The Annals of Mathematical Statistics
1968, Vol. 39, No. 3, 957–966

UPPER AND LOWER PROBABILITIES GENERATED BY A RANDOM CLOSED INTERVAL[1]

By A. P. Dempster

Harvard University

0. Summary. Within the class of models producing upper and lower probability systems, as discussed in Dempster (1967a), a simple and important subclass may be characterized by random intervals on the line. Detailed expressions are given here for the upper and lower probabilities of a general fixed closed interval determined by a general random closed interval. Such random closed intervals occur in the applications of the general class of models to statistical inference described in Dempster (1966, 1967b, 1968). The illustration given here concerns inference about binomial p and stresses the flexibility allowed in the introduction of prior information.

1. Introduction. Suppose that U and V are random variables such that $U \leqq V$ with probability one. U and V may be viewed as determining a random closed interval $[U, V]$ on the real line. Loosely speaking, the probability distributed over the half-plane of (U, V) points divides into three parts relative to any Borel set A on a real line: (i) probability associated with intervals $[U, V] \subset A$, (ii) probability associated with intervals $[U, V] \subset \bar{A}$ where $\bar{A}$ is the complement of A, and (iii) probability associated with intervals $[U, V]$ having non-empty intersections with both A and $\bar{A}$. The first part consists of probability which necessarily relates to outcomes in A and will be called the *lower probability* of A and denoted by $P_*(A)$. The second part cannot relate to any outcome in A while the third part may relate either to A or to $\bar{A}$. The sum of the first and third parts is the largest amount of probability which may relate to outcomes in A and will be called the *upper probability* of A and denoted by $P^*(A)$. Note that, if $U = V$ with probability one, the random interval reduces to a random point and the part (iii) above vanishes, so that $P^*(A) = P_*(A) = P(A)$ where $P(A)$ is the ordinary probability that the random point lies in A.

The theory just outlined is a specific instance of a general theory which was described in Dempster (1967a). In general, upper and lower probabilities are defined over subsets of a space S by a multivalued mapping Γ from X to S where X is a space having an ordinary probability measure μ over a class of subsets $\mathfrak{F}$. In Dempster (1967a) the case of finite S was emphasized. Here S is the real line, and the subsets $\Gamma x \subset S$ are closed intervals for each $x \varepsilon X$. For present purposes the points of X which lead to a common interval $[U, V]$ are an equivalence class, and hence X may be identified with the half-plane of points (U, V) where $U \leqq V$. In these terms the multivalued mapping Γ carries the point (U, V) in X into the closed interval $[U, V] \subset S$.

Received July 21, 1967.

[1] This research was supported by the Office of Naval Research. Distribution of the document is unlimited.

In many of the applications of models of this type to statistical inference, as proposed in Dempster (1966) and pursued in Dempster (1967b, 1968), the role of S is played by a k-dimensional parameter space and the subsets Γx are convex sets. Since convex sets are intervals when $k = 1$, the theory of this paper paves the way for some further examples concerning inference about a single parameter. The case of binomial p is used here as a simple first illustration.

2. Some details. The upper and lower probabilities of any Borel set A may be formally expressed as

$$P^*(A) = \Pr\{[U, V] \cap A \neq \varnothing\} \tag{2.1}$$

and

$$P_*(A) = \Pr\{[U, V] \subset A\}. \tag{2.2}$$

Emphasis will be laid on the computation of $P^*(A)$ and $P_*(A)$ when A is a closed interval $[a, b]$. It is easily seen that

$$P^*([a, b]) = 1 - \Pr\{U > b\} - \Pr\{V < a\} \tag{2.3}$$

and

$$P_*([a, b]) = \Pr\{U \geqq a, V \leqq b\}. \tag{2.4}$$

Another important set of probabilities is defined by

$$Q(A) = \Pr\{[U, V] \supset A\} \tag{2.5}$$

which for closed intervals $[a, b]$ may be written

$$Q([a, b]) = \Pr\{U \leqq a, V \geqq b\}. \tag{2.6}$$

These probabilities do not have a direct meaning like $P^*(A)$ and $P_*(A)$ relative to an uncertain outcome on the real line. However, they are technically important in the envisaged applications because they obey a simple product rule for the combination of the information in independent samples (Dempster (1967a)).

From (2.4) and (2.6) it is clear that knowledge of $P_*([a, b])$ for all (a, b) or of $Q([a, b])$ for all (a, b) determines the joint distribution of U and V and thence determines $P^*(A)$ and $P_*(A)$ for any Borel set A. A similar remark does not hold for $P^*([a, b])$ which is determined solely by the marginal distributions of U and V.

For the technical reason given above, the function

$$H(u, v) = Q([u, v]) = \Pr\{U \leqq u, V \geqq v\} \tag{2.7}$$

for $-\infty < u \leqq v < \infty$, is the preferred form of the bivariate cdf of U and V. By Stieltjes integration, one may find from $H(u, v)$ the marginal distributions of U and V which are conveniently characterized here by the modified cdf's

$$F(u) = \Pr\{U > u\} \tag{2.8}$$

and

$$G(v) = \Pr\{V < v\}. \tag{2.9}$$

Note, however, that

$$H(u, u) + F(u) + G(u) = 1 \tag{2.10}$$

so that only one of the margins need be determined from $H(u, v)$ by integration.

In terms of the functions F, G and H one may write

$$\begin{aligned} P^*([a, b]) &= 1 - F(b) - G(a) \\ &= H(b, b) + G(b) - G(a) \\ &= H(a, a) + F(a) - F(b), \end{aligned} \tag{2.11}$$

and

$$\begin{aligned} &P_*([a, b]) \\ &= 1 - F(b) - G(a) - H(b, b+) - H(a-, a) + H(a-, b+) \\ &= F(a-) - F(b) - H(b, b+) + H(a-, b+) \\ &= G(b+) - G(a) - H(a-, a) + H(a-, b+). \end{aligned} \tag{2.12}$$

The equivalence of the three right sides follows from (2.10) or its extensions involving different inclusion or exclusion of boundary points. Note also that

$$P^*([a, b]) - P_*([a, b]) = H(b, b+) + H(a-, a) - H(a-, b+). \tag{2.13}$$

In the special case of an absolutely continuous distribution of (U, V), the marginal density functions of U and V are respectively

$$H_1(u, u) = (\partial/\partial u)H(u, v)]_{v=u} \text{ and } H_2(v, v) = -[(\partial/\partial v)H(u, v)]_{u=v}, \tag{2.14}$$

and convenient formulas are

$$P^*([a, b]) = H(b, b) + \int_a^b H_2(t, t)\, dt = H(a, a) + \int_a^b H_1(t, t)\, dt, \text{ and} \tag{2.15}$$

$$\begin{aligned} P_*([a, b]) &= \int_a^b H_1(t, t)\, dt - [H(b, b) - H(a, b)] \\ &= \int_a^b H_2(t, t)\, dt - [H(a, a) - H(a, b)]. \end{aligned} \tag{2.16}$$

It might be sometimes necessary to compute upper and lower probabilities of intervals which are open at one or both ends. The appropriate formulas are simple modifications of those given above, and are not written out here. If the distribution of (U, V) is absolutely continuous, the probabilities are the same regardless of whether the intervals are open or closed.

A different possible extension would be to allow the random interval determined by (U, V) to be some mixture of open and closed intervals, allowing in general a different mixture of the four types of interval for each (u, v). The complications entering with such a formulation seem to outweigh its immediate

usefulness, and it is not pursued in detail here. Again the previous formulas remain valid in the presence of absolutely continuous (U, V).

3. Binomial sampling. Some formulas for inference about binomial p which were first given in Dempster (1966) are rederived here following the more general methods of this paper. Suppose that Z denotes an observable, taking the values 0 or 1, where the long run frequency of 1 is an unknown p on $0 \leqq p \leqq 1$. A random variable W uniformly distributed on (0, 1) is supposed to underlie each observation Z where

$$\begin{aligned} Z &= 0 \quad \text{if} \quad p < W \leqq 1 \\ &= 1 \quad \text{if} \quad 0 \leqq W < p \\ &= 0 \quad \text{or} \quad 1 \quad \text{if} \quad W = p. \end{aligned} \tag{3.1}$$

An observation of $Z = 0$ implies that $0 \leqq p \leqq W$, thus creating a random closed interval $[0, W]$ of p values governed by the uniform distribution of W. If the interval is denoted by $[U, V]$ as in Sections 1 and 2, then

$$H(u, v) = \Pr\{U \leqq u, V \geqq v\} = 1 - v \tag{3.2}$$

for $0 \leqq u \leqq v \leqq 1$. Similarly, an observation of $Z = 1$ leads to the random closed interval $[W, 1]$ for which

$$H(u, v) = u, \tag{3.3}$$

for $0 \leqq u \leqq v \leqq 1$. The systems of upper and lower probabilities relating to p generated by (3.2) or (3.3) are conditional probabilities, given the observations $Z = 0$ or $Z = 1$, respectively.

If n independent observations are made, yielding $n - r$ 0's and r 1's, then the rule for combining independent sources of information as given in Dempster (1967a, 1968) implies that the combination will again be based on a random closed interval whose $H(u, v)$ is proportional to the product of the $H(u, v)$ for the individual sources. In other words, a binomial sample with r 1's in n trials yields a system of conditional upper and lower probabilities for Borel sets of p values, governed by

$$H(u, v) = Ku^r(1 - v)^{n-r}, \tag{3.4}$$

where K is a normalizing constant and $0 \leqq u \leqq v \leqq 1$. The normalizing constant is found by characterizing that distribution over the triangle $0 \leqq u \leqq v \leqq 1$ for which $\Pr\{U \leqq u, V \geqq v\} = Ku^r(1 - v)^{n-r}$. Three cases need to be considered: $r = 0$, $r = n$ and $0 < r < n$. If $r = 0$, the distribution is concentrated on the line segment $u = 0$, $0 \leqq v \leqq 1$ and has the beta density $Kn(1 - v)^{n-1}$ along the line segment. If $r = n$, the distribution is concentrated on the line segment $0 \leqq u \leqq 1$, $v = 1$ and has the beta density Knu^{n-1} on the line segment. If $0 < r < n$, the distribution is absolutely continuous over $0 \leqq u \leqq v \leqq 1$ and has the density $Kr(n - r)u^{r-1}(1 - v)^{n-r-1}$ which is the density of the rth and $(r + 1)$st order statistics of a random sample of size n from a uniform (0, 1)

population. It follows that

$$K = \binom{n}{r} \quad \text{for} \quad n \geqq 1 \quad \text{and} \quad 0 \leqq r \leqq n. \tag{3.5}$$

Formulas (2.15) and (2.16) may now be applied for $0 < r < n$ to yield

$$\begin{aligned} P^*([a, b]) &= \binom{n}{r} b^r (1-b)^{n-r} + n!\,[r!\,(n-r-1)!]^{-1} \int_a^b t^r (1-t)^{n-r-1}\,dt \\ &= \binom{n}{r} a^r (1-a)^{n-r} \\ &\quad + n!\,[(r-1)!\,(n-r)!]^{-1} \int_a^b t^{r-1}(1-t)^{n-r}\,dt, \end{aligned} \tag{3.6}$$

and

$$\begin{aligned} &P_*([a, b]) \\ &= n![(r-1)!\,(n-r)!]^{-1} \int_a^b t^{r-1}(1-t)^{n-r}\,dt - \binom{n}{r}[b^r - a^r](1-b)^{n-r} \\ &= n![r!\,(n-r-1)!]^{-1} \int_a^b t^r (1-t)^{n-r-1}\,dt \\ &\quad - \binom{n}{r} a^r [(1-a)^{n-r} - (1-b)^{n-r}]. \end{aligned} \tag{3.7}$$

The cases $r = 0$ and $r = n$ require separate handling, but it is easily checked that

$$\begin{aligned} P^*([a, b]) &= \Pr\{V \geqq a\} = (1-a)^n \quad \text{if} \quad r = 0 \\ &= \Pr\{U \leqq b\} = b^n \quad \text{if} \quad r = n, \quad \text{and} \end{aligned} \tag{3.8}$$

$$\begin{aligned} P_*([a, b]) &= \Pr\{V \leqq b\} = 1 - (1-b)^n \quad \text{if} \quad a = 0 \quad \text{and} \quad r = 0 \\ &= 0 \quad \text{if} \quad a > 0 \quad \text{and} \quad r = 0 \\ &= \Pr\{U \geqq a\} = 1 - a^n \quad \text{if} \quad b = 1 \quad \text{and} \quad r = n \\ &= 0 \quad \text{if} \quad b < 1 \quad \text{and} \quad r = n. \end{aligned} \tag{3.9}$$

Another convenient approach is to use the first line of (2.11), noting that the marginal distributions of U and V are given by

$$\begin{aligned} F(u) &= \Pr\{U > u\} \\ &= n!\,[(r-1)!(n-r)!]^{-1} \int_u^1 t^{r-1}(1-t)^{n-r}\,dt \quad \text{if} \quad r > 0 \\ &= 0 \quad \text{if} \quad r = 0, \end{aligned} \tag{3.10}$$

for $0 \leqq u \leqq 1$ and

$$\begin{aligned} G(v) &= \Pr\{V < v\} \\ &= n!\,[r!\,(n-r-1)!]^{-1} \int_0^v t^r (1-t)^{n-r-1}\,dt \quad \text{if} \quad r < n \\ &= 0 \quad \text{if} \quad r = n, \end{aligned} \tag{3.11}$$

for $0 \leqq v \leqq 1$. Alternative forms for (3.10) and (3.11) are

$$\begin{aligned} F(u) &= \sum_{x=0}^{r-1} \binom{n}{x} u^x (1-u)^{n-x} \quad \text{if} \quad r > 0 \\ &= 0 \quad \text{if} \quad r = 0, \quad \text{and} \end{aligned} \tag{3.12}$$

$$\begin{aligned} G(v) &= \sum_{x=r+1}^{n} \binom{n}{x} v^x (1-v)^{n-x} \quad \text{if} \quad r < n \\ &= 0 \quad \text{if } r = n. \end{aligned} \tag{3.13}$$

Formulas (3.12) and (3.13) together with (2.11) provide a convenient route for computing $P^*([a, b])$. Having $P^*([a, b])$, one may compute $P_*([a, b])$ from (2.13).

NUMERICAL EXAMPLE. Suppose that 6 successes are observed in 10 trials. What are the upper and lower probabilities that $.25 \leqq p \leqq .75$ where p is the long run frequency of success?

$$F(.75) = \sum_{x=0}^{5} \binom{10}{x}(.75)^x(.25)^{10-x} = .0781$$
$$G(.25) = \sum_{x=7}^{10} \binom{10}{x}(.25)^x(.75)^{10-x} = .0035,$$

from which

$$P^*([.25, .75]) = 1 - .0781 - .0035 = .9184.$$

From (2.13),

$$\begin{aligned} P_*([.25, .75]) &= P^*([.25, .75]) - \binom{10}{6}(.75)^6(.25)^4 - \binom{10}{6}(.25)^6(.75)^4 \\ &\quad + \binom{10}{6}(.25)^6(.25)^4 \\ &= .9184 - .14600 - .01622 + .00020 \\ &= .7564. \end{aligned}$$

4. Prior information. The inferences just given do not formally incorporate prior information about p. But sometimes prior information may be commensurate with sample information, and then to ignore it would be unrealistic. If prior information about p is available and if it may be expressed in terms of a distribution over sets of values of p, then the prior information may be treated as a source of information independent of the sample observations and therefore combinable directly with the sample observations. Indeed, even the analysis of Section 3 which made no mention of prior information may be viewed as a combination of the sample information with a trivial form of prior information, namely the "informationless" prior assignment of unit probability to [0, 1] for which $H(u, v) = 1$ on $0 \leqq u \leqq v \leqq 1$.

Another limiting form of prior information is the familiar prior distribution of the Bayesian statistician. In this case, the prior information is expressed in terms of a random interval $[U, V]$ where $U = V$ with probability one. Consider a discrete prior distribution where

$$\Pr(p = p_i) = f_i, \tag{4.1}$$

for $i = 1, 2, \cdots$. The associated $H(u, v)$ function is very simple, namely,

$$\begin{aligned} H(u, v) &= f_i \quad \text{if} \quad (u, v) = (p_i, p_i) \qquad \text{for} \quad i = 1, 2, \cdots, \\ &= 0 \qquad \text{otherwise,} \end{aligned} \tag{4.2}$$

for $0 \leqq u \leqq v \leqq 1$. Combination of this source with the sample information

(3.4) leads to

$$H(u, v) = K' f_i p_i^r (1 - p_i)^{n-r} \quad \text{if} \quad (u, v) = (p_i, p_i) \quad \text{for} \quad i = 1, 2, \cdots, \tag{4.3}$$
$$= 0 \qquad \text{otherwise.}$$

From (4.3) it is clear that the result is the familiar Bayesian posterior distribution. Continuous prior and posterior distributions may be approached as limiting forms of the discrete case just given.

The chief aim of the present section is to illustrate the notion that more flexible forms of prior information may be useful. Consider a hypothetical experimenter engaged in screening drugs for their efficacy in treating a controlled condition experimentally induced in rats of a specified population. Suppose that each animal provides a binomial response, i.e., positive effect or no effect, and that the experimenter is willing to define an effective drug as one which would produce 75 per cent positive responses in a large sample of rats. Each drug is therefore postulated to have an unknown p and the question, often to be asked of a very small sample, is whether or not $p \geqq .75$. The experimenter might be willing to formalize his past experience by asserting that he would expect at least 1% and at most 10% of the drugs to be effective. He might not care to differentiate among the set of drugs to be screened. And he might not care to render prior judgments all along the p scale, but only on whether or not $p \geqq .75$. The prior information thus described could be expressed formally as

$$\begin{aligned} \Pr\{[U, V] = [0, .75 - \epsilon]\} &= .90, \\ \Pr\{[U, V] = [.75, 1]\} &= .01, \\ \Pr\{[U, V] = [0, 1]\} &= .09, \end{aligned} \tag{4.4}$$

which implies that

$$P^*(p \geqq .75) = .10 \quad \text{and} \quad P_*(p \geqq .75) = .01. \tag{4.5}$$

Consider the slightly more general form of prior information where

$$\begin{aligned} \Pr\{[U, V] = [0, c]\} &= \alpha, \\ \Pr\{[U, V] = [d, 1]\} &= \beta, \\ \Pr\{[U, V] = [0, 1]\} &= \gamma \end{aligned} \tag{4.6}$$

for given c, d, α, β, γ satisfying $0 \leqq c < d \leqq 1$, $\alpha \geqq 0$, $\beta \geqq 0$, $\gamma \geqq 0$, $\alpha + \beta + \gamma = 1$. The $H(u, v)$ function determined by (4.6) over the triangle $0 \leqq u \leqq v \leqq 1$ is given by

$$\begin{aligned} H(u, v) &= \alpha + \gamma \qquad \text{if} \quad v \leqq c \\ &= \beta + \gamma \qquad \text{if} \quad u \geqq d \\ &= \gamma \qquad \text{otherwise.} \end{aligned} \tag{4.7}$$

Combining this with the sample information (3.4) yields

$$
\begin{aligned}
H(u, v) &= K'(\alpha + \gamma)\binom{n}{r}u^r(1 - v)^{n-r} && \text{if } v \leqq c \\
&= K'(\beta + \gamma)\binom{n}{r}u^r(1 - v)^{n-r} && \text{if } u \geqq d \\
&= K'\gamma\binom{n}{r}u^r(1 - v)^{n-r} && \text{otherwise,}
\end{aligned} \tag{4.8}
$$

over $0 \leqq u \leqq v \leqq 1$.

The next task is to give a simple characterization of the distribution defined by the cdf (4.8). First consider the case $0 < r < n$. From (4.8)

$$\lim_{\epsilon \to 0}\{H(u, c) - H(u, c + \epsilon)\} = K'\alpha\binom{n}{r}u^r(1 - c)^{n-r}, \tag{4.9}$$

for $0 \leqq u \leqq c < 1$, which shows that the line segment (u, c) for $0 \leqq u \leqq c$ carries weight $K_\alpha'\binom{n}{r}c^r(1 - c)^{n-r}$ distributed according to the density $K'\alpha r\binom{n}{r}u^{r-1}(1 - c)^{n-r}$. Similarly the line segment (d, v) for $d \leqq v \leqq 1$ carries weight $K'\beta\binom{n}{r}\, d^r(1 - d)^{n-r}$ distributed according to the density $K'\beta(n - r)\binom{n}{r}\, d^r(1 - v)^{n-r-1}$. No isolated points and no other line segments carry positive probability, but the triangle itself carries a continuous bivariate distribution whose density function is formed by differentiating the expressions in (4.8). By integrating these bivariate densities it may be checked that the three subregions $0 \leqq u \leqq v \leqq c$, $d \leqq u \leqq v \leqq 1$, and $[u, v] \cap [c, d] \neq \varnothing$, carry weights

$$K'(\alpha + \gamma)\sum_{r+1}^{n}\binom{n}{x}c^x(1 - c)^{n-x},\ K'(\beta + \gamma)\sum_{0}^{r-1}\binom{n}{x}\, d^x(1 - d)^{n-x} \text{ and}$$

$$K'\gamma[1 - \textstyle\sum_{r+1}^{n}\binom{n}{x}c^x(1 - c)^{n-x} - \sum_{0}^{r-1}\binom{n}{x}\, d^x(1 - d)^{n-x}].$$

By summing the two linear pieces and the three planar pieces and setting the total equal to unity, one finds that

$$K' = [\alpha\textstyle\sum_{x=r}^{n}\binom{n}{x}c^x(1 - c)^{n-x} + \beta\sum_{x=0}^{r}\binom{n}{x}\, d^x(1 - d)^{n-x} + \gamma]^{-1}. \tag{4.10}$$

If $r = 0$, four pieces of probability may be distinguished, namely, $K'\alpha(1 - c)^n$ on the point $(0, c)$, $K'\beta(1 - d)^n$ distributed continuously along the line segment (d, v) for $d \leqq v \leqq 1$, $K'(\alpha + \gamma)[1 - (1 - c)^n]$ distributed continuously along the line segment $(0, v)$ for $0 \leqq v \leqq c$, and $K'\gamma(1 - c)^n$ distributed continuously along the line segment $(0, v)$ for $c \leqq v \leqq 1$. Similarly, if $r = n$, the four pieces of probability are $K'\beta\, d^n$ on the point $(d, 1)$, $K'\alpha c^n$ distributed continuously along the line segment (u, c) for $0 \leqq u \leqq c$, $K'(\beta + \gamma)[1 - d^n]$ distributed continuously along the line segment $(u, 1)$ for $d \leqq u \leqq 1$, and $K'\gamma\, d^n$ distributed continuously along the line segment $(u, 1)$ for $0 \leqq u \leqq d$. Formula (4.10) remains valid in the cases $r = 0$ and $r = n$.

It is now a straightforward exercise to compute upper and lower probabilities for any interval $[a, b]$ of values of p. Only one of the unfortunately many cases will be given in detail, because of its role in the hypothetical drug screening example used to motivate the prior information (4.4). By assigning the various pieces of probability correctly, it is easily checked that

$$P_*([d, 1]) = K'[\beta\binom{n}{r}\, d^r(1 - d)^{n-r} + (\beta + \gamma)\textstyle\sum_{x=0}^{r-1}\binom{n}{x}\, d^x(1 - d)^{n-x}] \tag{4.11}$$

while

$$(4.12)\quad P^*([d, 1]) = P_*([d, 1]) + K'\gamma[1 - \textstyle\sum_{x=r+1}^{n} \binom{n}{x} c^x(1 - c)^{n-x} - \sum_{x=0}^{r-1} \binom{n}{x} d^x(1 - d)^{n-x}],$$

for $0 < r < n$ and K' given by (4.10). Formulas (4.11) and (4.12) also apply when $r = 0$ or $r = n$ if sums with empty ranges are set to zero.

In the example $d = .75$ and $c = .75 - \epsilon$ or, in other words, c is regarded as d in the limit approached from below. Formula (4.12) may be simplified in such a situation to read

$$(4.13)\quad P^*([d, 1]) = P_*([d, 1]) + K'\gamma\binom{n}{r} d^r(1 - d)^{n-r}.$$

NUMERICAL EXAMPLE. Using $c = .75$- and $d = .75$, the prior information (.4) and $n = 6$ the following upper and lower probabilities of an effective drug were computed from (4.10), (4.11) and (4.13).

r	$P^*([.75, 1])$	$P_*([.75, 1])$
0	.000024	.0000024
1	.00047	.000069
2	.0038	.00080
3	.0177	.0053
4	.0553	.0236
5	.1420	.0867
6	.3843	.3228

For comparison, the case $\alpha = \beta = \frac{1}{2}$ was computed. This means prior probabilities of $\frac{1}{2}$ for effective and for not effective.

r	$P^*([.75, 1])$	$P_*([.75, 1])$
0	.0002	.0002
1	.0046	.0046
2	.0364	.0364
3	.1497	.1497
4	.3594	.3594
5	.6062	.6062
6	.8489	.8489

Finally, the case $\gamma = 1$ of no prior information is displayed.

r	$P^*([.75, 1])$	$P_*([.75, 1])$
0	.0002	0.0
1	.0046	.0002
2	.0376	.0046
3	.1694	.0376
4	.4661	.1694
5	.8220	.4661
6	1.0	.8220

The prior information used in the first set of inferences loads the scales away from [.75, 1], relative to the third set, as might be expected. Thus the experimenter who trusted his prior information could reject a drug even with 4 positive

responses out of 6 and have at most a 6% chance of error. Without the prior information, rejection of a drug with 3 positive responses out of 6 corresponds to an upper 17% chance of error.

The second set of inferences incorporates rather neutral prior information. The coincidence of upper and lower probabilities is a consequence of the general fact that, if one source provides $P^*(A) = P_*(A)$ for some event A, then $P^*(A) = P_*(A)$ still holds after combination with further independent sources. As might be expected from neutral prior information, the coincident probabilities of the second set lie between the upper and lower probabilities of the third set.

REFERENCES

DEMPSTER, A. P. (1966). New approaches for reasoning towards posterior distributions based on sample data. *Ann. Math. Statist.* **37** 355–374.

DEMPSTER, A. P. (1967a). Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.* **36** 325–339.

DEMPSTER, A. P. (1967b). Upper and lower probability inferences based on a sample from a finite univariate population. *Biometrika* **54** 515–528.

DEMPSTER, A. P. (1968). A generalization of Bayesian inference. To appear in *J. Roy. Statist. Soc. Ser. B* **30**.