

AD-A088 072

PRINCETON UNIV NJ DEPT OF STATISTICS
FORMULAS FOR A TWO-SAMPLE MONTE CARLO SWINDLE.(U)
NOV 79 K KAFADAR

F/6 12/1

UNCLASSIFIED

TR-153-SER-2

ARO-14244.13-M

DAA829-76-6-0298

NL

1 of 1
80 PRO 2

END
DATE
FILMED
9-80
DTIC

18. APR 81

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD A088072

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 19 14244.13-M	2. GOVT ACCESSION NO. AD-A088 072	3. RECIPIENT'S CATALOG NUMBER 7
4. TITLE (and Subtitle) 6 FORMULAS FOR A TWO-SAMPLE MONTE CARLO SWINDLE.		5. TYPE OF REPORT & PERIOD COVERED Technical report
7. AUTHOR(s) 10 Karen Kafadar		6. PERFORMING ORG. REPORT NUMBER 115
9. PERFORMING ORGANIZATION NAME AND ADDRESS Princeton University Princeton, NJ 08540		8. CONTRACT OR GRANT NUMBER(s) DAAG29-76-G-0298
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 1477-153-FER-2
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) 12 13		12. REPORT DATE 11 Nov 79
		13. NUMBER OF PAGES 20
		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) NA A		
18. SUPPLEMENTARY NOTES The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) probability sampling Monte Carlo methods mathematical formulas simulation methodology		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) In order to obtain accurate estimates of the percentage points of the two-sample F -like statistic, we use a Monte Carlo swindle based on the method of conditional probabilities. The formulas for this implementation are described in this report. ←		

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

406873

80 8 14 113

DDC FILE COPY

FORMULAS FOR A TWO-SAMPLE
MONTE CARLO SWINDLE

by

Karen Kafadar
Princeton University

Technical Report No. 153, Series 2
Department of Statistics/
Princeton University
November 1979

This research was supported in part by a contract with the U.S. Army Research Office, No. DAAG29-76-0298, awarded to the Department of Statistics, Princeton University, Princeton, New Jersey.

A B S T R A C T

In order to obtain accurate estimates of the percentage points of the two-sample "t"-like statistic, we use a Monte Carlo swindle based on the method of conditional probabilities. The formulas for this implementation are described in this report.

Approved For
R.F. [redacted]
and [redacted]
Classified
Classification _____

By _____
Date _____
Approved for Codes
_____ and/or
Special _____

A |

0. INTRODUCTION

Efficient methods of problem solving using the computer are necessary for both accuracy and speed. These requirements often depend upon partial reduction of the problem to an analytic solution. This is the goal of Monte Carlo simulation, as opposed to experimental sampling.

This report is an extension of the formulas derived in Gross ([1]) for estimating the tail distribution of a statistic. In this work, the statistic was the form of a Student's-t for one sample. Here we derive the corresponding formulas for the two-sample statistic.

A. Location statistic.

Let $x_1, \dots, x_{n_1}$ be a (pseudo-) random sample from population 1 having location parameter μ_1 , and let $x_1, \dots, x_{n_2}$ be a (pseudo-) random sample from population 2 having location parameter μ_2 . Following the generation method described in [2], each x_i is generated as z_i/y_i , where $z_i \sim N(0, 1)$, and y_i comes from a positive distribution. Let $A = \sum_{i=1}^{n_1} x_i^2 / \sum_{i=1}^{n_1} y_i^2$. Likewise, each x_i is generated as z_i/y_i , where $z_i \sim N(0, 1)$, y_i comes from a positive distribution, and $a = \sum_{i=1}^{n_2} x_i^2 / \sum_{i=1}^{n_2} y_i^2$. Notice that, conditional on $\{y_i\}$ (respectively, $\{y_i\}$), A (resp. a) $\sim N(0, 1/\sum y_i^2)$ (resp. $N(0, 1/\sum y_i^2)$).

Let $T(X)$ (resp. $t(x)$) be an estimate of μ_1 (respec-

tively, μ_2), and let S_T (resp. s_t) be an estimate of the width of the distribution of $T(X)$ (resp. $t(x)$). Finally, let $G = (X - A)$, and $g = (x - a)$. Note that since width estimates are generally location invariant (as ours will be here), $S_T = S_{T(G)}$, $s_t = s_{t(g)}$, and since location estimates are translation invariant, $T(X) = T(G) + A$, $t(x) = t(g) + a$. In addition, we will assume that T and t are SYMMETRIC functions.

In order to use

$$W = \frac{T(X) - t(x)}{\sqrt{S_T^2 + s_t^2}}$$

to test the null hypothesis $H_0: \mu_1 = \mu_2$ and subsequently to derive a confidence interval for $\mu_1 - \mu_2$, we determine the percentage points from the distribution of W as follows:

$$\begin{aligned} P_0(W \leq k) &= E_0[I(-\infty, k | W)] \\ &= E_0[I(-\infty, k | (T(X) - t(x)) / (S_T^2 + s_t^2)^{1/2})] \\ &= E_0 \left[I(-\infty, k | \frac{T(G) + A - t(g) + a}{(S_T^2 + s_t^2)^{1/2}}) \right] \\ &= E_{G, A, X, Y} \left[E_0 \left[I(-\infty, k | \frac{A - a + T(G) - t(g)}{(S_T^2 + s_t^2)^{1/2}}) \right] \right] \\ &= E_{G, A, X, Y} \left[P_0 \left(\frac{A - a + T(G) - t(g)}{(S_T^2 + s_t^2)^{1/2}} \leq k \mid G, A, X, Y \right) \right] \\ &= E_{G, A, X, Y} P_0((A - a) \leq k(S_T^2 + s_t^2)^{1/2} - T(G) + t(g) \mid G, A, X, Y) \end{aligned}$$

$$= E \left[\frac{k(S_1^2 + s_1^2)}{T(\zeta) - t(\zeta)} \right]^{1/2} \frac{n_2}{(1/\sum_{i=1}^{n_1} Y_i^2 + 1/\sum_{i=1}^{n_2} Y_{ij}^2)^{1/2}}$$

$$\Delta \frac{1}{N} \sum_{j=1}^N \left[\frac{k(S_1^2 + s_1^2)}{T(\zeta) - t(\zeta)} \right]^{1/2} \frac{n_2}{(1/\sum_{i=1}^{n_1} Y_i^2 + 1/\sum_{i=1}^{n_2} Y_{ij}^2)^{1/2}}$$

where M is the number of samples from each of populations 1 and 2. (The subscripts on "E" refer to the variables over which the expectation is conditioned.)

Denote the term being summed as Y_j , which in this case is

$$Y_j = \left[\frac{k(S_1^2 + s_1^2)}{T(\zeta) - t(\zeta)} \right]^{1/2} \frac{n_2}{(1/\sum_{i=1}^{n_1} Y_i^2 + 1/\sum_{i=1}^{n_2} Y_{ij}^2)^{1/2}}$$

Then $P_0(W \leq k)$ can be estimated unbiasedly by

$$P_0 = \frac{1}{N} \sum_{j=1}^N Y_j$$

the variance of which ($t = E(P_0^2) - [E(P_0)]^2$) can be estimated by

$$\text{Var}(P_0) \Delta \frac{1}{N} \left[\frac{1}{N} \sum_{j=1}^N Y_j^2 - (P_0)^2 \right]$$

The next step to consider is reducing the variance of P_0 without generating any more samples from either population. How would the estimate of P_0 change if it were based on the negatives of the (pseudo-) random samples? Since T and t are symmetric, $T(-X)$ (resp. $t(-x)$) estimates $-\mu_1$ (resp. $-\mu_2$). Therefore, evaluating the conditional probability in (1) based on $-X_1, \dots, -X_{n_1}$ and $-x_1, \dots, -x_{n_2}$ yields

$$E_{G, S, X, X} \left[\frac{k(S_1^2 + s_1^2)}{T(\zeta) - t(\zeta)} \right]^{1/2} \frac{n_2}{(1/\sum_{i=1}^{n_1} Y_i^2 + 1/\sum_{i=1}^{n_2} Y_{ij}^2)^{1/2}} \Delta \frac{1}{N} \sum_{j=1}^N \phi_{2j}(2)$$

where

$$\phi_{2j} = \left[\frac{k(S_1^2 + s_1^2)}{T(\zeta) - t(\zeta)} \right]^{1/2} \frac{n_2}{(1/\sum_{i=1}^{n_1} Y_i^2 + 1/\sum_{i=1}^{n_2} Y_{ij}^2)^{1/2}}$$

Now $P_0(W \leq k)$ can be estimated by

$$P_0 = \frac{1}{N} \sum_{j=1}^N \phi_{2j}$$

where

$$Y_j = \frac{1}{2}(\theta_{1j} + \theta_{2j})$$

Again,

$$\text{Var}(P_0) \triangleq \frac{1}{N} \sum_{j=1}^N Y_j^2 - (P_0)^2$$

but now the estimate of P_0 has a smaller variance, since each Y_j involves two quantities which are negatively correlated. Gross ([1]) derived this improvement by recognizing that $P_0(W \leq k) = P_0(W_2 \leq k)$ by the symmetry of W , and (2) is a direct estimate of this latter probability.

Further reductions in the variance of P_0 are possible if the symmetry of the distribution of t , say $G(z - \mu_2)$, is assumed. Just as there is no loss of generality in taking the difference between μ_1 and μ_2 to be 0 in the null hypothesis (the distribution of W is a function of the difference $(\mu_1 - \mu_2)$ only), there will be no loss of generality in taking μ_2 to be 0, and the free parameter is μ_1 . Furthermore, the symmetry assumption on the distribution of t says that

$$G(\mu_2 + z) = 1 - G(\mu_2 - z)$$

i.e.,

$$G(z) = 1 - G(-z)$$

$$\Rightarrow dG(z) = -dG(-z)$$

Let $F(z - \mu_1)$ denote the distribution of T , and let

$$S_{\text{samp}}^2 = \frac{1}{N} \sum_{j=1}^N (T_j - \bar{T})^2 + s^2 \quad \text{and} \quad S_{\text{res}}^2 = \frac{1}{N} \sum_{j=1}^N (T_j - \bar{T})^2 + s^2 \quad \text{Then we have that}$$

$$\begin{aligned} P(W \leq k) &= P\{(T-t)/S_{\text{samp}} \leq k\} \\ &= P\{T \leq kS_{\text{samp}} + t\} \\ &= \int_{-\infty}^{\infty} P_0\{T \leq kS_{\text{samp}} + z\} dG(z) \\ &= \int_{-\infty}^{\infty} F(kS_{\text{samp}} + z - \mu_1) dG(z) \\ &= - \int_{-\infty}^{\infty} F(kS_{\text{samp}} + z - \mu_1) dG(-z) \\ &= \int_{-\infty}^{\infty} F(kS_{\text{samp}} - v - \mu_1) dG(v) \\ &= \int_{-\infty}^{\infty} P\{T \leq kS_{\text{samp}} - v\} dG(v) \\ &= P\{T + t \leq kS_{\text{samp}}\} \\ &= P\{T(X) + t(X) \leq kS_{\text{samp}}\} \\ &= P\{T(X) + A + t(X) + a \leq kS_{\text{res}}\} \\ &= E_{G, S, X, X} P\{A + a \leq kS_{\text{res}} - T(X) - t(X) \mid G, S, X, X\} \end{aligned}$$

Under the null hypothesis,

$$A + a \sim N(\mu_1 + 0, \frac{1}{2Y_1^2} + \frac{1}{2Y_2^2}).$$

Hence,

$$P_0(W \leq k) = E_{\xi, \Omega, X, Y} \left[\frac{k S_1^2 s_1^2 - (T_1 - A_1) - (t_1 - a_1)}{\left[\frac{n_1}{1/\sum_{i=1}^n Y_{1i}^2} + \frac{n_2}{1/\sum_{i=1}^n Y_{2i}^2} \right]^{1/2}} \right]$$

$$\Delta \frac{1}{N} \sum_{j=1}^N \phi_{3j}$$

where

$$\phi_{3j} = \phi \left[\frac{k \left[\frac{n_1}{1/\sum_{i=1}^n Y_{1i}^2} - (T_1 - A_1) - (t_1 - a_1) \right]}{\left[\frac{n_1}{1/\sum_{i=1}^n Y_{1i}^2} + \frac{n_2}{1/\sum_{i=1}^n Y_{2i}^2} \right]^{1/2}} \right]$$

Again, recognizing the symmetry of the conditional distribution of $A + a$, we arrive at

$$P_0(W \leq k) \Delta \frac{1}{N} \sum_{j=1}^N \phi_{4j},$$

where

$$\phi_{4j} = \phi \left[\frac{k \left[\frac{n_1}{1/\sum_{i=1}^n Y_{1i}^2} + (T_1 - A_1) + (t_1 - a_1) \right]}{\left[\frac{n_1}{1/\sum_{i=1}^n Y_{1i}^2} + \frac{n_2}{1/\sum_{i=1}^n Y_{2i}^2} \right]^{1/2}} \right].$$

In taking our final estimate of P_0 as

$$\frac{1}{N} \sum_{j=1}^N \gamma_j, \quad \gamma_j = \frac{1}{4} (\phi_{1j} + \phi_{2j} + \phi_{3j} + \phi_{4j}) \quad (3)$$

it remains to check that all six pairs between the ϕ_{kj} are negatively correlated. It is clear that the two pairs ϕ_{1j}, ϕ_{2j} and ϕ_{3j}, ϕ_{4j} are, since they are based on antithetic sampling. That ϕ_{1j}, ϕ_{3j} and ϕ_{2j}, ϕ_{4j} are negatively correlated is shown by the fact that if $T(\xi)$ is held fixed, ϕ_{1j} (resp. ϕ_{4j}) is increasing in $t(\xi)$, while ϕ_{3j} (resp. ϕ_{2j}) decreases in $t(\xi)$. Likewise, holding $t(\xi)$ fixed, ϕ_{1j}, ϕ_{4j} and ϕ_{2j}, ϕ_{3j} are negatively correlated by the same reasoning on $T(\xi)$. Hence, the overall variance of γ_j has been reduced, and we can use (3) as our final estimate.

One further improvement may be worth investigating. As in equation (1),

$$P_0(W \leq k) = E_{\xi, \Omega, X, Y} \left[P(A - a \leq k S_{res} - T(C) + t(c) \mid \xi, \Omega, X, Y) \right]$$

$$= E_{\xi, \Omega, X, Y} \left[P(A \leq k S_{res} - T(C) + t(c) + a \mid \xi, \Omega, X, Y) \right]$$

Given ξ, Ω, X, Y , the distribution of A alone is $N(\mu_1, 1/2Y_1^2) = N(0, 1/2Y_1^2)$, and the conditional distribution of a is $N(\mu_2, 1/2Y_2^2) = N(0, 1/2Y_2^2)$ (under the null hypothesis). The inner conditional probability can then be evaluated as

$$P(A \leq kS_{res} - T(C) + t(c) + a | \xi, \eta, X, Y)$$

$$= \int_{-\infty}^{\infty} P(A \leq kS_{res} - T(C) + t(c) + a | \xi, \eta, X, Y) d\phi(z) \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{2\pi}} dz$$

$$= \int_{-\infty}^{\infty} \phi \left\{ (kS_{res} - T(C) + t(c) + a) \sqrt{\frac{2}{\pi}} \right\} d\phi(z) \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{2\pi}} dz$$

$$= (2\pi)^{-1/2} \int_{-\infty}^{\infty} \phi \left[\frac{u - (2\pi)^{1/2} (-kS_{res} + T(C) - t(c))}{(2\pi)^{1/2} \sqrt{2\pi}} \right] d\phi(u)$$

$$= (2\pi)^{-1/2} \int_{-\infty}^{\infty} \phi \left[\frac{u - \frac{1}{\eta}}{\eta} \right] d\phi(u)$$

where

$$\eta = (2\pi)^{1/2} (-kS_{res} + T(C) - t(c))$$

$$\eta^2 = \frac{2}{\pi} \sum_{i=1}^{n_1} \frac{1}{X_i^2}$$

An evaluation of this integral is numerically possible (see Note 1). The result will be an increasing function of

$$\eta = (2\pi)^{1/2} (-kS_{res} + T(C) - t(c)).$$

Thus,

$$\phi_{5j} = (2\pi)^{-1/2}.$$

$$\int_{-\infty}^{\infty} \phi \left[\frac{u - (2\pi)^{1/2} (-k(S_{res}^2 + s^2))^{1/2} + (T - A_1) - (t - a_1)}{(2\pi)^{1/2} \sqrt{2\pi}} \right] d\phi(u)$$

will be negatively correlated with each of the other ϕ_{kj} , $k=1, \dots, 4$. However, the additional computation required to

estimate $P_0(W_k)$ by

$$\hat{P}_0 = N^{-1} \sum_{j=1}^N Y_j, \quad Y_j = (\phi_{1j} + \dots + \phi_{5j})^{1/5}$$

may not be worth the possibly slight reduction in the variance of $\hat{P}_0$.

B. Location and Scale Swindle.

The location and scale swindle follows the same principle as the location swindle, but now the configurations are taken as

$$T(\xi) = T \left(\frac{X - A}{B} \right) = \frac{T(X) - A}{B}$$

and

$$t(\xi) = t \left(\frac{X - A}{B} \right) = \frac{t(X) - a}{B},$$

where

$$(n_1 + n_2 - 2)B^2 = \sum_{i=1}^{n_1} (Z_i - AY_i)^2 + \sum_{i=1}^{n_2} (z_i - ay_i)^2.$$

As illustrated in [4], the first term is distributed as $\chi^2_{(n_1-1)}$ and is independent of A; since the second term is based on an entirely different sample, B itself is independent of A. Likewise, the second term is distributed as $\chi^2_{(n_2-1)}$; by the same reasoning, B is also independent of a. By the reproductive property of the chi-squared distribution

$$(n_1 + n_2 - 2)S^2 = \chi^2_{(n_1+n_2-2)}$$

Now,

$$S^2 = \text{Var}(T(X)) = \text{Var}\left[\left(\frac{TX - S}{B}\right) \cdot B + A\right] \\ = \text{Var}(T(X)B + A) = B^2 \text{Var}(T(X)) = B^2 S^2_C$$

and similarly

$$s^2 = B^2 S^2_C$$

Hence,

$$S_{\text{amp}} = (S^2 + s^2)^{1/2} = BS_{\text{res}}$$

The location and scale swindle then proceeds as:

$$P_0(W \leq k) = P_0(T(X) - t(x) \leq k' S_{\text{amp}}) \\ = P_0\left(\left(\frac{TX - A}{B}\right)B + A\right) - \left[\left(\frac{TX - A}{B}\right)B + A\right] \leq k S_{\text{amp}} \\ = P_0(T(X)B - t(x)B \leq kBS_{\text{res}} - A + A) \\ = E_{G, A, X, Y} P_0\left(\frac{kS_{\text{res}}B - A + A}{B} \geq T(X) - t(x) \mid G, A, X, Y\right) \\ = E_{G, A, X, Y} P_0\left(kS_{\text{res}} - \frac{A - A}{B} \geq T(X) - t(x) \mid G, A, X, Y\right) \\ = E_{G, A, X, Y} P_0\left(kS_{\text{res}} - \frac{A - A}{B} \geq T(X) - t(x) \mid G, A, X, Y\right)$$

$$= E_{G, A, X, Y} P_0\left[\frac{A - A}{1/2} \leq \frac{kS_{\text{res}} - (T(X) - t(x))}{1/2} \mid G, A, X, Y\right]$$

The expression

$$\frac{kS_{\text{res}} - (T(X) - t(x))}{(1/2S^2 + 1/2s^2)^{1/2}}$$

depends only on the configuration and on X, Y . Also, since the conditional distribution of the numerator is

$$\frac{A - a}{(1/2S^2 + 1/2s^2)^{1/2}} \sim N(0, 1)$$

and that of the denominator is

$$B \sim \left[\chi^2_{(n_1+n_2-2)} / (n_1+n_2-2) \right]^{1/2}$$

the conditional probability is Student's t on $(n_1 + n_2 - 2)$ degrees of freedom. Thus,

$$P_0(W \leq k) = \frac{1}{N} \sum_{j=1}^N F_{t'} \left[\frac{k(S_C^2 + s_C^2)^{1/2} - \left(\frac{A - a}{B}\right) + \left(\frac{1 - a^2}{B}\right)^{1/2}}{(1/2S^2 + 1/2s^2)^{1/2}} \right]$$

where $F_{t'}$ is the cumulative distribution function of Student's- t on $U = n_1 + n_2 - 2$ degrees of freedom. As in the location swindle, the same improvements in the estimate of P_0 by using all possible combinations of sign on $T(X)$ and $t(x)$ are applicable here as well.

C. Power Estimates.

Using either swindle, the power of W against various points in the alternative, $H_1: \mu_1 - \mu_2 = \delta$, can be approximated. For the location swindle,

$$\begin{aligned} P_G(W|K) &= P_G(T(X) - t(X) \geq kS_{\text{saap}}) \\ &= P_G((T(\underline{Q}) + A) - (t(\underline{Q}) - a) \geq kS_{\text{res}}) \\ &= E_{\underline{Q}, \underline{Q}, \underline{X}, \underline{Y}} P_G(A - a - \delta \geq kS_{\text{res}} - T(\underline{Q}) + t(\underline{Q}) - \delta ; \underline{Q}, \underline{Q}, \underline{Y}, \underline{Y}) \end{aligned}$$

$$\hat{\Delta} = \frac{1}{N} \sum_{j=1}^N \left[\frac{(T_j - A_j) - (t_j - a_j) + \delta - kS_j}{(1/\sum_{i=1}^{n_1} Y_{1i}^2 + 1/\sum_{i=1}^{n_2} Y_{2i}^2)^{1/2}} \right]$$

where again all four possible combinations of sign on $(T_j - A_j)$ and $(t_j - a_j)$ may be utilized to achieve a more efficient estimate.

In the case of the location and scale swindle, however, the conditional probability is noncentral Student's t :

$$\begin{aligned} P_G(W|K) &= P_G((T(\underline{Q}) + A) - (t(\underline{Q}) - a) \geq kS_{\text{res}}) \\ &= P_G(\frac{A - a}{B} \geq kS_{\text{res}} - T(\underline{Q}) + t(\underline{Q})) \\ &= E_{\underline{Q}, \underline{Q}, \underline{X}, \underline{Y}} P_G \left[\frac{A - a}{B(1/\sum_{i=1}^{n_1} Y_{1i}^2 + 1/\sum_{i=1}^{n_2} Y_{2i}^2)^{1/2}} \geq \frac{kS_{\text{res}} - T(\underline{Q}) + t(\underline{Q})}{(1/\sum_{i=1}^{n_1} Y_{1i}^2 + 1/\sum_{i=1}^{n_2} Y_{2i}^2)^{1/2}} \right] \\ &\quad ; \underline{Q}, \underline{Q}, \underline{X}, \underline{Y} \end{aligned}$$

$$\hat{\Delta} = \frac{1}{N} \sum_{j=1}^N \left[1 - F_{\delta, U} \left[\frac{kS_{\text{res}} - T(\underline{Q}) + t(\underline{Q})}{(1/\sum_{i=1}^{n_1} Y_{1i}^2 + 1/\sum_{i=1}^{n_2} Y_{2i}^2)^{1/2}} \right] \right]$$

where $F_{\delta, U}$ is a tabulated or approximated noncentral- t with noncentrality parameter δ and degrees of freedom $U = n_1 + n_2 - 2$.

An alternative approach to achieving power estimates that does not require a noncentral Student's t would be to swindle on scale separately as follows:

Suppose $X \sim N(\delta, 1)$, $B \sim \sqrt{\chi_n^2/n}$. Then

$$\begin{aligned} P\left(\frac{X}{B} \geq r\right) &= 1 - F_{\delta, n}(r), \\ &= \int_0^{\infty} P(X \geq Br ; B=b) df^B(b). \end{aligned}$$

Since $g(B) = nb^2 - \chi_n^2$, $\partial g/\partial B = 2nB$,

$$\begin{aligned} f^B(b) &= (f^n(n/2) \cdot 2^{n/2})^{-1} \cdot (nb^2)^{n/2} \cdot 1 \cdot e^{-nb^2/2} \cdot I_{(0, \infty)}(b) \\ &= n^{n/2} (f^n(n/2) \cdot 2^{n/2})^{-1} \cdot b^{n-1} \cdot e^{-nb^2/2} \cdot I_{(0, \infty)}(b). \end{aligned}$$

Thus,

$$P\left(\frac{\lambda}{b} \geq r\right) = \int_0^\infty P\left\{X - 6 \geq br - 6\right\} dF^B(b) \\ = \int_0^\infty \phi\left(6 - br\right) \cdot n^{n/2} \left(\frac{n}{2}\right)^{n/2} \cdot e^{-nb^2/2} db.$$

Evaluation of the integral is numerically possible.

(The details will not be shown here, but the argument follows the same lines as that in Note 1.) Since r represented

$$\frac{kS_{res} - T(\Omega) - t(\Omega)}{(1/2Y_1^2 + 1/2Y_1^2)^{1/2}}.$$

additional efficiency can again be obtained by using all possibility of sign on $T(\Omega)$ and $t(\Omega)$.

Note 1. Evaluation of $\rho \cdot \int_{-\infty}^\infty \phi\left(\frac{u-b}{\eta}\right) d\phi(u)$, where

$$\rho = \left(\sum_{i=1}^{n_2} Y_1^2\right)^{-1/2} \\ \bar{b} = \rho^{-1} \cdot (-kS_{res} + T(\Omega) - t(\Omega)) \\ \eta^2 = \sum_{i=1}^{n_2} Y_1^2 / \sum_{i=1}^{n_2} Y_1^2.$$

First, let us note that:

$$(1) \text{ as } \bar{b}/\eta \rightarrow -\infty, \phi\left(\frac{u-b}{\eta}\right) \cdot d\phi(u) \rightarrow 0 \\ \rightarrow \int_{-\infty}^\infty \phi\left(\frac{u-b}{\eta}\right) \cdot d\phi(u) \rightarrow 0; \\ (11) \text{ for } \bar{b} = 0, \eta = 1, \int_{-\infty}^\infty d\phi(u) = 1/2; \\ (111) \text{ as } \bar{b}/\eta \rightarrow +\infty, \phi\left(\frac{u-b}{\eta}\right) d\phi(u) = \phi\left(\frac{u-b}{\eta}\right) \cdot 1 \\ \rightarrow \int_{-\infty}^\infty \phi\left(\frac{u-b}{\eta}\right) d\phi(u) \rightarrow 1.$$

The "sledgehammer" approach expands $\phi\left(\frac{u-b}{\eta}\right)$ in a Taylor series about 0 and uses the moment properties of the Gaussian distribution:

$$\int_{-\infty}^\infty \phi\left(\frac{u-b}{\eta}\right) d\phi(u) = \int_{-\infty}^\infty \sum_{k=0}^\infty \left(\frac{u-b}{\eta}\right)^k \frac{1}{k!} \phi^{(k)}(0) \cdot \frac{1}{\eta} d\phi(u) \\ = \sum_{k=0}^\infty \frac{\phi^{(k)}(0)}{\eta^k k!} \int_{-\infty}^\infty (u-b)^k d\phi(u)$$

(interchange OK: everything absolutely integrable)

$$= \sum_{k=0}^\infty \frac{\phi^{(k)}(0)}{\eta^k k!} \sum_{j=0}^k \binom{k}{j} (\bar{b})^{k-j} \int_{-\infty}^\infty u^j d\phi(u) \\ = \frac{1}{2} + \sum_{p=0}^\infty \frac{\phi^{(p)}(0)}{\eta^{p+1} (p+1)!} \sum_{j=0}^{p+1} \binom{p+1}{j} (\bar{b})^{p+1-j} \int_{-\infty}^\infty u^j d\phi(u)$$

Now,

$$\phi^{(p)}(0) = \begin{cases} 0 & p \text{ odd} \\ \frac{(-1/2)^{p/2} p!}{(2\pi)^{1/2} (p/2)!} & p \text{ even} \end{cases}$$

and

$$E(N^j) = \begin{cases} 0 & j \text{ odd} \\ \frac{1}{2^{j/2} (j/2)!} & j \text{ even} \end{cases}$$

Substituting in these formulas, and changing the index of summation,

$$\int_{-\infty}^{\infty} \phi(u) d\phi(u) = \frac{1}{2} + \sum_{m=0}^{\infty} \frac{(-1)^m (2m)!}{2^m m!} \sum_{p=0}^{\infty} \frac{(-1)^p (2p)!}{2^p p!} \frac{1}{2} \frac{(m-p)!}{(m+p)!}$$

which, unfortunately, looks difficult in providing a numerical answer.

Another more feasible approach is in terms of standardized Hermite polynomials. Following [2], the integral can be written as

$$\int_{-\infty}^{\infty} \phi(u) d\phi(u) = \int_{-\infty}^{\infty} \phi(u) d\phi(u)$$

$$\xi = \left(\sum_{i=1}^{n_1} \frac{1}{2} \right)^{1/2} (k - \tau(\xi) + i(\xi)) ; \quad \eta^2 = \sum_{i=1}^{n_1} \frac{1}{2} \frac{y_i^2}{x_i^2}$$

Now,

$$\int_{-\infty}^{\infty} \phi(u) d\phi(u) = \int_{-\infty}^{\infty} \phi(u) d\phi(u) \cdot \frac{1}{\phi_{0,1}(u)} \cdot \phi_{0,1}(u)$$

$$= \langle \phi_{0,1}, \frac{1}{\phi_{0,1}} \rangle$$

where $\langle \cdot, \cdot \rangle$ is the inner product in the L^2 space of functions spanned by $\phi_{0,1}(Y)dY$ as basis elements. Using the standardized Hermite polynomials

$$h_n(x) = (n!)^{-1/2} He_n(x),$$

this inner product may be written as

$$\sum_{p=0}^{\infty} \langle \phi_{0,1}, h_p \rangle \langle \frac{1}{\phi_{0,1}}, h_p \rangle.$$

Using Equation (3.7) from [2], $\langle \phi_{0,1}, h_p \rangle$ may be expressed as

$$\langle \phi_{0,1}, h_p \rangle = \begin{cases} \frac{2(-1)^{(p-1)/2} (p!)^{1/2}}{\sqrt{p!} ((p-1)/2)!} \frac{(p+1)/4}{p} & p \text{ odd} \\ 0 & p \text{ even} \end{cases}$$

and, using (5.4) and (5.5) from [2],

$$\langle \phi_{0,1}, h_p \rangle = \begin{cases} (1-\eta^2)^{p/2} h_p(\xi(1-\eta^2)^{-1/2}) & \eta^2 < 1 \\ \xi^{p-1} / p! & \eta^2 = 1 \end{cases}$$

Thus,

$$\int_{-\infty}^{\infty} \phi_{0,1}(x) d\phi(x) = \frac{2}{\eta^2} \sum_{p=0}^{\infty} \frac{(-1)^p (p-1)/2}{\sqrt{p!} ((p-1)/2)!} \frac{(p!)^{1/2}}{(p+1)/4} \cdot (1-\eta^2)^{p/2}$$

$$= h_p(\xi(1-\eta^2)^{-1/2}), \quad p \text{ odd}, \quad \eta^2 < 1$$

$$= \frac{2}{\sqrt{p!} ((p-1)/2)!} \frac{(p!)^{1/2}}{(p+1)/4} \cdot (1-\eta^2)^{p/2}$$

$$= h_{2r+1}(\xi(1-\eta^2)^{-1/2}), \quad \eta^2 < 1$$

$$= \frac{2}{\sqrt{p!} ((p-1)/2)!} \frac{(p!)^{1/2}}{(p+1)/4} \cdot (1-\eta^2)^{p/2}$$

$$= \frac{2}{\sqrt{p!} ((p-1)/2)!} \frac{(p!)^{1/2}}{(p+1)/4} \cdot (1-\eta^2)^{p/2} = 1, \quad \eta^2 = 1.$$

The restriction $\eta^2 \leq 1$ can be assured in the Gaussian case, either when the sample sizes are equal ($\sum y_1^2 = \sum y_2^2 = n$) or with unequal sample sizes, if population 2 is the one having the larger sample size ($n_1 < n_2$). When the distributions are different, no guarantee can be made that $\eta^2 \leq 1$; the samples would have to be compared in each case so that $(\sum_{i=1}^{n_1} y_{1i}^2 = \sum_{i=1}^{n_2} y_{2i}^2)$, for all $j=1, \dots, M = \#$ of samples. For anything other than the Gaussian situation, this is likely to be more trouble than it is worth. For the Gaussian situation, tabulated values of $n_{2r+1}(\frac{1}{2}(1 - n_1/n_2))^{-1/2}$ may assist in approximating the integral, from which a suitable cutoff limit in the infinite summation may be derived to obtain a sufficiently accurate finite approximation. (To date, no such results were attempted; hence, the improvement in the variance of the estimate of P_0 has not been determined.)

REFERENCES

- [1] Gross, A. M. (1973). A Monte Carlo Swindle for Estimators of location. *Appl. Statist.*, 22, No. 3, 347-353.
- [2] Kates, Louis K. (1977). Hermite Polynomials and the Gaussian Distribution. Unpublished manuscript, Princeton University.
- [3] Relles, D. (1970). Variance Reduction Techniques for Monte Carlo Sampling from Student Distributions. *Technometrics*, 12, 499-515.
- [4] Simon, Gary (1975). Swindles to Improve Computer Simulation, with Application of the Problems of Appraising Estimates of Location and Dispersion in Univariate Samples. Technical Report No. 91, Series 2, Princeton University.