ON THE RISK OF SOME STRATEGIES FOR OUTLYING OBSERVATIONS[1]

By Friedrich Gebhardt[2]

University of Connecticut

1. Introduction and summary. Many procedures have been proposed for handling outlying observations. Most authors use various statistics T to reject one or more observations, if T is too large (or too small), and compute percentage points of the distribution function of T. Some authors, as in [11], [13], [15], [16], find statistics with the optimum property of minimizing for certain alternative hypotheses the error of the second kind given the error of the first kind. The observations that are not rejected are used to estimate unknown parameters, e.g. the mean.

In this paper, we shall consider one of these procedures which gives rise to a one-parameter family of estimators for the mean and compare their risks with those of the Bayes solutions with respect to a one-parameter family of prior distributions. Two simplified models of outliers as will be specified in (2.1) and (2.2) and a quadratic loss function will be investigated. For sample sizes of 3, 6, and 10, the risks are tabulated and plotted. For Model (2.1) the risks of the first procedure are only a little larger than those of the Bayes solutions; in other words, to each admissible strategy there exists an estimator arising from the first procedure with approximately the same risk. Winsorizing the observations [18] should presumably result in an even better approximation; however, numerical results are not available. For Model (2.2), the risks differ considerably. Although this model is rather unrealistic it is included here because of its prominence in the literature.

References for tables of percentage points and the distribution functions of various statistics are collected in the table guide of J. A. Greenwood and H. O. Hartley [7]. Bayes solutions are considered by S. Karlin and D. Truax [10], B. de Finetti [3], and T. E. Ferguson [4]. Some other important papers are [1], [12], [14], [18].

In this paper, "stragglers" are random variables that obey a probability law with a larger variance or with a larger or smaller expectation than the "typical" random variables. An "outlier" is an observation that because of its magnitude may be suspected of being a straggler. Thus the term "outlier" is not precisely defined.

2. Bayes solutions. We shall now investigate two special cases in which there exists not a whole family of alternative hypotheses, but rather apart from permutations of the indices only one: We assume that all parameters of the alternative hypothesis are known except those also appearing in the null hypothesis.

Received March 13, 1963; revised June 17, 1964.

[1] Part of the research was done while the author was at the Statistische Arbeitsgruppe Prof. Dr. H. Kneser, Universität Tübingen.

[2] Now at Deutsches Rechenzentrum, Darmstadt.

Let the random variables $X_1, \cdots, X_n$ be normally and independently distributed with expected values $m_1, \cdots, m_n$ and variances $\sigma_1^2, \cdots, \sigma_n^2$. In the first model the straggler has a greater variance than the other variables (the notation "$A := B$" or "$B =: A$" defines A as B):

$$(2.1)\qquad \begin{aligned} H_0: \quad & m_1 = \cdots = m_n =: m, \\ & \sigma_1 = \cdots = \sigma_n = 1; \\ H_i: \quad & m_1 = \cdots = m_n = m, \\ & \sigma_j = 1\ (j \neq i), \sigma_i = \sigma > 1 \qquad (i = 1, \cdots, n). \end{aligned}$$

In the second model the straggler has a distinct expected value:

$$(2.2)\qquad \begin{aligned} H_0: \quad & m_1 = \cdots = m_n =: m, \\ & \sigma_1 = \cdots \sigma_n = 1; \\ H_i: \quad & m_j = m\ (j \neq i), m_i = m + a \\ & \sigma_1 = \cdots = \sigma_n = 1 \qquad (i = 1, \cdots, n). \end{aligned}$$

Let m be unknown and a and σ be known. In either case there is at most one straggler. The purpose of this restriction is to simplify the numerical calculations of Section 3 which otherwise would exceed the capacity of a medium sized computer. If a straggler is a rather rare event, our model can be expected to be a fairly good approximation to the general case where more than one straggler is admitted. We will discuss this question a little bit further at the ends of both this section and Section 3.

Usually the decisions

D_i: *H_i is regarded to be the underlying hypothesis* $(0 \leqq i \leqq n)$ are admitted together with a loss function $L(H_i, D_j) = L_{ij}$. Often the special case

$$\begin{aligned} L_{ij} &= 0 \quad \text{if} \quad i = j, \\ &= 1 \quad \text{if} \quad i \neq j \end{aligned}$$

and prior probabilities $\phi_1 = \cdots = \phi_n =: \phi$, $\phi_0 = 1 - n\phi$, for the hypotheses $H_1, H_2, \cdots, H_n, H_0$ are considered, as by S. Karlin and D. Truax [10] and T. E. Ferguson [4]. The class of admissible strategies is in this case just the class of optimum procedures mentioned in the first section, the particular one now being selected by the value of ϕ instead of the error of the first kind α.

Typically, one is not really interested in which observation is a straggler; rather one wants to estimate the unknown parameter m. Thus, if one knew that the correct hypothesis in (2.1) were H_1, one would use with advantage

$$(2.3)\qquad \frac{1}{n - 1 + \sigma^{-2}} \left(\sum_{j=2}^{n} X_j + \frac{1}{\sigma^2} X_1 \right) = \bar{X} - \frac{1 - \sigma^{-2}}{n - 1 + \sigma^{-2}} (X_1 - \bar{X})$$

which is an efficient estimator for m utilizing the information about m contained in X_1 as well as that in the other observations.

A Bayesian approach to the estimation of m is due to B. de Finetti [3]. Starting with a (subjective) prior probability density of the parameters involved, one gets the posterior density by multiplication with the likelihood. If the influence of some observations on the posterior distribution is weak or practically negligible, de Finetti calls these observations outliers.

In this paper, however, we are not concerned with the subjective probability distributions of the unknown parameters, rather we will use the Bayesian estimators to minimize the weighted sum of risks for a quadratic loss function.

Let $Z_i = X_i - \bar{X}$ and write X instead of $(X_1, \cdots, X_n)$ and correspondingly x, Z, z.

The theory of statistical decision functions [5], [17], [19] will now be used. Let us denote by $\bar{X} - b(Z, \bar{X})$ an estimator for m, and consider the quadratic loss

$$L(b) = (\bar{X} - b - m)^2. \tag{2.4}$$

The risk R_i under the hypothesis H_i is defined as the expectation of the loss and it depends in general also on m:

$$\begin{aligned} R_i(b; m) &:= E_i(L(b)) \\ &= \int [\bar{x} - b(z, \bar{x}) - m]^2 f_i(z, \bar{x} - m)\, dz_1 \cdots dz_{n-1}\, d\bar{x}. \end{aligned} \tag{2.5}$$

Here the probability densities f_i have already been written as functions of z and $\bar{x}$. In all formulas, z_n denotes the function $-\sum_{j=1}^{n-1} z_j$. For some simple estimators $\bar{X} - b(Z, \bar{X})$ and a quadratic loss function, the risk was approximately computed by F. J. Anscombe [1] who in addition deals with complex patterns of data, such as factorial designs.

If $\bar{X} - b(Z, \bar{X})$ is an unbiased estimator for m, then, when (2.4) is the loss, the risk is just the variance of $\bar{X} - b(Z, \bar{X})$.

As (simple) decisions D_b are admitted D_b: $\bar{x} - b$ *is the estimate for* m.

The decision functions (composite decisions) are then just the measurable functions $\bar{x} - b(\bar{x}, z)$. Let $(\phi_0, \phi_1, \cdots, \phi_n)$ be a prior distribution of the hypotheses $H_0, H_1, \cdots, H_n$. Within each of these hypotheses we consider an equal prior distribution (in the limit) of all values of m, i.e. we try to minimize the sum

$$\sum_{i=0}^{n} \phi_i \lim_{M \to \infty} 1/2M \int_{-M}^{+M} R_i(b, m)\, dm, \tag{2.6}$$

if the limit exists. First we shall show that we can restrict ourselves to decision functions $\bar{x} - b(z)$, where b is independent of $\bar{x}$, so that $R_i(b, m)$ is independent of m and the limit exists, trivially.

Let

$$A(M, z) = 1/2M \int_{-M}^{+M} dm \int_{-\infty}^{+\infty} d\bar{x} \sum \phi_i [\bar{x} - m - b(\bar{x}, z)]^2 f_i(z, \bar{x} - m).$$

For the sake of simplicity we shall make the assumption, that $|b(\bar{x}, z)|$ is bounded by a polynomial $B(z)$. This seems to be no essential restriction, since one would expect that the estimator $\bar{x} - b(\bar{x}, z)$ falls into the range of the observations x_i and since the final result does not depend on the particular function $B(z)$ as long as $B(z) > \max |z_i|$. A closer analysis shows indeed that this assumption is not necessary.

From our assumption follows that $A(M, z)$ is bounded by an integrable function $A^*(z)$. Therefore, if $\lim_{M\to\infty} A(M, z)$ exists, the order of integral and limit can be exchanged and (2.6) becomes

$$\lim_{M\to\infty} \int A(M, z)\, dz = \int \lim_{M\to\infty} A(M, z)\, dz. \tag{2.7}$$

(If $\lim_{M\to\infty} A(M, z)$ does not exist, all considerations and the inequality (2.8) still hold for any sequence $\liminf_{\nu\to\infty} A(M_\nu)$ where $\lim M_\nu = \infty$).

Writing $y = \bar{x} - m$ and dropping the variable z which for a while will be held fixed we find

$$A(M) = \frac{1}{2M} \int_{-\infty}^{+\infty} dy \int_{y-M}^{y+M} d\bar{x} \sum \phi_i[y - b(\bar{x})]^2 f_i(y).$$

The inner integral can be split into three integrals ranging from $y - M$ to $-M$, from $-M$ to $+M$, and from $+M$ to $y + M$, respectively. As $b(\bar{x})$ is bounded, so is

$$I_1(M) = \int_{-\infty}^{+\infty} dy \int_{y-M}^{-M} \sum \phi_i[y - b(\bar{x})]^2 f_i(y)$$

and therefore $\lim_{M\to\infty} I_1(M)/2M = 0$. The same holds for the third integral, so that finally

$$\begin{aligned} \lim_{M\to\infty} A(M) &= \lim_{M\to\infty} \frac{1}{2M} \int_{-\infty}^{+\infty} dy \int_{-M}^{+M} d\bar{x} \sum \phi_i[y - b(\bar{x})]^2 f_i(y) \\ &\geqq \min_b \int_{-\infty}^{+\infty} dy \sum \phi_i[y - b]^2 f_i(y) \end{aligned} \tag{2.8}$$

The value b_0 that minimizes (2.8) depends of course on z. $b_0(z)$ is a solution of the problem (2.6) provided $b_0(z)$ is measurable. This measurability can be expected because all functions involved are continuous. Consequently we will consider now decision functions $b(z)$, and because the Bayes solution will be seen to be measurable in the $(\bar{x}, z)$-space, it is a solution of (2.6).

As $b(z)$ is now a function of z only, (2.5) can easily be integrated with respect to $\bar{x}$. Let $\bar{X} - w_i(Z)$ be an efficient estimator for m under the hypothesis H_i; such a $w_i(Z)$ always exists in the cases of interest to us. Then

$$\begin{aligned} R_i(b) &= \operatorname{var}_i(\bar{X} - w_i(Z)) + \int [w_i(z) - b(z)]^2 F_i(z)\, dz_1 \cdots dz_{n-1} \\ &= R_i' + r_i(b) \end{aligned} \tag{2.9}$$

with $F_i(z) = \int f_i(z, \bar{x} - m)\, d(\bar{x} - m)$.

Thus $R_i(b)$ does not depend on m, and our problem is reduced to a finite number of simple hypotheses.

The term R_i' on the right side of (2.9) depends only on the hypothesis H_i, not, however, on the estimator $\bar{X} - b(Z)$. If one wishes to compare two different estimators, R_i' plays no part at all. Let us therefore concern ourselves first only with r_i.

To minimize $\sum \phi_i r_i$, one has to minimize $\sum \phi_i [w_i(z) - b(z)]^2 F_i(z)$ for fixed z with respect to b. The solution is

$$b_\phi(z) = (\sum \phi_i w_i(z) F_i(z))/(\sum \phi_i F_i(z)) \tag{2.10}$$

Let us introduce the following notations: $c := n(2\pi)^{-(n-1)/2}$, $k := (n-1)\sigma^2 + 1$, $K := [(n-1)\sigma^2 + 1]/[n(\sigma^2 - 1)]$. For the hypotheses (2.1) we have (compare (2.3))

$$\begin{aligned} w_0(z) &= 0, \\ w_i(z) &= z_i/nK \qquad (i \neq 0), \\ F_0(z) &= (c/(n)^{\frac{1}{2}}) \exp\{-\tfrac{1}{2}\sum z_\nu^2\}, \\ F_i(z) &= (c/(k)^{\frac{1}{2}}) \exp\{-\tfrac{1}{2}\sum z_\nu^2 + (1/2K)z_i^2\} \qquad (i \neq 0). \end{aligned}$$

In the case of a symmetrical prior distribution $(1 - n\phi, \phi, \cdots, \phi)$, defining $C := (k/n)^{\frac{1}{2}}(1 - n\phi)/\phi$, the Bayes solution

$$b_\phi(z) = \frac{\sum z_i \exp\{z_i^2/2K\}}{nK(C + \sum \exp\{z_i^2/2K\})} \tag{2.11}$$

results.

Writing $D = [(1 - n\phi)/\phi] \exp\{\frac{1}{2}[(n-1)/n]a^2\}$, the corresponding formulae for the hypotheses (2.2), again with the symmetrical prior distribution, are:

$$\begin{aligned} w_0(z) &= 0, \qquad w_i(z) = a/n \qquad (i \neq 0), \\ F_0(z) &= (c/(n)^{\frac{1}{2}}) \exp\{-\tfrac{1}{2}\sum z_\nu^2\}, \\ F_i(z) &= (c/(n)^{\frac{1}{2}}) \exp\{-\tfrac{1}{2}\sum z_\nu^2 + az_i - \tfrac{1}{2}[(n-1)/n]a^2\} \qquad (i \neq 0), \end{aligned}$$

$$b_\phi(z) = (a/n)(\sum \exp\{az_i\})/(D + \sum \exp\{az_i\}). \tag{2.12}$$

In both cases the Bayes solutions $\bar{x} - b_\phi(z)$ are rather complicated.

If all z_i are small, then $b_\phi \approx 0$. If one z_j is much larger than the rest, then (2.11) becomes $b_\phi(z) \approx z_j/nK$ and (2.12) becomes $b_\phi(z) \approx a/n$, that is $w_j(z)$ in either case.

In a similar way, the Bayes solution for a model permitting more than one straggler can be computed. Let ψ_r be the prior probability of each hypothesis assuming r stragglers; $\sum^*$ stands for a summation with respect to all r-tuples $i_1 < i_2 < \cdots < i_r$. Writing $S = 1 - \sigma^{-2}$

$$h(i_1, \cdots, i_r) = \exp\left\{\frac{S}{2}\sum_{i=i_1}^{i_r} z_i^2 + \frac{1}{2}\frac{S^2}{n - rS}\left[\sum_{i=i_1}^{i_r} z_i\right]^2\right\},$$

$$M_r = \psi_r \sigma^{-r}(n - rS)^{-3/2} \cdot \sum^* \left[h(i_1, \cdots, i_r) \sum_{i=i_1}^{i_r} z_i\right]$$

$$N_r = \psi_r \sigma^{-r}(n - rS)^{-1/2} \cdot \sum^* h(i_1, \cdots, i_r)$$

the Bayes solution for stragglers differing in the variance is

$$a_\phi(z) = S \sum_{r=1}^{n-1} M_r \Big/ \sum_{r=0}^{n} N_r . \tag{2.13}$$

(2.11) arises from (2.13) if $\psi_0 = 1 - n\phi$, $\psi_1 = \phi$, $\psi_2 = \cdots = \psi_n = 0$. Due to the factor $\psi_r \sigma^{-r}$, M_r and $N_r (r \geqq 2)$ are small as compared to M_1 and N_1 as long as all z_i or all but one are absolutely small. If r observations z_i are absolutely large, then M_r and N_r will predominate and $a_\phi(z) \approx S/(n - rS) \sum_{i=i_1}^{i_r} z_i$ so that these r observations have little weight in $\bar{x} - a_\phi(z)$.

3. Risks of several estimators for the mean. The risk of another estimator $\bar{X} - b^*(Z)$, which is formulated very simply, will be compared with the risk of the Bayes solution. This second estimate arises when the largest (or smallest) observation is not used in estimating m if it deviates too much from the sample mean. In particular, under the hypotheses (2.1) we use

$$\begin{aligned} \bar{X} - b^*(Z) &= \bar{X} && \text{if } \max |Z_i| \leqq \beta, \\ &= (1/(n-1)) \sum_{\nu \neq M} X_\nu = \bar{X} - (1/(n-1)) Z_M && \\ & && \text{otherwise, } |Z_M| = \max |Z_i|. \end{aligned} \tag{3.1}$$

Correspondingly, under the hypotheses (2.2) for $a > 0$

$$\begin{aligned} \bar{X} - b^*(Z) &= \bar{X} && \text{if } \max Z_i \leqq \beta, \\ &= \bar{X} - (1/(n-1)) Z_M && \text{otherwise, } \quad Z_M = \max Z_i . \end{aligned} \tag{3.2}$$

The risk of the Bayes solutions and of the estimators (3.1), (3.2) respectively were computed for various parameter values on the electronic computer Siemens 2002 of Tübingen University. The results are compiled in Tables 1–4. The error of the numerical computation is in some instances several units of the last indicated decimal.

Since the parameter ϕ of the Bayes solution has no relationship to the parameter β of the estimate functions (3.1) and (3.2), the risks of b_ϕ and b^* cannot be compared effectively by means of the tables. This comparison can be much better effected by means of Figures 1 and 2. The thick lines represent the risk of the Bayes solutions, the thin ones, the risk of b^*. In the case of the model (2.1) (where stragglers have a larger variance), the risk of b^* is not much larger than that of b_ϕ. If, just for the purpose of comparing the risks, the parameters are so chosen that $r_0(b_\phi) = r_0(b^*)$, then $r_1(b^*)$ is a fourth to a half larger than $r_1(b_\phi)$, and this fraction becomes smaller as n increases. This means that $R_1(b^*)$ is only a few percent larger than $R_1(b_\phi)$, e.g. about 2–5% for $n = 10$, while, by contrast, a decrease of n by one would increase both R_0 and R_1 by 11–12½%. In the case of the hypotheses (2.2), the differences are considerably greater. Choosing again the parameters so that $r_0(b_\phi) = r_0(b^*)$, then for $n = 3$, $R_1(b^*)$ is 30% to 50% larger than $R_1(b_\phi)$, for $n = 10$ still about 10%. This shows that considerable information is disregarded if the outlier is not used while the parameter a is known.

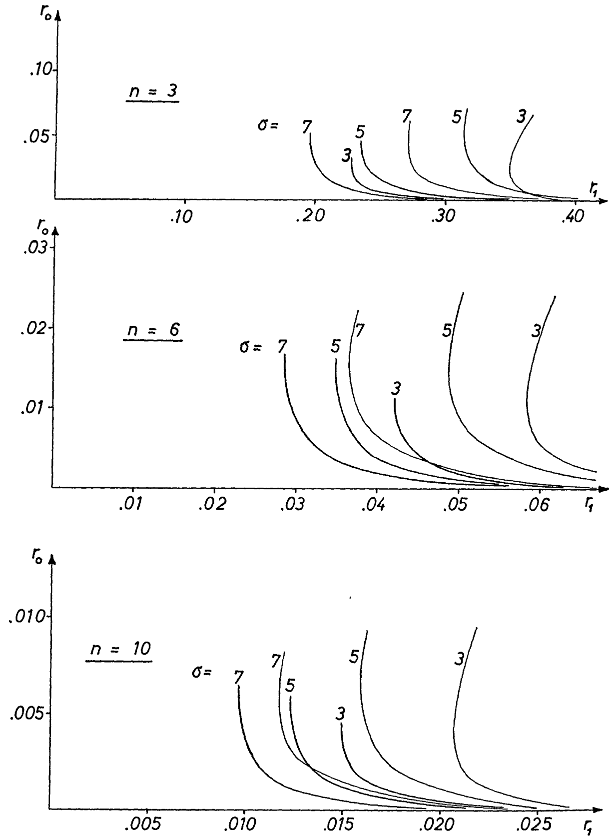

FIG. 1. Outliers with different variance. Bold lines: risk of the Bayes solution b_ϕ. Fine lines: risk of the estimate b^*.

While the Bayes solutions have a smaller risk than the statistics (3.1) and (3.2), the latter ones are more easily computed. There is an additional advantage to the statistics b^*: the straggler parameter a, σ respectively does not enter into $b^*(Z)$, i.e. $\bar{X} - b^*(Z)$ can be used even when this parameter is known only approximately or not at all, and in every case the risk is not appreciably larger than that of a Bayes solution for straggler parameters known, if only β is not too small.

If in the case of a straggler with distinct variance, β becomes too small, the risks $r_0(b^*)$ and $r_1(b^*)$ both increase again. From the diagrams and tables one sees that for $n = 3$ and $n = 10$, $\beta = 2.5$ and $\beta = 3$, respectively, are to be recommended; this corresponds to an error of the first kind of about $\alpha = 0.01$. The value of β corresponding to $\alpha = 0.05$ in any case already results in risks on the upper branch of the thin line curves, where both risks increase again. In the case of a straggler with a distinct expectation the values of β corresponding to

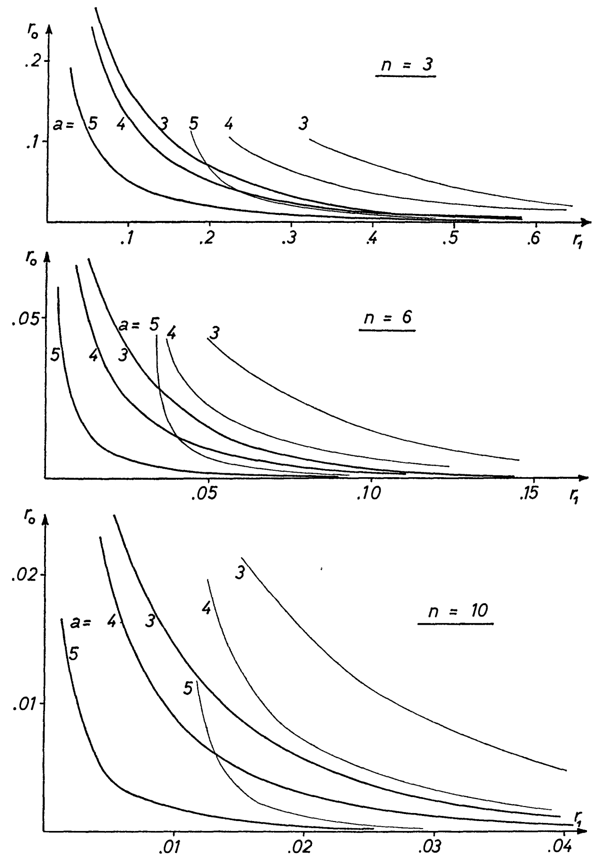

FIG. 2. Outliers with different mean. Bold lines: risk of the Bayes solution b_ϕ. Fine lines: risk of the estimate b^*.

$\alpha = 0.02$ to $\alpha = 0.05$ are recommended. For smaller α, $r_1(b^*)$ climbs steeply without $r_0(b^*)$ becoming appreciably smaller.

If the set of hypotheses (2.1) is extended to allow for any number of stragglers, the Bayes solution is $a_\phi(z)$ as given in (2.13) rather than $b_\phi(z)$, equation (2.11); likewise, $b^*(z)$ should be replaced by an estimate, $a^*(z)$, say, which permits more than one observation to be discarded. For the following discussion, let us denote the risk of a decision function $b(z)$ (symmetric in $z_1, \cdots, z_n$) under any hypothesis assuming j stragglers by $Q_j(b)$, so that e.g. $R_1(b) = \cdots = R_n(b) = Q_1(b)$. The risks $Q_j(a_\phi)$ and $Q_j(a^*)$ have not been computed, but under model (2.1) they again should be nearly equal (at least for $j = 0, 1$) for the following reasons: $Q_0(a_\phi)$ and $Q_1(a_\phi)$ must be larger than $Q_0(b_\phi)$ and $Q_1(b_\phi)$ (more exactly: the graph of a_ϕ must lie to the right of that of b_ϕ), and one will expect that $Q_0(a^*)$ and $Q_1(a^*)$ also are larger than $Q_0(b^*)$ and $Q_1(b^*)$. However, the differences should be small; for we have seen in Section 2 that $b_\phi(z)$ is a good approximation to $a_\phi(z)$ as long as there is at most one outlier, and the probability for two or more outliers is small; the same will be true for $b^*(z)$ and a reasonably chosen $a^*(z)$. None of the risks Q_j has been computed for $j \geqq 2$.

4. On the numerical calculation of the risks. To compute integrals such as (2.7) one often uses the Monte Carlo method, that is one computes the integrand for a large number of random arguments and takes the mean. This procedure converges as $N^{-\frac{1}{2}}$ when N is the number of computed function values. However, it is not necessary that the arguments be independent from each other provided they are evenly scattered. The following procedure was cited by Davis and Rabinowitz [2] to compute the integral

$$I = \int_0^1 \cdots \int_0^1 F(t_1, \cdots, t_n)\, dt_1 \cdots dt_n. \tag{4.1}$$

Let $p_1, \cdots, p_n$ be arbitrary real numbers (e.g., $p_1 = \cdots = p_n = 0$) and $q_1, \cdots, q_n$ real numbers such that no relation $\sum r_\nu q_\nu = r_0$ with rational numbers $r_0, r_1, \cdots r_n$ holds except for $r_0 = \cdots = r_n = 0$, and let

$$\begin{aligned} s_{i_\nu} &\equiv p_\nu + iq_\nu \pmod 1), \qquad 0 \leqq s_{i_\nu} < 1, \\ I_N &= (1/N) \sum_{i=1}^{N} F(s_{i_1}, \cdots, s_{i_n}). \end{aligned} \tag{4.2}$$

Then I_N converges to I and the error is almost of order N^{-1}. This has not been proven, but it has been verified for several examples.

In our case one reasonably proceeds from (2.5), not (2.9), writing this integral in terms of the original integration variables $x_1, \cdots, x_n$, with $m = 0$. Let us denote by $\Phi(x)$ the normal distribution function. Using the transformation $t = \Phi(x)$, it takes the form (4.1), namely

$$I = \int_0^1 \cdots \int_0^1 [\bar{x} - b(z)]^2\, dt_1 \cdots dt_n; \tag{4.3}$$

here $b(z)$ and $\bar{x}$ are to be regarded as functions of $t_1, \cdots, t_n$ by means of $z_j = x_j - \bar{x}$ and $t_j = \Phi(x_j)$.

Thus to be able to compute $b(z)$, one needs the inverse function for $t = \Phi(x)$. Since with moderate computing time (14 hours on the Siemens 2002 have actually been used) one can expect for I an accuracy of three decimals at the most, a moderate accuracy in the computation of $x = x(t)$ suffices; on the other hand, one is interested in a rapid procedure because of the great number of function values. This is accomplished in the method cited in [6].

For fixed n and σ, a, respectively, the values of $r_0(b_\phi)$, $r_1(b_\phi)$, $r_0(b^*)$ and $r_1(b^*)$ were computed simultaneously, and simultaneously for all values of ϕ and β, i.e. using for all these integrals the same values of s_{i_ν}. For the risks under the hypotheses (2.1), N was set equal to 10,000; for the risks under the hypotheses

(2.2), $N = 6{,}000$. In addition the partial sums for 1,000 arguments apiece were computed, in order to be able to estimate the accuracy of the approximation. Although $r_1(b)$ can be explicitly determined for $\phi = 0$, it has also been computed by the method described here to check its accuracy. The agreement was satisfactory.

Acknowledgment. I wish to thank Professor Leonard J. Savage for many helpful suggestions which led to a major revision of the earlier version of this paper and Professor K. Zeller for the permission to use the Siemens 2002 computer in the Computer Center of the Universität Tübingen.

TABLE 1

Risk of the Bayes solution under hypotheses (2.1)

	n:	3			6			10		
	ϕ σ:	3	5	7	3	5	7	3	5	7
r_0	.333	.029	.044	.049						
	.167	.015	.021	.020	.0109	.0146	.0167			
	.10	.009	.013	.012	.0062	.0069	.0073	.0043	.0059	.0065
	.05	.005	.007	.006	.0032	.0032	.0035	.0019	.0022	.0021
	.01	.001	.002	.001	.0007	.0006	.0009	.0004	.0004	.0004
	.005	.001	.001	.001	.0003	.0003	.0005	.0002	.0002	.0002
	.000	0	0	0	0	0	0	0	0	0
R_0'		.333	.333	.333	.1667	.1667	.1667	.1000	.1000	.1000
r_1	.333	.228	.233	.196						
	.167	.233	.242	.206	.0420	.0346	.0282			
	.10	.242	.253	.218	.0432	.0366	.0303	.0149	.0122	.0096
	.05	.259	.274	.238	.0470	.0414	.0351	.0157	.0134	.0109
	.01	.313	.343	.302	.0601	.0573	.0494	.0203	.0188	.0159
	.005	.341	.380	.335	.0678	.0658	.0568	.0228	.0216	.0184
	.000	.749	2.510	5.172	.1932	.635	1.301	.0702	.2294	.4691
R_1'		.474	.490	.495	.1957	.1984	.1992	.1098	.1106	.1109

TABLE 2

Risk of the estimates (3.1) under hypotheses (2.1). Approximate error of first kind, α, according to M. Halperin et al. [9], their formula (4.1)

n	β	α	r_0	r_1		
				$\sigma = 3$	$\sigma = 5$	$\sigma = 7$
3	2.00	.042	.050	.357	.314	.27
	2.25	.017	.026	.348	.321	.27
	2.50	.0066	.011	.353	.345	.30
	2.75	.0023	.004	.370	.379	.34
	3.00	.00071	.002	.396	.430	.38
6	2.25	.079	.0203	.060	.049	.037
	2.50	.036	.0109	.058	.049	.037
	2.75	.015	.0056	.060	.055	.041
	3.00	.0062	.0026	.066	.060	.048
	3.25	.0022	.0011	.073	.070	.056
10	2.50	.08	.0078	.0214	.0160	.012
	2.75	.037	.0040	.0208	.0165	.012
	3.00	.0156	.0019	.0213	.0180	.014
	3.25	.0061	.0009	.0233	.0208	.017
	3.50	.00225	.0003	.0260	.0235	.020

TABLE 3

Risk of the Bayes solution under hypotheses (2.2)

	n:	3			6			10		
	ϕ a:	3	4	5	3	4	5	3	4	5
r_0	.333	1.000	1.777	2.777						
	.25	.250	.205	.131						
	.167				.250	.444	.694			
	.15	.096	.086	.060	.114	.087	.045			
	.10	.055	.054	.040	.037	.029	.016	.0900	.1600	.2500
	.08							.0282	.0205	.0101
	.05	.024	.027	.022	.013	.011	.007	.0104	.0081	.0041
	.025				.0056	.0056	.0039	.0040	.0035	.0018
	.01	.004	.006	.006	.0019	.0023	.0019	.0014	.0013	.0008
	.005	.002	.003	.003	.0008	.0012	.0011	.0006	.0006	.0004
	.000	0	0	0	0	0	0	0	0	0
R_0'		.333	.333	.333	.1667	.1667	.1667	.1000	.1000	.1000
r_1	.333	0	0	0						
	.25	.063	.065	.043						
	.167				0	0	0			
	.15	.160	.140	.089	.0052	.0066	.0043			
	.10	.229	.195	.123	.0268	.0226	.0123	0	0	0
	.08							.0043	.0048	.0025
	.05	.340	.292	.187	.0564	.0442	.0227	.0132	.0107	.0052
	.025				.083	.065	.034	.0240	.0183	.0090
	.01	.583	.557	.381	.116	.095	.051	.0369	.0287	.0148
	.005	.676	.688	.490	.141	.120	.069	.0460	.0373	.0199
	.000	1.000	1.777	2.777	.250	.444	.694	.0900	.1600	.2500
R_1'		.333	.333	.333	.1667	.1667	.1667	.1000	.1000	.1000

TABLE 4

Risk of the estimates (3.2) under hypothesis (2.2). Error of first kind, α, according to F. G. Grubbs [8]

n	β	α	r_0	r_1		
				$a = 3$	$a = 4$	$a = 5$
3	1.50	.0992	.085	.358	.254	.188
	1.75	.0481	.051	.464	.337	.218
	2.00	.0215	.028	.584	.460	.275
	2.25	.0088	.015	.704	.627	.374
	2.50	.0033	.006	.809	.829	.531
6	1.50	.2874	.0425	.050	.037	
	1.75	.1625	.0297	.066	.043	.0345
	2.00	.0847	.0189	.089	.055	.0367
	2.25	.0410	.0107	.113	.076	.0420
	2.50	.0185	.0057	.142	.100	.052
	2.75	.0078	.0031			.064
10	1.50	.4842	.0235	.0134		
	1.75	.3001	.0176	.0182	.0129	
	2.00	.1686	.0117	.0245	.0155	.0117
	2.25	.0871	.0074	.0331	.0194	.0125
	2.50	.0418	.0044	.0420	.0262	.0141
	2.75	.0187	.0021		.0360	.0167
	3.00	.0078	.0008			.0227

REFERENCES

[1] ANSCOMBE, F. J. (1960). Rejection of outliers. *Technometrics* **2** 123–147.

[2] DAVIS, P. and RABINOWITZ, P. (1956). Some Monte Carlo experiments in computing multiple integrals. *Math. Tables Aids Comput.* **10** 1–8.

[3] DE FINETTI, B. (1961). The Bayesian approach to the rejection of outliers. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 199–210.

[4] FERGUSON, T. E. (1961). On the rejection of outliers. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 253–287.

[5] GEBHARDT, F. (1961). *Bayeslösungen des Ausreisserproblems.* Dissertation, München.

[6] GEBHARDT, F. (1964). Generating normally distributed random numbers by inverting the normal distribution function. *Math. Comput.* **18** 302–306.

[7] GREENWOOD, J. A. and HARTLEY, H. O. (1962). *Guide to Tables in Mathematical Statistics.* Princeton Univ. Press.

[8] GRUBBS, F. E. (1950). Sample criteria for testing outlying observations. *Ann. Math. Statist.* **21** 27–58.

[9] HALPERIN, M., GREENHOUSE, S. W., CORNFIELD, J. and ZALOKAR, J. (1955). Tables of percentage points for the studentized maximum absolute deviate in normal samples. *J. Amer. Statist. Assoc.* **50** 185–195.

[10] KARLIN, S. and TRUAX, D. (1960). Slippage problems. *Ann. Math. Statist.* **31** 296–324.

[11] KUDO, A. (1956–57). On the testing of outlying observations. *Sankhyā* **17** 67–76.

[12] LEHMANN, E. L. (1961). Some model I problems of selection. *Ann. Math. Statist.* **32** 990–1012.

[13] PAULSON, E. (1952). An optimum solution to the k-sample slippage problem for the normal distribution. *Ann. Math. Statist.* **23** 610–616.

[14] PFANZAGL, J. (1959). Ein kombiniertes Test- und Klassifikationsproblem. *Metrika* **2** 11–45.

[15] QUESENBERRY, C. P., and DAVID, H. A. (1961). Some tests for outliers. *Biometrika* **48** 379–387.

[16] RAMACHANDRAN, K. V. and KHATRI, C. G. (1957). On a decision procedure based on the Tukey statistic. *Ann. Math. Statist.* **28** 802–806.

[17] RICHTER, H. (1960). Über optimale mehrstufige Tests. *Trans. 2nd Prague Conference on Information Theory, Statist. Decision Functions, Random Processes.* 557–568.

[18] TUKEY, J. W. (1962). The future of data analysis. *Ann. Math. Statist.* **33** 1–67.

[19] WALD, A. (1950). *Statistical Decision Functions.* Wiley, New York and London.