NON-PARAMETRIC EMPIRICAL BAYES PROCEDURES[1]

By M. V. Johns, Jr.

Columbia University

1. Introduction and summary. In the usual formulation of problems in statistical decision theory the probability distribution of the observations is assumed to be a member of some specified class of distribution functions. No a priori information is ordinarily assumed to exist concerning which member of this class is the true distribution of the observations although a priori probability measures defined over this class may be introduced as a technical device for generating complete classes of decision functions, minimax decision rules, etc. However, in some experimental situations it may be reasonable to suppose that such an a priori probability measure actually exists in the sense that the distributions of observations occurring in different experiments made under similar circumstances may be thought of as having been selected from a specified class of distribution functions according to some probability law.

Such an assumption seems particularly apt in the case where measurements are made on an individual selected according to some probability law (e.g., "at random") from a population and where it is desired to make inferences about some characteristic of the individual on the basis of these measurements. If the class of probability distributions of the measurements for all individuals in the population and the law of selection are known, an optimum Bayes decision procedure can then be found. In general, however, such information will not be available to the experimenter, but there may be observations available on individuals previously selected in the same way from the same population and, under certain circumstances, these prior observations may be used to obtain approximations to the optimum Bayes decision procedure. The possibility of using prior observations to approximate Bayes procedures was first established for certain estimation problems in [1] by H. Robbins who coined the term "empirical Bayes procedures" to describe such approximations.

Robbins in [1] discusses the estimation, using a squared error loss function, of the value λ of a random variable Λ associated with a discrete valued observation X whose conditional probability function, given λ, is $p(x \mid \lambda)$, where $p(x \mid \lambda)$ is known for each λ but where the (a priori) distribution of Λ is unknown. For several specific parametric families of discrete probability functions $p(x \mid \lambda)$ Robbins shows that if prior independent observations $X_1, X_2, \cdots, X_n$, each having the same unconditional distribution as X are available, then an empirical Bayes estimator using $X_1, X_2, \cdots, X_n$ can be found which converges with probability 1 to the Bayes estimator as n increases, for any a priori distribution of Λ.

Received July 2, 1956.

[1] Work sponsored by the Office of Naval Research, Nonr Contract No. 266(33).

In Sec. 2 below a similar estimation problem is considered for the "nonparametric" case where the class of (conditional) probability distributions of X is not restricted to a particular parametric family, but is instead the class of all probability functions assigning probability 1 to some specified denumerable set of numbers. The quantity to be estimated is the value of a functional defined on this class of probability functions and it is assumed that there exists an unknown a priori probability measure defined on a suitable σ-algebra of subsets of this class. For this case it is shown that under certain circumstances prior observations may be used to construct empirical Bayes estimators having the property that, as the number of prior observations increases, the risks of the empirical Bayes estimators converge to the risk of the Bayes estimator for any a priori probability measure provided that certain moments exist. The rate of convergence of these risks is also investigated for two special cases.

In Sec. 3 the techniques of Sec. 2 are modified to apply to the case where the class of (conditional) distributions of X is the class of all absolutely continuous distribution functions, and similar results are obtained.

In Sec. 4 the results of the previous sections are used to obtain empirical Bayes solutions for certain two-decision problems of the hypothesis-testing type.

Throughout this paper certain elementary properties of conditional expectations are used which are immediate consequences of results contained, for example, in Chapter VII of [2].

2. Estimation: the discrete case. For a specified denumerable set of numbers $\chi = \{x\}$ let $\mathfrak{F} = \{F(x \mid \omega) : \omega \, \varepsilon \, \Omega\}$ be the class of all c.d.f.'s assigning probability 1 to χ, where $\Omega = \{\omega\}$ is an abstract indexing set. Let μ be an a priori probability measure defined on a σ-algebra $\mathfrak{A}$ of subsets of Ω, and let Y be the Ω-valued random variable which is the identity mapping of Ω onto itself. We may then define the random variables $X_1, X_2, \cdots, X_r$ so that they are conditionally independently and identically distributed with the common c.d.f. $F(x \mid \omega)$ given that $Y = \omega$. Finally, for a given measurable function $h(x)$ we define the random variable Λ by

$$\Lambda = \Lambda(Y) = E(h(X) \mid Y), \tag{2.1}$$

where X is a generic representative of the X_j's. We assume that the a priori probability measure space $(\Omega, \mathfrak{A}, \mu)$ is such that

$$Eh^2(X) < \infty. \tag{2.2}$$

Here Λ may be thought of as a functional defined on $\mathfrak{F}$ or, equivalently, as a function defined on Ω. We might, for example, define $h(x) \equiv x$ so that the value of $\Lambda(\omega)$ is the expected value of X given that the c.d.f. of X is $F(x \mid \omega)$. Another possibility would be to let $h(x) = 1$ if $x < c$, and $h(x) = 0$ otherwise, so that $\Lambda(\omega) = F(c \mid \omega)$.

Suppose that we wish to use the vector of observations

$$\underset{\sim}{X} = (X_1, X_2, \cdots, X_r)$$

to obtain an estimate of the value λ assumed by Λ, where the loss incurred when t is the estimated value is

$$L(t, \lambda) = (t - \lambda)^2. \tag{2.3}$$

Then the risk involved in using any estimator $\varphi(\underset{\sim}{x})$ is

$$R(\varphi) = EL(\varphi(\underset{\sim}{X}), \Lambda) = E[\varphi(\underset{\sim}{X}) - \Lambda]^2. \tag{2.4}$$

Now from (2.1) amd (2.2) we have

$$E\Lambda^2 = E\{E^2(h(X) \mid Y)\} \leqq Eh^2(X) < \infty, \tag{2.5}$$

so that $R(\varphi)$ may be written

$$\begin{aligned} &R(\varphi) \\ &= E\{[\varphi^2(\underset{\sim}{X}) - 2\varphi(\underset{\sim}{X})E(\Lambda \mid \underset{\sim}{X}) + E^2(\Lambda \mid \underset{\sim}{X})] + E(\Lambda^2 \mid \underset{\sim}{X}) - E^2(\Lambda \mid \underset{\sim}{X})\}. \end{aligned} \tag{2.6}$$

The expression in square brackets is a perfect square and hence is non-negative and equals zero if and only if $\varphi(\underset{\sim}{X}) = E(\Lambda \mid \underset{\sim}{X})$. Thus the Bayes estimator $\varphi_\mu(\underset{\sim}{x})$ which minimizes $R(\varphi)$ is given by

$$\varphi_\mu(\underset{\sim}{x}) = E(\Lambda \mid \underset{\sim}{X} = \underset{\sim}{x}) \tag{2.7}$$

for all $\underset{\sim}{x} = (x_1, x_2, \cdots, x_r) \varepsilon \chi^* = \{\underset{\sim}{x}: \text{Prob}\,(\underset{\sim}{X} = \underset{\sim}{x}) > 0\}$. The risk of the Bayes estimator is then

$$R(\varphi_\mu) = E\Lambda^2 - E\varphi_\mu^2(\underset{\sim}{X}) < \infty. \tag{2.8}$$

To obtain the Bayes estimator φ_μ we must, of course, know the a priori probability structure of the problem. Suppose now that this structure is *unknown* but that collateral information is available, the form of which is determined as follows:

Let $Y_1, Y_2, \cdots, Y_n$ be mutually independent Ω-valued random variables each of which is independent of Y and has the same distribution as Y. Then let the additional information be in the form of vectors of observations $\underset{\sim}{X}_i = (X_{i1}, X_{i2}, \cdots, X_{i,r+1})$, $i = 1, 2, \cdots, n$ where the $\underset{\sim}{X}_i$'s are mutually independent and independent of $\underset{\sim}{X}$ and where for each i the X_{ij}'s are conditionally independent and identically distributed according to $F(x \mid \omega_i)$ given that $Y_i = \omega_i$. Here although the $\underset{\sim}{X}_i$'s are independent of $\underset{\sim}{X}$ and Λ, they nevertheless contain useful information since they possess the same a priori probability structure as $\underset{\sim}{X}$. Thus, if we let $\underset{\sim}{X}_i^{(r)} = (X_{i1}, X_{i2}, \cdots, X_{ir})$ then $\underset{\sim}{X}_i^{(r)}$ and

$$E(h(X_{i,r+1}) \mid Y_i)$$

have the same joint distribution as $\underset{\sim}{X}$ and Λ so that

$$
\begin{aligned}
E(h(X_{i,r+1}) \mid \underset{\sim}{X}_i^{(r)} = \underset{\sim}{x}) &= E\{E(h(X_{i,r+1}) \mid Y_i, \underset{\sim}{X}_i^{(r)}) \mid \underset{\sim}{X}_i^{(r)} = \underset{\sim}{x}\} \\
&= E\{E(h(X_{i,r+1}) \mid Y_i) \mid \underset{\sim}{X}_i^{(r)} = \underset{\sim}{x}\} \\
&= E(\Lambda \mid \underset{\sim}{X} = \underset{\sim}{x}) = \varphi_\mu(\underset{\sim}{x}),
\end{aligned} \tag{2.9}
$$

for $\underset{\sim}{x} \varepsilon \chi^*$. This suggests the following empirical Bayes estimation procedure:

Let $\underset{\sim}{x}_{(q)}$, $q = 1, 2, \cdots, m(\underset{\sim}{x})$ represent the $m(\underset{\sim}{x})$ distinct vectors obtained by permuting the components $x_1, x_2, \cdots, x_r$ of $\underset{\sim}{x}$. Clearly for each $\underset{\sim}{x}$ we have $1 \leqq m(\underset{\sim}{x}) \leqq r!$. Now we define the random functions $M_i(\underset{\sim}{x})$, $i = 1, 2, \cdots, n$, and $\bar{M}_n(\underset{\sim}{x})$ by

$$
M_i(\underset{\sim}{x}) = \begin{cases} 1, & \text{if there exists a } q,\ 1 \leqq q \leqq m(\underset{\sim}{x}), \\ & \text{such that } \underset{\sim}{X}_i^{(r)} = \underset{\sim}{x}_{(q)}, \\ 0, & \text{otherwise,} \end{cases} \tag{2.10}
$$

and

$$
\bar{M}_n(\underset{\sim}{x}) = \sum_{i=1}^n M_i(\underset{\sim}{x}). \tag{2.11}
$$

Then we define an empirical Bayes estimator $\varphi_n(\underset{\sim}{x})$ by

$$
\varphi_n(\underset{\sim}{x}) = \begin{cases} \dfrac{1}{\bar{M}_n(\underset{\sim}{x})} \displaystyle\sum_{i=1}^n M_i(\underset{\sim}{x}) h(X_{i,r+1}), & \bar{M}_n(\underset{\sim}{x}) > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{2.12}
$$

In order to show that the risk involved in using φ_n approaches the risk for the Bayes estimator φ_μ as n becomes large we first prove two lemmas. For $n = 1, 2, \cdots$, let

$$
P_n(\underset{\sim}{x}) = \text{Prob}\ \{\bar{M}_n > 0\}, \tag{2.13}
$$

$$
V_n(\underset{\sim}{x}) = \begin{cases} \dfrac{1}{\bar{M}_n(\underset{\sim}{x})}, & \bar{M}_n(\underset{\sim}{x}) > 0, \\ 0, & \text{otherwise,} \end{cases} \tag{2.14}
$$

$$
\xi_n(\underset{\sim}{x}) = EV_n(\underset{\sim}{x}), \tag{2.15}
$$

and for $\underset{\sim}{x} \varepsilon \chi^*$,

$$
\theta(\underset{\sim}{x}) = E(h^2(X_{i,r+1}) \mid \underset{\sim}{X}_1^{(r)} = \underset{\sim}{x}). \tag{2.16}
$$

LEMMA 1. *For any fixed* $\underset{\sim}{x} \varepsilon \chi^*$,

$$
E\varphi_n(\underset{\sim}{x}) = \varphi_\mu(\underset{\sim}{x}) P_n(\underset{\sim}{x}), \tag{2.17}
$$

and

$$
E\varphi_n^2(\underset{\sim}{x}) = [\theta(\underset{\sim}{x}) - \varphi_\mu^2(\underset{\sim}{x})]\xi_n(\underset{\sim}{x}) + \varphi_\mu^2(\underset{\sim}{x}) P_n(\underset{\sim}{x}). \tag{2.18}
$$

Proof. For notational simplicity we let $H_i = h(X_{i,r+1})$ and suppress the fixed argument $\underline{x}$ ($\varepsilon\ \chi^*$) in all functions defined above. Then letting

$$\underline{U}_n = (M_1, M_2, \cdots, M_n)$$

and noting that for each i the joint distribution of the X_{ij}'s is invariant under permutations we have

$$E(M_iH_i \mid \underline{U}_n) = E(M_iH_i \mid M_i) = M_i\varphi_\mu, \tag{2.19}$$

by (2.9), and

$$E(M_i^2H_i^2 \mid \underline{U}_n) = E(M_i^2H_i^2 \mid M_i) = M_i\theta. \tag{2.20}$$

Also for $i \neq q$,

$$\begin{aligned} E(M_iM_qH_iH_q \mid \underline{U}_n) &= E\{M_iH_iE(M_qH_q \mid H_i, M_i, M_q) \mid M_i, M_q\} \\ &= E\{M_iH_iE(M_qH_q \mid M_q) \mid M_i, M_q\} \\ &= E(M_iH_i \mid M_i)E(M_qH_q \mid M_q) = M_iM_q\varphi_\mu^2. \end{aligned} \tag{2.21}$$

Hence

$$E\varphi_n = E\left\{\frac{1}{\bar{M}_n}\sum_{i=1}^n E(M_iH_i \mid \underline{U}_n) \,\middle|\, \bar{M}_n > 0\right\} P_n = \varphi_\mu P_n, \tag{2.22}$$

and

$$\begin{aligned} E\varphi_n^2 &= E\left\{\frac{1}{\bar{M}_n^2}\sum_{i=1}^n E(M_i^2H_i^2 \mid \underline{U}_n) \,\middle|\, \bar{M}_n > 0\right\} P_n \\ &\quad + E\left\{\frac{1}{\bar{M}_n^2}\sum_{i\neq q}^n\sum^n E(M_iM_qH_iH_q \mid \underline{U}_n) \,\middle|\, \bar{M}_n > 0\right\} P_n, \\ &= \theta E\left\{\frac{1}{\bar{M}_n} \,\middle|\, \bar{M}_n > 0\right\} P_n + \varphi_\mu^2 E\left\{\frac{1}{\bar{M}_n^2}\sum_{i\neq q}^n\sum^n M_iM_q \,\middle|\, \bar{M}_n > 0\right\} P_n. \end{aligned} \tag{2.23}$$

Now since

$$\frac{1}{\bar{M}_n^2}\sum_{i\neq q}^n\sum^n M_iM_q = 1 - \frac{1}{\bar{M}_n}, \tag{2.24}$$

and

$$\xi_n = EV_n = E\left\{\frac{1}{\bar{M}_n} \,\middle|\, \bar{M}_n > 0\right\} P_n, \tag{2.25}$$

we have

$$E\varphi_n^2 = \theta\xi_n + \varphi_\mu^2[P_n - \xi_n], \tag{2.26}$$

and the proof of the lemma is complete.

LEMMA 2. *For any* $\underline{x}\ \varepsilon\ \chi^*$,

$$\lim_{n\to\infty} P_n(\underline{x}) = 1, \tag{2.27}$$

and

(2.28) $$\lim_{n\to\infty} \xi_n(\underset{\sim}{x}) = 0.$$

Proof. Let

(2.29) $$p(\underset{\sim}{x}) = \text{Prob}\,\{M_1(\underset{\sim}{x}) = 1\}.$$

Then $\underset{\sim}{x} \varepsilon \chi^*$ implies $p(\underset{\sim}{x}) > 0$. Now for any fixed $\underset{\sim}{x}$, $\bar{M}_n(\underset{\sim}{x})$ is a binomial variable with parameters n and $p(\underset{\sim}{x})$. Hence for sufficiently large n, $\bar{M}_n(\underset{\sim}{x})$ will be arbitrarily large with probability arbitrarily close to 1 for any $\underset{\sim}{x} \varepsilon \chi^*$. This implies that for $\underset{\sim}{x} \varepsilon \chi^*$,

(2.30) $$\lim_{n\to\infty} P_n(\underset{\sim}{x}) = \lim_{n\to\infty} \text{Prob}\,\{\bar{M}_n(\underset{\sim}{x}) > 0\} = 1,$$

and

(2.31) $$V_n(\underset{\sim}{x}) \to 0, \qquad \text{in probability},$$

as $n \to \infty$. Now (2.31), together with the fact that $0 \leqq V_n(\underset{\sim}{x}) \leqq 1$, implies

(2.32) $$\lim_{n\to\infty} \xi_n(\underset{\sim}{x}) = \lim_{n\to\infty} EV_n(\underset{\sim}{x}) = 0,$$

for $\underset{\sim}{x} \varepsilon \chi^*$, which completes the proof of the lemma.

We now prove the following theorem:

THEOREM 1. *If the a priori probability measure space* $(\Omega, \mathfrak{A}, \mu)$ *is such that* (2.2) *is satisfied, then*

(2.33) $$\lim_{n\to\infty} R(\varphi_n) = R(\varphi_\mu).$$

Proof. We first observe that

(2.34) $$R(\varphi_n) = E[\varphi_n(\underset{\sim}{X}) - \Lambda]^2 = E\varphi_n^2(\underset{\sim}{X}) - 2E[\Lambda\varphi_n(\underset{\sim}{X})] + E\Lambda^2,$$

provided that all of the terms on the right are finite. Now since $\underset{\sim}{X}$ and Λ are independent of $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \cdots, \underset{\sim}{X}_n$, we have

(2.35) $$E(\varphi_n^2(\underset{\sim}{X}) \mid \underset{\sim}{X} = \underset{\sim}{x}) = E\varphi_n^2(\underset{\sim}{x}),$$

and

(2.36) $$\begin{aligned} E(\Lambda\varphi_n(\underset{\sim}{X}) \mid \underset{\sim}{X} = \underset{\sim}{x}) &= E\{\Lambda E(\varphi_n(\underset{\sim}{X}) \mid \Lambda, \underset{\sim}{X}) \mid \underset{\sim}{X} = \underset{\sim}{x}\} \\ &= E\{\Lambda E(\varphi_n(\underset{\sim}{X}) \mid \underset{\sim}{X}) \mid \underset{\sim}{X} = \underset{\sim}{x}\} \\ &= \varphi_\mu(\underset{\sim}{x}) E\varphi_n(\underset{\sim}{x}), \end{aligned}$$

for all $\underset{\sim}{x} \varepsilon \chi^*$. Hence by (2.17) and (2.18) of Lemma 1 together with (2.34), (2.35) and (2.36) we have

(2.37) $$R(\varphi_n) = E\Lambda^2 + E[\theta(\underset{\sim}{X})\xi_n(\underset{\sim}{X})] - E[\varphi_\mu^2(\underset{\sim}{X})(P_n(\underset{\sim}{X}) + \xi_n(\underset{\sim}{X}))].$$

Now by (2.27) and (2.28) of Lemma 2, for all $\underset{\sim}{x} \varepsilon \chi^*$,

$$\lim_{n\to\infty} \theta(\underset{\sim}{x})\xi_n(\underset{\sim}{x}) = 0, \tag{2.38}$$

and

$$\lim_{n\to\infty} \varphi_\mu^2(\underset{\sim}{x})(P_n(\underset{\sim}{x}) + \xi_n(\underset{\sim}{x})) = \varphi_\mu^2(\underset{\sim}{x}). \tag{2.39}$$

Also since $0 \leqq \xi_n(\underset{\sim}{x}) \leqq 1$ and $0 \leqq P_n(\underset{\sim}{x}) \leqq 1$, we have

$$|\theta(\underset{\sim}{x})\xi_n(\underset{\sim}{x})| \leqq \theta(\underset{\sim}{x}), \tag{2.40}$$

and

$$|\varphi_\mu^2(\underset{\sim}{x})(P_n(\underset{\sim}{x}) + \xi_n(\underset{\sim}{x}))| \leqq 2\varphi_\mu^2(\underset{\sim}{x}), \tag{2.41}$$

and furthermore since (2.2) is satisfied,

$$E\theta(\underset{\sim}{X}) = Eh^2(X) < \infty, \tag{2.42}$$

and

$$E\varphi_\mu^2(\underset{\sim}{X}) < \infty. \tag{2.43}$$

Hence by the Lebesgue Dominated Convergence Theorem we may assert

$$\lim_{n\to\infty} R(\varphi_n) = E\Lambda^2 - E\varphi_\mu^2(\underset{\sim}{X}) = R(\varphi_\mu), \tag{2.44}$$

which is the desired result.

This result is "non-parametric" in the sense that we have assumed that the unknown a priori probability measure is defined over the class $\mathfrak{F}$ of *all* c.d.f.'s assigning probability 1 to the set χ. If we are willing to assume that some specific parametric subclass of $\mathfrak{F}$ is assigned *a priori* probability 1 then we may be able to find empirical Bayes estimators such as those discussed by Robbins in [1] which (presumably) are more efficient than (2.12) for such cases.

It seems likely that the empirical Bayes estimator φ_n given by (2.12) is relatively inefficient when n is small or r is large relative to n, since, in this case, $\bar{M}_n(\underset{\sim}{X})$ is small with high probability so that relatively few of the $\underset{\sim}{X}_i$'s contribute useful information to φ_n. This difficulty may be offset to some extent by replacing φ_n by an estimate of Λ based on the value of $\underset{\sim}{X}$ when $\bar{M}_n(\underset{\sim}{X})$ is small. Thus, for example, we may define a modified empirical Bayes estimator $\varphi_n^{(c)}$ for fixed $c \geqq 0$ by

$$\varphi_n^{(c)}(\underset{\sim}{x}) = \begin{cases} \varphi_n(\underset{\sim}{x}), & \bar{M}_n(\underset{\sim}{x}) > c, \\ \dfrac{1}{r}\sum_{j=1}^{r} h(x_j), & \bar{M}_n(\underset{\sim}{x}) \leqq c. \end{cases} \tag{2.45}$$

It is not difficult to verify that $\varphi_n^{(c)}$ has the same asymptotic properties as φ_n and that for small n, $R(\varphi_n^{(c)})$ tends to be smaller than $R(\varphi_n)$ except when the distribution of Λ is concentrated near zero.

If the vectors of prior observations $\underset{\sim}{X}_i$ are of the form $\underset{\sim}{X}_i = (X_{i1}, X_{i2}, \cdots,$

$X_{i,r+k_i}$) where $k_i \geqq 1$, $i = 1, 2, \cdots, n$, we may make use of the additional information available when $k_i > 1$ by substituting $M_i(\underset{\sim}{x})(w_i/k_i) \sum_{j=r+1}^{k_i} h(X_{i,j})$ for $M_i(\underset{\sim}{x})h(X_{i,r+1})$ and $\sum_{i=1}^{n} w_i M_i(\underset{\sim}{x})$ for $\bar{M}_n(\underset{\sim}{x})$ in (2.11), where $w_1, w_2, \cdots, w_n$ are positive numerical weights depending on the k_i's. It can be shown by arguments similar to those of Theorem 1 that whenever the w_i's are uniformly bounded above and below by positive constants then the risk of the resulting estimator approaches $R(\varphi_\mu)$ as n becomes large. However, in general when the k_i's are not all the same the optimum choice of $w_1, w_2, \cdots, w_n$ depends on the unknown a priori probability structure and hence cannot be determined.

In practice it may happen that the numbers of components in the $\underset{\sim}{X}_i$'s and in $\underset{\sim}{X}$ are all the same ($= r + 1$, say) so that in order to use the estimator φ_n given by (2.12) we must discard one of the components of $\underset{\sim}{X}$. This seems undesirable and suggests the use of the modified empirical Bayes estimator

$$\tilde{\varphi}_n(\underset{\sim}{x}) = \frac{1}{r+1} \sum_{j=1}^{r+1} \varphi_n(\underset{\sim}{x}^{(j)}), \tag{2.46}$$

where $\underset{\sim}{x}^{(j)} = (x_1, x_2, \cdots, x_{j-1}, x_{j+1}, \cdots, x_{r+1})$, $j = 1, 2, \cdots, r+1$ and φ_n is given by (2.12). To evaluate the performance of $\tilde{\varphi}_n(\underset{\sim}{x})$ we compare its risk with that of $\varphi_n(\underset{\sim}{x}^{(1)})$ which uses only r components of $\underset{\sim}{X}$, as follows: For all n,

$$\begin{aligned} R(\tilde{\varphi}_n) &= E\Lambda^2 + E\tilde{\varphi}_n^2(\underset{\sim}{X}) - 2E[\Lambda\tilde{\varphi}_n(\underset{\sim}{X})] \\ &= E\Lambda^2 + \frac{1}{r+1} E\varphi_n^2(\underset{\sim}{X}^{(1)}) \\ &\quad + \frac{r}{r+1} E[\varphi_n(\underset{\sim}{X}^{(1)})\varphi_n(\underset{\sim}{X}^{(2)})] - 2E[\Lambda\varphi_n(\underset{\sim}{X}^{(1)})] \\ &= R(\varphi_n) - \frac{1}{2}\frac{r}{r+1} E[\varphi_n(\underset{\sim}{X}^{(1)}) - \varphi_n(\underset{\sim}{X}^{(2)})]^2 \\ &\leqq R(\varphi_n), \end{aligned} \tag{2.47}$$

with equality holding only in the degenerate case where $\underset{\sim}{X}^{(1)} = \underset{\sim}{X}^{(2)}$ with probability 1. By arguments similar to those of Theorem 1 it may be shown that

$$\lim_{n\to\infty} R(\tilde{\varphi}_n) = R(\tilde{\varphi}_\mu) \leqq R(\varphi_\mu), \tag{2.48}$$

where $R(\tilde{\varphi}_\mu)$ is the risk of the estimator

$$\tilde{\varphi}_\mu(\underset{\sim}{x}) = \frac{1}{r+1} \sum_{j=1}^{r+1} E(\Lambda \mid \underset{\sim}{X}^{(j)} = \underset{\sim}{x}^{(j)}), \tag{2.49}$$

and $R(\varphi_\mu)$ is the risk of the ordinary Bayes estimator $\varphi_\mu(\underset{\sim}{x}^{(1)}) = E(\Lambda \mid \underset{\sim}{X}^{(1)} = \underset{\sim}{x}^{(1)})$ based on only r of the $r + 1$ components of $\underset{\sim}{X}$. The inequality in (2.48) will again be strict except in degenerate cases.

It would be interesting to compute the value of $R(\varphi_n)$ for various values of n for some specific a priori probability distribution in order to obtain some idea of the rate at which $R(\varphi_n)$ converges to $R(\varphi_\mu)$. Unfortunately, such computations are extremely lengthy even in the simplest non-trivial cases. However, reasonably good upper and lower bounds for $R(\varphi_n)$ may be computed without much difficulty by using simple bounds for $\xi_n(\underset{\sim}{x})$ obtained as follows:

$$\xi_n(\underset{\sim}{x}) = EV_n(\underset{\sim}{x}) = \sum_{s=1}^{n} \frac{1}{s}\, b(s; n, p(\underset{\sim}{x})) \tag{2.50}$$

where $b(s; n, p(\underset{\sim}{x}))$ is the probability of s successes occurring in n independent trials where the probability of success at each trial is $p(\underset{\sim}{x}) = \text{Prob}\,\{M_1(\underset{\sim}{x}) = 1\}$. Also

$$\begin{aligned} \sum_{s=1}^{n} \frac{1}{s+1}\, b(s; n, p(\underset{\sim}{x})) &= \frac{1}{(n+1)p(\underset{\sim}{x})} \sum_{j=1}^{n} b(s+1; n+1, p(\underset{\sim}{x})) \\ &= \frac{1}{(n+1)p(\underset{\sim}{x})} \{1 - [1 + np(\underset{\sim}{x})][1 - p(\underset{\sim}{x})]^n\}, \end{aligned} \tag{2.51}$$

for $\underset{\sim}{x} \,\varepsilon\, \chi^*$. Hence, letting

$$b_n(\underset{\sim}{x}) = \frac{1}{(n+1)p(\underset{\sim}{x})} \{1 - [1 + np(\underset{\sim}{x})][1 - p(\underset{\sim}{x})]^n\}, \tag{2.52}$$

and noting that $1/(s + 1) < 1/s \leqq 2/(s + 1) < 1$ for all $s \geqq 1$, we have

$$b_n(\underset{\sim}{x}) < \xi_n(\underset{\sim}{x}) \leqq 2b_n(\underset{\sim}{x}) \leqq 1, \tag{2.53}$$

for $\underset{\sim}{x} \,\varepsilon\, \chi^*$.

Suppose now that we wish to estimate the expected value of a non-negative integer-valued random variable X (i.e., we set $h(x) \equiv x$ and $\chi = \{0, 1, 2, \cdots\}$), and suppose that $r = 1$ so that $\underset{\sim}{X} = X$ and $\underset{\sim}{X}_i = (X_{i1}, X_{i2})$ $i = 1, 2, \cdots, n$. For this problem we may compute the upper bounds based on (2.53) for the risks of the estimators φ_n and $\varphi_n^{(0)}$ given respectively by (2.12) and (2.45) with $c = 0$, for the particular a priori probability measure μ which assigns probability 1 to the family of Poisson c.d.f.'s

$$F(x \mid \lambda) = \sum_{t=0}^{[x]} \frac{1}{t!} e^{-\lambda t} \lambda^t; \qquad x \geqq 0, \qquad \lambda > 0, \tag{2.54}$$

and which induces the Γ-distribution

$$G(\lambda) = \int_0^\lambda \frac{\alpha^\beta}{\Gamma(\beta)} u^{\beta-1} e^{-\alpha u}\, du; \qquad \alpha, \beta > 0, \tag{2.55}$$

as the c.d.f. of Λ. In Table I below we compare the upper and lower bounds for $R(\varphi_n)$ and $R(\varphi_n^{(0)})$ with the risk $R(\varphi_\mu)$ of the Bayes estimator φ_μ and with the risk $R(x)$ of the classical estimator x. These quantities are computed for six values of n, for $\alpha = 2$ and $\beta = 10$ (i.e., $E\Lambda = 5$ and $\text{Var}\Lambda = 2.5$). In Table II we compare the same quantities for the case where μ assigns a priori probability 1 to the particular member of the family of c.d.f.'s (2.54) for which $\lambda = \lambda_0 = 5$ (i.e., $\Lambda \equiv \lambda_0$ with probability 1).

TABLE I

n	$R(\varphi_n)$		$R(\varphi_n^{(0)})$		$R(\varphi_\mu)$	$R(x)$
	Lower Bound	Upper Bound	Lower Bound	Upper Bound		
15	10.70	12.51	5.21	7.02	1.67	5.00
30	6.74	8.32	4.43	6.01	1.67	5.00
60	4.46	5.54	3.50	4.58	1.67	5.00
120	3.20	3.87	2.79	3.46	1.67	5.00
240	2.50	2.89	2.32	2.72	1.67	5.00
480	2.12	2.34	2.05	2.27	1.67	5.00

TABLE II

n	$R(\varphi_n)$		$R(\varphi_n^{(0)})$		$R(\varphi_\mu)$	$R(x)$
	Lower Bound	Upper Bound	Lower Bound	Upper Bound		
15	6.06	7.44	3.68	5.06	0	5.00
30	2.94	4.04	2.57	3.66	0	5.00
60	1.46	2.17	1.57	2.28	0	5.00
120	0.74	1.16	0.91	1.32	0	5.00
240	0.38	0.61	0.51	0.74	0	5.00
480	0.19	0.32	0.27	0.40	0	5.00

3. Estimation: the continuous case. In this section we extend the methods of Sec. 2 to cover the case where the observed X's possess absolutely continuous distribution functions. As before these results will be "non-parametric" in the sense that the unknown a priori probability measure is assumed to be defined over the class of all absolutely continuous c.d.f.'s subject only to some conditions on the existence of moments.

Let $\mathfrak{F} = \{F(x \mid \omega) : \omega \in \Omega\}$ be the collection of all absolutely continuous c.d.f.'s where $\Omega = \{\omega\}$ is an abstract indexing set, and let $(\Omega, \mathfrak{A}, \mu)$ be an a priori probability measure space where $\mathfrak{A}$ is a σ-algebra of subsets of Ω. Then there exists a function $f(u \mid \omega)$ defined on (reals) $\times\ \Omega$ such that

$$F(x \mid \omega) = \int_{-\infty}^{x} f(u \mid \omega)\, du, \tag{3.1}$$

for each $\omega \in \Omega$. We assume that $(\Omega, \mathfrak{A})$ is such that the function $f(u \mid \omega)$ is a measurable function on the product space (reals) $\times\ \Omega$.

Let $Y = Y(\omega)$ be the Ω-valued random variable which is the identity mapping of Ω onto itself. Let the (real) random variables $X_1, X_2, \cdots, X_r$ be conditionally independent and identically distributed according to $F(x \mid \omega)$ given that $Y = \omega$, and let $\underset{\sim}{X} = (X_1, X_2, \cdots, X_r)$. Then the unconditional joint c.d.f. of X_1, $X_2, \cdots, X_r$ will be

$$F(x_1, \cdots, x_r) = F(\underset{\sim}{x}) = \int_\Omega \prod_{j=1}^{r} F(x_j \mid \omega)\, d\mu$$

$$= \int_\Omega \prod_{j=1}^{r} \int_{-\infty}^{x_j} f(u_j \mid \omega)\, du_j\, d\mu \tag{3.2}$$
$$= \int_{-\infty}^{x_r} \cdots \int_{-\infty}^{x_1} f(u_1, \cdots, u_r)\, du_1 \cdots du_r,$$

where

$$f(x_1, \cdots, x_r) = f(\underset{\sim}{x}) = \int_\Omega \prod_{j=1}^{r} f(x_j \mid \omega)\, d\mu. \tag{3.3}$$

Thus $f(\underset{\sim}{x})$ is the joint unconditional probability density function of $X_1, \cdots, X_r$. Now for a given measurable function $h(x)$ we define the random variable Λ by

$$\Lambda = \Lambda(Y) = E(h(X) \mid Y), \tag{3.4}$$

where X is a generic representative of the X_j's. As in Sec. 2 we assume that

$$Eh^2(X) < \infty, \tag{3.5}$$

which implies $E\Lambda^2 < \infty$ and hence the existence of all conditional expectations with which we will be concerned. Now if we wish to estimate the value of Λ using $\underset{\sim}{X}$ where the risk is the expected squared error, then, as before, the essentially unique Bayes estimator is

$$\psi_\mu(\underset{\sim}{x}) = E(\Lambda \mid \underset{\sim}{X} = \underset{\sim}{x}). \tag{3.6}$$

(In this section when conditional expectations are regarded as functions of the values of random variables it will be understood that we mean the essentially unique Borel measurable version which is set equal to zero whenever arbitrariness is possible on sets of positive Lebesgue measure.)

As in Sec. 2 we introduce random variables $Y_1, \cdots, Y_n$ independent of each other and of Y such that each has the same distribution as Y. We also introduce the random vectors of prior observations $\underset{\sim}{X}_i = (X_{i1}, \cdots, X_{i,r+1})$, $i = 1, 2, \cdots, n$, where the $\underset{\sim}{X}_i$'s are independent of each other and of $\underset{\sim}{X}$ and where for each i the X_{ij}'s are conditionally independent and identically distributed according to $F(x \mid \omega_i)$ given that $Y_i = \omega_i$. As before we let $\underset{\sim}{X}_i^{(r)} = (X_{i1}, X_{i2}, \cdots, X_{ir})$, $i = 1, 2, \cdots, n$, and we note that for each i

$$E(h(X_{i,r+1}) \mid \underset{\sim}{X}_i^{(r)} = \underset{\sim}{x}) = E(\Lambda \mid \underset{\sim}{X} = \underset{\sim}{x}) = \psi_\mu(\underset{\sim}{x}). \tag{3.7}$$

In order to make use of the results of Sec. 2 we must discretize the X's in some way. To this end we consider the double sequence of half-open intervals

$$I_t^{(n)} = \left[\frac{tc}{n^{1-\delta/r}}, \frac{(t+1)c}{n^{1-\delta/r}}\right), \qquad t = 0, \pm 1, \pm 2, \cdots; \qquad n = 1, 2, \cdots, \tag{3.8}$$

where $c > 0$ and $0 < \delta < 1$. For each n we partition r-dimensional euclidean space into a countable sequence of non-overlapping hypercubes $C_j^{(n)}$, $j = 1, 2, \cdots$, of the form

$$C_j^{(n)} = I_{t_{1j}}^{(n)} \times I_{t_{2j}}^{(n)} \times \cdots \times I_{t_{rj}}^{(n)}, \qquad j = 1, 2, \cdots, \tag{3.9}$$

where the t_{ij}'s are suitably chosen integers. Then for each n and each r-component numerical vector $\underset{\sim}{x} = (x_1, x_2, \cdots, x_r)$ we let $C^{(n)}(\underset{\sim}{x})$ be the unique member of the sequence (3.9) containing $\underset{\sim}{x}$. As before, we designate by $\underset{\sim}{x}_{(q)}$, $q = 1, 2, \cdots, m(\underset{\sim}{x})$, $m(\underset{\sim}{x}) \geqq 1$, the distinct vectors obtained by permuting the components of $\underset{\sim}{x}$. Then proceeding by analogy with Sec. 2 we define the random functions

$$M_i^{(n)}(\underset{\sim}{x}) = \begin{cases} 1, & \text{if there exists a } q,\ 1 \leqq q \leqq m(\underset{\sim}{x}), \\ & \text{such that } \underset{\sim}{X}_i^{(r)} \varepsilon\, C^{(n)}(\underset{\sim}{x}_{(q)}), \\ 0, & \text{otherwise,} \end{cases} \tag{3.10}$$

for $i = 1, 2, \cdots, n$, and

$$\bar{M}^{(n)}(\underset{\sim}{x}) = \sum_{i=1}^{r} M_i^{(n)}(\underset{\sim}{x}), \tag{3.11}$$

and finally we define the empirical Bayes estimator $\psi_n(\underset{\sim}{x})$ by

$$\psi_n(\underset{\sim}{x}) = \begin{cases} \dfrac{1}{\bar{M}^{(n)}(\underset{\sim}{x})} \displaystyle\sum_{i=1}^{n} M_i^{(n)}(\underset{\sim}{x}) h(X_{i,r+1}), & \bar{M}^{(n)}(\underset{\sim}{x}) > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{3.12}$$

Before we can show that $\lim_{n\to\infty} R(\psi_n) = R(\psi_\mu)$ we must prove three preliminary lemmas.

We first let

$$\varphi^{(n)}(\underset{\sim}{x}) = E(h(X_{1,r+1}) \mid M_1^{(n)}(\underset{\sim}{x}) = 1), \qquad n = 1, 2, \cdots, \tag{3.13}$$

and

$$\theta^{(n)}(\underset{\sim}{x}) = E(h^2(X_{1,r+1}) \mid M_1^{(n)}(\underset{\sim}{x}) = 1), \qquad n = 1, 2, \cdots. \tag{3.14}$$

For each n, $\varphi^{(n)}(\underset{\sim}{x})$ and $\theta^{(n)}(\underset{\sim}{x})$ are analogous respectively to $\varphi_\mu(\underset{\sim}{x})$ and $\theta(\underset{\sim}{x})$ of Sec. 2. We now prove the following lemma:

LEMMA 3. *For almost all $\underset{\sim}{x}$ such that $f(\underset{\sim}{x}) > 0$,*

$$\lim_{n\to\infty} \varphi^{(n)}(\underset{\sim}{x}) = E(h(X_{1,r+1}) \mid \underset{\sim}{X}_1^{(r)} = \underset{\sim}{x}) = \psi_\mu(\underset{\sim}{x}), \tag{3.15}$$

and

$$\lim_{n\to\infty} \theta^{(n)}(\underset{\sim}{x}) = E(h^2(X_{1,r+1}) \mid \underset{\sim}{X}_1^{(r)} = \underset{\sim}{x}). \tag{3.16}$$

Proof. Let the joint density function for the $r + 1$ components of $\underset{\sim}{X}_i$ (for any fixed i) be written as

$$f(\underset{\sim}{x}, x_{r+1}) = f(x_1, x_2, \cdots, x_{r+1}) = \int_\Omega \prod_{j=1}^{r+1} f(x_j \mid \omega)\, d\mu. \tag{3.17}$$

Then since $f(\underset{\sim}{x}, x_{r+1}) = f(\underset{\sim}{x}_{(q)}, x_{r+1})$, $q = 1, 2, \cdots, m(\underset{\sim}{x})$, we have for almost all $\underset{\sim}{x}$

$$\varphi^{(n)}(\underset{\sim}{x}) = \frac{\int_{C^{(n)}(\underset{\sim}{x})} \int_{-\infty}^{\infty} h(v)f(\underset{\sim}{u}, v)\, dv\, d\underset{\sim}{u}}{\int_{C^{(n)}(\underset{\sim}{x})} f(\underset{\sim}{u})\, d\underset{\sim}{u}}, \tag{3.18}$$

provided the denominator is not zero. (If the denominator is zero, we conventionally set $\varphi^{(n)}(\underset{\sim}{x}) = 0$). Now for each $\underset{\sim}{x}$, $C^{(n)}(\underset{\sim}{x})$, $n = 1, 2, \cdots$, is a sequence of hypercubes converging regularly to the point $\underset{\sim}{x}$. Hence for almost all $\underset{\sim}{x}$

$$\lim_{n\to\infty} \frac{1}{\text{Volume } (C^{(n)}(\underset{\sim}{x}))} \int_{C^{(n)}(\underset{\sim}{x})} \int_{-\infty}^{\infty} h(v)f(\underset{\sim}{u}, v)\, dv\, d\underset{\sim}{u} = \int_{-\infty}^{\infty} h(v)f(\underset{\sim}{x}, v)\, dv, \tag{3.19}$$

by a well-known theorem on the differentiation of multiple Lebesgue integrals. A similar limit holds for the denominator of (3.18) so that

$$\begin{aligned} \lim_{n\to\infty} \varphi^{(n)}(\underset{\sim}{x}) &= \frac{\int_{-\infty}^{\infty} h(v)f(\underset{\sim}{x}, v)\, dv}{f(\underset{\sim}{x})} \\ &= E(h(X_{1,r+1}) \mid \underset{\sim}{X}_1^{(r)} = \underset{\sim}{x}) = \psi_\mu(\underset{\sim}{x}), \end{aligned} \tag{3.20}$$

for almost all $\underset{\sim}{x}$ such that $f(\underset{\sim}{x}) > 0$. Similarly

$$\begin{aligned} \lim_{n\to\infty} \theta^{(n)}(\underset{\sim}{x}) &= \lim_{n\to\infty} \frac{\int_{C^{(n)}(\underset{\sim}{x})} \int_{-\infty}^{\infty} h^2(v)f(\underset{\sim}{u}, v)\, dv\, d\underset{\sim}{u}}{\int_{C^{(n)}(\underset{\sim}{x})} f(\underset{\sim}{u})\, d\underset{\sim}{u}} \\ &= \frac{\int_{-\infty}^{\infty} h^2(v)f(\underset{\sim}{x}, v)\, dv}{f(\underset{\sim}{x})} \\ &= E(h^2(X_{1,r+1}) \mid \underset{\sim}{X}_1^{(r)} = \underset{\sim}{x}), \end{aligned} \tag{3.21}$$

for almost all $\underset{\sim}{x}$ such that $f(\underset{\sim}{x}) > 0$, so that (3.15) and (3.16) are verified.

For $n = 1, 2, \cdots$, let

$$V^{(n)}(\underset{\sim}{x}) = \begin{cases} \dfrac{1}{\bar{M}^{(n)}(\underset{\sim}{x})}, & \bar{M}^{(n)}(\underset{\sim}{x}) > 0, \\ 0, & \text{otherwise}, \end{cases} \tag{3.22}$$

$$\xi^{(n)}(\underset{\sim}{x}) = EV^{(n)}(\underset{\sim}{x}), \tag{3.23}$$

and

$$P^{(n)}(\underset{\sim}{x}) = \text{Prob } \{\bar{M}^{(n)}(\underset{\sim}{x}) > 0\}. \tag{3.24}$$

For each n, $V^{(n)}(\underset{\sim}{x})$, $\xi^{(n)}(\underset{\sim}{x})$, and $P^{(n)}(\underset{\sim}{x})$ are analogous respectively to $V_n(\underset{\sim}{x})$, $\xi_n(\underset{\sim}{x})$, and $P_n(\underset{\sim}{x})$ of Sec. 2.

LEMMA 4. *For almost all* $\underset{\sim}{x}$ *such that* $f(\underset{\sim}{x}) > 0$

$$\lim_{n\to\infty} \xi^{(n)}(\underset{\sim}{x}) = 0, \tag{3.25}$$

and

$$\lim_{n\to\infty} P^{(n)}(\underset{\sim}{x}) = 1. \tag{3.26}$$

Proof. Let

$$p^{(n)}(\underset{\sim}{x}) = \int_{C^{(n)}(\underset{\sim}{x})} f(\underset{\sim}{u})\, d\underset{\sim}{u}. \tag{3.27}$$

We have remarked that for almost all $\underset{\sim}{x}$

$$\lim_{n\to\infty} \frac{1}{\text{Volume }(C^{(n)}(\underset{\sim}{x}))} \int_{C^{(n)}(\underset{\sim}{x})} f(\underset{\sim}{u})\, d\underset{\sim}{u} = f(\underset{\sim}{x}). \tag{3.28}$$

By (3.8) and (3.9) we have

$$\text{Volume }(C^{(n)}(\underset{\sim}{x})) = \frac{c^r}{n^{1-\delta}}, \tag{3.29}$$

so that whenever $f(\underset{\sim}{x}) > 0$ we may write

$$p^{(n)}(\underset{\sim}{x}) = \frac{c^r}{n^{1-\delta}} f(\underset{\sim}{x})(1 + \epsilon_n(\underset{\sim}{x})), \tag{3.30}$$

where $\lim_{n\to\infty} \epsilon_n(\underset{\sim}{x}) = 0$ for almost all $\underset{\sim}{x}$.

Referring to the upper bound (2.53) obtained for $\xi_n(\underset{\sim}{x})$ in Sec. 2 and noting that $m(\underset{\sim}{x})p^{(n)}(\underset{\sim}{x})$ plays the same role as $p(\underset{\sim}{x})$ and that $m(\underset{\sim}{x}) \geqq 1$, we see that

$$\xi^{(n)}(\underset{\sim}{x}) \leqq \frac{2}{np^{(n)}(\underset{\sim}{x})} = \frac{2}{n^{\delta}c^r f(\underset{\sim}{x})(1 + \epsilon_n(\underset{\sim}{x}))}, \tag{3.31}$$

for all $\underset{\sim}{x}$ such that $f(\underset{\sim}{x}) > 0$. Hence

$$\lim_{n\to\infty} \xi^{(n)}(\underset{\sim}{x}) = 0, \tag{3.32}$$

for almost all $\underset{\sim}{x}$ such that $f(\underset{\sim}{x}) > 0$ and (3.25) is verified. Now

$$P^{(n)}(\underset{\sim}{x}) = \text{Prob }\{\bar{M}^{(n)}(\underset{\sim}{x}) > 0\} \geqq 1 - [1 - p^{(n)}(\underset{\sim}{x})]^n, \tag{3.33}$$

and since for all $u \leqq 1$, $e^{-u} \geqq 1 - u \geqq 0$, we may write

$$P^{(n)}(\underset{\sim}{x}) \geqq 1 - e^{-np^{(n)}(\underset{\sim}{x})}. \tag{3.34}$$

Hence for almost all $\underset{\sim}{x}$ such that $f(\underset{\sim}{x}) > 0$,

$$\liminf_{n\to\infty} P^{(n)}(\underset{\sim}{x}) \geqq 1 - \lim_{n\to\infty} e^{-n^{\delta}c^r f(\underset{\sim}{x})(1+\epsilon_n(\underset{\sim}{x}))} = 1, \tag{3.35}$$

which implies (3.26).

We now prove a simple convergence lemma.

LEMMA 5. *If* $(S, \mathcal{B}, \nu)$ *is a measure space,* $\{f_n\}$ *and* $\{g_n\}$ *are sequences of non-negative integrable functions,* f *is an integrable function and* g *is a function such that*

$$\text{(3.36)} \quad \begin{aligned} &\text{(i)} \quad \lim_{n\to\infty} f_n = f, \text{ a.e.}; \qquad \lim_{n\to\infty} g_n = g, \text{ a.e.};\\ &\text{(ii)} \quad g_n \leqq f_n, \quad \text{all} \quad n;\\ &\text{(iii)} \quad \limsup_{n\to\infty} \int f_n \, d\nu \leqq \int f \, d\nu; \end{aligned}$$

then g *is integrable and*

$$\text{(3.37)} \qquad \lim_{n\to\infty} \int g_n \, d\nu = \int g \, d\nu.$$

Proof. By (i), (ii) and Fatou's Lemma, g is integrable,

$$\text{(3.38)} \qquad \liminf_{n\to\infty} \int g_n \, d\nu \geqq \int g \, d\nu,$$

and (noting (iii)),

$$\text{(3.39)} \qquad \lim_{n\to\infty} \int f_n \, d\nu = \int f \, d\nu.$$

Furthermore,

$$\text{(3.40)} \qquad \begin{aligned} \limsup_{n\to\infty} \int (f_n - g_n) \, d\nu &\leqq \limsup_{n\to\infty} \int f_n \, d\nu - \liminf_{n\to\infty} \int g_n \, d\nu\\ &\leqq \int (f - g) \, d\nu, \end{aligned}$$

and by (i) and (ii), $\lim_{n\to\infty} (f_n - g_n) = f - g$ and $(f_n - g_n) \geqq 0$, so that applying Fatou's Lemma again we obtain

$$\text{(3.41)} \qquad \lim_{n\to\infty} \int (f_n - g_n) \, d\nu = \int f \, d\nu - \int g \, d\nu.$$

The desired result then follows from (3.39) and (3.41).

THEOREM 3. *If the a priori probability measure space* $(\Omega, \mathcal{A}, \mu)$ *is such that* (3.5) *is satisfied, then*

$$\text{(3.42)} \qquad \lim_{n\to\infty} R(\psi_n) = R(\psi_\mu).$$

Proof. Since $\underset{\sim}{X}$ and Λ are independent of $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \cdots, \underset{\sim}{X}_n$, we have (as in Sec. 2)

$$\text{(3.43)} \qquad E(\psi_n^2(\underset{\sim}{X}) \mid \underset{\sim}{X} = \underset{\sim}{x}) = E\psi_n^2(\underset{\sim}{x}),$$

and

$$\text{(3.44)} \qquad E(\Lambda\psi_n^2(\underset{\sim}{X}) \mid \underset{\sim}{X} = \underset{\sim}{x}) = \psi_\mu(\underset{\sim}{x}) E\psi_n(\underset{\sim}{x}),$$

for almost all $\underset{\sim}{x}$. Now applying Lemma 1 of Sec. 2 to $\psi_n(\underset{\sim}{x})$ (mutatis mutandis) we obtain

(3.45) $$E\psi_n(\underset{\sim}{x}) = \varphi^{(n)}(\underset{\sim}{x})P^{(n)}(\underset{\sim}{x}),$$

and

(3.46) $$E\psi_n^2(\underset{\sim}{x}) = [\theta^{(n)}(\underset{\sim}{x}) - \varphi^{(n)^2}(\underset{\sim}{x})]\xi^{(n)}(\underset{\sim}{x}) + \varphi^{(n)^2}(\underset{\sim}{x})P^{(n)}(\underset{\sim}{x}),$$

for almost all $\underset{\sim}{x}$. Hence

(3.47) $$E\psi_n^2(\underset{\sim}{X}) = E[\theta^{(n)}(\underset{\sim}{X})\xi^{(n)}(\underset{\sim}{X})] - E[\varphi^{(n)^2}(\underset{\sim}{X})\xi^{(n)}(\underset{\sim}{X})] + E[\varphi^{(n)^2}(\underset{\sim}{X})P^{(n)}(\underset{\sim}{X})],$$

and

(3.48) $$E[\Lambda\psi_n(\underset{\sim}{X})] = E[\psi_\mu(\underset{\sim}{X})\varphi^{(n)}(\underset{\sim}{X})P^{(n)}(\underset{\sim}{X})].$$

Now in terms of the hypercubes $C_j^{(n)}$, $j = 1, 2, \cdots$, defined by (3.9) we may write

(3.49) $$E\varphi^{(n)^2}(\underset{\sim}{X}) = \sum_{j=1}^{\infty} \frac{\left[\int_{C_j^{(n)}} \int_{-\infty}^{\infty} h(v)f(\underset{\sim}{u}, v)\,dv\,d\underset{\sim}{u}\right]^2}{\int_{C_j^{(n)}} f(\underset{\sim}{u})\,d\underset{\sim}{u}},$$

and

(3.50) $$\begin{aligned} E\psi_\mu^2(\underset{\sim}{X}) &= \int \frac{\left[\int_{-\infty}^{\infty} h(v)f(\underset{\sim}{u}, v)\,dv\right]^2}{f(\underset{\sim}{u})}\,d\underset{\sim}{u} \\ &= \sum_{j=1}^{\infty} \int_{C_j^{(n)}} \frac{\left[\int_{-\infty}^{\infty} h(v)f(\underset{\sim}{u}, v)\,dv\right]^2}{f(\underset{\sim}{u})}\,d\underset{\sim}{u}, \end{aligned}$$

where it is understood that the ratios appearing in these expressions are to be replaced by zero whenever their denominators vanish. Furthermore, for each j we have

(3.51) $$\left[\int_{C_j^{(n)}} \int_{-\infty}^{\infty} h(v)f(\underset{\sim}{u}, v)dv\,d\underset{\sim}{u}\right]^2 \leqq \int_{C_j^{(n)}} f(\underset{\sim}{u})\,d\underset{\sim}{u} \cdot \int_{C_j^{(n)}} \frac{\left[\int_{-\infty}^{\infty} h(v)f(\underset{\sim}{u}, v)\,dv\right]^2}{f(\underset{\sim}{u})}\,d\underset{\sim}{u},$$

by the Schwartz inequality. Hence

(3.52) $$E\varphi^{(n)^2}(\underset{\sim}{X}) \leqq E\psi_\mu^2(\underset{\sim}{X}), \qquad \text{all } n,$$

and since $0 \leqq P^{(n)}(\underset{\sim}{x}) \leqq 1$,

(3.53) $$E[\varphi^{(n)^2}(\underset{\sim}{X})P^{(n)}(\underset{\sim}{X})] \leqq E\psi_\mu^2(\underset{\sim}{X}), \qquad \text{all } n.$$

Now by (3.15) of Lemma 3 and (3.26) of Lemma 4

$$\lim_{n\to\infty} \varphi^{(n)^2}(\underset{\sim}{x})P^{(n)}(\underset{\sim}{x}) = \psi_\mu^2(\underset{\sim}{x}), \tag{3.54}$$

for almost all $\underset{\sim}{x}$ such that $f(\underset{\sim}{x}) > 0$. Hence by Lemma 5 (with $f_n = g_n = \varphi^{(n)^2}(\underset{\sim}{x})P^{(n)}(\underset{\sim}{x})$) we have

$$\lim_{n\to\infty} E[\varphi^{(n)^2}(\underset{\sim}{X})P^{(n)}(\underset{\sim}{X})] = E\psi_\mu^2(\underset{\sim}{X}). \tag{3.55}$$

Now

$$E\theta^{(n)}(\underset{\sim}{X}) = Eh^2(X_{1,r+1}) = E\{E(h^2(X_{1,r+1}) \mid \underset{\sim}{X}_1^{(r)})\}, \tag{3.56}$$

for all n, and by (3.16) of Lemma 3

$$\lim_{n\to\infty} \theta^{(n)}(\underset{\sim}{x}) = E(h^2(X_{1,r+1}) \mid \underset{\sim}{X}_1^{(r)} = \underset{\sim}{x}), \tag{3.57}$$

for almost all $\underset{\sim}{x}$ such that $f(\underset{\sim}{x}) > 0$. Also, $0 \leqq \xi^{(n)}(\underset{\sim}{x}) \leqq 1$ and by (3.25) of Lemma 4

$$\lim_{n\to\infty} \xi^{(n)}(\underset{\sim}{x}) = 0, \tag{3.58}$$

for almost all $\underset{\sim}{x}$ such that $f(\underset{\sim}{x}) > 0$, so that by Lemma 5 (with $f_n = \theta^{(n)}(\underset{\sim}{x})$ and $g_n = \theta^{(n)}(\underset{\sim}{x})\xi^{(n)}(\underset{\sim}{x})$) we have

$$\lim_{n\to\infty} E[\theta^{(n)}(\underset{\sim}{X})\xi^{(n)}(\underset{\sim}{X})] = 0. \tag{3.59}$$

Similarly, in view of (3.16) of Lemma 3 and (3.52),

$$\lim_{n\to\infty} E[\varphi^{(n)^2}(\underset{\sim}{X})\xi^{(n)}(\underset{\sim}{X})] = 0. \tag{3.60}$$

Hence by (3.47), (3.55), (3.59), and (3.60)

$$\lim_{n\to\infty} E\psi_n^2(\underset{\sim}{X}) = E\psi_\mu^2(\underset{\sim}{X}). \tag{3.61}$$

Now for any fixed n and j the functions $\varphi^{(n)}(\underset{\sim}{x})$ and $P^{(n)}(\underset{\sim}{x})$ are constant for all $\underset{\sim}{x} \, \varepsilon \, C_j^{(n)}$ and we may designate their values by $\varphi_j^{(n)}$ and $P_j^{(n)}$ respectively. Then by (3.48) we have for each n

$$\begin{aligned} E[\Lambda\psi_n(\underset{\sim}{X})] &= E[\psi_\mu(\underset{\sim}{X})\varphi^{(n)}(\underset{\sim}{X})P^{(n)}(\underset{\sim}{X})] \\ &= \sum_{j=1}^{\infty} \varphi_j^{(n)}P_j^{(n)} \int_{C_j^{(n)}} \psi_\mu(\underset{\sim}{x})f(\underset{\sim}{x})\, d\underset{\sim}{x} \\ &= \sum_{j=1}^{\infty} \varphi_j^{(n)}P_j^{(n)} \int_{C_j^{(n)}} \int_{-\infty}^{\infty} h(v)f(\underset{\sim}{x}, v)\, dv\, d\underset{\sim}{x} \\ &= \sum_{j=1}^{\infty} \varphi_j^{(n)^2}P_j^{(n)} \int_{C_j^{(n)}} f(\underset{\sim}{x})\, d\underset{\sim}{x} \\ &= E[\varphi^{(n)^2}(\underset{\sim}{X})P^{(n)}(\underset{\sim}{X})], \end{aligned} \tag{3.62}$$

so that by (3.55)

$$\lim_{n\to\infty} E[\Lambda\psi_n(\underset{\sim}{X})] = E\psi_\mu^2(\underset{\sim}{X}). \tag{3.63}$$

Now

$$R(\psi_n) = E\Lambda^2 + E\psi_n^2(\underset{\sim}{X}) - 2E[\Lambda\psi_n(\underset{\sim}{X})], \tag{3.64}$$

so that by (3.61) and (3.63) we have

$$\begin{aligned} \lim_{n\to\infty} R(\psi_n) &= E\Lambda^2 + \lim_{n\to\infty} E\psi_n^2(\underset{\sim}{X}) - 2\lim_{n\to\infty} E[\Lambda\psi_n(\underset{\sim}{X})] \\ &= E\Lambda^2 - E\psi_\mu^2(\underset{\sim}{X}) = R(\psi_\mu), \end{aligned} \tag{3.65}$$

which was to be proved.

The estimation procedure introduced in this section contains an element of arbitrariness arising from the fact that the definition of the sequence of intervals $\{I_t^{(n)}\}$ involves the two constants c and δ whose values must be specified. The problem of the proper choice of c and δ will not, however, be considered further here.

The remarks made in Sec. 2 concerning various modifications of the estimator φ_n apply as well to the analogous modifications of the estimator ψ_n.

4. Hypothesis testing. The empirical Bayes estimation procedures introduced in the preceding sections may be applied to certain two-decision problems of the hypothesis-testing type. This is illustrated by the following two examples:

Example 1. (One-sided alternatives): Suppose that we wish to test a hypothesis about the value λ of the random variable Λ associated with the vector of observations $\underset{\sim}{X}$. In particular, suppose that we wish to test the hypothesis $H_0:\lambda < a$ versus the alternative hypothesis $H_1:\lambda \geqq a$. Let A_0 represent the action of accepting H_0 and let A_1 represent the action of accepting H_1. Then we may define a loss function L as follows:

$$L(A_i, \lambda) = \begin{cases} \max(0, \lambda - a), & i = 0, \\ -\min(0, \lambda - a), & i = 1. \end{cases} \tag{4.1}$$

In the decision theoretic framework this loss function is certainly no less reasonable than the classical zero-one loss function usually postulated for hypothesis-testing problems. Now for any decision function $\delta(x) = \text{Prob}\{A_1 \mid \underset{\sim}{X} = \underset{\sim}{x}\}$ = probability of rejecting H_0 when $\underset{\sim}{x}$ is observed, the risk involved in using δ is given by

$$\begin{aligned} R(\delta) &= E\{\delta(\underset{\sim}{X})L(A_1, \Lambda)\} + E\{[1 - \delta(\underset{\sim}{X})]L(A_0, \Lambda)\} \\ &= EL(A_0, \Lambda) - E\{\delta(\underset{\sim}{X})[L(A_0, \Lambda) - L(A_1, \Lambda)]\} \\ &= EL(A_0, \Lambda) - E\{\delta(\underset{\sim}{X})[\Lambda - a]\} \\ &= EL(A_0, \Lambda) - E\{\delta(\underset{\sim}{X})[E(\Lambda \mid \underset{\sim}{X}) - a]\}. \end{aligned} \tag{4.2}$$

Hence the Bayes decision function $\delta_\mu(\underset{\sim}{x})$ minimizing $R(\delta)$ is

$$\delta_\mu(\underset{\sim}{x}) = \begin{cases} 1, & E(\Lambda \mid \underset{\sim}{X} = \underset{\sim}{x}) > a, \\ 0, & \text{otherwise.} \end{cases} \tag{4.3}$$

Now we have seen in the previous sections that when the a priori probability measure (and hence the joint distribution of Λ and $\underset{\sim}{X}$) is unknown we may still be able, under certain circumstances, to find an empirical Bayes estimator $\varphi_n(\underset{\sim}{x})$ based on prior independent observations, such that

$$\lim_{n\to\infty} E[\varphi_n(\underset{\sim}{x}) - E(\Lambda \mid \underset{\sim}{X} = \underset{\sim}{x})]^2 = 0, \tag{4.4}$$

for all $\underset{\sim}{x}$ in some set S which is assigned probability 1 under the distribution of $\underset{\sim}{X}$. Now (4.4) implies that as $n \to \infty$,

$$\varphi_n(\underset{\sim}{x}) \to E(\Lambda \mid \underset{\sim}{X} = \underset{\sim}{x}), \qquad \text{in probability,} \tag{4.5}$$

for all $\underset{\sim}{x} \varepsilon S$, so that if we define the empirical Bayes decision function $\delta_n(\underset{\sim}{x})$ by

$$\delta_n(\underset{\sim}{x}) = \begin{cases} 1, & \varphi_n(\underset{\sim}{x}) > a, \\ 0, & \text{otherwise,} \end{cases} \tag{4.6}$$

we will have

$$\lim_{n\to\infty} E\delta_n(\underset{\sim}{x}) = \lim_{n\to\infty} \text{Prob } \{\varphi_n(\underset{\sim}{x}) > a\} = \delta_\mu(\underset{\sim}{x}), \tag{4.7}$$

for all $\underset{\sim}{x} \varepsilon S$ with the possible exception of values of $\underset{\sim}{x}$ for which $E(\Lambda \mid \underset{\sim}{X} = \underset{\sim}{x}) = a$. Hence

$$\begin{aligned} \lim_{n\to\infty} E\{\delta_n(\underset{\sim}{X})[\Lambda - a] \mid \underset{\sim}{X} = \underset{\sim}{x}\} &= \lim_{n\to\infty} [E\delta_n(\underset{\sim}{x})][E(\Lambda \mid \underset{\sim}{X} = \underset{\sim}{x}) - a] \\ &= \delta_\mu(\underset{\sim}{x})[E(\Lambda \mid \underset{\sim}{X} = \underset{\sim}{x}) - a], \end{aligned} \tag{4.8}$$

for all $\underset{\sim}{x} \varepsilon S$. Also, since $0 \leqq \delta_n(\underset{\sim}{x}) \leqq 1$ for all values of $\underset{\sim}{x}$ and n, we have

$$|[E\delta_n(\underset{\sim}{x})][E(\Lambda \mid \underset{\sim}{X} = \underset{\sim}{x}) - a]| \leqq |E(\Lambda \mid \underset{\sim}{X} = \underset{\sim}{x})| + |a|, \qquad \text{all } \underset{\sim}{x}, n. \tag{4.9}$$

Hence by (4.8), (4.9) and the Lebesgue Dominated Convergence Theorem we have

$$\lim_{n\to\infty} R(\delta_n) = R(\delta_\mu), \tag{4.10}$$

whenever the a priori probability measure is such that (4.5) holds and $E|\Lambda| < \infty$.

Example 2. (Two-sided alternatives): Suppose now that we wish to test the hypothesis $H_0^*: \lambda \varepsilon (a - b, a + b)$, $b > 0$, versus the alternative hypothesis $H_1^*: \lambda \not\varepsilon (a - b, a + b)$, where, as before, λ is a value of the random variable

Λ which is associated with the vector of observations $\underline{X}$. Let the loss function L^* be defined by

$$(4.11) \qquad L^*(A_i, \lambda) = \begin{cases} \max (0, [(\lambda - a)^2 - b^2]), & i = 0, \\ -\min (0, [(\lambda - a)^2 - b^2]), & i = 1, \end{cases}$$

where, as before, A_i represents the action of accepting the hypothesis H_i^*, $i = 0, 1$. The graph of L^* is shown in Fig. 1. For any decision function $\delta(\underline{x}) = \text{Prob} \{A_1 \mid \underline{X} = \underline{x}\}$, the risk is

$$\begin{aligned} R(\delta) &= EL^* (A_0, \Lambda) - E\{\delta(\underline{X})[L^* (A_0, \Lambda) - L^* (A_1, \Lambda)]\} \\ (4.12) \qquad &= EL^* (A_0, \Lambda) - E\{\delta(\underline{X})[(\Lambda - a)^2 - b^2]\} \\ &= EL^* (A_0, \Lambda) - E\{\delta(\underline{X})[E(\Lambda^2 \mid \underline{X}) - 2aE(\Lambda \mid \underline{X}) + a^2 - b^2]\}. \end{aligned}$$

Hence the Bayes decision function $\delta_\mu^*(\underline{x})$ minimizing $R(\delta)$ is given by

$$(4.13) \qquad \delta_\mu^*(\underline{x}) = \begin{cases} 1, & E(\Lambda^2 \mid \underline{X} = \underline{x}) - 2aE(\Lambda \mid \underline{X} = \underline{x}) > b^2 - a^2, \\ 0, & \text{otherwise.} \end{cases}$$

Now if the a priori probability measure is not known we may still be able, under certain circumstances, to find empirical Bayes estimators $\varphi_n^{(1)}(\underline{x})$ and $\varphi_n^{(2)}(\underline{x})$ based on prior independent observations, such that as $n \to \infty$,

$$(4.14) \qquad \varphi_n^{(1)}(\underline{x}) \to E(\Lambda \mid \underline{X} = \underline{x}), \qquad \text{in probability,}$$

and

$$(4.15) \qquad \varphi_n^{(2)}(\underline{x}) \to E(\Lambda^2 \mid \underline{X} = \underline{x}), \qquad \text{in probability,}$$

for all $\underline{x} \in S$ where S, as before, is assigned probability 1 under the distribution of $\underline{X}$. Then if we define the empirical Bayes decision function $\delta_n^*(\underline{x})$ by

$$(4.16) \qquad \delta_n^*(\underline{x}) = \begin{cases} 1, & \varphi_n^{(2)}(\underline{x}) - 2a\varphi_n^{(1)}(\underline{x}) > b^2 - a^2, \\ 0, & \text{otherwise,} \end{cases}$$

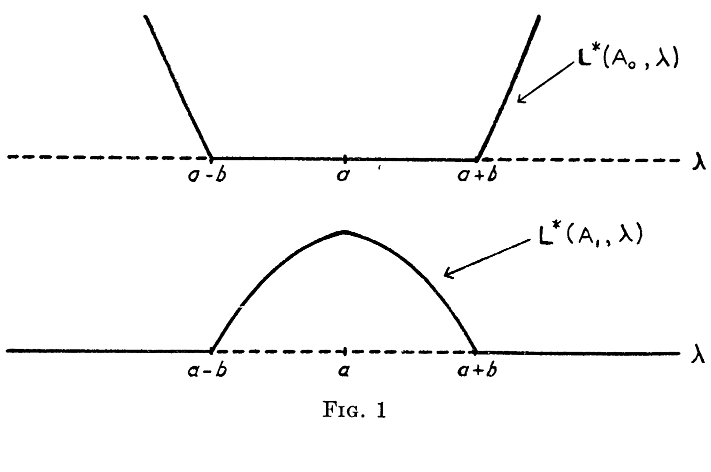

Fig. 1

we have (by the same argument as in Example 1),

$$\lim_{n\to\infty} R(\delta_n^*) = R(\delta_\mu^*), \tag{4.17}$$

for any a priori probability measure such that $E\Lambda^2 < \infty$ and (4.14) and (4.15) hold.

The existence of empirical Bayes estimators satisfying (4.14) follows directly (as in Example 1) from the results of the previous sections. We remark that if we assume that $Eh^4(X) < \infty$ in the cases treated in Secs. 2 and 3, then an empirical Bayes estimator satisfying (4.15) can be obtained whenever the number of components in the vector $\underline{X}_i$ exceeds the number in $\underline{X}$ by at least 2 for all i. To see this we observe that

$$E(h(X_{1,r+1})h(X_{1,r+2}) \mid \underline{X}_1^{(r)} = \underline{x}) = E(\Lambda^2 \mid \underline{X} = \underline{x}), \tag{4.18}$$

so that (in the notation of Sec. 2) if we let

$$\varphi_n^{(2)}(\underline{x}) = \begin{cases} \dfrac{1}{\bar{M}_n(\underline{x})} \sum_{i=1}^{n} M_i(\underline{x})h(X_{i,r+1})h(X_{i,r+2}), & \bar{M}_n(\underline{x}) > 0, \\ 0, \quad \text{otherwise}, \end{cases} \tag{4.19}$$

we can show (by arguments paralleling those of Sec. 2) that if $Eh^4(X) < \infty$ then

$$\lim_{n\to\infty} E[\varphi_n^{(2)}(\underline{x}) - E(\Lambda^2 \mid \underline{X} = \underline{x})]^2 = 0, \tag{4.20}$$

for all $\underline{x} \in S$, which implies (4.15).

5. General remarks. The methods of this paper clearly may be modified to apply to compound Bayes decision problems where the component problems are of one of the types considered above and where the compound risk is the average of the component risks. Robbins has conjectured in [3] that empirical Bayes solutions of such compound problems will often lead to asymptotically subminimax solutions for the corresponding compound decision problems where no a priori probability measure is assumed to exist. We may surmise therefore that suitable modifications of the techniques given here are applicable to such problems.

6. Acknowledgements. I wish to express my thanks to Professor Herbert Robbins for proposing this investigation and for his helpful suggestions during its progress. I also wish to thank my wife, Maryann Haertter Johns, for performing the computations necessary for Tables I and II.

REFERENCES

[1] H. Robbins, "An empirical Bayes approach to statistics," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1956, pp. 157–163.

[2] M. Loève, *Probability Theory*, D. Van Nostrand, 1955.

[3] H. Robbins, "Asymptotically subminimax solutions of compound statistical decision problems," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 131–148.