Biometrika (1967), **54**, 3 *and* 4, *p.* 451
With 1 *text-figure*
Printed in Great Britain

A supplementary sample non-parametric empirical Bayes approach to some statistical decision problems

By RICHARD G. KRUTCHKOFF
Virginia Polytechnic Institute

Summary

When an estimating problem is routine, it is often possible to consider the parameter being estimated as a random variable. The data obtained to estimate previous values of the parameter then contain information which can be used to advantage in estimating the present parameter. Besides these data it is assumed that there are supplementary estimates of the previous parameters, perhaps in the form of customer feedback. All the probability distributions are assumed to be unknown. The estimating procedure given here is shown to be asymptotically optimal, and by a Monte Carlo example to have good small sample properties.

1. Introduction

Consider the situation where a process, perhaps a machine, is routinely producing batches of items, say bolts. We assume that one batch parameter, λ, is of interest, say the proportion of defective items in the batch. Since a parameter depends on many things, such as machine operator, machine adjustment, room temperature, humidity, etc., which are often uncontrollable, one can assume that the parameter varies slightly and unpredictably from batch to batch. This being the case, one usually ignores information obtained from previous batches when estimating the parameter for the present batch. The parameter, by virtue of its unpredictability, may, however, be considered a random variable with an unknown probability distribution. If the distribution of the parameter were known, one would be in what is usually called a Bayesian situation. It was shown (Robbins, 1955) that there is sometimes sufficient information available from previous batches to obtain results which at least asymptotically supplant knowledge of the probability distribution. We will use this approach, called the empirical Bayes approach, and extend it to the situation where the batch information is of a non-parametric nature. As will become apparent, this non-parametric approach is often useful even when parametric information is available. To be specific, consider the case in which one takes a random sample of m bolts from the batch and counts the number of defectives. Since the total number of defectives, X, is a sufficient statistic for the proportion λ of defectives in the batch one can use X/m, the uniformly minimum variance unbiased estimator for λ.

In this example the form of the distribution of X can be considered known allowing one to find an estimator for λ with desirable properties. As an example where the distribution is not known consider the case where the batch represents missile propellant and the parameter is thrust per pound. The sample might consist of several static test firings. Unless one was willing to assume some form for the distribution one might simply give the average measured thrust as the estimator for λ without giving its properties; the estimator may be biased since the firing was static and the thrust of interest may be the dynamic thrust.

In order to obtain a non-parametric estimator, we will assume that there is supplementary information available. Consider the situation in which there is feedback from the purchaser of the batch of bolts. He conducted his own tests, and obtained his own estimate of λ. Perhaps he drew a sample of r bolts and used the number of defectives divided by r, call it y, as his estimate. In the propellant example y might represent the average thrust obtained after several of the missiles were fired dynamically. All that is required of the estimator Y of the supplementary information is that it be an unbiased estimator for the parameter. It should be emphasized here that this is supplementary information available only after the estimate of λ is given and, therefore, cannot be used to estimate λ. This supplementary information is available, however, for many, if not all, of the previous batches. The values of y obtained for previous batches are not estimates of the parameter for the present batch and would usually not be used to estimate λ. In spite of this, we will present an estimator which uses the values of X and Y from previous batches that has an expected mean squared error smaller than the variance of X/m.

2. An asymptotically optimal estimator

As indicated in the introduction, the situation we consider is the following. The unobservable outcome λ of a random variable Λ occurs according to the unknown distribution $G(\lambda)$. The observable outcome x of random variable X is obtained according to an unknown conditional distribution $F_\lambda(x)$. At this point one must estimate λ with, hopefully, a small squared error. After the estimate is made one is given the outcome y of a random variable Y which has an unknown conditional distribution $H_\lambda(y)$. It is known, however, that

$$E_\lambda(Y) = \int y\, dH_\lambda(y) = \lambda \tag{2·1}$$

and

$$E(Y^2) = \iint y^2\, dH_\lambda(y)\, dG(\lambda) < \infty. \tag{2·2}$$

The situation has occurred n times and as a consequence there is a vector $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n)$ of unobservables obtained independently according to the distribution $G(\lambda)$, a vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ of observations independently obtained from the distributions

$$F_{\lambda_i}(x) \quad (i = 1, 2, \ldots, n)$$

and a vector $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ of observations independently obtained from the distributions $H_{\lambda_i}(y)$ $(i = 1, 2, \ldots, n)$. After this one obtains a value x from the distribution $F_\lambda(x)$, where λ is the unobserved outcome from the distribution $G(\lambda)$. Note that these should be labelled x_{n+1} and λ_{n+1} but we will drop these subscripts as no confusion can result.

It can easily be seen from the independence that the uniformly minimum variance unbiased estimator for λ, if it exists, is a function only of x and not $\mathbf{x}$, $\mathbf{y}$. The minimum mean squared error estimator, the Bayes estimator for a squared error loss function, is similarly not a function of $\mathbf{x}$, $\mathbf{y}$. The Bayes estimator will, however, be unattainable since $G(\lambda)$ is unknown, but we will obtain an estimator which is at least asymptotically Bayes by using $\mathbf{x}, \mathbf{y}$.

The Bayes estimator $a^*(\cdot)$ is defined as one for which

$$E\{a^*(X) - \Lambda\}^2 = \min_{a(\cdot)} E\{a(X) - \Lambda\}^2, \tag{2·3}$$

where the expectation is taken over all the random variables and $a(\cdot)$ is any measurable function of X. It can easily be seen that

$$E\{a(X)-\Lambda\}^2 = E(E[\{a(X)-\Lambda\}^2|X = x]) \\ = E[\{a(X)-E(\Lambda|X)\}^2]+E\{\text{var}(\Lambda|X)\}, \tag{2·4}$$

where var $(\Lambda|X)$ is the posterior variance of Λ for given X. This is minimized when

$$a(X) = E(\Lambda|X = x) = a^*(X) \tag{2·5}$$

with a minimum mean squared error

$$R = E\{\text{var}(\Lambda|X)\}. \tag{2·6}$$

If we assume that
$$E(\Lambda^2) = \int\lambda^2 dG(\lambda) < \infty, \tag{2·7}$$

then $R < \infty$ and the difference between the mean squared error for any estimator $a(\cdot)$ and R is

$$r(a) = E[\{a(X)-E(\Lambda|X)\}^2], \tag{2·8}$$

which is called the regret in using $a(\cdot)$. Since the Bayes estimator $E(\Lambda|X)$ cannot, in general, be found without knowledge of both $G(\lambda)$ and $F_\lambda(x)$, which are unknown to us, we will attempt to find an estimator which is a function of $\mathbf{Z} = \mathbf{X}, \mathbf{Y}$ as well as X for which

$$\lim_{n\to\infty} r(a) = \lim_{n\to\infty} E[\{a(\mathbf{Z}, X)-E(\Lambda|X)\}^2] = 0. \tag{2·9}$$

Equation (2·9) is the definition of asymptotic optimality of a sequence of estimators $a(\mathbf{Z}, X)$. It should be noted that each estimator in this sequence is estimating a different parameter. That is, $a(\mathbf{z}_i, x_{i+1})$ estimates λ_{i+1}, whereas $a(\mathbf{z}_j, x_{j+1})$ estimates λ_{j+1} where $Z_i = (X_1, X_2, \ldots, X_i, Y_1, Y_2, \ldots, Y_i)$.

Since $r(a)$ is an unconditional expectation, we must be careful to take expectations over $\mathbf{Z} = (X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_n)$ as well as over X and Λ. The introduction of $\mathbf{Z}$ does not, however, require the redevelopment of R or $r(a)$ since by assumption, X and Λ are independent of $\mathbf{Z}$. At this point we will simplify matters by considering only discrete X, i.e. $dF_\lambda(x) = P_\lambda(x)$ on some countable set of values of x. The continuous case has been treated in my Columbia thesis.

In order to define the function $\phi_n(\mathbf{z}; x)$ which will be our choice of $a(\mathbf{z}; x)$ we let

$$\delta(x_i, x) = \begin{cases} 1 & \text{if} \quad x_i = x, \\ 0 & \text{if} \quad x_i \neq x \end{cases} \tag{2·10}$$

and
$$m_n(\mathbf{z}; x) = \sum_{i=1}^{n} \delta(x_i, x). \tag{2·11}$$

We now define
$$\phi_n(\mathbf{z}; x) = \begin{cases} \dfrac{1}{m_n(\mathbf{z}; x)} \sum_{i=1}^{n} \delta(x_i, x)\, y_i & \text{if} \quad m(\mathbf{z}; x) > 0, \\ 0 & \text{if} \quad m(\mathbf{z}; x) = 0. \end{cases} \tag{2·12}$$

Before continuing, we present an important equality which is essentially Theorem 1 of my thesis, namely

$$E(Y|x) = E(\Lambda|x) \text{ a.s.} \tag{2·13}$$

For
$$E(\Lambda|x) = E\{E(Y|\Lambda)|X = x\} \text{ a.s.} \\ = E\{E(Y|\Lambda, X)|X = x\} \text{ a.s.} \\ = E(Y|x) \text{ a.s.} \tag{2·14}$$

The second step is obtained by assuming that X and Y are conditionally independent, i.e. the joint distribution of X, Y, Λ is given by $F_\lambda(x)\,H_\lambda(y)\,G(\lambda)$. The third step uses integration with respect to Λ only. The assumption that $E(\Lambda^2) < \infty$ is sufficient to ensure that this interchange in order of taking expectations is permissible.

When $m_n(\mathbf{z}; x) = m > 0$, $\phi_n(\mathbf{z}; x)$ is simply the average of m independent unbiased estimates of $E(Y|x)$, which equals $E(\Lambda|x)$. Therefore if we let

$$\phi(x) = E(\Lambda|x) \tag{2·15}$$

we obtain

$$E\{\phi_n(\mathbf{Z}; x)|m_n = m > 0, X = x\} = \phi(x) \tag{2·16}$$

and

$$E[\{\phi_n(\mathbf{Z}; x) - \phi(x)\}^2|m_n = m > 0, X = x] = \frac{1}{m}\operatorname{var}(Y|x), \tag{2·17}$$

while

$$E\{\phi_n(\mathbf{Z}; x)|m_n = 0, X = x\} = 0 \tag{2·18}$$

and

$$E[\{\phi_n(\mathbf{Z}; x) - \phi(x)\}^2|m_n = 0, X = x] = \phi^2(x). \tag{2·19}$$

Iterating the expectation in (2·8), we obtain, after some manipulation with conditional expectations, that

$$\begin{aligned} r(\phi_n) &= E\{\phi_n(\mathbf{Z}; X) - \phi(x)\}^2 \qquad (2·20)\\ &= E\left(\operatorname{var}(Y|X)\,E\left[\sum_{m=1}^{n}\frac{1}{m}\operatorname{prob}\{m_n(\mathbf{Z}_n; X) = m\}|X\right]\right.\\ &\qquad\left. + \phi^2(X)\,E[\operatorname{prob}\{m_n(\mathbf{Z}_n; X) = 0\}|X]\right). \end{aligned}$$

Note that since both expected probabilities cannot exceed unity the quantity inside the outer expectation is dominated by $\operatorname{var}(Y|X) + \phi^2(X)$. This, however, equals $E(Y^2|X)$, which has finite expectation.

Now $\operatorname{prob}\{m_n(\mathbf{Z}; x) = m\}$ is the probability that in n independent trials with probability

$$P(x) = \int P_\lambda(x)\,dG(\lambda) \tag{2·21}$$

for success there are exactly m successes. Therefore,

$$\operatorname{prob}\{m_n(\mathbf{Z}; x) = m\} = \binom{n}{m}\{P(x)\}^m\{1 - P(x)\}^{n-m} \tag{2·22}$$

for any $\mathbf{Z}$ such that $m_n(\mathbf{Z}; x) = m$. From this we obtain

$$E[\operatorname{prob}\{m_n(\mathbf{Z}; X) = 0\}|X = x] = \{1 - P(x)\}^n \tag{2·23}$$

and therefore

$$\lim_{n\to\infty} E[\operatorname{prob}\{m_n(\mathbf{Z}; x) = 0\}|X = x] = 0 \text{ a.s.} \tag{2·24}$$

Also

$$\begin{aligned} &E\left[\sum_{m=1}^{n}\frac{1}{m}\operatorname{prob}\{m_n(\mathbf{Z}; X) = m\}|X = x\right] \qquad (2·25)\\ &= \sum_{m=1}^{n}\frac{1}{m}\binom{n}{m}\{P(x)\}^m\{1 - P(x)\}^{n-m}\\ &\leqslant \sum_{m=1}^{n}\frac{2}{m+1}\binom{n}{m}\{P(x)\}^m\{1 - P(x)\}^{n-m}\\ &\leqslant \frac{2}{(n+1)P(x)}\sum_{m=1}^{n}\binom{n+1}{m+1}\{P(x)\}^{m+1}\{1 - P(x)\}^{n-m} \leqslant \frac{2}{(n+1)P(x)}. \end{aligned}$$

Therefore $$\lim_{n\to\infty} E\left[\sum_{m=1}^{n} \frac{1}{m} \operatorname{prob}\{m_n(\mathbf{Z}; X) = m\}|X\right] = 0 \text{ a.s.} \tag{2·26}$$

Using (2·24), (2·26) and the remark above we obtain by the Lebesgue dominated convergence theorem that, as $n \to \infty$, $\lim r(\phi_n) = 0$. Thus the risk of $\phi_n(\mathbf{z}; x)$ attains the Bayes risk as n increases and therefore $\phi_n(\mathbf{z}; x)$ is an asymptotically optimal estimator.

3. An unrealistic but practical assumption

Before relating the results of the previous section to an example, we will consider the unrealistic assumption which will later prove to be of practical importance. Consider the assumption that the mean of the posterior distribution of Λ for given $X = x$ is a linear function of x. That is,

$$E(\Lambda|X) = \alpha + \beta X. \tag{3·1}$$

Note that this is, in general, not true. Thus in the binomial example it is easily shown (Robbins, 1955) that

$$E(\Lambda|X) = \left(\frac{X+1}{m+1}\right)\frac{P_{m+1}(X+1)}{P_m(X)}, \tag{3·2}$$

where $P_s(t)$ is the marginal probability that there will be t defectives found in a sample of size s.

Since $E(Y|X) = E(\Lambda|X)$ we obtain from (3·1)

$$E(Y|X) = \alpha + \beta X. \tag{3·3}$$

Thus, one obtains the regression model

$$y = \alpha + \beta X + \epsilon, \tag{3·4}$$

where ϵ is an error with zero mean and finite variance say σ^2. Thus, the previous values of x and y can be considered observations from a linear regression model with x as the controlled variable. One can obtain the least squares estimates of β and α as

$$b = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2}, \quad a = \bar{y} - b\bar{x}, \tag{3·5}$$

with the usual expectations and variances, given X.

Assuming that $(X_i - \bar{X})^2$ is not zero for most X_i while X remains finite, we can show that

$$\lim_{n\to\infty} \operatorname{var}(a + bX|\mathbf{X}, X) = 0 \text{ a.s.} \tag{3·6}$$

Let us now define

$$\phi^*(\mathbf{Z}, X) = \begin{cases} a + bX & \text{if } S_{XX} = \Sigma(X_i - \bar{X})^2 \neq 0, \\ 0 & \text{if } S_{XX} = 0. \end{cases} \tag{3·7}$$

Then

$$\begin{aligned} r(\phi^*) &= E[\{\phi^*(\mathbf{Z}, X) - \phi(X)\}^2] \\ &= E\{\operatorname{var}(a + bX|\mathbf{X}, X, S_{XX} \neq 0)\, P(S_{XX} \neq 0)\} + E\{\phi^2(X)\, P(S_{XX} = 0)\}. \end{aligned} \tag{3·8}$$

Since $S_{XX} = 0$ only for degenerate $P_\lambda(X)$, we can see that

$$\lim_{n\to\infty} P(S_{XX} = 0) = 0, \tag{3·9}$$

while
$$\lim_{n\to\infty} E\{\mathrm{var}\,(a+bX|\mathbf{X}, X, S_{XX} \neq 0)\,P(S_{XX} \neq 0)\}$$
$$\leqslant \lim_{n\to\infty} E\{\lim_{n\to\infty} \mathrm{var}\,(a+bX|\mathbf{X}, X, S_{XX} \neq 0)\} = 0. \tag{3·10}$$

Therefore
$$\lim_{n\to\infty} r(\phi^*) = 0 \tag{3·11}$$

and $\phi^*(\mathbf{Z}, X)$ is asymptotically optimal.

4. A NUMERICAL EXAMPLE

Consider the bolt problem of the introduction. Let us say that the proportion of defectives, Λ, was distributed as a normal variable with mean 0·2 and standard deviation 0·05. Let us say further that there were five bolts taken for inspection and the number of defectives called X. The uniformly minimum variance unbiased estimator would have a mean squared error given by

$$E\{\mathrm{var}\,(\tfrac{1}{5}X)\} = E\left\{\frac{\Lambda(1-\Lambda)}{5}\right\} = 0·0315. \tag{4·1}$$

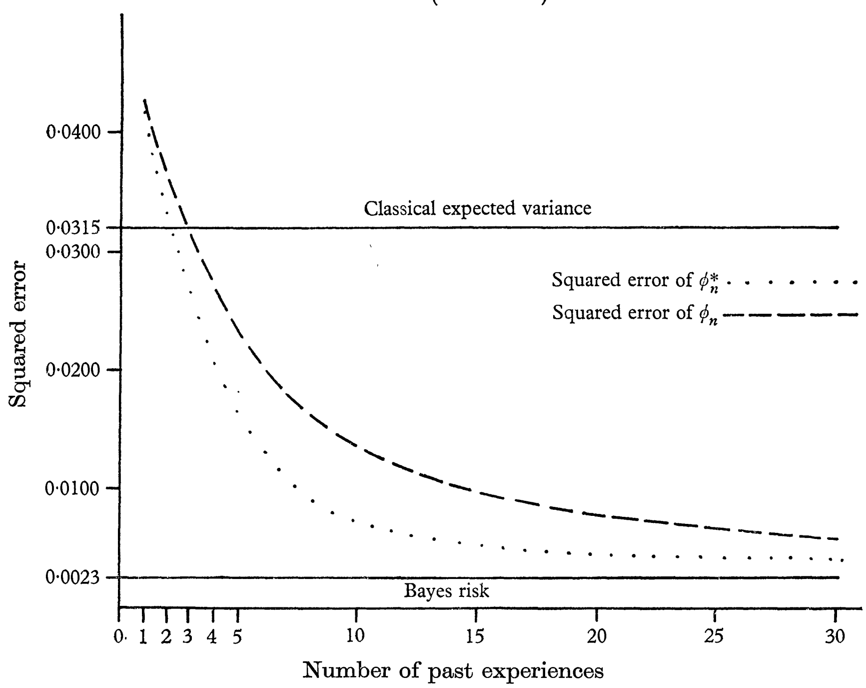

Fig. 1. A Monte Carlo comparison of two non-parametric empirical Bayes estimators with the classical estimator.

The Bayes estimator $E(\Lambda|x)$ has a mean squared error $E\{\mathrm{var}\,(\Lambda|X)\} = 0·0023$ obtained by numerical integration on a high speed computer. Since it is assumed that the distribution $G(\lambda)$, the normal here, is actually unknown, one cannot obtain the Bayes estimator. We therefore try $\phi(\mathbf{Z}, X)$ and $\phi^*(\mathbf{Z}, X)$, which do not require knowledge of $G(\lambda)$. We assume that y is the proportion of defectives in a supplementary sample, obtained after each estimate is made, of ten bolts. The plot, Fig. 1, was obtained by Monte Carlo simulation using 5000 replications. Notice that for as little as three previous experiences both estimators have a mean squared error smaller than that of the uniformly minimum variance unbiased estimator. By thirty previous experiences $\phi(\mathbf{Z}, x)$ has a mean squared error below

0·0060, whereas $\phi^*(\mathbf{Z}, X)$ has a mean squared error below 0·0037. Even though the linearity assumption does not hold here $\phi^*(\mathbf{Z}, X)$ has a smaller mean squared error than does $\phi(\mathbf{Z}, X)$. This may suggest a practical compromise. If the data look fairly linear one should use $\phi^*(\mathbf{Z}, X)$ until a sufficient number of previous occurrences have occurred to ensure at least several values of y for the given value of x, and then switch over to $\phi(\mathbf{Z}, X)$. This would ensure asymptotic optimality even when the linearity assumption did not hold.

5. An hypothesis testing problem

Sometimes one is interested not in estimating the parameter but rather in testing

$$H_0\colon \lambda < \lambda^*, \quad \text{or} \quad \lambda \leqslant \lambda^*, \quad \lambda^* \text{ being given,}$$

against its natural alternative. We can treat the hypothesis with the inequality reversed in a similar manner. The possible actions are a_0, to accept the hypothesis and a_1 to reject the hypothesis. We will use as our loss the function which is zero when we are correct and $|\lambda - \lambda^*|$ when we are incorrect. Note there are no fixed type I or type II errors here. Each error is considered equally costly and is weighted by how incorrect it is (Johns, 1957; Robbins, 1963; Samuel, 1963). Let

$$L(a_i, \lambda) = \begin{Bmatrix} L_0(\lambda) = |\lambda - \lambda^*| & \text{if} \quad i = 0, \quad \lambda \geqslant \lambda^*, \\ L_1(\lambda) = |\lambda - \lambda^*| & \text{if} \quad i = 1, \quad \lambda < \lambda^*, \\ 0 & \text{otherwise.} \end{Bmatrix} \tag{5·1}$$

Let
$$A(x) = \begin{cases} 0 & \text{if} \quad a(x) = a_0, \\ 1 & \text{if} \quad a(x) = a_1. \end{cases} \tag{5·2}$$

Then
$$\begin{aligned} L\{a(x), \lambda\} &= A(x)\, L_1(\lambda) + \{1 - A(x)\}\, L_0(\lambda) \\ &= L_0(\lambda) - A(x)\,\{L_0(\lambda) - L_1(\lambda)\}. \end{aligned} \tag{5·3}$$

It is easily seen that
$$L_0(\lambda) - L_1(\lambda) = \lambda - \lambda^* \tag{5·4}$$

for all λ and either decision and therefore

$$L\{a(x), \lambda\} = L_0(\lambda) - A(x)\,(\lambda - \lambda^*). \tag{5·5}$$

The risk or expected loss can be found by

$$\begin{aligned} R(a) &= E\{L_0(\Lambda)\} - E\{A(X)\,(\Lambda - \lambda^*)\} \\ &= E\{L_0(\Lambda)\} - E[A(X)\,\{E(\Lambda|X) - \lambda^*\}]. \end{aligned} \tag{5·6}$$

A function $a^*(x)$ which minimizes this can be seen to be the $a^*(x)$ corresponding to

$$A^*(x) = \begin{Bmatrix} 1 & \text{if} \quad \phi(x) \geqslant \lambda^*, \\ 0 & \text{if} \quad \phi(x) < \lambda^*, \end{Bmatrix} \tag{5·7}$$

where $\phi(x)$ was defined to be $E(\Lambda|x)$ in (2·15). The Bayes risk or minimum expected loss is then
$$R \leqslant E\{L_0(\Lambda)\} \leqslant E\{|\Lambda - \lambda^*|\}, \tag{5·8}$$

which by assumption is finite. The regret in using an arbitrary $a(x)$ with corresponding $A(x)$ is then
$$r(a) = E[\{A^*(X)\}\,\{\phi(X) - \lambda^*\}]. \tag{5·9}$$

If we now use the supplementary sample non-parametric empirical Bayes estimate $\phi_n(\mathbf{z}; x)$ defined in (2·12) we can define a supplementary sample non-parametric empirical Bayes test of hypothesis as the $a_n(\mathbf{z}; x)$ such that

$$A_n(\mathbf{z}; x) = \begin{cases} 1 & \text{if} \quad \phi_n(\mathbf{z}; x) \geqslant \lambda^*, \\ 0 & \text{if} \quad \phi_n(\mathbf{z}; x) < \lambda^*. \end{cases} \tag{5·10}$$

The regret becomes $$r(a_n) = E[\{A^*(X) - A_n(\mathbf{Z}; X)\}\{\phi(X) - \lambda^*\}]. \tag{5·11}$$

As in the estimation problem the introduction of $\mathbf{Z}$ causes no difficulties because of the independence of $\mathbf{Z}$ and (Λ, X). Note that

$$A^*(x) - A_n(\mathbf{z}; x) = 1 \quad \text{only when} \quad \phi(x) - \lambda^* \geqslant 0 \tag{5·12}$$

and $$A^*(x) - A_n(\mathbf{z}; x) = -1 \quad \text{only when} \quad \phi(x) - \lambda^* < 0. \tag{5·13}$$

Note also that $$A^*(x) - A_n(\mathbf{z}; x) \neq 0 \tag{5·14}$$

only when $\phi(x) - \lambda^* \geqslant 0$, and $\phi_n(\mathbf{z}; x) - \lambda^* < 0$, or when $\phi(x) - \lambda^* < 0$ and $\phi_n(\mathbf{z}; x) - \lambda^* \geqslant 0$. Now (5·12), (5·13), and (5·14) imply that

$$\begin{aligned} r(a_n) &\leqslant E\{|\phi(X) - \lambda^*|\} \\ &\leqslant E[|\{\phi(X) - \lambda^*\} - \{\phi_n(\mathbf{Z}; X) - \lambda^*\}|] \\ &\leqslant E\{|\phi_n(\mathbf{Z}; X) - \phi(X)|\}. \end{aligned} \tag{5·15}$$

By (2·8) and (2·27) we see that we have already proved that

$$\lim_{n\to\infty} E\{\phi_n(\mathbf{Z}; X) - \phi(X)\}^2 = 1. \tag{5·16}$$

This, however, implies that $\lim r(a_n) = 0$. Therefore, $a_n(\mathbf{z}; x)$ defined by (5·10) is an asymptotically optimal test of the hypothesis.

Similarly it is easily shown that under condition (3·1) $a_n^*(\mathbf{z}; x)$ defined by

$$A_n(\mathbf{z}; x) = \begin{cases} 1 & \text{if} \quad \phi_n^*(\mathbf{z}; x) \geqslant \lambda^*, \\ 0 & \text{if} \quad \phi_n^*(\mathbf{z}; x) < \lambda^* \end{cases} \tag{5·17}$$

is also an asymptotically optimal test of hypothesis.

REFERENCES

JOHNS, M. V. JR. (1957). Non-parametric Bayes procedures. *Ann. Math. Statist.* **28**, 649–69.

ROBBINS, H. (1955). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. and Prob.* University of California Press, pp. 157–64.

ROBBINS, H. (1963). The empirical Bayes approach to testing statistical hypotheses. *Rev. Inst. Internat. Statist.* **31**, 195–208.

SAMUEL, E. (1963). An empirical Bayes approach to the testing of certain parametric hypotheses. *Ann. Math. Statist.* **34**, 1370–85.

[*Received May* 1966. *Revised May* 1967]