

# Ask Dr. STATS

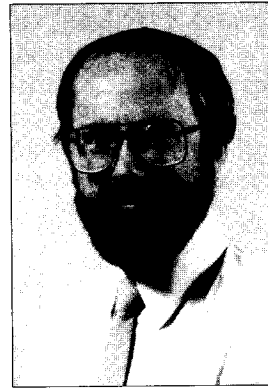
BOB STEPHENSON  
Professor of Statistics  
Iowa State University

## Preface

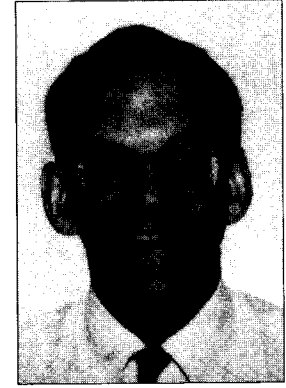
A while back Dr. STATS got an inquiry from Andrew Robinson, a graduate student in forest biometrics at the University of Minnesota. Andrew asked, "I wonder if you would be so kind as to discuss the distinctions between the Bayes/Empirical Bayes/Frequentist perspectives on a simple statistical process or two? For example, significance testing, point estimation, and interval construction for a sample from a normal population." Dr. STATS was stumped. Having been raised (academically) as a frequentist, with only a smattering of Bayes and Empirical Bayes, Dr. STATS turned to a colleague to field this question. *STATS* magazine is indeed fortunate to have the article "A Primer on the Bayesian Approach to Statistical Inference" by Hal S. Stern. It not only discusses the distinctions mentioned by Andrew but provides a clear and balanced exposition on a controversial topic in statistics.

---

*Hal S. Stern is professor of statistics at the Department of Statistics, Iowa State University. He teaches both introductory and graduate level courses in probability and statistics. His research interests include: applied Bayesian statistics, statistical computing, model diagnostics. He is interested in applications of statistics to the biological sciences, social sciences, and sports. He is an editor-elect of Chance magazine and also writes a column, "A Statistician Reads the Sports Pages," for Chance. Stern received a B.S. degree in mathematics from the Massachusetts Institute of Technology and an M.S. and a Ph.D. degree in statistics from Stanford University. His E-mail address is hstern@iastate.edu.*



**Bob Stephenson**



**Hal S. Stern**

## A Primer on the Bayesian Approach to Statistical Inference

HAL S. STERN  
Professor of Statistics  
Iowa State University

### ■ 1. Introduction

The Bayesian approach to statistical inference has become increasingly popular in recent years, statistical journals sometimes appear to be full of articles developing or applying Bayesian methods. Bayesian methods are also becoming more common in subject-area journals. The increase in activity is due largely to computational developments that have made it practical for people who have always found the theory attractive (and others that haven't) to now use Bayesian methods in working with data. Two recent books, Gelman et al. (1995) and Carlin and Louis (1996), provide many examples of applications of Bayesian methods and describe the computational issues in some detail. Despite the increased level of Bayesian activity there is a large audience, including advanced statistics undergraduates, beginning statistics graduate students, and researchers outside of statistics, who are not aware of exactly what the differences are between the Bayesian approach to inference and the procedures commonly used under the heading frequentist inference. The goal of this article is to reach this audience by focusing on a simple example. This style of article involves some risk because we will find that for simple one-parameter models the Bayesian approach does not differ very much from the frequentist approach, and readers may wonder what the fuss is about. A complex example, however, risks confusing readers on details not directly relevant to the Bayesian or frequentist ideas.

Before proceeding, in the interest of full disclosure, I must confess that I find the arguments in favor of the Bayesian paradigm quite compelling. I hope that the discussion below is relatively unbiased (in the colloquial rather than technical sense), but there are no guarantees.

## ■ 2. Frequentist Inference

Apologies aside, we begin by considering one of the simplest of all statistical problems. Assume that  $Y_1, \dots, Y_n$  are  $n$  independent random variables each having a Gaussian or normal distribution with mean  $\mu$  and variance  $\sigma^2$ . It may help to have a concrete example, so let's take the  $Y_i$ s to be repeated measurements of the weight of an object on a particular scale. The scale is thought to be accurate so that the mean  $\mu$  is the true, unknown weight of the object but there are random measurement errors due to factors like vibration. The measurement errors imply that there is variability (the measurements will not all be equal) and  $\sigma^2$  is a measure of this variability.

There is a minor problem here since the normal distribution assigns probability mass to the entire real line including negative values whereas weights must be positive. We don't worry about this issue, assuming that the objects we measure are relatively heavy compared to  $\sigma$ . To keep things incredibly simple we assume that  $\sigma^2$  is known. Although this is usually unrealistic, we might have a great deal of experience with this scale and be willing to treat the variance of the measurement errors as known. The goal is inference about the fixed but unknown parameter  $\mu$  based on a sample of observed values  $y_1, \dots, y_n$ . We follow classical notation in this section, allowing uppercase Roman letters to denote random variables and lowercase Roman letters to denote observed values.

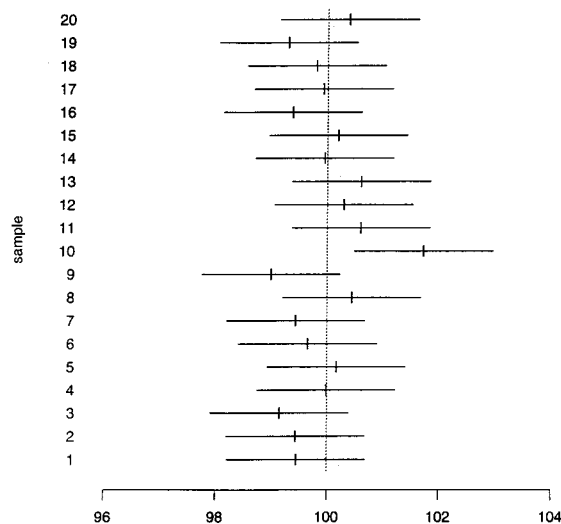
Before getting down to details we stop to discuss the title of this section. The term "frequentist" is actually derived from the argument that uses long-run frequencies of events to define probabilities. Hence the term "frequentist inference" is not exactly well defined. What we have are a collection of techniques or procedures that have good properties under the repeated sampling view that characterizes the frequentist definition of probability. It is these procedures that we review here in the context of point estimation, interval estimation, and significance testing. An exhaustive description is not attempted, instead we illustrate the kinds of statements that one makes from the frequentist perspective and point out some relevant features.

## 2.1 Point estimation

The sample mean,  $\bar{Y} = (1/n)\sum_i Y_i$ , is a natural estimator for the population mean  $\mu$ . How shall we decide if this is in fact a good estimator? The frequentist perspective evaluates estimators (or any other inference procedure) based on properties that would hold under repeated sampling from the same model with fixed values of the unknown parameters. Thus, for example  $\bar{Y}$  is said to be unbiased because the average value of  $\bar{Y}$  in repeated samples from a population with mean  $\mu$  is equal to  $\mu$  (we could write  $E(\bar{Y}|\mu) = \mu$  where the  $\mu$  is included as a conditioning argument to emphasize that  $\mu$  is being held fixed in the repeated samples). Unbiasedness is one property that supports the use of the sample mean as an estimator of  $\mu$ . We can also evaluate other properties, e.g.,  $\bar{Y}$  is minimum variance among unbiased estimators. It is also common to evaluate large-sample or asymptotic properties of the estimators (examples include properties like consistency and efficiency but we don't formally define these terms here).

## 2.2 Interval estimation

The sampling distribution of  $\bar{Y}$  is  $N(\mu, \sigma^2/n)$ . This is the distribution that would be observed in repeated samples of size  $n$  from our normal population with mean  $\mu$  and variance  $\sigma^2$ . From this sampling distribution it follows that the interval



**Figure 1.** Demonstration of the frequentist interpretation of confidence intervals. Twenty 95% confidence intervals based on 20 samples of size 10 from a normal population with mean 100 and standard deviation 2. The tick mark in the middle of each interval is located at the sample mean. The dashed vertical line indicates the true population mean. Nineteen of the 20 intervals contain the true population mean.

$$\left( \bar{Y} - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{Y} + 1.96 \frac{\sigma}{\sqrt{n}} \right)$$

will contain the true population mean in 95% of the repeated samples. This interval is known as a 95% confidence interval for the population mean  $\mu$ . It is emphasized in standard introductory textbooks (e.g., Moore, 1995, pg. 344) that the confidence level applies to the performance of our procedure in repeated samples. Under this view, it is not appropriate to say that the probability is 95% that the true mean  $\mu$  is in the interval for a single fixed data set (i.e., once the random variable  $Y$  has been replaced by the observed mean  $\bar{y}$ ). Our procedure for generating intervals is such that the interval will contain the true value 95% of the time; however, for a given data set the interval either contains the true mean or not with no randomness remaining. Figure 1 provides an illustration where  $\mu$  is fixed and separate intervals are created for each of 20 samples (19 of the 20 intervals contain the true value). This frequentist interpretation can be difficult to grasp for beginning students.

### 2.3 Significance testing

The traditional approach to significance testing begins with a null hypothesis and an alternative hypothesis. Then, for an appropriate test statistic  $T(Y_1, \dots, Y_n)$  we derive a procedure that allows us to determine the propriety of the null hypothesis. This is measured by the  $p$ -value which gives the probability that in repeated samples we would obtain a test statistic value as or more extreme than the observed value under the null hypothesis. Small  $p$ -values suggest that the data we have are unusual under the null hypothesis which suggests that the null hypothesis may be false (either that our sample was just an unlucky one). For example, suppose we are weighing a bar supposed to contain 100 lbs. of material and are concerned that the bar may be too light. We would then wish to test the null hypothesis that  $\mu = 100$  (or  $\mu \geq 100$ ) versus the alternative hypothesis  $\mu < 100$  in our normal model. We can use  $Y$  as our test statistic in which case the  $p$ -value for an observed sample mean  $\bar{y}$  is

$$p = \Phi \left( \frac{\sqrt{n}}{\sigma} (\bar{y} - 100) \right).$$

Several aspects of this procedure should be emphasized. First, the test does not treat the null hypothesis and the alternative hypothesis symmetrically. The  $p$ -value is computed assuming the null hypothesis is true and consequently it bears only on the appropriateness of the null hypothesis. The alternative hypothesis serves primarily to help

us decide what test statistic to use and which values of the test statistic should serve as evidence against the null hypothesis. Second, the  $p$ -value is a probability relevant to repeated samples from the population assuming that the null hypothesis is true. The  $p$ -value does not measure the probability that the null hypothesis is true, countless Statistics 101 student exam papers notwithstanding. A third point is that to get a reasonable conclusion we have included more extreme values of the test statistic in our definition of the  $p$ -value even though they weren't observed. Finally, it should be emphasized here that most statisticians (frequentist or Bayesian) now realize that an over-reliance on significance testing, especially on binary accept/reject decisions, is not useful for science. Confidence intervals are constructive statements about plausible values for the unknown parameter  $\mu$ ; tests are negative statements that rule out specific hypothesized values.

### 3. Bayesian inference

The key features of the Bayesian approach to inference are: (1) all unknowns are treated as random variables with probability distributions used to describe the state of our knowledge about these unknowns, and (2) inference about unknowns is derived using Bayes' rule (described below) to condition on the values of observed quantities. Qualitatively, the Bayesian approach begins with a probability distribution describing our state of knowledge about unknowns (usually parameters) before collecting data, and then uses observed data to update this distribution. We now introduce Bayesian terminology and the technical machinery of the Bayesian paradigm in the context of our example. As in the previous section we suppose that  $Y_1, \dots, Y_n$  are  $n$  independent random variables each having a normal distribution with unknown mean  $\mu$  and known variance  $\sigma^2$ . Under the Bayesian paradigm the unknown mean  $\mu$  is also a random variable. Because of this it would be more precise for a Bayesian to say that  $Y_1, \dots, Y_n$ , conditional on the unknown value of  $\mu$ , are  $n$  independent random variables each having a normal distribution with mean  $\mu$  and known variance  $\sigma^2$ . This is known as the data distribution, or more formally as the conditional distribution of the data given the model parameters; in the general case we write  $p(Y_1, \dots, Y_n | \mu)$ . This distribution, known as the likelihood function when it is viewed as a function of  $\mu$  for a given data set, is exactly the same distribution used to construct the frequentist procedures in the previous section. To complete the probabilistic description of all the random variables, the marginal distribution of the

unknown parameter  $\mu$  must be specified. The marginal distribution of  $\mu$  which we denote generically as  $p(\mu)$  is called the prior distribution; it describes our state of knowledge about  $\mu$  before seeing any data. We use  $p(\cdot)$  to represent all distributions in this paper, even though they can be very different functional forms. Readers without any background in statistical theory can think of the prior distribution  $p(\mu)$  as identifying those values of  $\mu$  that we most believe to be true before observing the data. Inference for the unknown  $\mu$  given observed values  $y_1, \dots, y_n$  is obtained from the laws of probability, namely Bayes' rule,

$$p(\mu | y_1, \dots, y_n) = \frac{p(y_1, \dots, y_n | \mu) p(\mu)}{p(y_1, \dots, y_n)},$$

where  $p(y_1, \dots, y_n)$  is the marginal distribution of the data and can be derived from  $p(y_1, \dots, y_n | \mu)$  and  $p(\mu)$ .

The result of Bayes' rule,  $p(\mu | y_1, \dots, y_n)$ , is known as the posterior distribution of  $\mu$  and it describes our state of knowledge about  $\mu$  after observing  $y_1, \dots, y_n$ . The posterior distribution is fundamental to Bayesian inference for  $\mu$ . The Bayesian approach makes probability statements on unknowns after fixing the things that have been observed at their known values. This means that we do not think about repeated samples from the same population — we fix our attention on the sample at hand.

The prior distribution appears to be the most troubling point for many statisticians. They ask where such prior distributions come from and wonder about the impact of the prior distribution on the conclusions that we reach. It is important to address this concern and we return to it in some detail later. For now we pick a convenient prior distribution and revisit our discussion of point estimation, interval estimation, and significance testing. Suppose the  $\mu$  is assumed to follow a normal distribution with mean  $\mu_0$  and variance  $\tau^2$  where  $\mu_0$  and  $\tau^2$  are specified constants. This means that *a priori* we expect the value of  $\mu$  to be near  $\mu_0$ , e.g., the probability under the prior distribution is 0.95 that  $\mu$  is between  $\mu_0 - \tau$  and  $\mu_0 + 2\tau$ . Some algebra is required to obtain the posterior distribution from Bayes' rule but the result is again a normal distribution,

$$\mu | y_1, \dots, y_n \sim N(\mu_n, V_n),$$

where

$$\mu_n = \frac{\frac{n}{\sigma^2} \bar{y} + \frac{1}{\tau^2} \mu_0}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}} \quad \text{and} \quad V_n = \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}$$

To interpret the posterior distribution we use the term precision for the reciprocal of the variance. The posterior mean,  $\mu_n$ , is a precision-weighted average of the sample mean,  $\bar{y}$ , and the prior mean,  $\mu_0$ . If the prior information is extremely

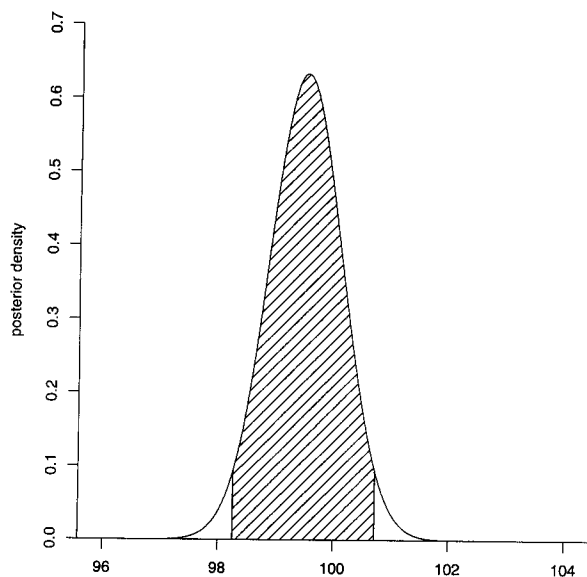
precise ( $\tau^2$  small), then the posterior mean for the unknown  $\mu$  will be heavily influenced by the prior mean ( $\mu_0$ ). If on the other hand our prior information is vague ( $\tau^2$  large), then our posterior mean will be primarily determined by the data. Finally, the precision of the posterior distribution,  $V_n^{-1}$ , is just the sum of the precisions of the prior distribution and the data distribution.

### 3.1 Point estimation

The posterior distribution describes our state of knowledge about  $\mu$  after observing the data, specifying which values are plausible and how likely each is. A point estimate is just a one-number summary of the posterior distribution. In general, the mean of the posterior distribution, the median of the posterior distribution, and the mode of the posterior distribution could all be reasonable point estimates for  $\mu$ . In this example, they all agree but this will not generally be the case. To choose a single point estimate under the Bayesian perspective, we require a loss function that specifies the cost of an estimation error, and then we choose as our estimate the value that minimizes the expected loss under the posterior distribution. Here again, some object to the formality required by the Bayesian paradigm, a loss function is required to define an optimal point estimate. It is possible to consider whether Bayesian point estimates possess desirable frequentist properties (e.g., determine if they are unbiased) but these properties are not essential to the Bayesian approach. The increasing emphasis on computation in Bayesian work has meant that we often display graphical/numerical summaries of the posterior distribution rather than relying on a particular loss function to define a particular point estimate.

### 3.2 Interval estimation

The posterior distribution allows us to identify intervals that contain  $\mu$  with any specified probability. These are called posterior intervals or credible sets. In our example,  $(\mu_n \pm 1.96\sqrt{V_n})$  is a 95% central posterior interval. Since  $\mu$  is viewed as a random variable under the Bayesian paradigm, it is permissible to say that for any given data set,  $\mu$  lies in the stated interval with 95% probability. Recall that such a statement is not possible using the frequentist definition of a confidence interval. The Bayesian approach makes explicit use of probability theory to obtain probability statements about the unknown parameter given the single sample at hand, whereas the frequentist approach makes probability statements about the performance of procedures in repeated samples.



**Figure 2.** Bayesian posterior distribution for  $\mu$ . The prior distribution is assumed to be a normal distribution with mean 110 and standard deviation 10. The data are 10 observations from a normal population with  $\bar{y} = 99.45$  and  $\sigma = 2$  (this is sample 1 that was used to generate one of the confidence intervals in Figure 1). The posterior distribution is normal with mean  $\mu_n = 99.49$  and standard deviation  $\sqrt{V_n} = 0.631$ . The shaded region is the central 95% region; it is quite similar to the confidence interval for sample 1 in Figure 1 because the prior distribution is vague.

Figure 2 shows the posterior distribution in one case with a central 95% interval indicated.

### 3.3 Significance testing

In our simple setting, a decision regarding whether  $\mu < 100$  (the weight of our bar is less than advertised) or  $\mu \geq 100$  could be made by computing the posterior probability that  $\mu \geq 100$ . This is easily done using a posterior distribution like the one pictured in Figure 2. Note that this is a probability that addresses directly the significance testing question about the unknown  $\mu$ . It differs dramatically in interpretation from the  $p$ -value, which is a probability concerned with repeated samples under a fixed, null value for the parameter. A more formal procedure exists for testing one hypothesis versus another in the Bayesian paradigm, known as the Bayes factor. Unfortunately, it will take us reasonably far afield to go into detail, so we do not discuss Bayes factors here.

## 4. Prior distributions

The Bayesian perspective avoids some of the conceptual difficulty associated with interpreting confidence intervals and  $p$ -values. Unfortunately, the "price" that is required to get these benefits,

namely the specification of a prior distribution for the unknown parameter, is one that many people are unwilling to pay. It can be disturbing that individuals with different prior distributions will obtain different answers, but we should remember that this occurs every day when individuals with different information make different decisions.

We now consider some statistical issues related to specifying the prior distribution. One key point is that in large samples the prior distribution becomes irrelevant. In our simple normal distribution example as we let  $n \rightarrow \infty$  for any prior distribution (i.e., any choice of  $\mu_0$  and  $\tau^2$ ) of the posterior distribution, the asymptotic behavior of the posterior distribution doesn't depend on the parameters of the prior distribution. Qualitatively, if we weigh an object hundreds or thousands of times then we would likely disregard our prior information in favor of the data from the scale. In the limit the posterior distribution would behave as if there were no prior information,  $\mu | y_1, \dots, y_n \sim N(\bar{y}, \sigma^2/n)$ . Note that this limiting result is similar to the usual sampling distribution result except of course that here  $\mu$  is random and  $\bar{y}$  is fixed. In fact, the result suggests that although Bayesian and frequentist conclusions may differ in finite samples, they will tend to agree asymptotically. This result holds more generally as well. Of course, data analysts can't generally rely on asymptotics, so we now briefly describe several approaches to developing prior distributions.

The Bayesian approach requires our prior distribution be an honest assessment of our prior beliefs about model parameters. Although people are hesitant to supply such subjective prior distributions it is often the case that some prior information is available. In our scale example if we know what kind of object is to be weighed, then we could likely supply a range of plausible weights. If it is possible to specify a prior distribution, then the Bayesian paradigm provides the natural way (in fact, the only reasonable way) to update our prior beliefs given new data. Tools for helping researchers develop prior distributions is one area of Bayesian research.

Often the choice of a prior distribution is made easier, although possibly less honest, by the existence of conjugate families, families of prior distributions that combine with a given data distribution to produce posterior distributions in the same family. The normal prior distribution in our example is a conjugate prior distribution for the normal data distribution, they combine to produce a normal posterior distribution. Conjugate prior distributions are convenient to use because they make calculations easy and, because they are typically well studied, they can be easily interpreted. Conjugate prior distributions are

capable of supporting a variety of prior opinions (e.g., by making difference choices of  $\mu_0, \tau^2$ ) but not all. Choosing specific values of  $\mu_0$  and  $\tau^2$  remains a bit of a stumbling block for some, so that automatic methods have occasionally been proposed. Empirical Bayes methods use the data to help choose the parameters of the prior distribution. Empirical Bayes techniques are usually applied in more complex models (e.g., analyses that incorporate random effects); we do not discuss such techniques any further in this article.

The desire to avoid using subjective prior information and/or arbitrary distributional forms has led to a great deal of work using vague or "non-informative" prior distributions.

Formally a vague prior distribution is one that assigns roughly equal probability to a wide range of possible values. In our normal example, the contribution of the prior distribution depends on its precision compared to the precision of the data distribution, choosing  $\tau^2 = 100^2$  (a very flat normal prior distribution) when  $\sigma^2 = 1^2$  would be considered a vague prior distribution. A vague prior distribution will not have a strong influence on the ultimate form of the posterior distribution. In the limit, vague prior distributions may become so vague as to no longer be proper distributions (they may not integrate to one!). It is "legal" to use improper prior distributions as long as we verify mathematically that the resulting posterior distribution is proper. Improper prior distributions are popular because sometimes they appear to be "non-informative" in that they reproduce classical frequentist results. It is probably best to think of improper prior distributions as approximations to real prior distributions.

If the improper prior distribution leads to a proper posterior distribution and sufficiently accurate conclusions then we might accept the resulting analysis. If not, then we would need to think harder about an appropriate prior distribution.

## ■ 5. Why use Bayesian methods?

The relatively simple example we have discussed makes it difficult to answer this question. It is tempting to conclude, since the posterior distribution of  $\mu$  in our example resembles the frequentist sampling distribution when we have large samples or vague priors, that the Bayesian analysis is only useful in cases where the prior information is strong. The argument might then continue, since data analysis problems for which there is strong prior information don't come along terribly often, why be Bayesian? I think there are several problems with this argument. First, the Bayesian approach provides a coherent method for

analyzing data with natural probability-based interpretation of the results, e.g., for interval estimates. Second, it is possible to draw valid Bayesian inferences in finite samples (although the computation might be a bit harder), whereas except for special cases frequentist tests and confidence intervals rely on asymptotic (large-sample) results. Finally, in many multiparameter problems, prior information is not as rare as one might think. For example, prior information may suggest to us that a set of parameters can be treated as a sample from a common population (e.g., a random effects model as described in the next paragraph).

More convincing motivation to use Bayesian methods can be found in models that are more sophisticated than our single normal sample. These sophisticated models are not necessarily terribly complex and are often quite realistic for solving scientific problems. We consider one example, the use of the mixed linear model in animal breeding applications. The standard model used in animal breeding includes a number of linear regression parameters relating animal characteristics to the outcome of interest, plus one parameter for each animal that describes the animal's unobservable genetic "breeding value." Studies with thousands of animals therefore include thousands of parameters. These animal breeding values are usually assumed to follow a normal distribution with known non-diagonal covariance matrix (i.e., they are treated as correlated random effects). To a Bayesian the normal distribution of these random effects is just a prior distribution for those parameters. The Bayesian paradigm can be used to draw conclusions about the random effects (e.g., about which animals have the largest breeding values). Conceptually we just apply Bayes' rule; in practice this may require sophisticated computing techniques, like Gibbs sampling or other Markov chain Monte Carlo algorithms.

As mentioned at the start, I find the Bayesian approach to data analysis quite attractive. For me, there are two compelling reasons. First, probability is the language that statisticians use to describe uncertainty and it seems natural to me to use this language in describing our state of knowledge about the value of unknown parameters. Second, it is generally quite obvious how to proceed in the face of added complexity (missing data, additional hypothesized structure in a model) or new data.

## ■ 6. Conclusions

The frequentist approach is generally still the dominant one in most, but not all, graduate statistics programs. It is not the intent of this article to convince anyone that Bayesian approaches are

always better than frequentist approaches. Instead, I have tried to introduce the Bayesian approach and demonstrate both similarities to the frequentist approach (use of parametric data distributions dependent on unknown parameters) and differences from the frequentist approach (use of probability distributions for unknowns).

This article is barely an introduction for interested readers. The Bayesian literature is vast and growing. The recent texts by Gelman et al. (1995) and Carlin and Louis (1996) contain a number of examples that better illustrate the difference between Bayesian and frequentist analyses. Moreover, we have not addressed at all those statisticians who attempt to draw inferences directly from the likelihood function (data distribution) without relying on either repeated sampling or prior information, nor have we considered nonparametric approaches to data analysis. Efron (1986) and Lindley (1990) provide excellent discussions comparing Bayesian and non-Bayesian approaches. Even better, both of these papers are discussed by a number of statisticians from a variety of Bayesian and non-Bayesian perspectives, providing an excellent survey of the many opinions concerning foundational issues in statistical inference.

## ■ Acknowledgements

I received comments on an earlier draft of this paper from Brad Carlin, Herman Chernoff, Herbert A. David, Dominic Dousa, Andrew Gelman, Mark Glickman, Deanne Reber, Don Rubin, Bob Stephenson, and Alan Zaslavsky. Their contributions are greatly appreciated.

## ■ Bibliography

- Carlin, B. P., and Louis, T. A. (1996), *Bayes and Empirical Bayesian Methods for Data Analysis*, London: Chapman and Hall.
- Efron, B. (1986), "Why Isn't Everyone a Bayesian?," *The American Statistician*, 40, 1-11 (including discussion).
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995), *Bayesian Data Analysis*, London: Chapman and Hall.
- Lindley, D. V. (1990), "The 1988 Wald Memorial Lectures: The Present Position of Bayesian Statistics," *Statistical Science*, 5, 44-89 (including discussion).
- Moore, D. S. (1995), *The Basic Practice of Statistics*, New York: W. H. Freeman and Co.

115

## Student Paper Competition

The Statistical Computing Section of the ASA is again sponsoring a Student Paper Session at the Joint Statistical Meetings in 1999. The topic of the session is Statistical Computing. Students are encouraged to submit a paper in this area, which might be original methodological research, some novel application, or any other suitable contribution (for example, a software related project). Four students will be selected to participate in this session, which will include a discussant nominated by the selection committee. Students at all levels (Undergraduate, Masters, and Ph.D.) are encouraged to participate. To be eligible, an applicant must be a registered student in the fall of 1998. The applicant must be the first author of the paper. The Section will pay all registration fees for the selected winners, as well as a substantial allowance for transportation to the meetings and housing (which in most cases covers these expenses completely). To be considered for selection in the session, students must submit an abstract, a six-page manuscript, a resume, and a letter of recommendation from a mentor familiar with their work. The manuscript should be single-spaced in a 10 point font with one inch margins (this is consistent with ASA's Proceedings guidelines.) All figures, tables and references should be included in the six-page limit. In the case of joint authorships, the mentor should indicate what fraction of the contribution is attributable to the applicant. All application materials **must be received** by 5:00 P.M. PST, Friday January 8, 1999 at the address below. They will be reviewed by the Statistical Computing Section Student Paper Competition Award committee. Selection will be based on a variety of criteria at the discretion of the selection committee, and will include novelty and significance of contribution, amongst others. Award announcements will be made in late January, 1999. The decision of the selection committee will be final. Applicants not selected for inclusion in the Session may submit their abstract and a registration fee to ASA by February 1, 1999 if they plan to attend the Joint Meetings. Those abstracts must be submitted following the ASA abstract submission instructions described in *Amstat News*. Students selected for inclusion in the session will receive further information about abstract submission and fee waivers from the award committee.

Inquires and application materials should be E-mailed or mailed to:

Lionel Galway, RAND Corporation, 1700 Main St., Santa Monica, CA 90407; galway@rand.org.

All electronic submissions of papers should be in Postscript.