A GENERALIZATION OF A PRELIMINARY TESTING PROCEDURE FOR POOLING DATA[1, 2, 3]

By D. V. Huntsberger

Iowa State College

1. Summary. This paper is concerned with a generalization of the sometimes-pool procedure for pooling two estimators which is based on a preliminary test of significance. A weighted estimator for one of the parameters is obtained by using weights which are determined by the observed value of the preliminary test statistic. The efficiencies of the weighting procedure and of the sometimes-pool procedure are compared for the special case where the estimators are normally distributed. Further, it is shown that the weighting procedure offers a greater degree of control over the disturbances which may result from pooling than does the sometimes-pool procedure. Some problems concerning the choice of a weighting function are discussed.

2. Introduction. The effects of preliminary tests of significance on subsequent statistical inferences have been studied in various special cases by Bancroft [1], [2], [3], Bechhofer [4], Mosteller [8], and Paull [9]. They found that the use of such tests introduces serious disturbances into the final inferences. These disturbances take the forms of biased estimates, losses of efficiency as regards estimation, or shifts in the sizes and powers of tests of hypotheses.

Preliminary testing procedures may be characterized as follows: A statistic, T, is evaluated from the data at hand. If T is not significant at some preassigned level of significance, a given procedure is used to estimate the parameter in question or to test the major hypothesis. If T is significant, an alternative procedure is used for obtaining estimates or for testing the hypothesis. In any event, the only information derived from T is that it does or does not fall into the region of rejection. If more of the information contained in T is utilized, it is possible to exert more control over the disturbances inherent in the preliminary testing procedures than is possible by merely altering the level of significance of the preliminary test.

Let $X_1, \cdots, X_n$ be a random sample with joint probability density function

$$f(X_1, \cdots, X_n; \theta_1, \cdots, \theta_k),$$

where the functional form is known, θ_1 and θ_2 are unknown parameters, and the last $k - 2$ θ's are parameters whose values may or may not be known. Let $\hat{\theta}_1$

Received December 14, 1954.

[1] Some of the material in this paper was presented to the joint meeting of the Institute of Mathematical Statistics, the American Statistical Association, and the Biometric Society at Washington, D. C., December 29, 1953.

[2] Based on research sponsored in part by Wright Air Development Center.

[3] These results are contained in a thesis submitted to the graduate faculty of Iowa State College in partial fulfillment of the requirements for the Ph.D. degree in Statistics.

and $\hat{\theta}_2$ be the best estimators of θ_1 and θ_2 as provided by statistical theory. If $\theta_1 = \theta_2$, a pooled estimator $g(\hat{\theta}_1, \hat{\theta}_2)$ will, in general, provide an estimator for θ_1 which is better in some sense than $\hat{\theta}_1$. When it is not known whether or not θ_1 and θ_2 are equal, a better estimate for θ_1 may still be obtained by making use of any information provided by $\hat{\theta}_2$.

Let T be the statistic which the theory indicates will provide the best test of the hypothesis that $\theta_1 = \theta_2$ against the class of alternatives $\theta_1 \neq \theta_2$. Evaluate T using the data at hand, and for an estimator of θ_1 use the function

$$W(T) = \phi(T)\hat{\theta}_1 + [1 - \phi(T)]g(\hat{\theta}_1, \hat{\theta}_2), \tag{2.1}$$

where $\phi(T)$ is a function of T only. If $\phi(T)$ is defined as

$$\begin{aligned} \phi(T) &= 0, \qquad T \subset A_\alpha \\ \phi(T) &= 1, \qquad T \subset R_\alpha \end{aligned} \tag{2.2}$$

where A_α and R_α are the acceptance and rejection regions for the test of H_0 with probability of type I error equal to α, then $W(T)$ reduces to the estimator, $SP(T)$, following from the "sometimes-pool" procedure based upon the preliminary test of significance.

In order to determine whether or not $W(T)$ offers any advantages over $SP(T)$ or $\hat{\theta}_1$ as an estimator for θ_1, the mean square deviation, D^2, about the true parameter value is used as a criterion of goodness.

3. Pooling normal estimators. Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two independent, unbiased, normally distributed estimators for θ_1 and θ_2 respectively. Let $\hat{\theta}_1$ and $\hat{\theta}_2$ have known variances, σ_1^2 and σ_2^2, respectively. A pooled estimator for θ_1 is obtained in this special case as

$$W(T) = \phi(T)\hat{\theta}_1 + [1 - \phi(T)]\,\frac{\sigma_2^2\hat{\theta}_1 + \sigma_1^2\hat{\theta}_2}{\sigma_1^2 + \sigma_2^2}, \tag{3.1}$$

where $\phi(T)$ is a function of T only and

$$T = \frac{\hat{\theta}_1 - \hat{\theta}_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \tag{3.2}$$

is normally distributed with mean

$$\gamma = \frac{\theta_1 - \theta_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \tag{3.3}$$

and variance one. Equation (3.1) may be written as

$$W(T) = \frac{\sigma_2^2\hat{\theta}_1 + \sigma_1^2\hat{\theta}_2}{\sigma_1^2 + \sigma_2^2} + \frac{\sigma_1^2\, T\phi(T)}{\sqrt{\sigma_1^2 + \sigma_2^2}}. \tag{3.4}$$

The mean square deviation of $W(T)$ about θ_1 is a function of the nuisance parameter γ and is given by

$$D_W^2(\gamma) = E\,[W(T) - \theta_1]^2,$$

$$D_W^2(\gamma) = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}\left\{\sigma_2^2 + \sigma_1^2 \int_{-\infty}^{\infty} [T\phi(T) - \gamma]^2 N(T - \gamma)\, dT\right\}, \tag{3.5}$$

where

$$N(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}.$$

The bias of $W(T)$ as an estimator for θ_1, is

$$B_W(\gamma) = \theta_1 - E[W(T)] = \frac{\sigma_1^2}{\sqrt{\sigma_1^2 + \sigma_2^2}}\left[\gamma - \int_{-\infty}^{\infty} T\phi(T)N(T - \gamma)\, dT\right] \tag{3.6}$$

4. Weighting functions. In order to use the estimator given by the weighting procedure (3.4), the weighting function $\phi(T)$ must first be selected. The choice of $\phi(T)$ will be restricted to the class of single-valued functions of T which are continuous except on a set of measure zero, which are defined for all T, and which satisfy the conditions:

(i) $0 \leqq \phi(T) \leqq 1$, for all T,

(ii) $\phi(-T) = \phi(T)$.

The class of functions so defined will be referred to as the admissible class of weighting functions.

The choice of a weighting function should be based on some criterion by means of which the relative merits of various alternative functions may be assessed. A possible criterion is unbiasedness; that is, if a function $\phi_u(T)$ exists such that for all γ

$$E[W_u(T)] = \theta_1$$

then $\phi_u(T)$ is an unbiased weighting function.

THEOREM 1. *Among the class of admissible weighting functions the only unbiased weighting function is* $\phi(T) = 1$.

PROOF. Because of the symmetry of $\phi(T)$ the bias of $W(T)$, equation (3.6), may be put into the form

$$B_W(\gamma) = \frac{\sigma_1^2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \int_0^{\infty} T[1 - \phi(T)][N(T - \gamma) - N(T + \gamma)]\, dT.$$

It is obvious that this is equal to zero when γ is not equal to zero if and only if $\phi(T)$ is identically equal to one.

A second desirable property would be uniformly minimum mean square error about θ_1. If $\phi_v(T)$ is an admissible weighting function such that $D_{\phi_v}^2(\gamma) \leqq D_{\phi'}^2(\gamma)$ for every ϕ' and every γ, with inequality holding for at least one γ and one ϕ', then $\phi_v(T)$ is a uniformly minimum mean square error weighting function. It will be shown in Theorem 3 that such a function does not exist.

A third criterion which might be proposed is one which selects a weighting function which yields an estimator whose efficiency is greatest when averaged in some sense for all γ. One such measure of overall efficiency is the area between the curves corresponding to $D_W^2(\gamma)$ and to the variance, σ_1^2, of the "neverpool" estimator, $\hat{\theta}_1$. The following theorem is obvious and is stated without proof.

THEOREM 2. *Among the class of admissible weighting functions, that one which maximizes the integral*

$$(4.1) \qquad I = \int_{-\infty}^{\infty} [\sigma_1^2 - D_W^2(\gamma)]\, d\gamma$$

is $\phi(T) = 1$.

As a consequence of Theorem 2 and the fact that if γ is zero the minimum variance estimator is obtained by letting $\phi(T) = 0$ the following theorem may be stated.

THEOREM 3. *Among the class of admissible weighting functions there exists no function* $\phi_v(T)$ *such that*

$$D_{\phi_v}^2(\gamma) \leqq D_{\phi'}^2(\gamma)$$

for every ϕ' *and every* γ.

Suppose that γ is fixed and we consider an estimate

$$A\hat{\theta}_1 + (1 - A)\,\frac{\sigma_2^2\hat{\theta}_1 + \sigma_1^2\hat{\theta}_2}{\sigma_1^2 + \sigma_2^2},$$

where A can be a function of γ. Its mean square deviation about θ_1 is

$$\frac{\sigma_1^4}{\sigma_1^2 + \sigma_2^2}\left[\frac{\sigma_2^2}{\sigma_1^2} + A^2 + (1 - A)^2\gamma^2\right],$$

and this is minimized with respect to A when $A = \gamma^2/(1 + \gamma^2)$.

Since γ is considered to be unknown and since T is an unbiased estimator for γ, it was decided to estimate A by

$$(4.2) \qquad \phi_0(T) = \frac{T^2}{1 + T^2}.$$

5. Mean square and bias when $\phi(T) = \phi_0(T)$. If $\phi_0(T) = T^2/(1 + T^2)$ is substituted for $\phi(T)$ in the expression (3.5) for the mean square deviation,

$$(5.1) \qquad D_{W_0}^2(\gamma) = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}\left\{\sigma_2^2 + \sigma_1^2 \int_{-\infty}^{\infty}\left[\gamma - \frac{T^3}{1 + T^2}\right]^2 N(T - \gamma)\, dT\right\}.$$

The integral in (5.1) can be put into the form

$$(5.2) \qquad I(\gamma) = 1 + 3H(\gamma) - 5G(\gamma),$$

where

$$H(\gamma) = \int_{-\infty}^{\infty} \frac{1}{1 + T^2} N(T - \gamma)\, dT, \tag{5.3}$$

$$G(\gamma) = \int_{-\infty}^{\infty} \frac{1}{(1 + T^2)^2} N(T - \gamma)\, dT. \tag{5.4}$$

A series solution for the integral $H(\gamma)$ was found by using a method given by Zemansky [10].

$$H(\gamma) = \sum_{n=0}^{\infty} \frac{e^{-\gamma^2/2}}{n!} \left(\frac{\gamma^2}{2}\right)^n A_n, \tag{5.5}$$

$$A_n = e^{1/2} \int_1^{\infty} \frac{e^{-z^2/2}}{z^{2n}}\, dz,$$

$$A_0 = e^{1/2} \int_1^{\infty} e^{-z^2/2}\, dz,$$

$$A_n = \frac{1}{2n - 1} (1 - A_{n-1}).$$

Since

$$G(\gamma) = [1/2] \left[1 - \gamma \frac{dH(\gamma)}{d\gamma} - \gamma^2 H(\gamma)\right],$$

a series solution for $G(\gamma)$ is obtained from that of $H(\gamma)$, (5.5).

$$G(\gamma) = \frac{e^{-\gamma^2/2}}{2} \left[1 + \sum_{n=1}^{\infty} \frac{\gamma^{2n}}{2^n n!} (A_{n-1} - A_n)\right], \tag{5.6}$$

and substitution into (5.1) gives

$$D^2_{W_0}(\gamma) = \sigma_1^2 + \frac{\sigma_1^4 e^{-\gamma^2/2}}{\sigma_1^2 + \sigma_2^2} \left[3A_0 - \frac{5}{2} + \sum_{n=1}^{\infty} \frac{\gamma^{2n}}{2^n n!} \left(\frac{11}{2} A_n - \frac{5}{2} A_{n-1}\right)\right]. \tag{5.7}$$

A similar procedure gives the bias of $W(T)$ when $\phi(T) = \phi_0(T)$ as

$$B_{W_0}(\gamma) = \frac{\sigma_1^2 \gamma}{\sqrt{\sigma_1^2 + \sigma_2^2}} \sum_{n=0}^{\infty} \frac{e^{-\gamma^2/2} \gamma^{2n} A_{n+1}}{2^n n!}. \tag{5.8}$$

6. Equal variances. If $\sigma_1^2 = \sigma_2^2 = \sigma^2$, the estimator for θ_1 reduces to

$$W(T) = \frac{\hat{\theta}_1 + \hat{\theta}_2}{2} + \frac{T\phi(T)\sigma}{\sqrt{2}}, \tag{6.1}$$

where

$$T = \frac{\hat{\theta}_1 - \hat{\theta}_2}{\sigma\sqrt{2}}.$$

Let $\phi(T) = T^2/(1 + T^2)$, then the mean square deviation of $W_0(T)$ is

$$D^2_{W_0}(\gamma) = \frac{\sigma^2}{2} \left\{2 + e^{-\gamma^2/2} \left[3A_0 - \frac{5}{2} + \sum_{n=1}^{\infty} \frac{\gamma^{2n}}{2^n n!} \left(\frac{11}{2} A_n - \frac{5}{2} A_{n-1}\right)\right]\right\}, \tag{6.2}$$

and the bias is

$$B_{W_0}(\gamma) = \frac{\gamma\sigma}{\sqrt{2}} e^{-\gamma^2/2} \sum_{n=0}^{\infty} \frac{\gamma^{2n}}{2^n n!} A_{n+1}. \tag{6.3}$$

The "sometimes-pool" estimator for θ_1 is obtained if we let

$$\phi(T) = \begin{cases} 0 \text{ for } T < t_\alpha, \\ 1 \text{ for } T \geqq t_\alpha, \end{cases}$$

where $P(|t| \geqq t_\alpha) = \alpha$ and t is the standard normal deviate. Mosteller [8] gave the mean square of $SP(T)$.

$$\begin{aligned} D_{SP}^2(\gamma) = \left(\frac{\sigma^2}{2}\right)\Big\{2 &+ (t_\alpha + \gamma)N(t_\alpha + \gamma) + (t_\alpha - \gamma)N(t_\alpha - \gamma) \\ &+ (\gamma^2 - 1)\int_{-t_\alpha-\gamma}^{t_\alpha-\gamma} N(y)\,dy\Big\}. \end{aligned} \tag{6.4}$$

The bias of $SP(T)$ is obtained from the results reported by Bennett [5].

$$B_{SP}(\gamma) = \frac{\sigma}{\sqrt{2}}\left[\gamma \int_{-t_\alpha-\gamma}^{t_\alpha-\gamma} N(y)\,dy + N(t_\alpha + \gamma) - N(t_\alpha - \gamma)\right]. \tag{6.5}$$

The two estimators were compared on the basis of their efficiencies relative to $\hat{\theta}_1$. These efficiencies are plotted as functions of $|\gamma|$ in Figure 1. Let γ^* be defined as the largest value of $|\gamma|$ such that for all $|\gamma| < \gamma^*$ the efficiency is greater

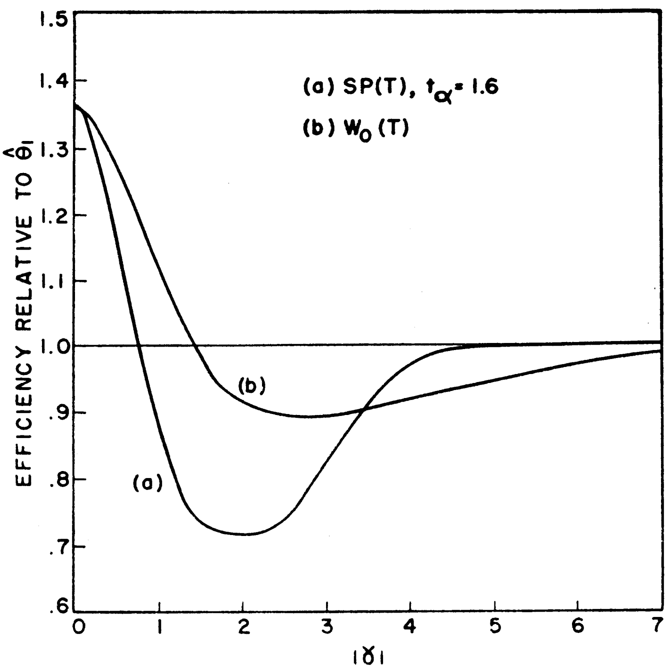

FIG. 1. Efficiencies of $SP(T)$ and $W_0(T)$

than one. γ^* will be referred to as the effective difference of the estimator. The curves of Figure 1 indicate that: (1) the maximum possible loss of efficiency is smaller for $W_0(T)$ than for $SP(T)$; (2) the effective difference is greater for $W_0(T)$; and (3) for larger values of $|\gamma|$, $SP(T)$ is more efficient than $W_0(T)$.

7. Two-parameter weighting functions. In order to study the effects of changing the shape of the weighting function curve, a two-parameter family of weighting functions is defined as

$$\phi(T; a, b) = 1 - ae^{-bT^2}, \tag{7.1}$$

where the parameters have ranges
(i) $0 \leqq a \leqq 1$,
(ii) $b \geqq 0$.
The mean square of $W_{ab}(T)$ is

$$D_{ab}^2(\gamma) = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}\left\{\sigma_2^2 + \sigma_1^2 \int_{-\infty}^{\infty} [\gamma - T(1 - ae^{-bT^2})]^2 N(T - \gamma)\, dT\right\}.$$

Integration yields

$$\begin{aligned} D_{ab}^2(\gamma) = \sigma_1^2 + \frac{\sigma_1^4}{\sigma_1^2 + \sigma_2^2}\Bigg\{&\frac{a^2}{(4b+1)^{3/2}}\left[1 + \frac{\gamma^2}{4b+1}\right] e^{-2b\gamma^2/(4b+1)} \\ &- \frac{2a}{(2b+1)^{3/2}}\left[1 - \frac{2b\gamma^2}{2b+1}\right] e^{-b\gamma^2/(2b+1)}\Bigg\}. \end{aligned} \tag{7.2}$$

For the special case where $\sigma_1^2 = \sigma_2^2 = \sigma^2$

$$\begin{aligned} D_{ab}^2(\gamma) = \frac{\sigma^2}{2}\Bigg\{&2 - \frac{2a}{(2b+1)^{3/2}}\left[1 - \frac{2b\gamma^2}{2b+1}\right] e^{-b\gamma^2/(2b+1)} \\ &+ \frac{a^2}{(4b+1)^{3/2}}\left[1 + \frac{\gamma^2}{4b+1}\right] e^{-2b\gamma^2/(4b+1)}\Bigg\}, \end{aligned} \tag{7.3}$$

and the efficiency of $W_{ab}(T)$ relative to $\hat{\theta}_1$ is $\sigma^2/D_{ab}^2(\gamma)$. This was evaluated for various values of a and b.

In Figure 2 the efficiency curves are plotted as functions of $|\gamma|$ for $a = 1.00$ and $b = 1.00, .50, .25, .10$. When a is fixed, decreasing b has the following effects: (1) for $|\gamma| \doteq 0$ the efficiency increases; (2) the maximum possible loss of efficiency is increased; and (3) the range of $|\gamma|$ for which large losses of efficiency may be sustained is increased without a corresponding increase in the effective difference.

The curves of Figure 3 are the efficiencies of $W_{ab}(T)$ for $b = .10$ and $a = 1.00$, .65, .42. They reveal that as a decreases, b fixed: (1) the efficiency at $|\gamma| = 0$ decreases; (2) the maximum possible loss decreases; and (3) the effective difference increases.

8. Comparison of $W_0(T)$, $W_{ab}(T)$, and $SP(T)$. The relative efficiencies of these three estimators for θ_1 in the case of equal variances are plotted as functions of

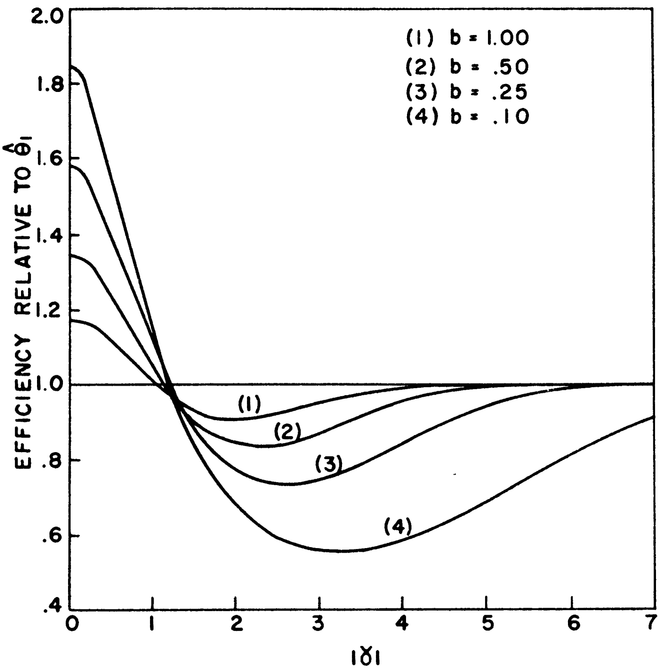

FIG. 2. Efficiency of $W_{ab}(T)$ for $a = 1.00$

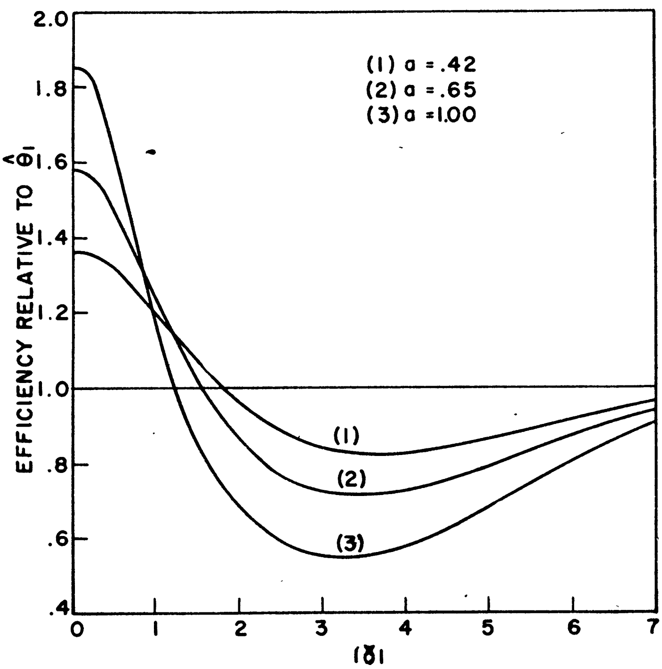

FIG. 3. Efficiency of $W_{ab}(T)$ for $b = .10$

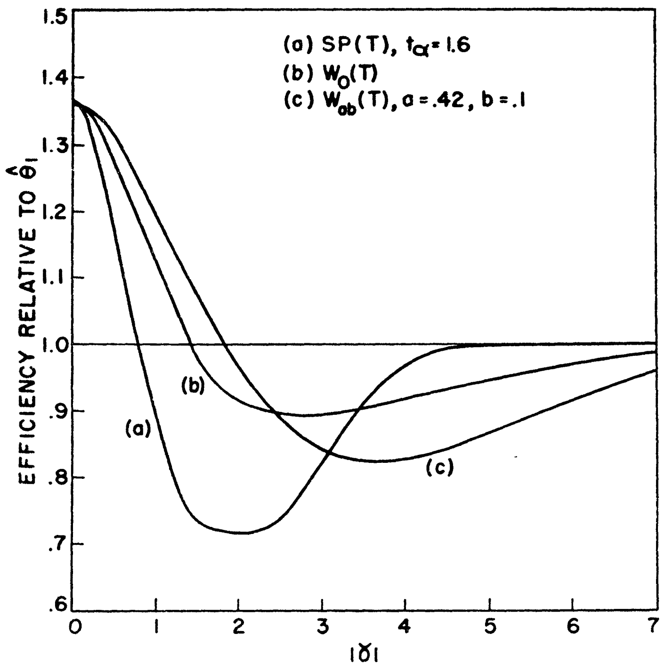

FIG. 4. Efficiencies of $SP(T)$, $W_0(T)$, and $W_{ab}(T)$

$|\gamma|$ in Figure 4. The constants involved were selected so that the efficiencies of all three would be very nearly equal for $|\gamma| = 0$. The following facts are apparent: (1) $W_{ab}(T)$ provides the greatest effective difference, $SP(T)$ the smallest; (2) the maximum loss is least for $W_0(T)$, greatest for $SP(T)$; and (3) the range of $|\gamma|$ for which large losses occur is shortest for $SP(T)$.

To compare these estimators on the basis of overall efficiency as defined in Section 4 the integral

$$J = -\frac{\sigma^2}{2} \int_{-\infty}^{\infty} T^2[1 - \phi(T)]^2 \, dT$$

was evaluated for each of the weighting functions plotted in Figure 4 with the following results, the largest value corresponding to the greatest overall efficiency; $J_0 = -.785\sigma^2$, $J_{ab} = -.875\sigma^2$, $J_{Sp} = -1.365\sigma^2$. Of the three, $W_0(T)$ has the greatest overall efficiency.

9. Discussion. The results of Sections 6, 7, and 8 indicate that in the case of normal estimators the generalized pooling procedure is effective in reducing the maximum loss of efficiency and increasing the effective difference.

Since it was shown in Section 4 that there is no uniformly minimum mean square error weighting function and no unbiased weighting function, the choice of a function, $\phi(T)$, might be based on one of the following criteria or on a combination of them:

(1) Select a weighting function which will provide a small maximum loss.
(2) Choose $\phi(T)$ so as to have a large overall efficiency.
(3) Select $\phi(T)$ to give a large effective difference.
(4) Select $\phi(T)$ so that the gain in efficiency is large when $\gamma = 0$.

It is readily apparent that these criteria are not independent and that minimizing the maximum loss or maximizing with respect to any one of the last three will, in general, lead to the never-pool estimator or will have adverse effects on the other characteristics of the estimator. Of the functions which were studied it appears that $\phi_0(T)$ is the best compromise when nothing is known concerning the size of the nuisance parameter γ. Any prior knowledge concerning γ might conceivably be used as an aid in selecting one of several possible alternative functions.

It is realized that only a beginning has been made on the applications and effects of the generalized pooling procedure and that the problems which were considered in this investigation belong to the simplest class of problems to which the procedure might be applied. The author feels, however, that the results which have been achieved here indicate that the procedure should be effective in controlling some of the disturbances which arise in other more complex applications of preliminary tests of significance. He feels that the advantages claimed for the weighting procedure in this study warrant further investigations along two lines: (1) An investigation should be made into the operating characteristics of the procedure when used in the other problems for which the effects of a preliminary test have been studied; and (2) a more rigorous examination of possible weighting functions and rules for their selection should be considered.

REFERENCES

[1] T. A. Bancroft, "On biases in estimation due to the use of preliminary tests of significance," *Ann. Math. Stat.*, Vol. 15 (1944), pp. 190–204.

[2] T. A. Bancroft, "Bias due to the omission of independent variables in ordinary multiple regression analysis" (abstract), *Ann. Math. Stat.*, Vol. 21 (1950), p. 142.

[3] T. A. Bancroft, "Preliminary tests and pool rules" (abstract), *J. Amer. Stat. Assn.*, Vol. 49 (1954), p. 348.

[4] R. E. Bechhofer, "The effect of preliminary tests of significance on the size and power of certain tests of univariate linear hypotheses," Unpublished Ph.D. thesis, Columbia Univ. Library.

[5] B. M. Bennett, "Estimation of means on the basis of preliminary tests of significance," *Ann. Inst. Stat. Math.*, Tokyo, Vol. 4 (1952), pp. 31–43.

[6] D. V. Huntsberger, "An extension of preliminary tests for pooling data" (abstract), *J. Amer. Stat. Assn.*, Vol. 49 (1954), p. 348.

[7] Tosio Kitagawa, "Successive process of statistical inference," *Mem. Faculty of Sci., Kyusyu Univ., Ser. A*, Vol. 5 (1950), No. 2, pp. 139–180.

[8] Frederick Mosteller, "On pooling data," *J. Amer. Stat. Assn.*, Vol. 43 (1948), pp. 231–242.

[9] A. E. Paull, "On a preliminary test for pooling mean squares in the analysis of variance," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 539–556.

[10] M. W. Zemansky, "Absorption and collision broadening of the mercury resonance line," *Phys. Rev.*, Vol. 36 (1930), pp. 219–238.