ON THE EFFECT OF STRAGGLERS ON THE RISK OF SOME MEAN ESTIMATORS IN SMALL SAMPLES[1]

By Friedrich Gebhardt

Deutsches Rechenzentrum, Darmstadt

1. Summary. In a previous paper [2], the risk (essentially, the variance) of certain mean estimators that in some way allow for a possible occurrence of stragglers (random variables with different mean or larger variance) have been numerically computed when the respective variances of stragglers and non-stragglers are given. This restriction is relaxed in the present paper; only the ratio of these variances is assumed to be known. A variety of cases as explained in detail in Sections 2 and 3 is being considered: The underlying distribution may be a normal (Gaussian), a logistic, or a generalized Cauchy distribution; there may be none, one, or two stragglers; the ratio of the standard deviations may be 3 or 6; finally, trimmed and Winsorized means and Bayes estimators are considered. The results are discussed in Section 5. They support the suggestion (J. W. Tukey [3]) to trim the sample by all observations that deviate substantially from the sample mean and to Winsorize those observations that deviate moderately, but trimming off exactly two observations is almost without reservations a strategy superior to always Winsorizing two. A very satisfactory behavior is exhibited by some estimators that formally are Bayes solutions for the Gaussian distribution but that also have been used with the other ones. This suggests to further pursue estimators with similar properties.

2. Examined distributions and definition of the risk. Let $X_1, \cdots, X_n$ be n independent random variables with common unknown mean m and standard deviations $\sigma_1, \cdots, \sigma_n$. Consider the following hypotheses on the variances (the notation "$A := B$" or "$B =: A$" defines A as B):

$$H_0: \sigma_1 = \cdots = \sigma_n =: \sigma,$$

$$H_1(i): \sigma_i = \sigma\tau, \quad \sigma_j = \sigma \quad \text{for} \quad j \neq i \qquad (i = 1, \cdots, n), \tag{2.1}$$

$$H_2(i, j): \sigma_i = \sigma_j = \sigma\tau, \quad \sigma_k = \sigma \quad \text{for} \quad k \neq i, j \quad (1 \leqq i < j \leqq n).$$

That is, one or two random variables may be stragglers with variances $\sigma^2\tau^2$ while all others have variance σ^2. Here τ is assumed to be known while σ is not. The restriction to no more than two stragglers has been made to keep the intended numerical computations within reasonable limits. 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 two stragglers may occur and it seemed more valuable to consider a larger variety of estimators and distributions instead.

Received 19 April 1965; revised 6 December 1965.

[1] Most of the work was done while the author was at the University of Connecticut. Supported in part by the National Science Foundation, Grant GP-1819.

Three types of distributions will be investigated:

(a) the normal distribution;

(b) the logistic distribution with cdf (for $m = 0, \sigma = 1$)

$$F(x) = 1/[1 + \exp\{-x/3^{\frac{1}{2}}\}],$$

(c) a generalized Cauchy distribution with density (again for $m = 0, \sigma = 1$)

$$f(x) = (2^{\frac{1}{2}}/\pi)(1 + x^4)^{-1}.$$

While for large x the logistic density converges to zero exponentially, the generalized Cauchy density is of order x^{-4} thus having moments only of order <3.

Let $\bar{X} := \sum X_i/n$, $S^2 := \sum (X_i - \bar{X})^2/(n-1)$, $T_i := (X_i - \bar{X})/S$, $i = 1, \cdots, n$. Furthermore, let $\bar{X} - a(\bar{X}, S, T_1, \cdots, T_n)$ be any estimator for the mean m. We shall replace in such expressions the sequence $T_1, \cdots, T_n$ just by T. The joint density functions of $\bar{X}, S, T_1, \cdots, T_{n-2}$ under Hypotheses $H_0, H_1(1), H_2(1, 2)$ will be denoted by $f_j(\bar{x}, s, t; m, \sigma, \tau), j = 0, 1, 2$ respectively. Any function of $t_1, \cdots, t_n$ is to be considered as a function of $t_1, \cdots, t_{n-2}$ only by means of $\sum_{i=1}^n t_i = 0$, $\sum_{i=1}^n t_i^2 = n - 1$; using all n arguments better exhibits the symmetry. Since the densities of X_i are of the general form $\sigma_i^{-1} f^*((x_i - m)/\sigma_i)$, it is easily verified that f_j can be converted into $f_j(\bar{x}, s, t; m, \sigma, \tau) = \sigma^{-2} g_j((\bar{x} - m)/\sigma, s/\sigma, t; \tau)$, the factor σ^{-2} arising from the Jacobian $|\partial(x_1, \cdots, x_n)/\partial(\bar{x}, s, t_1, \cdots, t_{n-2})| = n(n-1)s^{n-2}/|t_n - t_{n-1}|$ combined with the factor σ^{-n} stemming from the product of the densities of $X_1, \cdots, X_n$.

Now we introduce the quadratic loss

$$L(H_j, a; m) := (\bar{x} - a(\bar{x}, s, t) - m)^2 \sigma^{-2}, \tag{2.2}$$

if $\bar{x} - a$ is the estimated, and m is the true mean. Then the risks of the estimator $\bar{X} - a(\bar{X}, S, T)$ under hypotheses $H_0, H_1(1), H_2(1, 2)$ respectively are

$$\begin{aligned} R_j &= E_j[(\bar{X} - a(\bar{X}, S, T) - m)^2 \sigma^{-2}] \\ &= \int (\bar{x} - a(\bar{x}, s, t) - m)^2 \sigma^{-2} f_j(\bar{x}, s, t; m, \sigma, \tau)\, d\bar{x}\, ds\, dt_1 \cdots dt_{n-2}. \end{aligned}$$

The risk under hypothesis $H_1(i)$ for $i \neq 1$ is defined correspondingly; but as long as a is symmetric in $t_1, \cdots, t_n$, it is equal to R_1; the same holds for $H_2(i, k)$.

In general, R_j is a function not only of τ, but also of the unknown parameters m and σ. If however the estimator is of the form $a(\bar{X}, S, T) = Sb(T)$, a transformation $u = (\bar{x} - m)/\sigma$, $v = s/\sigma$ shows immediately that in this case R_j does not depend on m and σ. The estimators that usually are employed are of this form; we shall also investigate only such estimators. They have the special property of reflecting some symmetries of the problem: A linear transformation $X_i \to \alpha X_i + \beta$ induces the same transformation of the estimator.

One might argue about the justification of the denominator σ^2 in the definition of the loss. Its purpose is to measure the deviation of the estimator from the true value m in units of the standard deviation. As long as σ is regarded as an unknown but fixed constant, the denominator σ^2 is just a multiplicative constant and of no further importance since it does not change any relations between the

risks of different estimators; it becomes essential only in finding the Bayes solutions if estimators of the general form $\bar{X} - a(\bar{X}, S, T)$ are admitted since the risk is to be minimized with respect to a prior distribution not only of the hypotheses H_j but also of the dummy parameter σ.

If the variances of all variables are known (that is, if one knows which variables are stragglers) and if the variables are normally distributed, the minimum variance, unbiased estimator is $\bar{X} - a_0(X) = \sum (X_i/\sigma_i^2)/\sum \sigma_i^{-2}$. Especially, if $\sigma_1 = \cdots = \sigma_j = \sigma\tau$, $\sigma_{j+1} = \cdots = \sigma_n = \sigma$, then a_0 becomes

$$\begin{aligned} a_0 &= (1 - \tau^{-2})(n - j + j\tau^{-2})^{-1} \sum_{i=1}^{j} (X_i - \bar{X}) \\ &= (1 - \tau^{-2})(n - j + j\tau^{-2})^{-1} \sum_{i=1}^{j} ST_i . \end{aligned}$$

The risk of this estimator under the correct hypothesis is

$$R_j' = \sigma^{-2}/\sum \sigma_i^{-2} = 1/(n - j + j\tau^{-2}). \tag{2.3}$$

For any probability distribution with existing var (X) other than the normal distribution, (2.3) still gives the risk of $\bar{X} - a_0$ although this may no longer be the minimum variance unbiased estimator.

Merely for convenience, we introduce

$$r_j(b) := R_j(b) - R_j' \tag{2.4}$$

and we now call this the risk of the estimator $\bar{X} - Sb(T)$. Of course, this additive constant R_j' does not affect the comparison of different estimators or distributions with one another. The numerical values of R_j' as far as they concern us here, are

$$\begin{aligned} n = 6, \quad \tau = 6: \quad & R_0' = 0.1667, \quad R_1' = 0.1989, \quad R_2' = 0.2466; \\ n = 6, \quad \tau = 3: \quad & R_0' = 0.1667, \quad R_1' = 0.1957, \quad R_2' = 0.2368; \\ n = 10, \quad \tau = 6: \quad & R_0' = 0.1000, \quad R_1' = 0.1108, \quad R_2' = 0.1241; \\ n = 10, \quad \tau = 3: \quad & R_0' = 0.1000, \quad R_1' = 0.1098, \quad R_2' = 0.1216. \end{aligned}$$

3. Mean estimators. Next we shall describe the estimators investigated. The symbols that will be introduced refer to those that are used in the Diagrams.

T. All variables X_i for which $|T_i| > \beta$ are to be trimmed off; that is, the mean estimator is

$$\bar{X} - Sb_T = \sum{}^* X_i/\sum{}^* 1 = \bar{X} + S\sum{}^* T_i/\sum{}^* 1$$

where $\sum^*$ stands for the summation with respect to all values of i such that $|T_i| < \beta$. This estimator contains a parameter, β; therefore its risk appears in the diagrams as a line rather than a single point. If $\beta \geqq ((n - 1)/2)^{\frac{1}{2}}$, at most one $|T_i|$ can exceed β and b_T is equal to b_{T1} discussed below.

$T1$. If $\max |T_i| > \beta$, then the corresponding X_i is to be trimmed off. If $\beta > (n - 1)n^{-\frac{1}{2}}$, no such variable can exist; if $\beta < \beta_0 := ((n - 1)/n)^{\frac{1}{2}}$, there is always at least one T_i with $|T_i| > \beta$.

$T2$. The largest variable and the smallest one are always to be trimmed off without regard to their sizes. In the diagrams, $T2$ is marked by a cross.

$W1$. If $\max |T_i| = |T_j| > \beta$, then X_j is to be Winsorized, that is, it is to be replaced by the nearest variable that is not Winsorized. The mean of these altered n values is to be taken as an estimator for m. The curves $W1$ have definite left ends that correspond to $\beta = \beta_0$; in this case always exactly one observation is Winsorized.

$W2$. The largest observation and the smallest one are to be Winsorized without regard to their sizes. The risk of $W2$ is marked in the diagrams by a small circle.

$B1$. If the variables X_i are normally distributed, b_{B1} is the Bayes solution to the prior probability ϕ_0 of no straggler occurring and the prior probability ϕ_1 of X_i $(i = 1, \cdots, n)$ being the only one, $\phi_0 + n\phi_1 = 1$. Thus, the occurrence of more than one straggler has prior probability 0. The formula for b_{B1} as well as that for b_{B2} are derived in Section 4. The same estimator has been used with the other distributions, although it then is no longer a Bayesian estimator.

$B2$. For normally distributed variables X_i , this is again a Bayes estimator when $\phi_2 = \phi_1^2$ is the prior probability that for any pair i, j $(i < j)$, X_i and X_j are the only stragglers; $\phi_0 + n\phi_1 + n(n - 1)\phi_2/2 = 1$.

$B1'$, $B2'$. These estimators that have been used only with normal distributions and $n = 6$, are the same as $B1$ and $B2$, but the wrong value of τ has been inserted ($\tau = 3$ when the true parameter value is 6 and vice versa).

As already noted in the Summary, the risks have been computed for $\tau = 3$ and $\tau = 6$, for $n = 6$ and $n = 10$. The method used was in part a Monte Carlo integration, mostly a modification proposed by P. Davis and P. Rabinowitz [1], see also [2], that proved slightly more accurate. Both methods utilized 6000 or 10000 samples for $n = 10$ and $n = 6$ respectively; the standard deviations of Monte Carlo integrals seem to be 2 to 3 times those of the modified method.

4. Bayesian estimators. For a finite number of simple hypotheses H_j with prior probabilities ϕ_j , the Bayes estimator $\bar{X} - a_B(\bar{X}, S, T)$ is by definition that function that minimizes the weighted sum of risks

$$\sum_j \phi_j E_j[L(H_j , a)]. \tag{4.1}$$

Since however our hypotheses are composite, ϕ_j has to be replaced by its product with $\psi_j(m, \sigma)$, the conditional prior probability density of m and σ given H_j ; instead of (4.1) we now have to minimize

$$\sum_j \int d\sigma \int dm \phi_j \psi_j(m, \sigma) E_j[L(H_j , a; m, \sigma)]. \tag{4.2}$$

In order to simplify this expression, we shall impose the following restriction on the estimators: Only estimators of the form $\bar{X} - Sb(T)$ are admitted.

Then the expectation in (4.2) becomes

$$\begin{aligned} E_j[L(H_i , a; m, \sigma)] &= \int d\bar{x} \int ds \int dt (\bar{x} - sb(t) - m)^2 \sigma^{-2} \cdot \sigma^{-2} g_j((\bar{x} - m)/\sigma, s/\sigma, t; \tau) \\ &= \int du \int dv \int dt (u - vb(t))^2 g_j(u, v, t; \tau) \end{aligned}$$

with $u := (\bar{x} - m)/\sigma$, $v := s/\sigma$. This expression is independent of m and σ; therefore (4.2) can be integrated with respect to m and σ resulting in the omission of $\int \psi_i(m, \sigma)\, dm\, d\sigma$ that is equal to one. Thus we have to find a function $b_B(t)$ that minimizes

$$\sum_i \phi_i \int du \int dv \int dt (u - vb(t))^2 g_i(u, v, t; \tau). \tag{4.3}$$

The integration with respect to u and v can be performed explicitly and then the values of b minimizing the sum for fixed t can be found by differentiation. The final result is given below; it is a measurable function of t.

In analogy to results in [2] one might state the following conjecture: The minimum risk of all estimators $\bar{X} - a(\bar{X}, S, T)$ with respect to uniform prior density (in the limit) of m and $\log \sigma$ will also be attained by an estimator $\bar{X} - Sb_B(T)$.

Of course, the expression "in the limit" needs some further specification; possibly some weak regularity conditions must be added. It is hoped to give a more general proof in a subsequent paper.

If each hypothesis $H_1(i)$ stating that X_i is the only straggler ($i = 1, \cdots, n$) has prior probability ϕ_1 and each hypothesis $H_2(i, j)$ assuming X_i and X_j ($1 \leqq i < j \leqq n$) to be stragglers has prior probability ϕ_2, then the solution of (4.3) is

$$b_{B2}(T) = \frac{\phi_1 \sum T_i[1 - C_1 T_i^2]^{-(n+1)/2} + \phi_2 C_4 V_1}{\phi_0 C_2 + \phi_1 C_3 \sum [1 - C_1 T_i^2]^{-(n+1)/2} + \phi_2 C_5 V_2} \tag{4.4}$$

with

$$K = 1 - \tau^{-2}, \qquad L = n - 1 + \tau^{-2}, \qquad M = n - 2 + 2\tau^{-2},$$

$$C_1 = nK/(n-1)L, \qquad C_2 = \tau L^{3/2}/Kn^{\frac{1}{2}},$$

$$C_3 = L/K, \qquad C_4 = L/\tau M,$$

$$C_5 = L^{3/2}/\tau K M^{\frac{1}{2}}, \qquad C_6 = K/(n-1),$$

$$C_7 = K^2/(n-1)M,$$

$$V_1 = \sum_{i<j} (T_i + T_j)[1 - (T_i^2 + T_j^2)C_6 - (T_i + T_j)^2 C_7]^{-(n+1)/2},$$

$$V_2 = \sum_{i<j} [1 - (T_i^2 + T_j^2)C_6 - (T_i + T_j)^2 C_7]^{-(n+1)/2}.$$

Note the similarity of numerator and denominator; characteristically the terms of the denominator are multiplied by a linear expression in T to obtain the corresponding term of the numerator (this even holds for the first term $\phi_0 C_2$: multiplying it analogously by $\sum T_i$ yields 0).

From (4.4), b_{B1} is obtained by letting $\phi_2 = 0$; then the last terms of numerator and denominator vanish.

The prior probabilities must of course obey the equality $\phi_0 + n\phi_1 + n(n-1)\phi_2/2 = 1$.

5. Discussion of the results. The risks of the estimators to be considered are plotted in 12 diagrams. Note that r_2 has a smaller scale than r_0 and r_1. Diagrams for different distributions but equal n and τ have the same scale. Due to the

rather low accuracy of computation, some minor details in the diagrams may not be correct.

In the following discussion, estimators will be referred to by the letters used in the diagrams; e.g. we shall say $T1$ meaning $\bar{X} - Sb_{T1}(T)$. Statements like "$T2$ is better than $W2$" are to be taken liberally; to state the exact meaning of words like "better" in each case would only obscure the results while a glance at the diagrams will explain the situation.

As might be expected, the relative merits of different estimators change with sample size, type of distribution, and quotient of the variances. Nevertheless some relations seem to hold rather generally.

The easiest comparison is that between $T2$ and $W2$. In all cases, $T2$ has smaller risks under hypotheses H_1 and H_2 (sometimes they are considerably smaller) while under H_0 the risks of $T2$ and $W2$ are about equal. Since $T2$ in most cases also compares rather favorably with all other estimators and is very easily computed, it might justly be preferred in many situations.

The left ends of $W1$ correspond to the strategy of always Winsorizing one observation. If the parameter β exceeds β_0 (as introduced in Section 3) only moderately, the risks remain almost unaffected. This is however not surprising since $W1(\beta)$ and $W1(\beta_0)$ themselves are almost equal in this case. For $T1$, the situation is different. The curves again have definite ends corresponding to $\beta = \beta_0$. But if $\beta_0 < \max |T_i| < \beta$, the largest (or smallest) observation is in one case discarded; this causes a considerable difference in the risks of $T1(\beta)$ and $T1(\beta_0)$. $r_0(T1(\beta_0))$ is rather large and falls outside the range of the diagrams (except possibly for the generalized Cauchy distribution; the risks of $T1$ have not been computed for $\beta < 1.5$). But also $r_1(T1)$ increases for small values of β and, it seems, so does $r_2(T1)$. Thus, while always Winsorizing one observation still is a rather reasonable procedure, always trimming off one is not. If one decides on trimming off at most one observation, recommended values of β are about 1.7 for $n = 6$ and 2.0 to 2.1 for $n = 10$.

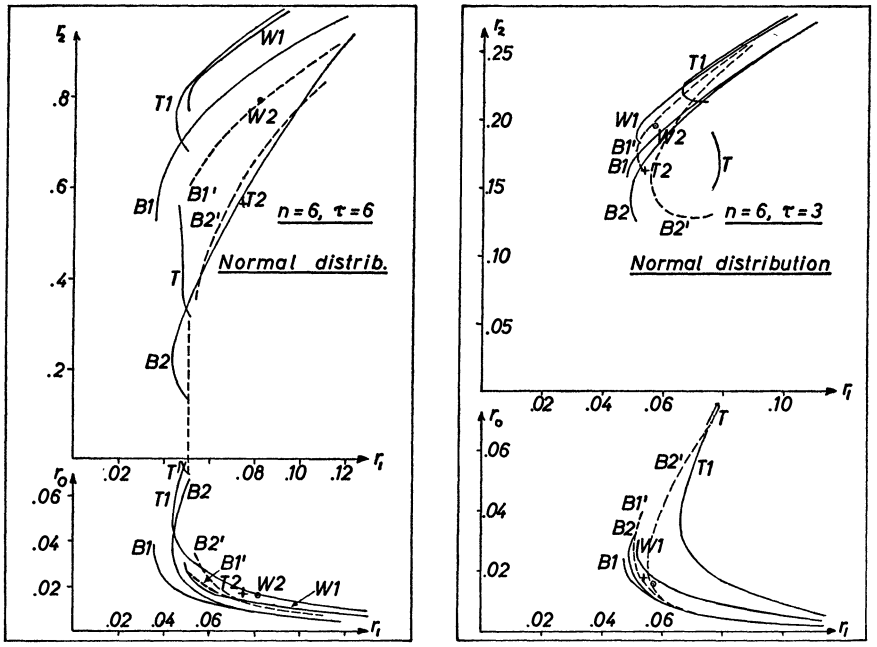
Comparing $W1$ with $T1$, it appears that $W1$ is to be preferred for $\tau = 3$, $T1$ for $\tau = 6$. This supports Tukey's proposal [3] to trim off observations that deviate substantially from the sample mean and to Winsorize those that deviate moderately.

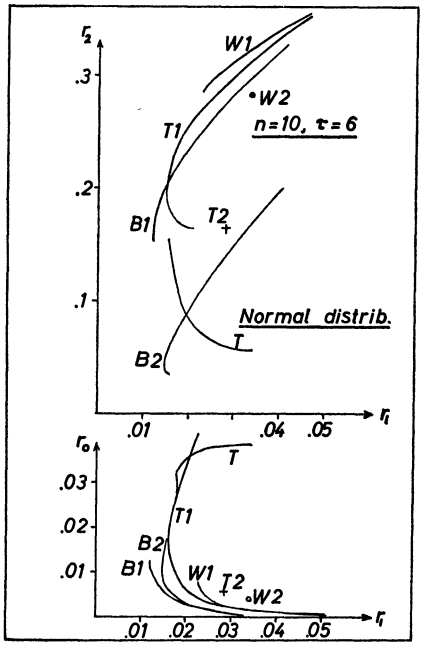
If $\beta > \beta_1 = ((n-1)/2)^{\frac{1}{2}}$, at most one observation can exceed β. For $\beta < \beta_1$, there may be more, and $T(\beta)$ is the strategy of trimming off all those. Obviously, $T(\beta)$ coincides with $T1(\beta)$ for $\beta > \beta_1$. The risks of T have been computed for $1.2 \leqq \beta \leqq 1.4$, if $n = 6$, and for $1.1 \leqq \beta \leqq 1.7$, if $n = 10$. It turns out that $r_0(T)$ is rather large in comparison with other estimators while $r_2(T)$ is satisfactorily small. The other features of T vary substantially with the underlying distribution. Due to the large risk $r_0(T)$ it should not be used unless a high straggler rate is expected.

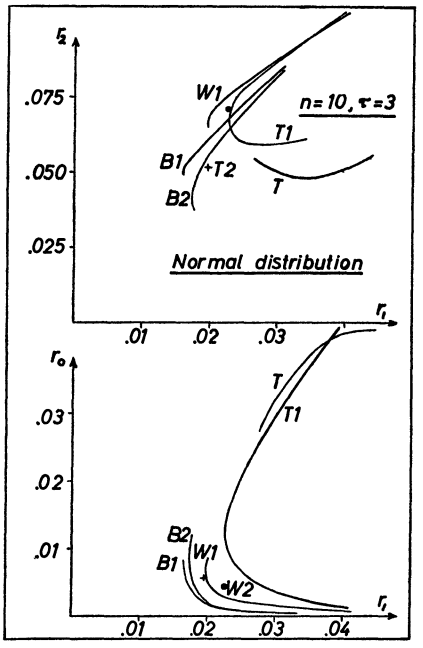
For normal distributions, the Bayes solution must of course be a reasonable estimator since it minimizes by definition the weighted sum of risks. But it came as a surprise to the author that this complicated, specialized estimator also works very well for other distributions (where it is no longer the Bayes

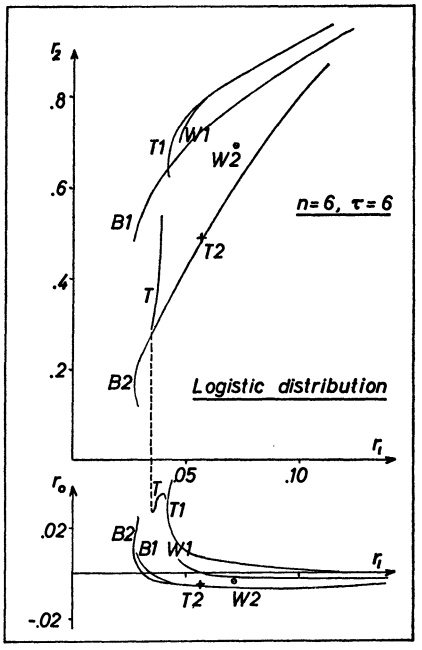
solution). While usually $B1$ has a somewhat smaller risk than $B2$ under hypotheses H_0 and H_1, the difference is not large; for the generalized Cauchy distribution, $r_0(B2)$ is even smaller than $r_0(B1)$. Under Hypothesis H_2, $B2$ has always a considerably smaller risk than $B1$, especially for $\tau = 6$. Whenever the tedious work of computing $B1$ or $B2$ is justified (generally this will imply using an electronic computer), it is highly recommended to choose $B2$ with not too small a value of ϕ (about .10 for $n = 6$ and about .05 for $n = 10$). This may be of advantage even if one is sure that there are no stragglers but if the distribution is non-normal. It may also be worthwhile to look for new estimators that are simpler than $B2$ but show a similar behavior.

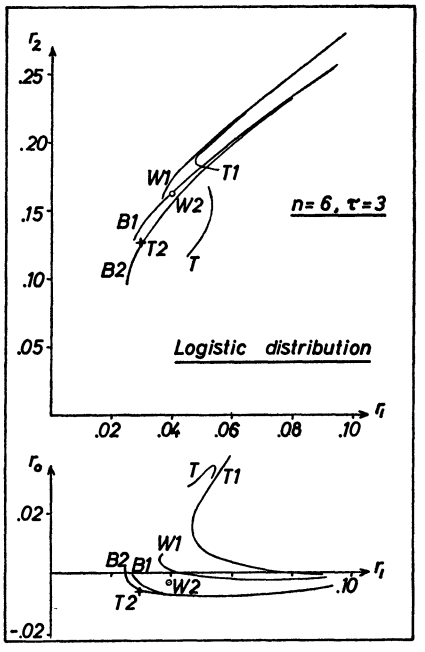
There is one drawback, however, with the Bayesian estimators. To employ them, one must know τ as has been assumed throughout this paper. All other estimators considered here do not contain τ. Thus they can be used even if τ is not known as will mostly be the case; nevertheless one knows something about their risks. To study the effect of a wrong choice of τ, the Bayes estimators have been computed for underlying normal distribution and $n = 6$ using $\tau = 3$ in the formulae for b_{B1} and b_{B2} when the true parameter is $\tau = 6$ and vice versa. The results are shown in the corresponding diagrams as broken curves (not to be confused with the broken straight lines that in some cases connect corresponding points in the upper and lower parts of a diagram). It seems that choosing too small a value of τ decreases r_0 and increases r_2 while too large a value of τ increases r_1 and r_2; $B2$ is much more affected than $B1$. It is not known if these results can be generalized to $n = 10$ or to other distributions.

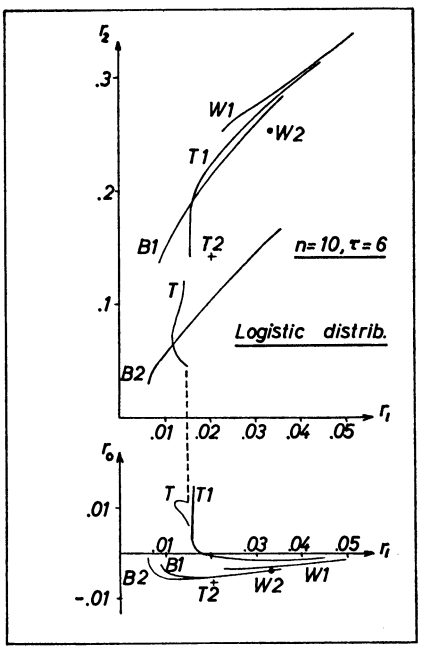


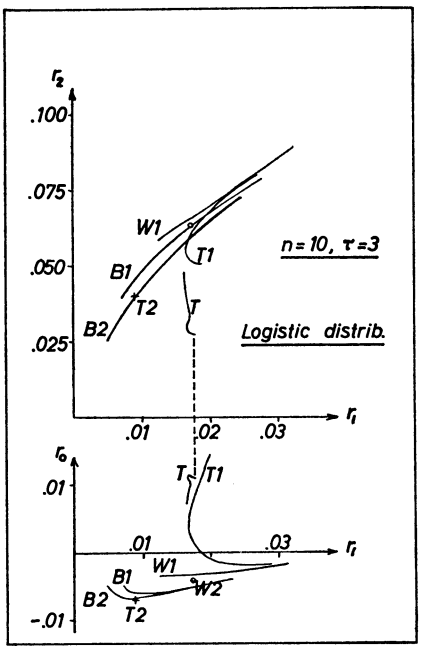

r2
.3
.2
.1
W1
W2
T1
n=10, τ=6
B1
T2
Normal distrib.
T
B2
.01
.04
.05
r1
ro
.03
.02
.01
T
T1
B2
B1
W1
T2
W2
.01
.02
.03
.04
.05
r1

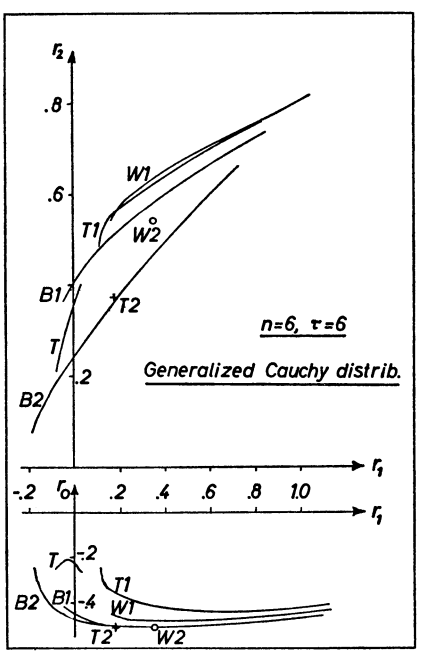

r2
.075
.050
.025
W1
n=10, τ=3
T1
B1
T2
T
B2
Normal distribution
.01
.02
.03
r1
ro
.03
.02
.01
T
T1
B2
B1
W1
W2
.01
.02
.03
.04
r1

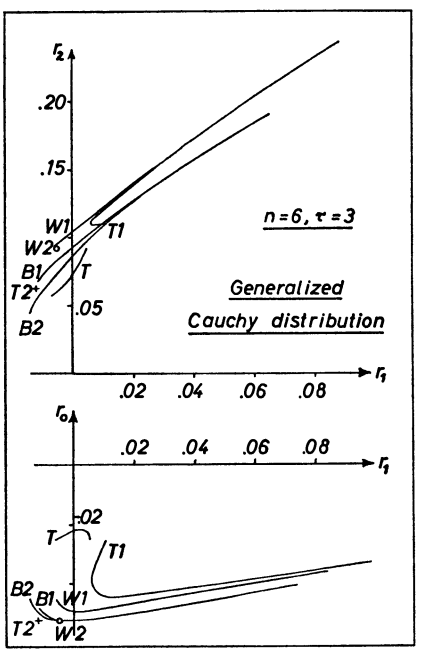

r2
.8
.6
.4
.2
T1
W1
W2
n=6, τ=6
B1
T2
T
B2
Logistic distribution
.05
.10
r1
ro
.02
-.02
T
T1
B2
B1
W1
T2
W2
r1

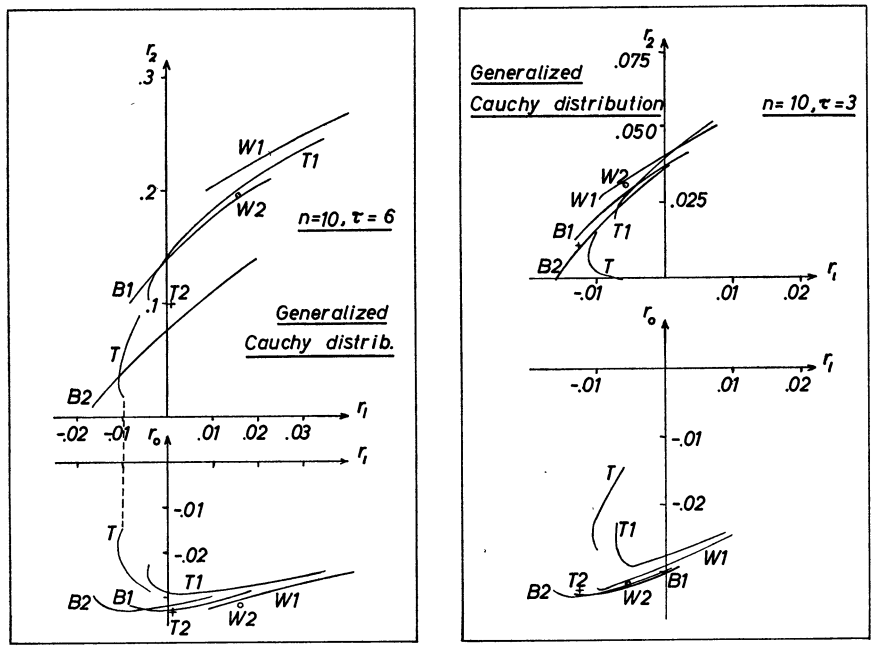

r2
.25
.20
.15
.10
.05
W1
T1
W2
n=6, τ=3
B1
T2
B2
T
Logistic distribution
.02
.04
.06
.08
.10
r1
ro
.02
-.02
T
T1
W1
B2
B1
W2
T2
.10
r1


r2
.3
.2
.1
W1
W2
T1
B1
T2
T
B2
n=10, τ=6
Logistic distrib.
.01 .02 .03 .04 .05
r1
r0
.01
-.01
T
T1
.01 .03 .04 .05
B1
B2
T2
W2
W1


r2
.100
.075
.050
.025
W1
T1
B1
T2
T
B2
n=10, τ=3
Logistic distrib.
.01 .02 .03
r1
r0
.01
-.01
T
T1
.01 .03
B1
W1
W2
B2
T2


r2
.8
.6
W1
T1
W2
B1
T2
T
.2
B2
n=6, τ=6
Generalized Cauchy distrib.
-.2
r0
.2 .4 .6 .8 1.0
r1
T
-.2
T1
B2
B1
-.4
W1
T2
W2


r2
.20
.15
W1
T1
W2
B1
T
T2
.05
B2
n=6, τ=3
Generalized
Cauchy distribution
.02 .04 .06 .08
r1
r0
.02 .04 .06 .08
r1
-.02
T
T1
B2
B1
W1
T2
W2



6. Diagrams. Risks r_0, r_1, and r_2 under hypotheses H_0, H_1, and H_2, resp., of several estimators introduced in Section 3. The hypotheses are defined in (2.1), the risks, in (2.4). The risks of $T2$ and $W2$ are marked by a cross and a circle, resp. Broken curves indicate risks of $B1'$ and $B2'$. Broken straight lines connect corresponding points in upper and lower parts of the diagrams.

The diagrams do not indicate which points of the curves correspond to which values of the parameters ϕ and β respectively. Since the tables with the numerical results are rather extensive, they are not published here but may be obtained from the author on request.

7. Acknowledgment. The author wishes to thank Prof. J. Löf (Computer Center, University of Connecticut) for the permission to use extensively the IBM 7040 computer.

REFERENCES

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

[2] GEBHARDT, F. (1964). On the risk of some strategies for outlying observations. *Ann. Math. Statist.* **35** 1524–1536.

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

[4] TUKEY, J. W. and MCLAUGHLIN, D. H. (1963). Less vulnerable confidence and significance procedures for location based on a single sample: Trimming/Winsorization. *Sankhyā Ser. A* **25** 331–352.