

---

# Bayesian Methods in Conservation Biology

PAUL R. WADE

Office of Protected Resources, National Marine Fisheries Service, c/o National Marine Mammal Laboratory, 7600 Sand Point Way, N.E., Seattle, WA 98155, U.S.A., email paul.wade@noaa.gov

---

**Abstract:** *Bayesian statistical inference provides an alternate way to analyze data that is likely to be more appropriate to conservation biology problems than traditional statistical methods. I contrast Bayesian techniques with traditional hypothesis-testing techniques using examples applicable to conservation. I use a trend analysis of two hypothetical populations to illustrate how easy it is to understand Bayesian results, which are given in terms of probability. Bayesian trend analysis indicated that the two populations had very different chances of declining at biologically important rates. For example, the probability that the first population was declining faster than 5% per year was 0.00, compared to a probability of 0.86 for the second population. The Bayesian results appropriately identified which population was of greater conservation concern. The Bayesian results contrast with those obtained with traditional hypothesis testing. Hypothesis testing indicated that the first population, which the Bayesian analysis indicated had no chance of declining at >5% per year, was declining significantly because it was declining at a slow rate and the abundance estimates were precise. Despite the high probability that the second population was experiencing a serious decline, hypothesis testing failed to reject the null hypothesis of no decline because the abundance estimates were imprecise. Finally, I extended the trend analysis to illustrate Bayesian decision theory, which allows for choice between more than two decisions and allows explicit specification of the consequences of various errors. The Bayesian results again differed from the traditional results: the decision analysis led to the conclusion that the first population was declining slowly and the second population was declining rapidly.*

## Métodos Bayesianos en Conservación Biológica

**Resumen:** *Las inferencias de la estadística Bayesiana proveen una forma alternativa de analizar datos que probablemente sea más apropiada para la conservación biológica que los métodos estadísticos tradicionales. Comparé técnicas Bayesianas con técnicas tradicionales de prueba de hipótesis, mediante el uso de ejemplos aplicables a la conservación. Utilicé un análisis de tendencias de dos poblaciones hipotéticas para ilustrar lo fácil que es entender los resultados Bayesianos, los cuales se obtienen en forma de probabilidades. El análisis Bayesiano de tendencias indicó que las dos poblaciones tenían probabilidades muy diferentes de disminuir a tasas biológicamente importantes. Por ejemplo, la probabilidad de que la primera población disminuyera más de un 5% por año fue de 0.0, comparada con una probabilidad de un 0.86 para la segunda población. Los resultados Bayesianos identifican adecuadamente cual población era la de mayor interés para conservar. Los resultados Bayesianos se contrastan con aquéllos obtenidos mediante pruebas de hipótesis tradicionales. Las pruebas de hipótesis indicaron que la primera población (la cual fue indicada por las pruebas Bayesianas como teniendo una probabilidad de disminución mayor al 5% por año) disminuía significativamente debido a que la disminución ocurría a una tasa muy baja y las estimaciones de abundancia eran precisas. A pesar de la alta probabilidad de que la segunda población experimente una disminución seria, la prueba de hipótesis falló al rechazar la hipótesis nula de no-disminución, debido a que las estimaciones de abundancia eran poco precisas. Extendí el análisis de tendencias para ilustrar la teoría de decisión Bayesiana, la cual permite decidir entre más de dos posibilidades y permite una especificación explícita de las consecuencias de varios errores. Los resultados Bayesianos difirieron una vez más de los resultados tradicionales: el análisis de decisiones condujo a la conclusión de que la primera población disminuía despacio y la segunda disminuía rápidamente.*

---

*Paper submitted September 2, 1999; revised manuscript accepted March 15, 2000.*

1308

## Introduction

Conservation scientists gather and analyze data with the goal of improving resource management. Therefore, the analysis of data should lead to results that are easy to understand and useful for making conservation decisions. Biologists often analyze their data with standard statistical procedures that test hypotheses. Such tests may not detect what the data could potentially tell us about a conservation problem and in some cases may even mislead conservation action because results may inadequately treat uncertainty. Bayesian statistical inference provides an alternate way to analyze data that remedies many of the problems inherent in standard hypothesis testing and, most important, allows the incorporation of uncertainty.

Bayesian methods calculate the probability of different values of a parameter given the data. Consequently, Bayesian methods have practical advantages for conservation biologists because (1) Bayesian analyses are simple to explain and present and automatically include the uncertainty of the estimate; (2) probability statements better represent the state of a population than  $p$  values generated from hypothesis tests; (3) Bayesian analyses relate directly to biological relevance in contrast to significance tests, in which biological importance usually plays no role; (4) Bayesian decision theory can be used, which allows consideration of the relative consequences of making incorrect decisions; (5) uncertainty from important but unknown parameters can be included; (6) uncertainty in model choice can be formally incorporated into analysis results by combining the results from different plausible models via the Bayes factor; and (7) uncertainty can be reduced by incorporating additional information in a formal and transparent way, including combining various types of data, updating an analysis after collection of additional data, or subjectively using information from similar populations or species.

I use three examples to illustrate Bayesian methods. The first provides a visual example as an introduction to Bayesian analysis, the second is a hypothetical trend analysis that illustrates the first three general advantages, and the third illustrates a Bayesian decision analysis. Given that good introductions to the use of Bayesian methods in ecology already exist (Reckhow 1990; Ellison 1996), I focus on examples that illustrate specific benefits for conservation issues.

## What Are Bayesian Statistics?

Statistical methods based on Bayes's theorem (Bayes 1763) represent a different school of statistical inference and a different statistical philosophy from the standard statistics that most scientists are taught (Jeffreys 1939, 1961; Berger 1985; Howson & Urbach 1989, 1991; Lee 1989; Press 1989). Bayesian methods calculate the prob-

ability of the value of a parameter given the observed data. In contrast, conventional statistical analyses (called frequentist statistics) calculate the probability of observing data given a specific value for a parameter, such as a null hypothesis. In simplest terms, the data are what is known, the value of the parameter is what is unknown, and Bayesians therefore focus on what the data tell about the parameter (Lindley 1986). Nearly any statistical analysis, such as linear regression, can be carried out as a Bayesian analysis. Standard statistical methods generally use significance tests to make conclusions from data. Bayesian methods rely on probability statements that are made from a distribution that describes the probability of all parameter values, given the data.

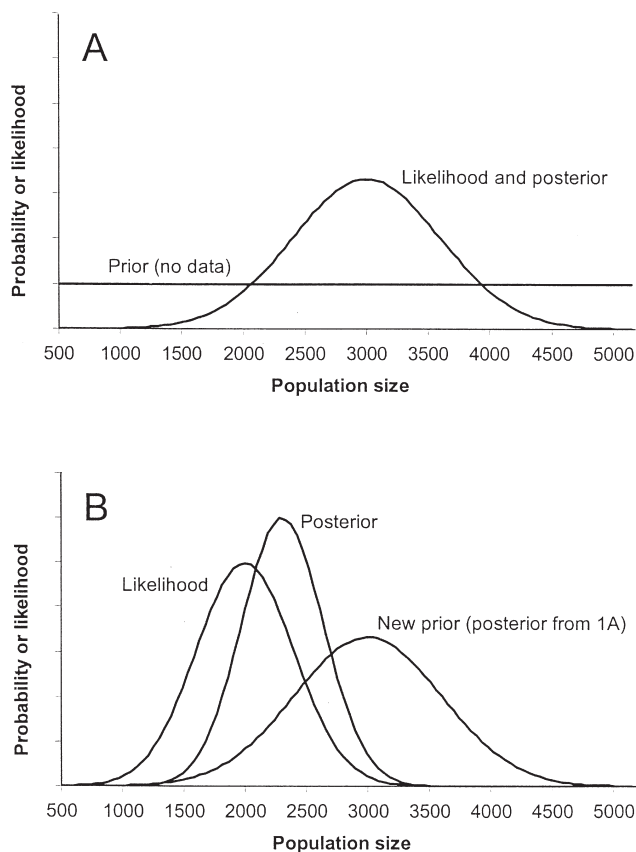
Both Bayesian and frequentist statistics use the common tool of sampling distributions to calculate the probability of observing data given specific values of parameters. Frequentist methods use this sampling distribution directly. For example, a  $p$  value is the probability of observing data as extreme as or more extreme than the data that were observed, given that the null hypothesis is true, on repeated sampling of the data. This explains the origin of the name: the performance of frequentist statistics is measured by their long-run frequencies under repeated sampling of data. In addition to hypothesis testing, a frequentist parameter estimate can also be made from the sampling distribution by calculating the likelihood function, which is formed by calculating the probability of observing the data for every possible value of the parameter. This function is then interpreted in frequentist statistics to represent the relative likelihood of different parameter values. The function, however, does not represent the probability of different parameter values; it represents the probability of observing the data given different parameter values. A maximum likelihood estimate is the value of the parameter that maximizes the probability of the observed data (the peak of the likelihood function).

Bayesian methods also use the likelihood function, but in a different way. Bayesian analyses calculate a posterior probability distribution for the parameter as the integral of the product of the likelihood function with a prior probability distribution for the parameter. The prior distribution represents a probability distribution for the parameter before consideration of the data, and the posterior represents a probability distribution for the parameter after consideration of the data. All statistical inference is then made from the posterior distribution. Further technical details on Bayesian methods are provided in the Appendix.

## An Example of Bayesian Methods

I use a simple example to introduce the likelihood function, the prior distribution, and how they interact to pro-

duce the posterior distribution. A survey is conducted and data collected that result in an abundance estimate of 3000, which is assumed to have a normal sampling distribution. The likelihood function is the probability of observing the data (the abundance estimate) given different values of the parameter (the true population size). In this case, the likelihood function (Fig. 1a) is produced by successively calculating the probability of observing an estimate of 3000 if the true population size is 500, 501, 502, and so on up to 5000. Understandably, the probability of observing an estimate of 3000 would be greatest if the true population size was 3000, and the likelihood function is therefore a normal distribution centered on that value.



**Figure 1.** An (a) Bayesian analysis with a uniform prior distribution and a normal likelihood function. The posterior distribution that results is itself a normal distribution. The likelihood function is on a different scale than the probability distributions; here it has been scaled arbitrarily to the posterior distribution and so is identical to the posterior distribution. A (b) second analysis with new data based on the posterior distribution from Fig. 1a as the prior distribution for the second analysis. The likelihood function for the second data set (arbitrarily scaled to a similar size) is centered at 2000. The new posterior distribution is intermediate to the new prior and the new likelihood.

Because nothing is known about the abundance of the population except for our current data, we choose a noninformative prior distribution in which any positive value of the parameter is equally likely (a uniform distribution). If the prior distribution is uniform and the likelihood function is normal, then the posterior distribution is itself a normal distribution (Fig. 1a). The posterior is the product of the two distributions scaled to be a probability distribution (i.e., the area under the curve equals one). This posterior distribution represents a statement about how probable different values of the parameter are in light of the data and the prior distribution. The posterior distribution automatically describes the uncertainty of the abundance estimate and provides the point estimate, which is the mean of the posterior. The posterior distribution and the likelihood function have exactly the same shape in this example (and are drawn as perfectly overlapping lines in Fig. 1a), but the likelihood function is independent of scale. It provides only a relative measure of each parameter value that cannot be interpreted as a probability.

Now suppose an independent second abundance survey is performed immediately after the first. The Bayesian analysis uses the knowledge gained from the previous survey: the posterior distribution from the previous analysis serves as the prior distribution for the new analysis (Fig. 1b). Because the new prior distribution results from analysis of data, it is called a "data-based prior." A new likelihood distribution is calculated from the new data (shown in Fig. 1b, arbitrarily scaled to a similar size for presentation purposes). If the likelihood function and the prior distribution are both normal, then the posterior distribution is also a normal distribution that can be solved analytically (Fig. 1b; Iversen 1984). The new posterior distribution is intermediate to the prior and the likelihood and becomes zero where either the prior or likelihood becomes zero. Inference from the posterior distribution will be explained in the next example.

## A Bayesian Regression Example

To highlight the different interpretations resulting from Bayesian and frequentist approaches, I analyzed trends in abundance for two artificial data sets. If population growth is exponential, the trend of a population can be estimated by a linear regression on the natural log of a series of abundance estimates. A frequentist linear regression estimates the slope and intercept, and common practice usually focuses on a significance test of the slope. The null hypothesis of a zero slope (no trend) is rejected when  $p$  values are below a specified critical level (usually 0.05 or 0.01).

I created 10 years of abundance data for two hypothetical populations. In population 1 (Fig. 2a), a frequentist test of whether the slope was different than 0.0 was sig-

nificant at the 0.05 level ( $p = 0.048$ ), whereas the test for population 2 (Fig. 2b) was not significant ( $p = 0.053$ ). Therefore, if a statistically significant decline must occur before conservation action is taken, then action would be taken for population 1 but not for population 2.

But which hypothetical population likely poses a conservation problem? Setting aside the results of the significance test, most people would view population 2 as a potential conservation problem, whereas population 1 would appear fairly stable. Population 2 is estimated to be declining at 10% per year (slope( $s$ ) =  $-0.10$ ) but with much uncertainty, whereas population 1 is fairly precisely estimated to be declining at about  $\frac{1}{3}$  of 1% per year ( $s = -0.0036$ ). In other words, population 2 is estimated to be declining at a rate about 30 times greater than population 1. The frequentist tests did not help identify the population at greater risk; in fact, the results were the opposite of the conclusion that most people would reach without using statistics.

Why did these results occur? There is no mystery: the data from population 1 are significant because they are

more precise, which provides more statistical power to detect a trend. The data from population 2 are much less precise and provide little statistical power to detect a trend. In practice, most researchers set Type I error, the probability of rejecting a true null hypothesis, a priori to a pre-specified level (here, the critical value of 0.05) without explicitly considering Type II error. Several authors have pointed out the importance of calculating Type II error levels or statistical power, which are rarely calculated in practice (Peterman 1990; Taylor & Gerrodette 1993; Steidl et al. 1997). Even though retrospective power analyses can demonstrate a lack of power and help interpretation, they cannot be used to change the results (Hayes & Steidl 1997). In particular, a calculation of statistical power does little to help interpretation of the conservation status of population 1 because a finding of a significant decline will not be changed by a calculation of power.

In a Bayesian analysis of these same hypothetical data, all inference is drawn from the posterior distribution. Using noninformative (uniform) distributions for both the slope and the intercept, the posterior distribution for the slope is a  $t$  distribution with degrees of freedom equal to the sample size minus two (Press 1989; Bernardo & Smith 1994). The confidence limits for the slope in the frequentist analysis use this same  $t$  distribution.

The posterior distributions for the slope differ dramatically for the two populations (Fig. 3). For visual convenience, I present the two distributions as equal in height, which means the scales differ. If they were shown on the same scale, the area under each curve would equal one and the posterior distribution for population 1 would be much taller. The visual impression given by examining the two posterior distributions matches our previous intuition: population 2 is likely to be at greater risk than population 1. The posterior distribution can immediately

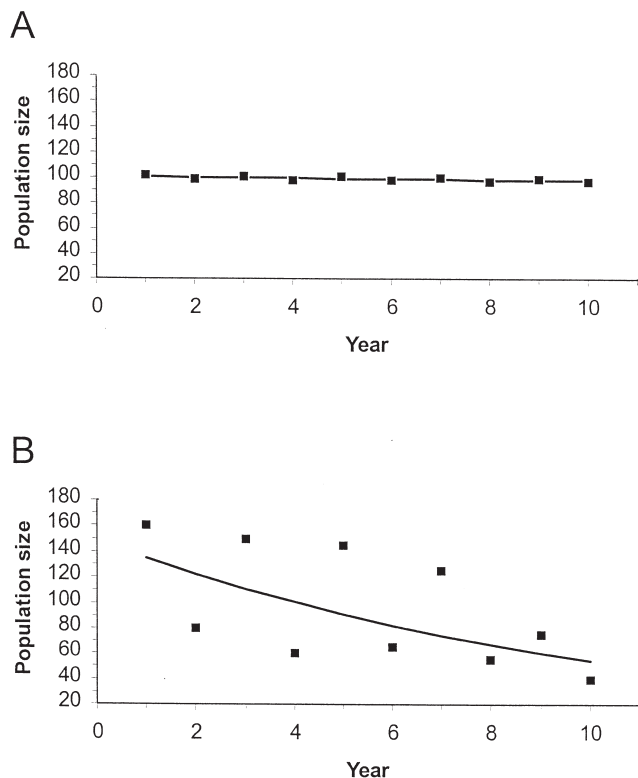


Figure 2. Hypothetical data for (a) population 1 and the estimated linear regression on log abundance using traditional frequentist methods (slope =  $-0.03$ ,  $p = 0.048$ ), and hypothetical data for (b) population 2 and the estimated linear regression on log abundance using traditional frequentist methods (slope =  $-0.10$ ,  $p = 0.053$ ).

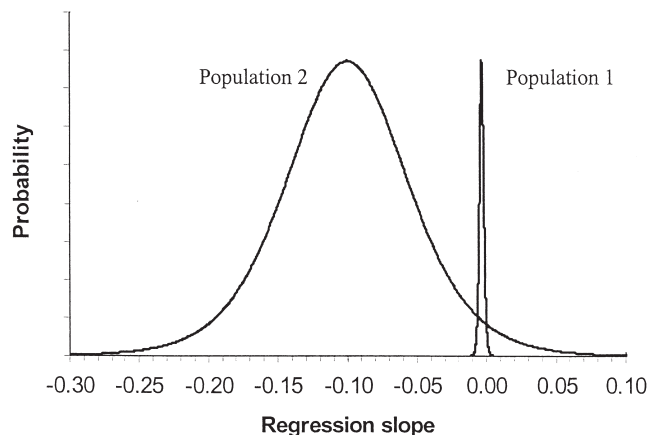


Figure 3. Posterior distributions for population 1 and population 2 from a Bayesian linear regression on log abundance. The two posterior distributions are scaled differently to be of similar height.

be “queried” for biologically important questions. For example, assuming that declines of more than 5% represent an undesired risk level, one can easily ask “What is the probability that the population is declining faster than 5%?” For population 1, the probability is zero, whereas for population 2 it is 0.86, or fairly high (Table 1).

In Bayesian inference, hypotheses can be compared by means of posterior odds, or the Bayes Factor (Reckhow 1990; Kass & Raftery 1995; Ellison 1996). The Bayes factor is the ratio of the posterior odds to the prior odds and thus represents a measure of whether the data have increased or decreased the likelihood of one hypothesis relative to another (Kass & Raftery 1995). Interpretation of the Bayes factor is placed into four broad categories, from weak (“not worth a bare mention”) to positive, strong, or very strong evidence for one hypothesis over another (Jeffreys 1961; Kass & Raftery 1995). Where two hypotheses have equal prior probability, the Bayes factor is equal to the posterior odds and is a function of only the data. To compare the simple hypothesis that  $s = 0.0$  to the simple hypothesis that  $s = -0.05$ , assuming equal prior probability, the Bayes factor is essentially the ratio of the height of the posterior distribution at the two locations (Fig. 3). The Bayes factor for  $s = 0.0$  versus  $s = -0.05$  for population 1 approaches infinity, indicating that the data provide strong evidence that the population is stable versus declining at 5% per year. The Bayes factor for  $s = -0.05$  versus  $s = 0.00$  would be approximately 5 for population 2, meaning that the data provide positive evidence for a 5% decline relative to a stable population. With this as the posterior odds ratio (because of the assumption of equal prior odds), the hypothesis of a decline of 5% is five times more probable than the hypothesis of a stable population.

Another useful approach compares composite hypotheses. For a composite hypothesis, the Bayes factor is still the ratio of the posterior odds to the prior odds, but these probabilities are now calculated by integrating across the posterior and prior distributions, rather than just evaluating them at a single value. For the sake of these comparisons, I again assume equal prior odds for the two hypotheses. Comparing the hypothesis that  $s < -0.05$  to the hypothesis that  $s > -0.05$  leads to the conclusion that population 1 shows evidence that it is declining slowly versus declining rapidly. On the other hand, population 2 has positive evidence for a biologi-

cally important decline. Alternatively, we could compare the hypothesis that the population is declining ( $s < 0.0$ ) to the hypothesis it is increasing ( $s > 0.0$ ), which results in strong evidence that both populations are declining. In summary, the Bayes factor leads to the conclusions that there is (1) strong evidence that population 1 is declining, but not at a biologically important rate, and (2) strong evidence that population 2 is declining and positive evidence that the decline is biologically important.

### Using the Bayesian Regression Results in a Decision Analysis

Drawing conclusions about the conservation status of populations is an inherent part of conservation biology. Conventional frequentist trend analysis concludes that a population is declining only when the data results in a slope that is significantly different from 0.0. Thus decisions depend on obtaining statistically significant results. Managers are often unaware that accepting conventional standards of statistical proof may result in potentially unacceptable under-protection errors.

Bayesian decision theory provides an alternative framework for making decisions (Berger 1985). Often the finished product from the scientist will be the posterior distribution. The manager, however, can use that distribution in further analyses to make decisions through the use of a “loss function.” A manager creates a loss function by specifying values that represent the relative undesirability of various wrong decisions—such as under-protecting or overprotecting a population—according to agency policy or his or her personal beliefs. Once these loss functions are specified, the “Bayes expected loss” is calculated for each possible decision as the integral of the product of the posterior distribution for the slope and the loss function for that decision. The manager can then choose the decision with the lowest loss, because this decision minimizes the Bayes expected loss (Berger 1985; Lee 1989).

Decision theory allows for multiple decision possibilities, rather than just the two states (significant or not) allowed by frequentist hypothesis tests. Returning to our example of the trend of the two hypothetical populations, a policymaker could consider three possible decisions: (1) to conclude that a population is declining rapidly, and thus managers would presumably initiate direct conservation action to slow the rate of decline; (2) to conclude that a population is declining slowly or that its trend is uncertain, and thus managers would presumably wish to collect additional data to confirm the trend and study potential ways of directly benefiting the population; or (3) to conclude that a population is not declining and is possibly increasing, and thus managers would simply maintain baseline monitoring at a level appropri-

**Table 1.** Posterior probabilities for three specific ranges of the slope parameter ( $s$ ) calculated from the areas under the curves in Fig. 3.

Posterior probabilities	$s < -0.05$	$-0.05 \leq s \leq 0.0$	$s > 0.0$
Population 1	0.00	0.976	0.024
Population 2	0.855	0.118	0.027

ate for a population considered of no conservation concern. As an example, I specify a simple loss function for each of these three possible decisions (Table 2).

The values used in loss functions result from policy decisions that quantify the relative gravity of various errors. I provide an example of the logic that could be used to create Table 2. There are three different possible states of nature: (1)  $s \leq -0.05$ , (2)  $-0.05 \leq s \leq 0.00$ , and (3)  $s \geq 0.0$ . These correspond to the three decisions stated above that represent the correct decision (from the policymaker's point of view) for each of those three states of nature. If the true state matches the decision, then a value of 0.0 is given to signify no loss under that decision (diagonal values, Table 2). For decision 1, to conclude that there is a rapid decline, a moderate penalty (loss of 0.5) is given when a population is actually declining at  $<5\%$ , and a strong penalty (loss of 1.0) is given if a population is actually increasing. These loss values represent the belief that if one concludes that a population is declining rapidly, then it would be twice as bad to do so when the population is increasing than when it is declining slowly. It should be recognized that the particular loss function values of 0.0, 0.5, and 1.0 have meaning only relative to one another, so it would be equivalent to use values of 0.0, 1.0, and 2.0, for example. For decision 2, to conclude that there is a slow decline, a moderate penalty is given if the population is actually declining rapidly or if the population is increasing. Finally, for decision 3, to conclude that there is no decline, a strong penalty is given if the population is truly declining at  $>5\%$ , and a moderate penalty is given if the population is truly declining at a rate between 5% and 0%. The penalties in Table 2 are symmetrical, which means that equal loss is assigned to over- and underprotecting a population. A policy maker could choose a precautionary approach by making the values above the diagonal greater than those below.

The policymaker chooses the decision that minimizes the Bayes expected loss and therefore concludes that population 1 is declining slowly and that population 2 is declining rapidly (footnote c, Table 2). Therefore, a manager would initiate direct action to attempt to slow or stop the decline of population 2 and would call only

for the collection of additional data on population 1. This Bayesian analysis leads to conclusions opposite of those based on the frequentist hypothesis tests, in which population 1 has a significant decline and population 2 does not.

In the interest of making the example clear, simple step functions have been used here for the loss functions. Nevertheless, these loss functions can be continuous functions of the parameter. For example, the relative loss could continue to increase with greater rates of decline. Taylor et al. (1996) provide an example of a continuously changing loss function for extinction risk.

## Discussion

Conservation research will be more effective if results can be communicated clearly to managers, stakeholders, and policymakers. A Bayesian posterior distribution is easy to understand and communicate and contrasts with the often convoluted interpretation of frequentist statistics (Lindley 1986; Berger & Berry 1988). In particular, a probability distribution communicates uncertainty in a parameter estimate in a visual manner, which is simpler to interpret than a confidence interval. In the trend analysis, I demonstrated that a probability distribution was more useful for identifying a population of conservation concern than a frequentist significance test. In certain circumstances, a population might go extinct before a significant decline could be detected (Taylor & Gerrodette 1993). Additional critiques of the frequentist hypothesis testing framework, from a non-Bayesian point of view, have been offered by Edwards (1992) and Royall (1997).

Bayesian methods allow decision theory to be applied to conservation problems. Policy makers can then explicitly state that some errors are worse than others. Decision theory also allows a richer set of responses. In the example, two different responses were possible for a declining population—to take immediate conservation action or to collect additional data—depending on the magnitude and the certainty of the result. A set of possible actions allow a more appropriate response than just

**Table 2.** Loss functions and expected loss for three possible decisions in a Bayesian decision analysis applied to trend data.

Decisions	Loss functions <sup>a</sup>			Expected loss for population 1 <sup>b</sup>	Expected loss for population 2 <sup>b</sup>
	$s < -0.05$	$-0.05 \leq s \leq 0.0$	$0.0 < s$		
(1) Conclude population is declining rapidly	0.0	0.5	1.0	0.512	0.086 <sup>c</sup>
(2) Conclude population is declining slowly	0.5	0.0	0.5	0.012 <sup>c</sup>	0.441
(3) Conclude population is not declining	1.0	0.5	0.0	0.488	0.914

<sup>a</sup>The slope parameter is  $s$ .

<sup>b</sup>The expected loss is the sum across the three slope categories of the product of the loss function and the appropriate posterior probability from Table 1. For example, the expected loss for decision 1 for population 1 of 0.512 was calculated as  $(0.0 \times 0.000) + (0.5 \times 0.976) + (1.0 \times 0.024)$ .

<sup>c</sup>The decision with the smallest expected loss is chosen, which in this case is to conclude that population 1 is declining slowly and that population 2 is declining rapidly.

being able to conclude whether a population is declining or not.

Bayesian methods also allow uncertainty to be incorporated directly into analyses. No standard frequentist method exists for incorporating uncertainty for parameters for which no data exist. Often such parameters are fixed at single values, ignoring the high degree of uncertainty when no species-specific data are available. In contrast, Bayesian methods allow a range of plausible values to be incorporated by specifying a prior distribution for an unknown parameter. Examples of this approach include specifying a prior distribution for the value of an environmental variance parameter (Taylor et al. 1996) and specifying a prior distribution for a density-dependence parameter (Wade 1994, 1999, 2001). A similar ad hoc frequentist approach has been proposed (Restrepo et al. 1992), but it has been shown that this approach generally performs no better, and sometimes much worse, than Bayesian methods (Poole et al. 1999).

Another advantage not illustrated here is the use of Bayesian methods for comparing models and incorporating model uncertainty (Kass & Raftery 1995). The Bayes factor, used here for comparing hypotheses, can also be used to compare how well different models fit the data in a Bayesian framework. There are two main advantages of Bayesian methods over frequentist methods for model comparison (Kass & Raftery 1995). First, the models that are compared do not have to be nested as they need to be in many frequentist analyses, such as step-wise multiple regression. Second, rather than just choosing the single best model when several models have some probability of fitting the data reasonably well, their results can be proportionally mixed to form a posterior distribution for the parameter of interest which takes into account model uncertainty. As an example, Wade (2001) used the Bayes factor to compare how well different models (e.g., simple vs. age-structured) fit data on gray whale abundance.

Given these advantages, one may wonder why Bayesian methods have not been taught or used more often. Although Bayesian theory has older roots (Bayes 1763; Laplace 1774), frequentist methods have dominated in applied statistics until recently. This is partly because Fisher, Neyman, Pearson, and others established practical methods for using frequentist statistics well before Jeffreys (1939) laid the foundation for applied Bayesian statistics. Most scientists receive training only in frequentist methods, so it is natural for frequentist methods to dominate common practice.

Computational difficulties are another reason for the lack of a widespread use of Bayesian methods. Some scientists interested in using Bayesian methods have not used them because of technical difficulties. A major problem has been the computational difficulties in solving integrations for anything but the most simple problems. This problem is quickly disappearing because

many new solutions are now available through numerical methods for integration, which are possible because of increased computer speeds. Examples include the sampling-importance-resampling method (Rubin 1988; Smith & Gelfand 1992) and the Markov chain Monte Carlo (Geyer 1992). Tanner (1993) and Gelman et al. (1995) provide a comprehensive look at various numerical integration methods.

In addition, some scientists have been exposed to Bayesian methods but have chosen, for a variety of reasons, to use other methods. Some have raised objections to the potential for introducing subjective opinion into prior distributions (Efron 1986; Dennis 1996). It is true that some Bayesian practitioners treat probability as a subjective quantity and include expert opinion in analyses (e.g., Wolfson et al. 1996). This is not the only way to use Bayesian methods, however, and there is nothing inherently subjective or unscientific about using a prior distribution (Cox & Hinkley 1974). Efron (1986) recognized that there are both subjective Bayesian methods and what he called "objective Bayesian" methods. In 1812, Laplace proposed the "principle of insufficient reason," in which all values of the unknown parameter are taken to be equally likely, a priori, unless there is a reason to the contrary (Press 1989). Following Laplace, Jeffreys (1939, 1961) discussed objective Bayesian inference and the use of "non-informative" or "vague" priors, which in simple cases can be just uniform distributions (e.g., Fig. 1).

Press (1989) recommended that non-informative priors be used when public policy may be influenced by the outcome of an analysis, and conservation biology issues likely fall into this category. Such priors can be objective in the sense that any person applying the same criteria will form the same prior distribution. Most Bayesian statistics textbooks contain large sections on the use of non-informative prior distributions (e.g., Jeffreys 1961; Iversen 1984; Berger 1985; Lee 1989; Press 1989; Gelman et al. 1995).

Although some Bayesian advocates have pointed out the potential for subjectivity in frequentist methods (Berger & Berry 1988; Press 1989), this does not negate the fact that there sometimes are genuine difficulties involved in specifying prior distributions that do not influence the results. Problems can arise because non-informative distributions (such as a uniform) may no longer be so if a model is re-parameterized and transformed to a new parameter space. It can also be difficult to specify non-informative prior distributions for several parameters at once if those parameters are interrelated through the model; for example, in combination, uniform distributions for model parameters may result in a nonuniform distribution for an output quantity of interest that is a function of the model parameters.

Practical solutions often exist to the problem of specifying non-informative prior distributions. Savage (1962)

has pointed out that the prior distribution needs to be non-informative only over the range at which the likelihood function is non-negligible; outside this range the prior is irrelevant because it will not influence the results. Where so few data are available that the choice of a non-informative prior influences the results, Gelman et al. (1995) recommend putting relevant information into the prior distribution by means of a hierarchical model. Ideally, one hopes to avoid these complications by using a data-based prior (Press 1989), which is a prior based on an analysis of previously available data (not the data to be used in the likelihood function in the current analysis).

The attractions of Bayesian methods can be seen in their increasing use in many fields of applied science, including ecology. In fisheries biology, an increasing number of applied assessments have used Bayesian methods in recent years (e.g., Hilborn & Walters 1992; Thompson 1992; McAllister et al. 1994; Walters & Ludwig 1994; McAllister & Ianelli 1997), and they have become common enough to justify a review (Punt & Hilborn 1997). Similarly, applied Bayesian methods are increasingly being used in the assessment of whale populations (e.g., Givens et al. 1993, 1995; Raftery et al. 1995; Punt & Butterworth 1997, 1999; Wade 2001). Other examples of Bayesian methods in ecology include analyses by Ellison (1996); Gazey and Staley (1986); Reckhow (1990); Shaughnessy et al. (1995); Pascual and Kareiva (1996); and Omlin and Reichert (1999). Within conservation biology, a few examples exist of Bayesian population viability analyses (Ludwig 1996; Taylor et al. 1996). It is likely that the use of Bayesian methods will continue to increase in applied science, including disciplines such as conservation biology.

## Acknowledgments

The Bayesian regression analysis was calculated with the program First Bayes V1.3, excellent teaching software available from A. O'Hagen, Department of Probability and Statistics, University of Sheffield, Hicks Building, Sheffield, SR 73H, United Kingdom (<http://www.shef.ac.uk/~st1ao.index.html>). The manuscript benefitted greatly from helpful reviews by A. M. Ellison, C. Fowler, J. Laake, K. Ralls, B. Taylor, and an anonymous reviewer.

## Literature Cited

- Bayes, T. 1763. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions, Royal Society of London* **53**: 370–418.
- Berger, J. O. 1985. *Statistical decision theory and Bayesian analysis*. Springer-Verlag, New York.
- Berger, J. O., and D. A. Berry. 1988. Statistical analysis and the illusion of objectivity. *American Scientist* **76**:159–65.
- Bernardo, J. M., and A. F. M. Smith. 1994. *Bayesian theory*. Wiley, Chichester, United Kingdom.
- Cox, D. R., and D. V. Hinkley. 1974. *Theoretical statistics*. Chapman and Hall, London.
- Dennis, B. 1996. Discussion: should ecologists become Bayesians? *Ecological Applications* **6**:1095–1103.
- Edwards, A. W. F. 1992. *Likelihood*. The Johns Hopkins University Press, Baltimore, Maryland.
- Efron, B. 1986. Why isn't everyone a Bayesian? *The American Statistician* **40**(1):1–5.
- Ellison, A. M. 1996. An introduction to Bayesian inference for ecological research and environmental decision making. *Ecological Applications* **6**:1036–1046.
- Gazey, W. J., and M. J. Staley. 1986. Population estimation from mark-recapture experiments using a sequential Bayes algorithm. *Ecology* **67**:941–951.
- Gelman, A. J. Carlin, H. Stern, and D. Rubin, 1995. *Bayesian data analysis*. Chapman and Hall, London.
- Geyer, C. J. 1992. Practical Markov chain Monte Carlo. *Statistical Science* **7**:473–511.
- Givens, G. H., A. E. Raftery, and J. E. Zeh. 1993. Benefits of a Bayesian approach for synthesizing multiple sources of evidence and uncertainty linked by a deterministic model. Report of the International Whaling Commission **43**:495–503.
- Givens, G. H., J. E. Zeh, and A. E. Raftery. 1995. Assessment of the Bering-Chukchi-Beaufort Seas stock of bowhead whales using the BALEEN II model in a Bayesian synthesis framework. Report of the International Whaling Commission **45**:345–364.
- Hayes, J. P., and R. J. Steidl. 1997. Statistical power analysis and amphibian population trends. *Conservation Biology* **11**:273–275.
- Hilborn, R., and C. J. Walters. 1992. *Quantitative fisheries stock assessment: choice, dynamics and uncertainty*. Chapman and Hall, New York.
- Howson, C., and P. M. Urbach. 1989. *Scientific reasoning: the Bayesian approach*. Open Court, Chicago.
- Howson, C., and P. M. Urbach. 1991. Bayesian reasoning in science. *Nature* **350**:371–374.
- Iversen, G. R. 1984. *Bayesian statistical inference*. Sage university paper series on quantitative applications in the social sciences. Sage Publications, Beverly Hills, California.
- Jeffreys, H. 1939. *Theory of probability*. 1st edition. Clarendon Press, Oxford, United Kingdom.
- Jeffreys, H. 1961. *Theory of probability*. 3rd edition. Clarendon Press, Oxford, United Kingdom.
- Kass, R. E., and A. E. Raftery. 1995. Bayes factors. *Journal of the American Statistics Association* **90**:773–795.
- Laplace, P. S. 1774. Memoir on the probability of the cause of events. *Memoires de l'Académie des Sciences de Paris* **6**:621–656 (reprinted in translation in *Statistical Science*, 1986, **1**:359–378).
- Lee, P. M. 1989. *Bayesian statistics: an introduction*. Oxford University Press, New York.
- Lindley, D. V. 1986. Comment. *The American Statistician* **40**(1):6–7.
- Ludwig, D. 1996. Uncertainty and the assessment of extinction probabilities. *Ecological Applications* **6**:1067–1076.
- McAllister, M. K., and J. N. Ianelli 1997. Bayesian stock assessment using catch-age data and the sampling-importance-resampling algorithm. *Canadian Journal of Fisheries and Aquatic Sciences* **54**:284–300.
- McAllister, M. K., E. K. Pikitch, A. E. Punt, and R. Hilborn. 1994. A Bayesian approach to stock assessment and harvest decisions using the sampling/importance resampling algorithm. *Canadian Journal of Fisheries and Aquatic Sciences* **51**:2673–2687.
- Omlin, M., and P. Reichert. 1999. A comparison of techniques for the estimation of model prediction uncertainty. *Ecological Modeling* **115**:45–59.
- Pascual, M. A., and P. Kareiva 1996. Predicting the outcome of competition using experimental data: maximum likelihood and Bayesian approaches. *Ecology* **77**:337–349.
- Peterman, R. M. 1990. Statistical power analysis can improve fisheries research and management. *Canadian Journals of Fisheries and Aquatic Sciences* **47**:2–15.



- Poole, D., G. H. Givens, and A. E. Raftery. 1999. A proposed stock assessment method and its application to bowhead whales, *Balaena mysticetus*. *Fisheries Bulletin* 97:144-152.
- Press, S. J. 1989. *Bayesian statistics: principles, models, and applications*. Wiley, New York.
- Punt, A. E., and D. S. Butterworth. 1997. Assessments of the Bering-Chukchi-Beaufort Seas stock of bowhead whales (*Balaena mysticetus*) using maximum likelihood and Bayesian methods. Report of the International Whaling Commission 47:603-618.
- Punt, A. E., and D. S. Butterworth. 1999. An assessment of Bering-Chukchi-Beaufort Seas stock of bowhead whales (*Balaena mysticetus*) using a Bayesian approach. *Journal of Cetacean Research and Management* 1:53-71.
- Punt, A., and R. Hilborn. 1997. Fisheries stock assessment and decision analysis: the Bayesian approach. *Reviews in Fish Biology and Fisheries* 7:35-63.
- Raftery, A. E., G. H. Givens, and J. E. Zeh. 1995. Inference from a deterministic population dynamics model for bowhead whales. *Journal of the American Statistical Association* 90:402-416.
- Reckhow, K. H. 1990. Bayesian inference in non-replicated ecological studies. *Ecology* 71:2053-2059.
- Restrepo, V. R., J. M. Hoenig, J. E. Powers, J. W. Baird, and S. C. Turner. 1992. A simple simulation approach to risk and cost analysis, with application to swordfish and cod fisheries. *Fisheries Bulletin, United States* 90:736-48.
- Royall, R. 1997. *Statistical evidence: a likelihood paradigm*. Chapman and Hall, London.
- Rubin, D. B. 1988. Using the SIR algorithm to simulate posterior distributions. Pages 395-402 in J. M. Bernardo et al., editors. *Bayesian statistics 3*. Clarendon Press, Oxford, United Kingdom.
- Savage, L. J. 1962. *The foundations of statistical inference*. Wiley, New York.
- Shaughnessy, P. D., J. W. Testa, and R. M. Warneke. 1995. Abundance of Australian fur seal pups, *Arctocephalus pusillus doiferus*, at Seal Rocks, Victoria, in 1991-92 from Peterson and Bayesian estimators. *Wildlife Research* 22:625-632.
- Smith, A. F. M., and A. E. Gelfand. 1992. Bayesian statistics without tears: a sampling-resampling perspective. *The American Statistician* 46(2):84-88.
- Steidl, R. J., J. P. Hayes, and E. Schaubert. 1997. Statistical power analysis in wildlife research. *Journal of Wildlife Management* 61:270-279.
- Tanner, M. A. 1993. *Tools for statistical inference: methods for the exploration of posterior distributions and likelihood functions*. Springer-Verlag, New York.
- Taylor, B. L., and T. Gerrodette. 1993. The uses of statistical power in conservation biology: the vaquita and Northern Spotted Owl. *Conservation Biology* 7:489-500.
- Taylor, B., P. R. Wade, R. Stehn, and J. Cochrane. 1996. A Bayesian approach to classification criteria for Spectacled Eiders. *Ecological Applications* 6:1077-1089.
- Thompson, G. G. 1992. A Bayesian approach to management advice when stock-recruitment parameters are uncertain. *Fisheries Bulletin* 90:561-573.
- Wade, P. R. 1994. Abundance and population dynamics of two eastern Pacific dolphins, *Stenella attenuata* and *Stenella longirostris orientalis*. Ph.D. dissertation. Scripps Institution of Oceanography, University of California, San Diego.
- Wade, P. R. 1999. A comparison of statistical methods for fitting population models to data. Pages 249-270 in G. W. Garner, S. C. Amstrup, J. L. Laake, B. F. J. Manly, L. L. McDonald, and D. G. Robertson, editors. *Marine mammal survey and assessment methods*. Balkema, Rotterdam.
- Wade, P. R. 2001. A Bayesian stock assessment of the eastern Pacific gray whale using abundance and harvest data from 1967 to 1996. *Journal of Cetacean Research and Management Special Volume: in press*.
- Walters, C., and D. Ludwig. 1994. Calculation of Bayes posterior probability distributions for key population parameters. *Canadian Journal of Fisheries and Aquatic Sciences* 51:713-722.
- Wolfson, L. J., J. B. Kadane, and M. J. Small. 1996. Bayesian environmental policy decisions: two case studies. *Ecological Applications* 6:1056-1066.

## Appendix

### Bayesian Statistics

The goal of Bayesian analysis is to calculate the probability of a specific value of a parameter ( $\theta$ ) given the data ( $x$ ), written as  $p(\theta|x)$ . Bayes's theorem is a conditional probability statement that proves that  $p(\theta|x)$  is proportional to the sampling distribution for the data,  $p(x|\theta)$ , multiplied by an independent probability distribution for the parameter,  $p(\theta)$  (independent, in this case, of the specific data  $x$ ).

In this relationship, Bayesians have named  $p(\theta)$  the prior probability for  $\theta$  and  $p(\theta|x)$  the *posterior* probability for  $\theta$  (in the sense that it summarizes what is known about  $\theta$  prior and posterior to the examination of the data  $x$ ). In this context, the sampling distribution for the data,  $p(x|\theta)$ , is often referred to as the likelihood function. In more formal terms,

$$p(\theta|x) = p(x|\theta) \times p(\theta) \times c, \quad (1)$$

where  $p(\theta|x)$  is the posterior probability for the parameter  $\theta$  given the data  $x$ ;  $p(x|\theta)$  is the likelihood function, which provides the probability of the data  $x$  given  $\theta$  (sometimes written as  $L(\theta|x)$ );  $p(\theta)$  is the prior probability for  $\theta$ ; and  $c$  is the normalizing constant, which is the inverse of the integral of  $p(x|\theta) \times p(\theta)$  over the parameter space of  $\theta$ . To calculate  $c$ , one needs to calculate the entire distribution of  $p(x|\theta) \times p(\theta)$ , so one automatically calculates the entire distribution for  $p(\theta|x)$  as well. Therefore, one usually speaks in terms of the posterior and prior distributions. All statistical inference is then based on the posterior distribution. The mean of the posterior distribution can serve as a point estimate for the parameter. Uncertainty in the point estimate is expressed directly in the posterior distribution and can be summarized either as percentiles of the posterior distribution or as what is termed the highest posterior density interval.

To calculate the posterior distribution, one has to integrate the product of the prior distribution and the likelihood function. In some simple cases, the integral can be calculated directly (in these cases it is said to have an analytical or "closed-form" solution). If no analytical solution is available, the integration can be done by numerical methods.