THE DENSITY OF THE t-STATISTIC FOR NON-NORMAL DISTRIBUTIONS

by

Raymond Clayton Sansing

Technical Report No. 99
Department of Statistics ONR Contract

April 6, 1971

Research sponsored by the Office of Naval Research
Contract N00014-68-A-0515
Project NR 042-260

Reproduction in whole or in part is permitted
for any purpose of the United States Government.

This document has been approved for public release
and sale; its distribution is unlimited.

DEPARTMENT OF STATISTICS
Southern Methodist University

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

Although there is no general agreement on a strict definition of robustness, most writers implicitly accept a qualitative definition. A general statement of this definition is as follows: a statistical test is said to be robust (with respect to an underlying assumption for some class of alternatives to the assumption and for a fixed sample size) if the power function of the test (under any member of the class of alternative assumptions) is not excessively larger than the power function of the test under the original assumption for parameter values where the null hypothesis is true and not excessively smaller where the alternative hypothesis is true. A great many tests are based on an assumption of normality of the parent population and robustness with respect to non-normal parent populations has received by far the most study.

Classically, tests are formulated such that a type I error, rejecting the null hypothesis when it is true, is the most critical error and probabilities of this type of error are strictly controlled. Very often the sample size is controlled by physical considerations and, hence, the probability of the other type of error is beyond the control of the experimentor. For these reasons, changes in the probability of a type I error for changes in the assumptions are of more importance than other points on the power function and most robustness studies have been limited to the null hypothesis point on the power function. It should be observed that the central issue is the distribution of test statistic under the alternative assumptions being considered. The specific concern in this work is the density of the one sample t-statistic without the assumption of normality of the parent population.

An annotated bibliography of robustness studies in general has been given by Govindarajulu and Leslie [8]. A survey of robustness studies of the Student t-tests, both one sample and two sample, has been given by Hatch and Posten [9]. We will adopt the convention that references to the Student t-statistic or test have the underlying assumption of normality of the parent population, while references to the t-statistic or test include no such assumption.

Surprisingly little has been accomplished in deriving the exact density or mass function of the t-statistic. Rider [15] derived the density for samples of size 2 for a uniform population as well as the mass function for various discrete uniform samples of size 2, 3, and 4. Perlo [14] has given the density of the t-statistic for samples of size 3 from a uniform parent population. Geary [7] derived the t-density for double exponential samples and Baker [1] treated the compound normal case with equal variances; both for sample size 2. Hotelling [10] derived the tails of the t-density for samples of size 2 from a Cauchy parent population. Laderman [11] derived the t-density for samples of size 2 from an arbitrary density with mean zero by geometric arguments. His result is derived analytically here without the assumption about the mean but for parent densities positive on the entire real line and can easily be derived for the other cases with results presented here.

Various approximations for the t-density have been given. Bartlett [2], Geary [7], Gayen [6] and others have used the first few terms of an

Edgeworth type A series or a Gram-Charlier series as the density of the parent population and proceeded to derive the associated t-density. A thorough account of many such works is given by Hatch and Posten [9]. Bradley [4] worked in quite a different way. He wrote the distribution function of the t-statistic as an integral of the joint density of the observation over the appropriate subset of Euclidean n-space, then manipulated the n-fold integral. After making simplifying assumptions about the parent density, similar to those made in this work, he approximated the t-density with the first few terms of a series representation for it. He developed a computational technique and illustrated it with Cauchy and logistic parent populations for sample sizes of 2, 3, 4 and 5.

The general approach taken here is to recursively derive the joint density of the sample mean and sum of squares of deviation about the sample mean. For $n \geq 3$, the recursion relation requires an integration which is accomplished by application of the mean value theorem. This technique produces exact results for only a certain class of functions but gives an approximation for others. This type of application of the mean value theorem to carry out integrals promises to be a powerful statistical technique with further study, which is indicated in Appendix A. A transformation from this density to the density of the t-statistic is given here, along with a symmetry property for the t-density.

The formulas derived here are illustrated when a member of the generalized normal family is the parent density. The approximation of the t-density is derived for these parent densities for the case where $\mu = 0$ and for all sample sizes. Tables of type I error probabilities are given for several specific members of this family for sample sizes $2, 3, \cdots, 31$.

CHAPTER II

THE RECURSION RELATION

Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of independent identically distributed random variables with density function f where $f(x) > 0$, a.e. Let

$$T_1^{(n)} = n^{-1} \sum_{i=1}^{n} X_i \quad \text{and} \quad T_2^{(n)} = \sum_{i=1}^{n} [X_i - T_1^{(n)}]^2 \quad .$$

The recursion relations

$$\begin{aligned} T_1^{(n+1)} &= \frac{n}{n+1} T_1^{(n)} + \frac{1}{n+1} X_{n+1} \\ T_2^{(n+1)} &= T_2^{(n)} + \frac{n}{n+1} [T_1^{(n)} - X_{n+1}]^2 \end{aligned} \tag{2.1}$$

can be verified directly. The superscript on T_1 and T_2 will be suppressed where there is no ambiguity.

Suppose $f_n(t_1, t_2)$ is the joint density of $T_1^{(n)}$ and $T_2^{(n)}$. X_{n+1} is independent of $(X_1, X_2, \cdots, X_n)$ and hence, is independent of $T_1^{(n)}$ and $T_2^{(n)}$. Therefore, the joint density of $T_1^{(n)}$, $T_2^{(n)}$ and X_{n+1} is $f_n(t_1, t_2) f(x_{n+1})$. Using the relations in (2.1), we can transform $[T_1^{(n)}, T_2^{(n)}, X_{n+1}] \rightarrow [T_1^{(n+1)}, T_2^{(n+1)}, U]$ with the auxiliary variable being defined by $U = T_1^{(n)} - X_{n+1}$. The inverse of this transformation is

$$T_1^{(n)} = T_1^{(n+1)} + \frac{1}{n+1} U$$

$$T_2^{(n)} = T_2^{(n+1)} - \frac{n}{n+1} U^2$$

$$X_{n+1} = T_1^{(n+1)} - \frac{n}{n+1} U$$

and the jacobian is $J = -1$. Then the density of $[T_1^{(n+1)}, T_2^{(n+1)}, U]$ is

$$f_n\left(t_1 + \frac{1}{n+1} u, t_2 - \frac{n}{n+1} u^2\right) f\left(t_1 - \frac{n}{n+1} u\right) ,$$

$$-\infty < t_1 < \infty , t_2 > 0 ,$$

and the density of $[T_1^{(n+1)}, T_2^{(n+1)}]$ is

$$f_{n+1}(t_1, t_2) = \int_S f_n\left(t_1 + \frac{1}{n+1} u, t_2 - \frac{n}{n+1} u^2\right) f\left(t_1 - \frac{n}{n+1} u\right) du ,$$

$$-\infty < t_1 < \infty , t_2 > 0 ,$$

where S is the open interval

$$S = \left(- \sqrt{(n+1)t_2/n} , \sqrt{(n+1)t_2/n}\right) .$$

In order to make the range of integration independent of t_2 and to put the recusion relation in a more useful form we can transform

$$u = \sqrt{(n+1)t_2/n} \; v .$$

Then the joint density of $T_1^{(n+1)}$ and $T_2^{(n+1)}$ is

$$f_{n+1}(t_1,t_2) = \sqrt{\frac{n+1}{n}\,t_2}\int_{-1}^{1}\left[f_n\left(t_1 + u\sqrt{\frac{t_2}{n(n+1)}}\,,\ t_2(1-u^2)\right)f\left(t_1 - u\sqrt{\frac{nt_2}{n+1}}\right)\right]du,$$

$$-\infty < t_1 < \infty\ ,\ t_2 > 0\ . \qquad (2.2)$$

CHAPTER III

APPLICATION OF THE RECURSION RELATION

For $n = 2$, the sample mean and sum of squares of deviation can be expressed as $T_1 = \frac{1}{2}(X_1 + X_2)$ and $T_2 = \frac{1}{2}(X_1 - X_2)^2$. The transformation is not 1-1 and the sample space must be broken down into the subspaces

$$A_1 = \left\{(X_1, X_2) \mid X_1 \geq X_2\right\}$$

$$A_2 = \left\{(X_1, X_2) \mid X_1 < X_2\right\} .$$

On A_1, the inverse transformation is

$$X_1 = T_1 + \sqrt{T_2/2} \quad , \quad X_2 = T_1 - \sqrt{T_2/2}$$

with $J_1 = -1/\sqrt{2T_2}$. On A_2, the inverse transformation is

$$X_1 = T_1 - \sqrt{T_2/2} \quad , \quad X_2 = T_1 + \sqrt{T_2/2}$$

with $J_2 = 1/\sqrt{2T_2}$. Then

$$f_2(t_1, t_2) = \sqrt{2}\, t_2^{-1/2} f\left(t_1 + \sqrt{\frac{t_2}{2}}\right) f\left(t_1 - \sqrt{\frac{t_2}{2}}\right) ,$$

$$-\infty < t_1 < \infty \ , \ t_2 > 0 \ . \tag{3.1}$$

The only changes required in (3.1) for parent densities that are positive on (a, b) or $(0, \infty)$ are changes in the limits on t_1 and t_2. For densities positive on (a, b), $a < t_1 < b$ and $0 < \sqrt{t_2} < \sqrt{2}\min(t_1-a, b-t_1)$. For densities positive on $(0, \infty)$, $0 < t_1 < \infty$ and $0 < \sqrt{t_2} < \sqrt{2}\, t_1$. These results are also proven by Craig [5].

Applying the relation (2.2) we have

$$f_3(t_1, t_2) = \sqrt{3}\int_{-1}^{1}\left\{(1-u^2)^{-1/2} f\left(t_1 + \sqrt{t_2}\left[\frac{u + \sqrt{3(1-u^2)}}{\sqrt{6}}\right]\right)\right.$$

$$\left. \times f\left(t_1 + \sqrt{t_2}\left[\frac{u - \sqrt{3(1-u^2)}}{\sqrt{6}}\right]\right) f\left(t_1 + \sqrt{t_2}\left[-\frac{2u}{\sqrt{6}}\right]\right)\right\} du \ . \tag{3.2}$$

Let

$$\begin{aligned} a_{13}(u) &= \frac{u}{\sqrt{6}} + \frac{1}{\sqrt{2}}\sqrt{1-u^2} \\ a_{23}(u) &= \frac{u}{\sqrt{6}} - \frac{1}{\sqrt{2}}\sqrt{1-u^2} \\ a_{33}(u) &= -u\sqrt{\frac{2}{3}} \end{aligned} \tag{3.3}$$

then we can rewrite $f_3(t_1, t_2)$ in the more compact form

$$f_3(t_1, t_2) = \sqrt{3}\int_{-1}^{1} (1-u^2)^{-1/2} \prod_{i=1}^{3} f\left[t_1 + \sqrt{t_2}\, a_{i3}(u)\right] du \ . \tag{3.4}$$

If we further restrict f to be everywhere continuous, the integrand is the product of compositions of continuous functions on the interval $(-1, 1)$. Hence, the integrand is a continuous function on $(-1, 1)$ that is unbounded at the end points. Then for each fixed (t_1, t_2) we can apply the modified mean value theorem for integrals, a statement and proof of which is given in Appendix A, to say there exists $\xi_3 \varepsilon (-1, 1)$ such that

$$f_3(t_1, t_2) = 2\sqrt{3}\ (1 - \xi_3^2)^{-1/2} \prod_1^3 f\left[t_1 + \sqrt{t_2}\ a_{i3}(\xi_3)\right] , \qquad (3.5)$$

where $\sum_1^3 a_{i3}(\xi_3) = 0$ and $\sum_1^3 a_{i3}^2(\xi_3) = 1$ are easily verified identities in ξ_3, as well as the fact that $-1 < a_{i3} < 1$ for $i = 1, 2, 3$.

The existence of a value ξ_3 is guaranteed for each (t_1, t_2), hence, $\xi_3 = \xi_3(f, t_1, t_2)$. The part of the integrand involving f in (3.4) can be written

$$\prod_1^3 f\left(\sqrt{t_2}\left[\frac{t_1}{\sqrt{t_2}} + a_{i3}(u)\right]\right) .$$

For densities that are symmetric about zero and sufficiently smooth, $t_1/\sqrt{t_2}$ large moves the range of integration out in the tail of the density and the multiplicative factor tends to play a smaller role as the ratio increases. Then for smooth densities symmetric about zero $\xi_3 \approx \xi_3\left[f, \left|t_1/\sqrt{t_2}\right|\right]$. Also, for densities that can be written with a scale parameter, the scale can be made large and the mass concentrated about zero so that the ratio $t_1/\sqrt{t_2}$ is always large and hence, $\xi_3 \approx \xi_3(f)$ since $T_1/\sqrt{T_2}$ is a scale invariant random variable. In any case, we will treat ξ_3 as a constant and derive the values of the coefficients $\{a_{i3}\}$ from other considerations. Then the following results will be exact only for the class of density functions where ξ_3 is independent of t_1 and t_2 . For the class of densities where ξ_3 is not independent of t_1

and t_2 results that follow will yield approximations for the densities of (T_1, T_2) and T .

Suppose we have recursively derived, for all integers up to and including n,

$$f_n(t_1, t_2) = 2^{n-2} \sqrt{n} \left\{ \prod_3^n (1-\xi_i^2)^{\frac{i-4}{2}} \right\} \left\{ t_2^{\frac{n-3}{2}} \prod_1^n f(t_1 + \sqrt{t_2}\, a_{in}) \right\} ,$$

$$-\infty < t_1 < \infty \ , \ t_2 > 0 \tag{3.6}$$

where

$$a_{in} = \frac{\xi_n}{\sqrt{n(n-1)}} + a_{i,n-1} \sqrt{1-\xi_n^2} \quad , \quad i = 1, 2, \cdots , n-1$$

$$a_{nn} = -\, \xi_n \sqrt{\frac{n-1}{n}}$$

$$\sum_1^n a_{in} = 0 \quad , \quad \sum_1^n a_{in}^2 = 1 \ . \tag{3.7}$$

And for each $i \leq n$, ξ_i is the constant whose existence is guaranteed by the mean value theorem. Applying the recursion relation (2.2), we have

$$f_{n+1}(t_1, t_2) = 2^{n-2} \sqrt{n+1} \left(\prod_3^n (1-\xi_i^2)^{\frac{i-4}{2}} \right) t_2^{\frac{n-2}{2}} \int_{-1}^{1} \left\{ (1-u^2)^{\frac{n-3}{2}} \right.$$

$$\left. \times \prod_1^n f\left(t_1 + \sqrt{t_2} \left[\frac{u}{\sqrt{n(n+1)}} + a_{in} \sqrt{1-u^2} \right] \right) f\left(t_1 + \sqrt{t_2} \left[-u \sqrt{\frac{n}{n+1}} \right] \right) \right\} du ,$$

$$-\infty < t_1 < \infty \ , \ t_2 > 0 \ . \tag{3.8}$$

Then we define

$$a_{i,n+1}(u) = \frac{u}{\sqrt{n(n+1)}} + a_{in}\sqrt{1-u^2} \quad , \; i = 1, 2, \cdots , n$$

$$a_{n+1,n+1}(u) = -u\sqrt{\frac{n}{n+1}} \tag{3.9}$$

and rewrite (3.8) as

$$f_{n+1}(t_1, t_2) = 2^{n-2}\sqrt{n+1}\left\{\prod_{3}^{n}(1-\xi_i^2)^{\frac{i-4}{2}}\right\} t_2^{\frac{n-2}{2}}$$

$$\times \int_{-1}^{1}\left\{(1-u^2)^{\frac{n-3}{2}} \prod_{1}^{n+1} f[t_1 + \sqrt{t_2}\, a_{i,n+1}(u)]\right\} du \; ,$$

$$-\infty < t_1 < \infty \; , \; t_2 > 0 \quad .$$

Again the mean value theorem can be applied to provide the existence of $\xi_{n+1} \varepsilon [-1, 1]$ which is assumed to be independent of t_1 and t_2, using the conditions stated following equation (3.5). The identities $\sum_{1}^{n} a_{i,n+1} = 0$ and $\sum_{1}^{n} a_{i,n+1}^2 = 1$ can be verified directly.

Then by the strong principal of finite induction, the density of (T_1, T_2) is given by (3.6) and the recursion relation for the coefficients $\{a_{in}\}$ in terms of the constants $\{\xi_i\}$ is given by (3.7) for all values of $n \geq 3$. If we take vacuous products to be 1, (3.6) is valid for $n \geq 2$.

CHAPTER IV

THE DENSITY OF THE t-STATISTIC

The appropriate form of the t-statistic is

$$T = \frac{\sqrt{n}\, T_1}{\sqrt{\frac{T_2}{n-1}}} = \sqrt{n(n-1)}\; T_1 T_2^{-1/2} \quad .$$

Using the auxiliary random variable $U = \sqrt{T_2}$, we are transforming $(T_1, T_2) \to (T, U)$. The inverse of the transformation is

$$T_1 = [n(n-1)]^{-1/2}\; U\, T$$

$$T_2 = U^2$$

and the jacobian is $J = [n(n-1)]^{-1/2}\, 2U^2$. Then the density of the t-statistic is

$$Q_n(t) = \frac{2^{n-1}}{\sqrt{n-1}} \prod_{i=3}^{n} (1-\xi_i^2)^{\frac{i-4}{2}} \int_0^\infty u^{n-1} \prod_{i=1}^{n} f\left(\left[\frac{t}{\sqrt{n(n-1)}} + a_{in}\right] u\right) du \quad ,$$

$$-\infty < t < \infty \quad , \tag{4.1}$$

where $\{\xi_i\}$ and $\{a_{in}\}$ are defined in the previous chapter.

$Q_n(t)$ is an even function whenever f is an even function. This will be verified by the fact that for every $n \geq 2$, $f_n(t_1, t_2) = f_n(-t_1, t_2)$ for all t_1 and t_2, which will first be established by induction. Since f is an even function, the property holds for n = 2 from (3.1). Supposing the property holds for n, applying (2.2), we have for all t_1 and t_2

$$f_{n+1}(-t_1, t_2) = \sqrt{\frac{n+1}{n}\, t_2} \int_{-1}^{1} f_n\left(t_1 - u\sqrt{\frac{t}{n(n+1)}}\,,\ t_2(1-u^2)\right) f\left(t_1 + u\sqrt{\frac{nt_2}{n+1}}\right) du \; .$$

Then with the change of variable $u \to -u$, we see the property holds for n+1 and hence the property holds for $n \geq 2$.

The density of $(T_1, \sqrt{T_2}) = (X, Y)$ is $2yf_n(x, y^2)$ and we have, since $Y > 0$,

$$\Pr\left\{\frac{X}{Y} \leq -t\right\} = \Pr\left\{X \leq -tY\right\} = \int_0^{\infty}\int_{-\infty}^{-ty} 2yf_n(x, y^2)\,dxdy \quad .$$

Then using the symmetry property of f_n and making the change of variable $x \to -x$, we have

$$\Pr\left\{\frac{X}{Y} \leq -t\right\} = \Pr\left\{\frac{X}{Y} \geq t\right\} \quad , \text{ for all } t \quad .$$

Hence, $T_1/\sqrt{T_2}$ is a symmetric random variable and Q_n is an even function.

Comparing $Q_n(t)$ and $Q_n(-t)$ in (4.1), one implication of this symmetry might be that $\{a_{in} | i = 1, 2, \cdots , n\} = \{-a_{in} | i = 1, 2, \cdots , n\}$ for some f's. When this implication is not true, the set $\{a_{in}\}$ that is symmetric about zero could be used to approximate the t-density with Q_n normalized to make it a density function, since the symmetric set of coefficients does make Q_n an even function. This approximation could also be applied to parent densities that are not even functions. The coefficient set $\{a_{in}\}$ that is symmetric is dispersed on the interval (-1, 1) and hence would be a reasonable approximation of the true coefficient set in the absence of additional information about the true set.

CHAPTER V

DETERMINATION OF SYMMETRIC $\{a_{in}\}$

By the relation (3.7), we can see that the coefficient set $\{a_{in} | i = 1, 2, \cdots , n\}$ has the property $a_{jn} - a_{in} = (a_{j,n-1} - a_{i,n-1})\sqrt{1-\xi_n^2}$ for $i,j = 1, 2, \cdots , n-1$ and hence any ordering of $\{a_{i,n-1} | i = 1, 2, \cdots , n-1\}$ must give the same ordering of $\{a_{in} | i = 1, 2, \cdots n-1\}$. When we have solved for ξ_{n-1} and $\{a_{i,n-1}\}$ we will order $\{a_{i,n-1}\}$ and then we will have only three considerations to determine the position of a_{nn} in the order. For $n = 2k+1$, the $(k+1)^{st}$ value in the completely ordered set must be zero, i.e., $a_{k,n} = 0$, $a_{k+1,n} = 0$ or $a_{n,n} = 0$. For $n = 2k$, the extreme values in the completely ordered set must differ exactly in sign, i.e., $a_{1n} = -a_{n-1,n}$, $a_{1n} = -a_{nn}$, or $a_{nn} = -a_{n-1,n}$.

For $n = 3$, the coefficients are given, as functions of ξ_3, by (3.3) and the conditions above yield the solutions $\xi_3 = 0, -\sqrt{3}/2, \sqrt{3}/2$. All three values of ξ_3 yield the coefficient set $\{a_{i3}\} = \left\{1/\sqrt{2}, 0, -1/\sqrt{2}\right\}$. We can discriminate between $\xi_3^2 = 0$ or $3/4$ by requiring $Q_3(t)$ to be a density function. For $n = 4$, the coefficients are given by (3.7) and the conditions above yield $\xi_4 = 0, -\sqrt{3/5}, \sqrt{3/5}$. The values $\xi_4 = \sqrt{3/5}, -\sqrt{3/5}$ yield the coefficient set $\left\{3/\sqrt{20}, 1/\sqrt{20}, -1/\sqrt{20}, -3/\sqrt{20}\right\}$ and the value $\xi_4 = 0$ yields the coefficient set $\left\{1/\sqrt{2}, 0, 0, -1/\sqrt{2}\right\}$. We will assume that $\xi_n \neq 0$, $n \geq 4$, and handle the other cases later. By inspection of these cases, we can set up the induction hypotheses

$$a_{in} = \frac{\sqrt{3}\ (n-2i+1)}{\sqrt{(n-1)n(n+1)}}\ , \quad i = 1, 2, \cdots, n$$

$$\xi_k^2 = \frac{3}{k+1}\ , \quad k = 3, 4, \cdots, n \quad . \tag{5.1}$$

The coefficients $\{a_{i,n+1}\}$ are given by (3.9) but as functions of ξ_{n+1}. Considering the cases n, an even and an odd integer, separately, the conditions yield the solutions $\xi_{n+1} = 0, \sqrt{3/(n+2)}\ , -\sqrt{3/(n+2)}$ and for $\xi_{n+1} \neq 0$, the coefficient set

$$a_{i,n+1} = \frac{\sqrt{3}\ (n-2i+2)}{\sqrt{n(n+1)(n+2)}}\ , \quad i = 1, 2, \cdots, n$$

$$a_{n+1,n+1} = \begin{cases} \dfrac{\sqrt{3}\ (n)}{\sqrt{n(n+1)(n+2)}}\ , & \xi_{n+1} = -\sqrt{\dfrac{3}{n+2}} \\[2ex] \dfrac{-\sqrt{3}\ (n)}{\sqrt{n(n+1)(n+2)}}\ , & \xi_{n+1} = \sqrt{\dfrac{3}{n+2}} \end{cases} .$$

In either case, the induction hypotheses (5.1) are verified for the case where $\xi_n \neq 0,\ n \geq 4$. When $\xi_n \neq 0,\ n \leq n_0$ and $\xi_n = 0,\ n > n_0$, the coefficients for n_0 are duplicated with $n-n_0$ additional values of zero. Additional cases, where the ξ's are zero with some irregular spacing, can be handled recursively.

CHAPTER VI

THE GENERALIZED NORMAL DISTRIBUTION

The generalized normal distribution has been considered by various writers in connection with robustness studies. The density is

$$f(x|\mu, \sigma, \beta) = \left\{ 2^{\frac{3+\beta}{2}} \Gamma\left(\frac{3+\beta}{2}\right)\sigma \right\}^{-1} \exp\left\{ -\frac{1}{2} \left| \frac{x-\mu}{\sigma} \right|^{\frac{2}{1+\beta}} \right\} ,$$

$$-\infty < x < \infty \ , \ -1 < \beta \leq 1 \ , \ -\infty < \mu < \infty \ , \ \sigma > 0 \ .$$

Due to the scale invariance of T, we will consider $\sigma = 1$, without loss of generality. We will further consider the null case where $\mu = 0$, and f is an even function. Applying (4.1) we have

$$Q_n(t|\beta) = c_n(\beta) \left\{ \sum_1^n \left| \frac{t}{\sqrt{n(n-1)}} + a_{in} \right|^{\frac{2}{1+\beta}} \right\}^{-n\left(\frac{1+\beta}{2}\right)} , \qquad (6.1)$$

where

$$c_n(\beta) = \frac{(1+\beta)\Gamma\left[n\left(\frac{1+\beta}{2}\right)\right] \prod_3^n (1-\xi_i^2)^{\frac{i-4}{2}}}{4\sqrt{n-1}\ \Gamma^n\left(\frac{3+\beta}{2}\right)} .$$

Since Q_n must be an even function, we also have

$$Q_n(t|\beta) = c_n(\beta) \left\{ \sum_1^n \left| \frac{t}{\sqrt{n(n-1)}} - a_{in} \right|^{\frac{2}{1+\beta}} \right\}^{-n\left(\frac{1+\beta}{2}\right)} . \qquad (6.2)$$

Many of the robustness studies that have been made have led people to conclude that kurtosis of the parent population is the primary factor determining the usefulness of using the Student t-distribution as an approximation of the density of the t-statistic for the population under consideration. This generalized normal density not only is a one parameter variation of the normal density, but that one parameter is also a kurtosis parameter. The kurtosis of the generalized normal is

$$\kappa = \frac{\Gamma\left(\frac{1+\beta}{2}\right)\Gamma\left[5\left(\frac{1+\beta}{2}\right)\right]}{\Gamma^2\left[3\left(\frac{1+\beta}{2}\right)\right]} ,$$

which is an increasing function of β .

The normal density corresponds to $\beta = 0$. With this value we can simplify (6.1) and compare the result with the normal theory t-density to see that

$$\prod_{3}^{n} (1-\xi_i^2)^{\frac{i-4}{2}} = \frac{\pi^{\frac{n-1}{2}}}{2^{n-2}\Gamma\left(\frac{n-1}{2}\right)}$$

and hence, for $n \geq 3$,

$$(1-\xi_n^2)^{\frac{n-4}{2}} = \frac{\sqrt{\pi}\,\Gamma\left(\frac{n-2}{2}\right)}{2\Gamma\left(\frac{n-1}{2}\right)}$$

gives a sequence of values $\{\xi_n\}$ satisfying the conditions of Chapter III. Therefore, the normal density is a member of the class of densities where $Q_n(t)$ is the exact density of the t-statistic.

When $2/(1+\beta)$ is an even integer > 2, we can expand $\sum\left|\frac{t}{\sqrt{n(n-1)}} + a_{in}\right|^{2/1+\beta}$ by the binomial theorem. Since Q_n is an even function, the coefficients of all odd powers of t must be zero. The solutions of $\sum_1^n a_{in}^3 = 0$ are $\xi_n = 0, \pm\sqrt{\frac{3}{n+1}}$ which are the values that produce the symmetric constants discussed in the previous chapter. Clearly, these symmetric constants have the property that $\sum_1^n a_{in}^{2k+1} = 0$ for $k = 0, 1, 2, \cdots$. When $2/(1+\beta) = 2k+1$, we can compare (6.1) and (6.2) to see that

$$\sum_1^n \left| \frac{t}{\sqrt{n(n-1)}} + a_{in} \right|^{2k+1} = \sum_1^n \left| \frac{t}{\sqrt{n(n-1)}} - a_{in} \right|^{2k+1} \tag{6.3}$$

is an identity in t . For $k = 0$, we can go through a tedious process of letting $t/\sqrt{n(n-1)}$ take values between the a_{in}'s to show that the set of coefficients $\{a_{in}\}$ must be symmetric. For $k \geq 1$, we can let $t/\sqrt{n(n-1)} > \max_i\{|a_{in}|\}$ and expand the terms by the binomial theorem. Comparing terms in (6.3) we can again see that $\sum_1^n a_{in}^{2k+1} = 0$ for $k = 1, 2, \cdots$, and hence that the symmetric set $\{a_{in}\}$ must be the proper one. Then the symmetric set of constants must apply for all $\beta \neq 0$ such that $2/(1+\beta)$ is an integer.

The set of symmetric coefficients was used in (4.1) for the values of β where $2/(1+\beta) = 1, 4/3, 3/2, 7/3, 5/2, 4, 16$. The appropriate values of $\xi_3^2 = 0, 3/4$ where used to numerically integrate $Q_n(t|\beta)$, since this must be a density function. Neither of the values gave a density function. This shows that $f(x|0, 1, \beta)$ is not in the class of parent densities for which (4.1) is the exact t-density, but (4.1) does represent an approximation for this density.

The set $\{\xi_n\}$, with the associated set of coefficients $\{a_{in}\}$, was fitted sequentially for the values of β given above in such a way as to make the integral of $Q_n(t|\beta)$ closest to 1 . The sample sizes considered were n = 2, 3, $\cdots$, 31 and, of course, for n = 2 the density of $Q_2(t|\beta)$ is exact for all β . The set $\{\xi_n\}$ used was $\xi_n^2 = 3/(n+1)$ except for the β's where $2/(1+\beta) < 2$ and in this case, $\xi_7 = \xi_{10} = \xi_{13} = \xi_{19} = \xi_{29} = 0$ were the only changes.

The tables that follow are tables of

$$\tilde{\alpha}(n, \beta) = \frac{\int_{t_\alpha(n-1)}^{\infty} Q_n(t|\beta)\,dt}{\int_{-\infty}^{\infty} Q_n(t|\beta)\,dt} , \qquad (6.4)$$

where $t_\alpha(n-1)$ is the critical point of the Student t-density for a normal parent population. The $t_\alpha(n-1)$ were taken from [12], for the five decimal place accuracy, where available and from [13], otherwise. $\tilde{\alpha}$ is interpreted as the true probability of a type I error under the approximation $Q_n(t|\beta)$ when α was the advertised probability and the false assumption of normality was utilized. The computer program that was used to carry out the calculations is listed and explained in Appendix B.

TABLE 1

$$\frac{2}{1+\beta} = 1$$

n-1	α=.100	α=.050	α=.025	α=.010	α=.005
1	.08129	.03962	.01969	.00786	.00393
2	.08885	.03705	.01707	.00651	.00321
3	.10313	.04453	.01801	.00620	.00291
4	.11142	.05186	.02237	.00675	.00296
5	.11945	.05812	.02620	.00831	.00336
6	.09308	.04741	.02211	.00778	.00341
7	.11259	.05742	.02952	.01143	.00497
8	.12572	.07042	.03744	.01439	.00647
9	.09688	.05148	.02760	.01103	.00529
10	.11456	.06460	.03424	.01440	.00698
11	.13114	.07277	.04011	.01692	.00872
12	.10234	.05411	.02889	.01214	.00615
13	.11645	.06465	.03498	.01544	.00827
14	.12540	.06987	.03948	.01828	.00966
15	.13046	.07618	.04405	.02012	.01085
16	.14016	.08252	.04719	.02175	.01185
17	.14570	.08531	.04888	.02301	.01268
18	.11893	.06603	.03672	.01651	.00904
19	.12571	.07091	.03978	.01857	.01003
20	.12967	.07412	.04259	.01961	.01069
21	.13223	.07695	.04360	.02015	.01112
22	.13677	.07853	.04467	.02102	.01169
23	.13763	.07917	.04562	.02161	.01219
24	.13855	.08067	.04667	.02250	.01273
25	.13911	.08134	.04762	.02304	.01322
26	.14047	.08307	.04870	.02395	.01372
27	.14290	.08472	.05021	.02466	.01429
28	.11954	.06776	.03850	.01822	.01027
29	.12415	.07087	.04079	.01947	.01105
30	.12768	.07392	.04263	.02048	.01155

TABLE 2

$$\frac{2}{1+\beta} = \frac{4}{3}$$

n-1	α=.100	α=.050	α=.025	α=.010	α=.005
1	.08936	.04395	.02188	.00874	.00437
2	.09323	.04220	.02006	.00778	.00385
3	.10125	.04637	.02074	.00759	.00365
4	.10652	.05066	.02314	.00799	.00370
5	.11082	.05425	.02542	.00890	.00398
6	.09771	.04866	.02330	.00855	.00393
7	.10771	.05458	.02734	.01049	.00483
8	.11485	.06063	.03100	.01199	.00562
9	.10044	.05165	.02646	.01045	.00504
10	.10926	.05801	.02994	.01214	.00593
11	.11717	.06237	.03276	.01344	.00673
12	.10336	.05355	.02766	.01130	.00565
13	.11010	.05839	.03058	.01283	.00657
14	.11496	.06153	.03288	.01412	.00725
15	.11833	.06462	.03496	.01505	.00780
16	.12260	.06741	.03647	.01580	.00825
17	.12543	.06899	.03743	.01637	.00860
18	.11317	.06049	.03219	.01378	.00719
19	.11601	.06251	.03351	.01457	.00761
20	.11806	.06409	.03471	.01510	.00794
21	.11965	.06548	.03543	.01549	.00819
22	.12157	.06644	.03605	.01590	.00846
23	.12248	.06709	.03661	.01624	.00870
24	.12322	.06783	.03716	.01661	.00893
25	.12382	.06839	.03766	.01690	.00914
26	.12460	.06914	.03817	.01724	.00934
27	.12565	.06990	.03874	.01754	.00954
28	.11541	.06259	.03393	.01502	.00805
29	.11691	.06368	.03472	.01545	.00832
30	.11830	.06481	.03543	.01583	.00852

TABLE 3

$$\frac{2}{1+\beta} = \frac{3}{2}$$

n-1	α=.100	α=.050	α=.025	α=.010	α=.005
1	.09258	.04573	.02279	.00911	.00455
2	.09512	.04442	.02141	.00838	.00416
3	.10070	.04732	.02194	.00824	.00400
4	.10450	.05033	.02360	.00855	.00405
5	.10748	.05284	.02520	.00919	.00426
6	.09896	.04913	.02379	.00893	.00420
7	.10559	.05327	.02655	.01025	.00484
8	.11044	.05716	.02890	.01125	.00537
9	.10107	.05142	.02602	.01027	.00499
10	.10683	.05556	.02835	.01139	.00559
11	.11191	.05847	.03019	.01226	.00611
12	.10304	.05284	.02699	.01092	.00544
13	.10741	.05595	.02889	.01191	.00603
14	.11070	.05813	.03041	.01273	.00647
15	.11311	.06014	.03173	.01333	.00682
16	.11578	.06189	.03270	.01382	.00710
17	.11763	.06295	.03335	.01418	.00732
18	.10994	.05770	.03016	.01264	.00650
19	.11169	.05896	.03098	.01311	.00676
20	.11305	.05999	.03172	.01347	.00697
21	.11416	.06089	.03223	.01374	.00714
22	.11536	.06156	.03267	.01400	.00731
23	.11605	.06205	.03305	.01423	.00746
24	.11661	.06255	.03341	.01446	.00761
25	.11709	.06297	.03375	.01465	.00774
26	.11763	.06344	.03407	.01486	.00786
27	.11828	.06391	.03441	.01503	.00797
28	.11200	.05948	.03155	.01357	.00713
29	.11281	.06009	.03199	.01381	.00728
30	.11362	.06074	.03240	.01404	.00740

TABLE 4

$$\frac{2}{1+\beta} = \frac{7}{3}$$

n-1	α=.100	α=.050	α=.025	α=.010	α=.005
1	.10361	.05220	.02615	.01047	.00523
2	.10268	.05298	.02703	.01096	.00551
3	.10003	.05160	.02668	.01103	.00561
4	.09802	.05012	.02588	.01082	.00556
5	.09649	.04888	.02508	.01048	.00542
6	.09538	.04791	.02440	.01014	.00524
7	.09450	.04715	.02384	.00983	.00507
8	.09383	.04653	.02338	.00957	.00491
9	.09326	.04603	.02300	.00934	.00477
10	.09281	.04562	.02269	.00915	.00465
11	.09244	.04528	.02242	.00899	.00455
12	.09212	.04498	.02219	.00885	.00446
13	.09184	.04473	.02199	.00873	.00438
14	.09161	.04451	.02182	.00863	.00431
15	.09141	.04432	.02167	.00853	.00425
16	.09121	.04415	.02154	.00845	.00420
17	.09106	.04400	.02142	.00838	.00415
18	.09091	.04388	.02132	.00831	.00411
19	.09079	.04375	.02122	.00825	.00407
20	.09068	.04364	.02114	.00820	.00403
21	.09056	.04354	.02106	.00815	.00400
22	.09047	.04345	.02099	.00810	.00397
23	.09037	.04336	.02092	.00806	.00394
24	.09030	.04329	.02086	.00802	.00392
25	.09023	.04322	.02081	.00799	.00390
26	.09015	.04315	.02075	.00795	.00388
27	.09009	.04309	.02071	.00792	.00386
28	.09003	.04304	.02066	.00789	.00383
29	.08998	.04298	.02062	.00787	.00382
30	.08992	.04294	.02058	.00784	.00380

TABLE 5

$$\frac{2}{1+\beta} = \frac{5}{2}$$

n-1	α=.100	α=.050	α=.025	α=.010	α=.005
1	.10512	.05315	.02666	.01067	.00534
2	.10386	.05428	.02795	.01140	.00574
3	.10013	.05233	.02743	.01150	.00589
4	.09724	.05023	.02629	.01119	.00582
5	.09505	.04847	.02515	.01070	.00561
6	.09343	.04708	.02418	.01021	.00536
7	.09215	.04599	.02339	.00978	.00511
8	.09117	.04510	.02274	.00941	.00488
9	.09035	.04439	.02220	.00909	.00469
10	.08970	.04379	.02175	.00882	.00452
11	.08914	.04330	.02137	.00860	.00438
12	.08868	.04287	.02105	.00840	.00425
13	.08827	.04251	.02077	.00823	.00415
14	.08793	.04219	.02053	.00808	.00405
15	.08762	.04192	.02032	.00795	.00397
16	.08735	.04168	.02013	.00783	.00389
17	.08712	.04146	.01996	.00774	.00383
18	.08690	.04127	.01981	.00764	.00377
19	.08672	.04109	.01968	.00756	.00371
20	.08655	.04093	.01956	.00749	.00366
21	.08638	.04079	.01945	.00742	.00362
22	.08625	.04066	.01934	.00735	.00358
23	.08611	.04054	.01925	.00730	.00354
24	.08600	.04043	.01917	.00724	.00351
25	.08589	.04033	.01909	.00720	.00348
26	.08577	.04023	.01901	.00715	.00345
27	.08568	.04014	.01895	.00711	.00342
28	.08560	.04007	.01888	.00707	.00340
29	.08552	.03999	.01882	.00703	.00337
30	.08544	.03992	.01877	.00700	.00335

TABLE 6

$$\frac{2}{1+\beta} = 4$$

n-1	α=.100	α=.050	α=.025	α=.010	α=.005
1	.11349	.05898	.02983	.01197	.00599
2	.11074	.06206	.03382	.01447	.00743
3	.10170	.05718	.03222	.01472	.00794
4	.09384	.05170	.02912	.01365	.00760
5	.08763	.04701	.02612	.01225	.00691
6	.08275	.04321	.02356	.01093	.00617
7	.07888	.04015	.02144	.00977	.00549
8	.07575	.03765	.01970	.00880	.00489
9	.07318	.03560	.01827	.00799	.00439
10	.07104	.03390	.01707	.00731	.00397
11	.06925	.03246	.01606	.00674	.00361
12	.06771	.03124	.01521	.00625	.00331
13	.06638	.03018	.01447	.00584	.00305
14	.06523	.02927	.01384	.00548	.00283
15	.06421	.02847	.01328	.00518	.00264
16	.06331	.02776	.01280	.00491	.00247
17	.06252	.02713	.01237	.00467	.00233
18	.06180	.02657	.01198	.00446	.00220
19	.06116	.02607	.01164	.00427	.00208
20	.06058	.02561	.01133	.00411	.00198
21	.06005	.02520	.01105	.00396	.00189
22	.05957	.02482	.01080	.00382	.00181
23	.05912	.02448	.01057	.00370	.00174
24	.05872	.02416	.01036	.00358	.00167
25	.05834	.02387	.01016	.00348	.00161
26	.05799	.02360	.00998	.00339	.00156
27	.05767	.02335	.00982	.00330	.00150
28	.05737	.02312	.00966	.00322	.00146
29	.05709	.02290	.00952	.00315	.00142
30	.05682	.02270	.00939	.00308	.00138

TABLE 7

$$\frac{2}{1+\beta} = 16$$

n-1	α=.100	α=.050	α=.025	α=.010	α=.005
1	.12187	.06730	.03534	.01442	.00724
2	.11510	.06936	.04097	.01946	.01065
3	.10222	.06194	.03781	.01935	.01139
4	.09037	.05386	.03290	.01724	.01048
5	.08024	.04668	.02819	.01479	.00909
6	.07167	.04059	.02410	.01252	.00771
7	.06440	.03547	.02066	.01057	.00647
8	.05820	.03117	.01778	.00893	.00542
9	.05287	.02754	.01538	.00757	.00455
10	.04827	.02446	.01338	.00645	.00383
11	.04428	.02184	.01169	.00552	.00324
12	.04079	.01960	.01027	.00474	.00275
13	.03771	.01766	.00907	.00410	.00234
14	.03501	.01599	.00804	.00355	.00201
15	.03260	.01453	.00716	.00310	.00172
16	.03045	.01325	.00640	.00271	.00149
17	.02853	.01212	.00575	.00238	.00129
18	.02681	.01113	.00518	.00210	.00112
19	.02525	.01025	.00468	.00186	.00098
20	.02384	.00947	.00424	.00165	.00086
21	.02256	.00876	.00386	.00147	.00076
22	.02139	.00814	.00352	.00132	.00067
23	.02032	.00757	.00322	.00118	.00059
24	.01935	.00706	.00295	.00106	.00053
25	.01845	.00660	.00271	.00096	.00047
26	.01762	.00618	.00249	.00086	.00042
27	.01685	.00580	.00230	.00078	.00037
28	.01615	.00545	.00213	.00071	.00034
29	.01549	.00513	.00197	.00065	.00030
30	.01488	.00484	.00183	.00059	.00027

CHAPTER VII

CONCLUSIONS

As was previously mentioned, many authors have concluded that kurtosis is of primary importance in the robustness of the Student t-test. With our approximation, we might extend this to say that for parent populations with kurtosis > 3, which is the normal density value, the tests are optimistic in the type I error probabilities; that is, the real α value is larger than the advertised value. For parent populations with kurtosis < 3, the tests are conservative; that is, the real α value is smaller than the advertised value.

Box and Tiao [3] concluded that the Student t-test is a good approximation for all members of the generalized normal family. They used Bayesian techniques with specific families of prior densities for μ and β in their work. The present work seems to reinforce their conclusions since the same conclusions follow from very different approaches.

It should be observed that the conclusions stated here are based on an approximation of the t-density for a particular family of parent densities. Further, the approximation was thoroughly investigated for only a few members of this family and no estimate is given of the precision of the approximation.

APPENDIX A

For the sake of completeness, the mean-value theorem in its standard form is stated. The proof is omitted since it can be found in most elementary calculus textbooks.

Theorem: If g is a continuous function on the interval $[a, b]$, then there exists $\xi \varepsilon [a, b]$ such that $\int_a^b g(x)\,dx = (b-a)\,g(\xi)$.

The slightly modified version of this theorem that is applied for equation (3.5) is as follows.

Theorem: If g is a continuous function on the interval (a, b) , $\lim_{x \to a} g(x) = \lim_{x \to b} g(x) = \infty$ and $\int_a^b g(x)\,dx < \infty$, then there exists $\xi_1, \xi_2 \varepsilon (a, b)$ such that $\int_a^b g(x)\,dx = (b-a)\,g(\xi_1) = (b-a)\,g(\xi_2)$ where $\xi_1 \neq \xi_2$.

Proof: Since g is continuous on (a, b) and positively unbounded at the endpoints, g is bounded below on (a, b). Hence, g has an absolute minimum on (a, b), that is, there exists $\eta \varepsilon (a, b)$ such that $g(x) \geq g(\eta)$ for all $x \varepsilon (a, b)$. Then

$$g(\eta) \leq \frac{1}{b-a} \int_a^b g(x)\,dx \quad . \tag{A.1}$$

Suppose equality holds in (A.1) and consider the function $h(x) = g(x) - g(\eta)$, which is bounded below by zero and continuous on (a, b) and positively unbounded at the end points. We can easily verify that

$$\int_a^b h(x)\,dx = 0 \quad ,$$

which is clearly impossible. Hence,

$$g(\eta) < \frac{1}{b-a} \int_a^b g(x)\,dx \quad . \tag{A.2}$$

Since g is continuous, g takes on all values $[g(\eta), \infty)$ and there exists $\xi_1 \varepsilon (\eta, b)$ and $\xi_2 \varepsilon (a, \eta)$, such that the conditions of the theorem are satisfied.

The application of these two theorems is actually a direct corollary and several questions are introduced in these corollaries. The exact situation and some of the questions, which are basically mathematical in nature, need to be specified. Let $\underline{t}$ be a p-dimensional vector and $g\colon R_{p+1} \to R$ be suitably continuous. The mean-value theorem guarantees the existence of $\xi\colon R_p \to R$ where $\int_a^b g(\underline{t}, u)\,du = g[\underline{t}, \xi(\underline{t})]$ and $a \leq \xi(\underline{t}) \leq b$. Obviously, the optimal situation is to be able to exhibit a ξ for the particular g. When this is not possible facts such as the variation of ξ, the differentiability of ξ or even the continuity of ξ, given sets of conditions on g, would be useful.

APPENDIX B

The computer program used to calculate the probabilities given in the tables was run on a UNIVAC 1108. For each β considered, the running time was slightly under 1 minute. The program given here is not optimal for the machine, but it is in a form general enough to accommodate most considerations of the generalized normal distribution with $\mu = 0$.

The transformation $x = u/(1+u)$ was made to make the range of integration finite where u is the original variable of integration. The function given in (6.1) was factored slightly and the coefficients $b_{in} = a_{in}\sqrt{n(n-1)}$ were used.

```
      DOUBLE PRECISION T, BETA, BETAS, PROB, GAMALN, B, E, XIP, DN,
     1GLN1, GLN2, C, FCT, ANS, UL
      DIMENSION T(150) BETA(  ) PROB(  ) E(31)
      COMMON BETA, N, B(31)
      EXTERNAL FCT
      DATA T/(tα(n-1) values)
      DATA BETAS/(β values to be considered)
      DATA E/(ξn values, n=2,···,31 with ξ2=0)
      DO 5M=1, (no. of β values)
      XIP=1.0D0
      K=0
      BETA=BEATAS(M)
      B(1)=1.0D0
      B(2)=-B(1)
      DO 4N=2, 31
      DN=DBLE(N)
      IF(N.LE.2) GO TO 2
      J=N-1
      DO 1 I=1,J
      B(I)=E(N)+B(I)*DSQRT((DN*(1-E(N)*E(N)))/(DN-2))
    1 CONTINUE
      B(N)=-(DN-1)*(E(N)
```

```
    2 GLN1=GAMALN(DN*(1+BETA)/2)
      GLN2=-DN*GAMALN((3+BETA)/2)
      XIP=XIP*DEXP(((DN-2)/4)*DLOG(1-E(N)*E(N)))
      C=XIP*((1+BETA)/4)*DEXP((DN/2)*DLOG(N)+((DN-1)/2)
     1*DLOG(DN-1)+GLN1+GLN2
      CALL QUAD 48(FCT, 1.0D0, 0.0D0, ANS)
      PROB(1)=2*C*ANS
      DO 3 J=2, (no. of α's being considered)
      UL=T(K+J-1)/(T(K+J-1)+1)
      CALL QUAD 48(FCT, 1.0D0, UL, ANS)
      PROB(J)=C*ANS/PROB(1)
    3 CONTINUE
      WRITE(6, 1000)N, (PROB(J), J=1, (no. of α's +1))
 1000 FORMAT(3x, I4, 10x, (no. of α's +1)F10.5,/)
      K=K+(no. of α's)
    4 CONTINUE
    5 CONTINUE
      STOP
      END


      FUNCTION FCT(U)
      DOUBLE PRECISION U, X, Y, FCT, BETA, B
      COMMON BETA, N, B(31)
      FCT=0.0D0
      X=U/(1-U)
      DO 1 I=1,N
      Y=DABS(X+B(I))
      FCT=DEXP((2/(1+BETA))*DLOG(Y))+FCT
    1 CONTINUE
      FCT=DEXP(-(N*(1+BETA)/2)*DLOG(FCT))/((1-U)*(1-U))
      RETURN
      END
```

QUAD 48(FCT, UL, LL, ANS) is a 48 point Gauss-Legendre integration subroutine where FCT carries the function values, UL is the upper limit of integration, LL is the lower limit of integration and ANS returns the value of the integral. GAMALN(X) is a function that calculates the natural logarithm of the gamma function of x, using Bernoulli numbers.

LIST OF REFERENCES

[1] Baker, G. A. "Distribution of the means divided by the standard deviations of samples from non-homogeneous populations," Annals of Mathematical Statistics, 3, 1-9, 1932.

[2] Bartlett, M. S. "The effect of non-normality on the t-distribution," Proceedings of the Cambridge Philosophical Society, 31, 223-31, 1935.

[3] Box, G. E. P. and Tiao, G. C. "A further look at robustness via Bayes' theorem," Biometrika, 49, 419-32, 1962.

[4] Bradley, R. A. "The distribution of the t and F statistics for a class of non-normal populations," Virginia Journal of Science, 3, 1-32, 1952.

[5] Craig, Allen T. "The simultaneous distribution of mean and standard deviation in small samples," Annals of Mathematical Statistics, 3, 126-40, 1932.

[6] Gayen, A. K. "The distribution of 'Student's' t in random samples of any size drawn from non-normal universes," Biometrika, 36, 353-69, 1949.

[7] Geary, R. C. "On the distribution of 'Student's' ratio for samples of three drawn from a rectangular distribution," Biometrika, 25, 203-04, 1933.

[8] Govindarajulu, Z. and Leslie, R. T. Annotated Bibliography on Robustness Studies. The University of Kentucky, Department of Statistics, Technical Report #7, Sept. 1970.

[9] Hatch, L. O. and Posten, H. O. Robustness of the Student-Procedure: A Survey. The University of Connecticut, Department of Statistics, Research Report #24, August 1966.

[10] Hotelling, H. "The behavior of some standard statistical tests under nonstandard conditions," Proceedings of the Fourth Berkeley Symposium, I, 319-59, 1961.

[11] Laderman, J. "The distribution of 'Student's' ratio for samples of two items drawn from non-normal universes," Annals of Mathematical Statistics, 10, 376-79, 1939.

[12] Owen, D. B. Handbook of Statistical Tables, Addison-Wesley, 1962.

[13] Owen, D. B. "The power of Student's t-test," JASA, 60, 320-33, 1965.

[14] Perlo, V. "On the distribution of 'Student's' ratio for samples of three drawn from a rectangular distribution," Biometrika, 25, 203-04, 1933.

[15] Rider, P. R. "On small samples from certain non-normal universes," Annals of Mathematical Statistics, 2, 48-65, 1931.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
SOUTHERN METHODIST UNIVERSITY

2a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

2b. GROUP
UNCLASSIFIED

3. REPORT TITLE
The Density of the t-Statistic for Non-normal Distributions

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Technical Report

5. AUTHOR(S) (First name, middle initial, last name)
Raymond C. Sansing

6. REPORT DATE
April 6, 1971

7a. TOTAL NO. OF PAGES
36

7b. NO. OF REFS
15

8a. CONTRACT OR GRANT NO.
N00014-68-A-0515

b. PROJECT NO.
NR 042-260

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)
99

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT
This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY
Office of Naval Research

13. ABSTRACT
The joint density function of the sample mean and sample variance is recursively derived for samples from a population with density function f where $f(x) > 0$ almost everywhere, everywhere continuous and has certain integral properties. For populations where f does not have these integral properties, this joint density is an approximation. This joint density function is used to derive the density function of the t-statistic for samples from f. The family of generalized normal density functions is used for an example. The approximation for the t-density is given for that family. For some specific members of the family, the true α probabilities for the approximations are tabled.

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6801

UNCLASSIFIED
Security Classification