COMMUN. STATIST.-SIMULA., 15(3), 715-732 (1986)

TESTING THE MEAN OF AN ASYMMETRIC POPULATION: JOHNSON'S MODIFIED t TEST REVISITED

J.P.C. Kleijnen, G.L.J. Kloppenburg
and F.L. Meeuwsen

Catholic University Tilburg
School of Business and Economics
5000 LE Tilburg, Netherlands

Key Words and Phrases: *t variable; skewness; power; geometric analysis.*

ABSTRACT

Confidence intervals for the mean of an asymmetric distribution can be based on Student's t statistic or on Johnson's modified t statistic; Johnson's statistic has two variants, based on a linear and a quadratic approximation respectively. The quadratic approximation is complicated and is first investigated geometrically, which results in new insight. Next Monte Carlo experiments yield estimates of the coverage and power of several variations on Johnson's test. These experiments show that the quadratic approximation is superior.

 0361-0918/86/1503-0715$3.50/0

1. INTRODUCTION

We investigate several confidence intervals for the mean of an asymmetric distribution. Norman Johnson (1978, p. 537) modified Student's t statistic, explicitly accounting for skewness μ_3:

$$t_1 = [(\bar{x}-\mu) + \frac{\mu_3}{6\sigma^2 N} + \frac{\mu_3}{3\sigma^4}(\bar{x}-\mu)^2]\ [s^2/N]^{-\frac{1}{2}} \tag{1.1}$$

where all symbols are standard (mean μ, variance σ^2, third central moment μ_3, sample mean $\bar{x}$, unbiased sample variance s^2, sample size N) and where in practice σ^2 and μ_3 are estimated by the unbiased moment estimators s^2 and

$$\hat{\mu}_3 = \frac{N}{(N-1)(N-2)} \sum_1^N (x_i - \bar{x})^3; \tag{1.2}$$

see Kenney and Keeping (1954, p. 100). The skewness measure μ_3 is unknown in general, except when we can prove analytically that the distribution is symmetric ($\mu_3 = 0$). In the latter (rare) case we do not need Johnson's modification; also see section 3 (sub v) and section 4 (subsection 3).

Johnson (1978, p. 538) further stated "... the effect of the term involving $(\bar{x}-\mu)^2$ is small order ...neglecting the term involving $(\bar{x}-\mu)^2$ .... reduces [t_1] to the variable t_1'...". Our preliminary experiments, however, showed that neglecting $(\bar{x}-\mu)^2$ definitely affects the coverage and power of the test. Those experiments also demonstrated that distribution-free alternatives, like the sign and Wilcoxon's signed rank test, do not work for an asymmetric distribution (and as the sample size N increases, these statistics perform worse). Therefore we shall compare several versions of Johnson's statistic to the classical t statistic. First we shall present analytical results; next we shall discuss Monte Carlo estimates of coverage and power.

2. CONFIDENCE INTERVALS REANALYZED

To analyze confidence intervals based on the quadratic term $(\bar{x}-\mu)^2$, we develop a graphical representation that seems new. For didactic reasons we first discuss Student's t test; see Fig. 1(a). The t statistic with ν degrees of freedom satisfies the following equation:

$$1 - \alpha = P\,[t_\nu^2 \leq t_{\nu,\alpha/2}^2\,]$$

$$= P\,[-t_{N-1,\alpha/2}s/\sqrt{N} \leq \bar{x} - \mu \leq t_{N-1,\alpha/2}s/\sqrt{N}]. \qquad (2.1)$$

If we define

$$f_1(\mu) = \bar{x} - \mu, \qquad c = t_{N-1,\alpha/2}s/\sqrt{N},$$

$$\mu_{(1)} = \bar{x} - c, \qquad \mu_{(2)} = \bar{x} + c\,, \qquad (2.2)$$

then eq. (2.1) implies with probability $1-\alpha$

$$\{f_1(\mu) \leq c\} \cap \{f_1(\mu) \geq -c\} \qquad (2.3)$$

where $\{\ \}$ denotes a set and $\cap$ denotes intersection; (2.3) is equivalent to

$$\{\mu_{(1)} \leq \mu \leq \mu_{(2)}\}\,, \qquad (2.4)$$

which is equivalent to the well-known Student confidence interval.

Now we consider Johnson's linear approximation t_1'. We have pictured a positive μ_3 in Fig. 1(b). We define

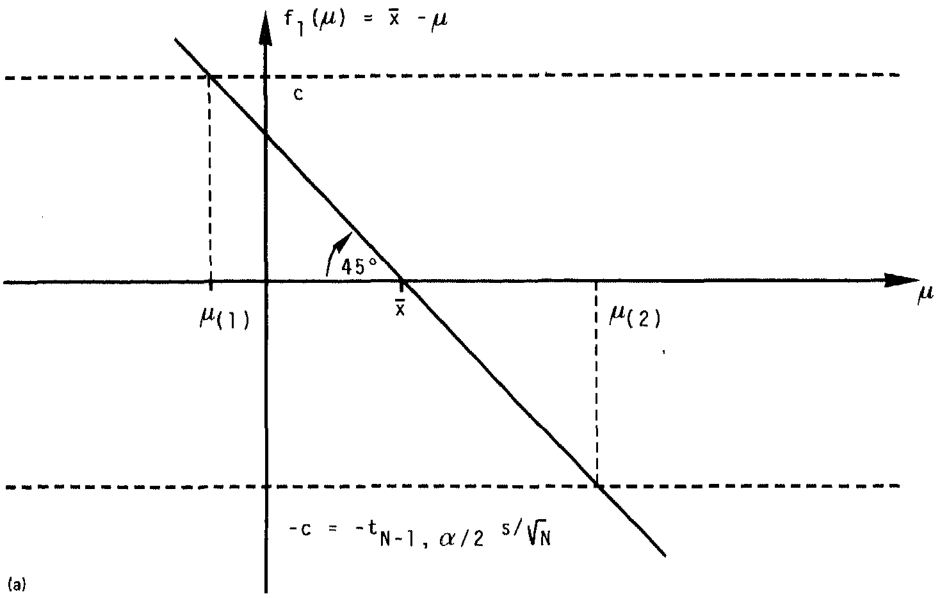


FIG. 1(a). Student statistic t.

$$\mu'_{(1)} = \bar{x} - c + \frac{\hat{\mu}_3}{6s^2N}\ , \tag{2.5}$$

$$\mu'_{(2)} = \bar{x} + c + \frac{\hat{\mu}_3}{6s^2N}\ . \tag{2.6}$$

The resulting interval is not centered around the sample mean $\bar{x}$; it moves to the right when compared to the Student interval in (2.4); it covers $\bar{x}$ unless a small N yields a high $\hat{\mu}_3$. For negative skewness analogous results hold; computationally, if $\mu_3 < 0$ then we may apply the formulas derived for $\mu_3 > 0$, to $-x$.

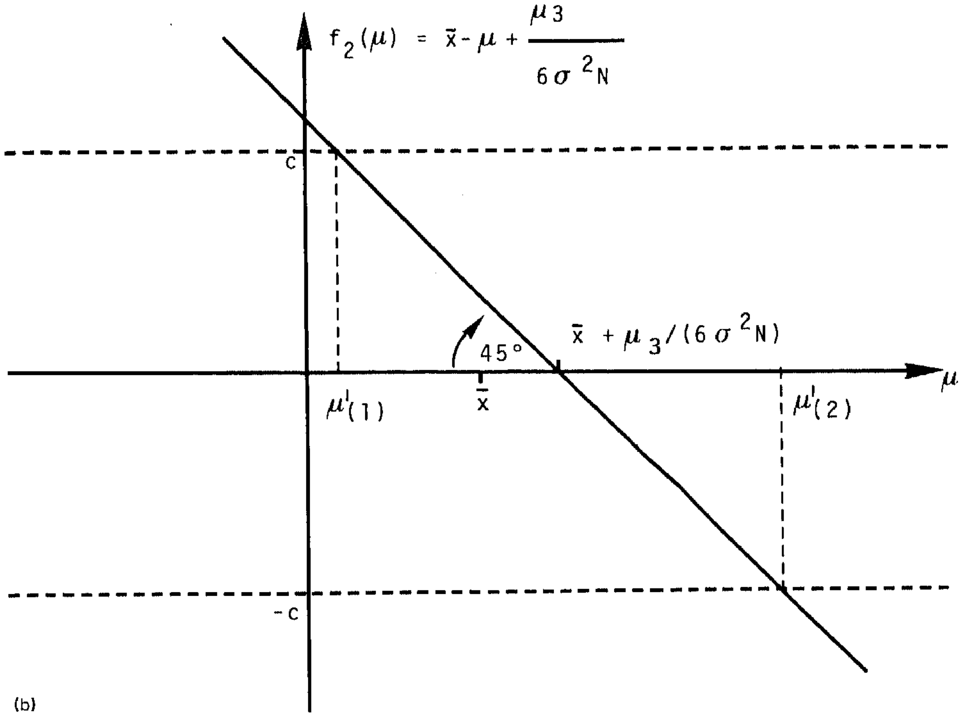


FIG.1(b). Johnson's first-order modification t'_1

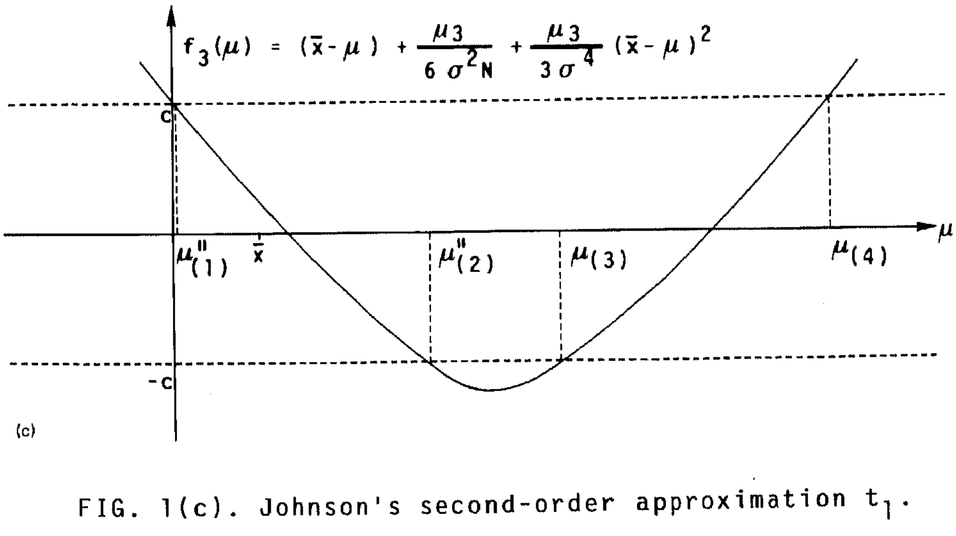


FIG. 1(c). Johnson's second-order approximation t_1.

Concerning the quadratic approximation t_1 Johnson (1975, p. 538) stated: "Use of the variable t_1 does not lead to a simple expression for confidence intervals for μ since the numerator of t_1 in nonlinear in μ"; he does not elaborate. We analyze possible complications, using the graphical representation of Fig. 1(c) where $f_3(\mu)$ equals the first factor in eq. (1.1), i.e., $f_3(\mu)$ is a second-degree polynomial in μ. The mathematical analysis yields

$$\{f_3(\mu) < c\} = \{\mu''_{(1)} \leq \mu \leq \mu_{(4)}\} \tag{2.7}$$

where Fig. 1(c) illustrates a situation with $\mu_3 > 0$ and

$$\mu''_{(1)} = \bar{x} + \frac{1}{2c_2} - \frac{\sqrt{d_1}}{2c_2}, \quad \mu_{(4)} = \bar{x} + \frac{1}{2c_2} + \frac{\sqrt{d_1}}{2c_2} \tag{2.8}$$

where we define $c_1 = \hat{\mu}_3/(6s^2N), c_2 = \hat{\mu}_3/(3s^4)$, $d_1 = 1-4c_2\ (c_1-c)$. The set of (2.7) should be intersected with

$$\{f_3(\mu) > -c\} = \{\mu \leq \mu''_{(2)}\} \cup \{\mu \geq \mu_{(3)}\} \tag{2.9}$$

where $\cup$ denotes union and

$$\mu''_{(2)} = \bar{x} + \frac{1}{2c_2} - \frac{\sqrt{d_2}}{2c_2}, \quad \mu_{(3)} = \bar{x} + \frac{1}{2c_2} + \frac{\sqrt{d_2}}{2c_2} \tag{2.10}$$

where $d_2 = 1 - 4\ c_2\ (c_1 + c)$. The mathematical analysis of t_1 (to be distinguished from the statistical solution, discussed below) yields

$$\{\mu''_{(1)} \leq \mu \leq \mu''_{(2)}\} \cup \{\mu_{(3)} \leq \mu \leq \mu_{(4)}\} . \tag{2.11}$$

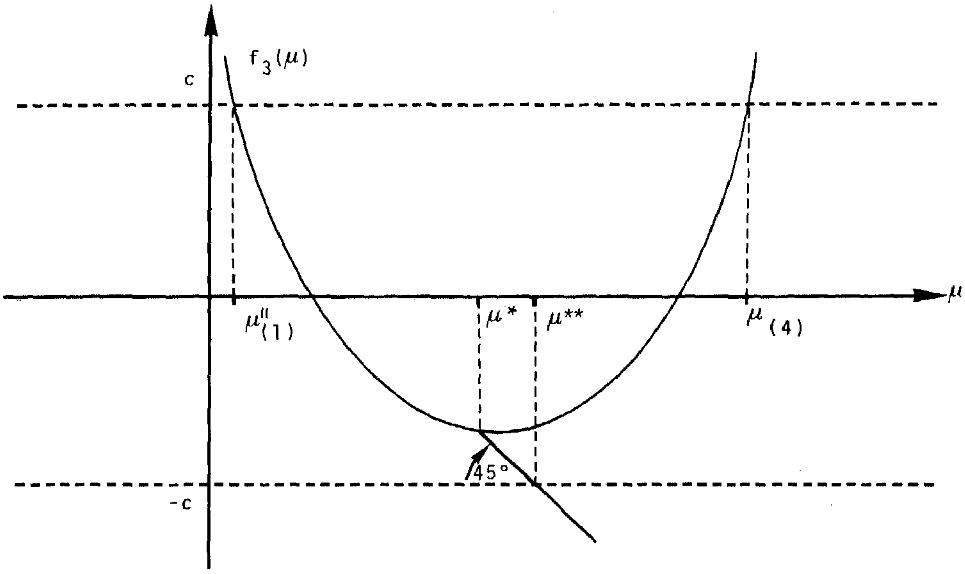


FIG. 2. No intersection with -c.

The disjunct interval $[\mu_{(3)}, \mu_{(4)}]$ does not seem to make sense statistically, so that it seems better to eliminate this interval. In the Monte Carlo experiment we shall compare procedures with and without elimination of the disjunct interval that does not cover $\bar{x}$. If the skewness is negative, then the results are analogous (see below).

There is a further complication, namely $f_3(\mu)$ may <u>not intersect</u> the lower horizontal line, i.e.,

$$\min_{\mu} [f_3(\mu)] > -c . \tag{2.12}$$

We investigate two heuristic solutions; see Fig. 2. The first solution makes the upper limit of the confidence interval covering $\bar{x}$ equal to μ^*, the value where $f_3(\mu)$ reaches its minimum.

Obviously μ^* lies halfway $\mu''_{(1)}$ and $\mu_{(4)}$ of eq. (2.8):

$$\mu^* = \bar{x} + \frac{1}{2c_2} \tag{2.13}$$

The second solution uses Fig. 1(a) and 1(b), where $f(\mu)$ is a straight line with slope -1, i.e., in Fig. 2 we replace $f_3(\mu)$ by $-\mu + c_3$ for $\mu > \mu^*$ where c_3 is some constant. So we replace μ^* by

$$\mu^{**} = \bar{x} + \frac{1}{2c_2} + c_1 + c \tag{2.14}$$

Obviously $\mu^{**} > \mu^*$, so that the coverage (probability that μ lies within the confidence interval) is higher for the second solution, and the power is smaller. The (absolute, not relative) values of the coverages and power functions of the two solutions are unknown. Therefore we shall resort to Monte Carlo experimentation in the next section. We never experienced $\hat{f}_3(\mu^*) > c$ (also see the comment on eq. 2.6 concerning a small N resulting in a high $\hat{\mu}_3$.)

We estimate μ_3 through the unbiased estimator $\hat{\mu}_3$. Especially if N is small, $\hat{\mu}_3$ may have the wrong sign. In that case (see Fig. 3), if $f_3(\mu)$ is a polynomial with a minimum then $\hat{f}_3(\mu)$ has a maximum. The confidence interval based on $\hat{f}_3(\mu)$ still covers the sample mean (at $\mu = \bar{x}$ the slopes of $f_3(\mu)$ and $\hat{f}_3(\mu)$ equal -1).

We summarize the different procedures as follows:

(1) Student statistic t; see Fig. 1(a) and eq. (2.2).

(2) Johnson's linear statistic t_1' with σ^2 and μ_3 estimated by s^2 and $\hat{\mu}_3$; see Fig. 1(b) and eqs. (2.5) and (2.6).

(3) Johnson's quadratic statistic t_1 (again with estimated σ^2 and μ_3) which yields the 2 disjunct intervals $[\mu''_{(1)}, \mu''_{(2)}]$ and $[\mu_{(3)}, \mu''_{(4)}]$ of (2.11) if $\min [f_3(\mu)] \le -c$ and one long inter-

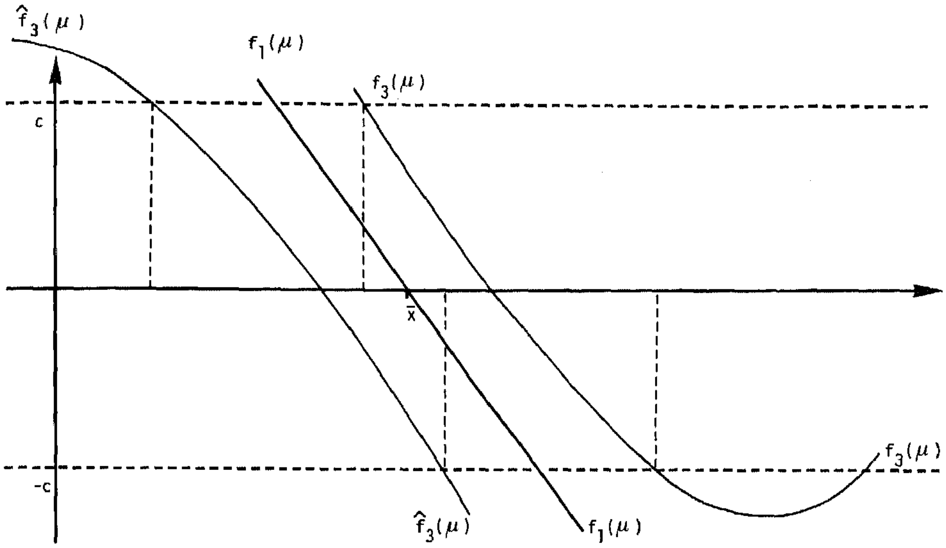


FIG. 3. Wrong sign of μ_3

val $[\mu_{(1)},\mu_{(4)}]$ if (2.12) holds; see Fig. 1(c) and Fig. 2.

(4) Johnson's quadratic statistic t_1 (with estimated σ^2 and μ_3) with elimination of the interval that does not cover the sample mean $\bar{x}$ and, if no intersection occurs, with the upper limit equal to μ^* (see Fig. 2 and eq. 2.13); obviously we halve the interval of procedure (3).

(5) We proceed as in procedure (4) but we replace μ^* by μ^{**}, using partial linearization of $f_3(\mu)$; see eq. (2.14) and the line with slope - 1 in Fig. 2.

3. DESIGNING THE MONTE CARLO EXPERIMENT

(1) Statistical procedures: We use 5 different statistics to derive a confidence interval for the mean, as listed at the end of section 2.

(ii) Random number generator: We use the standard multiplicative congruential generator for the ICL 2960 computer. This generator was developed by NAG (Numerical Algorithms Group) in England. It uses the multiplier 13^{13} and the modulus 2^{59}. The seeds are generated randomly by the computer itself, using the internal clock. All Monte Carlo results are independent (different seeds), except for the fact that each sample is analyzed through 5 different statistics which yields dependent results.

(iii) Sample size N: Johnson (1978) used N equal to 13 and 25. We pick N equal to 10, 16, 25 and 50 in the coverage study, and 10 and 25 in the more expensive power study.

(iv) α level: We select α as 0.10, 0.05 and 0.01 in the coverage study, and 0.10 and 0.05 in the power study.

(v) Type of distribution: We could have selected many asymmetric distributions. We choose the exponential and the lognormal distributions. Given the random numbers r, we sample from the exponential and the lognormal distributions, using standard procedures available on our computer. So we sample from the exponential distribution using the logarithmic transformation $-(\ln r)/\lambda$ where $\mu = \sigma = 1/\lambda$. It is well-known that if y has a normal distribution with mean μ_y and variance σ_y^2, then x = exp (y) has a lognormal distribution with mean

$$\mu_x = \exp(\mu_y + \sigma_y^2/2) \tag{3.1}$$

and variance

$$\sigma_x^2 = [\exp(2\mu_y + \sigma_y^2)][\exp(\sigma_y^2) - 1]. \tag{3.2}$$

We sample y from the normal distribution using the standard Box-Muller transformation.

Obviously changes in the exponential parameter λ do not affect the results of the various statistics. Therefore we fix

λ at the value 1. For the lognormal distribution all combinations of μ_x and σ_x with a fixed ratio μ_x/σ_x yield identical results so that we study only 3 combinations: $\sigma_x = \mu_x/3$, $\sigma_x = \mu_x$ and $\sigma_x = 3\mu_x$. The exponential might seem to be the skewest distribution since its mode is at the extreme left; actually its standardized skewness μ_3/σ^3 equals 2 whereas the lognormal has values 1, 4 and 36 for $\sigma_x = \mu_x/3$, $\sigma_x = \mu_x$ and $\sigma_x = 3\mu_x$. Johnson (1978, p. 538) selected the χ^2_2 and χ^2_{10} distributions. The χ^2_2 is identical to the exponential; obviously the χ^2_{10} is less asymmetric. So we cover more extreme forms of asymmetry.

We also apply the different statistics to the normal distribution. In that situation the assumptions of Student's statistic are satisfied, i.e., the expected coverages should equal the prespecified nominal values $1-\alpha$. We use the normal distribution not only to verify our computer program, but also to examine whether the modified t statistics with or without neglect of $(\bar{x}-\mu)^2$ work when the distribution is actually symmetric (namely Gaussian).

(vi) Number of Monte Carlo replications R: The more often we repeat the Monte Carlo experiment, the more accurate our results become. Unfortunately, Monte Carlo experimentation requires much computer time. Obviously the number of replications R needed to estimate the actual α-error within 10% with 90% probability, is

$$R = 100\ (1.6449)^2\ (1-\alpha)/\alpha. \tag{3.3}$$

Hence if α is 0.10, 0.05, 0.01 (see iv) then R is 2435, 5140, 26786 respectively. Such high R values are prohibitive, given our computer budget, so that we use R equal to 2500 to estimate coverage and R equal to 400 to estimate power functions. Fortunately the experimental noise turns out to be small relative to the systematic effects, so that we can detect certain patterns (see next section).

4. MONTE CARLO RESULTS

We do not bother the reader with the raw data of the Monte Carlo experiment. (These data were made available to the referees, and interested readers may write the authors for these details.) Instead we present the information we derived from these data (see Fig. 4, to which we shall return). These data showed that the results of procedure (4) are very close to those of procedure (5), so that we display only the version with better coverage, namely procedure (5).

4.1 Coverage: Exponential Distribution

Student's statistic t and Johnson's linear statistic t_1' give significantly low coverage, for all 12 combinations of α and N except one (N = 50; α = 0.10); we test this significance through the binomial distribution with parameters α and R and significance level 0.10 (no normal approximation). Based on Johnson's quadratic statistic we derived 3 procedures. Procedure (3) defined at the end of Section 2, has the highest coverage, but it may have 2 disjunct intervals which seems statistically unacceptable (also see the power study in subsection 4.3). Obviously the coverage of procedure (5) exceeds that of procedure (4). Procedure (5) gives significantly low coverage in 3 out of 12 cases; these significant values are not dramatically low (0.979 for α = 0.01; 0.941 for α = 0.05 and 0.889 for α = 0.10). The Monte Carlo experiment also shows how often no intersection of $\hat{f}_3(\mu)$ and $-c$ $(=-t_{N-1,\alpha/2}\, s/\sqrt{N})$ occurs. Obviously intersection occurs more often, as N and α increase. "No intersection" occurs with an estimated probability of 70% if α is 0.01 and N is 10, and 3% if α is 0.10 and N is 50. The estimate $\hat{\mu}_3$ has the wrong sign with estimated probability of 5% if N is 10 and 0% if N is 50 (evidently α has no effect).

4.2 Coverage: Lognormal Distributions

If μ equals 1 and σ equals 1/3, then the lognormal has smaller μ_3/σ^3 than the exponential has. The Monte Carlo results confirm the expectation that the coverage of Student's statistic is larger for this lognormal than it is for the exponential, especially as N and α increase. Johnson's linear statistic hardly improves the coverage. The quadratic procedure (5) gives the best coverage; only 3 out of 12 situations show significantly low coverage.

If μ and σ equal 1, then the lognormal shows $\mu_3/\sigma^3 = 4$. Student's and Johnson's linear statistics give significantly low coverages in all 12 situations. The quadratic statistic always improves the coverage; sometimes (3 out of 12) the coverage is not too low; never is the coverage dramatically low; for example, estimated coverage is 0.969 instead of 0.99 (= $1-\alpha$) if N is 10.

If μ equals 1 and σ equals 3, then μ_3/σ^3 is extreme, namely 36. Student's and Johnson's linear statistics perform very poorly. The quadratic statistics do relatively better; their coverages are significantly low, for example, if α is 0.01 and N is 10 then estimated coverage is 0.88 and if α becomes 0.10 then coverage becomes 0.77 (t and t_1' give 0.66 and 0.67).

4.3 Power

We perform a two-sided test of the hypothesis H_0: $\mu = \mu_0 = 1$ and estimate the power function at 21 values of the true μ, using 400 replications per value (and 2500 replications at μ_0). We investigate 13 combinations of different distributions, N values and α values. We display only 4 representative examples in Fig. 4. The estimated power functions of Student's and Johnson's linear statistics are so close that they cannot be distin-

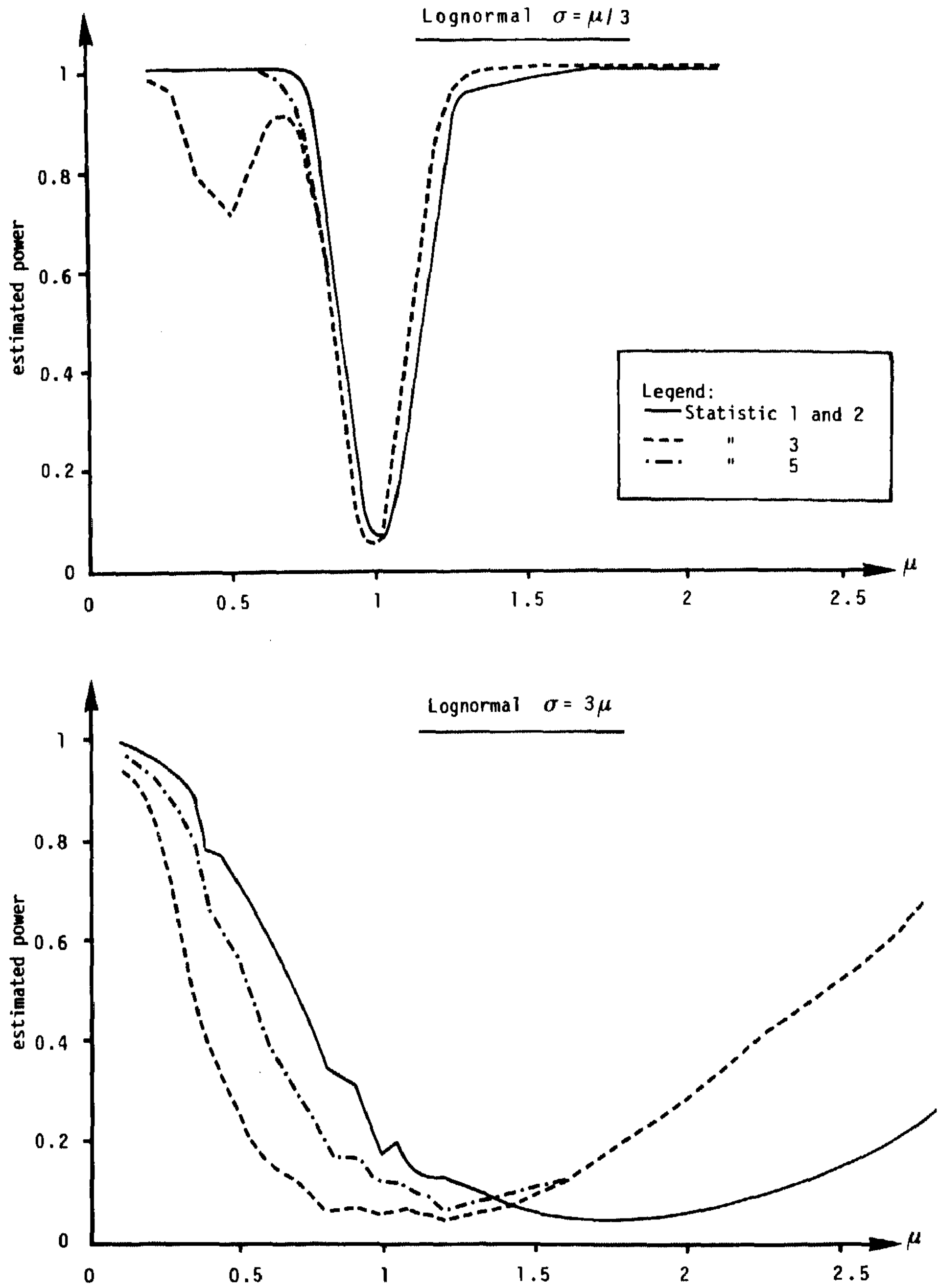
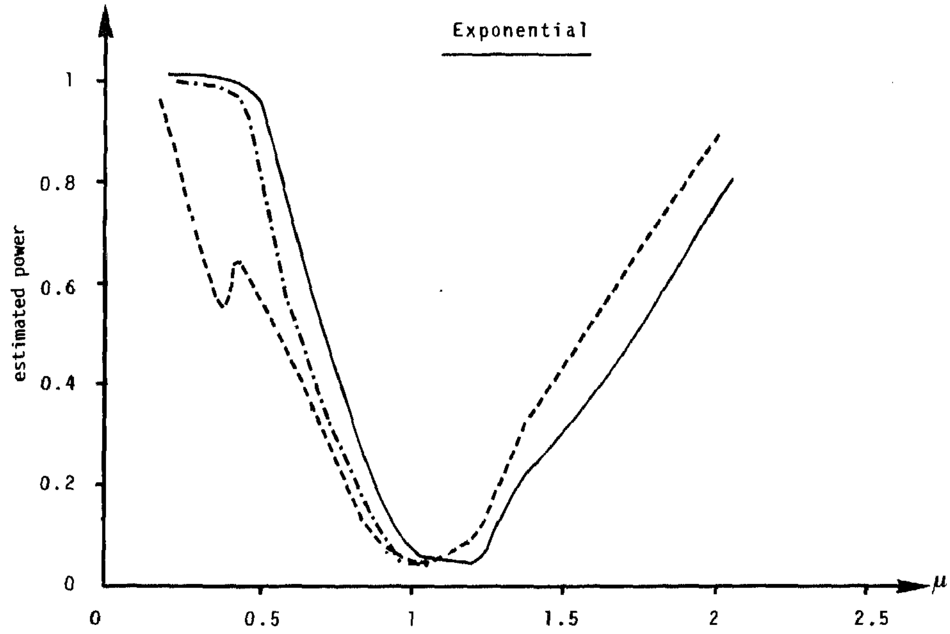


FIG. 4. Estimated power functions (N = 25; α = 0.05)



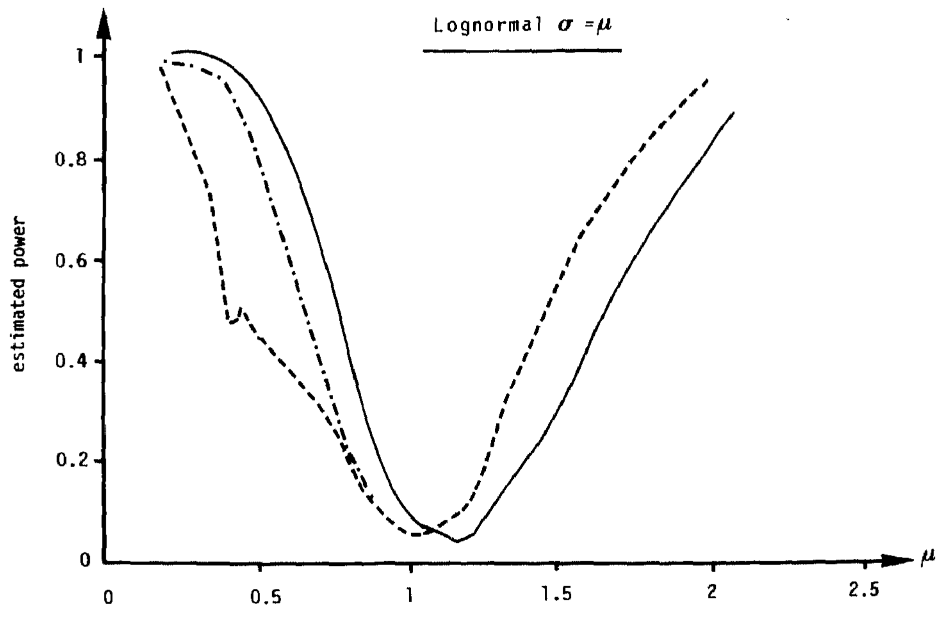


Figure 4 Continued

guished in the resulting pictures; see "statistic 1 and 2" in Fig. 4 which corresponds to the procedures (1) and (2) in section 2. The quadratic statistic with 2 disjunct intervals (statistic 3) shows curious behavior, i.e., the power shows a dip in a certain area left of μ_0. We have already mentioned that disjunct intervals seem statistically unacceptable; the dip of the power function emphasizes the statistical misbehavior of this procedure. The quadratic statistic with partial linearization (statistic 5) has an estimated power function that is very close to that of statistic 3 in certain intervals of μ; in that case Fig. 4 displays only the function for statistic 3. The estimated power function of statistic 5 reaches its minimum at μ equal to μ_0 whereas the linear statistics reach their minimum when μ exceeds μ_0. The values of these minima are roughly equal. An extremely skew lognormal distribution ($\sigma = 3\mu$)gives an estimated power function with a shape that differs from the other distributions. If the distribution is actually normal (not displayed), then all 5 procedures give nearly identical power functions. So as Johnson (1978, p. 539) has already noted, the modified t test also works if the distribution happens to be normal.

5. CONCLUSION

Many studies claim that Student's statistic t is robust (at least in the two-sided case); see Cressie (1980) and Kleijnen (1986). Yet Johnson (1978) proposed a modified t statistic. We find that his linear statistic t_1', which neglects $(\bar{x}-\mu)^2$, does not improve the t statistic. The statistic t_1, which includes the $(\bar{x}-\mu)^2$ term, gives excellent results in the exponential case while the linear statistics (t and t_1') then fail. In the lognormal cases the quadratic statistic t_1 does not give perfect results; nevertheless its coverage is quite close to the prespecified nominal value $1-\alpha$, and its results are better than the linear statistics (t and t_1'). If the distribution is actually normal, then all statistics give the desired coverage. So it is good practice to modify the classical t statistic, as proposed by Johnson (1978), provided we do include the $(\bar{x}-\mu)^2$ term; the price we pay is a slight increase in computation which, however, is negligible when using a (micro) computer. If 2 disjunct confidence intervals result, then the interval not covering the sample mean $\bar{x}$, should be eliminated (see Fig. 1c and eqs. 2.8 and 2.10). If the mathematical solution of the second-degree polynomials gives a single interval, then that interval should be halved (see μ^* in Fig. 2) or, still better, reduced by partial linearization (see μ^{**} in Fig. 2 and eq. 2.14).

ACKNOWLEDGEMENT

The authors wish to thank H. Moors (statistician in the Department of Econometrics) for bringing Johnson's article to their attention, C. Cuypers and J. Van Rooijen (graduate students) for preliminary research on Johnson's test, P. Nederstigt (Dutch State Mines) for suggesting the elimination of 1 of the 2 disjunct intervals, and E. Crow (associate editor/U.S. Department of Commerce) for comments leading to a better presentation.

REFERENCES

CRESSIE, N. (1980), "Relaxing assumptions in the one sample t-test", Australian Journal of Statistics, 22, 143-153.

JOHNSON, N.J. (1978), "Modified t Tests and Confidence Intervals for Asymmetric Populations", Journal of the American Statistical Association, 73, 536 - 544.

KENNEY, J.F. AND KEEPING, E.S. (1954), Mathematics of Statistics, Part One, D. Van Nostrand, New York.

KLEIJNEN, J.P.C. (1986), Statistical Tools for Simulation Practitioners, Marcel Dekker, New York.

Received by Editorial Board member, September, 1985; Revised April, 1986.

Recommended by Edwin L. Crow, U. S. Department of Commerce, Boulder, CO.

Refereed Anonymously.