k] = \frac{\theta^k}{k!} \exp(-\theta).$$

Here $\exp(-\theta)$ is Euler's constant, approximately 2.718, raised to the power $-\theta$. Using a Taylor series, one can show that the average number of events is equal to θ .

To form a conjugate family, the prior belief about θ must be represented by a gamma distribution. The investigator would have to form that prior based on personal experience, other data, or input from experts. For example, if the nurse in question worked with geriatric patients, then one would expect more natural deaths than if he worked in the pediatric unit. And hospital administrators could provide data on death rates in similar units at similar hospitals. The investigator would combine this information, with her own subjective beliefs, to specify her prior.

The gamma family is a group of distributions indexed by two parameters, α and β , both of which must be greater than zero. The gamma density function for a random variable θ is written as

$$f(\theta) = \begin{cases} \beta^\alpha \theta^{\alpha-1} \exp(-\beta\theta) / \Gamma(\alpha) & \text{for } \theta > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Be warned that some textbooks write this density function with an alternate parameterization, so the mathematical expression, although equivalent, looks a little different. If one is working with a gamma density, be sure to check which parameterization is being used.

The gamma family includes the exponential distributions and the chi-squared distributions as special cases. Figure 3.2 illustrates some of the beliefs about θ that can be described through a gamma distribution.

The mean of a gamma distribution is α/β , and the variance is α/β^2 . When specifying a subjective prior, it is often helpful to try to quantify one's uncertainty by thinking about the mean ν of the random variable θ and one's uncertainty about its value (perhaps expressed as the range $[L, U]$ that one believes has probability 0.95 of including the true mean of the distribution). Then one can solve the system of equations

$$\nu = \alpha/\beta \text{ and } [(U - L)/4]^2 = \alpha/\beta^2 \quad (3.4)$$

to find values for α and β that reflect one's personal beliefs. This strategy for eliciting one's belief is simple and based upon confidence intervals for normal distributions. But there are alternatives—the literature upon elicitation of Bayesian subjective beliefs is large. A standard reference is O'Hagan et al. (2006).

If there are data on the number of deaths for nurses who are not under suspicion, say x_1, \dots, x_n , then the analyst can update her subjective prior through Bayes' Rule to find her posterior belief about the distribution of θ for innocent nurses. Suppose the initial subjective prior has parameters α and β . Then it turns out that the posterior will be a gamma density with parameters $\alpha^* = \alpha + \sum x_i$ and $\beta^* = \beta + n$. This calculation can be a little tricky, since one must use data from other nurses who work with similar patients (e.g., geriatric or pediatric), and one must take care to adjust for the differing number of hours each nurse worked during the year.

Returning to the hypothetical case, suppose the investigator wanted to find the posterior distribution for the annual death rate θ of the suspect nurse. As will be seen, this is not the best thing to do, but it is a starting point for a naive Bayesian analysis.

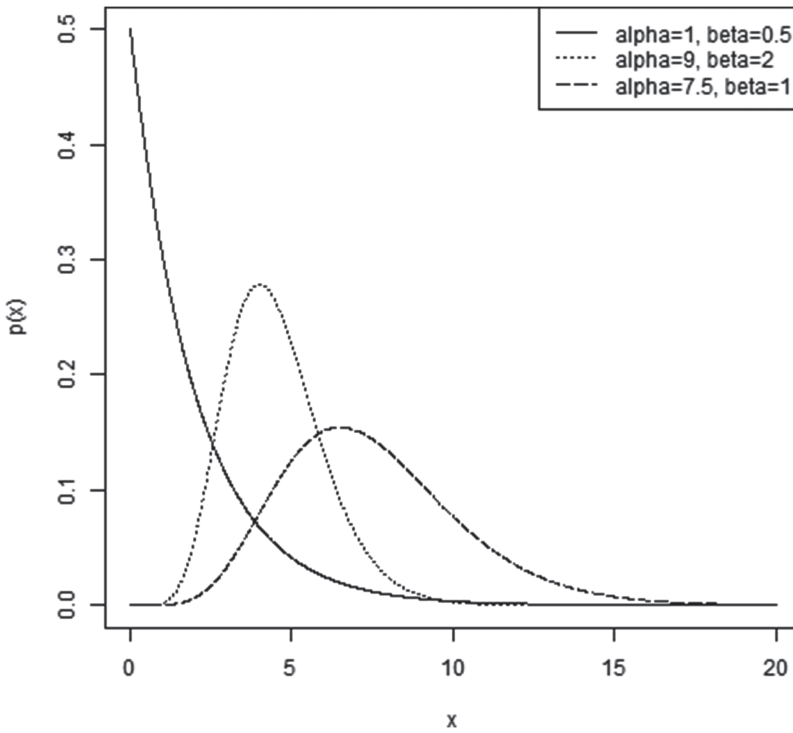


FIGURE 3.2

This figure shows three different densities for the gamma distribution, which illustrate the kinds of prior belief that may be represented. The case $\alpha = 1, \beta = 0.5$ corresponds to an exponential distribution; the other two cases show different means, variances, and skewness.

Her use of a prior implicitly accords with the fact that not all nurses have the same value of θ . A well-trained and experienced nurse might have a θ value slightly smaller than a less experienced nurse. This is related to random effects models that are discussed in the next section.

If x people have died on the suspect nurse’s watch, then the conjugacy property shows that the posterior distribution for the suspect nurse’s annual death rate θ has gamma distribution with parameters $\alpha^* = \alpha + x$ and $\beta^* = \beta + 1$. If this density has a mean that is very much larger than the prior mean α/β , then that could be interpreted as evidence against the nurse.

A frequentist might want to decide whether the nurse’s annual death rate θ is less than or equal to θ_0 , the average value for all similar nurses, or perhaps a value that is set by hospital policy as a proficiency standard. The frequentist would then test the null hypothesis that the suspect nurse’s θ is less than or equal to θ_0 against the alternative hypothesis that the nurse’s θ is greater than θ_0 . This entails calculation of a test statistic and, from that, a significance probability. If the significance probability were very low, say less than 0.01, then the frequentist would reject the null hypothesis and conclude that the nurse’s mean annual death rate was greater than θ_0 (see Chapter 2). Technically, the significance probability is the chance of observing a death count as high or higher than the suspect nurse’s count when the null hypothesis is true.

From a legal standpoint, one must be careful about the conclusion drawn from rejection of the null hypothesis. The investigator decides that the suspect nurse’s fatality rate is

larger than the specified null value, but this could be due to incompetence rather than criminality, or a management policy that assigns the most senior nurse to the sickest patients. Nonetheless, such analyses are sometimes undertaken as part of an investigation (for cases and discussions, see Fienberg and Kaye, 1991; Loue, 2010).

Although it may not be the best approach, the Bayesian analogue to this hypothesis test would calculate the posterior probability that the nurse's θ was greater than θ_0 . The Bayesian calculates the posterior distribution for the nurse's θ , and then integrates over the parameter space from θ_0 to infinity, to find

$$\mathbb{P}[\theta > \theta_0|x] = \int_{\theta_0}^{\infty} \pi^*(\theta|x) d\theta,$$

which is the posterior probability that the nurse's θ exceeds the hospital's threshold of θ_0 . Note that (1) the Bayesian analysis sidesteps the awkwardly expressed significance probability, (2) the solution was found by integrating over the parameter space, and (3) the uncertainty about θ is directly expressed as a probability, which frequentists may not do.

A different analysis might consider the probability of observing x or more deaths if the distribution for the nurse's death rate θ were a draw from the prior distribution. This is not actually a Bayesian analysis, since, among other things, Bayesians do not condition on unobserved events (i.e., Bayesians do not find the probability seeing more deaths than were counted). Instead, it is an attempt to harness a Bayesian tool to a frequentist purpose. However, note that in this situation, the frequentist logic forces the analyst to calculate the probability of x or more deaths, since the probability of seeing exactly x deaths can be misleadingly small; the probability of seeing exactly ten ants in a garden is low, but the chance of seeing ten or more could be large.

As before, this calculation requires integration over the parameter space. This is not permitted in frequentist analyses, but it allows the Bayesian investigator to calculate the probability of observing exactly x deaths when the uncertainty about the death rate is expressed as a distribution, which can then be repeated for $x + 1$, $x + 2$, and so forth. This hybrid Bayes-frequentist approach incorporates the fact that not all nurses have the same value θ —innocent nurses presumably have θ values that are close to some lower value than a guilty nurse.

Specifically, the Bayesian investigator solves

$$\begin{aligned} \mathbb{P}[x \text{ deaths}] &= \int_0^{\infty} \mathbb{P}[x \text{ deaths} | \theta] \pi(\theta) d\theta \\ &= \int_0^{\infty} \frac{\theta^x}{x!} \exp(-\theta) \beta^\alpha \theta^{\alpha-1} \exp(-\beta\theta) / \Gamma(\alpha) d\theta \\ &= \frac{\beta^\alpha}{x! \Gamma(\alpha)} \int_0^{\infty} \theta^{\alpha+x-1} \exp[-(\beta+1)\theta] d\theta \\ &= \frac{\beta^\alpha}{x! \Gamma(\alpha)} \frac{\Gamma(\alpha+x)}{(\beta+1)^{\alpha+x}}. \end{aligned}$$

The last step follows by recognizing the kernel of the gamma density in the integral, and recalling that the gamma density, integrated from zero to infinity, equals 1. This calculation can be repeated for the values $x + 1, x + 2, \dots$, and the sum of a small number of these probabilities quickly approximates $\mathbb{P}[x \text{ or more deaths}]$.

If the probability of observing a nurse with x or more deaths in, say, a year is very small, then this could be considered evidence of homicide. A numerical example is helpful. Suppose the investigator assumes that nurses in this kind of ward will have an average of 4 patient deaths per year, with a standard deviation of 2. Then, from (3.4), the gamma prior has $\alpha = 1$ and $\beta = 0.5$. Suppose that the suspect nurse has had 12 patients die on during his shift this year. Then, for this prior, the probability of exactly that many deaths is

$$\frac{\beta^\alpha}{x! \Gamma(\alpha)} \frac{\Gamma(\alpha + x)}{(\beta + 1)^{\alpha+x}} = \frac{0.5^1}{12! \Gamma(1)} \frac{\Gamma(13)}{(1.5)^{13}} = 0.00257. \tag{3.5}$$

So the chance of a good nurse having exactly 12 deaths in one year is very small.

To find the probability of 12 or more deaths in a year, the investigator calculates

$$\mathbb{P}[x \text{ or more deaths}] = \sum_x \frac{\beta^\alpha}{x! \Gamma(\alpha)} \frac{\Gamma(\alpha + x)}{(\beta + 1)^{\alpha+x}}. \tag{3.6}$$

For our conveniently chosen values of α and β , (3.6) is a geometric series and can be summed exactly; it equals 0.00771. For less convenient values, the summation can be easily approximated, but that detail lies outside the ambit of this discussion.

However, recall that there are probably many, many nurses who work in hospitals that monitor staff for unusual fatality patterns. Just by chance, one unlucky nurse will surely accumulate a suspicious number of deaths. To correct for this, one must calculate the probability that among m monitored nurses who are innocent, at least one of them will have 12 or more fatalities during the year.

There are different strategies for making this correction. In principle, a well-calibrated Bayesian need not make any adjustment, because her prior for the guilt of the nurse takes account of the fact that there are many nurses being monitored and thus the chance that this specific nurse is criminal is automatically low. Alternatively, the *False Discovery Rate* of Benjamini and Hochberg (1995) admits a Bayesian interpretation (cf. Muller et al., 2006). Scott and Berger (2006) develop a Bayesian solution for multiple testing when each of the hypotheses is independent, in the sense that the guilt or innocence of one nurse is unaffected by the guilt or innocence of another, and Gelman et al. (2012) tackles the more complicated case when dependence is present. These approaches are generally somewhat technical.

For an easier path, consider a Bayesian perspective on adjustment for multiple testing (Westfall et al., 1997). Let p be the probability that an innocent nurse has k or more fatalities in a year. Then among m monitored nurses who are innocent of murder, the probability that all of them will have fewer than k fatalities is $(1 - p)^m$. This is simple probability, and only assumes that the nurses are independent. From this, it follows that the probability that one or more nurses has k or more fatalities is $1 - (1 - p)^m$.

Recall the mock example, in which the nurse's probability of having 12 or more deaths in one year is 0.00771. If 10 comparable and innocent nurses are being monitored, then the probability that at least one of them reaches this level of suspicion is 0.0745. If there are 20 comparable nurses, then the probability that one or more seems suspicious is 0.1434, and if there are 100, then the probability is 0.5388. In a large hospital system, it seems quite likely that many nurses are monitored, and so the evidence in this example is probably not determinative on its own.

Finally, before leaving this example, consider the classical Bayesian approach, which avoids the awkward hybrids and compromises discussed previously. It is directly comparable to the ELISA example. Either the nurse is innocent or guilty—that provides a finite partition. The investigator (or a juror) has a personal subjective prior probability for innocence or guilt, which is based upon all that they know about the circumstances before seeing the data (i.e., the number of fatalities during the year on that nurse’s shift).

To find the posterior probability of guilt, a Bayesian must build a probability model for the number of deaths for an innocent nurse and the number of deaths for a guilty nurse. Different people would likely build that model in different ways. Earlier in this section, there was a model that found the probability of exactly 12 deaths for an innocent nurse as 0.00257. Similar reasoning could build a model for the probability of a guilty nurse having 12 fatalities in one year. It is likely that conjugate families would not seem reasonable to some of these Bayesians, in which case they would use Markov chain Monte Carlo methods, as discussed in the next section.

In any case, after eliciting their personal priors for innocence and guilt, and calculating the probability of exactly 12 deaths under the model for innocent and guilty nurses, all of the Bayesians would then find the posterior probability as

$$\begin{aligned} & \mathbb{P}[\text{guilt} | 12 \text{ deaths}] \\ &= \frac{\mathbb{P}[\text{guilt}]\mathbb{P}[12 \text{ deaths} | \text{guilt}]}{\mathbb{P}[\text{guilt}]\mathbb{P}[12 \text{ deaths} | \text{guilt}] + \mathbb{P}[\text{innocence}]\mathbb{P}[12 \text{ deaths} | \text{innocence}]} \end{aligned}$$

Here $\mathbb{P}[\text{guilt}]$ is the prior probability of guilt and $\mathbb{P}[\text{innocence}] = 1 - \mathbb{P}[\text{guilt}]$. Note that this traditional Bayesian analysis did not attempt to find the probability of observing 12 or more deaths. Instead, it simply used the observed count.

In Bayesian hypothesis testing, one decides which of two models (e.g., those for guilt and innocence) is more likely. The *Bayes factor* measures the degree to which the data support one or the other model, and it corresponds to a likelihood ratio (see Chapter 2). Formally, the Bayes factor $BF(H_0, H_A)$ is the ratio of the posterior odds of the null hypothesis H_0 and alternative hypothesis H_A to the prior odds of the null and alternative. Specifically,

$$BF(H_0, H_A) = \frac{\mathbb{P}[H_A|D]/\mathbb{P}[H_0|D]}{\mathbb{P}[H_A]/\mathbb{P}[H_0]}$$

where D indicates the data. It is not obvious from this expression, but it is nonetheless true, that when two models or two hypotheses explain the data equally well, the Bayes factor automatically penalizes the more complex model, and thus applies Occam’s Razor (Kass and Raftery, 1995).

A Bayes factor of, say, 5 means that after seeing the data, the alternative hypothesis is five times more likely relative to the null hypothesis than it was *a priori*, before seeing the data. In scientific settings, a Bayes factor of 10 or more is generally considered to be strong evidence for the alternative hypothesis (Jeffreys, 1998 [1961], p. 432). But in forensic settings, the standard is generally much more stringent. The Department of Justice (2018, p. 4) proposed uniform language for legal testimony in which a Bayes factor between 2 and 100 is described as “limited support,” a factor between 100 and 10,000 corresponds to moderate support, a factor between 10,000 and 1,000,000 is strong support, and a value greater than 1,000,000 is very strong support.

The logarithm of the Bayes factor is called the *weight of evidence*. Good (1985) has a slightly theoretical discussion of this concept, partly in the context of the legal system. Also, for more extensive discussions of the use of Bayesian reasoning in court cases, see Fienberg and Kadane (1983); Kaye et al. (2011).

3.3 Markov Chain Monte Carlo

Conjugacy is sometimes helpful, but most realistic applications are more complex. It is likely that distribution of the data and distribution of the honest Bayesian belief do not correspond to a conjugate family, and thus, in general, there is no solution that can be expressed as a finite number of standard mathematical terms—some numerical approximation is needed, and these can be computationally intractable or difficult to interpret. This limited the use of Bayesian methods for almost a century.

Gelfand and Smith (1990) found a way forward. Instead of trying to approximate the posterior density analytically, they developed a computer-intensive procedure that could sample repeatedly from the posterior distribution. Variants of this procedure are called Gibbs Sampling and Markov chain Monte Carlo (MCMC). The former is a special case of the latter, and these procedures have been further developed and improved by many researchers.

When one can draw large samples from the posterior density, it is possible to estimate the density with as much accuracy as the application requires. With a sample of size n , one can put weight $1/n$ on each observation to obtain an estimator that is a (possibly multivariate) histogram that converges to the true density function. If more accuracy is needed, one can simply increase n .

These procedures are not foolproof, and a number of technical problems that can arise (although these pathologies generally do not arise in reasonably posed problems—MCMC is robust).

One potential problem is that the Markov chain can take a long time to converge to its stationary distribution, the point at which the samples that are drawn are taken from the true posterior distribution. The usual solution is to let the chain run for a several thousand iterations, discard those draws, and then begin recording the sample values (this initial period is called the *burn-in*). To confirm that the chain has reached its stationary distribution, the statistician will typically monitor *trace plots*. These plots help diagnose when convergence has been reached, and other technical problems that can occur.

A second issue is that, at convergence, the sampled observations are draws from the posterior, but they are not an independent (random) sample. Consecutive draws are correlated, and so one needs to draw very large samples to ensure that the empirical distribution closely approximates the true posterior distribution.

A third potential issue arises when the support of the posterior density (i.e., the region where the density is greater than zero or, equivalently, the region where there it is possible to observe a data point) has a complex shape. Recall that the support of the beta density was the interval from 0 to 1, and the support of the gamma density was the positive numbers. These are simple supports. But in multivariate settings, certain geometries of the support lead to slow mixing (i.e., slow exploration of the support), so the analyst could be deceived

into thinking that the histogram of the data accurately reflects the true distribution when it does not.

Multivariate parameter spaces arise when measuring the amounts of several trace elements in a glass fragment or bullet lead. For example, the FBI used to perform a test called Compositional Analysis of Bullet Lead (CABL). The test estimated the proportion of seven trace elements in a bullet: antimony, copper, arsenic, silver, tin, bismuth, and cadmium. This implies a seven-dimensional parameter space, and the goal is to estimate $(\theta_1, \dots, \theta_7)$, the proportions of the trace elements in bullet. When these means were very similar to the estimates from another bullet, the FBI experts used to testify that the bullets were manufactured on the same day, from the same “melt” of lead.

CABL testimony is no longer given, following scientific criticism (National Research Council, 2004), but it is illustrative to note that in this example the support of the posterior density will be a seven-dimensional simplex, since $\theta_i \geq 0$ for all $i = 1, \dots, 7$ and $\sum \theta_i \leq 1$. This simplicial region is not problematic for MCMC, but more exotic regions are.

Suppose, for example, that the support of the multivariate density looked like a dumbbell. Then the MCMC algorithm might easily wander around in one of the end weights for a very long time before finding the bar that connects to the other end weight. In that case the traceplot would look fine for a long time, and the analyst might wrongly conclude that convergence had been achieved, even though less than half of the support had been explored. (But, as is probably evident, this kind of pathology arises very rarely, and should not be problematic for routine Bayesian applications.)

The main message is that Bayesian statisticians are now able to analyze problems that are much more complex and realistic than used to be possible. It has been used in many cases to interpret complex DNA mixtures (cf. Chapter 11). MCMC, the computational algorithm for doing such analyses, is a mature and reliable technique, but it should be implemented by a trained statistician since there are certain kinds of pitfalls that require special expertise to avoid.

For additional discussion of MCMC, there are a number of possible books, at a range of technical detail. One that is both popular and fairly accessible is Gelman et al. (2013). Gamerman and Lopes (2006) is a bit older and more theoretical. McElreath (2018) is recent, fairly introductory, and has many examples.

3.4 Broad Applications

Modern Bayesian analysis can address essentially any question that frequentist methods do, but there are some areas where the Bayesian perspective is especially elegant. These situations are probably rare in legal contexts, but it may be useful to sketch a few examples.

First, consider *random effects* models. To return to the automobile repair example, suppose one suspects that a particular shop routinely overcharges for, say, an engine tune-up. A sample of data from different repair shops reveals what each has charged for tune-ups performed during the previous year. Some shops have done many repairs, but for others, there are only a few observations. The goal is to calculate each shop’s average cost, borrowing strength from those with many observations to improve estimates for those with few.

Let Y_{ij} be the charge for the i th tune-up at the j th shop. Then the *fixed effects* model is

$$Y_{ij} = \mu + b_j + \epsilon_{ij}$$

where μ is the average of all shops, b_j is the effect of the j th shop, and $\epsilon_{ij} \sim N(0, \sigma^2)$. The fixed effects estimate of b_j is $\bar{Y}_j - \bar{Y}$.

In a fixed effects model, it is very difficult to borrow strength. But in a random effects model, it is easy. Fixed effects models are not Bayesian hierarchical models, but random effects models are. One can use the Durbin-Wu-Hausman test (Hausman, 1978) to decide whether to use a fixed effects or random effects model.

The random effects model is similar to the fixed effects model but allows the parameters to be random. The model is

$$Y_{ij} = \mu + b_j^* + \epsilon_{ij}$$

where the prior is that $b_j^* \sim N(0, \tau^2)$. In this case, the Best Linear Unbiased Predictor (BLUP) of b_j^* is

$$\hat{b}_j^* = \frac{\tau^2}{(\sigma^2/n + \tau^2)} (\bar{Y}_j - \bar{Y}).$$

So the estimate shrinks towards zero by an amount that depends on the magnitudes of τ^2 and σ^2 . More completely, one obtains a posterior distribution for each shop's effect:

$$b_j^* | Y \sim N \left(\frac{\tau^2}{(\sigma^2/n + \tau^2)} (\bar{Y}_j - \bar{Y}), \frac{\tau^2 \sigma^2 / n}{\tau^2 + (\sigma^2 / n)} \right).$$

Shrinkage borrows strength by saying that the charges at the j th shop tend to be similar to those at the other shops. This could enable an attorney understand how much variation there is in average tune-up fees across different shops, and thus to better assess whether the suspect shop's charges are extreme. This methodology would be directly applicable to cases alleging wage discrimination, excessive pollution, or medical malpractice, where the observations would be salaries in similar jobs, daily emissions of nitrogen oxide, or survival times of patients following a procedure, respectively.

Besides fixed effects and random effects models, there are *mixed effects* models (cf. Sahai and Ageel, 2012, for a discussion of all three types of models). These arise when some of the variables are not random, but others are. In the repair shop example, where older vehicles would be expected to require more labor, there might be information on the age of the vehicle, A_i , which is a fixed effect. In that case, the model is

$$Y_{ij} = \mu + b_j^* + \beta A_i + \epsilon_{ij}$$

where β is a non-random unknown fixed effect and the b_j^* are random effects.

Let the vector of scores be Y , the vector of ages be a , and the vector of errors (individual effects) be ϵ (all are $n \times 1$). There are k shops, and let the $k \times 1$ vector of shop effects be b^* . Suppose $b^* \sim N(0, \tau^2 I)$, $\epsilon \sim N(0, \sigma^2 I)$, and that these two random vectors are independent.

With these assumptions (which can be made much more general), the point estimates are:

$$\begin{pmatrix} \hat{\beta} \\ \hat{b}^* \end{pmatrix} = \begin{bmatrix} a'a/\tau^2 I & a'Z/\tau^2 I \\ Z'a/\tau^2 I & Z'Z/\tau^2 I + \sigma^{-2} I \end{bmatrix}^{-1} \begin{bmatrix} a'Y/\tau^2 I \\ Z'Y/\tau^2 I \end{bmatrix}$$

where I is the identity matrix (i.e., a square matrix that has ones on the diagonal from top left to bottom right, and zeroes for all other entries) and Z is the $n \times k$ design matrix that assigns vehicles to shops. Specifically, the i, j th entry of Z is zero unless vehicle i is tuned-up at shop j , in which case the entry is 1.

The important point is that one can improve estimates using Bayesian methods to borrow strength from similar cases, while still incorporating known covariates. Thus, an age discrimination suit could control for fixed effects, such as years of education, while still using random effects models to identify inequities. Or one could use fixed effects such as hours of operation to assess the amount of pollution from a factory, or the severity of illness when looking at patient survival times.

A second useful technique is hierarchical Bayesian models. In these, one places a prior over the parameters in the prior distribution, and one can extend this to higher levels by placing a third prior on the parameters in the second prior (forming a hierarchy). These models tend to be robust, in the sense that the posterior distribution is less sensitive to the prior specification. Additionally, there are many situations where this approach facilitates the construction of reasonable models.

The random and mixed effects repair-shop cases were two-level hierarchical models. There was a prior on the the distribution of the effect, which was input to the model for the data. In general, a three-level hierarchical model finds

$$\mathbb{P}(\theta, \phi | X) = \frac{\mathbb{P}[X | \theta, \phi] \mathbb{P}[\theta, \phi]}{\mathbb{P}[X]} = \frac{\mathbb{P}[x | \theta] \mathbb{P}[\theta | \phi] \mathbb{P}[\phi]}{\mathbb{P}[X]}.$$

There can be many more levels, but the mathematics is essentially the same—there are just more parameters.

A third area in which Bayesian methods are attractive is variable selection. A famous example in which this would apply is the case brought against Harris Trust & Savings Bank for wage discrimination against women (U.S. Department of the Treasury v. Harris Trust and Savings Bank, 1979). The bank denied discriminating and contended that wages were based only on seniority, experience, education, and age.

The case was complex and had many facets, but a statistical analysis by Harold Roberts presented during the trial used a multiple linear regression model in which the response variable was salary and the explanatory variables were seniority, experience, education, age, and gender. Roberts did a standard frequentist t -test to determine whether the gender coefficient was significantly different from zero. The test showed that the coefficient was very different from zero, which strongly suggested that gender was indeed a factor in determining wages. (Although experts continue to use such multiple linear regression in discrimination litigation, there are better tools.)

The Bayesian alternative (not used in the case), has a prior that puts probability p on the value of the zero for the gender coefficient, and probability $1 - p$ on the gender coefficient being a random draw from a very dispersed distribution, such a normal distribution with mean 0 and standard deviation 5,000. The Bayesian would use the salary data to get a posterior distribution on the coefficient, and presumably would find that the posterior probability p^* that the coefficient is zero is quite small, and that the posterior probability that the coefficient is negative (indicating discrimination against women) and large. The analyst would also calculate the posterior mean of the coefficient, which would also be useful in assessing damages.

The Bayesian world is big, and there are many other kinds of methods and applications that are relevant to legal questions. But Bayesian tools should not be applied without

guidance from experts with advanced degrees in statistics. Nor, for that matter, should frequentist analyses be done by non-professionals.

3.5 Summary

Bayesian inference is a versatile and well-established method of modern statistical inference. People are generally more likely to express opinions in terms of probability, as a Bayesian does, than in the contorted language of frequentist exposition.

However, Bayesian methods are less common in legal cases than they are in published scientific articles. There are several reasons for this: courts have accepted the use of other methods in the past; realistic applications are computer-intensive; and using prior information in framing an analysis and defending its conclusions can create difficulties for lawyers.

A much richer discussion of the Bayesian perspective in the context of the law is given in the book *Statistics and the Law*, edited by De Groot et al. (1986). It is dated only in that it precedes the invention of MCMC techniques, but the fundamental issues and philosophy are addressed from many angles. Case law and evidentiary issues are discussed in Kaye et al. (2011).

Acknowledgments

This work was done as part of the research program on Forensic Statistics that was organized by the Statistical and Applied Mathematical Sciences Institute, with support from the National Science Foundation.

References

- Bayes, T. and Price, R. An essay towards solving a problem in the doctrine of chance. By the Late Rev. Mr. Bayes, Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. *Philosophical Transactions of the Royal Society of London*, **53**, 370–418, 1763.
- Benjamini, Y. and Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, **57**, 289–300, 1995.
- CDC. Estimated HIV incidence and prevalence in the United States, 2010–2016. HIV Surveillance Supplemental Report 2019;24(No. 1), 2017.
- Chou, R., Huffman, L.H., Fu, R., Smits, A.K., and Korthuis, P.T. Screening for HIV: A review of the evidence for the U.S. preventive services task force. *Annals of Internal Medicine*, **143**(1), 55–73, July 2005.
- DeGroot, M.H., Fienberg, S.E., and Kadane, J.B. (Eds.). *Statistics and the Law*. New York: Wiley, 1986.
- Department of Justice. Uniform language for testimony and reports using forensic autosomal DNA examinations using probabilistics genotyping systems. <https://www.justice.gov/olp/page/file/1095961/download>. Verified on January 12, 2020, 2018.

- Derksen, T. and Meijsing, M. The fabrication of facts: the lure of the incredible coincidence. In *Legal Evidence and Proof Statistics, Stories, Logic*, pages 39–70, 2009.
- Fienberg, S.E. and Kadane, J.B. The presentation of Bayesian statistical analyses in legal proceedings. *Journal of the Royal Statistical Society: Series D (The Statistician)*, **32**(1–2), 88–98, 1983.
- Fienberg, S.E. and Kaye, D.H. Legal and statistical aspects of some mysterious clusters. *Journal of the Royal Statistical Society: Series A*, **154**, 61–74, 1991.
- Gamerman, D. and Lopes, H.F. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Boca Raton, FL: Chapman and Hall/CRC, 2006.
- Gelfand, A.E. and Smith, A.F. Sampling-based approaches to calculating marginal densities. *Journal of the American Atatistical Association*, **85**, 398–409, 1990.
- Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., and Rubin, D.B. *Bayesian Data Analysis*. Boca Raton, FL: Chapman and Hall/CRC, 2013.
- Gelman, A., Hill, J., and Yajima, M. Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, **5**, 189–211, 2012.
- Good, I.J. "Weight of evidence: A brief survey" (with discussion). In *Bayesian Statistics 2*, J. Bernardo, M. DeGroot, D. Lindley, and A. Smith, editors, pages 249–269. Amsterdam: Elsevier Science Publishers, 1985.
- Hausman, J. Specification tests in econometrics. *Econometrica*, **46**, 1251–1271, 1978.
- Jeffreys, H. *The Theory of Probability* (3rd ed.). Oxford University Press, England, 1998 [1961].
- Kadane, J.B., Girón, J., Peña, D., Fishburn, P., French, S., Lindley, D.V., Parmigiani, G., and Winkler, R.L. Several Bayesians: A review. *Test*, **2**, 1–32, 1993.
- Kass, R.E. and Raftery, A.E. Bayes factors. *Journal of the American Statistical Association*, **90**, 773–795, 1995.
- Kass, R.E. and Wasserman, L. Formal rules for selecting prior distributions: A review and annotated bibliography. *Journal of the American Statistical Association*, **91**, 1343–1370, 1996.
- Kaye, D.H. The numbers game: Statistical inference in discrimination cases. *Michigan Law Review*, **80**, 833–856, 1982.
- Kaye, D.H., Bernstein, D.E., and Mnookin, J.L. *The New Wigmore: A Treatise on Evidence: Expert Evidence* (2nd ed.). Aspen Pub., New York, 2011.
- Loue, S. *Forensic Epidemiology: Integrating Public Health and Law Enforcement*. Jones & Bartlett Learning, 2010.
- McElreath, R. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Chapman and Hall/CRC, 2018.
- Muller, P., Parmigiani, G., and Rice, K. FDR and Bayesian multiple comparisons rules, Johns Hopkins University, Department of Biostatistics Working Paper 115, 2006.
- National Research Council. *Forensic Analysis: Weighing Bullet Lead Evidence*. The National Academies Press, Washington, DC, 2004. <https://doi.org/10.17226/10924>.
- O'Hagan, A., Buck, C.E., Daneshkhah, A., Eiser, J.R., Garthwaite, P.H., Jenkinson, D.J., Oakley, J., and Rakow, T. *Uncertain Judgements: Eliciting Experts' Probabilities*. John Wiley & Sons, 2006.
- Sahai, H. and Ageel, M.I. *The Analysis of Variance: Fixed, Random and Mixed Models*. Springer Science and Business Media, 2012.
- Savage, L.J. *The Foundations of Statistics*. Wiley, New York, NY, 1954.
- Scott, J.G. and Berger, J.O. An exploration of aspects of Bayesian multiple testing. *Journal of Statistical Planning and Inference*, **136**, 2144–2162, 2006.
- U.S. Department of the Treasury v. Harris Trust and Savings Bank. U.S. Department of Labor Administrative Proceeding, Case No. 78-OFCCP-2 (Burrow, ALJ), 1979.
- Westfall, P., Johnson, W., and Utts, J. A Bayesian perspective on Bonferroni adjustment. *Biometrika*, **84**, 419–427, 1997.
- Young, G.A., Severini, T.A., and Smith, R.L. *Essentials of Statistical Inference*, Vol. 16. Cambridge University Press, 2005.