Biometrika (1961), **48**, 3 *and* 4, *p*. 281
Printed in Great Britain

Bayesian sequential analysis

BY G. B. WETHERILL
Birkbeck College

1. DISTRIBUTIONS CLOSED UNDER SAMPLING

Suppose we have batches of items presented for acceptance inspection, and each batch is to be sentenced independently by a sequential procedure. Let the result of inspection of an item be a random variable x with a density function $\phi(x|\theta)$, where $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$. In the simplest case where items are classified as effective or defective, $\phi(x|\theta)$ is a binomial distribution with probability of a defective equal to θ.

In acceptance inspection, the prior knowledge is represented by the process curve, which is a distribution of θ specifying the relative frequency with which batches are produced with quality θ.

In general let the prior knowledge be represented by a density function which we shall call a parameter distribution, denoted $\xi(\theta|\alpha)$, depending on a known constant α (where α may be a vector). For convenience throughout §§ 1 to 3 of the paper, equations will be formulated for continuous density functions $\xi(\theta|\alpha)$ and $\phi(x|\theta)$. A fully rigorous discussion of the theory will not be attempted.

After observations $x_1, x_2, \ldots, x_n$, have been taken the posterior distribution of θ is

$$\frac{\xi(\theta|\alpha).\phi(x_1|\theta).\phi(x_2|\theta)\ldots\phi(x_n|\theta)}{\int \xi(\theta|\alpha).\phi(x_1|\theta).\phi(x_2|\theta)\ldots\phi(x_n|\theta)\,d\theta}.$$

Suppose now that this posterior distribution is of the same functional form as $\xi(\theta|\alpha)$, say $\xi(\theta|\beta)$. Our total knowledge at any point is then represented as a distribution of this type. We shall call a distribution possessing this property a distribution closed under sampling, with respect to observations having a distribution $\phi(x|\theta)$.

*Formal definition (due to Barnard)**

A parametric family $\xi(\theta|\alpha)$ is said to be closed under sampling with respect to observations having a density function $\phi(x|\theta)$, if for all $x_1, x_2, \ldots, x_n$, for which $\phi(x_i|\theta) \neq 0$, there exists a β such that

$$\xi(\theta|\beta) = \frac{\xi(\theta|\alpha).\phi(x_1|\theta).\phi(x_2|\theta)\ldots\phi(x_n|\theta)}{\int \xi(\theta|\alpha).\phi(x_1|\theta).\phi(x_2|\theta)\ldots\phi(x_n|\theta)\,d\theta}, \tag{1}$$

where $\beta = \beta(\alpha, x_1, x_2, \ldots, x_n)$. To make this clear we shall consider one or two examples.

(*a*) Consider inspection of batches by attributes, and suppose that items are of two kinds, effective and defective. Suppose that the probability of a defective in each batch is p, and that p varies from batch to batch according to the Beta distribution $kp^s(1-p)^t$. This is our

* Dr M. Stone has independently studied the idea of distributions closed under sampling in an unpublished Ph.D. thesis.

parameter density function representing our prior knowledge. The posterior distribution after observing r defectives in n items is

$$\frac{kp^s q^t . p^r q^{n-r}}{k\int_0^1 p^s q^t . p^r q^{n-r} dp} = k' p^{r+s} q^{n-r+t}, \quad \text{where} \quad q = 1-p,$$

which is also a Beta distribution. The family $\xi(p|\beta)$ is therefore the class of all Beta distributions $kp^{\beta_1}q^{\beta_2}$, where $\beta = \{\beta_1, \beta_2\}$, and $\beta_1 = r+s$, $\beta_2 = n-r+t$.

Thus a Beta distribution is closed under sampling with respect to binomially distributed observations. The total knowledge at any point in the sampling can be represented in a two-dimensional space with co-ordinates (β_1, β_2).

(*b*) Suppose that our observations are distributed according to any probability distribution $\phi(x|\theta)$, where θ varies from batch to batch, but may only take certain given values θ_i $(i = 1, 2, \ldots, k)$. Our prior probability is then $(a_1, a_2, \ldots, a_k)$, where $\sum_1^k a_i = 1$. The posterior distribution is of the same form $(b_1, b_2, \ldots, b_k)$, where

$$b_i = \frac{a_i \prod_j \phi(x_j|\theta_i)}{\sum_{i=1}^k a_i \prod_j \phi(x_j|\theta_i)} \quad \text{and} \quad \sum_1^k b_i = 1.$$

The set of values θ does not change, and sampling merely alters the set of probabilities a_i. Therefore a prior distribution, which takes its values at a finite set of points θ, is closed under sampling with respect to observations distributed according to any probability function $\phi(x|\theta)$. The family $\xi(\theta|\alpha)$ is the class of all discrete distributions where the random variable θ may only take on the values θ_i $(i = 1, 2, \ldots, k)$.

(*c*) If our observations are distributed according to the Poisson distribution

$$\phi(x|\theta) = e^{-\theta}.\theta^x/x!.$$

and if our prior knowledge is given by the Beta-type distribution

$$\xi(\theta|\alpha) = K\theta^s(1-\theta)^t,$$

where $0 < \theta < 1$, and the case $t = 0$ corresponds to the improper distribution $k\theta^s$ over all positive θ: then after taking observations $x_1, \ldots, x_n$, the posterior distribution is

$$\frac{K\theta^s(1-\theta)^t . e^{-n\theta} . \theta^m / \Pi_j x_j!}{K\int \theta^s(1-\theta)^t . e^{-n\theta} . \theta^m d\theta / \Pi_j x_j!}, \quad \text{where} \quad m = \Sigma_j x_j,$$

$$= \frac{\theta^{s+m}(1-\theta)^t e^{-n\theta}}{\int \theta^{s+m}(1-\theta)^t e^{-n\theta} d\theta}.$$

This is not of the same form as $\xi(\theta|\alpha)$, and hence here we have an example of a distribution not closed under sampling. It can be made closed by replacing the Beta distribution by the probability density

$$k\theta^{\alpha+x}(1-\theta)^\beta e^{-\gamma\theta}/x!$$

and if we use this as our prior distribution, we see that it is closed under sampling with respect to observations having a Poisson distribution.

The great simplification introduced by considering distributions closed under sampling is that the problem of sequential decision is reduced from dealing with a random walk in infinitely many dimensions, to being a random walk in finitely many.

2. The ξ-space

If we have a parameter distribution $\xi\,(\theta|\alpha)$, where α may be a vector, and this distribution is closed under sampling, we can define some space ξ, in which each member of the family of distributions of the type $\xi(\theta|\alpha)$ is represented by a point. Our total knowledge at any stage in the sampling is a point in this space.

For examples of ξ-space we refer to examples (*a*) and (*b*) in § 1. The ξ-space for (*a*) is a plane with axes β_1 and β_2 and that for (*b*) is a k-dimensional simplex with vertices at the unit points.

Now, assuming that we are dealing with a prior distribution which is closed under sampling, consider some form of sequential inspection in which there are only two possible terminal decisions—to accept or reject the batch under inspection. Suppose also we can define the losses associated with making the terminal decisions when the true quality of the batch is specified by some particular value of θ. We define the risk of a terminal decision at any point in the ξ-space to the expected value of the loss with respect to θ. One terminal decision is preferred to the other if the risk associated with making this terminal decision is less than the risk associated with making the other terminal decision. If the quality of the batches remains unaltered during inspection, then the risk associated with making a terminal decision depends only upon the posterior distribution function, and hence on the position in the ξ-space. We can therefore associate with each point in the ξ-space the risks of taking either of the two kinds of terminal decision, and depending upon which of these is the smaller at each point, mark the ξ-space out into two regions, one in which acceptance is preferred and the other in which rejection is preferred. Clearly, apart from possible discontinuities, the boundary between these regions is the locus of all points such that the risks of the two kinds of terminal decision are equal. This boundary will be referred to as the neutral boundary.

This division of the ξ-space into terminal decision regions does not taking sampling into account. In fact, we shall have extremes in the ξ-space where the expected quality of the batch is so good or so bad that it should be accepted or rejected outright, without any sampling. In between there will be a region of doubt where it pays to take further observations. We thus have three regions, a continuation region and two terminal decision regions, each defined uniquely by position in the ξ-space.

Four boundaries are of interest: the boundaries between the terminal decision regions, and the continuation region; the boundary between the two terminal decision regions, and the boundary at which the continuation region and the two-terminal decision regions meet. Clearly, some of these boundaries may not exist, for example, in a likelihood ratio sequential procedure, the two terminal decision regions do not meet. We proceed to derive equations for each of these four boundaries.

3. Equations for the boundaries

We shall refer to the two terminal decisions as decisions 1 and 2; the boundary in the ξ-space between the decision 1 (or 2) region and the other terminal decision region or the continuation region, will be referred to as the decision 1 (or 2) boundary. Let the losses of taking decisions 1 and 2 when θ specifies the true quality be denoted $W_1(\theta)$ and $W_2(\theta)$, respectively, where the unit of losses is the cost of a sample, which is assumed constant.

We now consider the boundary where the three decision regions meet; this is often a point and will then be referred to as the meeting point. Interest centres in the meeting point because if it exists and can be determined, the next section describes a method by which, in principle, the optimum terminal decision boundaries can be worked out backwards from the meeting point.

The neutral boundary is the locus of points such that the decision risks (where risk is defined as expected loss) of taking either terminal decision are equal, and this is given by α satisfying

$$\int \xi(\theta|\alpha)\,[W_2(\theta)-W_1(\theta)]\,d\theta = 0. \tag{2}$$

Clearly, if the decision 1 and 2 boundaries meet, they will meet on the neutral boundary.

Suppose now that the decision 1 and 2 boundaries meet the neutral boundary at the same point, then at this point the risk of making a terminal decision is equal to the cost of one more observation, plus the terminal decision risks associated with this further observation (one more observation from the meeting point can only lead to a terminal decision). If at the meeting point we make decision 2 (the resulting equation is the same whichever terminal decision is used here), then for a given θ, the risk of a terminal decision at the meeting point is simply $W_2(\theta)$. After one more observation is taken, the terminal decision risks are

$$q(\alpha,\theta)\,W_1(\theta)+[1-q(\alpha,\theta)]\,W_2(\theta),$$

where $q(\alpha,\theta)$ is the probability that for an α on the meeting point boundary, one observation leads to points in terminal decision 1 region. Thus for the meeting point, α should satisfy

$$\int [q(\alpha,\theta)\,W_1(\theta)+(1-q(\alpha,\theta))\,W_2(\theta)+1]\,\xi(\theta|\alpha)\,d\theta = \int \xi(\theta|\alpha)\,W_2(\theta)\,d\theta$$

which reduces to

$$\int q(\alpha,\theta)\,[W_2(\theta)-W_1(\theta)]\,\xi(\theta|\alpha)\,d\theta = 1. \tag{3}$$

Two points remain to be clarified, first, that the two terminal decision boundaries do in fact meet the neutral boundary at the same point, and secondly, that the locus of α satisfying equations (2) and (3) is a complete definition of the meeting point boundary. Both of these points can be established in the following way. First, we derive equations for the decision 1 and 2 boundaries. Simultaneous solution of both of these equations defines the boundary where the two terminal decision boundaries meet, and it will be seen that the resulting equations can be reduced to equations (2) and (3) above.

Let points on the decision 1 and 2 boundaries be denoted by α' and α'', respectively. For sampling starting from position α in the continuation region of the ξ-space, with given boundaries, and the value of θ specified, let $S(\alpha,\theta)$ and $A(\alpha,\theta)$ be the expected further sample size, and the probability that sampling eventually terminates with decision 1, respectively. For points α' and α'' on the terminal decision boundaries, $S(\alpha',\theta)$, $A(\alpha',\theta)$, etc., are defined to be the values obtained if sampling is continued, but otherwise with the same terminal decision boundaries.

Consider any point α' on the decision 1 boundary, then the risk of continuing sampling is equal to the risk of taking terminal decision 1 immediately, see Barnard (1954).

To decide immediately has risk $\int \xi(\theta|\alpha')\, W_1(\theta)\, d\theta.$

To continue sampling has risk

$$\int [S(\alpha',\theta) + (1-A(\alpha',\theta))\, W_2(\theta) + A(\alpha',\theta)\, W_1(\theta)]\, \xi(\theta|\alpha')\, d\theta.$$

The equation for the value(s) of α' on the decision 1 boundary is obtained by equating these expressions,

$$\int \xi(\theta|\alpha')\, W_1(\theta)\, d\theta = \int [S(\alpha',\theta) + (1-A(\alpha',\theta))W_2(\theta) + A(\alpha',\theta)\, W_1(\theta)]\xi(\theta|\alpha')\, d\theta \tag{4}$$

which reduces to

$$\int [S(\alpha',\theta) + (1-A(\alpha',\theta))\, \{W_2(\theta) - W_1(\theta)\}]\, \xi(\theta|\alpha')\, d\theta = 0. \tag{5}$$

Similarly for points α'' on the decision 2 boundary we obtain the equation

$$\int \xi(\theta|\alpha'')\, W_2(\theta)\, d\theta = \int [S(\alpha'',\theta) + A(\alpha'',\theta)\, W_1(\theta) + (1-A(\alpha'',\theta))\, W_2(\theta)]\, \xi(\theta|\alpha'')\, d\theta \tag{6}$$

which reduces to $$\int [S(\alpha'',\theta) + A(\alpha'',\theta)\, \{W_1(\theta) - W_2(\theta)\}]\, \xi(\theta|\alpha'')\, d\theta = 0. \tag{7}$$

It was pointed out above that the meeting-point boundary will be defined by $\alpha = \alpha' = \alpha''$ simultaneously satisfying equations (5) and (7). If such a solution exists, then provided the next step from points on this boundary do not land back in the continuation region (see §§4 and 8 for discussion of this and other assumptions), we have $S(\alpha,\theta) = 1$ and $A(\alpha,\theta) = q(\alpha,\theta)$. On substituting these values in equations (5) and (7), equation (7) becomes equation (3), and equation (5) can be reduced to equation (2). Thus if a solution to equations (2) and (3) exists, then this solution defines the locus of points where both the terminal decision boundaries meet the neutral boundary.

Equations (5) and (7) can be transformed into each other by using the equation for the neutral boundary. Therefore, no points in the ξ-space exist where the neutral boundary meets one only of the terminal decision boundaries.

Equations (5) and (7) are difficult to solve because of the complex form of $S(\alpha,\theta)$ and $A(\alpha,\theta)$. On the other hand, the function $q(\alpha,\theta)$ is usually very simple, so that the neutral boundary and the meeting-point boundary can be determined. The next section describes the numerical procedure by which the terminal decision boundaries can be worked out backwards from the meeting-point boundary.

Equations (5) and (7) can usually be solved and used in the following special case. Consider points α' and α'' on the terminal decision boundaries which are such that they lead only to points on the meeting-point boundary or to terminal decision points. For such points, $S(\alpha',\theta)$ and $S(\alpha'',\theta)$ are both unity, $A(\alpha',\theta)$ and $A(\alpha'',\theta)$ are of the form $q(\alpha,\theta)$, so that for points on the decision 2 boundary equation (3) holds, with a similar equation for the decision 1 boundary.

4. Determination of the boundaries from the meeting-point boundary

If we are considering a point β in the ξ-space, then the terminal decision risk at this point is

$$R_d(\beta) = \min\left\{\int \xi(\theta|\beta)\, W_1(\theta)\, d\theta,\ \int \xi(\theta|\beta)\, W_2(\theta)\, d\theta\right\} \tag{8}$$

and the continuation risk is

$$R_c(\beta) = 1+\iint \xi(\theta|\beta)\,\phi(x|\theta)\,R(\beta')\,dx\,d\theta, \tag{9}$$

where β' denotes the posterior distribution for a prior distribution specified by β and an observation x, and where $R(\beta)$ is the risk at β, defined

$$R(\beta) = \min\{R_c(\beta), R_d(\beta)\} \tag{10}$$

so that if $R(\beta) = R_c(\beta)$, β is a continuation point, and if $R(\beta) = R_d(\beta)$, β is a terminal decision point.

In this section we make the following assumptions (in § 8 these assumptions are discussed again in some detail).

Assumption A. Given any particular origin in the ξ-space, if any point can be reached in n steps, then it cannot be reached in any other than n steps.

Assumption B. Given any particular origin in the ξ-space, of all points which can be reached in n steps, there is at most one which satisfies equation (2) for the neutral boundary. (This excludes such cases as testing two normal populations with unequal variances. This assumption is not necessary, and is introduced here for simplification.)

Assumption C. If a meeting-point boundary exists, then for a given origin in the ξ-space the maximum of the sample sizes associated with points on this boundary has some finite value, say N. (See § 8 for a further explanation of this assumption.)

With regard to assumption C, no points of the subset designated by N—other than the one on the meeting-point boundary—can be continuation points. This follows, since for a continuation point, it must be possible to proceed sampling and reach either terminal decision region, and this implies the existence of a point on the meeting-point boundary for some $N' > N$. Therefore, we can write down all the risks for the points of the subset designated by N. Further, points of the subset designated by $(N-1)$ all lead to points of the subset N, the risks of which are known. We can therefore apply equations (8), (9), and (10) to all the points of the subset $(N-1)$, calculate their risks and classify them as decision or continuation points. This process can now be repeated for the points of the subset $(N-2)$ and so on.

This procedure is quite simple in practice, and in principle can always be done numerically, by the method known as dynamic programming, see Bellman (1957). A large amount of computing will often be required, but the problem is ideally suited to an electronic computer.

5. Example 1. The mixed binomial distribution

Suppose that the distribution $\phi(x|\theta)$ is a binomial distribution with a probability θ of a defective item, and suppose that the parameter distribution for θ is a k-point binomial distribution with $k \geqslant 3$, so that prob $(\theta = \theta_i) = a_i$ $(i = 1, 2, \ldots, k)$, where $\sum_1^k a_i = 1$. The terminal decisions are to accept or reject the batch under inspection, and these are denoted decisions 1 and 2 respectively. The loss associated with taking decision i when θ_j is true is written W_{ij}.

Equation (2) gives the neutral boundary

$$\sum_1^k a_i \theta_i^r (1-\theta_i)^{n-r} (W_{1i} - W_{2i}) = 0, \tag{11}$$

where n is the number of items inspected, and r the number of defectives found. Similarly, in equation (3) the function $q(\alpha, \theta_i)$ is the probability, given θ_i, that one more observation from the meeting point leads to acceptance. The next observation can be either a good item or a bad item, and both lead to terminal decision points. One more good item can only increase the risk of rejection and decrease the risk of acceptance, so that $q(\alpha, \theta_i) = 1-\theta_i$, and equation (3) is

$$\sum_1^k a_i \theta_i^r (1-\theta_i)^{n-r} [1+(1-\theta_i)(W_{1i}-W_{2i})] = 0. \tag{12}$$

Wetherill (1960) has pointed out that equations such as these take on a particularly simple form if we assume the prior distribution to be such that $\theta_i/(1-\theta_i) = \lambda^i$, for $i = 1, 2, \ldots, k$, and it is also shown that these equations are easily solved by Horner's method. If we make this assumption, equations (11) and (12) for the meeting point become

$$\sum_1^k a_i (1-\theta_i)^n x^i (W_{1i}-W_{2i}) = 0 \tag{13}$$

and

$$\sum_1^k a_i (1-\theta_i)^n x^i [1+(1-\theta_i)(W_{1i}-W_{2i})] = 0, \tag{14}$$

where $x = \lambda^r$.

Numerical example. Suppose that we have $\lambda = \frac{1}{3}$, so that $\theta_1 = \frac{1}{4}$, $\theta_2 = \frac{1}{10}$, $\theta_3 = \frac{1}{28}$; also suppose that $W_{1j} = \max\{0, 500(\theta_j - 0{\cdot}07)\}$ and $W_{2j} = \max\{0, 500(0{\cdot}07-\theta_j)\}$, so that $W_{11} = 90$, $W_{12} = 15$, $W_{23} = 17{\cdot}14$; and take the values of a_i to be $a_1 = 0{\cdot}05$, $a_2 = 0{\cdot}25$, and $a_3 = 0{\cdot}70$.

For this example, equations (13) and (14) are quadratic equations, and they have a common solution for an x corresponding to $r \simeq 1{\cdot}75$, $n = 15$. (For the method of solution see Wetherill (1960).) Because of the discreteness of the problem, the last point in the continuation region must correspond to integral n and r both less than or equal to the values given for the meeting point, so that for this example, $r = 0$ or 1, and $n \leqslant 15$. We can therefore proceed by determining the position of the acceptance and rejection boundaries for $r = 1$. This can be done by the method outlined in § 4, or alternatively it is quite simple for the stage immediately prior to the meeting point, to employ equations (5) and (7).

Consider equation (5) for the acceptance boundary. When $r = 1$ one further observation leads to a terminal decision, so that $S(\alpha, \theta_i) = 1$, and $A(\alpha, \theta_i) = 1-\theta_i$, and the equation becomes

$$\sum_1^k a_i \theta_i^1 (1-\theta_i)^{n-1} [1+\theta_i(W_{1i}-W_{2i})] = 0,$$

from which the value of n giving the position of the acceptance boundary at $r = 1$ can be determined. At the rejection boundary for $r = 1$, one further good item will not always lead to acceptance (if there is a point in the continuation region), and therefore $S(\alpha, \theta_i) \geqslant 1$, and $A(\alpha, \theta_i) \leqslant 1-\theta_i$. When equality holds in these expressions, equation (7) for the rejection boundary reduces to equation (12), and it follows that for $r = 1$, all points for which the expression of equation (12) is positive are certainly rejection points. Any points for $r = 1$ not thus classified by these two boundaries, may be either continuation points or rejection points, and these can be checked numerically by the method given below, or by using an equation similar to (12) in which we may have to introduce a more complicated function for $S(\alpha, \theta_i)$. For this example ($n = 9, r = 1$) is the last continuation point, and ($n = 10, r = 1$) and ($n = 7, r = 1$), are acceptance and rejection points, respectively.

For the problem outlined above, the procedure for working back to find the optimum decision boundaries is as follows. Denote the posterior probability of θ_i at any point (n, r) by a'_i, so that at (n, r)

$$a'_i = a_i \theta_i^r (1-\theta_i)^{n-r} \Big/ \sum_1^k a_i \theta_i^r (1-\theta_i)^{n-r}. \tag{15}$$

(Actually we should write this as $a'_i(n, r)$, but it simplifies the notation to use simply a'_i.) The terminal decision risk at (n, r), $R_d(n, r)$, is defined

$$R_d(n, r) = \min \left\{ \sum_1^k a'_i W_{1i}, \sum_1^k a'_i W_{2i} \right\}. \tag{16}$$

The probability of a bad item at (n, r) is $\sum_1^k a'_i p_i = b(n, r)$, say. Then the risk at any point is

$$R(n, r) = \min \{R_d(n, r), R_c(n, r)\}, \tag{17}$$

where $R_c(n, r)$ is the continuation risk defined by

$$R_c(n, r) = 1 + b(n, r) R(n+1, r+1) + [1 - b(n, r)] R(n+1, r). \tag{18}$$

On using these equations for the above numerical example we have the following sequential scheme:

Accept the batch when sampling reaches $n = 9$, $r = 0$.

Reject the batch when sampling reaches $r = 1$, and $n \leqslant 7$.

If sampling reaches $n = 8$, $r = 1$, inspect another item, and reject if it is bad.

If sampling reaches $n = 9$, $r = 1$, inspect another item, reject if it is bad, and accept if it is good.

The total risk of this scheme is 4·96, as against the risk of 7·2 of accepting without sampling.

6. Example 2. A beta prior distribution

Suppose that the distribution $\phi(x|\theta)$ is a binomial distribution with a probability θ of a defective item, and that the parameter distribution for θ is a Beta distribution $\theta^{s-1}(1-\theta)^{t-1}/\beta(s, t)$. Let the decision losses associated with wrongly accepting or rejecting a batch be $k|\theta - \theta_0|$.

Denote the number of bad items found during inspection by y, and the number of good items found by x, then equations (5) and (7), or alternatively (2) and (3), reduce to

$$(y+s)(1-\theta_0)/\theta_0 = x+t, \tag{19}$$

$$(y+s+x+t+1)(y+s+x+t)k^{-1} = (x+t)[\theta_0(x+t+1) - (1-\theta_0)(y+s)]. \tag{20}$$

From these equations it follows that the meeting point is at (X, Y), where

$$Y = k\theta_0^2(1-\theta_0) - \theta_0 - s, \tag{21}$$

$$X = (Y+s)(1-\theta_0)/\theta_0 - t. \tag{22}$$

Champernowne (1953) has written a paper on the most economic boundaries for the Beta distribution, but the values he gives for the meeting points are very different from the values given by the above equations. Actually Champernowne tabulates boundaries, not meeting points, and many of the differences are accounted for when we realize that the farthest point reachable in sampling may be considerably nearer the origin than the meeting point of the optimum boundaries.

Suppose the meeting point is at (x, y), then it may well be that at $(x, y-1)$ it pays to accept, and at $(x-1, y)$ it pays to reject, while $(x-1, y-1)$ is a continuation point. Then clearly sampling will never reach (x, y), for if $(x-1, y-1)$ is reached, the next item sampled will always lead to a decision. In order to find the last reachable point in sampling we examine the points immediately off the neutral line, towards the origin from the meeting point of the boundaries, to see when they start being continuation points. If we are examining to see if (x, y) is the last reachable point in sampling, we compare the continuation and terminal decision risks for $(x, y-1)$ and $(x-1, y)$; if one (or both) of these is a continuation point, then (x, y) can be reached in sampling.

For any point, the terminal decision and continuation risks can be compared by comparing the left- and right-hand sides of either equation (4) or (6) (or alternatively using the expressions of equations (5) and (7)). Now suppose (x, y) is a continuation point, and both $(x+1, y)$ and $(x, y+1)$ are terminal decision points, then for $(x-1, y)$, given that sampling is continued, the probability of acceptance is $(1-\theta_0)^2$, so that the expected further sample size is $1.\theta+2(1-\theta) = 2-\theta$. Thus for α corresponding to $(x-1, y)$, we have

$$A(\alpha, \theta) = (1-\theta)^2 \quad \text{and} \quad S(\alpha, \theta) = 2-\theta.$$

We now substitute into equation (6), and if the left-hand side is greater than the right-hand side then $(x-1, y)$ is a continuation point, otherwise it is a terminal decision 2 point. The point $(x, y-1)$ can be examined in a similar way, and we proceed in this manner, examining points immediately off the neutral line until we find one or both of such points to be continuation points. We shall then have determined the last reachable point in sampling, as described above.

An example has been worked out by this method for $\theta_0 = \frac{1}{2}, k = 200$, so that the boundaries and decision risks are symmetrical. Equations (21) and (22) give the meeting point as $X = Y = 23{\cdot}5$. The last reachable point in sampling is (6, 6), since by the above method (5, 6) and (6, 5) were found to be the farthest off-neutral line points which are continuation points. Champernowne has given boundaries for this case, and they give the same point, (6, 6), as the last reachable point (his boundaries cease here). For this example, the boundaries determined by my method agree with those given by Champernowne with the exception of (2, 0) which he gives as a continuation point, while he gives (0, 2) as a decision point (which is correct). Clearly, the boundaries must be symmetrical. This is the only difference I have found between my boundaries and Champernowne's.

7. Example 3. A normal prior distribution

Suppose that the distribution $\phi(x|\theta)$ is normal with a mean m and unit variance, and that the parameter distribution for m is normal with mean μ and known variance σ^2. The two decisions to be made are to accept (decision 1), or to reject (decision 2), the batch under consideration, and we shall take the loss function to be $\alpha\,|m-m_0|$.

Clearly, the normal parameter distribution is closed under sampling with respect to normally distributed observations. If a sample of n observations has a mean $\bar{x}_n$, the posterior distribution of m is well known to be normal with a mean $(n\sigma^2\bar{x}_n+\mu)/(n\sigma^2+1)$ and variance $\sigma^2/(1+n\sigma^2)$, and this will be written $N(\mu_n, \sigma_n)$.

The neutral line is therefore

$$\frac{\alpha}{\sigma_n\sqrt{(2\pi)}}\int_{-\infty}^{\infty}(m-m_0)\exp-\frac{1}{2}\left\{\frac{m-\mu_n}{\sigma_n}\right\}^2 dm = 0,$$

which is

$$\mu_n = (n\sigma^2\bar{x}_n+\mu)/(n\sigma^2+1) = m_0, \tag{23}$$

or

$$\bar{x}_n = m_0+(m_0-\mu)/n\sigma^2. \tag{24}$$

Thus for all n, given a position (μ_n, n) on the neutral line, the value of x_{n+1} required to remain on the neutral line at $(n+1)$ is m_0.

In order to determine the meeting point of the boundaries we must evaluate equation (7), and for this example $q(\alpha,\theta) = \Phi(m_0|m)$, where $\Phi(x|m)$ is the normal probability integral with mean m and unit variance. Equation (7) becomes

$$\frac{\alpha}{\sigma_n\sqrt{(2\pi)}}\int_{-\infty}^{\infty}(m-m_0)\,\Phi(m_0|m)\exp-\frac{1}{2}\left\{\frac{m-\mu_n}{\sigma_n}\right\}^2 dm = 1. \tag{25}$$

The meeting point must be on the neutral line (23), and using this, equation (25) can be written

$$\frac{\alpha\sigma_N}{\sqrt{(2\pi)}}\int_{-\infty}^{\infty}\Phi(m_0|m)\,d\exp-\frac{1}{2}\left\{\frac{m-m_0}{\sigma_N}\right\}^2 = 1,$$

where N denotes the sample size at the meeting point. This equation is clearly

$$\frac{\alpha\sigma_N^2}{\sqrt{(2\pi)}\sqrt{(1+\sigma_N^2)}} = 1. \tag{26}$$

On substituting for σ_N^2, we have a quadratic equation for N,

$$\alpha^2\sigma^4 = 2\pi[1+(N+1)\sigma^2][1+N\sigma^2]$$

and the only root of this which can be positive is

$$N = \sqrt{\left(\frac{\alpha^2}{2\pi}+\frac{1}{4}\right)}-\frac{1}{2}-\frac{1}{\sigma^2}. \tag{27}$$

The meeting point is therefore specified by $\mu_N = m_0$, where N is given by (27). The sample mean $\bar{x}_N$ corresponding to μ_N is obtained from (24) and (27),

$$\bar{x}_N = m_0+(m_0-\mu)/\{\sigma^2(\tfrac{1}{2}\alpha^2+\tfrac{1}{4})-\tfrac{1}{2}\sigma^2-1\}. \tag{28}$$

When N is large, we have approximately

$$N \simeq \frac{\alpha}{\sqrt{(2\pi)}}-\frac{1}{\sigma^2} \tag{29}$$

hence

$$\bar{x}_N \simeq m_0+\sqrt{(2\pi)}\,(m_0-\mu)/\alpha\sigma^2. \tag{30}$$

These functions appear to be of the right form. The sample size at the meeting point, N, is an increasing function of the losses α, and also of σ^2 which represents the precision of the prior distribution. The effect of the parameter μ of the prior distribution on $\bar{x}_N$ decreases with increasing N, and also decreases with increasing variance σ^2.

8. Conditions and assumptions

It is clearly of interest to know what are the necessary and sufficient conditions that (i) the procedure for working back to find the optimum boundaries can be performed, and (ii) that a meeting point exists. We shall reconsider assumptions A, B and C of § 4, and also consider a fourth assumption, which we shall label D, that no sequence of sampling can result in repeating a posterior distribution which has been reached before.

First, it is clearly not necessary for there to be a meeting point for (i) to be possible, as it is sufficient if there is a 'last reachable point' in the continuation region, in the sense described in example 2. Thus if the decision boundaries tend asymptotically to the neutral line, (i) will be possible for discrete distributions $\phi(x|\theta)$. Further, assumption A is not necessary either, for (i) is possible if sampling can jump a stage from, say, a point in the set n to a point in the set $(n+2)$; however, it must not be possible for a point to lead to points in previous subsets, and assumption D is sufficient to ensure this, but may not be necessary.

Assumption B is really irrelevant, the essential point being that once a meeting point has been determined, we wish to be able to write down the risks for all points in the subset containing the meeting point. If what has been described as the meeting point is in fact a line or surface containing a (possible infinite) number of points, then for (i) to be possible all that is necessary is that a 'last reachable point' exists. If the neutral line has two or more branches, (i) is possible provided sampling is closed on each branch.

Assumption C is introduced to cover the case where the ξ-space has at least three dimensions, so that the meeting-point boundary is a line or surface, and it may be that for a given origin in the ξ-space, the number of steps taken to reach this boundary is finite for only part of the ξ-space, and otherwise infinite. For this situation, unless a last reachable point exists, (i) will be possible for at most only parts of the ξ-space. For an example of this consider testing among three binomial populations with probabilities θ_0, θ_1 and θ_2, where $\theta_0 > \theta_1 > \theta_2$, and three associated terminal decisions. There are parts of the ξ-space in which θ_0 or θ_2 have very small prior probabilities, so that the sequential procedure reduces to a test of two binomial populations, and clearly for the two separate cases of this, sampling may be closed for one and not for the other. For examples 1, 2 and 3 of this paper, the ξ-space is two-dimensional, and the question does not arise.

Now let us consider examples 1, 2 and 3. For example 3, the posterior distribution has a variance which is a decreasing function of n, and assumptions A, B and D hold. For example 2, similar remarks apply and for both examples 2 and 3, there exist explicit formulae for the meeting points.

The k-point binomial distribution of example 1 is more complicated. Although assumptions A, B and D all hold, there is not necessarily a meeting point, see Vagholkar & Wetherill (1960) and Wetherill (1957). The situation can be given briefly as follows. Suppose the loss functions are $k\,|\theta_i - \theta_0|$, then if we are sampling from a population with θ close to θ_0, there can be a stage in the sampling where all θ_i's except those either side of θ_0 are negligible. In such a situation, the optimum boundaries must approximate to those for a two-point binomial, which are open and parallel. To exclude this possibility it is necessary to put a condition on the loss functions, to ensure that it is not worth sampling for the two-point binomial scheme based on these two particular ordinates (i.e. those on either side of p_0) see Vagholkar & Wetherill (1960) and Vagholkar (1955).

It is interesting to notice that the two-point prior distribution (e.g. $k = 2$ in example 1), is such that for nearly every possible prior distribution there is a sequence of sample values (n, r) which will repeat it. This is easily proved, for the posterior distribution is a simple function of the likelihood ratio, λ, where

$$\log \lambda = r \log (\theta_1(1-\theta_2)/\theta_2(1-\theta_1)) + n \log ((1-\theta_1)/(1-\theta_2))$$

and there is a sequence of values of (n, r) which yield any given value of λ, provided the logarithms on the right-hand side have a rational ratio.

Now if a closed sequence of points is possible in the ξ-space, sampling can enter a closed loop, the sample size can be infinite, and the boundaries must be open. It would appear, therefore, that assumption C is necessary for a meeting point to exist.

I conjecture that, with 'reasonable' loss functions, and continuous probability densities $\xi(\theta|\alpha)$, assumption D is both necessary and sufficient for a meeting point to exist. (By 'reasonable' loss functions, I mean loss functions which are either monotonically increasing or decreasing functions of θ.)

The discussion in this section is incomplete, but it is hoped that sufficient has been said to give a lead to further work.

9. Generalizations

We have assumed throughout that we have only two terminal decisions, but there is no difficulty in extending the theory to three or more decisions. With three decisions, for instance, the ξ-space can be divided into four regions, and there are in general two neutral lines.

The restriction to prior distributions closed under sampling could be abandoned, but this would yield much more complex results which are more difficult to apply, and I suspect that there is not a corresponding gain in generality.

Sections 1 to 3 are based on work in my Ph.D. thesis, and I am greatly indebted to Prof. G. A. Barnard for his generous help throughout. I am also indebted to Prof. D. R. Cox for some helpful discussions during preparation of the paper, and to Prof. D. V. Lindley for some useful comments on an earlier draft of this paper.

References

Barnard, G. A. (1954). Sampling inspection and statistical decisions. *J. R. Statist. Soc.* Ser. B, **16**, 151–74.

Bellman, R. (1957). *Dynamic Programming*. Princeton University Press, Princeton, New Jersey.

Champernowne, D. G. (1953). The economics of sequential sampling procedures for defectives. *Appl. Statist.* **2**, 118–30.

Vagholkar, M. K. (1955). Application of statistical decision theory to sampling inspection schemes. Ph.D. thesis, University of London.

Vagholkar, M. K. & Wetherill, G. B. (1960). The most economical binomial sequential probability ratio test. *Biometrika*, **47**, 103–9.

Wetherill, G. B. (1957). Application of the theory of decision functions to sampling inspection with special reference to cost of inspection. Ph.D. thesis, University of London.

Wetherill, G. B. (1960). Some remarks on the Bayesian solution of the single sample inspection scheme. *Technometrics*, **2**, 341–52.