AD-A038 966 WISCONSIN UNIV MADISON MATHEMATICS RESEARCH CENTER F/G 12/1

AN EMPIRICAL BAYES APPROACH TO OUTLIERS.(U)

MAR 77 E DE ALBA, J VAN RYZIN DAAG29-75-C-0024

UNCLASSIFIED MRC-TSR-1734 NL

1 OF 1

AD A038966

END

DATE FILMED

5-77

MRC Technical Summary Report #1734

AN EMPIRICAL BAYES APPROACH TO OUTLIERS

Enrique de Alba and J. Van Ryzin

Mathematics Research Center
University of Wisconsin–Madison
610 Walnut Street
Madison, Wisconsin 53706

March 1977

(Received September 18, 1974)
(Revised February 25, 1977)

See Form 1473

Sponsored by

U.S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

and

National Institutes of Health
Bethesda, Maryland 20014

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

AN EMPIRICAL BAYES APPROACH TO OUTLIERS

Enrique de Alba and J. Van Ryzin[1]

Technical Summary Report # 1734
March 1977

ABSTRACT

A formulation of the problem of detecting outliers as an empirical Bayes problem is studied. In so doing what arises is a non-standard empirical Bayes problem for which the notion of average risk asymptotic optimality (a.r.a.o.) of procedures is defined. Some general theorems giving sufficient conditions for a.r.a.o. procedures are developed. These general results are then used in various formulations of the outlier problem for underlying normal distributions to give a.r.a.o. empirical Bayes procedures. Rates of convergence results are also given using the methods of Johns and Van Ryzin (1971, 1972).

AMS (MOS) Subject Classifications: Primary 62C99, 62F05.

Key Words: Empirical Bayes procedure
detection of outliers
average risk asymptotic optimality
rates of convergence

Work Unit Number 4 (Probability, Statistics, and Combinatorics)

---

[1] Research of this author was sponsored in part by DHEW, PHS, National Institutes of Health under Grant 5 R01 CA 18332-02 and by the United States Army under Contract No. DAAG29-75-C-0024.

AN EMPIRICAL BAYES APPROACH TO OUTLIERS

Enrique de Alba and J. Van Ryzin[1]

1. Introduction.

It is not uncommon to find samples in which some of the observations appear suspiciously far away from the main group. Such observations are often called outliers.

The problem appears in the literature as early as 1838, when Bessel mentioned the simple rule of not rejecting any observations. There is a large number of results on the topic, which consider the causes of outliers and present different solutions. There are several articles which include excellent historical reviews of the work done in this area. See for example de Alba (1974), Rider (1933), Ferguson (1961a, 1961b) and Guttman and Smith (1969).

The outlier problem actually presents two aspects: i) identify any particular observation (or observations) which come from a distribution other than the one which has been assumed to explain the main body of the observations: spurious observations, ii) obtain a procedure for the analysis of the data which is not very much affected (if at all) by the presence of spurious observations or by the rejection of non-spurious observations. This paper only considers the first aspect of the problem.

[1] Research of this author was sponsored in part by DHEW, PHS, National Institutes of Health under Grant 5 R01 CA 18332-02 and by the United States Army under Contract No. DAAG29-75-C-0024.

Three basic approaches have been used: the premium-protection, the significance test and the decision theoretic approaches (see de Alba (1974) and Guttman and Smith (1969)). We shall use the empirical Bayes (e.B.) approach of Robbins (1964). Our solution can be classified under the decision theoretic approach.

We shall use the following model. Out of a sample of size n, $(n-k)$ of the random variables are normal with mean μ and variance σ^2, while the remaining m r.v.'s, also normal, have the same mean μ but their variance is $\lambda_i\sigma^2$, $(\lambda_i > 1)$, where i ranges over the subscripts of the m r.v.'s. If we let the symbol "$\sim$" mean "is distributed as", then

$$\begin{aligned} &X_{i_1}, X_{i_2}, \ldots, X_{i_{(n-k)}} \sim N(\mu, \sigma^2) \\ &X_{i_j} \sim N(\mu, \lambda_j\sigma^2), \qquad j = n-k+1, \ldots, n, \end{aligned} \tag{1.1}$$

where $(i_1, i_2, \ldots, i_n)$ is some permutation of the subscripts $(1, 2, \ldots, n)$ and λ_j is the multiple of the variance of the non-spurious observations and $\lambda_j\sigma^2$ is the variance of spurious observation i_j.

2. Empirical Bayes Terminology.

We begin by stating the basic elements of decision theory which we will be using:

i) The parameter space Θ.

ii) A set A of possible actions.

iii) A loss function $L(a,\lambda) \geq 0$, defined on $A \times \Theta$. For any point $(a,\lambda) \varepsilon A \times \Theta$, $L(a,\lambda) < \infty$ is the loss that results when $\theta \varepsilon \Theta$ is the true value of the parameter and we take the action $a \varepsilon A$.

iv) The sample space $\mathcal{X}$ which is taken to be a finite dimensional Euclidean space. For each $\lambda \varepsilon \Theta$ there is defined a c.d.f. $F_\lambda(x)$.

We shall be working with behavioral decision rules. A behavioral decision rule, $t(x)$, is a function which gives, for each x in the sample space, a probability distribution over A. We thus have the average loss when using $t(x)$ given by

$$L(t(x),\lambda) = EL(Z,\lambda)$$

where the expectation on Z is taken with respect to $t(x)$. Further, the risk function is defined as

$$r(t,\lambda) = E_\lambda L(t(X),\lambda)$$

where E_0 denotes the expectation of X when the value of the parameter is λ. The minimum Bayes risk with respect to prior distribution G on λ, or Bayes envelope functional, is:

$$r(G) = r^*(t_G, G) = \inf_t r^*(t,G) = \inf_t \int r(t,\lambda)dG(\lambda) ,$$

where t_G is a Bayes rule.

Suppose we are confronted with the same decision problem repeatedly and independently. Let $(\Lambda_1,X_1),(\Lambda_2,X_2),\ldots,(\Lambda_n,X_n)$ be a sequence of mutually independent pairs of random variables, where

$\Lambda_1, \Lambda_2, \ldots, \Lambda_n$ all have an unknown common (prior) distribution G defined on Θ and the conditional distribution of X given $\Lambda_r = \lambda$ is specified by the p.d.f. $f_\lambda(x) = (dF_\lambda/d\tau)(x)$, $r = 1, \ldots, n$ assuming that the family $\{F_\lambda : \lambda \in \Theta\}$ is dominated by a σ-finite measure τ on $\mathcal{X}$.

The empirical Bayes approach constructs a decision procedure concerning Λ_{n+1} (unobservable), based on the values that have been observed for $X_1, \ldots, X_{n+1}$, i.e., using $x_1, x_2, \ldots, x_{n+1}$. The $(\Lambda_1, \ldots, \Lambda_n)$ remain unobservable throughout. For a decision about Λ_{n+1}, a function $t_n(x_{n+1})$, whose form depends on the values $x_1, \ldots, x_n$ is used. An e.B. decision procedure is a sequence $T = \{t_n\}$ of such functions.

Robbins (1963, 1964) defined a sequence $T = \{t_n\}$ as being asymptotically optimal (a.o.) relative to G if

$$\lim_{n \to \infty} r^*(t_n, G) = r(G) .$$

He also gives conditions under which a rule is a.o.

A question that arises when dealing with empirical Bayes decision rules is "how fast, relative to n, does $r^*(t_n, G)$ converge to the minimum risk ?". Johns and Van Ryzin (1971, 1972) have studied the problem and given rates in different situations. We shall also consider the rate problem in relation to our rules for testing outliers.

3. An Extension of Empirical Bayes Procedures.

The results presented for testing outliers are based on a non-standard empirical Bayes framework which can be described as follows. Let $(X_1, \Lambda_1), (X_2, \Lambda_2), \ldots, (X_n, \Lambda_n)$ be mutually independent pairs of random variables, $X_r (r = 1, \ldots, n)$ is defined on a sample space $\mathcal{X}$ and $\Lambda_r (r = 1, \ldots, n)$ on a parameter space Θ. The Λ_r, $r = 1, \ldots, n$, are assumed to have a common prior distribution G on Θ and the conditional density of X_r given that $\Lambda_r = \lambda$ is $f_\lambda(x_r)$, $r = 1, \ldots, n$, w.r.t. a σ-finite measure τ.

We now define the empirical Bayes rule for the r^{th} problem, $r = 1, \ldots, n$, denoted $t_n^{(r)}(X_r)$. Let $t_n(x) = t_n(X_1, \ldots, X_n; x)$. Define

$$\underline{X}_r(x) = (X_1, \ldots, X_{r-1}, x, X_{r+1}, \ldots, X_n)$$

and

$$t_n^{(r)}(x) = t_n(X_1, \ldots, X_{r-1}, x, X_{r+1}, \ldots, X_n; x) = t_n(\underline{X}_r(x); x)$$

so that the form of the r^{th} decision rule depends on $\underline{X}_r(x)$. Note $t_n^{(r)}(X_r) = t_n(X_r)$ since $t_n^{(r)}(x)$ differs from $t_n(x)$ in that the first has a fixed value of x in the place of the r^{th} random variable.

This particular use of the e. B. method is non-standard. We have two differences:

i) We are not actually working with a sequential decision procedure, and

ii) We do not use a decision rule whose form depends only on the first $(n-1)$ observations.

If $t_n^{(r)}(x_r)$ is obtained for each $r = 1, \ldots, n$ and we denote its risk relative to G by $r^*(t_n^{(r)}, G)$, we obtain an n-vector, $\underline{t}_n = \{t_n^{(r)}(x_r); r = 1, \ldots, n\}$ (or $\underline{t}_n = \{t_n^{(r)}(x_r)\}$) of decision functions which we shall call an e. B. decision procedure. In our application such a procedure will give us a rule to determine whether each X_r, $r = 1, \ldots, n$, is spurious or not. We make the following definition.

DEFINITION 1. Let $(X_1, \Lambda_1), \ldots, (X_n, \Lambda_n)$ be mutually independent pairs of random variables and $t_n^{(r)}(x) = t_n(\underline{X}_r(x); x)$, $r = 1, \ldots, n$. If

$$r^*(\underline{t}_n, G) = (1/n) \sum_{r=1}^{n} r^*(t_n^{(r)}, G) \to r(G) \quad \text{as} \quad n \to \infty \tag{3.1}$$

then the e. B. procedure $\underline{t}_n = \{t_n^{(r)}(X_r)\}$ is said to be "average risk asymptotically optimal" (a. r. a. o.) relative to G.

The symbol $-P\to$ will be used to denote convergence in probability. We now state and prove the following lemma in the case where $\Theta = \{\theta_0, \theta_1\}$ and $A = \{a_0, a_1\}$.

A Bayes rule against G in this case can be written as (see Robbins (1964)),

$$t_G(x) = \begin{cases} 1 \text{ if } & \Delta_G(x) \geq 0 \\ 0 \text{ if } & \Delta_G(x) < 0 \end{cases} \tag{3.2}$$

where $t(x) = \Pr\{\text{taking action } a_1 | X = x\}$ and

$$\Delta_G(x) = \int \{L(a_0, \lambda) - L(a_1, \lambda)\} f_\lambda(x) dG(x) . \tag{3.3}$$

LEMMA 1. Let $\Delta_n(x) = \Delta_n(\underline{X}; x)$ be such that as $n \to \infty$, $\Delta_n(x) - P \to \Delta_G(x)$ a.e. (τ). Define

$$t_n^{(r)}(x) = \begin{cases} 0 & \text{if } \Delta_n^{(r)}(x) \geq 0 \\ 1 & \text{if } \Delta_n^{(r)}(x) < 0 , \end{cases} \tag{3.4}$$

where

$$\Delta_n^{(r)}(x) = \Delta_n(\underline{X}_r(x);x) .$$

Let

$$r^*(\underline{t}_n, G) = (1/n) \sum_{r=1}^{n} r^*(t_n^{(r)}, G) .$$

Then, for any $0 < d_1 \leq 1$, $0 < d_2 \leq 1$, and each fixed r,

$$r^*(t_n^{(r)}, G) - r(G) \leq 2^{d_1} \int |\Delta_G(x)|^{1-d_1} E|\Delta_n^{(r)}(x) - \Delta_n(x)|^{d_1} d\tau(x) +$$

$$2^{d_2} \int |\Delta_G(x)|^{1-d_2} E|\Delta_n(x) - \Delta_G(x)|^{d_2} d\tau(x), \quad r = 1, \ldots, n .$$

E denotes the expectation under $X_1, \ldots, X_n$.

PROOF. From the definition of $t_n^{(r)}(x)$ and $r^*(t_n^{(r)}, G)$ we have

$$r^*(t_n^{(r)}, G) - r(G) = E \int \Delta_G(x)[t_n^{(r)}(x) - t_G(x)]d\tau(x) \tag{3.5}$$

with $t_G(x)$ and $\Delta_G(x)$ as in (3.2) and (3.3). But from (3.2) and (3.4), we see that

$$|t_n^{(r)}(x) - t_G(x)| \leq \begin{cases} 1 & \text{if } |\Delta_n^{(r)}(x) - \Delta_G(x)| \geq |\Delta_G(x)| \\ 0 & \text{otherwise.} \end{cases}$$

and hence

$$E|t_n^{(r)}(x) - t_G(x)| \leq \Pr\{|\Delta_n^{(r)}(x) - \Delta_G(x)| \geq |\Delta_G(x)|\} . \tag{3.6}$$

This result and using Fubini's theorem in (3.5) gives

$$r^*(t_n^{(r)}, G) - r(G) \leq \int |\Delta_G(x)| \Pr\{|\Delta_n^{(r)}(x) - \Delta_n(x)| \geq |\Delta_G(x)|/2\} d\tau(x)$$

$$+ \int |\Delta_G(x)| \Pr\{|\Delta_n(x) - \Delta_G(x)| \geq |\Delta_G(x)|/2\} d\tau(x) .$$

Markov's inequality applied to the first term with $0 < d_1 \leq 1$ and to the second with $0 < d_2 \leq 1$ yields the required result.

Q. E. D.

The following result is an extension of Corollary 1.2 of Robbins (1964), to a.r.a.o. decision rules.

THEOREM 1. Let $(X_1, \Lambda_1), \ldots, (X_n, \Lambda_n)$ be mutually independent pairs of random variables. Let $A = \{a_0, a_1\}$, let G be such that

$$\int L(a_i, \lambda) dG(\lambda) < \infty, \quad i = 0,1 . \tag{3.7}$$

Let

$$t_n^{(r)}(x) = \begin{cases} 0 & \text{if } \Delta_n^{(r)}(x) \geq 0 \\ 1 & \text{if } \Delta_n^{(r)}(x) < 0 , \end{cases}$$

where

$$\Delta_n^{(r)}(x) = \Delta_n(\underline{X}_\tau(x); x) .$$

Assume that

$$\Delta_n(x) = Q(\phi_1(\underline{X}), \ldots, \phi_m(\underline{X}); x), \quad m \geq 1$$

where the following conditions are true:

a) $Q(y_1, \ldots, y_m; x)$ is continuous in every y_j, $j = 1, \ldots, m$, a.e. (τ).

b) $\phi_j(\underline{X}_1(x)) - \phi_j(\underline{X}) = \phi_j(x, X_2, \ldots, X_n) - \phi_j(X_1, \ldots, X_n) \xrightarrow{P} 0$ as $n \to \infty$, $j = 1, \ldots, m$, a.e. (τ).

c) $\phi_j(X_1, \ldots, X_n) = \phi_j(X_{\nu_1}, \ldots, X_{\nu_n})$, $j = 1, \ldots, m$, where $(\nu_1, \ldots, \nu_n)$ is any permutation of the subscripts $(1, \ldots, n)$, i.e. the ϕ_j's are symmetric in $(X_1, \ldots, X_n)$.

d) $\Delta_n(x) \xrightarrow{P} \Delta_G(x)$ as $n \to \infty$, a.e. (τ).

Then

$$\underline{t}_n = \{t_n^{(r)}(X_r)\} \quad \text{is a.r.a.o.}$$

PROOF. First note that for $r = 1$

$$\Delta_n^{(1)}(x) = Q(\phi_1(\underline{X}_1(x)), \ldots, \phi_m(\underline{X}_1(x));x) =$$
$$\{Q(\phi_1(\underline{X}_1(x)), \ldots, \phi_m(\underline{X}_1(x));x) - Q(\phi_1(\underline{X}), \ldots, \phi_m(\underline{X});x)\} + \Delta_n(x) .$$

Now conditions a) and b) imply that the term in brackets converges to zero in probability. Thus, with condition d), as $n \to \infty$, a.e. (τ)

$$\Delta_n^{(1)}(x) \xrightarrow{P} \Delta_G(x), \quad x \text{ fixed.}$$

From Equation (3.6), we have

$$r^*(t_n^{(r)}, G) - r(G) \le \int |\Delta_G(x)| \Pr\{|\Delta_n^{(r)}(x) - \Delta_G(x)| \ge |\Delta_G(x)|\} d\tau(x) ,$$

so that

$$\sup_{1 \le r \le n} \{r^*(t_n^{(r)}, G) - r(G)\} \le \int |\Delta_G(x)| \sup_{1 \le r \le n} \{\Pr\{|\Delta_n^{(r)}(x) - \Delta_G(x)| \ge |\Delta_G(x)|\}\} d\tau(x).$$

From the symmetry of the ϕ_j's (condition (c)) we get

$$\Pr\{|\Delta_n^{(r)}(x) - \Delta_G(x)| \ge |\Delta_G(x)|\} = \Pr\{|\Delta_n^{(1)}(x) - \Delta_G(x)| \ge |\Delta_G(x)|\} ,$$

for $r = 1, \ldots, n$. Hence

$$\sup_{1 \le r \le n} \{r^*(t_n^{(r)}, G) - r(G)\} \le \int |\Delta_G(x)| \Pr\{|\Delta_n^{(1)}(x) - \Delta_G(x)| \ge |\Delta_G(x)|\} d\tau(x) .$$

Now

$$|\Delta_G(x)| \Pr\{|\Delta_n^{(1)}(x) - \Delta_G(x)| \ge |\Delta_G(x)|\} \le |\Delta_G(x)| ,$$

and (3.7) implies $\Delta_G(x)$ is integrable, since as $n \to \infty$, $\Delta_n^{(1)}(x) \xrightarrow{P} \Delta_G(x)$, implies

$$|\Delta_G(x)|\Pr\{|\Delta_n^{(1)}(x) - \Delta_G(x)| \geq |\Delta_G(x)|\} \to 0 \quad \text{as} \quad n \to \infty, \quad \text{a.e. } (\tau),$$

we can apply the Dominated Convergence Theorem to get

$$\sup_{1 \leq r \leq n} \{r^*(t_n^{(r)}, G) - r(G)\} \to 0, \quad \text{as} \quad n \to \infty .$$

Since

$$(1/n) \sum_{r=1}^{n} \{r^*(t_n^{(r)}, G) - r(G)\} \leq \sup_{1 \leq r \leq n} \{r^*(t_n^{(r)}, G) - r(G)\} ,$$

we have

$$r^*(t_n, G) - r(G) \to 0, \quad \text{as} \quad n \to \infty, \quad \text{so that } \underline{t}_n \text{ is a.r.a.o.}$$

Q.E.D.

It is convenient to note at this point that we only have to verify the conditions of Theorem 1 in order to prove a.r.a.o. The following lemma will prove useful for deriving results on rates of convergence.

LEMMA 2. If $\Delta_n(x) = \Delta_n(\underline{X};x)$ is symmetric in $(X_1, \ldots, X_n)$ and

$$r^*(t_n^{(1)}, G) - r(G) = O(n^{-s}), \quad \text{for some} \quad s > 0 ,$$

then

$$r^*(\underline{t}_n, G) - r(G) = O(n^{-s}) .$$

The proof follows from (3.6) and the symmetry of $\Delta_n(x)$.

4. An e.B. Test for Outliers.

Consider n independent random variables $X_1, \ldots, X_n$, where given $\Lambda_r = \lambda$, X_r is normally distributed with mean μ and variance $\sigma^2\lambda$, $r = 1, \ldots, n$, i.e.

$$X_r \sim N(\mu, \sigma^2\lambda), \quad r = 1, \ldots, n . \tag{4.1}$$

μ and σ^2 are assumed known and Λ_r has a prior distribution $G = \{p, 1-p\}$ on $\Theta = \{1, \lambda_0\}$, defined as follows

$$\Pr\{\Lambda_r = 1\} = G(1^+) - G(1) = p \ ,$$

$$\Pr\{\Lambda_r = \lambda_0\} = G(\lambda_0^+) - G(\lambda_0) = 1 - p, \quad 0 < p < 1$$

for some $\lambda_0 > 1$, known, and $r = 1, \ldots, n$. We can use empirical Bayes methods to test the null hypothesis, for each r, $r = 1, \ldots, n$,

$$H_0 : X_r \sim N(\mu, \sigma^2), \text{ vs.}$$

$$H_1 : X_r \sim N(\mu, \lambda_0 \sigma^2) \ .$$

The procedure is given in the following theorem.

THEOREM 2. Let X_r, $r = 1, \ldots, n$ be defined as above, $A = \{a_0, a_1\}$ where the action a_i is defined as a_i = "decide in favor of H_i", $i = 0, 1$ and the loss function, be such that

$$L(a_0, 1) = L(a_1, \lambda_0) = 0 \ .$$

Furthermore, define $\Delta_n(x)$ as

$$\Delta_n(x) = (1 - \hat{p})\, L(a_0, \lambda_0) f_{\lambda_0}(x) - \hat{p} L(a_1, 1) f_1(x)$$

where $f_\lambda(x)$ is the p.d.f. of x obtained when $\Lambda_r = \lambda, \lambda = 1$ or λ_0. If $\hat{p} = \hat{p}(\underline{X})$ is a consistent estimator for p, symmetric in $\underline{X}$, and a.e. (τ)

$$\hat{p}(\underline{X}_1(x)) - \hat{p}(\underline{X}) \xrightarrow{P} 0, \quad \text{as } n \to \infty \ ,$$

then the e.B. decision rule

$$t_n(X_r) = \begin{cases} 0 & \text{if } \Delta_n(X_r) \geq 0 \\ 1 & \text{otherwise,} \end{cases} \tag{4.2}$$

$r = 1, \ldots, n$, is a.r.a.o. for testing H_0 vs. H_1. ∎

Under H_0, X_r has the p.d.f.

$$f_1(x_r) = (1/\sigma\sqrt{2\pi})\exp\{-(x_r - \mu)^2/2\sigma^2\} \tag{4.3}$$

and under H_1,

$$f_{\lambda_0}(x_r) = (1/\sigma\sqrt{2\pi\lambda_0})\exp\{-(x_r - \mu)^2/2\lambda_0\sigma^2\}, \quad r = 1, \ldots, n . \tag{4.4}$$

The proof of the theorem is a direct consequence of Theorem 1 for $m = 1$. One particular case of a consistent (symmetric) estimator $\hat{p}$ is given by

$$\hat{p} = (1/n) \sum_{r=1}^{n} \xi_0(X_r) , \tag{4.5}$$

where

$$\xi_0(X_r) = \frac{\sigma^2\lambda_0 - (X_r - \mu)^2}{\sigma^2(\lambda_0 - 1)} .$$

A question that arises when estimating G is that of identifiability. Maritz (1970) gives results on identifiability for location parameters under normality. A result for scale parameters can be found in de Alba (1974).

As pointed out earlier, the usefulness of e.B. procedures in statistical applications depends on how fast the Bayes risk of each successive decision problem approaches the minimum Bayes risk. In relation to Theorem 2 we have the following result.

THEOREM 3. Assume the conditions of Theorem 2 hold. Let

$\hat{p} = (1/n) \sum_{r=1}^{n} \xi_0(X_r)$, where $\xi_0(X_r) = \{\sigma^2\lambda_0 - (X_r - \mu)^2\}/\sigma^2(\lambda_0 - 1)$, then

$$0 \leq r^*(\underline{t}_n, G) - r(G) = O(n^{-1/2}) .$$

The proof is straightforward and follows from applying Lemma 1, with $d_1 = 1$ and $d_2 = 1$, and from Lemma 2, (see de Alba (1974) for details).

In this section we have introduced our approach to the outlier problem, applying it to a particular case. In the following sections we will give some extensions.

5. Small Versus Large Outliers.

Suppose we are interested in a test for "small outliers" against "large outliers", i.e.

$$\begin{aligned} H_0&: \lambda \leq \lambda_0 \quad \text{for some } \lambda_0 > 1, \quad \text{vs.} \\ H_1&: \lambda > \lambda_0 \end{aligned} \tag{5.1}$$

where $\lambda \geq 1$. The "largeness" criterion is determined by the value of λ_0. A reasonable loss function for this test is given by

$$L(a_0, \lambda) = \begin{cases} 0 & \lambda \leq \lambda_0 \\ 1/\lambda_0 - 1/\lambda, & \lambda > \lambda_0 \end{cases} \tag{5.2}$$

$$L(a_1, \lambda) = \begin{cases} 0 & \lambda > \lambda_0 \\ 1/\lambda - 1/\lambda_0, & \lambda \leq \lambda_0 . \end{cases} \tag{5.3}$$

Define now

$$b(\lambda) = L(a_0, \lambda) - L(a_1, \lambda) ,$$

so that

$$\Delta_G(x) = \int \{L(a_0, \lambda) - L(a_1, \lambda)\} f_\lambda(x) dG(\lambda) = (1/\lambda_0) f_G(x) + \{\sigma^2/(x-\mu)\} f'_G(x) ,$$

with $f_G(x) = \int f_\lambda(x) dG(\lambda)$ and $f'_G(x)$ is obtained from the definition of $f_G(x)$, by differentiating under the integral sign. The Bayes rule is:

$$t_G(x) = \begin{cases} 0 & \text{if } \Delta_G(x) = (1/\lambda_0)f_G(x) + \{\sigma^2/(x-\mu)\}f'_G(x) \geq 0 \\ 1 & \text{otherwise} . \end{cases}$$

The e.B. rule may be derived by getting consistent estimators for $f_G(x)$ and $f'_G(x)$. We have the following theorem.

THEOREM 4. Let $X_1, \ldots, X_n$ be independent random variables defined as in (4.1), the parameter space $\Theta = [1, \infty)$ and let G be any c.d.f. defined on Θ. The action space is $A = \{a_0, a_1\}$, H_0 and H_1 are given by (5.1) and the loss function by (5.2)-(5.3). Let $f_n(x) = f_n(\underline{X};x)$ and $f'_n(x) = f'_n(\underline{X};x)$ be any consistent estimators of $f_G(x)$ and $f'_G(x)$ respectively, for all x, symmetric in $\underline{X}$, and such that

$$f_n(\underline{X}_1(x);x) - f_n(\underline{X};x) \xrightarrow{P} 0, \quad \text{for all } x \tag{5.5}$$

$$f'_n(\underline{X}_1(x);x) - f'_n(\underline{X};x) \xrightarrow{P} 0, \quad \text{for all } x \tag{5.6}$$

as $n \to \infty$, and

$$\Delta_n(x) = (1/\lambda_0)f_n(x) + \{\sigma^2/(x-\mu)\}f'_n(x) .$$

Then

$$t_n(X_r) = \begin{cases} 0 & \text{if } \Delta_n(X_r) \geq 0 \\ 1 & \text{otherwise,} \end{cases} \tag{5.7}$$

$r = 1, \ldots, n,$ is a.r.a.o. relative to G.

The proof follows from Theorem 1 with $m = 2$.

A particular choice of $f_n(x)$ and $f'_n(x)$ is given by Johns and Van Ryzin (1972). With this particular choice, conditions (5.5) and (5.6) are satisfied (see de Alba (1974).) Theorem 3 of Johns and Van Ryzin (1972) will be very useful for proving our rate theorems.

We now define $\kappa_1 = \kappa_1(u_1)[\kappa_2 = \kappa_2(u_2)]$ as the class of all real-valued measurable functions on the real line satisfying the conditions of the definition of $f_n(x)[f'_n(x)]$ given in Johns and Van Ryzin (1972).

Also note that since we are assuming μ and σ^2 are known, in (4.1) we can define $y = (x - \mu)^2/2\sigma^2$ and restate our problem for the density

$$f_\lambda(y) = \begin{cases} (1/\sqrt{\pi y \lambda})\exp\{-y/\lambda\}, & y > 0, \ \lambda \geqq 1 \\ 0 \ \text{elsewhere} . \end{cases} \tag{5.8}$$

We can now prove the following rate theorem.

THEOREM 5. Let $Y_i = (X_i - \mu)^2/2\sigma^2$, $i = 1, \ldots, n$, where the X_i are independent random variables defined as in (4.1). Let $f_n(y)$ and $f'_n(y)$ be estimates of $f(y) = \int f_\lambda(y)dG(\lambda)$ and its derivative given as in [7] with $t_n(y)$ defined by (5.7) for $\Delta_n(y) = (\lambda_0^{-1} + (2y)^{-1})f_n(y) + f'_n(y)$. For any $\ell \geq 2$, if we choose $h_n = O(n^{-1/(2\ell+1)})$ and $K_i \in \kappa_i$ in defining $f_n(y)$ and $f'_n(y)$ such that

$$\int u^{j+i-1}K_i(u)du = 0 \quad \text{for} \quad j = 1, \ldots, \ell-1, \ i = 1, 2 , \tag{5.9}$$

and if for some d, $0 < d < 1/(2\ell + 3)$,

$$E(\Lambda^{(1+t)d/(2-d)}) < \infty, \quad \text{for some} \quad t > 0 , \tag{5.10}$$

then

$$r(\underline{t}_n, G) - r(G) = O(n^{-d(\ell-1)/(2\ell+1)}) .$$

PROOF. If in (5.8) we let $\theta = 1/\lambda$, then

$$f_\theta(y) = \begin{cases} (\sqrt{\theta}/\sqrt{\pi y})\exp\{-y\theta\} & \text{for } \; y > 0, \; 0 < \theta \le 1 \\ 0 & \text{otherwise} . \end{cases}$$

Furthermore, if

$$\beta(\theta) = \sqrt{\theta} \quad \text{and} \quad h(y) = 1/\sqrt{\pi y} \tag{5.11}$$

then $f_\theta(y)$ falls into the theory for Case I of Johns and Van Ryzin (1972). Now the loss function (5.2)-(5.3) may be written in terms of θ as

$$L(a_0, \theta) = \begin{cases} 0 & \theta \ge \theta_0 \\ \theta_0 - \theta & \theta < \theta_0 \end{cases}$$

and

$$L(a_1, \theta) = \begin{cases} 0 & \theta < \theta_0 \\ \theta - \theta_0 & \theta \ge \theta_0 . \end{cases}$$

The hypothesis to be tested will be $H_0^*: \theta \ge \theta_0$ vs. $H_1^*: \theta < \theta_0$.

From Lemma 1 with $d_1 = d_2 = d$ we see that

$$\begin{aligned} r^*(t_n^{(1)}, G) - r(G) \le 2^d \int |\Delta_G(y)|^{1-d} E|\Delta_n^{(1)}(y) - \Delta_n(y)|^d dy \, + \\ 2^d \int |\Delta_G(y)|^{1-d} E|\Delta_n(y) - \Delta_G(y)|^d dy , \end{aligned} \tag{5.12}$$

where $\Delta_G(y) = (\lambda_0^{-1} + (2y)^{-1})f(y) + f'(y)$.

Consider the second term on the right hand side of (5.12). We now verify the conditions of Theorem 3 in [7] to obtain the rate of convergence for this term.

Note that from (5.11), we have

$$h^{(\ell)}(y) = (1/\sqrt{\pi})(-1/2)^\ell(1 \cdot (1+2) \cdot (1+4) \ldots \cdot (1+2\ell))y^{-(2\ell+1)/2} . \tag{5.13}$$

Hence $h^{(\ell)}(y)$ exists and is continuous for any $\ell \ge 1$. From this and

Lemma 2 in Johns and Van Ryzin (1972) we know that $f^{(\ell)}(y)$ exists and is continuous for any $\ell \geq 1$ and $y > 0$. Furthermore, since θ is defined on $(0,1]$, $E\theta < \infty$ is always true as required by their Theorem 3.

We now verify (3.3) of their Theorem 3. Note that

$$f_\theta^{(\ell)}(y) = (-1)^\ell f_\theta(y) \sum_{j=1}^{\ell} a_j(\theta + 1/2y)^j y^{j-\ell} \tag{5.14}$$

where the a_j's are non-negative constants. Throughout, when writing the expression for $f_\theta^{(r)}(y)$, the summation will be assumed to be from $j = 1$ to $j = \ell$ unless stated otherwise. $f_\theta(y)$ is a decreasing function of y, for $y > 0$. Hence

$$f_\varepsilon^*(y) = \sup_{0 \leq t \leq \varepsilon} f_G(y + t) = f_G(y) \tag{5.15}$$

for all $y > 0$ and $\varepsilon > 0$. From this we get, using $0 < \theta \leq 1$,

$$q_\varepsilon^{(\ell)}(y) = \sup_{0 \leq t \leq \varepsilon} |f_G^{(\ell)}(y + t)| = \sup_{0 \leq t \leq \varepsilon} \int |f_\theta^{(\ell)}(y + t)|dG(\theta)$$

$$= \int |f_\theta^{(\ell)}(y)|dG(\theta) = |(-1)|^\ell \int f_\theta(y) \sum a_j(\theta + 1/2y)^j y^{j-\ell} dG(\theta) = |f_G^{(\ell)}(y)|$$

$$\leq \int f_\theta(y) \sum a_j(1 + 1/2y)^j y^{j-\ell} dG(\theta) = f_G(y) \sum a_j(1 + 1/2y)^j y^{j-\ell}, \tag{5.16}$$

where $f_G^{(\ell)}(y)$ is obtained by repeated differentiation under the integral sign. Also

$$|\Delta_G(y)| = |\theta_0 f_G(y) - \int \theta f_\theta(y)dG(\theta)| \leq (\theta_0 + 1)f_G(y) . \tag{5.17}$$

By (5.16) and (5.17)

$$\int |\Delta_G(y)|^{1-d}\{q_\varepsilon^{(\ell)}(y)\}^d dy \leq (1 + \theta_0)^{1-d} E(\sum a_j^*(2Y + 1)^j Y^{-\ell})^d \tag{5.18}$$

where $a_j^* = a_j/2^j$ and the expectation is taken with respect to $f_G(y)$. Repeated application of the c_r-inequality yields

$$(1+\theta_0)^{1-d} E\{\sum a_j^*(2Y+1)^j Y^{-\ell}\}^d \leq (1+\theta_0)^{1-d} \sum b_j E|(2Y+1)^j Y^{-\ell}|^d \quad (5.19)$$

where $b_j = a_j^* c_d^j$, $j = 1, \ldots, \ell - 1$ and $b_\ell = a_\ell^* c_d^{\ell-1}$. Now by the Hölder inequality

$$E|(2Y+1)^j Y^{-\ell}|^d \leq \{E(2Y+1)^j\}^d \{EY^{-\ell d/(1-d)}\}^{1-d} .$$

The first factor on the right is always finite so we need only verify that the second factor is also finite. By Fubini's theorem

$$E\{Y^{-\ell d/(1-d)}\} = \int_0^1 \{\theta^{\ell d/(1-d)}/\Gamma(1/2)\}\{\int_0^\infty (\theta/y)^{1/2-\ell d/(1-d)} \exp\{-\theta y\}dy\}dG(\theta) .$$

Hence, provided

$$1/2 - \ell d/(1-d) > 0 , \quad (5.20)$$

the expectation will be finite. But this requires $d < 1/(2\ell + 1)$ and we have as a condition in the theorem that $d < 1/(2\ell + 3)$. Hence condition (3.3) of Theorem 3 in Johns and Van Ryzin (1972) holds.

We next verify (3.4) of their Theorem 3. From (5.16), (5.17) and $v(y) = h'(y)/h(y) = -1/2y$ we get

$$\int |\Delta_G(y)|^{1-d}\{|v(y)|q_\varepsilon^{(\ell)}(y)\}^d dy \leq (1+\theta_0)^{1-d} E\{\sum a_j^*(1+2Y)^j Y^{-s}\}^d ,$$

where $s = \ell + 1$. Hence we can apply the same argument from (5.19) to (5.20) with s instead of ℓ and the condition will be true provided

$$1/2 - sd/(1-d) = 1/2 - (\ell + 1)d/(1-d) > 0$$

which requires $d < 1/(2\ell + 3)$. Hence, condition (3.4) in Johns and Van Ryzin (1972) is satisfied.

To verify condition (3.2) of their theorem we use $v(y) = -1/2y$, (5.15) and (5.17) to obtain

$$\int |\Delta_G(y)|^{1-d} |v(y)|^d \{f_\varepsilon^*(y)\}^{d/2} dy \leq$$

$$2^{-d}(1+\theta_0)^{1-d}\{\int_0^1 y^{-d}\{f_G(y)\}^{1-d/2}dy + \int_1^\infty y^{-d}\{f_G(y)\}^{1-d/2}dy\} =$$

$$2^{-d}(1+\theta_0)^{1-d}[A_I + A_{II}],$$

where A_I and A_{II} are defined in an obvious manner. Using the fact that $y \geqq 1$ and applying the Hölder inequality we get, for $t > 0$,

$$A_{II} \leq \int_1^\infty \{f_G(y)\}^{1-d/2}dy \leq \{\int_1^\infty y^{-(1+t)}dy\}^{d/2}\{\int_0^\infty y^{\eta_1} f_G(y)dy\}^{1-d/2} \tag{5.21}$$

with

$$\eta_1 = (1+t)d/(2-d).$$

Clearly the first factor on the right is finite. Finiteness of the second factor follows from the definition of $f_G(y)$, Fubini's Theorem, the use of Laplace transforms and application of condition (5.10).

A_I can be shown to be finite by noting that

$$\int_0^1 y^{-d} f_G(y)^{1-d/2} \leqq \pi^{-1/2}\int_0^1 y^{(2-d)/4}dy.$$

Hence condition (3.2) in Johns and Van Ryzin (1972) holds.

Condition (3.1) of Johns and Van Ryzin (1972) can be shown to hold by arguments which are essentially the same as those used so far.

This completes the proof that under certain conditions and if the consistent estimators of $f_G(y)$ and $f'_G(y)$ are used as given in

Johns and Van Ryzin (1972), then the rate of convergence of the second term in (5.12) is $O(n^{-(\ell-1)d/(2\ell+1)})$. We shall now give the rate of convergence of the first term in (5.12).

From the definition of $\Delta_n^{(r)}(y)$, $f_n(y)$ and $f_n'(y)$ we have

$$\Delta_n^{(1)}(y) = (\theta_0 + 1/2y)f_n^{(1)}(y) + f_n^{(1)'}(y)$$

where

$$f_n^{(1)}(y) = (1/nh_n) \sum_{j=2}^{n} (1/2)\{K_1\{(Y_j - y)/h_n\} + K_1\{(y - Y_j)/h_n\}\} ,$$

$$f_n^{(1)'}(y) = (1/nh_n) \sum_{j=2}^{n} \{(1/2h_n)K_2\{(Y_j - y)/2h_n\} - K_2\{(y - Y_j)/h_n\}\} ,$$

since $K_i(0) = 0$, $i = 1, 2$.

Using the c_r-inequality, $K_i^* = \sup_u |K_i(u)| < \infty$, $i = 1, 2$ and (5.17) we get

$$\begin{aligned}\int |\Delta_G(y)|^{1-d} E|\Delta_n^{(1)}(y) - \Delta_n(y)|^d dy &\leq \\ 2(K_1^*/2nh_n)^d(1 + \theta_0)^{1-d} &\int (\theta_0 + 1/2y)^d\{f_G(y)\}^{1-d}dy \\ + (K_2^*/nh_n^2)^d(2^{-d} + 1)(1 + \theta_0)^{1-d} &\int \{f_G(y)\}^{1-d}dy . \qquad (5.21)\end{aligned}$$

Arguments similar to those used above can be used to prove that both of the integrals that appear in (5.21) are finite. This, together with the particular choice of $h_n = O(n^{-1/(2\ell+1)})$, yields

$$\int |\Delta_G(y)|^{1-d} E|\Delta_n^{(1)}(y) - \Delta_n(y)|^d dy \leq O(n^{-d(2\ell-1)/(2\ell+1)}) .$$

The proof of the theorem is completed by using Lemma 2.

Q.E.D.

This theorem completes the results on testing "small outliers" against "large outliers". In the next section we shall consider another variation of tests for outliers under the e. B. approach.

6. Unknown Mean and Variance.

All the results derived up to now have assumed μ and σ^2 are both known. In this section we shall relax this assumption. The first result we present corresponds to the situation given in Theorem 2 without the assumption that μ and σ^2 are known.

THEOREM 6. Assume all the conditions of Theorem 2 except that μ and σ^2 are unknown and

$$\tilde{p} = \min\{1, p^*\}, \qquad p^* = \{(m_2 - \bar{X}^2 - \lambda\tilde{\sigma}^2)/(1-\lambda)\tilde{\sigma}^2\}^+$$

where $A^+ = \max\{0, A\}$,

$$m_s = (1/n)\sum_{r=1}^{n} X_r^s; \qquad \bar{X} = m_1 ,$$

and

$$\tilde{\sigma}^2 = \{3(m_2 - \bar{X}^2)(1+\lambda) \pm \sqrt{\{(m_2 - \bar{X}^2)^2(1+\lambda)^2 9 - 12\lambda(m_4 - 6\bar{X}^2 m_2 + 5\bar{X}^4)\}^+} \quad . \tag{6.1}$$

Also let

$$\tilde{f}_1(x) = (1/\sqrt{2\pi\tilde{\sigma}^2})\exp\{-(x-\bar{X})^2/2\tilde{\sigma}^2\} , \tag{6.2}$$

$$\tilde{f}_\lambda(x) = (1/\sqrt{2\pi\tilde{\sigma}^2\lambda})\exp\{-(x-\bar{X})^2/2\tilde{\sigma}^2\lambda\} , \tag{6.3}$$

and

$$\Delta_n(x) = (1-\tilde{p})L_0(\lambda)\tilde{f}_\lambda(x) - \tilde{p}L_1(1)\tilde{f}_1(x) .$$

Then the e.B. decision rule

$$t_n(X_r) = \begin{cases} 0 & \text{if } \Delta_n(X_r) \geq 0 \\ 1 & \text{otherwise,} \end{cases}$$

$r = 1, \ldots, n,$ is a.r.a.o. for testing H_0, relative to $G = \{p, q\}$.

As in Theorem 2, (4.3) and (4.4) are true and by the same argument given there we have that (3.4) is satisfied. The remaining conditions of Theorem 1 are easily verified by using Slutzky's theorem and the fact that the sample moments are consistent and such that, as $n \to \infty$,

$$m_4 \xrightarrow{P} 3\sigma^4[p + (1-p)\lambda^2] + 6\mu^2\sigma^2[p + (1-p)\lambda] + \mu^4,$$

$$m_2 \xrightarrow{P} \sigma^2\{p + (1-p)\lambda\} + \mu^2,$$

$$m_1 = \bar{X} \xrightarrow{P} \mu. \tag{6.4}$$

If only σ^2 is unknown similar arguments can be used. We must use μ instead of $\bar{X}$ and we do not need (6.4).

Notice that (6.1) is written with both signs (+ and -). The question of which root to use can be answered as follows. In the expression under the square root sign, as $n \to \infty$,

$$9(m_2 - \bar{X}^2)^2(1+\lambda)^2 - 12\lambda(m_4 - 6\bar{X}^2 m_2 + 5\bar{X}^4) \xrightarrow{P}$$

$$9\sigma^4\{2\lambda - p - (1-p)\lambda - \lambda p - (1-p)\lambda^2\}^2 = 9\sigma^4\{p + (1-p)\lambda + p\lambda + (1-p)\lambda^2 - 2\lambda\}^2 .$$

When we take the positive or negative signed term, the two possibilities with respect to $\tilde{\sigma}^2$ that we have as $n \to \infty$ are

$$\tilde{\sigma}^2 \xrightarrow{P} \sigma^2 \quad \text{or} \quad \tilde{\sigma}^2 \xrightarrow{P} \sigma^2(\lambda - \lambda p + p/\lambda) .$$

But

$$\sigma^2(\lambda - \lambda p + p/\lambda) < \sigma^2 \quad \text{if} \quad p > \lambda/(1+\lambda)$$
$$> \sigma^2 \quad \text{if} \quad p < \lambda/(1+\lambda) .$$

Now, in practical applications $(1 - p)$ would usually be small and $(1 - p) = 0.1$ is already considered too extreme. Box and Tiao (1968), take the extreme value $\lambda = 100$. On the other hand it seems reasonable to think that the larger the discrepancy between the variances (i.e. the larger λ) the smaller we would expect $(1 - p)$ to be. So that in general we can expect p to be such that $p < \lambda/(1+\lambda)$. This can be taken as an indication that the estimate of σ^2 is the largest of the two values obtained. This in turn means we should take the positive root in (6.1).

We present the following rate theorem for the situation where σ^2 is known and μ is unknown.

THEOREM 7. Assume the conditions of Theorem 2 except that μ is unknown and $\tilde{p}$ is defined as

$$\tilde{p} = \min\{1, p^*\}, \qquad p^* = \{(m_2 - \bar{X}^2 - \lambda\sigma^2)/(1-\lambda)\sigma^2\}^+ .$$

$f_{\lambda^*}(x)$, $\lambda^* = 1, \lambda$ is defined as in (6.2) and (6.3) but with σ^2 instead of $\tilde{\sigma}^2$ and $\Delta_n(x) = (1 - \tilde{p})L_0(\lambda)\tilde{f}_\lambda(x) - \tilde{p}L_1(1)\tilde{f}_1(x)$. Then,

$$r^*(\underline{t}_n, G) - r(G) = O(n^{-1/2}) .$$

PROOF. From Lemma 1, with $d_1 = d$, $3/4 < d < 1$, and $d_2 = 1$, we have

$$r^*(t_n^{(1)}, G) - r(G) \le 2 \int |\Delta_G(x)|^{1-d} E|\Delta_n^{(1)}(x) - \Delta_n(x)|^d dx +$$
$$2 \int E|\Delta_n(x) - \Delta_G(x)| dx . \tag{6.5}$$

The definition of $\Delta_n(x)$ and $\Delta_G(x)$ along with repeated use of the c_r-inequality yields

$$\int E|\Delta_n(x) - \Delta_G(x)|dx \leq |L(a_0, \lambda)(1 - p)| \int E|\tilde{f}_\lambda(x) - f_\lambda(x)|dx +$$

$$|L(a_1, 1)p| \int E|\tilde{f}_1(x) - f_1(x)|dx + |L(a_0, \lambda)| \int E|\tilde{f}_\lambda(x)(p - \tilde{p})|dx +$$

$$|L(a_1, 1)| \int E|\tilde{f}_1(x)(p - \tilde{p})|dx .$$

Now let

$$A_n = \int E|\tilde{f}_\lambda(x) - f_\lambda(x)|dx, \qquad B_n = \int E|\tilde{f}_1(x) - f_1(x)|dx ,$$

$$C_n = \int E|\tilde{f}_\lambda(x)(p - \tilde{p})|dx \quad \text{and} \quad D_n = \int E|\tilde{f}_1(x)(p - \tilde{p})|dx .$$

If in B_n we take the Taylor Series expansion of $f_1(x)$ about $\tilde{\mu}$, we get

$$f_1(x) = \tilde{f}_1(x) + (\mu - \tilde{\mu})f_1'(x, \mu^*) \tag{6.6}$$

where

$$f_1'(x, \mu^*) = \{df_1(x)/d\mu\}_{\mu=\mu^*} = \{f_1(x)\}_{\mu=\mu^*} \cdot (x - \mu^*)/\sigma^2 \tag{6.7}$$

and $\mu^* = \mu + \zeta(\mu - \tilde{\mu})$, $0 < \zeta < 1$. Hence

$$\int E|\tilde{f}_1(x) - f_1(x)|dx = E(|\mu - \tilde{\mu}|/\sigma^2) \int |x - \mu^*|f_1(x, \mu^*)dx , \tag{6.8}$$

with

$$f_{\lambda^*}(x, \mu^*) = \{f_{\lambda^*}(x)\}_{\mu=\mu^*}, \quad \lambda^* = 1, \lambda .$$

Now assume $\mu < \tilde{\mu}$. μ and $\tilde{\mu}$ are fixed so that as x changes, μ^* will shift but always staying within the interval $[\mu, \tilde{\mu}]$, hence for

a) $x < \mu$ $\quad |x - \mu| \le |x - \mu^*|$ and $|x - \tilde{\mu}| \ge |x - \mu^*|$ (6.9)

b) $x > \tilde{\mu}$ $\quad |x - \mu| \ge |x - \mu^*|$ and $|x - \tilde{\mu}| \le |x - \mu^*|$ (6.10)

c) $\mu \le x \le \tilde{\mu}$ $\quad 0 \le |x - \mu^*| \le |\mu - \tilde{\mu}|$. (6.11)

The integral in (6.8) can be broken into three integrals, each one taken over one of the intervals indicated in a), b) and c). The inequalities given in (6.9)-(6.11) can be used (de Alba (1974)) in the corresponding integrals to obtain

$$(1/\sigma) \int |x - \mu^*|(1/\sigma\sqrt{2\pi})\exp\{-(x - \mu^*)^2/2\sigma^2\}dx \le$$
$$\nu_1^\circ + |\mu - \tilde{\mu}|/\sigma + (\tilde{\mu} - \mu)^2/(\sigma^2\sqrt{2\pi}) , \tag{6.12}$$

where ν_1° is the first absolute moment of a standard normal random variable. The same result is true if $\tilde{\mu} > \mu$. In general, since $E\tilde{\mu} = \mu$,

$$\int E|\tilde{f}_1(x) - f_1(x)|dx = (1/\sigma)E\{\nu_1^\circ|\mu - \tilde{\mu}| + (\mu - \tilde{\mu})^2/\sigma + |\mu - \tilde{\mu}|^3/(\sigma^2\sqrt{2\pi})\} \le$$
$$(1/\sigma)\{\nu_1^\circ\sigma_x n^{-1/2} + (\sigma_x^2/\sigma)n^{-1} + (\sigma_x^3/\sigma^2\sqrt{2\pi})n^{-3/2}\} = O(n^{-1/2}) , \tag{6.13}$$

with $\sigma_x^2 = p\sigma^2 + (1 - p)\lambda\sigma^2$. A similar argument can be used to prove $A_n = O(n^{-1/2})$.

Now in D_n we have, by Fubini's Theorem,

$$\int E|\tilde{f}_1(x)(p - p)|dx = E\{|p - \tilde{p}| \int \tilde{f}_1(x)dx\} = E|p - \tilde{p}| \le E|p - p'| , \tag{6.14}$$

where

$$p' = (m_2 - \bar{X}^2 - \lambda\sigma^2)/(1 - \lambda)\sigma^2 . \tag{6.15}$$

It can be shown (de Alba (1974)) that $E|p - p'| \le O(n^{-1/2})$ and so $D_n = O(n^{-1/2})$. Here also, a similar argument can be used to prove $C_n = O(n^{-1/2})$. Thus

$$\int E|\Delta_n(x) - \Delta_G(x)|dx = O(n^{-1/2}) . \tag{6.16}$$

This completes the proof that the second term in (6.5) is $O(n^{-1/2})$.

We shall now find the rate of convergence of the first term. Using $\tilde{p}_1 = \tilde{p}(\underline{X}_1(x))$ and $\tilde{\mu}_1 = \tilde{\mu}(\underline{X}_1(x))$ along with the facts

$$|\tilde{f}_{\lambda^*}(x, \tilde{\mu}_1)| = |1/\sigma\sqrt{2\pi}| \quad \text{and} \quad |\tilde{f}_{\lambda^*}(x)| \leq |1/\sigma\sqrt{2\pi}|, \; \lambda^* = 1, \lambda$$

we get

$$|\Delta_n^{(1)}(x) - \Delta_n(x)|^d \leq |1/\sigma\sqrt{2\pi}|^d |(1 - \tilde{p}_1)L(a_0, \lambda) - \tilde{p}_1 L(a_1, \lambda) - (1 - \tilde{p})L(a_0, \lambda) +$$

$$\tilde{p}L(a_1, \lambda)|^d = |1/\sigma\sqrt{2\pi}|^d |L(a_0, \lambda) + L(a_1, \lambda)|^d |\tilde{p}_1 - \tilde{p}|^d = A_0 |\tilde{p}_1 - \tilde{p}|^d ,$$

where A_0 is defined in an obvious manner. Then

$$E|\Delta_n^{(1)}(x) - \Delta_n(x)|^d \leq A_0 E|\tilde{p}_1 - \tilde{p}|^d \leq A_0 E|p_1' - p'|^d . \tag{6.17}$$

Repeated use of the c_r-inequality gives

$$|(1 - \lambda)\sigma^2|^d E|p_1' - p'|^d \leq ((n - 1)/n^2)^d (x^{2d} + EX_1^{2d}) +$$

$$2^d n^{-2d}(|x|^d \sum_{i=2}^{n} E|X_i|^d + \sum_{i=2}^{n} E|X_1 X_i|^d) . \tag{6.18}$$

Let

$$\nu_d' = E|X_i|^d < \infty, \quad i = 1, \ldots, n . \tag{6.19}$$

Substitution of (6.19) in (6.18) and using the Schwarz inequality, together with (6.17), yields

$$E|\Delta_n^{(1)}(x) - \Delta_n(x)|^d \leq A_0 |\sigma^2(1 - \lambda)|^{-d} \{[(n - 1/n^2)^d (x^{2d} + \nu_{2d}') +$$

$$2^d n^{-2d}(|x|^d (n - 1)\nu_d' + (n - 1)\nu_{2d}')]\} . \tag{6.20}$$

Further, if $d > 1/2$,

$$\int |\Delta_G(x)|^{1-d} E|\Delta_n^{(1)}(x) - \Delta_n(x)|^d dx \leq A_0'(\tau_{2d} + \nu'_{2d})\{(n-1)/n\}^d n^{-d} +$$

$$A_0' 2^d(\tau_d \nu'_d + \nu'_{2d})\{(n-1)/n\} n^{1-2d} = O(n^{-(2d-1)}), \qquad (6.21)$$

where A_0' is a constant and $\tau_a = \int |x|^a f_{\lambda/(1-d)}(x)dx < \infty$. Substitution of (6.16) and (6.21) in (6.5) yields

$$r^*(t_n^{(1)}, G) - r(G) = O(n^{-1/2}) .$$

From Lemma 2 the proof is complete.

Q.E.D.

7. Concluding Remarks.

In this paper we have presented a first approach via e.B. methods to the problem of detecting outliers. Other alternatives are surely possible. A first one would be to use a different model, perhaps considering as outliers those observations which have a shifted mean rather than a larger variance. An interesting problem would be to determine a criterion to estimate the value of the increase in variance (λ) for the spurious observations. Other articles on some of these extensions are being prepared.

REFERENCES

(1) de Alba, E. (1974). An empirical Bayes approach to the detection of spurious observations and to inferences about a covariance matrix. U. of Wisconsin, Ph. D. thesis.

(2) Box, G. E. P. and Tiao, G. C. (1968). A Bayesian approach to some outlier problems. Biometrika 55, 119-129.

(3) Ferguson, T. S. (1961a). On the rejection of outliers. In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability.

(4) Ferguson, T. S. (1961b). Rules for rejection of outliers. Rev. Inst. Internat. Statist. 29, 29-43.

(5) Guttman, I. and Smith, P. E. (1969). Investigation of rules for dealing with outliers in small samples from the normal distribution: I. Estimation of the mean. Technometrics 11, 527-550.

(6) Johns, M. V. and Van Ryzin, J. (1971). Convergence rates for empirical Bayes two-action problems I. Discrete case. Ann. Math. Statist. 42, 1521-1539.

(7) Johns, M. V. and Van Ryzin, J. (1972). Convergence rates for empirical Bayes two-action problems II. Continuous case. Ann. Math. Statist. 43, 934-947.

(8) Maritz, J. S. (1970). Empirical Bayes Methods, Methuen, London.

(9) Rider, P. L. (1933). Criteria for rejection of observations. St. Louis, Washington Univ. Studies, new series, Science and Technology, No. 8.

(10) Robbins, H. (1963). The empirical Bayes approach to testing statistical hypotheses. Rev. Inst. Internat. Statist. 31, 195-208.

(11) Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. Ann. Math. Statist. 35, 1-19.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: MRC-TSR-1734

2. GOVT ACCESSION NO.: (9) Technical

3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): (6) AN EMPIRICAL BAYES APPROACH TO OUTLIERS

5. TYPE OF REPORT & PERIOD COVERED: Summary Report - no specific reporting period

6. PERFORMING ORG. REPORT NUMBER:

7. AUTHOR(s): (10) Enrique de Alba a[redacted]d J. Van Ryzin

8. CONTRACT OR GRANT NUMBER(s): (15) DAAG29-75-C-0024, PHS [redacted] CA18332-02

9. PERFORMING ORGANIZATION NAME AND ADDRESS: Mathematics Research Center, University of Wisconsin, 610 Walnut Street, Madison, Wisconsin 53706

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 4 (Probability, Statistics, and Combinatorics)

11. CONTROLLING OFFICE NAME AND ADDRESS: See Item 18 below. (12) 32p.

12. REPORT DATE: (11) Mar[redacted]977

13. NUMBER OF PAGES: 29

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):

15. SECURITY CLASS. (of this report): UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

National Institutes of Health
Bethesda, Maryland
20014

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Empirical Bayes procedure
detection of outliers
average risk asymptotic optimality
rates of convergence

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A formulation of the problem of detecting outliers as an empirical Bayes problem is studied. In so doing what arises is a non-standard empirical Bayes problem for which the notion of average risk asymptotic optimality (a.r.a.o.) of procedures is defined. Some general theorems giving sufficient conditions for a.r.a.o. procedures are developed. These general results are then used in various formulations of the outlier problem for underlying normal distributions to give a.r.a.o., empirical Bayes procedures. Rates of convergence results are also given using the methods of Johns and Van Ryzin (1971, 1972).

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED 221 200

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)