The Annals of Statistics
1979, Vol. 7, No. 2, 400–413

UPPER AND LOWER PROBABILITY INFERENCES FOR THE LOGISTIC FUNCTION

BY SANDRA A. WEST

Worcester Polytechnic Institute

A general system of inference which leads to upper and lower posterior distributions based on sample data has been proposed by Dempster (1967). This general theory of inference is applied to the two-parameter logistic function, given the data from independent binomial populations. Inferences are developed for fixed regions about the two parameters and about interesting combinations of these parameters. The resulting upper and lower probabilities are generated by a random polygonal-type region, or more exactly by specific extreme points of this region. For these extreme points, the exact marginal and joint distributions are derived; approximate distributions are also derived.

1. Introduction. Dempster's general theory (1967, 1968) is applied in this paper to derive methods of inference for the two parameter logistic function. The data consist of independent samples of successes and failures from each of t binomial populations. For population i, the probability of success, p_i, is given by

$$p_i = [1 + \exp(-\alpha - \beta T_i)]^{-1}, \qquad i = 1, 2, \cdots, t,$$

where α and β are unknown parameters and T_i is a numerical value, say "level," associated with the population. Such a model arises in bio-assay, lifetesting, and the study of psychosensory response systems, more generally, in analyzing experiments yielding quantal responses and so is broadly applicable.

Dempster's theory of inference, a generalization of Bayesian inference, allows prior information to be included, if available, but it is not essential as in standard Bayesian analysis. Dempster's system yields upper and lower probabilities rather than a single posterior probability, here for α–β event sets, for example, those generated by $\text{logit}(p) = \log p(1-p)^{-1} = \alpha + \beta T$ or $-\alpha/\beta$, the value of T for which $p = .5$.

Dempster's theory is outlined in Section 2. The upper and lower probabilities for α–β event sets are found by integrating over these sets, appropriate densities derived in Section 3. The probabilities themselves for certain particular cases are determined in Section 4. Because these probabilities will be somewhat difficult to evaluate in practice, approximations to them are derived in Section 5. The accuracy of these approximations is investigated in Section 6 using Monte Carlo methods. Section 7 contains some general comments on the upper and lower probability inferences.

Received January 1975; revised October 1977.

AMS 1970 *subject classifications.* Primary 62A99.

Key words and phrases. Upper and lower probability inferences, two parameter logistic function, upper and lower posterior probabilities, marginal and joint distributions of extreme points of a random polygonal region, exact and approximate distributions.

2. The model and method of inference. Before developing the new inference model a brief description of the general inference system is given, and the model for the binomial parameter p is presented. For a more detailed and rigorous explanation, the reader is referred to Dempster (1967, 1968).

The basic objects in Dempster's theory are a pair of spaces, say U and S. The first space U carries an ordinary probability measure μ but the events of interest are identified with subsets of S. In application the space U generally represents an underlying and unobservable population being sampled, and the randomness associated with a sample individual drawn from U is represented by the probability measure μ. S is generally the product space of an observation space Y and a parametric space Ω. A bridge is provided between U and S by a mathematical transformation Γ, which maps each point of U onto a subset of S. Given (U, S, μ, Γ), upper and lower probabilities are determined for subsets of S.

The probability distribution over U induces a distribution of random subsets of S. The subset Γu may be viewed as a random set in S generated by the random point u in U. Subject to the condition that Γu is nonempty, Γu generates the desired inferences. Letting $V = \Gamma u$ and F denote any fixed subset of S, then conditional on $V \neq \phi$ the upper and lower probabilities of F are defined as

$$P^*(F) = P\{(F \cap V) \neq \phi\}/P\{V \neq \phi\} \tag{2.1}$$

$$P_*(F) = P\{V \subset F, V \neq \phi\}/P\{V \neq \phi\}, \tag{2.2}$$

assuming $P\{V \neq \phi\} > 0$.

A probability model like (U, S, μ, Γ) may be modified into other models of the same general type by conditioning on subsets of S. A simple but important class of models can be characterized by random intervals on the line. As an illustration and for future reference the model producing upper and lower probabilities for the binomial parameter p is presented.

Let y_i be an observable taking the values 0, 1 with probabilities $(1 - p)$, p respectively. Suppose that underlying each observation y_i is a random variable u_i, which is uniformly distributed on [0, 1] and where

$$\begin{aligned} y_i &= 0 \quad \text{if} \quad p \leqslant u_i \leqslant 1 \\ &= 1 \quad \text{if} \quad 0 \leqslant u_i \leqslant p. \end{aligned} \tag{2.3}$$

That is, a single binomial observable y_i is represented before observation by the model $(U, Y \times P, \mu, \Gamma)$, where $U = \{u_i | 0 \leqslant u_i \leqslant 1\}$, $Y = \{y_i | y_i = 0 \text{ or } y_i = 1\}$, $P = \{p | 0 \leqslant p \leqslant 1\}$, μ is the uniform distribution over U, and

$$\Gamma u_i = \{(y_i, p) | y_i = 0 \quad \text{and} \quad 0 \leqslant p \leqslant u_i \qquad \text{or} \quad y_i = 1 \quad \text{and} \quad u_i \leqslant p \leqslant 1\}.$$

An observation of $y_i = 0$ implies that $0 \leqslant p \leqslant u_i$, thus creating a random closed interval $[0, u_i]$ of p values governed by the uniform distribution of u_i. Similarly, an observation of $y_i = 1$ leads to the random closed interval $[u_i, 1]$.

Suppose n independent observations are made yielding $y_1, y_2, \cdots, y_n$, where $\Sigma_{i=1}^n y_i = r$. The rule for combining independent sources of information, as given in Dempster (1967, 1968), implies that the combination will again be based on a random interval, which is the intersection of the individual random intervals arising from the observations y_i. Specifically, the random interval which contains all possible p values consistent with the data is

$$u_{(r)} \leqslant p \leqslant u_{(r+1)}, \tag{2.4}$$

where $u_{(1)} \leqslant u_{(2)} \leqslant \cdots u_{(n)}$ denote the ordered random variables $u_1, u_2, \cdots, u_n$ and where $u_{(0)} = 0$ and $u_{(n+1)} = 1$.

The random interval in (2.4), which is nonempty with probability one, generates the desired upper and lower probabilities, which should be interpreted as conditional probabilities given the data. This is just one of many applications of the theory of random closed intervals (Dempster 1968b, West 1977a). The simplicity of random closed intervals on the line is quickly lost when one considers random sets in a plane.

The specific model developed in this paper lies between these two classes. The upper and lower probability inferences are indeed generated by a random region in the plane. However, this two dimensional region is induced by a number of independent random intervals on the line. For the rest of this section the new inference model is developed and essential random variables are defined.

Suppose $n_1, n_2, \cdots, n_t$ trials are performed at t levels, say $T_1, T_2, \cdots, T_t$, respectively, where $0 \leqslant T_1 < T_2 \cdots < T_t$. Consider the situation in which the number of successes R_i, at level T_i, has a binomial distribution with the probability p_i for success, where

$$p_i = [1 + \exp(-\alpha - \beta T_i)]^{-1}, \qquad \text{for} \quad i = 1, 2, \cdots, t.$$

Based on the observed values of $R_i = r_i$, $i = 1, 2, \cdots, t$, upper and lower probability inferences for interesting regions about α and β are desired. These probabilities are generated by the random region in the space of the α, β parameters which is induced by the random intervals in the separate binomial parameters p_i. That is, the t independent binomial parameters give rise to t independent random closed intervals

$$U_{(r_i)} \leqslant p_i \leqslant U_{(r_i+1)}, \qquad i = 1, 2, \cdots, t.$$

Using the logit transformation each random interval of p_i values maps into a random strip of (α, β) values

$$X_i \leqslant \alpha + \beta T_i \leqslant Z_i, \qquad i = 1, 2, \cdots, t,$$

where $X_i = \ln U_{(r_i)}(1 - U_{(r_i)})^{-1}$ and $Z_i = \ln U_{(r_i+1)}(1 - U_{(r_i+1)})^{-1}$. (Note that if $r_i = 0$ or n_i the random strips are actually half-planes.) These t random α, β strips must intersect if all the random p intervals are to be consistent with at least one (α, β). The intersection of these random strips is a random region, denoted by V (illustrated in Figure 1). This region may be described as the range of α, β values for which a logit curve can be made to pass through the intervals

$$[U_{(r_1)}, U_{(r_1+1)}], [U_{(r_2)}, U_{(r_2+1)}], \cdots, [U_{(r_t)}, U_{(r_t+1)}],$$

at levels $T_1, T_2, \cdots, T_t$, respectively.

Accordingly, V is the important random subset of (α, β) values, and the conditional distribution of V given that it is nonempty is the key to this paper. Let

$$F = \{(\alpha, \beta) | c \leqslant f(\alpha, \beta) \leqslant d\},$$

where c, d are any real numbers, and f is some monotone function. With probability one,

$$(2.5)\quad (V \cap F) = \phi \Leftrightarrow \sup_{(\alpha\,\beta) \in V} f(\alpha, \beta) < c \quad \text{or} \quad \inf_{(\alpha, \beta) \in V} f(\alpha, \beta) > d$$
$$V \subset F \Leftrightarrow \inf_{(\alpha, \beta) \in V} f(\alpha, \beta) \geqslant c \quad \text{and} \quad \sup_{(\alpha, \beta) \in V} f(\alpha, \beta) \leqslant d,$$

thus from (2.1) and (2.2)

$$(2.6)\quad P^*(F) = 1 - P\{\sup_{(\alpha, \beta) \in V} f(\alpha, \beta) < c\} - P\{\inf_{(\alpha, \beta) \in V} f(\alpha, \beta) > d\}$$

$$(2.7)\quad P_*(F) = P\{\inf_{(\alpha, \beta) \in V} f(\alpha, \beta) \geqslant c \quad \text{and} \quad \sup_{(\alpha, \beta) \in V} f(\alpha, \beta) \leqslant d\}.$$

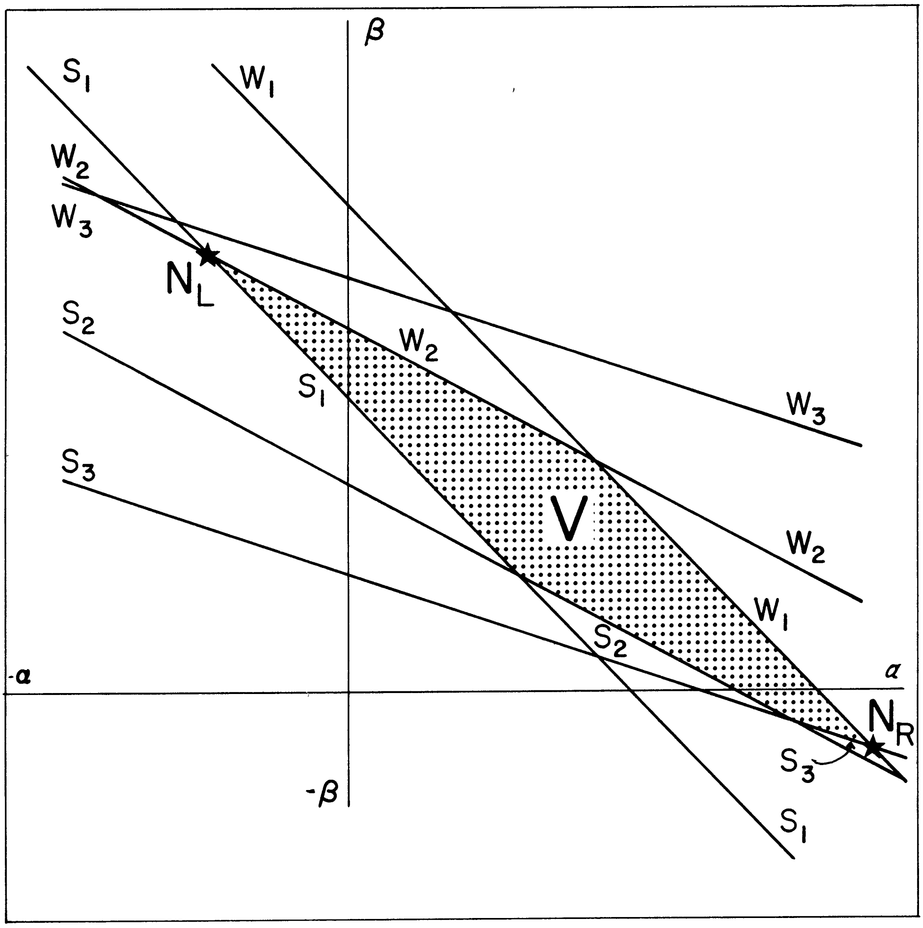

FIG. 1. *The shaded region is a realization of the random region V for the case t = 3. In this instance, the "left- and right-most" points of V, N_L and N_R, are formed by the lines S_1 and W_2 and the lines S_3 and W_1 respectively.*

For simplicity the following notation is now introduced. In the α, β plane,

$$\text{random line} \quad S_i\colon \quad \beta = -\alpha T_i^{-1} + X_i T_i^{-1}$$

$$\text{random line} \quad W_i\colon \quad \beta = -\alpha T_i^{-1} + Z_i T_i^{-1}$$

$$S_i \cap W_j, \quad \text{point of intersection of lines} \quad S_i \quad \text{and} \quad W_j,$$

$$\text{random strip} \quad V_i\colon \quad X_i \leqslant \alpha + \beta T_i \leqslant Z_i$$

for $i = 1, 2, \cdots, t$,

$$V = V_i \cap V_2 \cap V_3 \cdots \cap V_t.$$

Any particular instance of the region V which contains at least one point is a closed polygon. Specific vertices of V will be of special interest in the discussion of certain upper and lower probabilities; two such vertices that generate inferences about each of the parameters will now be defined.

DEFINITION 2.1. The "left and right most" points of V, denoted by $N_L = (\alpha_s, \beta_m)$ and $N_R = (\alpha_m, \beta_s)$ respectively, designate the vertices of V satisfying

$$\alpha_s \leqslant \alpha \leqslant \alpha_m \qquad \text{for all} \quad \alpha \in V,$$

and since all $T_i \geqslant 0$, it follows that

$$\beta_s \leqslant \beta \leqslant \beta_m \qquad \text{for all} \quad \beta \in V.$$

In deriving the densities of these extreme points it is necessary to know the list of possibilities for each vertex. It is easily checked that N_L is formed by the intersection of two lines such as S_i and W_j, and N_R is the intersection of S_j and W_i where $i < j$. That is, for each vertex there are $t(t-1)/2$ mutually exclusive candidates of the form

$$N_L = S_i \cap W_j, \qquad N_R = S_j \cap W_i, \qquad 1 \leqslant i < j \leqslant t. \tag{2.8}$$

Additional vertices needed for inferences pertaining to functions of α and β, such as $\alpha + \beta T$ or $-\alpha/\beta$, are discussed briefly in Section 4 and developed fully in West (1977b).

3. Exact densities of extreme points of V. In this section the conditional densities of N_L and N_R, given V nonempty are derived.

THEOREM 3.1. *Given $R_i = r_i$, $i = 1, 2, \cdots, t$ (where at least one $r_k \neq 0, n_k$, $k = 2, 3, \cdots, t-1$) and denoting the conditional densities of N_L and N_R given V nonempty, by $f_{01}(\alpha, \beta)$ and $f_{10}(\alpha\ \beta)$ respectively, then*

$$f_{01}(\alpha, \beta) = C \cdot L(r_1, r_2 \cdots r_t; \alpha, \beta) \cdot P_{01}(\alpha, \beta) \tag{3.1}$$

$$f_{10}(\alpha, \beta) = C \cdot L(r_1, r_2 \cdots r_t; \alpha, \beta) \cdot P_{10}(\alpha, \beta) \tag{3.2}$$

$$-\infty < \alpha, \beta < \infty,$$

where C is a normalizing factor. Specifically,

$$C^{-1} = P\{V \neq \phi\} \tag{3.3}$$

$$L(r_1, r_2, \cdots, r_t; \alpha, \beta) = \Pi_{k=1}^{t} \frac{\binom{n_k}{r_k}[\exp(\alpha + \beta T_k)]^{r_k}}{[1 + \exp(\alpha + \beta T_k)]^{n_k}}, \tag{3.4}$$

which is simply the likelihood function from t independent binomial samples, and

$$P_{01}(\alpha, \beta) = \Sigma_{j>i} \frac{(T_j - T_i) r_i (n_j - r_j)}{[1 + \exp(\alpha + \beta T_i)][1 + \exp(-\alpha - \beta T_j)]} \tag{3.5}$$

$$P_{10}(\alpha, \beta) = \Sigma_{j>i} \frac{(T_j - T_i)(n_i - r_i) r_j}{[1 + \exp(-\alpha - \beta T_i)][1 + \exp(\alpha + \beta T_j)]}. \tag{3.6}$$

Note that in the above notation the dependence on the data has been suppressed. The derivation of $f_{01}(\alpha, \beta)$ will be outlined. First assume $r_i \neq 0, n_i, i = 1, 2, \cdots, t$. By definition, $f_{01}(\alpha, \beta)$ may be symbolically written as

$$P\{N_L = (\alpha, \beta) | V \neq \phi\} = C \quad P\{N_L = (\alpha, \beta) \quad \text{and} \quad V \neq \phi\},$$

which can be found directly by summing the contributions from the $t(t - 1)/2$ mutually exclusive possibilities given in (2.8). That is, letting N_{ij} denote $S_i \cap W_j$, then

$$\begin{aligned} f_{01}(\alpha, \beta) &= C\Sigma_{j>i} P\{N_L = N_{ij} \quad \text{and} \quad N_{ij} = (\alpha, \beta) \quad \text{and} \quad V \neq \phi\} \\ &= C\Sigma_{j>i} P_{ij}(\alpha, \beta) \cdot G_{ij}(\alpha, \beta), \end{aligned} \tag{3.7}$$

where

$$P_{ij}(\alpha, \beta) = P\{N_L = N_{ij} \quad \text{and} \quad V \neq \phi | N_{ij} = (\alpha, \beta)\}, \tag{3.8}$$

and

$$G_{ij}(\alpha, \beta) = P\{N_{ij} = (\alpha, \beta)\}. \tag{3.9}$$

$P_{ij}(\alpha, \beta)$ is most easily derived by a geometrical argument, which essentially involves obtaining the probability that a given point lies within independent pairs of random lines. From (3.8), $P_{ij}(\alpha, \beta)$ denotes the probability that the point of intersection (α, β), formed by the lines S_i and W_j, is the "left-most" point of V. However, if N_{ij} is a point of V then its only possible position in V is as the "left-most" vertex; thus it is only necessary to find the probability that this point belongs to V. Recalling that $V = V_1 \cap V_2, \cdots \cap V_t$, where $V_1, V_2, \cdots V_t$ are independent random strips, then the probability that the given point belongs to V is simply the probability that this point belongs to each strip; that is,

$$P_{ij}(\alpha, \beta) = \Pi_{k=1}^{t} P\{(\alpha, \beta) \in V_k | N_{ij} = (\alpha, \beta)\}. \tag{3.10}$$

Recalling that S_k and W_k are the bounds of V_k, then for $k = i$ or j, the given point (α, β) belongs to the random strip V_k with probability one. In the case of $k \neq i, j$ the point (α, β) belongs to V_k if and only if $(\alpha + \beta T_k)$ is included between X_k and Z_k. Thus, for $k \neq i, j$

$$P\{(\alpha, \beta) \in V_k | N_{ij} = (\alpha, \beta)\} = P\{X_k \leqslant \alpha + \beta T_k \leqslant Z_k\}.$$

Recalling the inverse logit transformation, this last probability is equivalent to

$$P\left\{ U_{(r_k)} \leqslant [1 + \exp(-\alpha - \beta T_k)]^{-1} \leqslant U_{(r_k+1)} \right\}$$

and hence,

$$P_{ij}(\alpha, \beta) = \Pi^t_{k=1, k \neq i, j} \frac{\binom{n_k}{r_k} \exp[(\alpha + \beta T_k) r_k]}{[1 + \exp(\alpha + \beta T_k)]^{n_k}}. \tag{3.11}$$

The density $G_{ij}(\alpha, \beta)$ is simply obtained from the joint density of X_i and Z_j, which in turn is obtained from the joint density of $U_{(r_i)}$ and $U_{(r_j+1)}$, which are independent beta random variables. Letting $f_{ij}(\alpha, \beta) = P_{ij}(\alpha, \beta) \cdot G_{ij}(\alpha, \beta)$, it follows that

$$f_{ij}(\alpha, \beta) = \frac{K_{ij} L(r_1, r_2, \cdots, r_t; \alpha, \beta)}{[1 + \exp(\alpha + \beta T_i)][1 + \exp(-\alpha - \beta T_j)]}, \tag{3.12}$$

where $K_{ij} = (T_j - T_i) r_i (n_j - r_j)$. Summing $f_{ij}(\alpha, \beta)$ for $j > i$, one obtains $f_{01}(\alpha, \beta)$ as stated in (3.1).

In the case of some $r_i = 0$ or n_i, say $r_e = 0$ and $r_k = n_k$, the preceding derivation is changed only in that $P_{ij}(\alpha, \beta) = 0$, for $i = e, j > e$, and for $i < k, j = k$. The expression for $f_{01}(\alpha, \beta)$ is as stated in (3.1), with $r_e = 0$ and $r_k = n_k$. The derivation of $f_{10}(\alpha, \beta)$ is similar.

Note that the quantity C^{-1} has certain interesting features and deserves some attention on its own right. For example, C^{-1} which is equal to the probability of V nonempty, may be useful in determining the goodness of fit of the logit model. Also the effect of removing outlying observations may be assessed by comparing the two different values of C^{-1}. Approximations of C^{-1} will be given in Section 5.

In passing it is noted that the marginal densities in Theorem 3.1 bear a resemblance to Bayesian posterior probabilities with "prior densities" P_{01} and P_{10}, which are weighted averages of priors, the weights depending on the data. These densities are obtained via a Bayesian analysis in West (1977b).

4. Exact upper and lower probabilities. This section is concerned with upper and lower probabilities for certain types of regions in the (α, β) plane. All such probabilities are conditional on V nonempty.

4.1. *Marginal P^* and P_* for a and β.*

Consider the fixed region of β values of the form: $\{\beta | \beta \geqslant b\}$, for any real number b. Projecting the random region V onto the β axis results in the random interval $[\beta_s, \beta_m]$ of β values, and hence,

$$P^*\{\beta \geqslant b\} = P\{\beta_m \geqslant b\} = \int_b^\infty g_{01}(\beta)\, d\beta \tag{4.1.1}$$
$$P_*\{\beta \geqslant b\} = P\{\beta_s \geqslant b\} = \int_b^\infty g_{10}(\beta)\, d\beta,$$

where $g_{01}(\beta)$ and $g_{10}(\beta)$ are the marginal densities obtained from $f_{01}(\alpha, \beta)$ and $f_{10}(\alpha, \beta)$, respectively.

In particular $b = 0$ is of special interest, since it is an indication of whether the sequence of p's is increasing or decreasing. Similarly,

$$P^*\{\alpha \geqslant a\} = \int_a^\infty h_{10}(\alpha) d\alpha, \qquad P_*\{\alpha \geqslant a\} = \int_a^\infty h_{01}(\alpha) d\alpha, \tag{4.1.2}$$

where $h_{10}(\alpha)$ and $h_{01}(\alpha)$ are the marginal densities of α_m and α_s, respectively.

Next consider the two sided interval of α-values of the form: $\{\alpha | a_1 \leqslant \alpha \leqslant a_2\}$. The upper probability is determined solely by the marginal densities of α_s and α_m; that is, from (2.6)

$$P^*[a_1, a_2] = 1 - P\{\alpha_s > a_2\} - P\{\alpha_m < a_1\}, \tag{4.1.3}$$

which may be computed from (4.1.2). From (2.7),

$$P_*[a_1, a_2] = P\{\alpha_s \geqslant a_1, \alpha_m \leqslant a_2\}. \tag{4.1.4}$$

The lower probability requires the joint density of α_s and α_m, which is given in the appendix. However, it is fairly easy to show that the lower probability lies in the following interval

$$P_*(-\infty, a_2] - P^*(-\infty, a_1] \leqslant P_*[a_1, a_2] \leqslant P_*(-\infty, a_2] - P_*(-\infty, a_1].$$

Similar expressions can be determined for a two sided interval of β-values.

In some situations it may seem more appropriate to condition on some subspace of the parameter space. Suppose it is decided to condition on $I = \{(\alpha, \beta) | \beta > 0\}$. Letting $F = \{(\alpha, \beta) | \alpha > a\}$, the upper and lower conditional probabilities of F given I are defined as,

$$P^*(F|I) = P^*(F \cap I)/P^*(I) \tag{4.1.5}$$
$$P_*(F|I) = 1 - P^*(\bar{F}|I).$$

Applying (2.1) and (2.2) it follows that

$$P^*(F|I) = P\{\alpha_m > a, \beta_m > 0\}/P\{\beta_m > 0\} \tag{4.1.6}$$
$$P_*(F|I) = P\{\alpha_s > a, \beta_m > 0\}/P\{\beta_m > 0\}.$$

In addition to the P^* and P_*, one might also consider the upper and lower expected values. The expected values are defined as

$$E^*(\beta) = E(\beta_m), \qquad E_*(\beta) = E(\beta_s)$$

and similarly,

$$E^*(\alpha) = E(\alpha_m), \qquad E_*(\alpha) = E(\alpha_s).$$

It is shown in West (1971), that for each parameter the maximum likelihood estimate lies between the upper and lower expected values; that is,

$$E_*(\alpha) < \hat{\alpha} < E^*(\alpha), \qquad E_*(\beta) < \hat{\beta} < E^*(\beta).$$

4.2. *Other inferences.* Consider the fixed region $E = \{(\alpha, \beta) | \alpha + \beta T \geqslant c\}$. Let $A = (\alpha_0, \beta_0)$ and $B = (\alpha_0', \beta_0')$ denote the vertices of V such that, $(\alpha_0 + \beta_0 T) \leqslant (\alpha + \beta T) \leqslant (\alpha_0' + \beta_0' T)$, for all $(\alpha, \beta) \in V$. Thus

$$(4.2.1) \qquad P^*(E) = P\{\alpha_0' + \beta_0' T \geqslant c\} = \int_{-\infty}^{\infty} \int_{c-\beta T}^{\infty} f_B(\alpha, \beta)\, d\alpha\, d\beta$$

$$P_*(E) = P\{\alpha_0 + \beta_0 T \geqslant c\} = \int_{-\infty}^{\infty} \int_{c-\beta T}^{\infty} f_A(\alpha, \beta)\, d\alpha\, d\beta,$$

where $f_A(\alpha, \beta)$ and $f_B(\alpha, \beta)$ are the densities of A and B respectively. These densities are derived in West (1977b). Note that the coordinates of A and B are functions of T. If $T \leqslant T_1$ (or $T \geqslant T_t$) then A and B are simply N_L and N_R (or N_R and N_L) respectively. If $T_k \leqslant T < T_{k+1}$, $k = 1, 2, \cdots, t$, then $A = A_k$ and $B = B_k$ where

$$\begin{aligned} A_k &= S_i \cap W_j, & B_k &= S_j \cap W_i, & \text{for} \quad & k < i < j \leqslant t \\ &= S_i \cap S_j, & &= W_j \cap W_i, & & 1 \leqslant i \leqslant k, \quad j > k \\ &= S_j \cap W_i, & &= S_i \cap W_j, & & 1 \leqslant i < j \leqslant k. \end{aligned}$$

These results follow from the general properties of V and are used in deriving the densities of A and B.

If $R = \{(\alpha, \beta) | (-\alpha/\beta) \leqslant k\}$, the $P^*(R)$ and $P_*(R)$ can be expressed in terms of the coordinates of A and B. It is shown in West (1977b) that

$$(4.2.2) \qquad P^*(R) = P\{\underline{\theta} \leqslant 0, \bar{\theta} \geqslant 0\} + P\{\bar{\theta} < 0, \beta_s \leqslant 0\} + P\{\underline{\theta} > 0, \beta_m \geqslant 0\},$$

$$P_*(R) = 1 - P\{\underline{\theta} < 0, \beta_m > 0\} - P\{\bar{\theta} > 0, \beta_s < 0\} + P\{\underline{\theta} < 0, \bar{\theta} > 0, \beta_s < 0, \beta_m > 0\},$$

where $\bar{\theta} = \alpha_0' + \beta_0' k$ and $\underline{\theta} = \alpha_0 + \beta_0 k$.

5. Normal approximations. By expanding the ln(likelihood) in a second order Taylor series about the maximum likelihood estimates $(\hat{\alpha}, \hat{\beta})$, and the ln "prior" in a first order series about $(\hat{\alpha}, \hat{\beta})$, the following normal approximations are obtained.

Letting $\mathbf{X} = [\alpha, \beta]'$, $\hat{\mathbf{X}} = [\hat{\alpha}\ \hat{\beta}]'$, then the normal approximation for $f_{01}(\mathbf{X})$ is

$$\tilde{f}_{01}(\mathbf{X}) = N(\tilde{\boldsymbol{\mu}}^0, \tilde{\boldsymbol{\Sigma}}^0) \qquad \text{with} \quad \tilde{\boldsymbol{\mu}}^0 = \hat{\mathbf{X}} + \mathbf{I}^{-1}\mathbf{M}^0,\ \tilde{\boldsymbol{\Sigma}}^0 = \mathbf{I}^{-1}, \tag{5.1}$$

where $\mathbf{I}$ is the 2×2 information matrix with elements

$$I_{11}(\hat{\mathbf{X}}) = \Sigma_{i=1}^t n_i \hat{p}_i \hat{q}_i, \qquad I_{12}(\hat{\mathbf{X}}) = \Sigma_{i=1}^t n_i T_i \hat{p}_i \hat{q}_i, \qquad I_{22}(\hat{\mathbf{X}}) = \Sigma_{i=1}^t n_i T_i^2 \hat{p}_i \hat{q}_i,$$

and $\mathbf{M}^0$ is the matrix of first partial derivatives of $\ln P_{01}(\mathbf{X})$ evaluated at $\hat{\mathbf{X}}$, the elements of which are

$$M_\alpha^0(\hat{\mathbf{X}}) = \left[P_{01}(\hat{\mathbf{X}})\right]^{-1} \Sigma_{j>i} K_{i,j} \hat{p}_j \hat{q}_i (\hat{q}_j - \hat{p}_i)$$
$$M_\beta^0(\hat{\mathbf{X}}) = \left[P_{01}(\hat{\mathbf{X}})\right]^{-1} \Sigma_{j>i} K_{i,j} \hat{p}_j \hat{q}_i (\hat{q}_j T_j - \hat{p}_i T_i).$$

An approximation to the constant C is obtained as

$$\tilde{C}_0 = |\mathbf{I}|^{\frac{1}{2}} \exp(-2^{-1}\mathbf{M}^{0\prime}\mathbf{I}^{-1}\mathbf{M}^0) / \left[2\Pi L(r_1, \cdots, r_t; \hat{\mathbf{X}}) P_{01}(\hat{\mathbf{X}})\right]. \tag{5.2}$$

Analogous to $f_{01}(\mathbf{X})$, one obtains the following normal approximation for $f_{10}(\mathbf{X})$.

$$\tilde{f}_{10}(\mathbf{X}) = N(\tilde{\boldsymbol{\mu}}^1, \tilde{\boldsymbol{\Sigma}}^1) \qquad \text{with} \quad \tilde{\boldsymbol{\mu}}^1 = \hat{\mathbf{X}} + \mathbf{I}^{-1}\mathbf{M}^1,\ \tilde{\boldsymbol{\Sigma}}^1 = \mathbf{I}^{-1}, \tag{5.3}$$

where $\mathbf{M}^1$ is the matrix of first partial derivatives, analogous to $\mathbf{M}^0$, the elements of which are

$$M_\alpha^1(\hat{\mathbf{X}}) = \left[P_{10}(\hat{\mathbf{X}})\right]^{-1} \Sigma_{j>i} K'_{j,i} \hat{p}_i \hat{q}_j (\hat{q}_i - \hat{p}_j)$$
$$M_\beta^1(\hat{\mathbf{X}}) = \left[P_{10}(\hat{\mathbf{X}})\right]^{-1} \Sigma_{j>i} K'_{ji} \hat{p}_i \hat{q}_j (\hat{q}_i T_i - \hat{p}_j T_j).$$

Another approximation to the constant C is obtained as

$$\tilde{C}_1 = |\mathbf{I}|^{\frac{1}{2}} \exp(-2^{-1}\mathbf{M}^{1\prime}\mathbf{I}^{-1}\mathbf{M}^1) / \left[2\Pi L(r_1, \cdots, r_t; \hat{\mathbf{X}}) P_{10}(\hat{\mathbf{X}})\right]. \tag{5.4}$$

Note that both $\tilde{C}_0^{-1}$ and $\tilde{C}_1^{-1}$ provide an approximation to the probability that the random region V is nonempty.

The justification for these approximations are given in West (1971). The Monte Carlo studies, discussed in Section 6, indicate that the normal approximations are quite good, but can be improved upon by the following adjustment to the expected values. Letting $\boldsymbol{\delta} = 2^{-1}\mathbf{I}^{-1}(\mathbf{M}^0 + \mathbf{M}^1)$, then the adjusted expected values, denoted by, $\dot{\boldsymbol{\mu}}^0$ and $\dot{\boldsymbol{\mu}}^1$, are

$$\dot{\boldsymbol{\mu}}^0 = \tilde{\boldsymbol{\mu}}^0 - \boldsymbol{\delta}, \qquad \dot{\boldsymbol{\mu}}^1 = \tilde{\boldsymbol{\mu}}^1 - \boldsymbol{\delta}. \tag{5.5}$$

The accuracy of these approximations is illustrated in the next section.

Note that if the ln "priors" are approximated by quadratic functions, the resulting normal approximations are generally not as good as (5.1) and (5.3).

5.1. *A numerical example.* The example involves three levels, $T_i = (i - 1)$ for $i = 1, 2, 3$, with 10 independent trials at each level. The observed number of successes at each level was: $r_1 = 3$, $r_2 = 8$, $r_3 = 6$. For this data the maximum likelihood estimates of α and β are $\hat{\alpha} = -.345$ and $\hat{\beta} = .631$.

From (5.1) α_s and β_m are approximately distributed as the bivariate normal density $\tilde{f}_{01}(\mathbf{X})$ with

$$(5.1.1) \qquad \tilde{\boldsymbol{\mu}}^0 = [-.473, .738]' \quad \text{and} \quad \tilde{\boldsymbol{\Sigma}}^0 = \begin{bmatrix} .345 & -.212 \\ -.212 & .225 \end{bmatrix}.$$

Similarly, from (5.3) α_m and β_s are approximately distributed as $\tilde{f}_{10}(\mathbf{X})$ with mean and covariance matrix

$$(5.1.2) \qquad \tilde{\boldsymbol{\mu}}^1 = [-.160, .423]' \quad \text{and} \quad \tilde{\boldsymbol{\Sigma}}^1 = \tilde{\boldsymbol{\Sigma}}^0,$$

respectively. The estimates of C relative to these densities are from (5.2) and (5.4): $\tilde{C}_0 = 11.218$ and $\tilde{C}_1 = 10.467$.

For comparison, an array of numerical values corresponding to each exact density was constructed. For each array the value of C was computed, with the result that from both arrays the same number $C = 11.19669$ was obtained. Comparing this value of C with the two approximate values, it is clear that $\tilde{C}_0$ is the better estimate. (Generally, it appears that whenever $\hat{\beta} > 0$, $\tilde{C}_0$ is the better estimate and whenever $\hat{\beta} < 0$, $\tilde{C}_1$ is better.)

Using numerical analysis on the appropriate array of exact values, various quantities were computed. For example, the upper and lower expected values for β are

$$E^*\beta = E(\beta_m) = .798, \qquad E_*\beta = E(\beta_s) = .479.$$

Noting that the normal means are both underestimated by about the same amount and the near symmetry of the means about $\hat{\beta}$, it would seem that the adjusted means, defined in (5.5), would lead to better estimates. In this case the adjusted means are

$$(5.1.3) \qquad \dot{\tilde{E}}\beta_m = .790, \qquad \dot{\tilde{E}}\beta_s = .473,$$

which do give closer estimates; however even the original normal densities lead to good approximations. The P^* and P_* computed by numerical integration and by using the normal approximations, in (5.1.1) and (5.1.2), are recorded in Table 5.1.1, for a number of events.

TABLE 5.1.1
Exact and normal approximations of $P^*(F)$ and $P_*(F)$

F	$P^*(F)$		$P_*(F)$	
	Exact	*Approx.*	*Exact*	*Approx.*
$\alpha > -1.43$	.98	.98	.95	.95
$\beta > 0$	.94	.94	.83	.82
$\beta < \hat{\beta}$	.66	.67	.41	.41
$\alpha + \beta T_3 \geq 0$	.95	.95	.88	.87
$\alpha + \beta T_1 \geq 0$	.37	.37	.16	.16

6. Monte Carlo studies of P^* and P_* inferences. Monte Carlo studies were undertaken to investigate, under varying conditions, the behavior of the P^* and P_* inferences, and also to investigate the accuracy of several possible approximations for the exact probabilities. For these studies the design parameters are t, the number of binomial parameters or equivalently the number of equally spaced T-levels; W, the range of p; P_m, the median p; and N the size of the sample drawn at each level. Studies were made for designs with $t = 5, 10$; $W = .17, .6, .94$; $P_m = .5$; and $N = 2, 6, 20$. The results are summarized as follows.

1. In each study the array of exact values of the posterior densities resembled an array of numbers that could have come from a bivariate normal density, with individual mean values slightly smaller (if α_s or β_s) or slightly larger (if α_m or β_m) than the corresponding maximum likelihood estimate.

2. For each parameter, the upper and lower expected values were nearly equidistant from the maximum likelihood estimate, which they approached as the sample size increased.

3. The upper and lower probabilities of a p-level one-sided confidence interval were above and below p respectively.

4. The inferences became more precise with increased sample size.

5. The main difference between two experiments having the same value for the sufficient statistic $(\Sigma R_i, \Sigma R_i T_i)$, was in the value of C^{-1}. Of the two experiments, the one that resulted in sample values that gave more evidence to support a logistic model, had a larger value for C^{-1}. (This is not surprising since C^{-1} may be interpreted as the probability of V being nonempty.)

In each study four normal approximations were considered. One set of approximations, denoted by A1, is the set of normal densities given in (5.1) and (5.3). A second set, denoted by A2, is the set of normal densities obtained by approximating the ln "priors" by quadratic functions. In the third set of approximations, A3, the first two moments of the normal density are equal to the corresponding moments of the exact density. The fourth set of normal densities, A1′, is the modified version of A1, given in (5.5). The inferences obtained by the numerical integration of the exact densities were compared to the resulting inferences using the normal approximations with the following results.

6. Between A1 and A2, the results tend to generally favor A1. In every design A1 is as good as, or better than A2 for almost all the quantities. It is clear that A3 does better than A1, most noticeable for $N = 2$; although in most cases even A1 does a reasonably good job. The results indicate that A1′ is better than A1 and as good as A3. A1′ is much more desirable than A3, since the latter involves the computation of the mean and variances from the exact grid.

The details of these Monte Carlo studies are recorded in West (1971).

7. Remarks. The types of inferences discussed in this paper involve essentially posterior probability statements about regions of parameter values for the logistic function. These fixed regions may be viewed as hypotheses whose validity is to be tested.

The decision maker may think the upper and lower probabilities too conservative to be very useful, especially if the bounds are widely separated. But the fact that they are widely separated indicates that there is not much information on which to base a decision. The difference between the bounds acts as a safeguard against over confidence. It would seem that the upper and lower probability inferences could be quite useful in an area such as bio-assay, especially in view of the approximate nature of standard inferences.

The upper and lower probabilities provide a flexible means of summarizing the current state of knowledge about some hypothesis, where a state of complete ignorance is represented by an upper and lower probability of 1 and 0 respectively. Moreover, the difference between the bounds provides a kind of second order uncertainty describing the quality of the inference. For a given hypothesis, the difference, $(P^* - P_*)$, decreases as the sample size increases, indicating the improvement of the quality of the inference as a result of the increase in the number of observations. One gradually goes from no inferences to more and more firm inferences as the data accumulates. For extremely large sample size the upper and lower probabilities will coincide and both will be derivable from a posterior density proportional to the likelihood. It is noted that the difference, $P^* - P_*$, also reflects the design of the experiment; that is, relative to a given hypothesis the more efficient the design the smaller the difference between bounds.

The inferences derived in this paper depend on densities which are generally not integrable in closed form. Consequently, in application one has to either use numerical integration or some other approximation. Fortunately, the posterior densities are approximately normal functions even for moderate size samples, with the result that for all practical purposes, the upper and lower probability inferences require only tables of the normal density function. Approximations for the joint density of two vertices of V were not investigated, but it would seem that for moderately large samples a good approximation would be a multivariate normal density function.

APPENDIX

The joint density of N_L and N_R. In the computation of certain inferences the joint densities of certain pairs of vertices are needed, at least theoretically. Assuming $V \neq \phi$, the joint density of $N_L = (\alpha, \beta)$ and $N_R = (\alpha', \beta')$ defined over $\Omega = \{(\alpha, \beta, \alpha', \beta') | \alpha < \alpha', \beta > \beta'\}$ is given in the following theorem. (The density for any other pair of vertices is similar.)

THEOREM A1. *Denoting the joint density of $N_L = (\alpha, \beta)$ and $N_R = (\alpha', \beta')$ by $f(\mathbf{Z})$ where $\mathbf{Z} = [\alpha, \beta, \alpha', \beta'] \in \Omega$ then*

$$\begin{aligned} f(\mathbf{Z}) &= 0 \quad \text{if} \quad (\alpha' - \alpha)/(\beta - \beta') \leqslant T_1 \quad \text{or} \quad \geqslant T_t \\ &= f_{s-1,s}(\mathbf{Z}) \quad \text{if} \quad T_{s-1} < (\alpha' - \alpha)/(\beta - \beta') < T_s \\ &= f_s(\mathbf{Z}) \quad \text{if} \quad (\alpha' - \alpha)/(\beta - \beta') = T_s \end{aligned}$$

for $s = 2, 3, \cdots, t$, *where*

$$f_{s-1,s}(\mathbf{Z}) = H(\mathbf{Z}) \cdot P_{s-1,s}(\mathbf{Z}), \qquad f_s(\mathbf{Z}) = H(\mathbf{Z})P_s(\mathbf{Z})$$

with

$$H(\mathbf{Z}) = \Pi_{e=1}^{s-1}\binom{n}{r_e}(p_e)^{r_e}(q_e')^{n-r_e}\Pi_{e=s}^{t}\binom{n}{r_e}(p_e')^{r_e}(q_e)^{n-r_e}$$

$$P_{s-1,s}(\mathbf{Z}) = \left[\Sigma_{i=1}^{s-1}\Sigma_{k=s}^{t}(T_k - T_i)r_i(n - r_k)q_i p_k\right] \cdot \left[\Sigma_{j=1}^{s-1}\Sigma_{l=s}^{t}(T_l - T_j)(n - r_j)r_l q_l' p_j'\right]$$

$$P_s(\mathbf{Z}) = (n - r_s)p_s'\left[\Sigma_{i=1}^{s-1}(T_s - T_i)r_i q_i\right]\left[\Sigma_{l=s+1}^{t}(T_l - T_s)r_l q_l'\right] + r_s q_s\left[\Sigma_{i=1}^{s-1}(T_s - T_i)(n - r_i)p_i'\right]\left[\Sigma_{l=s+1}^{t}(T_l - T_s)(n - r_l)p_l\right],$$

and p_e, p_e' *denote* $[1 + \exp(-\alpha - \beta T_e)]^{-1}, [1 + \exp(-\alpha' - \beta' T_e)]^{-1}$ *respectively. This density is derived in West* (1971).

Acknowledgment. The author wishes to thank the Associate Editor and the referee for their helpful suggestions.

REFERENCES

[1] Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.* **38** 325–39.

[2] Dempster, A. P. (1968a). A generalization of Bayesian inference. *J. Roy. Statist. Soc. Ser. B* **30** 205–47.

[3] Dempster, A. P. (1968b). Upper and lower probabilities generated by a random closed interval. *Ann. Math. Statist.* **39** 957–66.

[4] Fieller, E. C. (1954). Some problems on interval estimation. *J. Roy. Statist. Soc. Ser. B* **16** 175–85.

[5] Rao, C. R. (1961). Asymptotic efficiency and limiting information. *Proc. Fourth Berkeley Symp. Math. Statist. Probability* **1** 531–45.

[6] West, Sandra A. (1971). Upper and lower probability inferences for the logistic function. Ph.D. Thesis, Dept. Statist., Harvard Univ., Cambridge, Massachusetts.

[7] West, Sandra A. (1977a). Upper and lower probability inferences pertaining to poisson processes. *J. Amer. Statist. Assoc.* **72** 448–53.

[8] West, Sandra A. (1977b). Bayesian-type inferences for the logistic function. Unpublished research report.

Department of Mathematics
Worcester Polytechnic Institute
Worcester, Massachusetts 01609