# 5 The Bayesian Vantage for Dealing with Uncertainty

*D. A. Evans, M. C. Newman,*
*M. Lavine, J. S. Jaworska, J. Toll,*
*B. W. Brooks, and T. C. M. Brock*

## 5.1 INTRODUCTION

Bayesian approaches are discussed throughout this book. Unfortunately, because frequentist methods are typically presented in introductory statistics courses, most environmental scientists do not clearly understand the basic premises of Bayesian methods. This lack of understanding could hamper appreciation for Bayesian approaches and delay the adaptation of these valuable methods for analyzing uncertainty in risk assessments.

Bayesian statistics are applicable to analyzing uncertainty in all phases of a risk assessment. Bayesian or probabilistic induction provides a quantitative way to estimate the plausibility of a proposed causality model (Howson and Urbach 1989), including the causal (conceptual) models central to chemical risk assessment (Newman and Evans 2002). Bayesian inductive methods quantify the plausibility of a conceptual model based on existing data and can accommodate a process of data augmentation (or pooling) until sufficient belief (or disbelief) has been accumulated about the proposed cause–effect model. Once a plausible conceptual model is defined, Bayesian methods can quantify uncertainties in parameter estimation or model predictions (predictive inferences). Relevant methods can be found in numerous textbooks, e.g., Carlin and Louis (2000) and Gelman et al. (1997).

Bayesian fundamentals are reviewed here because several chapters in this volume apply these methods in complex ways to assessing uncertainty. The goal is to create enough understanding so that methods described in later chapters can be fully appreciated.

## 5.2 CONVENTIONAL (FREQUENTIST) INFERENCE METHODS

The standard tools of statistical inference, including the concept and approaches of constructing a null hypotheses and associated $p$ values, are based on the frequentist view of probability. From a frequentist perspective, the probability of an event is defined as the fraction of times that the event occurs in a very large number of trials (known as a probability limit). Given a hypothesis and data addressing it, the classical procedure is to calculate from the data an appropriate statistic, which is typically

a single number. Based on the hypothesis being true and other assumptions, the probability distribution of this statistic is a known function.

This distribution, together with the numerical value of the statistic, allows an assessment of how "unusual" the data are, assuming that the hypothesis is valid. The $p$ value is the probability that the observed value of the statistic (or values even more extreme) occur. The data are declared significant at a particular level ($\alpha$); if $p < \alpha$, the data are considered sufficiently "unusual" relative to the hypothesis and the hypothesis is rejected. Standard, albeit arbitrary, values of $\alpha$ are taken as 0.05 and 0.01. Let us suppose that a particular data set gives $p = 0.02$. From the frequentist vantage, this means that, if the hypothesis were true and the whole experiment were to be repeated many times under identical conditions, in only 2% of such trials would the value of the statistic be "more unusual or extreme" than the value actually observed. One then prefers to believe that the data are not, in fact, "unusual"* and concludes that the assumed hypothesis is untenable.

It is important to note that the conclusion drawn from the observed data is based on a comparison with virtual data that might have been collected in other identical experiments but were never really observed. In fact, a judgement is made on the data rather than directly on the model or hypothesis. No consideration is given to the plausibility of the original hypothesis or specific alternatives. It is an erroneous assumption that the $p$ value is a measure of the validity of the null hypothesis. As noted, $p$ merely makes a statement about the data on the assumption that the hypothesis is valid.

While this is an almost universally used technique for testing hypotheses, the procedure can produce some odd or ambiguous conclusions. The following example, from the suggestion of Lindley and Phillips (1976), is quoted by Carlin and Louis (2000). We test the null hypothesis $H_0$:$\theta = 0.5$ for the probability that a given coin will turn up "heads" after a toss; the alternative hypothesis is $H_a$:$\theta > 0.5$, i.e., the coin is biased toward "heads." Twelve independent tosses result in 9 heads and 3 tails. In this case, the choice of test statistic is simple; it is the number of heads, denoted by $r$. The binomial distribution gives the probability of obtaining $r$ heads in 12 tosses as the following:

$$p\left(r\middle|\theta, n=12\right)=\binom{12}{r}\theta^{r}\left(1-\theta\right)^{12-r} \tag{5.1}$$

The $p$ value is given by

$$\sum_{r=9}^{r=12} p\left(r\middle|\theta=0.5, n=12\right)=0.073 \tag{5.2}$$

* That is, the observed statistic is considered to be a sample from the "center" of some (unknown) distribution whose form depends upon the true (unknown) hypothesis.



This is the probability of obtaining the observed number of heads, or more extreme (i.e., larger) values, when $H_0$ is assumed true ($\theta = 0.5$). Thus, $H_0$ is not rejected at the 5% level; to observe 9 heads, or more, in 12 tosses, is not sufficiently unusual for a coin with $\theta = 0.5$.

The above treatment has implicitly assumed that the experimental design was such that the number of trials was fixed at 12 and the observation was the number of heads. However, an alternative design could have been to continue tossing the coin until 3 tails were obtained, and the observation would be $n$, the number of tosses required to produce the 3 tails. In this case, the statistic for judging the data is just $n$. But the distribution of $n$, the number of tosses to produce 3 tails, is given by the negative binomial:

$$p\left(n\middle|\theta, r=3\right)=\binom{n-1}{2}\theta^{n-3}\left(1-\theta\right)^{3} \tag{5.3}$$

and the $p$ value for the experiment is given by

$$\sum_{n=12}^{\infty} p\left(n\middle|\theta=0.5, r=3\right)=0.033 \tag{5.4}$$

This is the probability of a result of $n = 12$, or more extreme values, given that $\theta = 0.5$ ($H_0$ assumed true). The result calls for rejection of $H_0$ at the 5% level.

The difference arises because the identification of which of the data element is the random variable differs between the 2 designs. It is $r$, the number of heads, in the first case and $n$, the number of tosses, in the second. The $p$ values compare the actually observed data with the data from an infinite number of virtual experiments (the frequentist approach). In the first case, all these experiments have 12 tosses and varying numbers of heads: in the second, they all have 3 tails and varying numbers of tosses.

Critics of the frequentist approach consider this disturbing. The actual observations: "in 12 tosses of a coin, 9 heads and 3 tails were observed" should not lead to 2 different conclusions dependent only upon the choice of when to stop the experiment (at 12 tosses or at 3 tails).

## 5.3 EXPERIMENTS CHANGE THE STATE OF KNOWLEDGE

The basic premise of the Bayesian approach is that observations change the state of knowledge of a system. Let us suppose for simplicity that the item of interest is some parameter, $\theta$, describing a state of nature (as in the above example, where $\theta$ was a property of the coin and the conditions under which it was tossed). Figure 5.1 indicates symbolically the development of knowledge.

The extent of knowledge about $\theta$ can be quantified by showing that probability also can be interpreted as "degree of belief" (Lindley 1965), "measure of plausibility" (Loredo 1990), or "personal probability" (O'Hagan 2001). Early workers such as

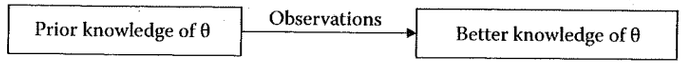


**FIGURE 5.1** Observations contribute to knowledge.

Bernoulli (1713) held this view of probability. Laplace (1812, 1820, 1951) described probability theory as "commonsense reduced to calculation." And, in Laplace's epistemic context, probability "expresses numerically degrees of uncertainty in light of data" (Howson and Urbach 1989). A large part of the motivation for the initial studies by workers such as Bernoulli and Laplace derived from the sponsorship of gambling noblemen. In fact, the problems addressed might now be called risk assessment because the noblemen wished to conduct their gaming so as to reduce their risk of loss and increase the "risk" of winning. Probability theory was intended to assist such decision making.

The above approach, which was attacked as being too vague to be the starting point of any theory of probability, led eventually to the frequentist approach, where probability was defined in a manner that assigns a numerical value, albeit a value that cannot ever be measured, since it requires an infinite number of trials

The numerical properties of probability and degrees of belief can be defined effectively and sensibly using a few axioms.

## 5.4 RULES OF PROBABILITY

Probability can be defined as a limiting case of a frequency ratio, and from this view the various rules of probability can be derived. An alternative approach is an axiomatic one that states that there is a quantity called probability associated with events and that it possesses assigned properties. The former is largely the frequentist point of view, the axiomatic approach is shared by Bayesians and non-Bayesians alike.

Probability values lie continuously in the range 0 to 1 inclusive, where the endpoints zero and unity are identified with impossibility and certainty, respectively. This follows immediately for the frequentist; for the axiomatic approach it is adopted as an axiom, but one imbued with Laplace's "commonsense." Any other range could be chosen at the cost of greater difficulty of interpretation.

Suppose $A$ and $B$ are events, then,

$$p(AB) = p(A|B) \cdot p(B) \tag{5.5}$$

where "$AB$" means that both events occur, or both propositions are valid. The notation introduces $p(A|B)$, the probability of $A$ conditional on $B$. For the frequentist, it means $\lim_{N\to\infty}\left(n(AB)/n(B)\right)$, being the frequency ratio of occurrence of $A$ on all the occasions when $B$ occurred ($n$ => $N$). If $p(B) \neq 0$ then, as the total number of trials ($N$) tends to infinity, so does $n(B)$. The above relationship then follows quite directly:



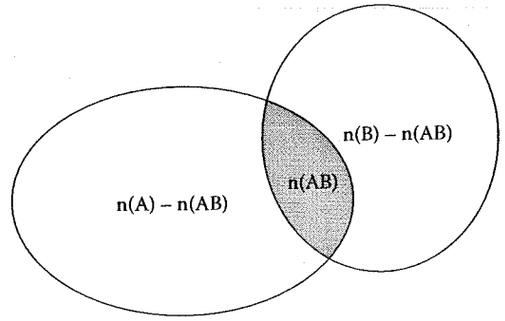


**FIGURE 5.2** Venn diagram illustrating the development of conditional probability.

$$\begin{aligned} p(AB) &= \lim_{N\to\infty}\left(\frac{n(AB)}{N}\right) = \lim_{N\to\infty}\left(\frac{n(AB)}{n(B)} \cdot \frac{n(B)}{N}\right) \\ &= \lim_{N\to\infty}\left(\frac{n(AB)}{n(B)}\right) \cdot \lim_{N\to\infty}\left(\frac{n(B)}{N}\right) \\ &= p(A|B) \cdot p(B) \end{aligned} \tag{5.6}$$

For the Bayesian, the relationship is taken as an axiom, but its motivation reflects the real world with the foreshadowing of rules implied by the above frequentist treatment.

Given the 2 events or propositions, $A$ and $B$, then

$$p(A \text{ or } B) = p(A) + p(B) - p(AB) \tag{5.7}$$

where "$A$ or $B$" means the inclusive "or," i.e., at least 1 of $A$ and $B$ occur. In Figure 5.2, it corresponds to the union of the 2 areas. The frequentist's numbers are shown in the various categories. From the figure it can be seen that

$$\begin{aligned} n(A \text{ or } B) &= \{n(A) - n(AB)\} + \{n(B) - n(AB)\} + n(AB) \\ &= n(A) + n(B) - n(AB) \end{aligned} \tag{5.8}$$

from which the result follows. The Bayesian takes the result as an axiom, motivated by the real world.

## 5.5 BAYES' THEOREM

The result $p(AB) = p(A|B) \cdot p(B)$ is symmetrical in $A$ and $B$ on the left side. It could equally well be written $p(BA)$, but

$$p(BA) = p(B|A) \cdot p(A) \tag{5.9}$$

thus

$$p(A|B) \cdot p(B) = p(B|A) \cdot p(A)$$

$$p(A|B) = \frac{p(B|A) \cdot p(A)}{p(B)} \tag{5.10}$$

Equation (5.10) is a statement of Bayes' theorem. Since the theorem is proved using results or axioms valid for both frequentist and Bayesian views, its use is not limited to Bayesian applications. Note that it relates 2 conditional probabilities where the events $A$ and $B$ are interchanged.

Bayesian interpretation and application of the theorem quantifies the development of information. Suppose that $A$ is a statement or hypothesis, and let $p(A)$ stand for the degree of belief in the statement or hypothesis $A$, based on prior knowledge, it is called the prior probability. Let $B$ represent a set of observations, then $p(B|A)$ is the probability that those observations occur given that $A$ is true. This is called the "likelihood" of the data and is a function of the hypothesis. The left side, $p(A|B)$, is the new degree of belief in $A$, taking into account the observations $B$, it is called the posterior probability. Thus Bayes' theorem tracks the effect that the observations have upon the changing knowledge about the hypothesis. The theorem can be expressed thus:

$$\text{posterior probability} \propto \text{likelihood function} \times \text{prior probability} \tag{5.11}$$

Figure 5.3 is a copy of Figure 5.1 showing the portions of the Bayes formulation pertaining to each part of the development of knowledge.

The argument can be extended to treat multiple hypotheses. Suppose $A_1$ and $A_2$ are competing hypotheses. Then Bayes' theorem gives the following:

$$p(A_1|B) = \frac{p(B|A_1) \cdot p(A_1)}{p(B)}; \quad p(A_2|B) = \frac{p(B|A_2) \cdot p(A_2)}{p(B)} \tag{5.12}$$

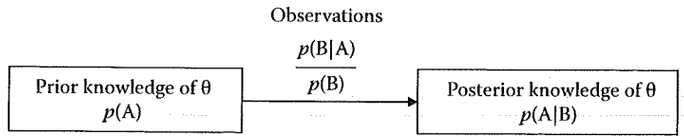


**FIGURE 5.3** The contributions of the components of Bayes' theorem to the development of knowledge.



giving the ratio

$$\frac{p(A_1|B)}{p(A_2|B)} = \frac{p(B|A_1)}{p(B|A_2)} \cdot \frac{p(A_1)}{p(A_2)} \tag{5.13}$$

which can be written in words as

$$\text{posterior odds ratio} = \text{likelihood ratio} \times \text{prior odds ratio.}$$

The coin-tossing experiment can be analyzed using this approach. As before, let $\theta$ be the probability of "heads." The hypothesis $\theta = 0.5$ is essentially meaningless because $\theta$ is a continuous parameter. Let the 2 hypotheses $A_1$ and $A_2$ be $\theta < 0.5$ and $\theta > 0.5$, respectively. The "prior odds ratio" represents an initial assessment of the relative probabilities or degree of belief of the 2 hypotheses. In the absence of any previous knowledge, a "noninformative prior" is used; in this case, we may assume it equally likely that the coin is biased to heads or tails, i.e., $p(A_1) = p(A_2)$. Consequently, the prior odds ratio is unity. Recognizing the earlier ambiguity whether the binomial or negative binomial distribution is applicable, we shall calculate the likelihood function, $p(B|A_i)$ using each.

For the binomial distribution the likelihood function is

$$p(9|n = 12, \theta) = \binom{12}{9} \theta^9 (1-\theta)^3 \tag{5.14}$$

and for the negative binomial it is

$$p(12|\theta, r = 3) = \binom{11}{2} \theta^9 (1-\theta)^3 \tag{5.15}$$

The functional dependence upon $\theta$ is identical for the 2 distributions. The coefficients cancel out when the likelihood is used in Bayes' theorem since they also appear in

$$p(B) = \int_0^1 p(B|\theta) d\theta$$

they also cancel out in the likelihood ratio. The approach does not suffer from ambiguity depending upon the design of the experiment; only the data are important in conformance with the likelihood principle, which states that the likelihood function expresses all the information that can be inferred about the parameter, $\theta$, from the observed data.

Suppressing the unimportant coefficients, the likelihoods for the 2 hypotheses are obtained by integrating over the values of θ covering the appropriate range:

$$A_1,\quad p\left(r=9\middle|n=12,\theta<0.5\right)\propto\int_0^{0.5}\theta^9\left(1-\theta\right)^3 d\theta$$

$$=1.613\times10^{-5}$$

$$A_2,\quad p\left(r=9\middle|n=12,\theta>0.5\right)\propto\int_{0.5}^{1}\theta^9\left(1-\theta\right)^3 d\theta \qquad (5.16)$$

Thus the likelihood ratio is 0.048. The new state of knowledge concerning θ is then expressed by the posterior odds:

$$\frac{p\left(A_1\middle|B\right)}{p\left(A_2\middle|B\right)}=0.048 \qquad (5.17)$$

This is a statement of the relative plausibility of the 2 hypotheses based on the observations. If one were a betting person, one would offer odds of 19 to 1 against the coin being biased toward "tails."

## 5.6 EXAMPLES RELEVANT TO UNCERTAINTY IN RISK ASSESSMENT QUANTIFYING PLAUSIBILITY OF A CAUSE–EFFECT MODEL

Central to any risk assessment is a model of causality. At the onset, a conceptual model is needed that identifies a plausible cause–effect relationship linking stressor exposure to some effect. Most ecological risk assessments rely heavily on weight-of-evidence or expert opinion methods to foster plausibility of the causal model. Unfortunately, such methods are prone to considerable error (Lane et al. 1987; Hutchinson and Lane 1989; Lane 1989), and attempts to quantify that error are rare. Although seldom used in risk assessment, Bayesian methods can explicitly quantify the plausibility of a causal model.

Let's use a fictitious example to illustrate the application of Bayes' theorem to quantifying the level of belief warranted in a causal model. A fishkill is observed below a discharge and the question is asked, "Did a toxic release (e.g., greater than LC10) from the point source cause the fishkill?" From the literature, one gathers evidence to assess this causal hypothesis. From a toxicological study of the major chemical of concern in the discharge, the likelihood of a fishkill if the discharge concentration was greater than LC10 (i.e., $p$(Fishkill|Release) > LC10) is calculated to be 0.400. From historical discharge records, it is also calculated that the probability of a discharge toxicant concentration greater than LC10 (i.e., $p$(Release) > LC10) is 0.005. From records of fishkills in this and similar streams of the region, the



likelihood of a fishkill ($p$(Fishkill)) is 0.003. This information can be applied with Bayes' theorem (Equation (5.10) to estimate the probability that there was a toxic discharge (>LC10) given the observed fishkill:

$$p(\text{Release}\mid\text{Fishkill})=\frac{p(\text{Fishkill}\mid\text{Release})\cdot p(\text{Release})}{p(\text{Fishkill})}$$

Inserting the above estimates into the right side of the equation gives a $p$(Release|Fishkill) of 0.666. Based on the evidence, the odds are 2 to 1 that the discharge caused the fishkill. Is this evidence sufficient to take regulatory action? Likely, it is not. One would need to gather more information in order to produce a level of belief sufficient to decide whether or not regulatory action was required. Assume that a characteristic lesion was found on the dead fish and that we know from the literature that $p$(Lesion|No Toxicant Exposure) = 0.010 and that $p$(Lesion|Exposure to the Discharge Toxicant) = 0.540. The likelihood ratio is 0.54:0.01 or 54:1. The posterior odds of 2:1 just calculated can become our new prior odds, and, based on this new evidence and Equation (5.11), new posterior odds of the toxic release having caused the fishkill can be calculated:

$$\text{Posterior Odds} = \text{Likelihood Ratio} \times \text{Prior Odds} = 54 \times 2 = 108{:}1$$

Based on this evidence, the odds that a toxic release caused the fishkill is a convincing 108 to 1. The level of belief is now sufficiently high for a reasonable person to take regulatory action. Bayes' theorem allowed optimal use of evidence to define the belief warranted in the causal hypothesis that a toxic release caused the fishkill: evidence changed our state of knowledge about the fishkill.

### 5.6.1 Estimating Indoor Radon Exposure

Empirical Bayes methods were applied to estimate geometric means (GM) of indoor radon concentrations for Minnesota counties (Price et al. 1996). Data were collected unevenly among counties, with some counties having very low numbers of samples. Consequently, counties with low sample numbers had more error in GM estimates than adequately sampled counties. Bayesian methods allowed estimation of GM and associated variance despite these differences in county sample sizes. Even if no measurements were available for a given county, there is nonetheless some knowledge about the county GM. Denoting the logarithm of the GM by θ, the GM were assumed to be log-normally distributed among the counties based on existing data, i.e., the state of knowledge of θ is represented by $p(\theta) = N(\mu, \sigma^2)$, where μ = the "true" mean of the logarithm of radon concentration over all counties. The $p(\theta)$ is the informative prior distribution. Also, radon concentrations were judged to be log-normally distributed within counties based on results for amply sampled counties: $N(\mu, \sigma^2)$ for the logarithm of radon concentration. New estimates of county GM were then estimated with Bayes' theorem,

$$p(\theta|y) = \frac{p(\theta)\,p(y|\theta)}{p(y)} \tag{5.18}$$

where $p(\theta|y)$ = the probability that the true mean is $\theta$ given the data $y$, and $p(y|\theta)$ = the likelihood or probability of the data set, $y$, given $\theta$. The $p(y)$ is a constant that, in practice, is estimated such that the right side of the equation integrates to unity (O'Hagan 2001). The "true" GM's of county radon concentrations were estimated with a modification of this equation and sample-size weighting of county geometric means. The informative prior distribution as modified by the likelihood of getting the data, $y$, for a county given $\theta$ and a better estimate of $\theta$ was produced: one "learned" from a particular county's data to produce a better estimate. The value of $\theta$ that maximizes Equation 5.18 can be considered a "best guess" of the true value of $\theta$.

### 5.6.2 Species Sensitivity Example

Suppose that we wish to know the species sensitivity distribution (SSD) for a new pesticide, chemical A. Specifically, we wish to know the collection of LC50 values for many species. Unfortunately, chemical A has been tested on only a very limited number of species. For each species, an LC50 value has been estimated. Suppose also that pesticide B, having similar chemical structure and identical mode of action, has been tested on many species. Can we use the information about B to help us estimate the SSD for A and, if so, how? One way, of course, is informal. We take our knowledge of B and our subject matter knowledge, cogitate for a while, and come up with our best guess for the species sensitivity curve for A. However, a Bayesian approach provides a more formal, quantitative method for using the information about B.

We begin with a model for the shape of the SSD. For the sake of argument, we will assume that the SSD of B is approximately normal. That is, the histogram of the LC50 values for pesticide B looks approximately like a normal density with mean $\mu_B$ and variance $\sigma_B^2$. We may reasonably expect the SSD of A also to be normal with unknown mean $\mu_A$ but the same variance, $\sigma_A^2 = \sigma_B^2$. Standard statistical theory tells us how to estimate $\mu_A$ and $\sigma_A^2$ from the few species that have been tested with A. But Bayesian statistics goes a bit further by telling us also how to use the information about pesticide B.

The fact that A and B are so similar chemically suggests that their SSDs will also be similar. We can model that by saying that $\mu_A$ is likely to be within a range of $\mu_B$ plus or minus, for example, 200. This is formally expressed by a statement such as

$$\mu_A \sim N(\mu_B, 100^2) \tag{5.19}$$

i.e., $\mu_A$ has a normal distribution with mean $\mu_B$ and variance $100^2$. The $100^2$ arises because we treat the range of ±200 as about 2 standard deviations; so 1 standard deviation is 100 and the variance is $100^2$. Equation 5.19 is the prior probability distribution for $\mu_A$. Suppose that each species tested with chemical A yields an LC50 value. Then Bayes' theorem and Bayesian statistics provide the formula for combining the



prior distribution of $\mu_A$ with the data to yield the posterior distribution. Suppose that there were 4 species tested with A that yielded LC50 values of $y_1, y_2, y_3, y_4$.

The likelihood function, $p(\text{data}|\mu_A)$, for these data is

$$\begin{aligned} p(\{y_1,y_2,y_3,y_4\}|\mu_A) &= p(y_1|\mu_A)\cdot p(y_2|\mu_A)\cdot p(y_3|\mu_A)\cdot p(y_4|\mu_A) \\ &\propto e^{-(y_1-\mu_A)^2/2\sigma_B^2}\cdot e^{-(y_2-\mu_A)^2/2\sigma_B^2}\cdot e^{-(y_3-\mu_A)^2/2\sigma_B^2}\cdot e^{-(y_4-\mu_A)^2/2\sigma_B^2} \\ &\propto e^{-\left[\sum(y_i-m_A)^2/\sigma_B^2\right]/2} \\ &\propto e^{-\left[(\bar{y}-\mu_A)^2/(\sigma_B^2/4)\right]/2} \end{aligned} \tag{5.20}$$

where $\bar{y} = (1/4)(y_1 + y_2 + y_3 + y_4)$, i.e., the mean of the observations. It is assumed that the observations are independent samples from a normal distribution with variance $\sigma_B^2$. The likelihood function of $\mu_A$ is a normal curve with the maximum at $\bar{y}$, and a variance of $\sigma_B^2/4$. Bayes' theorem gives the posterior distribution of $\mu_A$ as

$$\begin{aligned} p(\mu_A|\{y_1,y_2,y_3,y_4\}) &= p(\{y_1,y_2,y_3,y_4\}|\mu_A)\cdot p(\mu_A) \\ &\propto e^{-\left[(\bar{y}-\mu_A)^2/(\sigma_B^2/4)\right]/2}\cdot e^{-\left[(\mu_A-\mu_B)^2/100^2\right]/2} \\ &\propto e^{-\left[\left[(\bar{y}-\mu_A)^2/(\sigma_B^2/4)\right]+\left[(\mu_A-\mu_B)^2/100^2\right]\right]/2} \end{aligned} \tag{5.21}$$

Some algebra reveals that the posterior distribution of $\mu_A$ is normal with mean

$$\frac{\left(y_1+y_2+y_3+y_4/\sigma_B^2\right)+\left(\mu_B/100^2\right)}{\left(4/\sigma_B^2\right)+\left(1/100^2\right)} \tag{5.22}$$

and variance

$$\frac{1}{\left(4/\sigma_B^2\right)+\left(1/100^2\right)} \tag{5.23}$$

These equations illustrate a common feature of Bayesian analysis: the posterior mean is a compromise between the prior mean and the data. In our example, as in every simple example with normally distributed data, the posterior mean is a weighted average of the prior mean and the data points. Each data point is weighted by the reciprocal of its variance, $1/\sigma_B^2$, just as the prior mean is weighted by the reciprocal of its variance, $1/100^2$. Because the reciprocal of a variance is such a useful concept, it is given a special name, precision. The posterior mean is just the weighted average

of the prior mean and the data; the weights are the precisions. The general formula is the following:

$$\text{posterior mean} = \frac{\sum(\text{data precision} \times \text{data value}) + \text{prior precision} \times \text{prior mean}}{\sum(\text{data precision}) + \text{prior precision}}$$

$$\text{posterior precision} = N \times \text{data precision} + \text{prior precision}$$

$$\text{posterior variance} = \frac{1}{\text{posterior precision}} \tag{5.24}$$

Terms in the formula get more or less weight according to their precision, i.e., according to how accurate they are as measures of $\mu_A$. The posterior precision measures how accurately we know $\mu_A$. It is the sum of the prior precision and the precisions of each of the data points. In our example that is

$$\frac{1}{100^2} + \frac{1}{\sigma_B^2} + \frac{1}{\sigma_B^2} + \frac{1}{\sigma_B^2} + \frac{1}{\sigma_B^2} \tag{5.25}$$

This simple example illustrates principles of Bayesian analysis and how it accommodates information from different sources. Real situations and real analyses can be more complicated than our example. For example, when species are tested with chemical A, we might not know their LC50 values exactly; instead, we might have estimates of LC50 values. Or we may have data on another similar chemical C. In each case, we would adjust the analysis to accommodate the more complicated situation.

### 5.6.3 Inference about Confidence Intervals

Confidence intervals are interpreted differently by frequentists and Bayesians. The 95% confidence interval derived by a frequentist suggests that the "true" value of some parameter ($\theta$) will be contained within the interval 95% of the time in an infinite number of trials. Note that each trial results in a different interval because the data are different. This statement is dependent on the assumed conditions under which the calculations were done, e.g., an infinite number of trials and identical conditions for each trial (O'Hagan 2001). Nothing can be said about whether or not the interval contains the "true" $\theta$.

The Bayesian approach reverses the role of the sample and model: the sample is fixed and unique, and the model itself is uncertain. This viewpoint corresponds more closely to the practical situation facing the individual researcher: there is only 1 sample, and there are doubts either what model to use, or, for a specified model, what parameter values to assign. The model uncertainty is addressed by considering that the model parameters are distributed. In other words Bayesian interpretation of a confidence interval is that it indicates the level of belief warranted by the data: the



posterior probability is 0.95 that the "true" $\theta$ is within the stated 95% confidence interval. Statements are made about $\theta$ based on the data alone, not an infinite number of virtual trials.

The classical or frequentist approach to probability is the one most taught in university courses. That may change, however, because the Bayesian approach is the more easily understood statistical philosophy, both conceptually as well as numerically. Many scientists have difficulty in articulating correctly the meaning of a confidence interval within the "classical" frequentist framework. The common misinterpretation: the probability that a parameter lies between certain limits is exactly the correct one from the Bayesian standpoint.

Apart from this pedagogical aspect (cf. Lee 1989, preface), there is a more technical reason to prefer the Bayesian approach to the confidence approach. The Bayesian approach is the more powerful one eventually, for extending a model into directions necessary to deal with its weaknesses. These are various relaxations of distributional assumptions. The conceptual device of an infinite repetition of samples, as in the frequentist viewpoint, does not yield enough power to accomplish these extensions.

Confidence intervals using frequentist and Bayesian approaches have been compared for the normal distribution with mean $\mu$ and standard deviation $\sigma$ (Aldenberg and Jaworska 2000). In particular, data on species sensitivity to a toxicant was fitted to a normal distribution to form the species sensitivity distribution (SSD). Fraction affected (*FA*) and the hazardous concentration (*HC*), i.e., percentiles and their confidence intervals, were analyzed. Lower and upper confidence limits were developed from *t* statistics to form 90% 2-sided classical confidence intervals. Bayesian treatment of the uncertainty of $\mu$ and $\sigma$ of a presupposed normal distribution followed the approach of Box and Tiao (1973, chapter 2, section 2.4). Noninformative prior distributions for the parameters $\mu$ and $\sigma$ specify the initial state of knowledge. These were constant $c$ and $1/\sigma$, respectively. Bayes' theorem transforms the prior into the posterior distribution by the multiplication of the classic likelihood function of the data and the joint prior distribution of the parameters, in this case $\mu$ and $\sigma$ (Figure 5.4).

The Bayesian equivalent to the frequentist 90% confidence interval is delineated by the 5th and 95th percentiles of the posterior distribution. Bayesian confidence intervals for SSD (Figures 5.4 to 5.5), 5th percentile, i.e., HC5 and fraction affected (Figures 5.4 to 5.6) were calculated from the posterior distribution. Thus, the uncertainties of both *HC* and *FA* are established in 1 consistent mathematical framework: *FA* estimates at the $\log_{10}$ HC lead to the intended protection percentage, i.e., $FA^{50}(\log_{10} HC_p^{50}) = p$ where $p$ is a protection level. Further full distribution of *HC* and *FA* uncertainty can be very easily extracted from posterior distribution for any level of protection and visualized (Figures 5.5 to 5.7).

For the normal distribution there are analytical solutions allowing the assessment of both *FA* and *HC* using frequentist statistics. In contrast, Bayesian solutions are numerical. This highlights the flexibility of the Bayesian approach since it can easily deal with any distribution, which is not always possible with the frequentist approach.

Aldenberg and Jaworska (2000) demonstrate that frequentist statistics and the Bayesian approach with noninformative prior results in identical confidence intervals for the normal distribution. Generally speaking, this is more the exception than the rule.

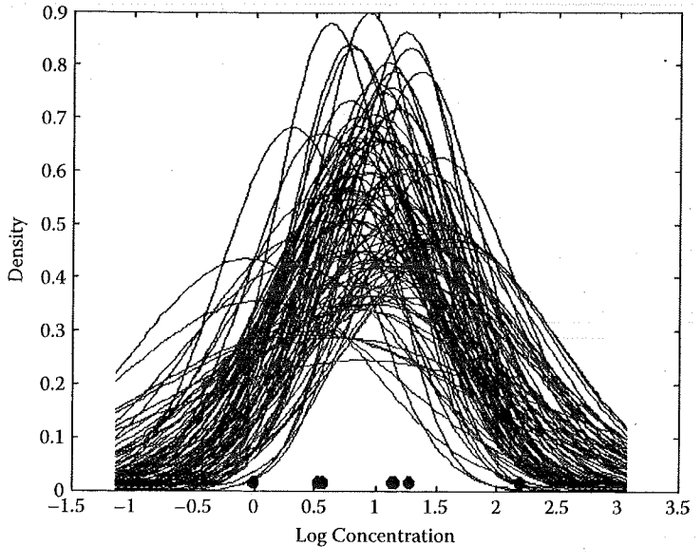


**FIGURE 5.4** Bayesian normal density "spaghetti plot": random sample of 100 normal probability density functions (pdfs) drawn from the posterior distribution of μ and σ, given 7 cadmium NOEC toxicity data (dots) from Aldenberg and Jaworska (2000).

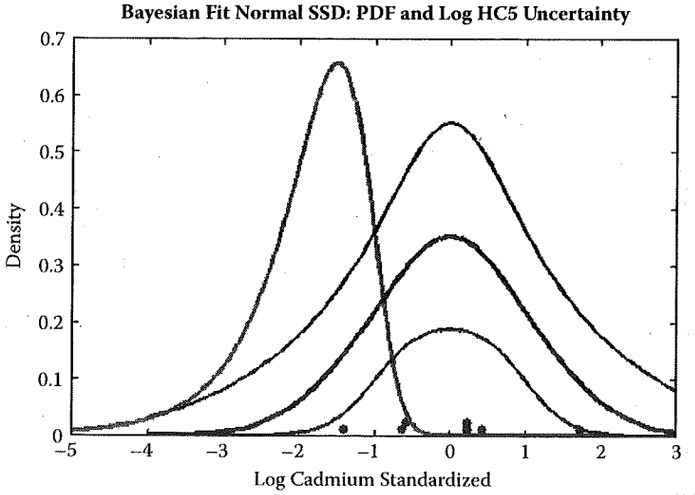


**FIGURE 5.5** Bayesian posterior normal probability density function values for SSD for cadmium and its Bayesian confidence limits: 5th, 50th, and 95th percentiles (black) and Bayesian posterior probability density of the HC5 (gray).



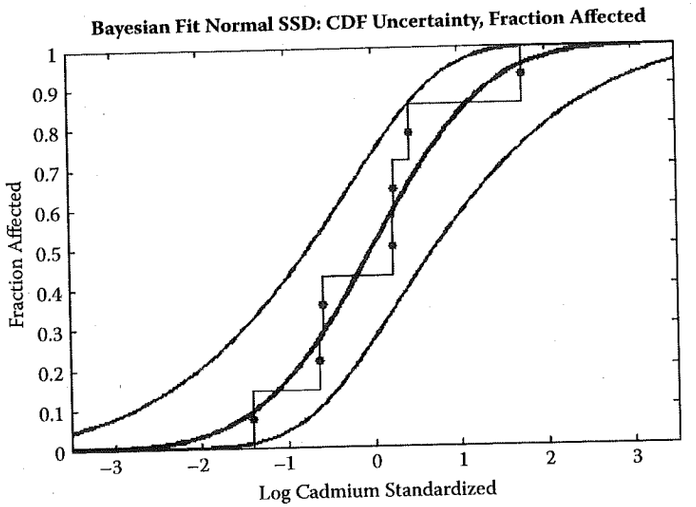


**FIGURE 5.6** Bayesian confidence limits of the fraction affected: percentiles (5th, 50th, and 95th) of posterior normal cdfs for cadmium. Data plotted cumulatively at $(i - 0.5)/n$ , with $i$ rank order, and $n$ the number of species tested.

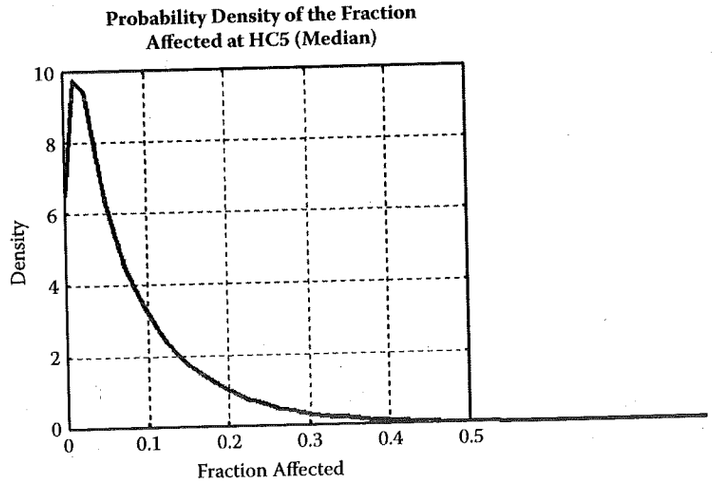


**FIGURE 5.7** Bayesian posterior probability density of the fraction affected at median log (HC5) for cadmium.

For those who feel more confident with the frequentist approach and find the Bayesian approach controversial to some extent, it is advantageous that both approaches yield the same answers in this simplest case. This might add confidence in the Bayesian approach for some practitioners.

## 5.7 CONCLUSION

The general Bayesian context is presented in this brief chapter with the intent of building sufficient understanding so that the reader can fully appreciate the methods presented with more complexity in following chapters. The distinction between the frequentist and Bayesian vantages was made using contrasting analyses of the outcomes of simple coin-toss trials. Then, the Bayesian theorem and associated concepts were explored briefly. Three examples relevant to uncertainty in risk assessments were given: estimation of the level of belief warranted for a causal model, estimation of exposure concentrations based on uneven sampling of a study area, and interpretation of confidence intervals. Hopefully, more involved Bayesian methods applied in later chapters will now be more easily understood.

## 5.8 REFERENCES

Aldenberg T, Jaworska J. 2000 Uncertainty of the hazardous concentration and fraction affected for normal species sensitivity distributions. Ecotox Env Saf 46:1–18.

Bernoulli J. 1713. Ars Conjectandi. Basel.

Box GEP, Tiao GC. 1973. Bayesian inference in statistical analysis. New York: Wiley Classics.

Carlin BP, Louis TA. 2000. Bayes and empirical Bayes methods for data analysis. 2nd ed. Boca Raton (FL): Chapman and Hall/CRC.

Gelman A, Carlin JB, Stern HS, Rubin DB. 1997. Bayesian data analysis. Boca Raton (FL): Chapman and Hall/CRC.

Howson C, Urbach P. 1989. Scientific reasoning. The Bayesian approach. La Salle (IL): Open Court.

Hutchinson TA, Lane DA. 1989. Assessing methods for causality assessment of suspected adverse drug reactions. J Clin Epidemiol 42:5–16.

Lane DA. 1989. Subjective probability and causality assessment. Appl Stochastic Model Data Anal 5:53–76.

Lane DA, Kramer MS, Hutchinson TA, Jones JK, Naranjo C. 1987. The causality assessment of adverse drug reactions using a Bayesian approach. Pharm Med 2:265–283.

Laplace PS. 1812. Theorie analytique des proabilitiés. Paris: Courcier.

Laplace PS. 1820. Essai Philosophique sur les Probabilitiés.

Laplace PS. 1951. *Philosophical essay on probability*. New York: Dover Publications (originally published as the introduction to Laplace 1812, Paris: Courcier).

Lee PM. 1989. Bayesian statistics: An introduction. Oxford (UK): Oxford University Press.

Lindley DV. 1965. Introduction to probability and statistics from a Bayesian viewpoint. 2 vols. Cambridge (UK): Cambridge University Press.

Lindley DV, Phillips LD. 1976. Inference for a Bernoulli process (a Bayesian view). Am Stat 30:112–119.

Loredo TJ. 1990. From Laplace to Supernova SN 1987A: A Bayesian inference in astrophysics. In: Fougère PF, editor. Maximum entropy and Bayesian methods. Dordrecht (DE): Kluwer.



Newman MC, Evans DA. 2002. Causal inference in risk assessments: cognitive idols or Bayesian theory? In: Newman MC, Roberts M, Hale R, editors. Coastal and estuarine risk assessment. Boca Raton (FL): CRC Press, p 73–96.

O'Hagan A. 2001. Uncertainty in toxicological predictions: the Bayesian approach to statistics. In: Rainbow PS, Hopkin SP, Crane M, editors. Forecasting the environmental fate and effects of chemicals. Chichester (UK): John Wiley, p 25–41.

Price PN, Nero AV, Gelman A. 1996. Bayesian prediction of mean indoor radon concentrations for Minnesota counties. Health Phys 71:922–936.

**Other Titles from the Society of Environmental Toxicology and Chemistry (SETAC)**

*Linking Aquatic Exposure and Effects: Risk Assessment of Pesticides*
Brock, Alix, Brown, Capri, Gottesbüren, Heimbach, Lythgo, Schulz, Streloke, editors
2009

*Derivation and Use of Environmental Quality and Human Health Standards for Chemical Substances in Water and Soil*
Crane, Matthiessen, Maycock, Merrington, Whitehouse, editors
2009

*Aquatic Macrophyte Risk Assessment for Pesticides*
Maltby, Arnold, Arts, Davies, Heimbach, Pickl, Poulsen, editors
2009

*Veterinary Medicines in the Environment*
Crane, Boxall, Barrett
2008

*Relevance of Ambient Water Quality Criteria for Ephemeral and Effluent-dependent Watercourses of the Arid Western United States*
Gensemer, Meyerhof, Ramage, Curley
2008

*Extrapolation Practice for Ecotoxicological Effect Characterization of Chemicals*
Solomon, Brock, de Zwart, Dyev, Posthumm, Richards, editors
2008

*Environmental Life Cycle Costing*
Hunkeler, Lichtenvort, Rebitzer, editors
2008

*Valuation of Ecological Resources: Integration of Ecology and Socioeconomics in Environmental Decision Making*
Stahl, Kapustka, Munns, Bruins, editors
2007

For information about SETAC publications, including SETAC's international journals, Environmental Toxicology and Chemistry and Integrated Environmental Assessment and Management, contact the SETAC office nearest you:

SETAC
1010 North 12th Avenue
Pensacola, FL 32501-3367 USA
T 850 469 1500 F 850 469 9778
E setac@setac.org

SETAC Office
Avenue de la Toison d'Or 67
B-1060 Brussells, Belguim
T 32 2 772 72 81 F 32 2 770 53 86
E setac@setaceu.org

***www.setac.org***
**Environmental Quality Through Science®**

# Application of Uncertainty Analysis to Ecological Risks of Pesticides

Edited by
William J. Warren-Hicks
Andy Hart

Coordinating Editor of SETAC Books
Joseph W. Gorsuch
Copper Development Association, Inc.
New York, NY, USA


CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

Published in collaboration with the Society of Environmental Toxicology and Chemistry (SETAC)
1010 North 12th Avenue, Pensacola, Florida 32501
Telephone: (850) 469-1500 ; Fax: (850) 469-9778; Email: setac@setac.org
Web site: www.setac.org



Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number: 978-1-4398-0734-7 (Hardback)



**Library of Congress Cataloging-in-Publication Data**

Application of uncertainty analysis to ecological risks of pesticides / editors, William J. Warren-Hicks, Andy Hart.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-4398-0734-7 (hardcover : alk. paper)
1. Pesticides--Environmental aspects--Mathematical models. 2. Ecological risk assessment. 3. Probabilities. I. Warren-Hicks, William J. II. Hart, Andy, 1956-

QH545.P4.A67 2010
577.27'9015118--dc22 2009038019

**Visit the Taylor & Francis Web site at**
**http://www.taylorandfrancis.com**

**the CRC Press Web site at**
**http://www.crcpress.com**

**and the SETAC Web site at**
**www.setac.org**

# SETAC Publications

Books published by the Society of Environmental Toxicology and Chemistry (SETAC) provide in-depth reviews and critical appraisals on scientific subjects relevant to understanding the impacts of chemicals and technology on the environment. The books explore topics reviewed and recommended by the Publications Advisory Council and approved by the SETAC North America, Latin America, or Asia/Pacific Board of Directors; the SETAC Europe Council; or the SETAC World Council for their importance, timeliness, and contribution to multidisciplinary approaches to solving environmental problems. The diversity and breadth of subjects covered in the series reflect the wide range of disciplines encompassed by environmental toxicology, environmental chemistry, hazard and risk assessment, and life-cycle assessment. SETAC books attempt to present the reader with authoritative coverage of the literature, as well as paradigms, methodologies, and controversies; research needs; and new developments specific to the featured topics. The books are generally peer reviewed for SETAC by acknowledged experts.

SETAC publications, which include Technical Issue Papers (TIPs), workshop summaries, newsletter (*SETAC Globe*), and journals (*Environmental Toxicology and Chemistry* and *Integrated Environmental Assessment and Management*), are useful to environmental scientists in research, research management, chemical manufacturing and regulation, risk assessment, and education, as well as to students considering or preparing for careers in these areas. The publications provide information for keeping abreast of recent developments in familiar subject areas and for rapid introduction to principles and approaches in new subject areas.

SETAC recognizes and thanks the past coordinating editors of SETAC books:

A.S. Green, International Zinc Association
Durham, North Carolina, USA

C.G. Ingersoll, Columbia Environmental Research Center
US Geological Survey, Columbia, Missouri, USA

T.W. La Point, Institute of Applied Sciences
University of North Texas, Denton, Texas, USA

B.T. Walton, US Environmental Protection Agency
Research Triangle Park, North Carolina, USA

C.H. Ward, Department of Environmental Sciences and Engineering
Rice University, Houston, Texas, USA