BAYESIAN BIO-ASSAY[1]

By Charles H. Kraft and Constance van Eeden

University of Minnesota

1. Introduction. The basic bio-assay problem can be stated as follows. Let F be a distribution function and let $Y = (Y_1, \cdots, Y_k)$ be a set of k independent random variables, each of which is binomial with parameters $(n_i, F(t_i))$. The numbers $t_1, \cdots, t_k$ are called dosage levels. The experimenter knows the n_i and the t_i, can observe Y and wants to make some inference about F.

The approach we discuss here is Bayesian, that is F is random and the distribution of Y just described is taken to be the conditional, given F, distribution of Y.

In the following sections a characterization of the class of all (a priori) distributions for F is given, the corresponding Bayes' estimates for a class of loss functions are found and the results of LeCam [5] are applied to show completeness of the closure of this class of estimates for a certain topology. A special case is given for which the estimates are explicitly computed.

2. A priori distributions for F. The set of distribution functions, F, is often taken to be a parametric family. In such a case an a priori distribution can be given for the parameters. It is, however, possible to be completely general. Namely let $D = \{d_i\}$ be a countable dense subset of the real line, and let a sequence of probability laws $\mathfrak{L}[F(d_1)]$, $\mathfrak{L}[F(d_2) \mid F(d_1)]$, $\cdots$, $\mathfrak{L}[F(d_n) \mid F(d_1), \cdots, F(d_{n-1})]$, $\cdots$ be given such that $P(F$, on D, is a distribution function$) = 1$. Then, defining for $x_0 \not\in D$, $F(x_0) = \lim_{x \to x_0^+, x \in D} F(x)$, defines P for F with $P(F$ is a distribution function$) = 1$. It is clear that this construction yields a separable process and, also, that any process that produces, with probability one, distribution functions will have such a separable representation.

The above given construction specifies an a priori distribution for F by giving the joint distribution of the ordinates of F at certain fixed abcissa. Another way to specify a distribution for F is to give, consistently, the joint distribution of the percentiles of F.

3. The loss functions. The loss functions we wish to consider are the following. Let $W(x)$ be an arbitrary (fixed) distribution function. If G, a non-decreasing bounded between 0 and 1 function on the real line, is the statistician's decision and F the distribution determining the distribution of Y, then the loss $L(F, G) = \int (F - G)^2 \, dW$. For this loss the Bayes' estimate is the conditional, given Y, expectation of the process. The proof follows immediately from the usual pointwise (in Y) construction of Bayes' procedures.

Received 19 March 1963; revised 10 August 1963.

[1] This research was supported in part by the Office of Naval Research under Grant Number Nonr 710(31), by the National Science Foundation under Grant Number G-19126 and by the National Institutes of Health under Grant Number 2G-43(C7+8).

4. Completeness of the set of Bayes' rules. The space of decisions G is the Helly space and is compact in the usual relative product topology (cf. e.g. Kelley [3] p 164). Decision functions, $\hat{F}$, are maps from the observation space to the decision space and the space of decision functions is given the product topology in the product space whose coordinate spaces are the spaces of decisions indexed by the finitely many points in the space of Y.

For the question of completeness of the set of rules it will be shown that the Assumptions 8, 9 and 10 of LeCam ([5] p 77) are satisfied and concluded, from his Theorem 4, that the closure of the class of Bayes' procedures is a complete class.

For the loss $L(F, G)$ at the decision G and nature's strategy F we have $\int (F - G)^2 \, dW$, so that, for a decision rule $\hat{F}$ and a strategy of nature F, we take $R(F, \hat{F}) = E_Y \int (F - \hat{F}_Y)^2 \, dW$, where E_Y denotes average with respect to the binomial distribution given to Y by F.

Assumption 8. $\inf_{\hat{F}} R(F, \hat{F}) > -\infty$ *for every* F.

Proof. Obvious.

Assumption 9. *For every pair* $(\hat{F}_1, \hat{F}_2)$ *of decision functions and every* α, $0 \leqq \alpha \leqq 1$, *there exists a decision function* $\hat{F}$ *such that*

$$R(F, \hat{F}) \leqq \alpha R(F, \hat{F}_1) + (1 - \alpha) R(F, \hat{F}_2).$$

Proof. Take $\hat{F} = \alpha \hat{F}_1 + (1 - \alpha)\hat{F}_2$.

Assumption 10. *The set of decision functions is a compact topological space and* $R(F, \hat{F})$ *is, for each* F, *lower semi-continuous on this space.*

Proof. The topological space of decision functions is a finite product of compact spaces and is therefore compact.

That $R(F, \hat{F})$ is continuous on this space can be seen as follows. Write $W = W_d + W_c$, where the only points of increase of W_d are jumps and W_c is continuous. Since

$$R(F, \hat{F}) = E_Y \left[\int (F - \hat{F}_Y)^2 \, dW_d + \int (F - \hat{F}_Y)^2 \, dW_c \right]$$

it suffices to prove that for each of the finitely many values of Y, (a) $\int (F - \hat{F})^2 \, dW_d$ is continuous and (b) $\int (F - \hat{F})^2 \, dW_c$ is continuous. (For simplicity the subscript Y is omitted.)

(a) Let $A_N = (x_1, \cdots, x_N)$ be the set where the first (in order of magnitude) N jumps of W_d occur, where N is such that $\int_{A_N} dW_d > 1 - \epsilon/4$. Then if $\vartheta = \{\hat{F}' \mid |\hat{F}'(x) - \hat{F}(x)| < \epsilon/4 \text{ for all } x \in A_N\}$ we have $\int |\hat{F}' - \hat{F}| \, dW_d < \epsilon/2$ for all $\hat{F}' \in \vartheta$ and so $|\int (F - \hat{F}')^2 \, dW_d - \int (F - \hat{F})^2 \, dW_d| < \epsilon$ for all $\hat{F}' \in \vartheta$.

(b) Let x_0 and x_0' be continuity points of $\hat{F}$ so that

$$\int_{-\infty}^{x_0} d\hat{F} - \inf_x \hat{F}(x) < \epsilon/6 \quad \text{and} \quad \sup_x \hat{F}(x) - \int_{x'_0}^{\infty} d\hat{F} < \epsilon/6.$$

Further let $B_N = (x_1, \cdots, x_N)$ be the set where the first (in order of magnitude) N jumps of $\hat{F}$ occur in the open interval (x_0, x_0') and $B = \lim_{N \to \infty} B_N$. Choose

N so that $\int_{B-B_N} d\hat{F} < \epsilon/6$. Then choose $(a_1, \cdots, a_N)$ so that the open intervals

$$\{(-\infty, x_0), (x_i - a_i, x_i + a_i) i = 1, \cdots, N, (x_0', \infty)\}$$

are disjoint and $W_c(x_i + a_i) - W_c(x_i - a_i) < \epsilon/6N$. Now choose $\delta > 0$ so that $|\hat{F}(x) - \hat{F}(y)| < \epsilon/6$ for all x, y with $|x - y| < \delta$ in the same one of the finitely many disjoint closed intervals whose union is the difference Δ of the closed interval $[x_0, x_0']$ and the union of the intervals $\{(x_i - a_i, x_i + a_i), i = 1, \cdots, N\}$. Finally let

$$\vartheta' = \{\hat{F}' \mid |\hat{F}(x) - \hat{F}'(x)| < \epsilon/6 \quad \text{for all} \quad x \in C\},$$

where $C = \{v_1 < v_2 < \cdots < v_n\}$ is any finite set of points in Δ which contains the points $x_0, x_0', x_i \pm a_i$ $(i = 1, \cdots, N)$ so that $v_{j+1} - v_j < \delta$ if v_j and v_{j+1} are in the same closed interval of Δ.

For $\hat{F}' \in \vartheta'$ it follows that $\int |F - F'| \, dW_c < \epsilon/2$, so that

$$\left| \int (F - \hat{F}')^2 \, dW_c - \int (F - \hat{F})^2 \, dW_c \right| < \epsilon.$$

5. Special case. The class of examples we want to describe will depend on a method we will call z-interpolation. Let z be a random variable with a distribution on the unit interval and let $Ez = \mu$, $\sigma_z^2 = \sigma^2$. Let $\{z_i\}$ be a sequence of independent random variables each distributed like z. A process F on the unit interval is now defined as follows. The distribution of $F(\frac{1}{2})$ will be that of z_1. The conditional, given $F(\frac{1}{2})$, distribution of $F(\frac{1}{4})$ is that of $z_2 F(\frac{1}{2})$. The conditional, given $F(\frac{1}{2})$ and $F(\frac{1}{4})$, distribution of $F(\frac{3}{4})$ is that of $F(\frac{1}{2}) + z_3(1 - F(\frac{1}{2}))$. The distributions of $F(\frac{1}{8})$, $F(\frac{3}{8})$, $F(\frac{5}{8})$ and $F(\frac{7}{8})$ are defined, in the same manner, by interpolation between 0, $F(\frac{1}{4})$, $F(\frac{1}{2})$, $F(\frac{3}{4})$ and 1, and those for $F(k/2^n)$ (k odd, $n > 3$) are defined by the obvious induction.

For processes whose law is given by z-interpolation the mean $m(\mu, x)$ at x can be found by noting that

1. $m(\mu, x) = \mu m(\mu, 2x)$ for $x \leqq \frac{1}{2}$,
2. $m(\mu, x) = 1 - m(1 - \mu, 1 - x)$.

These functions have been discussed by Dubins and Savage [2]. It is interesting to note, as have they, that $\{m(\mu, x)\}$ is a family of mutually singular distribution functions. In particular, if $\mu \neq \frac{1}{2}$, $m(\mu, x)$ is singular with respect to Lebesgue measure.

The variance var (μ, σ, x) at x of the process can be found by noting that

1. $\text{var}(\mu, \sigma, x) = (\sigma^2 + \mu^2) \text{var}(\mu, \sigma, 2x) + \sigma^2\{m(\mu, 2x)\}^2$ for $x \leqq \frac{1}{2}$,
2. $\text{var}(\mu, \sigma, x) = \text{var}(1 - \mu, \sigma, 1 - x)$.

6. Example. We have computed the Bayes' estimates, $E(F \mid Y)$, for a process defined by z-interpolation, with z taken to be uniform (0, 1). The dosage levels are at $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$, and the observations at these points are, in this order, (y_2, y_1, y_3). Further $N = \sum_{i=1}^{3} n_i$. These estimates are

$$E(F(\tfrac{1}{4}) \mid Y) = \frac{y_2+1}{n_2+2} \frac{\sum_{j=y_2+2}^{n_2+2} \sum_{i=0}^{y_3} \binom{n_2+2}{j}\binom{n_3+1}{i} B(y_1+i+j, N+3-i-j-y_1)}{\sum_{j=y_2+1}^{n_2+1} \sum_{i=0}^{y_3} \binom{n_2+1}{j}\binom{n_3+1}{i} B(y_1+i+j, N+2-i-j-y_1)}$$

$$E(F(\tfrac{1}{2}) \mid Y) = \frac{\sum_{j=y_2+1}^{n_2+1} \sum_{i=0}^{y_3} \binom{n_2+1}{j}\binom{n_3+1}{i} B(y_1+i+j+1, N+2-i-j-y_1)}{\sum_{j=y_2+1}^{n_2+1} \sum_{i=0}^{y_3} \binom{n_2+1}{j}\binom{n_3+1}{i} B(y_1+i+j, N+2-i-j-y_1)}$$

$$E(F(\tfrac{3}{4}) \mid Y) = \frac{y_3+1}{n_3+2} \frac{\sum_{j=y_2+1}^{n_2+1} \sum_{i=0}^{y_3+1} \binom{n_2+1}{j}\binom{n_3+2}{i} B(y_1+i+j, N+3-i-j-y_1)}{\sum_{j=y_2+1}^{n_2+1} \sum_{i=0}^{y_3} \binom{n_2+1}{j}\binom{n_3+1}{i} B(y_1+i+j, N+2-i-j-y_1)}.$$

The estimate for any point between the dosage levels is found by linear interpolation between these estimates. This is clear since $E(F \mid F(\frac{1}{4}), F(\frac{1}{2}), F(\frac{3}{4}), Y)$ is linear, so that $E[E(F \mid F(\frac{1}{4}), F(\frac{1}{2}), F(\frac{3}{4}), Y) \mid Y] = E(F \mid Y)$ is linear.

7. Remarks.

1. In the example just preceding we have used as an a priori distribution for F a process whose mean value is the cumulative of the uniform distribution on (0, 1). This process can be adapted to a process, whose mean value is any given continuous cumulative H_0. The corresponding estimate $\hat{H}$ is defined by $\hat{H}(x) = E\{\hat{F}(H_0(x)) \mid Y\}$, where $\hat{F}$ is the estimate of the above example.

2. In the foregoing example we have assumed that the dosage levels were at the $(k/2^n)$-tiles of nature's average strategy. To apply this method of estimation the experimenter needs some idea of this average strategy. The question of the sensitivity of these estimates to this assumption needs further exploration. However in Bayesian problems with simpler parameter spaces the performance of the Bayes' estimate is, for moderate n, relatively insensitive to the a priori distribution. It seems not unreasonable to expect the same here.

3. In many bio-assay problems the question is one of identifying F_θ, where $F_\theta(x) = F_0(x - \theta)$ for some fixed F_0. The usual loss for these problems is $L(F_{\theta_1}, F_{\theta_2}) = (\theta_1 - \theta_2)^2$. Since $(F_{\theta_2}(u) - F_{\theta_1}(u))^2 = (\theta_1 - \theta_2)^2(F'_\xi(u))^2$, $(\theta_1 < \xi < \theta_2)$, it follows that our loss $\int (F_{\theta_1} - F_{\theta_2})^2\, dW = (\theta_1 - \theta_2)^2 \int (F'_\xi(u))^2 dW$, so that the two functions are in local agreement if $\int (F'_\xi(u))^2\, dW < M$.

8. Some related results. Since we submitted this paper it has come to our attention that there are further results relating to some of the problems raised in this paper.

Dubins and Freedman [1] have studied a class of processes constructed with a method of interpolation which includes both the interpolation of Section 5 and a percentile interpolation mentioned in Section 2. There they prove that, for a priori distributions given by the interpolation of Section 5, the Bayes' estimates are consistent.

Kiefer and Wolfowitz [4] have given conditions under which the maximum likelihood estimate has the asymptotic properties of normality when the parameter space is infinitely dimensional. As an example they consider the problem of estimating a distribution function F.

Acknowledgment. The authors thank the referee for his many helpful suggestions.

REFERENCES

[1] Dubins, L. E. and Freedman, David A. (1963). Random distribution functions. *Bull. Amer. Math. Soc.* **69** 548–551.
[2] Dubins, L. E. and Savage, L. J. (1963). *How to Gamble If You Must.* (Unpublished.)
[3] Kelley, J. L. (1961). *General Topology.* Van Nostrand, Princeton.
[4] Kiefer, J. and Wolfowitz, J. Note on asymptotic efficiency of M. L. estimators in non-parametric problems. Unpublished report.
[5] LeCam, L. (1955). An extension of Wald's theory of statistical decision functions. *Ann. Math. Statist.* **26** 69–81.