203

SOME BAYESIAN STATISTICAL TECHNIQUES USEFUL IN ESTIMATING FREQUENCY AND DENSITY

Some Bayesian Statistical Techniques Useful in Estimating Frequency and Density

by Douglas H. Johnson

Some Bayesian Statistical Techniques Useful in Estimating Frequency and Density

by

Douglas H. Johnson

U.S. Fish and Wildlife Service
Northern Prairie Wildlife Research Center
P.O. Box 1747, Jamestown, North Dakota   58401

Abstract

This paper presents some elementary applications of Bayesian statistics to problems faced by wildlife biologists. Bayesian confidence limits for frequency of occurrence are shown to be generally superior to classical confidence limits. Population density can be estimated from frequency data if the species is sparsely distributed relative to the size of the sample plot. For other situations, limits are developed based on the normal distribution and prior knowledge that the density is non-negative, which insures that the lower confidence limit is non-negative. Conditions are described under which Bayesian confidence limits are superior to those calculated with classical methods; examples are also given on how prior knowledge of the density can be used to sharpen inferences drawn from a new sample.

Biologists often require estimates of distribution and abundance for a species of animal or plant. Many approaches, such as mark-recapture techniques, complete censuses, counts along prescribed transects, and animal signs, have been used to estimate frequency of occurrence or density (e.g., Overton 1969). These methods provide an index to abundance which may allow comparisons between years or between areas, but some do not permit valid estimates of the absolute population size. Variance estimates, which are necessary to assess the precision of a quantity, are typically not available. Because time and expense limit the amount of data that can be collected, the investigator generally desires estimators with as much precision as possible.

This paper will explore methods of analyzing counts of organisms to obtain accurate estimators of population size and frequency of occurrence, as well as to gauge the precision of these estimators. The techniques described here apply the Bayesian philosophy of statistical inference (e.g., Good 1965, Schmitt 1969, Lindley 1971) to description of frequency and density. Other authors have discussed applications of Bayesian theory to a variety of problems; see DeGroot (1970) for an extensive bibliography. The methods described here will be illustrated by examples of censuses of breeding pairs of birds (Stewart and Kantrud 1972), but the methods are applicable to a broad class of populations. The term "pair" is used for these examples, but in other situations the word "individual" might be substituted.

Stewart and Kantrud (1972), who studied the populations of breeding birds in North Dakota, drew a random sample of land areas (quarter-sections, 160 acres or 64.75 ha), and counted pairs of all avian species breeding on them. They wanted to estimate the statewide breeding population of each species and to measure the precision of those estimates. They assumed that a census of a quarter-section was essentially perfect; no pairs were missed and none counted more than once. Under this assumption, a census of all quarter-sections in the State would provide the exact population total, and no statistics would be required. Because complete censuses were not feasible for such a large area, however, sampling had to be employed.

A Bayesian Approach

We now consider some estimation procedures which employ the Bayesian philosophy of statistics. A distinguishing feature of this philosophy is the concept of a parameter. Whereas in classical thought a parameter is considered to be a fixed constant with unknown value, Bayesian method treats a parameter as a random variable, the distribution of which depends on available knowledge about the

1

parameter. The parameter does not vary, but our degree of belief in its value changes with increasing information, and the probability function reflects the belief. We will use this philosophy in two ways. First, it will be convenient to combine information about the parameter from more than one source. The Bayesian approach provides a rigorous framework in which information about a parameter can be applied either from experiments performed simultaneously or sequentially, or from knowledge available before the experiment. Second, the concept of a confidence interval will be replaced by a more natural statement about the probability that a parameter is contained within an interval. A classical 95% confidence interval (a, b) for a parameter θ consists of two functions of the observations, $a(x_1, x_2, \ldots, x_n)$ and $b(x_1, x_2, \ldots, x_n)$, such that, if the experiment were repeated an infinite number of times under identical conditions, 95% of the intervals obtained would include the true value of the parameter θ. A comparable Bayesian confidence interval is more simply the interval (a, b) such that 95% of the probability content lies in that interval.

To oversimplify the Bayesian method, we will say that it combines information about a parameter θ from two sources. One source is the experiment (or, in the present case, the random sample of counts). Information from the experiment is contained in the probability function of the observations which, considered as a function of θ, is called the "likelihood" function. The other source of information is that which is available before the experiment occurs, and is contained in the "prior" distribution of θ, denoted $p(\theta)$. We combine the two sources of information by multiplying the prior and likelihood functions; the product is termed the "posterior" distribution function. The Bayesian draws his inference from the posterior distribution, but the classical statistician generally uses only the likelihood function. Often our prior knowledge is lacking, so we take $p(\theta)$ to be noninformative. A Bayesian employing a noninformative prior obtains results closely paralleling classical results, although the Bayesian may do so more directly. We will later demonstrate how even a modest amount of prior knowledge can sharpen the inferences drawn.

Application of the Bayesian Method to Frequency of Occurrence

To put this concept in focus, let X be a random variable taking the value 1 if breeding mallards *(Anas platyrhynchos)* are present in a quarter-section selected at random, and 0 if they are absent. Then X is distributed binomially with $\Pr\{X = 1\} = \theta$ and θ, the expected value of X,

represents the frequency of occurrence of mallards per quarter-section.

Useful prior knowledge may be available, either from censuses conducted previously on these quarter-sections or from a knowledge of how well the habitat requirements of the mallard are satisfied on each quarter-section. But, for the moment, assume prior knowledge is lacking; we then have to select a prior distribution for θ which expresses our ignorance.

The Beta family of prior distributions
$$p(\theta) = \text{constant} \quad \theta^\alpha (1-\theta)^\beta$$
with $\alpha > -1$, $\beta > -1$ is appealing because it is of the same form as the likelihood function, so that its use results in a posterior distribution belonging to the same family. In addition, the Beta family is very flexible; by varying α and β the curve can assume a variety of shapes reflecting various degrees of prior belief. The choice of exponents $\alpha = \beta = -\frac{1}{2}$ has been suggested (Jeffreys 1967; Good 1965) on the basis of invariance under transformation of the parameter θ. That is, if we have no prior knowledge of θ, then we are equally ignorant of θ^{-1}, log θ, etc., and the prior should reflect this invariance. The choice of exponents α, β will have little effect on our conclusions as long as they are small in absolute value relative to z and $n - z$, (defined below). The prior distribution with $\alpha = \beta = -\frac{1}{2}$ is graphed in Fig. 1. Note how little it varies over most of the range of θ.

We next take our random sample of censuses on n quarter-sections, yielding $z = \sum x_i$ quarter-sections with mallards present. The average frequency of occurrence is z/n and the likelihood function of θ, given the data, is
$$L(\theta \mid x_1, x_2, \ldots, x_n) = \theta^{\sum x_i}(1-\theta)^{n - \sum x_i}$$
$$= \theta^z(1-\theta)^{n-z}.$$

For example, Stewart and Kantrud (1972) found mallards breeding in 79 of their 130 sample quarter-sections. An estimate of the frequency of occurrence is then given by $79/130 = 0.608$, and the likelihood function is
$$L(\theta \mid x_1, x_2, \ldots, x_n) = \theta^{79} (1-\theta)^{51}.$$

We next combine our prior information with the information from the sample by multiplying the corresponding probability functions, to obtain the posterior distribution of θ, given the observations, with probability density denoted by
$$p(\theta \mid x_1, x_2, \ldots, x_n) = \text{constant}$$
$$\times L(\theta \mid x_1, x_2, \ldots, x_n) \, p(\theta),$$
where the constant is such that
$$\int_0^1 p(\theta \mid x_1, x_2, \ldots, x_n) \, d\theta = 1.$$
The term "posterior" is in reference to the fact that it represents information available "after" the experi-
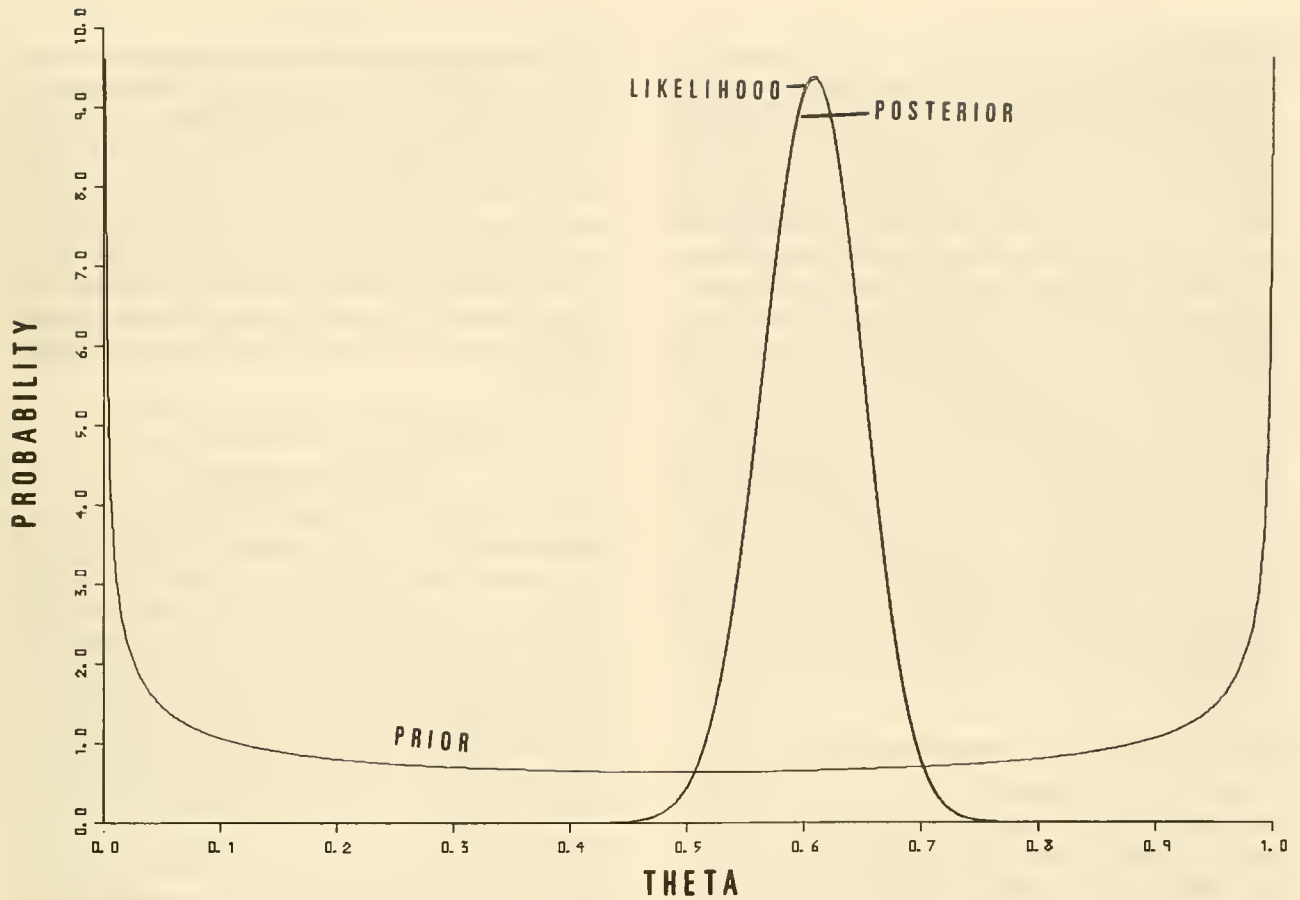
Fig. 1. Prior, likelihood, and posterior functions for frequency of occurrence of mallards. "Probability" means probability density.

ment (or counts). In the present case of a binomial variate,

$$L(\theta \mid x_1, x_2, \ldots, x_n) = \text{constant } \theta^z(1-\theta)^{n-z} ,$$

and

$$p(\theta) = \text{constant } \theta^{-1/2}(1-\theta)^{-1/2}$$

so the posterior distribution becomes

$$p(\theta \mid x_1, x_2, \ldots, x_n) = \text{constant } \theta^{z-1/2}(1-\theta)^{n-z-1/2}$$

or, in the mallard example,

$$p(\theta \mid x_1, x_2, \ldots, x_n) = \text{constant } \theta^{78.5}(1-\theta)^{50.5}.$$

The appendix describes how the constant is evaluated. Because of the large sample size and moderately large mean of the mallard data, the likelihood function dominates the posterior distribution to such an extent that graphs of the functions cannot readily be distinguished (Fig. 1). We can now find points θ_1 and θ_2 so that $\Pr\{\theta_1 < \theta < \theta_2\} = 1 - \alpha$. The interval (θ_1, θ_2) is the Bayesian counterpart of a classical confidence

interval. A two-sided confidence interval is ordinarily chosen so that $\alpha/2$ of the likelihood is eliminated from each tail of the distribution. On the other hand, a Bayesian seeks the shortest interval (θ_1, θ_2) containing $1 - \alpha$ of the probability by selecting θ_1 and θ_2 such that:

(1) no point outside the interval has its probability function greater than any point within, and

$$(2) \quad \int_{\theta_1}^{\theta_2} p(\theta \mid x_1, x_2, \ldots, x_n) \, d\theta = 1 - \alpha .$$

Such an interval is called a Bayesian confidence region or a Highest Posterior Density (HPD) region (Box and Tiao 1965). The 95% HPD region for frequency of occurrence of mallards was (0.523–0.690). The calculation of HPD regions is demonstrated in the appendix.

The usual method for obtaining a confidence interval for a binomial parameter employs the normal approximation, which is fairly good for large samples as long as z (the number of samples in which

the species was present) and $n - z$ (the number of samples in which the species was absent) are not too small. The normal approximation gives limits on the mean as plus or minus two standard errors. Exact limits can be obtained by using the relationship between a binomial sum and the incomplete beta function (Johnson and Kotz 1969), which allows equal-tailed intervals to be obtained from tables of the F distribution (Bliss 1967). For mallards, with a moderate value of z ($= \Sigma x = 79$) as well as a large sample size ($n = 130$), all three procedures yielded very similar confidence limits (Table 1). Consider, however, the burrowing owl *(Speotyto cunicularia)*, which was present on only 2 of the 130 sample quarter-sections. The normal approximation fails here because the number present was small, and the Bayesian confidence interval is shorter than the exact interval (Table 1). Although there are some non-Bayesian methods for reducing the size of confidence regions, e.g., Hudson (1971), they do not follow as naturally as in the Bayesian framework and there is no unified approach to using available prior knowledge.

Application to Estimating Population Density

In certain circumstances the probability limits derived for the frequency of occurrence can also be used to estimate the density (Dice 1952). Suppose the sample areas censused are small relative to the territory size of a species, so that the probability of more than one pair on an area is zero. Then the frequency of occurrence is identical to the average density per unit, and the total number of pairs is equal to the number of sample units containing a pair. Thus the method described above will also give probability limits for the population total.

To equate density to frequency, we have to assume that $\Pr\{$ more than one pair per sample unit $\}$ is negligible. For a territorial species, the validity of this assumption can be assured by reducing the size of the sample unit, although the resultant unit may be impracticably small. The assumption can be exam-

ined if a large number of units are censused by noting whether or not any unit contains more than one pair. If not, then $\Pr\{$ more than one pair per sample unit $\}$ is not likely to be appreciable. Even if more than one pair can occur in a unit, the estimator will be only slightly affected as long as $\Pr\{$ more than one pair per sample unit $\}$ is small and the probability of a large number of pairs on a sample unit is zero. For a colonial species this latter probability is nonzero, so the relationship between frequency of occurrence and density does not hold, and this method should not be attempted.

Modified Normal Limits for Population Density

If the counts are reasonably large and sufficiently numerous, we have some confidence that the central limit theorem applies. That is, we expect the average of the counts to be nearly normally distributed. We will indicate how inferences can be sharpened by using the slight prior knowledge available. Even before sampling takes place, we know that the count on any sample area will be non-negative. Since the normal distribution extends into negative values, it is more general then we need, although this becomes a problem only when the average is small relative to its standard error.

Let θ be the mean and σ^2 the variance of the counts. The usual (Jeffreys 1967) noninformative prior distribution for (θ, σ^2) is

$$p(\theta, \sigma^2) = p(\theta)p(\sigma^2) ,$$

with

$$p(\theta) = \text{constant for all } \theta ,$$

and

$$p(\sigma^2) = \text{constant } \sigma^{-1} \text{ for } \sigma > 0 .$$

But for counts, we know $\theta \geqslant 0$, so we can take

$$p(\theta) = \begin{cases} \text{constant} & \theta \geqslant 0 \\ \\ 0 & \theta < 0 , \end{cases}$$

to place all of the probability in the range where $\theta \geqslant 0$. The likelihood function obtained from the counts is $L(\theta, \sigma^2 \mid x_1, x_2, \ldots, x_n)$

$$= \text{constant } \sigma^{-n} \exp \left\{ -\Sigma(x_i - \theta)^2/2\sigma^2 \right\} .$$

Combining our prior distribution with the likelihood function yields the posterior distribution, given by $p(\theta, \sigma^2 \mid x_1, x_2, \ldots, x_n)$

$$= \begin{cases} \text{constant } \sigma^{-n-1} \exp \left\{ -\Sigma(x_i - \theta)^2/2\sigma^2 \right\} & \theta \geqslant 0 \\ \\ 0 & \theta < 0 . \end{cases}$$

Table 1.—*Confidence limits for frequency of occurrence of breeding pairs of mallards and burrowing owls on 130 sample quarter-sections in North Dakota.*

Method	Mallard	Burrowing owl
Normal approximation to binomial	(.522, .693)	(−.006, .037)
Exact binomial	(.519, .690)	(.002, .054)
Bayesian	(.523, .690)	(.001, .042)

Then the distribution of the mean θ is found by integrating out σ^2, so

$$p(\theta \mid x_1, x_2, \ldots, x_n)$$

$$= \int_0^{\infty} p(\theta, \sigma^2 \mid x_1, x_2, \ldots, x_n) \, d\sigma^2$$

$$= \begin{cases} \text{constant } [1 + n(\theta - \bar{x})^2/(n-1)s^2]^{-(n-1)/2} & \theta \geqslant 0 \\ 0 & \theta < 0 \end{cases}$$

On the positive axis, the distribution function of θ is proportional to Student's t with mean $\bar{x}$ and variance s^2/n, which is nearly normal for moderately large sample size. Hence, the posterior distribution of θ approaches the normal, except that the negative tail is eliminated and the portion of the probability mass from the negative range is redistributed.

For large samples with a moderate coefficient of variation, the HPD regions obtained from the above distribution are similar in numerical value, but not in interpretation, to classical normal confidence limits. If the mean is small relative to the standard error, however, the Bayesian method may give considerably shorter intervals because knowledge that θ is non-negative contributes more information when $\bar{x}$ is small than when $\bar{x}$ is large. For example, the density of the mallard, whose habitat requirements are so general that they are easily met in much of North Dakota, was 1.68 pairs per quarter-section. A standard error of 0.152 resulted in a 95% confidence interval of (1.38, 1.98), which is no different from the HPD interval. The long-billed marsh wren *(Telmatodytes palustris)*, on the other hand, is mostly restricted to semipermanent marshes, so that its distribution is much more localized than that of the mallard. But where it occurs, it is often abundant, so that the variation in the counts is large. The 95% confidence interval for this species, with an average of 0.39 pair per quarter-section and a standard error of 0.223, was (-0.05, 0.84). The HPD region (0.00, 0.76) is appreciably shorter because of the large coefficient of variation. Fig. 2 illustrates for the long-billed marsh wren the likelihood function on which ordinary confidence limits are based and the posterior distribution from which HPD regions are determined.

Blumenthal (1970) developed interval estimators of a normal mean restricted to be non-negative, in the case of known variance.

The Use of Prior Information

Thus far we have restricted our attention to the case in which we lack prior knowledge, except for the fact that counts are non-negative. One of the features of Bayesian statistics is the ease with which prior knowledge can be incorporated into the analysis. In this section we will illustrate this by an example.

The mallard density in the primary waterfowl area of North Dakota for 1957-66 varied from 0.92 to 2.97 pairs per quarter-section (Table 2). How can these values be used in combination with the 1967 counts? One possibility is to use the early data to develop a prior distribution for estimation of the 1967 count. Take the prior distribution for the 1967 density to be normal with mean = 2.11 and variance = 0.365. Combining this prior with the likelihood function for the 1967 data, a normal with mean = 2.45 and variance = 0.044, results in a normal posterior distribution, with mean = (2.45/0.044 + 2.11/0.365) $\div$ (1/0.044 + 1/0.365) = 2.41. The variance is $(1/0.044 + 1/0.365)^{-1} = 0.039$. The details of the calculation are shown in Box and Tiao (1973). Note that the posterior mean is a weighted average of the prior mean and the sample mean, with weights inversely proportional to the corresponding variances. Thus, the posterior mean will shift from the observed mean toward the prior mean; the amount of shift depends upon the relative precision of the two means. In the example, the shift was minor because the variance of the 1967 mean was small relative to the variation among years in the prior data. Nevertheless, some improvement in precision is indicated by the 11% reduction in the posterior variance as compared with the sample variance (0.039 vs. 0.044).

Although it may be contrary to the intuition of biologists to use data from other years to estimate the 1967 count, the theoretical justification is considerable. For example, Lindley (1971) and Box and Tiao (1973) developed Bayesian estimators for a vector of normal means such that information from all means is used in the estimation of each mean. A sampling theory justification was given by James and Stein (1961), who demonstrated the surprising result that, in estimating a vector of normal means, the maximum likelihood estimator (the sample mean) has a larger mean square error than does a weighted average of the sample mean and the overall mean.

Judgment is necessary, of course, in determining whether combining should be done or not. In the mallard example, if water conditions were uniformly excellent in 1957-66, but known to be poor in 1967, we should be reluctant to use the early data to form an opinion on the 1967 count. Efron and Morris (1973) discussed strategies for combining data and proposed a method that allows the data to determine the extent of combination: larger weights are given to means that are "close" (in the likelihood sense) to the observed mean being estimated. So, for example, in estimating the 1967 mallard density, higher weights
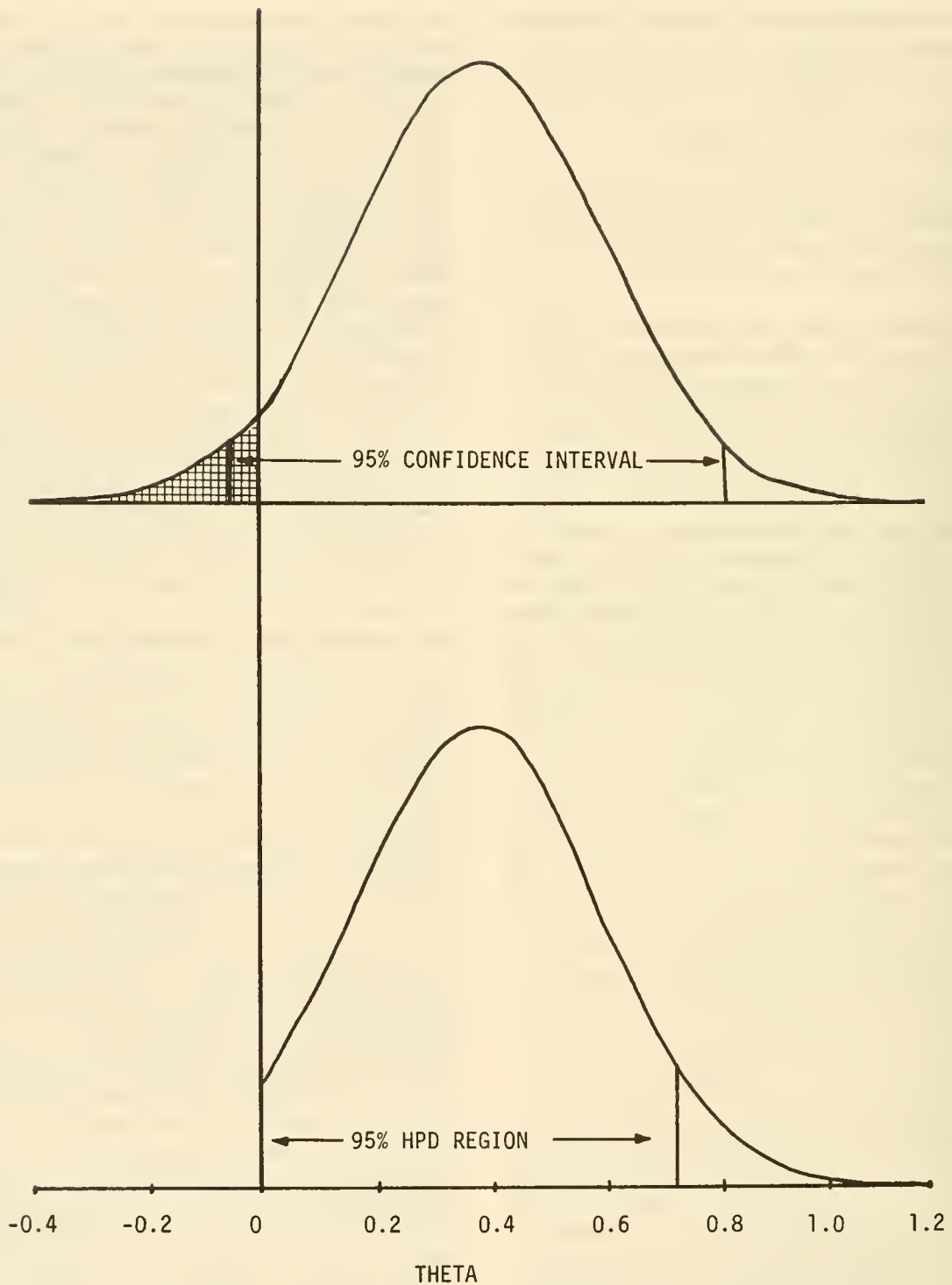
Fig. 2a. (TOP) Normal likelihood function for long-billed marsh wren, showing 95% confidence
limits (-0.05, 0.84). Shading indicates the region to which the prior distribution function assigns
zero probability.

Fig. 2b. (BOTTOM) Posterior probability function for long-billed marsh wren, showing 95% HPD
region (0.00, 0.76).

Table 2.—*Mallard densities (pairs per quarter-section) in primary waterfowl area of North Dakota, 1957-66.*

Year	Density
1957	2.50
1958	2.45
1959	0.92
1960	1.84
1961	1.58
1962	1.68
1963	2.97
1964	2.36
1965	2.23
1966	2.61
Average	2.11
Variance	0.365

would be given to years that had densities closer to the 1967 value. Their procedure is one of a class called "Empirical Bayes" methods, which are intermediate between Bayesian and sampling theory methods. They are similar to Bayesian methods in that a prior distribution is combined with a likelihood to obtain a posterior distribution, but, unlike Bayesian practice, the prior distribution is estimated from the data, rather than assigned subjectively. See Krutchkoff (1972) for a brief introduction to Empirical Bayes methods.

Most of the literature in this area is new and detailed practical procedures have not yet been fully worked out. Applications of Empirical Bayes procedures will no doubt increase in the near future, but it is beyond the scope of this article to provide a comprehensive review of the subject.

Acknowledgments

References

Beyer, W. H. (ed.). 1968. Handbook of tables for probability and statistics, 2nd ed. The Chemical Rubber Co., Cleveland. 642 pp.

Bliss, C. I. 1967. Statistics in biology: statistical methods for research in the natural sciences, vol. 1. McGraw-Hill, New York. 558 pp.

Blumenthal, S. 1970. Interval estimation of the normal mean subject to restrictions, when the variance is known. Nav. Res. Logist. Q. 17(4):485-505.

Box, G. E. P., and G. C. Tiao. 1965. Multiparameter problems from a Bayesian point of view. Annu. Math. Stat. 36(5):1468-1482.

Box, G. E. P., and G. C. Tiao. 1973. Bayesian inference in statistical analysis. Addison-Wesley, Reading, Mass. 588 pp.

Brown, K. M. 1973. Computer oriented algorithms for solving systems of simultaneous nonlinear algebraic equations. Pages 281-348 *in* G. D. Byrne and C. A. Hall, eds. Numerical solution of systems of nonlinear algebraic equations. Academic Press, Inc., New York.

DeGroot, M. H. 1970. Optimal statistical decisions. McGraw-Hill, New York. 489 pp.

Dice, L. R. 1952. Natural communities. Univ. of Michigan Press, Ann Arbor. 547 pp.

Efron, B., and C. Morris. 1973. Combining possibly related estimation problems (with discussion). J. Royal Stat. Soc. B 35(3):379-421.

Feller, W. 1966. An introduction to probability theory and its applications, vol. II. John Wiley and Sons, New York. 626 pp.

Good, I. J. 1965. The estimation of probabilities: an essay on modern Bayesian methods. MIT Press, Cambridge. 109 pp.

Hudson, D. J. 1971. Interval estimation from the likelihood function. J. Royal Stat. Soc. B 33(2):256-262.

International Business Machines Corporation. 1970. System/360 Scientific Subroutine Package Version III programmer's manual, 5th ed. IBM Corp., White Plains, N.Y. 454 pp.

James, W., and C. Stein. 1961. Estimation with quadratic loss. Proc. 4th Berkeley Symposium Math. Stat. and Probability. Univ. California Press 1:361-379.

Jeffreys, H. 1967. Theory of probability, 3rd ed. (corrected). Clarendon Press, Oxford, England. 459 pp.

Johnson, N. L., and S. Kotz. 1969. Discrete distributions. Houghton Mifflin, Boston. 328 pp.

Krutchkoff, R. G. 1972. Empirical Bayes estimation. Am. Statistician 26(5):14-16.

Lindley, D. V. 1971. Bayesian statistics, a review. Soc. Ind. Appl. Math., Philadelphia. 83 pp.

Overton, W. S. 1969. Estimating the numbers of animals in wildlife populations. Pages 403-455, *in* R. H. Giles, Jr., ed. Wildlife management techniques. 3rd ed. (revised). The Wildlife Society, Washington, D.C.

Pearson, K. (ed.) 1968. Tables of the incomplete beta-function, 2nd ed. Cambridge Univ. Press, New York. 505 pp.

Schmitt, S. A. 1969. Measuring uncertainty: an elementary introduction to Bayesian statistics. Addison-Wesley, Reading, Mass. 400 pp.

Stewart, R. E., and H. A. Kantrud. 1972. Population estimates of breeding birds in North Dakota. Auk 89(4):766-788.

Appendix.
Calculation of Highest Probability Density Regions

For a posterior distribution $p(\theta \mid x_1, x_2, \ldots, x_n)$, a $(1 - \alpha)$ Highest Posterior Density (HPD) region (θ_1, θ_2) must satisfy the following two conditions:

If $\theta \in (\theta_1, \theta_2)$ and $\theta^* \notin (\theta_1, \theta_2)$, then

$$p(\theta \mid x_1, x_2, \ldots, x_n) \geq p(\theta^* \mid x_1, x_2, \ldots, x_n) \ , \ (1)$$

and

$$\int_{\theta_1}^{\theta_2} p(\theta \mid x_1, x_2, \ldots, x_n) \ d\theta = 1 - \alpha \ . \quad (2)$$

We will develop HPD regions for a binomial parameter and for a normal mean restricted to non-negative values.

HPD Regions for The Binomial Parameter

If $\{X_i\}$ are distributed binomially with parameter θ, then the posterior distribution of θ corresponding to the noninformative prior specified earlier is given by

$$p(\theta \mid x_1, x_2, \ldots, x_n) = \text{constant} \ \theta^{z-1/2}(1-\theta)^{n-z-1/2}, (3)$$

where $z = \sum x_i$. The constant term is such that

$$1 = \int_0^1 p(\theta \mid x_1, x_2, \ldots, x_n) \ d\theta$$

$$= \text{constant} \int_0^1 \theta^{z-1/2} (1 - \theta)^{n-z-1/2} \ d\theta$$

$$= \text{constant} \ B(z+\tfrac{1}{2}, \ n-z+\tfrac{1}{2}) \ ,$$

where $B(a, b)$ is the beta function (e.g., Feller 1966). Thus, the required constant is the reciprocal of $B(z+\tfrac{1}{2}, n-z+\tfrac{1}{2})$.

A $(1-\alpha)$ HPD interval is given by (θ_1, θ_2) satisfying equations (1) and (2) above. For a unimodal distribution the first requirement is equivalent to

$$p(\theta_1 \mid x_1, x_2, \ldots, x_n) = p(\theta_2 \mid x_1, x_2, \ldots, x_n) \ ,$$

which, using (3) and noting that the constant occurs on both sides and thus cancels, becomes

$$\theta_1^{z-1/2} (1 - \theta_1)^{n-z-1/2} = \theta_2^{z-1/2} (1 - \theta_2)^{n-z-1/2},$$

which has the same solution as does

$$f_1(\theta_1, \theta_2) = 0 \ , \quad (4)$$

where

$$f_1(\theta_1, \theta_2) = (z-\tfrac{1}{2}) (\log \theta_1 - \log \theta_2)$$
$$+ (n-z-\tfrac{1}{2}) (\log(1-\theta_1) - \log (1-\theta_2)).$$

It follows from (2) that

$$\frac{1}{B(z+\tfrac{1}{2}, \ n-z+\tfrac{1}{2})} \int_{\theta_1}^{\theta_2} \theta^{z-1/2}(1-\theta)^{n-z-1/2} d\theta = 1 - \alpha \ ,$$

or

$$\frac{1}{B(z+\tfrac{1}{2}, \ n-z+\tfrac{1}{2})} \int_0^{\theta_2} \theta^{z-1/2}(1-\theta)^{n-z-1/2} d\theta$$

$$- \frac{1}{B(z+\tfrac{1}{2}, \ n-z+\tfrac{1}{2})} \int_0^{\theta_1} \theta^{z-1/2}(1-\theta)^{n-z-1/2} d\theta = 1 - \alpha$$

or

$$f_2(\theta_1, \theta_2) = 0 \quad (5)$$

where

$$f_2(\theta_1, \theta_2) = I_{\theta_2} (z+\tfrac{1}{2}, \ n-z+\tfrac{1}{2}) - I_{\theta_1} (z+\tfrac{1}{2}, \ n-z+\tfrac{1}{2})$$
$$- (1 - \alpha) \ ,$$

and $I_w(a, b)$ is the incomplete beta integral with parameters a and b evaluated at the point w (Beyer 1968). This function has been tabled (Pearson 1968), and computer routines are available for its calculation (International Business Machines Corporation 1970).

We thus have two equations, (4) and (5), that θ_1 and θ_2 must simultaneously satisfy. Because both equations are nonlinear, iterative techniques are necessary. Two such approaches have been successfully used, although either can fail under certain extreme conditions.

The first and more direct approach is to use a computer algorithm designed to solve a system of nonlinear equations. In the present instance the system consists of two equations, (4) and (5). Not all computer facilities have such routines, but Brown (1973) presented an algorithm and Fortran program which can be used for this purpose.

The second approach requires only computer routines that are commonly available. The equations (4) and (5) can be solved by minimizing the quantity

$$\mid f_1(\theta_1, \theta_2) \mid + \mid f_2(\theta_1, \theta_2) \mid$$

with respect to θ_1 and θ_2. The minimum is zero, and is attained for values of θ_1 and θ_2 such that equations (4) and (5) are satisfied. Numerous algorithms have been developed to find the minimum of a function. Most computing centers have one or more available.

HPD Regions for the Normal Mean

The approximate posterior distribution of the mean of non-negative random variables $\{X_i\}$ is asymptotically

$$p(\theta \mid x_1, x_2, \ldots, x_n) = \begin{cases} \text{constant } s^{-1} \times \\ \exp\left[-n(\theta-\bar{x})^2/2s^2\right] & \text{if } \theta \geq 0 \\ 0 & \text{if } \theta < 0. \end{cases}$$

The value of the constant will depend upon the excluded portion of the normal curve (the shaded area in Fig. 2a). Let A be this value. Then

$$A = \int_{-\infty}^{0} (2\pi \, s^2/n)^{-1/2} \exp[-n(\theta-\bar{x})^2/2s^2] \, d\theta$$

$$= \int_{-\infty}^{-\sqrt{n}\bar{x}/s} (2\pi)^{-1/2} \exp\left[-\tfrac{1}{2}W^2\right] dW$$

$$= Pr\left\{W < -\sqrt{n}\bar{x}/s\right\} \text{ where } W \text{ is a unit normal deviate}$$

$$= \Phi(-\sqrt{n}\bar{x}/s) ,$$

$\Phi(\cdot)$ being the cumulative normal distribution function. Because

$$\int_{-\infty}^{\infty} p(\theta \mid x_1, x_2, \ldots, x_n) \, d\theta = 1 ,$$

the distribution for positive values of θ is $1/(1-A)$ times that of a normal variate with mean $\bar{x}$ and variance s^2/n, i.e.,

$$p(\theta \mid x_1, x_2, \ldots, x_n)$$

$$= (1-A)^{-1} (2\pi s^2/n)^{-1/2} \exp\left[-n(\theta-\bar{x})^2/2s^2\right] \text{ for } \theta \geq 0.$$

As with a binomial variable, we want values θ_1 and θ_2 which satisfy equations (1) and (2). The form that the HPD interval takes will depend on whether or not $A < \alpha/2$ holds.

Case I: $A < \alpha/2$. Here the interval will be symmetric about $\bar{x}$. Equation (2) implies

$$1 - \alpha = \int_{\theta_1}^{\theta_2} (1-A)^{-1} (2\pi s^2/n)^{-1/2} \times \exp\left[-n(\theta-\bar{x})^2/2s^2\right] d\theta ,$$

or

$$(1-\alpha)(1-A) = \int_{\theta_1}^{\theta_2} (2\pi s^2/n)^{-1/2} \times \exp\left[-n(\theta-\bar{x})^2/2s^2\right] d\theta$$

$$= \int_{\sqrt{n}(\theta_1-\bar{x})/s}^{\sqrt{n}(\theta_2-\bar{x})/s} (2\pi)^{-1/2} \exp\left[-\tfrac{1}{2}W^2\right] dW$$

$$= \Phi[\sqrt{n}(\theta_2-\bar{x})/s] - \Phi[\sqrt{n}(\theta_1-\bar{x})/s] .$$

Equation (1) requires

$$p(\theta_1 \mid x_1, x_2, \ldots, x_n) = p(\theta_2 \mid x_1, x_2, \ldots, x_n)$$

which implies symmetry about $\bar{x}$, so (θ_1, θ_2) simply corresponds to an ordinary confidence interval with a confidence coefficient of $(1-\alpha)(1-A)$ rather than $(1-\alpha)$.

For the mallard, having a mean of 1.68 and a standard error of 0.152, the calculated value of $A = \Phi(-11.05)$ is so near zero that it can be ignored; thus, the HPD region coincides with the ordinary confidence interval.

Case II: $A \geq \alpha/2$. Here Equation (1) insures that the lower limit will be zero, so we need to exclude α of the probability content from the upper range. Equation (2) becomes simply

$$1 - \alpha = \int_{0}^{\theta_2} (1 - A)^{-1} (2\pi s^2/n)^{-1/2} \times$$
$$\exp \left[-n(\theta-\bar{x})^2/2s^2\right] d\theta,$$

or

$$\alpha = \int_{\theta_2}^{\infty} (1 - A)^{-1} (2\pi s^2/n)^{-1/2} \times$$
$$\exp \left[-n(\theta-\bar{x})^2/2s^2\right] d\theta,$$

or

$$\alpha(1 - A) = \int_{\theta_2}^{\infty} (2\pi s^2/n)^{-1/2} \exp \left[-n(\theta-\bar{x})^2/2s^2\right] d\theta$$

$$= \int_{\sqrt{n}(\theta_2-\bar{x})/s}^{\infty} (2\pi)^{-1/2} \exp \left[-\tfrac{1}{2} W^2\right] dW$$

$$= 1 - \Phi \left[\sqrt{n}(\theta_2-\bar{x})/s\right].$$

So

$$\sqrt{n}(\theta_2-\bar{x})/s = \Phi^{-1} \left[1 - \alpha(1 - A)\right],$$

or

$$\theta_2 = \bar{x} + (s/\sqrt{n}) \, \Phi^{-1} \left[1 - \alpha(1 - A)\right],$$

and

$$\theta_1 = 0 .$$

Case II is exemplified by the long-billed marsh wren. A mean of 0.39 and standard error of 0.223 resulted in $A = \Phi(-0.39/0.223) = 0.040 > \alpha/2$. Hence, $\theta_1 = 0.0$ and $\theta_2 = 0.39 + 0.223 \, \Phi^{-1} (0.952) = 0.76$.

Calculating HPD regions for the normal mean requires simply a routine for the normal probability function (to determine A) and one for the inverse normal probability function (to determine the confidence limits).

GPO 837-724

Mailing lists are computerized. Please return address label with change of address.