Ecological Applications, 6(4), 1996, pp. 1095–1103

DISCUSSION: SHOULD ECOLOGISTS BECOME BAYESIANS?[1,2]

BRIAN DENNIS
Department of Fish and Wildlife Resources and Division of Statistics, University of Idaho, Moscow, Idaho 83844-1136 USA

Abstract. Bayesian statistics involve substantial changes in the methods and philosophy of science. Before adopting Bayesian approaches, ecologists should consider carefully whether or not scientific understanding will be enhanced. Frequentist statistical methods, while imperfect, have made an unquestioned contribution to scientific progress and are a workhorse of day-to-day research. Bayesian statistics, by contrast, have a largely untested track record. The papers in this special section on Bayesian statistics exemplify the difficulties inherent in making convincing scientific arguments with Bayesian reasoning.

Key words: *Bayesian inference*; *ecological statistics*; *frequentist statistics*; *likelihood principle*; *science philosophy*; *scientific method*; *statistical ecology*; *subjective probability.*

INTRODUCTION

Ecologists should be aware that Bayesian methods constitute a radically different way of doing science. Bayesian statistics is not just another new tool to be added into ecologists' repetoire of statistical methods. Instead, Bayesians categorically reject various tenets of statistics and the scientific method that are currently widely accepted in ecology and other sciences. The Bayesian approach has split the statistics world into warring factions (ecologists' "density independence" vs. "density dependence" debates of the 1950s pale by comparison), and it is fair to say that the Bayesian approach is growing rapidly in influence.

The full implications of adopting Bayesian statistics in ecology are not readily apparent from the papers under discussion here. Therefore, before commenting on the individual papers I will attempt a brief survey of some of the major points of contention between Bayesian and "frequentist" (the kind of statistics most of us learned in college) statistics.

The controversy over Bayesian statistics is not over Bayes' theorem. Bayes' theorem remains true for frequentists and Bayesians alike. Let X and Y be random variables, with probability density functions (pdfs) $p_X(x)$ and $p_Y(y)$, respectively. Bayes' theorem, when expressed for random variables, states that

$$p_Y(y \mid x) = p_X(x \mid y)p_Y(y)/p_X(x), \qquad (1)$$

where $p_Y(y \mid x)$ is the conditional pdf for Y given $X = x$, and $p_X(x \mid y)$ is the conditional pdf for X given $Y = y$. The "theorem" just restates the definition of conditional probability of the event B given the event A: $P(B \mid A) = P(A, B)/P(A) = P(A \mid B)P(B)/P(A)$.

Rather, the controversy revolves around the use of the theorem. To the Reverend Bayes, and his contemporary followers, $p_Y(y)$ in the theorem represents the investigator's subjective prior beliefs about whether a fixed parameter takes the value y (see Stigler 1986, Trader 1989). The parameter is not a random variable; rather, it is an *unknown* variable, and the "prior pdf" denoted by $p_Y(y)$ measures the relative strengths of belief about possible values of the parameter (Lindley 1982, 1990). The investigator selects the form of the prior pdf that quantifies his or her best guesses about the parameter (the parameter can be a vector, and $p_Y(y)$ is then a joint pdf). In the theorem, x is the observed outcome of the experiment (vector of data), and $p_X(x \mid y)$ is the likelihood function familiar in frequentist statistics (pdf of the data, x, given the value of the parameter, y). The conclusions about the parameter after the experiment are summarized in the posterior pdf of y given x: $p_Y(y \mid x)$. The posterior pdf expresses how the investigator's subjective beliefs have been altered by the advent of the data, x.

In the Bayesian approach, all conclusions about the value of the parameter are embodied in the posterior pdf. The data enter the conclusions only through the likelihood function, $p_X(x \mid y)$. In particular, no sample-space probabilities, other than the actual realized value of the likelihood function, are admitted into the conclusions (Lindley 1982, 1990). Sample-space probabilities, such as the probability that a test statistic might exceed a critical value, involve "data that didn't happen," and are excluded from the analysis. The principle of including only the actual data in the analysis and excluding consideration of all other sample-space possibilities is known as the "likelihood principle."

Thus, there are two key elements in the Bayesian approach to statistics. First is the quantification of prior beliefs about a parameter in the form of a probability distribution and the incorporation of those beliefs into the data analysis. Second is the acceptance of the likelihood principle and the concomitant rejection of all sample-space probabilities from inferential conclusions about the parameter. The battle lines between Bayesians and frequentists are drawn around these two elements.

[1] Manuscript received 29 December 1995; revised and accepted 27 March 1996.

[2] For reprints of this group of papers on Bayesian inference, see footnote 1, page 1034.

Required Changes under Bayesianism

Long-standing statistical and scientific practices would change under the Bayesian framework.

P values.—*P* values are an anathema to Bayesians. To Bayesians, *P* values embody all that is logically incorrect about frequentist approaches. A *P* value is the probability, provided the null hypothesis is true, that the test statistic would have been more extreme than the actual observed value of the statistic. In other words, a *P* value is a probability of data that didn't occur. As such, use of a *P* value to draw conclusions violates the likelihood principle because it involves a sample-space probability. Bayesians point out that the use of *P* values has startling consequences, such as the dependence of the statistical analysis on the intentions of the investigator (Berger and Sellke 1987, Berger and Berry 1988).

Bayesians instead assign a prior subjective probability to each statistical hypothesis. The data modify the belief in a given hypothesis through the likelihood function. A posterior belief probability for each hypothesis emerges from the analysis. The hypothesis with the largest posterior probability is deemed by the investigator the most likely to be true. A pair of hypotheses can be compared with the ratio of their posterior probabilities, which is proportional to the likelihood ratio or "Bayes factor" (Lee 1989, Kass and Raftery 1995).

Under the Bayesian procedure, scientists with different prior beliefs are invited to draw their own conclusions from the data, using their own priors. Consensus would supposedly emerge when most scientists' priors became swamped by large amounts of data (i.e., when their posterior beliefs become nearly identical).

The concepts of significance level and test power are sample-space probabilities and play no role in Bayesian statistics.

Randomization.—Because of the likelihood principle, randomization is irrelevant to Bayesian inference (Basu 1975, 1980) and poses ethical dilemmas in human clinical trials (Royall 1991). To those ecologists who consider randomization to be a cornerstone of the scientific method, the Bayesians' rejection of randomization might at first seem surprising. The argument, briefly caricatured, is as follows. Suppose you have 10 chickens. Five are assigned at random to a treatment group; five to a control. Three of the chickens have oozing, festering sores. By chance, the three are assigned to the control group. The Bayesians would say that only the *outcome* of the randomization procedure has any relevance to the data analysis. The frequentist counts on the randomization to spread the variability (seen or unseen) in experimental units more or less uniformly among the treatments. The frequentist's statistical analysis takes into account all the possible treatment assignments, not just the actual one that happened. Mindless applications of frequentist procedures yields a potentially misleading conclusion: that the treatment improved some measurable characteristic of the chickens. The Bayesian incorporates beliefs about the actual chickens into the prior (healthy chickens would get a prior for control and a prior for treatment, say, and oozing chickens would get different priors for control and treatment). Only the actual treatment assignments would enter the analysis (via the likelihood function). (I can't help interjecting that a frequentist *scientist,* upon discovering the condition of the three chickens, would throw the data in the garbage and begin again.)

Sample surveys.—"Design-based" sample surveys are the standard sampling procedures used by ecologists. Simple random sampling, stratified random sampling, unequal probability sampling (Horvitz–Thompson estimator), and cluster sampling are examples of design-based sampling (see Overton and Stehman 1995 for a contemporary discussion). Bayesians reject design-based sampling (Basu 1971) or are ambivalent about it (Royall 1976, 1988). Under the likelihood principle, the probabilities in probability sampling contribute no information to the estimate of the quantity of interest. The frequentist considers all possible outcomes of the sampling procedure in drawing conclusions about the population sampled. The Bayesian considers only the actual sample drawn. It is not necessary to draw the sample at random; if nonrandom sampling occurred, the Bayesian forms an appropriate prior and proceeds as usual. To the Bayesian, the issue is belief. A random sampling experiment is *ancillary* to the population parameter of interest and hence contributes no information to beliefs about that parameter.

Confidence intervals.—Confidence intervals, as interpreted by Neyman (1937), are a central concept of frequentist statistics. As any instructor of basic statistics classes knows, confidence intervals are one of the most difficult concepts in statistics to understand properly. Even quantitative ecologists stumble on the concept (Poole 1974:49). Confidence intervals seductively suggest more than is actually delivered. What is delivered is an interval, say (9.5, 12.3). What is suggested is that the probability that the parameter is in the interval is (or is approximately) 0.95. Under frequentist probability, it just ain't so: the parameter is either in the interval or out of the interval, period.

Even Fisher, one of the founding fathers of frequentist statistics, chafed under the concept. Realizing that the statement $P(9.5 < \theta < 12.3) = 0.95$ made no sense under the frequentist definition of probability, Fisher devised a new type of probability, which he termed fiducial probability (Fisher 1935). Fiducial probability was a conceptual tent big enough to admit such probability statements as sensical, but it is fair to say that fiducial probability confused rather than clarified the issues (see Kendall and Stuart 1979:147, Buehler 1983, Edwards 1983, Stone 1983). Because fiducial probability has hints of subjectivity, Bayesians are quick to

claim that Fisher saw the right problem but attacked it with the wrong means.

That frequentists can become so confused over one of their most important concepts is a source of continuing amusement to Bayesians.

Furthermore, in the earlier days, confidence intervals were heavily based upon asymptotic approximations. A scientist was sometimes faced with a confidence interval that extended beyond the known range of the parameter (recall the old joke that when the confidence interval for a survival probability had a lower bound extending below zero it meant "a fate worse than death"). Confidence intervals came under intensive fire from Bayesians for this and other failures (Jaynes 1976). Modern-day computational approaches such as profile likelihoods and bootstrapping have rendered some of these complaints moot.

To Bayesians, the interpretation of a confidence interval involves long-run frequencies over a sample space and thus violates the likelihood principle. Instead, the posterior pdf can be used to construct a 95% belief interval (Bayesian probability interval or highest density region) for the parameter in question (see Jaynes 1976, Lee 1989).

Randomization tests.—Randomization tests are gaining popularity in the life sciences (Manly 1991, Sokal and Rohlf 1995). Such tests are utterly repudiated by Bayesians (Basu 1975, 1980). The dreaded P values are not the Bayesians' only objection; the act of basing a test on an ancillary random experiment, after the data are recorded, is a first-degree violation of the likelihood principle. Tests based on other resampling methods such as bootstrapping or jackknifing do not fare any better in the Bayesian view.

Purported advantages of Bayesianism.—Bayesians point to several practical advantages to their approach. First, combining data from several studies, and adding more data in a sequential study, is simple and straightforward under a Bayesian framework. Second, multiple hypotheses (more than two) are easy to sort out, whereas the pairwise hypothesis-testing framework of frequentists is awkward for such sorting (recall the difficult material on pairwise comparisons and linear contrasts in analysis of variance courses). Third, "nuisance parameters," that is, parameters that are unknown but not of direct interest to the investigator, are naturally and elegantly handled by the Bayesian approach. [The standard example of a nuisance parameter is σ^2 when the interest is in the mean μ of a normal $\mathcal{N}(\mu, \sigma^2)$ distribution.] Finally, perhaps most importantly, the designation of probability as a measure of belief unifies the concepts of prediction and estimation.

New developments in statistics are allowing most of these problems to be addressed in a frequentist framework and are being applied in ecological work. Meta-analysis has become an effective tool for combining studies and is gaining acceptance in ecology (Gurevitch et al. 1992). Information criteria, such as that of Akaike (see Sakamoto et al. 1986, Bozdogan 1987) have been used in ecological work to select models from many competing statistical models (see Kemp and Dennis 1991, Lebreton et al. 1992, Anderson et al. 1994, Hooten et al. 1995). Increasing uses in ecology of profile likelihoods (Lebreton et al. 1992, Dennis et al. 1995), randomization and Monte Carlo tests (Manly 1991), and parametric bootstrapping (Dennis and Taper 1994) are sidestepping the "nuisance parameter" problem.

The prediction/estimation dispute between Bayesians and frequentists requires some elaboration. If scientists adopt probability as a measure of belief, as Bayesians would advocate, then prediction and estimation are almost identical concepts. If a coin is tossed, it makes little difference to a Bayesian whether the coin has not landed yet (outcome random) or whether the coin has landed but the outcome has not yet been viewed (outcome fixed, but unknown). To a frequentist, the difference is crucial (and is a major source of confusion among beginning statistics students about confidence intervals, P values, and the like).

A frequentist estimates a fixed but unknown quantity, and predicts a random quantity. These are distinct because the frequentist uses sample-space properties of a model of a random process. To estimate the parameter y, the frequentist collects data x (the outcome of a random process) and builds a likelihood function, $p_X(X|y)$. The likelihood function is a hypothetical model of the random process. The frequentist uses a statistic, $s(x) = \hat{y}$ to estimate y. The estimated model, $p_X(x|\hat{y})$, is used to estimate hypothetical sample-space properties of the statistic $s(x)$ such as how variable the statistic would be under repetitions of the random process. The estimated model is used also to make predictions (i.e., to estimate hypothetical sample space properties) about a future outcome X of the random process.

To the frequentist, the focus must be on the appropriateness of the hypothetical model $p_X(x|y)$. Estimating y and predicting X are almost incidental; the investigator does not have confidence in the estimate or prediction until the form of the model itself is challenged and found to be an adequate representation of the random process.

Therein lies the key. The frequentist, it seems to me, is often not fundamentally interested in estimation or prediction; rather, the frequentist is interested in *explanation.* The *form* of the likelihood function is a central part of the theory being constructed. The parameter y is just an unknown portion of the theory and is not the only purpose for collecting the data. The prediction X is interesting, and indeed pleasing if it turns out to be accurate, but it is not the whole purpose of the investigation. The main purpose of prediction is to challenge the model. A good prediction is good because the outcome is not unreasonable under the model. A bad prediction is bad because the outcome was extreme enough as to be unreasonable under the model. In other words, bad predictions, in a sample-space framework,

come from bad models. To a frequentist, models are disposable and are screened by comparing predictions to models based on the sample-space properties of the models.

It is hard to see how a model can be adequately evaluated without resorting to sample-space properties. It is hard to see how a Bayesian scientific method can systematically produce satisfying explanations. Sample-space probability is a key tool scientists have for ridding themselves of bad theories. If the distinction between estimation and prediction is as a result confusing and difficult, perhaps it is the price we must pay for progress.

Bayesians, you see, are not allowed to look at their residuals. It violates the likelihood principle to judge an outcome by how extreme it is under a model. To a Bayesian, there are no bad models, just bad beliefs.

Bayesian Statistics in Ecology

As can be gathered from the above discussion, the context of the Bayesian ecology papers published here is laden with controversies. In light of these tensions, I now provide some remarks about the particular papers.

Taylor et al. (1996).—The Bayesian population viability analysis by Taylor et al. (1996) illustrates why scientists need frequentist model evaluation techniques. The lack of serious diagnostic analyses or data-based criticism of their model is telling. The main model component, wherein occur the most interesting biological assumptions about Spectacled Eider population growth, is located in the prior.

Taylor et al. (1996) treat population size N_t as an unknown parameter that changes through time according to a stochastic population model (stochastic exponential growth/decline). The parameters in the growth model are given prior distributions. The data, $N_I(t)$, are abundance indexes arising from surveys and are modeled essentially as N_t + (normal sampling error). The viability assessments are based on the resulting posterior distributions for the parameters, in particular, that of r.

The stochastic exponential growth/decline model is off the table, so to speak, when it comes to evaluating their analyses. It is a part of their prior beliefs. We do not know if their growth model "fits" or not. We also do not know if their sampling model is adequate. The data, in Bayesian analysis, are not permitted to shed light on model adequacy.

Consider an alternative frequentist analysis. The stochastic exponential growth/decline model in population viability analysis (PVA) was extensively discussed by Dennis et al. (1991) (a paper uncited by Taylor et al. 1996). Dennis et al. stressed: (1) the use of diagnostic procedures for model evaluation, (2) the incorporation of uncertainty in parameter estimation into population risk estimates; and (3) the use of "quasi-extinction" thresholds of population size (akin to Taylor et al.'s [1996] level of 250 Spectacled Eiders), rather than true extinction levels (<1), in order to minimize errors due to Allee effects, extinction vortexes, and so on.

I applied the stochastic exponential growth/decay model of Dennis et al. (1991) directly to the breeding pair time series (Taylor et al. 1996:Table 1). The model includes environmental variability but not sampling variability; it is $N_t = N_{t-1}\exp(\mu + \sigma Z_t)$, where N_t is the population abundance as indexed, Z_t is standard normal noise, and μ and σ^2 are parameters. I obtained estimates of the model's two unknown parameters: $\hat{\mu} = -0.0622$, $\tilde{\sigma}^2 = 0.265$ (notation of Dennis et al. 1991). I performed parametric bootstrapping, that is, I generated bootstrap values $\hat{\mu}^*$ and $\tilde{\sigma}^{2*}$ from the estimated sampling distributions of $\hat{\mu}$ and $\tilde{\sigma}^2$ and then simulated trajectories from the model using the bootstrap values. With the bootstrapping procedure, I estimated the probability, $G(t)$, of the time series reaching the level 250 starting from 2363, within t years. For various values of t, the estimates (along with bootstrapped 95% confidence intervals) were: $\hat{G}(10) = 0.246$ (0.054, 0.513), $\hat{G}(20) = 0.462$ (0.103, 0.866), $\hat{G}(30) = 0.580$ (0.142, 0.956), $\hat{G}(40) = 0.663$ (0.124, 0.984), $\hat{G}(50) = 0.686$ (0.106, 0.993).

According to Taylor et al. (1996), the population is in little jeopardy of reaching 250 until well after 50 yr in the future (their Fig. 6). My quick analysis above indicates that the *breeding pair time series* has a substantial risk of reaching 250 within even 10 yr. Note how the uncertainty in my estimates of first passage probabilities increases as time increases.

The point is, the model forms matter. But the real difference between our analyses is not the model forms (we could have easily used different models), but in the approach and scientific philosophy used. My analysis is *vulnerable* to being tossed out by model evaluation methods. It is straightforward to perform residual-based diagnostic analyses to ascertain whether or not the model I used adequately describes the variability in the data (briefly: the normal-based noise on the logarithmic scale does a pretty good job, but there is some first-order autocorrelation, and the 1987–1988 jump is an outlier). It is straightforward to use hypothesis testing techniques (Dennis and Taper 1994) and model selection criteria (Hooten et al. 1995, Taper et al. 1995) to compare the model with other less or more biologically detailed models. The model of Taylor et al. (1996), by contrast, is not vulnerable to falsification in the frequentist (and scientifically Popperian) sense.

Ver Hoef (1996).—It is debatable whether parametric empirical Bayes (PEB) methods are Bayesian. Purists point out that PEB methods violate the likelihood principle (Lindley 1983). At best, PEB methods provide an improvement to the likelihood function for a random process. Parameters in a stochastic model, formerly taken as fixed, are in turn modeled as arising from some stochastic mechanism. Thus, Ver Hoef takes the time

sequence of mean abundances $\theta_1, \theta_2 \ldots, \theta_k$, and models them as arising from a time series model (linear trend with noise). The θ_i's must be estimated by population sampling procedures (mark–recapture, etc.), yielding observations $\hat{\theta}_1, \hat{\theta}_2, \ldots \hat{\theta}_k$. The linear trend "hypermodel" is combined with the sampling model to form a likelihood function for the observations. The hypermodel plays somewhat of the role of a prior, except that the hyperparameters (intercept, slope, and variance in the linear trend model) are estimated from the data in frequentist fashion via the likelihood function. The combined sampling/hypermodel likelihood function is presumably a more accurate representation of the variability in the data, and, not surprisingly, confidence intervals for population abundances have better statistical coverage properties.

This is frequentism, pure and simple. Build a better sample-space model; get better confidence intervals. In frequentist statistics, it is standard practice to improve models by assuming that various parameters arise from their own random process. Random effects models in analysis of variance is one example (effects parameters arising from normal distributions). Ver Hoef's model of an unobservable time series process and an observable sampling process is an example of a frequentist "state-space" model (Carlin et al. 1992).

Wolfson et al. (1996).—Wolfson et al. (1996) recite the claim, often made, that environmental management decisions are better made in a Bayesian framework. Indeed, the mathematical theory of decisions has become increasingly a Bayesian one (Berger 1985, Lindley 1985). The mathematical theory can be characterized as a personal decision theory; it represents the sort of decisions individuals do when, for example, playing poker. The long-run chance that your opponent has a great set of cards is extremely low. However, you must base your immediate betting decisions in part on prior information (facial expressions, opponent's bluffing history), or lose. Wolfson et al. (1996) analyze examples from environmental pollution involving "stakeholders" (local residents potentially affected by the pollution, for instance) and a government regulatory agency (Environmental Protection Agency, EPA).

A recurring problem in environmental decision making is the lack of adequate data. The regulator must make a "decision" as to, say, whether or not an area is polluted enough to warrant expensive remediation measures. The data are typically inconclusive, from a frequentist standpoint. Decision theory states that the regulator's (predata) beliefs about the degree of pollution should be quantified into a prior distribution, and that the data and prior should be mixed into a posterior distribution of beliefs. The decision is calculated from the posterior, using a function (the loss function) representing the losses associated with each action.

The Bayesian manager, in other words, instead of saying "I don't know whether the area is polluted," says something like "there is a 20% chance that the area is polluted."

I object to calling this science. Science is not about decisions; science is about making convincing conclusions. If the information is inadequate and a policy decision must be made, the regulator should *take responsibility,* clearly admit that the information is inadequate, institute an interim, cautious policy until better data become available, and not pass off personal (or anyone's) beliefs disguised by fancy statistical analysis as science.

Also, use of personal decision theory for public decisions is controversial, to say the least. It is understandable that the stakeholders might be biased. It is understandable that polluters might be biased. For the EPA regulators not to enter the investigation with open minds, though, is a violation of public trust. A prior probability of no contamination equal to 0.95, as recommended in Wolfson et al. (1996), is not open-mindedness. A prior probability of p, where p is any number between zero and one, is not open-mindedness. It is the EPA's job to find out how much pollution there is and act accordingly.

Moreover, for the EPA to admit the special interests' opinions into the data analysis is an invitation to chaos. Consider, for instance, the loss function proposed by Wolfson et al. (1996) in their case study number 1 (radioactive groundwater contamination). In the loss function there is a quantity, ν_2, representing the value placed by the stakeholders on unnecessary public spending by the EPA on sampling and remediation. This quantity is "elicited" from the stakeholders prior to the data analysis. Just wait until stakeholders learn this game! If they express extreme enough views to the EPA investigators, they will *influence* the *conclusions* of the sampling! (Bayesians think this is reasonable, logical, and desirable!)

My view is this: ν_2 is a nonexistent quantity. It varies with the wind. It has no mean, no variance. It is nonstationary. It, and other quantities like it, have no place in scientific data analysis.

Ludwig (1996).—Ludwig's (1996) analysis illustrates an operational difficulty with Bayesian methods that frequentists find worrisome. The problem concerns how to quantify ignorance into a prior. Consider, for illustration purposes, a simpler version of Ludwig's population model. Suppose log-population size, X_t, undergoes a simple drift-free random walk with variability parameter σ^2. Thus, X_t would have a normal distribution with a mean of x_0 (log-initial population size) and a variance of $\sigma^2 t$. Let ψ be the probability that X_t attains a lower level a before hitting an upper level b, where $a \leq x_0 \leq b$. The quantity ψ is the probability that quasi-extinction occurs before recovery, and its estimation is a typical objective of PVA (Dennis et al. 1991). For fixed values of a, b, and x_0, ψ is a function of σ^2.

Now, ignorance about σ^2, quantified for instance into

a "noninformative prior" for $\log \sigma^2$, does *not* transform into ignorance about the risk parameter ψ. A noninformative prior for σ^2 will produce one posterior distribution for ψ, along the lines of Ludwig's analysis. Suppose, however, a noninformative prior is assumed for ψ instead. After all, ψ is the main parameter of interest, and we might reasonably suppose in this problem that all we really know is that ψ is somewhere between 0 and 1. A *different* posterior distribution for ψ results! Which of these beliefs should we believe? On which posterior should we bet the species?

Ludwig's model has more parameters (primarily for density dependence), but the basic problem remains. Prior lack of knowledge about ψ is not the same as prior lack of knowledge about any other parameter.

Criticizing maximum likelihood (ML) methods for producing only point estimates of ψ is a red herring. The frequentist approach concentrates on the point estimate of ψ *and its sample-space variability.* Dennis et al. (1991) wrote extensively about the need for quantifying the uncertainty in ML point estimates, and suggested approaches for calculating standard errors. An accurate analysis of the variability of the ML estimate of ψ for Ludwig's model and data would likely yield conclusions about our knowledge of ψ similar to those of Ludwig's.

The quantity ψ is, in fact, a sample-space probability. It is the probability of occurrence of a set of extreme population trajectories. For a living species, ψ is the probability of "data that didn't happen." The likelihood principle forbids the use of such quantities in inferential conclusions. If you think confidence intervals are tough to understand, try listening to how Bayesians estimate forbidden quantities.

As noted earlier in my discussion of the Taylor et al. (1996) paper, a major practical challenge in PVA is choosing an adequate model. Density-dependent models and density-independent models, for instance, can give substantially different risk estimates (Stacey and Taper 1992). Greater attention in PVA to diagnostic methods seems essential (Dennis et al. 1991, Dennis and Taper 1994). Methods for model selection based on information criteria are yielding promising results (Hooten et al. 1995, Taper et al. 1995). One hopes that Bayesian approaches to PVA will concentrate on building better likelihood functions.

Ellison (1996).—Ellison's paper explicitly admits that Bayesianism means abandoning the Popperian falsificationist approach to science. It is one of the first Bayesian expositions I have read to do so. Ecologists take note.

What scientific philosophy is offered by Bayesians as a replacement? Ellison does not give it a label; I would argue that the Bayesian philosophy of science is scientific relativism.

To scientific relativists, truth is subjective. Scientific theories are socially "warranted" constructs that do not necessarily make progress toward uncovering universal truths. Scientific discourse is a social power game in which the very rules of evidence are socially warranted by the ruling clique and change with fashion, a sort of intellectual Calvinball. There is no truth, only beliefs. Scientific relativists have published vigorous scholarly attacks on scientists' claims of objectivity in journals of feminist and multicultural studies (see Levitt and Gross 1994). So-called "creation science" has its basis in scientific relativism (for example, Campbell 1996).

Bayesians want to incorporate such beliefs directly into the conclusions drawn from data. They claim that explicitly stating prior beliefs and mixing the beliefs into the data analysis is more desirable than having the beliefs hidden and mixed into the data in uncontrolled, unacknowledged ways.

No scientist I know would deny that science is a human enterprise and is fraught with human imperfections. Scientists have careers to build, families to feed, grants to renew. Scientists are biased; some are petty jerks, some are racists, some are deluded. We all know scientists who have stubborn, data-resistant beliefs.

Frequentist statistics may indeed have hidden subjective biases. What is important, however, is that it seeks to rid itself of such biases. It seeks to remove the scientist's beliefs from the conclusions as much as possible and let the data do the talking. It seeks to understand and improve the best, most effective parts of the scientific method.

In particular, an enduring and effective argument in science for convincing skeptics is to make the skeptics' beliefs blow up in the skeptics' faces. The much-maligned "null hypothesis," if used properly, is a powerful argumentative tool and has been valuable for reducing the noise level in scientific discourse (Bross 1971).

What is amazing is that the process works so well. Science has made spectacular progress under the reductionist, hypothetico-deductive, falsificationist way of thinking. The Fisher–Neyman–Pearson–Wald (and, may I add, –Rao–Efron) frequentist contributions toward building that way of thinking into our data analysis have accelerated that progress immensely. Progress, advancement, or whatever we call that overall tendency of our theories and understandings to get better, is not predicted by scientific relativism. Bayesian statistics, though it has been in full-fledged existence since Laplace, and investigated intensively by mathematical statisticians for 60 yr, has played virtually no role in this progress. Bayesianism means never having to say you're wrong.

Ellison points out the introspection about scientific methodology that has characterized ecology in recent years. The fields of community and evolutionary ecology, for instance, have had major crises of confidence, as born witness by the famous November 1982 issue of *The American Naturalist.* Ellison paints a portrait of continuing malaise in our science, stemming from

the unattainable ideals of falsificationist epistemology and the lack of utility to decision-makers of our scientific conclusions. I couldn't disagree more.

I see a renewed, vigorous science emerging from a period of soul-searching and self-criticism. Ecologists have formed a clearer picture of the types of evidence it takes to make a convincing explanation of some aspect of the biological world. There is now increased, not decreased, importance attached to classical scientific method. Numerous observational and experimental studies published recently have emphasized rigor, originality, cleverness, and connection to general theories (a much shortened list of my exemplary favorites: Tilman and Wedin 1991, Grant and Grant 1993, Gross et al. 1993, Hanski et al. 1993, Lindström et al. 1994, Marquis and Whelan 1994, McLaren and Peterson 1994, Barry et al. 1995, Schoener and Spiller 1995, Valone and Brown 1995). Gurevitch et al. (1992) searched just six journals for the previous 10 yr and found 42 articles with data on a particular aspect of competition (biomass change) of sufficient quality for inclusion in a meta-analysis. Applied ecology and natural resource management have heard calls toward increased scientific rigor (e.g., Romesburg 1981, 1991). Ecologists and environmental scientists are presenting policy-makers and the public with remarkably strong scientific conclusions on an enormous array of environmental problems, from endangered species status to global change (Newman 1993). That ecologists' recommendations are frequently unheeded is not, as Ellison claims, a failing of the science, but rather of a political system that allows wealthy stakeholders to have an inordinate influence on environmental policy. Ecology is not a sick science, but a healthy and vital science making slow but steady progress on really, really difficult problems.

Ellison's claim that 96% of the authors of the papers he sampled from *Ecology* unwittingly considered the probability of their alternative hypotheses to be high is unjustified. In a well-designed study, H_1 is the only plausible alternative to H_0; if the mechanism H_0 cannot reasonably have given rise to the data, one does not have to be a Bayesian to state that the data support H_1. In some cases, the authors might simply have used careless wordings: why not ask the authors what they really meant? Frequentist statistical concepts are hard; wordings are delicate and treacherous. Many ecologists' grasp of the concepts is shaky. When ecologists at large attain command of the differences between frequentist and Bayesian approaches, and of the implications of attaching personal probabilities to hypotheses, I seriously doubt we will find that our science is full of closet Bayesians.

Attaining that command will be tough. Statistics education in ecology is already misdirected and ill-conceived. Too much time is spent in "methods" courses, that is, courses without calculus prerequisites. Statistics is a postcalculus topic; a student who takes only methods courses is not likely to gain any level of comfort with the concepts in frequentist *or* Bayesian statistics (a 1-yr course at the level of, say, Rice 1988 is far preferable for gaining confidence in statistical concepts than the conventional 2-yr parade of methods courses). However, ecologists cannot now hope to avoid these issues and go about business as usual; the Bayesian foot is in the door. What will the traditionally schooled ecologist do when his/her paper is refereed by a Bayesian? How will Bayesians get their studies past frequentist editors? How will ecologists respond to a Bayesian PVA of the Northern Spotted Owl put forth by the forest products industry?

Bayesian statistics, as I have tried to demonstrate here, is not just a new set of tools for ecologists to use. It is a whole different way of doing business. Bayesian and frequentist statistics cannot logically coexist.

The present group of papers offer fine expositions and nice, illustrative ecological examples (interested readers should also see Gazey and Staley 1986, Johnson 1989, and Raftery et al. 1995). However, the burden of proof is still on Bayesians to show that ecology can continue its progress with subjective probability approaches. Frequentism, like the peer review system, is imperfect but has a proven track record. We need to see a room full of community ecologists making progress and convincing each other using Bayesian arguments. Until I see some new compelling Bayesian understandings of nature, I will not be a believer.

ACKNOWLEDGMENTS

My thanks to Subash Lele and Mark Taper for their insightful comments on the paper, and for countless hours of discussions. The opinions expressed are my own.

LITERATURE CITED

Anderson, D. R., K. P. Burnham, and G. C. White. 1994. AIC model selection in overdispersed capture-recapture data. Ecology **75**:1780–1793.

Barry, J. P., C. H. Baxter, R. D. Sagarin, and S. E. Gilman. 1995. Climate-related, long-term faunal changes in a California rocky intertidal community. Science **267**:672–675.

Basu, D. 1971. An essay on the logical foundations of survey sampling, part one (with comments). Pages 203–242 *in* V. P. Godambe and D. A. Sprott, editors. Foundations of statistical inference. Hole, Rinehart, and Winston, Toronto, Canada.

———. 1975. Statistical information and likelihood (with discussion). Sankyā **A37**:1–71.

———. 1980. Randomization analysis of experimental data: the Fisher randomization test (with comments). Journal of the American Statistical Association **75**:575–595.

Berger, J. O. 1985. Statistical decision theory and Bayesian analysis. Springer-Verlag, Berlin, Germany.

Berger, J. O., and D. A. Berry. 1988. Statistical analysis and the illusion of objectivity. American Scientist **76**:159–165.

Berger, J. O., and T. Sellke. 1987. Testing a point null hypothesis: the irreconcilability of P values and evidence (with comments). Journal of the American Statistical Association **82**:112–139.

Bozdogan, H. 1987. Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. Psychometrika **52**:345–370.

Bross, I. D. J. 1971. Critical levels, statistical language, and

scientific inference (with discussion). Pages 500–519 *in* V. P. Godambe and D. A. Sprott, editors. Foundations of statistical inference. Holt, Rinehart, and Winston, Toronto, Canada.

Buehler, R. J. 1983. Fiducial inference. Pages 76–81 *in* S. Kotz, N. L. Johnson, and C. B. Read, editors. Encyclopedia of statistical sciences. Volume 3. Wiley, New York, New York, USA.

Campbell, J. A. 1996. John Stuart Mill, Charles Darwin, and the culture wars: resolving a crisis in education. Intercollegiate Review **31**:44–51.

Carlin, B. P., N. G. Polson, and D. G. Stoffer. 1992. A Monte Carlo approach to nonnormal and nonlinear state-space modeling. Journal of the American Statistical Association **87**:493–500.

Dennis, B., R. A. Desharnais, J. M. Cushing, and R. F. Costantino. 1995. Nonlinear demographic dynamics: mathematical models, statistical methods, and biological experiments. Ecological Monographs **65**:261–281.

Dennis, B., P. L. Munholland, and J. M. Scott. 1991. Estimation of growth and extinction parameters for endangered species. Ecological Monographs **61**:115–143.

Dennis, B., and M. L. Taper. 1994. Density dependence in time series observations of natural populations: estimation and testing. Ecological Monographs **64**:205–224.

Edwards, A. W. F. 1983. Fiducial distributions. Pages 70–76 *in* S. Kotz, N. L. Johnson, and C. B. Read, editors. Encyclopedia of statistical sciences. Volume 3. Wiley, New York, New York, USA.

Ellison, A. M. 1996. An introduction to Bayesian inference for ecological research and environmental decision-making. Ecological Applications **6**:1036–1046.

Fisher, R. A. 1935. The fiducial argument in statistical inference. Annals of Eugenics **6**:391–398.

Gazey, W. J., and M. J. Staley. 1986. Population estimation from mark-recapture experiments using a sequential Bayes algorithm. Ecology **67**:941–951.

Grant, B. R., and P. R. Grant. 1993. Evolution of Darwin's finches caused by a rare climatic event. Proceedings of the Royal Society of London **B251**:111–117.

Gross, J. E., L. A. Shipley, N. T. Hobbs, D. E. Spalinger, and B. A. Wunder. 1993. Functional response of herbivores in food-concentrated patches: tests of a mechanistic model. Ecology **74**:778–791.

Gurevitch, J., L. L. Morrow, A. Wallace, and J. S. Walsh. 1992. A meta-analysis of competition in field experiments. American Naturalist **140**:539–572.

Hanski, I., P. Turchin, E. Korpimäki, and H. Henttonen. 1993. Population oscillations of boreal rodents: regulation by mustelid predators leads to chaos. Nature **364**:232–235.

Hooten, M. M., M. L. Taper, W. P. Kemp, and B. Dennis. 1995. The form of density dependence in 125 time series of natural populations identified using information criteria. Bulletin of the Ecological Society of America (Supplement, Part Two) **76**:120.

Jaynes, E. T. 1976. Confidence intervals vs Bayesian intervals (with discussion). Pages 175–257 *in* W. L. Harper and C. A. Hooker, editors. Foundations of probability theory, statistical inference, and statistical theories of science. Volume II. D. Reidel, Dordrecht, Holland.

Johnson, D. H. 1989. An empirical Bayes approach to analyzing animal surveys. Ecology **70**:945–952.

Kass, R. E., and A. E. Raftery. 1995. Bayes factors. Journal of the American Statistical Association **90**:773–795.

Kemp, W. P., and B. Dennis. 1991. Toward a general model of rangeland grasshopper (Orthoptera: Acrididae) phenology in the steppe region of Montana. Environmental Entomology **20**:1504–1515.

Kendall, M., and A. Stuart. 1979. The advanced theory of statistics. Volume 2. Fourth edition. Oxford University Press, New York, New York, USA.

Lebreton, J.-D., K. P. Burnham, J. Clobert, and D. R. Anderson. 1992. Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. Ecological Monographs **62**:67–118.

Lee, P. M. 1989. Bayesian statistics: an introduction. Oxford University Press, New York, New York, USA.

Levitt, N., and P. R. Gross. 1994. Higher superstition: the academic left and its quarrels with science. Johns Hopkins University Press, Baltimore, Maryland, USA.

Lindley, D. V. 1982. Bayesian inference. Pages 197–204 *in* S. Kotz, N. L. Johnson, and C. B. Read, editors. Encyclopedia of statistical sciences. Volume 1. Wiley, New York, New York, USA.

———. 1983. Comment. (Discussion of a paper by C. N. Morris: Parametric empirical Bayes inference: theory and applications.) Journal of the American Statistical Association **78**:47–65.

———. 1985. Making decisions. Second edition. Wiley, New York, New York, USA.

———. 1990. The 1988 Wald memorial lectures: the present position in Bayesian statistics (with comments). Statistical Science **5**:44–89.

Lindström, E. R., H. Andrén, P. Angelstam, G. Cederlund, B. Hörnfeldt, L. Jäderberg, P.-A. Lemnell, B Martinsson, K. Sköld, and J. E. Swenson. 1994. Disease reveals the predator: sarcoptic mange, red fox predation, and prey populations. Ecology **75**:1042–1049.

Ludwig, D. 1996. Uncertainty and the assessment of extinction probabilities. Ecological Applications **6**:1067–1076.

Manly, B. F. J. 1991. Randomization and Monte Carlo methods in biology. Chapman and Hall, New York, New York, USA.

Marquis, R. J., and C. J. Whelan. 1994. Insectivorous birds increase growth of white oak through consumption of leaf-chewing insects. Ecology **75**:2007–2014.

McLaren, B. E., and R. O. Peterson. 1994. Wolves, moose, and tree rings on Isle Royale. Science **266**:1555–1558.

Newman, E. I. 1993. Applied ecology. Blackwell Scientific, Oxford, UK.

Neyman, J. 1937. Outline of a theory of statistical estimation based on the classical theory of probability. Philosophical Transactions of the Royal Society of London **A236**:333–380.

Overton, W. S., and S. Stehman. 1995. The Horvitz–Thompson theorem as a unifying perspective for probability sampling: with examples from natural resource sampling. American Statistician **49**:261–268.

Poole, R. 1974. Introduction to quantitative ecology. McGraw-Hill, New York, New York, USA.

Raftery, A. E., G. H. Givens, and J. E. Zeh. 1995. Inference from a deterministic population model for bowhead whales (with discussion). Journal of the American Statistical Association **90**:402–430.

Rice, J. A. 1988. Mathematical statistics and data analysis. Wadsworth, Belmont, California, USA.

Romesburg, H. C. 1981. Wildlife science: gaining reliable knowledge. Journal of Wildlife Management **45**:293–313.

———. 1991. On improving the natural resources and environmental sciences. Journal of Wildlife Management **55**: 744–756.

Royall, R. M. 1976. Current advances in sampling theory: implications for human observational studies. American Journal of Epidemiology **104**:463–474.

———. 1988. The prediction approach to sampling theory. Pages 399–413 *in* P. R. Krishnaiah and C. R. Rao, editors. Handbook of statistics. Volume 6. Elsevier, New York, New York, USA.

———. 1991. Ethics and statistics in randomized clinical trials. Statistical Science **6**:52–88.

Sakamoto, Y., M. Ishiguro, and G. Kitagawa. 1986. Akaike information criterion statistics. KTK Scientific, Tokyo, Japan.

Schoener, T. W., and D. A. Spiller. 1995. Effect of predators and area on invasion: an experiment with island spiders. Science **267**:1811–1813.

Sokal, R. R., and F. J. Rohlf. 1995. Biometry: the principles and practice of statistics in biological research. Freeman, New York, New York, USA.

Stacey, P. B., and M. Taper. 1992. Environmental variation and the persistence of small populations. Ecological Applications **2**:18–29.

Stigler, S. M. 1986. The history of statistics: the measurement of uncertainty before 1900. Belknap Press, Cambridge, Massachusetts, USA.

Stone, M. 1983. Fiducial probability. Pages 81–86 *in* S. Kotz, N. L. Johnson, and C. B. Read, editors. Encyclopedia of statistical sciences. Volume 3. Wiley, New York, New York, USA.

Taper, M. L., M. M. Hooten, B. Dennis, S. Lele, and W. P. Kemp. 1995. The use of information criteria in the identification of density dependence and independence in biological time series data. Bulletin of the Ecological Society of America (Supplement, Part Two) **76**:259–260.

Taylor, B. L, P. R. Wade, R. A. Stehn, and J. F. Cochrane. 1996. A Bayesian approach to classification criteria for Spectacled Eiders. Ecological Applications **6**:1077–1089.

Tilman, D., and D. Wedin. 1991. Oscillations and chaos in the dynamics of a perennial grass. Nature **353**:653–655.

Trader, R. L. 1989. Bayes, Thomas. Pages 14–17 *in* S. Kotz, N. L. Johnson, and C. B. Read, editors. Encyclopedia of statistical sciences. Supplement volume. Wiley, New York, New York, USA.

Valone, T. J., and J. H. Brown. 1995. Effects of competition, colonization, and extinction on rodent species diversity. Science **267**:880–883.

Ver Hoef, J. M. 1996. Parametric empirical Bayes methods for ecological applications. Ecological Applications **6**: 1047–1055.

Wolfson, L. J., J. B. Kadane, and M. J. Small. 1996. Bayesian environmental policy decisions: two case studies. Ecological Applications **6**:1056–1066.