AD 670743

DEPARTMENT OF STATISTICS
HARVARD UNIVERSITY

STUDENT'S T-TEST UNDER NON-NORMAL CONDITIONS

BRADLEY EFRON

TECHNICAL REPORT NO. 21
MAY 22, 1968
CONTRACT NONR 1866(37), NR-042-097

THIS PAPER WAS SUPPORTED BY THE OFFICE OF NAVAL RESEARCH.

u0

0. Abstract.

The size and power of Student's t-test are discussed under weaker than normal conditions. It is shown that assuming only a symmetry condition for the null hypothesis leads to effective bounds on the dispersion of the t-statistic. (The symmetry condition is weak enough to include all cases of independent but not necessarily identically distributed observations, each symmetric about the origin.) The connection between Student's test and the usual non-parametric testsis examined, as well as power considerations involving Winsorization and permutation tests. Simultaneous use of different one-sample tests is also discussed.

1. The Geometry of Student's One-sample t-statistic.

The distribution of Student's one-sample t-statistic

$$T_n = \frac{\sum_{i=1}^{n} X_i}{\sqrt{n}} \Bigg/ \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}}$$

is usually derived under normal sampling theory, with the X_i assumed to be independent, identically distributed normal random variables,

$$X_i \overset{ind}{\sim} N(\mu,\sigma^2), \qquad i = 1,2,\ldots,n.$$

In his essentially geometrical derivation of Student's distribution, Fisher [6] showed that the rotational symmetry of the random vector

$$X = (X_1, X_2, X_3, \ldots, X_n)$$

under the null hypothesis $\mu = 0$ is sufficient to yield the standard null distribution for T_n.

To be more precise, let U be the unit vector $U = X/\|X\|$, so that

$$U_i = X_i \Bigg/ \sqrt{\sum_{1}^{n} X_i^2} \qquad i = 1,2,\ldots,n.$$

Under the null hypothesis $X_i \overset{ind}{\sim} N(0,\sigma^2)$, U will be uniformly distributed on the surface of S_n, the unit sphere in Euclidean n-space E^n,

$$S_n = \{u: \sum_{i=1}^{n} u_i^2 = 1\}.$$

For any set A on S_n, $P(U \epsilon A) = \lambda\{A\}$, where λ is the usual measure of n-1 dimensional "area" on S_n, normalized so that $\lambda\{S_n\} = 1$.

[This follows from the fact that the density of X in E^n depends only on $\|X\|$.]

Student's statistic is a monotonic function of

$$S_n = \sum_{i=1}^{n} X_n \Big/ \sqrt{\sum_{i=1}^{n} X_i^2} = \sum_{i=1}^{n} U_i,$$

$(T_n = S_n \Big/ \sqrt{\frac{n-1}{n-S_n^2}})$. If we let

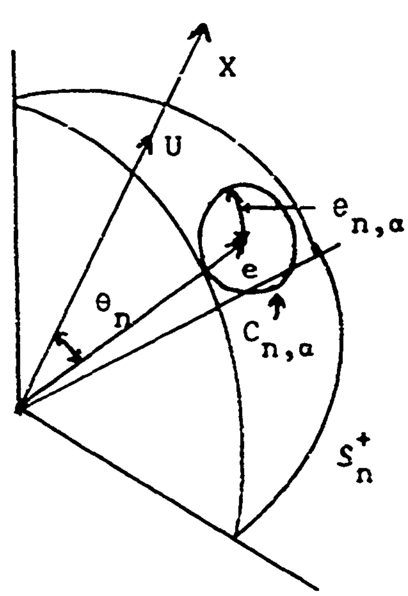
$$e = (\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}})$$

represent the unit main diagonal, then

$$S_n = \sqrt{n} \text{ Cos } \theta_n,$$

where θ_n is the angle between U and e, (see Figure 1), so that S_n is a decreasing function of θ_n.

Figure 1



The distribution for S_n now follows from the known formula for the area of a spherical cap on S_n. In particular, if we wish to choose $s_{n,\alpha}$ such that

$$P(S_n \geq s_{n,\alpha}) = \alpha,$$

it is equivalent to find the angular radius $\theta_{n,\alpha}$ of a spherical cap $C_{n,\alpha}$ on S_n having

$$\lambda\{C_{n,\alpha}\} = \alpha.$$

The rejection set for student's one-sided t-test is then $U \varepsilon C_{n,\alpha}$, where $C_{n,\alpha}$ has radius $\theta_{n,\alpha}$ and center e. The value $s_{n,\alpha}$ is given by $s_{n,\alpha} = \sqrt{n}\,\mathrm{Cos}\,\theta_{n,\alpha}$.

For reasonable values of n and α, the critical angle $\theta_{n,\alpha}$ tends to be quite large. If $\alpha = .025$ for instance, we have the following table of values:

$n =$	6	11	26	51	∞
$\theta_{n,\alpha} =$	41°	55°	68°	74°	90°

Figure 1 is misleading since it shows $C_{n,\alpha}$ entirely contained in the positive orthant of S_n. As a crucial part of our discussion we will see that ordinarily $C_{n,\alpha}$ will extend far outside the positive orthant. For example, when $n = 20$ and $\alpha = .05$, $C_{n,\alpha}$ contains the center points of 60,460 of the 2^{20} orthants [c.f. Section 3].

2. A Summary of This Paper.

The geometry of Section 1 shows that the normal theory for the null hypothesis distribution of Student's t-statistic remains valid under the weaker assumption of rotational symmetry, i.e. that $U = X/\|X\|$ is uniformly distributed over S_n. Thus, for example, the null hypothesis that X is uniformly distributed within some sphere centered at the origin can be tested at level α by rejecting for values of T_n greater than the tabled upper α point of Student's distribution (with n-1 degrees of freedom, in the standard terminology).

Unfortunately the usual sampling procedures almost never* yield rotational symmetry for the normalized vector U except in the case $X_i \overset{ind}{\sim} N(0,\sigma^2)$. If, for instance, the X_i are independently +1 or -1 with probabilities $\frac{1}{2}$, then U is always the center point of one of the orthants of S_n.

The central purpose of this paper is to discuss Student's t-statistic under a much weaker symmetry condition, which is satisfied under the null hypotheses of many standard sampling situations:

Definition: The random vector $U = (U_1, U_2, \ldots, U_n)$ is said to have ORTHANT SYMMETRY if it has the same distribution as $U_\delta = (\delta_1 U_1, \delta_2 U_2, \ldots, \delta_n U_n)$ for every choice of $\delta_i = \pm 1$, $i = 1,2,\ldots,n$.

*A very special "lucky" case is given in Section 5. It is possible to construct examples where T_n has the t distribution without U having rotational symmetry.

In particular, orthant symmetry obtains for $U = X/\|X\|$ whenever the components X_i of X are independent and each has a symmetric distribution about the origin. It is not necessary that the components have identical distributions.

Our main results are presented in Section 3, and can be roughly paraphrased as follows: orthant symmetry guarantees that Student's statistic, in the form $S_n = \sum_{i=1}^{n} X_i \Big/ \sqrt{\sum_{i=1}^{n} X_i^2}$, is less dispersed about the origin than the random variable $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Delta_i$, where the Δ_i are independent and equal +1 or -1 with probabilities $\frac{1}{2}$. That is, among all cases of orthant symmetry, the centered binomial case is the worst, in a sense to be described. We suggest that the size of Student's one-sample t-test is robust under the null hypothesis of orthant symmetry, and as a matter of fact, the type I error tends to decrease from the nominal α level under such "bad" conditions as the X_i having Cauchy distributions.

Sections 4 and 5 contain heuristic discussions of this point, as well as an Edgeworth-type expansion to help assess the magnitude of the decrease. Section 5 is particularly concerned with the effects of long-tailed error laws, such as the Cauchy, on the Student's ratio. As an aid to intuition, a particularly tractable long-tailed error law is introduced and examined in detail.

Orthant symmetry is preserved under many familiar statistical operations: taking signs, ranks, censoring, Winsorizing, etc. In Section 6 we use this fact to discuss the sign test and Wilcoxon's signed rank test as "generalized Student's tests".*

*Our definition of this term is not that of Hajek [8].

Section 7 extends this concept to Winsorization and permutation tests, and shows how orthant symmetry allows some "cheating" for increased power, that is looking at the data before choosing the test statistic, without compromising the α level.

The use of more than one test on the same data, for instance Student's test and the sign test, is discussed in Section 8, and a method of evaluating the α level of the simultaneous testing procedure is suggested. The question of conditional versus unconditional tests, which is largely ignored in most of the paper, is discussed in Section 9 in relation to another geometrical distribution, that of the angle between X and a vector other than e. We conclude with a discussion of references. Mathematical details are collected in an appendix to the paper.

3. The Main Results.

We will work with Student's statistic in the form*
$S_n = \sum_{i=1}^{n} X_i \Big/ \sqrt{\sum_1^n X_i^2}$. Our main assumption will be that $X = (X_1, X_2, \ldots, X_n)$ has orthant symmetry as defined in Section 2, and to avoid trivialities we also assume that $P(X_i = 0) = 0$ for $i = 1,2,\ldots,n$.

Let λ_n be the probability distribution of $U = X/\|X\|$ on the unit sphere S_n (so that $\lambda_n = \lambda$, the uniform distribution, if $X_i \overset{ind}{\sim} N(0,\sigma^2)$). Orthant symmetry is equivalent to the statement that λ_n is identical over each of the 2^n orthants of S_n. In

*Use of S_n rather than the traditional T_n almost obviates the need for special tables in the standard case $X_i \overset{ind}{\sim} N(0,\sigma^2)$. The upper 5% point of S_6 for instance is 1.640, as compared to 1.645 for a N(0,1) variable. (c.f. Section 4).

particular we can consider only the positive orthant S_n^+,

$$S_n^+ = \{\xi = (\xi_1, \xi_2, \ldots, \xi_n) : \xi_i > 0, \sum_{i=1}^{n} \xi_i^2 = 1\},$$

and define the probability measure λ_n^+ on S_n^+ by

$$\lambda_n^+\{A\} = 2^n \lambda_n\{A\}$$

for $A \subset S_n^+$. We see that λ_n^+ determines λ on all of S_n via the orthant symmetry. In the case $\lambda_n = \lambda$ we have $\lambda_n^+ = \lambda^+$, the uniform distribution on S_n^+.

Definition: Let $\xi \in S_n^+$. Then $S_\xi = \sum_{i=1}^{n} \xi_i \Delta_i$, where the Δ_i are independent random variables taking values +1 or -1 with probabilities $\frac{1}{2}$, will be called a generalized binomial random variable. The case $S_e = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Delta_i$ will be called the centered binomial random variable.

The following simple lemma is basic to our results.

Lemma: Under orthant symmetry S_n is a mixture of generalized binomial random variables with λ_n^+ as the mixing distribution. That is, $P(S_n < s) = \int_{S_n^+} P(S_\xi < s) d\lambda_n^+\{\xi\}$ for every value of s.

The lemma is only a statement of the fact that if we condition on the vector of normalized absolute values $(\xi_1, \xi_2, \ldots, \xi_n)$ defined by

$$\xi = (|U_1|, |U_2|, \ldots, |U_n|) ,$$

then orthant symmetry guarantees that each of the 2^n possible unit vectors

$$U = (\delta_1 \xi_1, \delta_2 \xi_2, \ldots, \delta_n \xi_n)$$

where $\qquad \delta_i = \pm 1 \qquad i = 1, 2, \ldots, n,$

is equally likely.

Notice that the lemma expresses the distribution of S_n in terms of an integrand $P(S_\xi < s)$ that does not depend on the distribution of the observations X. (This distribution enters only through the induced measure λ_n^+.) Therefore, anything we can prove about the class of generalized binomial random variables yields a general theorem about Student's statistic under orthant symmetry. A simple but not very useful example is Tchbycheff's inequality: $P(|S_\xi| > c) < \frac{1}{c^2}$ for every ξ, since S_ξ has mean 0 and variance 1. Therefore $P(|S_n| > c) < \frac{1}{c^2}$ under orthant symmetry.

Moments are convenient to work with here, since they pass easily through the mixture process. For every value of ξ, $ES_\xi^2 = 1$, $ES_\xi^\nu = 0$ for ν odd, so that the same statements hold for S_n. (In particular, S_n has mean 0 and variance 1.) Our main result is a bound on the higher even moments of S_n.

<u>Theorem</u>: Under orthant symmetry, $ES_n^\nu \leq ES_e^\nu$ for $\nu = 4,6,8,\ldots$, with equality if and only if the X_i are identical independent centered Bernoulli trials, $X_i = \pm c$ with probabilities $\frac{1}{2}$ for $i = 1,2,\ldots,n$.

Recall that $S_e = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Delta_i$ where the Δ_i independently equal +1 or -1 with probabilities $\frac{1}{2}$. The theorem says that the ν^{th} central moment of S_n is bounded by the corresponding moment of the centered binomial random variable, which equals 2^ν times the moment of a standard binomial random variable with n trials and $p = \frac{1}{2}$.

The theorem follows from the lemma by showing that $ES_\xi^\nu < ES_e^\nu$ for $\xi \neq e$. A proof of this statement is given in the appendix to this paper.

<u>Corollary</u>: ES_n^{ν} is less than the corresponding moment of a N(0,1) random variable.

By the theorem, it is necessary and sufficient to prove this for ES_e^{ν}, which is done in the appendix.

Our theorem bounds the moments of S_n rather than the type I error probabilities $P(S_n > s_{n,\alpha})$. In the next section we will use the mixture lemma to develop an Edgeworth expansion for the distribution of S_n. The random variable S_n has mean 0 and variance 1, and differs from a N(0,1) distribution by an Edgeworth sum whose leading term depends on the kurtosis* of S_n,

$$\text{kurt}(S_n) = ES_n^4 - 3.$$

The next corollary provides some justification for the statement in the summary that Student's test tends to behave conservatively (smaller than nominal α level) under orthant symmetry.

<u>Corollary</u>: S_n has negative kurtosis under orthant symmetry. Under the additional assumption that the variables $\underset{\sim}{U}_i$ are exchangable (index symmetry),

$$\text{kurt}(S_n) = -2nEU_1^4.$$

We calculate directly

$$ES_\xi^4 = 3 - 2\sum_i^n \xi_i^4,$$

and therefore $ES_n^4 = 3 - 2E\sum_{i=1}^n \xi_i^4 = 3 - 2nEU_1^4$ under exchangeability.

---

*Many writers call this the "coefficient of excess" rather than the kurtosis.

Note that U will have both symmetries if the random variables X_i are independent and identically distributed symmetrically about the origin.

Negative kurtosis <u>tends</u> to give a distribution that has smaller than the N(0,1) probability of exceeding any constant s greater than $\sqrt{3}$. The tabled values of $s_{n,\alpha}$, derived from normal theory, are close to the α points for a N(0,1) distribution. These two facts together would indicate that $P(S_n > s_{n,\alpha}) \leq \alpha$ under orthant symmetry for the usual values of α. This statement is not actually true in general, but the violations of the α-level seem to be slight, particularly in the case of i.i.d. observations. We will examine this phenomena more closely the next two sections.

R. R. Bahadur and J. Eaton have communicated the following interesting bound on $P(S_n > s)$, and have been kind enough to allow me to include it in this paper:

<u>Theorem (Bahadur and Eaton)</u>: Under orthant symmetry, $P(S_n \geq s) \leq e^{-\frac{1}{2}s^2}$.

The proof follows from the mixture lemma and the fact that $P(S_\xi \geq s) \leq Ee^{s(S_\xi - s)} \leq e^{-1/2\ s^2}$.

4. <u>Edgeworth Expansion For Student's Statistic.</u>

We can obtain an expansion for the c.d.f. of S_n from the mixture lemma in the following way: we expand the generalized binomial c.d.f. in a standard Edgeworth series ([2], Chapter 17),

$$P(S_\xi < s) = \Phi(s) + k_4(\xi)\Phi^{(4)}(s) + k_6(\xi)\Phi^{(6)}(s) + \ldots,$$

where Φ is the standard N(0,1) c.d.f., $\Phi^{(4)}(s)$ its fourth derivative, etc. The constants k_j depend on ξ, and hence on n, and vanish for odd values of j by symmetry. The second corollary of Section 3 guarantees that $k_4(\xi) = \mathrm{kurt}(S_\xi)$ is negative.

The mixture lemma now yields

$$P(S_n < s) = \Phi(s) + Ek_4(\xi)\Phi^{(4)}(s) + Ek_6(\xi)\Phi^{(6)}(s) + \ldots ,$$

where the expectations are with respect to λ_n^+,

$$Ek_j(\xi) = \int_{S_n^+} k_j(\xi)d\lambda_n^+(\xi)$$

(recall that $\xi_i = |U_i| = |X_i|/\|X\|$ for $i = 1,2,\ldots,n$). These expectations become particularly simple when U has index symmetry (exchangeable coordinates) as well as orthant symmetry.

<u>Edgeworth Expansion</u>: assuming that $U = X/\|X\|$ has orthant and index symmetry,

$$\begin{aligned} P(S_n < s) = \Phi(s) &- \{\tfrac{1}{12}(EnU_1^4)\Phi^{(4)}(s)\} \\ &+ \{\tfrac{1}{45}(EnU_1^6)\Phi^{(6)}(s) + \tfrac{1}{288}(EnU_1^8 + En(n-1)U_1^4U_2^4)\Phi^{(8)}(s)\} \\ &+ \ldots \end{aligned}$$

(See the appendix for the derivation of this formula.) Here the first bracketed term comes from the "$\frac{1}{n}$" term in the Edgeworth series for S_ξ, while the second bracketed term is "$\frac{1}{n^2}$". The quotation marks are necessary since, as we shall see in Section 5, these terms will approach non-zero limits if the X_i have long tails.

The usefulness of an asymptotic (non-convergent) series, such as the expansion given above, can usually be determined only by experience. Cramer suggests in [2] that the Edgeworth expansion not be used beyond the first correction term. Since in our case the term is either negligible or positive for values of s in the usual testing range, say $1.5 \leq s \leq 2.5$, the simple approximation of S_n's c.d.f. by $\Phi(s)$ would seem to err usually in the conservative direction.

A simple test case for the expansion formula is that where the X_i are independent $N(0,1)$, i.e. in the case of the genuine Student's distribution.

$s \rightarrow$	.50	.75	1.00	1.50	2.00	.50	1.00	2.00	2.50
Actual Value $P(S_n<s)$	.665	.743	.813	.928	.992	.675	.824	.984	.999
$\Phi(s)$	.692	.773	.841	.933	.977	.692	.841	.977	.994
$\Phi(s)-\frac{1}{12}(EnU_1^4)\Phi^{(4)}(s)$	.674	.754	.824	.928	.981	.679	.829	.980	.997
	n=5 (4 degrees of freedom)					n=8 (7 degrees of freedom)			

Chung [1] has given an expansion for S_n directly from the moments of the X_i, rather than through the normalized vector U. His expansion does not require orthant symmetry. On the other hand, it does require the existence of higher order moments of the X_i. while the formula given here does not. We can therefore apply our expansion to such interesting cases as the $X_i \overset{ind}{\sim}$ Cauchy. We discuss long-tailed error laws in some detail in the next section.

5. Long-tailed Error Laws.

Looking again at Figure 1, let us imagine computing the c.d.f.

$$F_\xi(s) = P(S_\xi < s)$$

of the generalized binomial random variable S_ξ (defined in Section 3) for each value of ξ in S_n^+. If n is even moderately large, say $n \geq 10$, F_ξ will be reasonably well approximated by the N(0,1) c.d.f. Φ or values of s in the usual testing range, as long as ξ is near the main diagonal point $e = (\frac{1}{\sqrt{n}},\ldots,\frac{1}{\sqrt{n}})$. The central limit theorem will fail more and more drastically as ξ approaches the corners* of S_n^+, which is to say as the components of ξ become more unequal in magnitude. The extreme case is $\xi = (1,0,0,\ldots,0)$, which yields $S_\xi = \pm 1$ with probabilities $\frac{1}{2}$.

As we have discussed, the deviations of F_ξ from Φ will always be in the platykurtic direction, kurtosis $(S_\xi) = -2 \sum_{i=1}^{n} \xi_i^4$, with a general tendency for $F_\xi(z_\alpha)$ to exceed the nominal value $\Phi(z_\alpha) = 1-\alpha$ for the customary α values. (Computer experimentation has shown that for small values of n this tendency to err in the conservative direction is more drastic than indicated by the Edgeworth correction.)

Now let us consider the case where the X_i are independent and identically distributed random variables, symmetric about 0. If the X_i have finite variance, then writing S_n as

$$S_n = \frac{\sum_i^n X_i}{\sqrt{n}} \Bigg/ \sqrt{\frac{\sum_i^n X_i^2}{n}}$$

shows that S_n is asymptotically N(0,1), since the numerator approaches $N(0,\sigma^2)$ by the central limit theorem while the

---

*It should be remembered that in higher-dimensional space there are "zero-dimensional corners", "one-dimensional corners", "two-dimensional corners", etc. We use "corner" here for a low dimensional boundary of S_n^+, in a sense which will be made explicit.

denominator approaches σ by the law of large numbers. In terms of our picture, this means that the mixing distribution λ_n^+ must, for large n, put nearly all of its mass in that portion of S_n^+ sufficiently near e for the central limit theorem to yield good approximations.*

If σ^2 is infinite our derivation of limiting normality for S_n fails in both the numerator and denominator. Intuitively, we expect that if the X_i have a long-tailed error law, $\sigma^2 = \infty$, then λ_n^+ will put much more of its mass near the corners of S_n^+. The very term "long-tailed" implies occasional freakishly large values for the X_i, which result in ξ vectors near these corners. From the mixture lemma we then expect S_n to have a much more platykurtic distribution, the most extreme case being kurt$(S_n) = -2$ if ξ always has only one non-zero component.

Two asymptotic results supporting these intuitive arguments can be reported for the case where the X_i have a stable distribution law of order α, $0 < \alpha < 2$. (We are still assuming that the X_i are i.i.d. and symmetric about 0.) Darling [3] shows that, in our notation, $E\left(\frac{1}{\max\limits_{1\le i\le n} \xi_i^2}\right) \to 1/(1-\frac{\alpha}{2})$ as $n \to \infty$, which implies

$\underline{\lim}\, E\left(\max\limits_{1\le i\le n} \xi_i^2\right) \ge 1 - \frac{\alpha}{2}$. We cannot expect limiting normality, therefore, since λ_n^+ must give high probability to ξ vectors whose

*In the case of the uniform distribution $\lambda_n^+ = \lambda^+$, for example, it is easy to prove that for any $\varepsilon, \delta > 0$ we have $\lambda^+\{\xi : \sup\limits_{-\infty<s<\infty} |F_\xi(s) - \Phi(s)| > \varepsilon\}$ $< \delta$ for all n sufficiently large.

components are not uniformly small (UAN). In an unpublished paper [14], Logan, Mallows, Rice, and Shepp* show that S_n must approach a limiting law when the X_i are i.i.d. stable variates, and give an integral expression for the characteristic function of the limit. This law has kurtosis = $-2 + \alpha$, so as α goes to 0 we approach the degenerate case.

We conclude this section with an example of a long-tailed error law very closely related to the normal law. This example has the advantage of easy calculation of λ_n^+ for any value of n, and is helpful in picturing many of the bizarre sampling effects of long-tailed distributions, such as Darling's result above.

We let

$$X_i = \frac{1}{\tilde{X}_i} \qquad \text{for } i = 1,2,\ldots,n,$$

where the $\tilde{X}_i$ are independent N(0,1) random variables. Then X_i has the density

$$f(X_i) = \frac{1}{\sqrt{2\pi}} \frac{1}{X_i^2} e^{-\frac{1}{2X_i^2}} ,$$

and has Cauchy-like tail behavior, being attracted to a stable law of order $\alpha = 1$. (The square $Z_i = X_i^2$ has density

$$f(Z_i) = \frac{1}{\sqrt{2\pi}} Z_i^{-\frac{3}{2}} e^{-\frac{1}{2Z_i}} ,$$

which is exactly the positive stable law of order $\frac{1}{2}$, a fact we use below. See [5], page 170.)

*I am grateful to the authors for allowing me to report these results.

Consider the mapping $X = g(\tilde{X})$ which takes vectors into vectors by inverting each component, $X_i = \frac{1}{\tilde{X}_i}$, $i = 1,2,\ldots,n$. Since $g(c\tilde{X}) = \frac{1}{c}\, g(\tilde{X})$ for any $c > 0$, this mapping induces a mapping on rays in E^n, and in particular induces a mapping of S_n^+ onto itself, say $\xi = g^+(\tilde{\xi})$. (For example, if $n = 3$ and $\tilde{X} = (2,1,2)$, so $\tilde{\xi} = (\frac{2}{3},\frac{1}{3},\frac{2}{3},)$, then $g(\tilde{X}) = (\frac{1}{2},1,\frac{1}{2})$ and $g^+(\tilde{\xi}) = (\frac{1}{\sqrt{6}},\frac{2}{\sqrt{6}},\frac{1}{\sqrt{6}}).)$

We see that the distribution λ_n^+ induced on S_n^+ by taking the X_i to be inverted normals, as defined above, is obtained by "inverting" the uniform distribution λ^+ via g^+,

$$\lambda_n^+\{g^+(A)\} = \lambda\{A\} \qquad \text{for } A \subset S_n^+,$$

(since $\tilde{X}$ itself induces λ^+). By putting coordinates on S_n^+ we can easily calculate $\frac{d\lambda_n^+\{\xi\}}{d\lambda^+\{\xi\}}$ for any value of ξ, from the properties of g^+. (This is done in the appendix.) Let us just note here that $\frac{d\lambda_n^+\{e\}}{d\lambda^+\{e\}} = 1$, which is not surprising since e is the fixed point of g^+, and more importantly, $\frac{d\lambda_n^+\{\xi\}}{d\lambda^+\{\xi\}} \to \infty$ as ξ approaches any corner of S_n^+ of dimension* $d < \frac{n}{2} - 1$. This is clear since g^+ maps any boundary line of S_n^+ into the opposite corner. Roughly speaking, g^+ maps points ξ near high dimensional boundaries of S_n^+, where S_ξ is nearly normal, into points ξ near low dimensional corners of S_n^+, where S_ξ tends to be non-normal in the platykurtic sense.

The formula for the kurtosis of S_n given in Section 3 can be evaluated explicitly here by making use of the fact that X_i^2 is stable of order $\frac{1}{2}$:

*In the case $n = 2$, $\lambda_n^+ = \lambda^+$ and there are no poles.

$$\mathrm{kurt}(S_n) = -\frac{4n}{\pi}\int_0^{\frac{\pi}{2}} [\frac{1}{1+(n-1)^2 \tan^2\theta}]^2 \, d\theta$$

$$= -1 - \frac{1}{n-1} + O(\frac{1}{n^2}).$$

The approach to the limiting value -1 given by Logan et. al. is seen to be rather rapid. (Details given in appendix.)

6. The Sign Test and Wilcoxon's Signed Rank Test.

Orthant symmetry of the vector $U = X/\|X\|$ is preserved under mar· amiliar statistical operations. In general we can define

$$\tilde{U} = g(U)$$

by simply specifying that g map every orthant into itself in a manner defined by a mapping

$$\tilde{\xi} = g^+(\xi)$$

of S_n^+ into itself. If U has orthant symmetry with measure λ_n^+ on S_n^+, then the mapped vector $\tilde{U}$ will also have orthant symmetry with induced measure

$$\tilde{\lambda}_n^+\{g^+(A)\} = \lambda_n^+\{A\}$$

for every $A \subset S_n^+$. The theorems and heuristics of the previous sections then apply as well to the statistic

$$\tilde{S}_n = \sum_{i=1}^{n} \tilde{U}_i$$

as to the original Student's statistic $S_n = \sum_{i=1}^{n} U_i$. To emphasize this point, we will call $\tilde{S}_n$ a generalized Student's statistic.

Generalized Student's statistics include most of the common nonparametric tests for the one sample problem. For instance, if $g^+(\xi) = e$, that is g maps U into $\frac{1}{\sqrt{n}}(\Delta_1,\Delta_2,\ldots,\Delta_n)$ where $\Delta_i = \text{sign}(U_i)$, then $\tilde{S}_n$ is simply the "sign test" for symmetry about the origin. If

$$g^+(\xi) = \sqrt{\frac{2}{n(n+1)}}\,(R_1,R_2,\ldots,R_n),$$

where R_i is the rank of ξ_i among $\{\xi_1,\xi_2,\ldots,\xi_n\}$, (the smallest having rank 1), then $\tilde{S}_n$ is Wilcoxon's signed rank test.

The effect of these transformations is to move ξ away from the corners of S_n^+. For the two transformations given above, it is easy to see that we move close enough to the center point e so that limiting normality is guaranteed under orthant symmetry*.

A reasonable question at this point is "why worry about limiting normality if the type I errors tend to be in the conservative direction in any case?" The answer, of course, is that we are also interested in the power of the test, which for the unmodified t-test may be nil in long-tailed cases. Power considerations are discussed, in an abbreviated manner, in the next two sections. It should be noticed that the question of power involves a property of the X vector which we have ignored up until now--namely its length, $\|X\|$. To see this, consider what happens to the distribution of the generalized binomial S_ξ if we move the center of orthant symmetry from the origin to the point

---

*For the sign test this is immediate, while for the signed rank test it follows from the Lindeberg condition, [5], p. 256.

$\delta e = (\frac{\delta}{\sqrt{n}}, \frac{\delta}{\sqrt{n}}, \ldots, \frac{\delta}{\sqrt{n}})$, $\delta > 0$, i.e. we undergo a component-wise translation to the right. The rough effect is to translate the distribution of S_ξ a random amount $\frac{\sqrt{n}\delta}{L_\xi}$ to the right, where L_ξ has the conditional distribution of $\|X\|$ given ξ, (calculated for the case $\delta = 0$). In the normal case $X_i \overset{ind}{\sim} N(\frac{\delta}{\sqrt{n}}, 1)$, L_ξ has a $\sqrt{\chi_n^2}$ distribution, independent of ξ, and all the S_ξ distributions translate in the same way. In general, this will not be the case. For translations of the inverted normals of Section 5, for instance, it is easy to show that $L_\xi \sim c(\xi)/\sqrt{\chi_n^2}$, where $c(\xi) = \sqrt{\sum_{i=1}^{n} \frac{1}{\xi_i^2}}$. Thus the translation effect on S_ξ, which yields the power of the test, is largest for $\xi = e$, and decreases as indicated as ξ moves away from e.

7. Legalized Cheating for Increased Power: Winsorization.

The two examples of generalized Student's statistics given in Section 6 relate to rank tests for the one-sample problem. Other familiar statistical operations can be discussed from this point of view. The example of this section relates permutation tests to Winsorization via generalized Student's statistics.

Consider the problem of testing the null hypothesis

$$H_o : X_i \text{ i.i.d. random variables, symmetric about } 0$$

versus the specific alternative

$$H_1 : X_i \text{ i.i.d. with density } \frac{1}{2\sigma} e^{-\frac{1}{\sigma}|X_i - \mu|}, \quad \mu > 0.$$

It is well known that if we want a genuine level α test for H_0, we must content ourselves with a permutation test (see [7], page 203, problem 11). That is, we must condition on

$$Y = (Y_1, Y_2, \ldots, Y_n), \quad Y_i = |X_i| \quad i = 1,2,\ldots,n,$$

and for each Y choose 100α% of the 2^n possible vectors $X = (\delta_1 Y_1, \delta_2 Y_2, \ldots, \delta_n Y_n)$, $\delta_i = \pm 1$, as rejection points.

Just as in the more familiar two-sample permutation test ([7], page 175), we maximize the conditional power of the test if we reject for those X with the maximum probability density under H_1,

$$f_{H_1}(X) = (2\sigma)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma}\sum_{i=1}^{n}|\delta_i Y_i - \mu|} .$$

Since we can write

$$\sum_{i=1}^{n} |\delta_i Y_i - \mu| = \sum_{i=1}^{n} (Y_i - Y_i^{[\mu]}) + n\mu - \sum_{i=1}^{n} \delta_i Y_i^{[\mu]} ,$$

where $Y_i^{[\mu]} = \min(Y_i, \mu)$, it is equivalent to reject for those choices of $\delta = (\delta_1, \delta_2, \ldots, \delta_n)$ maximizing

$$\sum_{i=1}^{n} \delta_i Y_i^{[\mu]} .$$

Now suppose we do not know the correct value of μ, so we "cheat" by first looking at the values of $Y_1, Y_2, \ldots, Y_n$, and calculating some scale invariant estimate of μ, say $\hat{\mu}(Y)$,

$$\hat{\mu}(cY) = c\hat{\mu}(Y) \qquad \text{for all } c > 0.$$

[An example would be to choose $\hat{\mu}$ to maximize the number of observed Y_i points in $\hat{\mu} \pm \varepsilon \|Y\|$, where ε is a positive constant; i.e. choose

$\hat{\mu}$ to be a modal point of the Y_i's. Or we could use maximum likelihood estimation, choosing $(\hat{\mu},\hat{\sigma})$ to maximize $\prod_{i=1}^{n} \frac{1}{2\sigma}(e^{-\frac{1}{\sigma}|Y_i-\mu|} + e^{-\frac{1}{\sigma}|Y_i+\mu|})$.]

The mapping which takes Y into $Y^{[\hat{\mu}]} = (Y_1^{[\hat{\mu}]}, Y_2^{[\hat{\mu}]}, \ldots, Y_n^{[\hat{\mu}]})$ takes rays into rays, that is it takes cY into $cY^{[\hat{\mu}]}$ for any $c > 0$. It therefore induces a mapping on S_n^+ of the form discussed in Section 6:

$$\tilde{\xi} = g^+(\xi) = \xi^{[\hat{\mu}(\xi)]}/\|\xi^{[\hat{\mu}(\xi)]}\|.$$

This is turn induces a mapping $\tilde{U} = g(U)$ on all of S_n by copying the map g^+ in each of the 2^n-1 other orthants. The corresponding generalized student's test, which rejects for

$$\tilde{S}_n = \sum_{i=1}^{n} \tilde{U}_i > \tilde{s}_{\xi,\alpha}$$

is seen to be an approximation to the most powerful level α test for H_0 versus H_1. Theoretically $\tilde{s}_{\xi,\alpha}$ should be chosen to give $\alpha \cdot 2^n$ values of $\sum_{i=1}^{n} \delta_i \tilde{\xi}_i$, $\delta_i = \pm 1$, greater than $\tilde{s}_{\xi,\alpha}$, but from our previous discussion of generalized binomials we feel safe in choosing $\tilde{s}_{\xi,\alpha} = s_{n,\alpha}$, the upper α point of S_n under normality, or even more simply $\tilde{s}_{\xi,\alpha} = z_\alpha$, the upper α point of a $N(0,1)$ random variable. (Note that the mapping $\tilde{\xi} = g(\xi)$ once again moves us away from the corners of S_n^+.) We know that the generalized Student's test we have constructed will have approximate size α for the null hypothesis of orthant symmetry, which includes H_0, and if the estimate $\hat{\mu}$ of μ is at all accurate, it should have good power under H_1.

There is nothing particularly compelling about the choice of a double exponential distribution for the X_i, except that it leads to a Student's test based on Winsorized values of the observations, rather than the raw values. What is striking, though, is how little one may deviate from the normal translation model without inducing drastic changes in the form of the appropriate test statistic (c.f. [12]).

In general, suppose we are testing H_0 versus

$$H_1 : X_i \text{ i.i.d. with density } \frac{1}{\sigma} f\left(\frac{X_i - \mu}{\sigma}\right), \quad \mu > 0.$$

Defining $Y_i = |X_i|$, $i = 1,2,\ldots,n$, as before, the most powerful level α test is a permutation test which rejects for the 100 % largest values of $\sum_{i=1}^{n} \delta_i \tilde{Y}_i$, $\delta_i = \pm 1$, where

$$\tilde{Y}_i = \frac{1}{2} \log \frac{f\left(\frac{Y_i - \mu}{\sigma}\right)}{f\left(\frac{-Y_i - \mu}{\sigma}\right)}, \qquad i = 1,2,\ldots,n.$$

If we do not know the parameters μ and σ, we can estimate them from the absolute values $Y_1, Y_2, \ldots, Y_n$ in any way we want, subject only to the restriction that the resulting mapping $Y \rightarrow \tilde{Y}$ takes rays into rays (maximum likelihood estimation will always have this property). As in the exponential case, we are led to a generalized Student's test which rejects for large values of S_n. The mapping $\tilde{\xi} = g^{+}(\xi)$ which determines which generalized Student's test we use is given by

$$\tilde{\xi}_i = \frac{1}{2} \log \frac{f\left(\frac{\xi_i - \hat{\mu}}{\hat{\sigma}}\right)}{f\left(\frac{-\xi_i - \hat{\mu}}{\hat{\sigma}}\right)}, \qquad i = 1,2,\ldots,n,$$

where $(\hat{\mu},\hat{\sigma})$ is the estimate of μ and σ given that $Y = \xi$. We have not actually cheated on our α-level, since mappings based on Y preserve orthant symmetry under the null hypothesis.

If f is the normal density then the procedure above yields the ordinary Student's test, but with other even slightly different kernals, far different tests are called for. This approach is not limited to translation and scale parameter families, but the author has not investigated the more interesting problem of obtaining useful estimates of f from the absolute values Y_i in general situations.

8. <u>Simultaneous Use of Student's Test and The Sign Test</u>.

Another approach to safeguarding the power of a one-sample test is to use more than one test on the data. For instance, we might use Student's statistic $S_n = \sum_{i=1}^{n} U_i$ in conjunction with the sign test $\tilde{S}_n = \sum_{i=1}^{n} \mathrm{Sign}(U_i)/\sqrt{n}$. In the language of Section 6, we would be simultaneously using two generalized Student's statistics, one based on ξ itself, the other on $\tilde{\xi} = e$.

If $\tilde{S}_n$ is any generalized Student's statistic, based on the mapping $\tilde{\xi} = g^+(\xi)$, then the vector $(S_n,\tilde{S}_n)$ can be expressed as

$$(S_n,\tilde{S}_n) = \sum_{i=1}^{n} \Delta_i(\xi_i,\tilde{\xi}_i)$$

where, as before, $\Delta_i = \mathrm{Sign}(X_i)$, $i = 1,2,\ldots,n$. Conditioning on the value of ξ, the Δ_i are independently ± 1 with probabilities $\frac{1}{2}$ under the null hypothesis of orthant symmetry. Conditionally, the random vector will have mean (0,0) and covariance matrix

$$\begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} ,$$

where $r = \sum_{i=1}^{n} \xi_i \tilde{\xi}_i = \xi \cdot g^+(\xi)$. In the case of the sign test, $r = \sum_{i=1}^{n} \xi_i/\sqrt{n}$. By the central limit theorem, $(S_n, \tilde{S}_n)$ will have, approximately, a normal distribution

$$(S_n, \tilde{S}_n) \sim N\left((0,0), \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}\right).$$

the approximation holding best for ξ and $\tilde{\xi}$ near e.

Now if we wish to use both S_n and $\tilde{S}_n$ simultaneously on the same set of data, we can accept the normal approximation as being sufficiently accurate, and base our decision on

$$\hat{S}_n = \max(S_n, \tilde{S}_n)$$

whose distribution can be read out of standard bivariate normal tables. Here is a small table for the approximate upper 5% point of $\hat{S}_n$:

$r = \sum_{i=1}^{n} \xi_i \tilde{\xi}_i$	.50	.55	.60	.65	.70	.75	.80	.85	.90	.95	1.00
Approximate 5% Point	1.91	1.91	1.90	1.89	1.88	1.86	1.85	1.83	1.80	1.76	1.65

These numbers should be compared with 1.96, the upper 5% point if you use the usual bound $P(\hat{S}_n < s) \geq 1 - P(S_n > s) - P(\tilde{S}_n > s)$, and 1.65, the upper 5% point if you perform either one of the tests seperately.

The value of η depends only on ξ, and so, as in the last section, we can compute it before we decide whether or not we want to use a simultaneous test. In the case where $\tilde{S}_n$ is the sign test, $r = \sum_{i=1}^{n} \xi_i/\sqrt{n}$, the computed value of r should ordinarily be quite large. If the X_i are i.i.d. with $X_i \sim X_0 + \frac{\varepsilon}{\sqrt{n}}$ for $i = 1,2,\ldots,n$, where X_0 has a finite second moment and is symmetrically distributed about the origin, then r will approach in probability the constant $E|X_0|/\sqrt{EX_0^2}$ as n goes to infinity. If $X_0 \sim N(0,\sigma^2)$ this limit is $\sqrt{2/\pi} = .798$, while if X_0 is double exponential the limit equals $\frac{1}{\sqrt{2}} = .707$. The variance of r in the case of normal components is about $\frac{.2}{n}$. If the computed value of r is not large, we have a strong indication of non-normality, and it is probably best not to use S_n at all.

The normal approximation to the conditional joint distribution of $(S_n,\tilde{S}_n)$ given ξ matches exactly the first and second moments (i.e. the mean vector and covariance matrix), and is conservative with respect to the higher moments exactly as in the theorem of Section 3: we can consider the general case of k simultaneous generalized Student's statistics,

$$\tilde{S} = (\tilde{S}_n(1),\tilde{S}_n(2),\ldots,\tilde{S}_n(k)) = \sum_{i=1}^{n} \Delta_i(\tilde{\xi}_i(1),\tilde{\xi}_i(2),\ldots,\tilde{\xi}_i(k)),$$

where ξ determines the k vectors $\tilde{\xi}(j)$ via $\tilde{\xi}(j) = g_j^+(\xi)$, and given ξ, the Δ_i are independently +1 or -1 with probabilities $\frac{1}{2}$ as before.

$\tilde{S}$ has conditional mean vector $(0,0,\ldots,0)$ and covariance matrix $\ddagger_\xi = [\tilde{\xi}(j_1)\cdot\xi(j_2)]_{j_1,j_2} = 1,2,\ldots,k$. For any vector $V = (V_1,V_2,\ldots,V_k)$, we then have

$$E(V\cdot\tilde{S})^\nu < (V\ddagger_\xi V')^{\frac{\nu}{2}}\, EN^\nu(0,1)$$

for $\nu = 4,6,8,\ldots$. (This follows from Section 3 by noting that $V\cdot\tilde{S}$ is itself a generalized binomial scaled by a factor $(V\ddagger V')^{1/2}$.)

The expectation here is conditional with respect to the observed value of ξ. The inequality may not hold with respect to the unconditional distribution of $\tilde{S}$, which has mean vector 0 and covariance matrix $\ddagger = E\ddagger_\xi$. This brings up an interesting point: there is no particular reason to approximate the unconditional distribution of $\tilde{S}$ with a multivariate normal, since it is really a mixture of such approximations with different covariance matrices. Asymptotic normality of $\tilde{S}$ comes from the fact that under certain conditions $\ddagger_\xi$ will go to a limiting matrix in probability as n gets large. However, for moderate n it seems more sensible to work directly with the conditional distribution, which is a fortiori approximately normal. This point is made more emphatically in the next section.

9. Conditional Versus Unconditional Distribution: Angle From an Arbitrary Vector.

So far we have been able to gloss over the distinction between applying the generalized Student's tests conditionally (conditional on ξ) as opposed to unconditionally. This was primarily because the conditional random variable S_ξ had the same mean and variance for all values of ξ. We can destroy the pleasant situation, and further explore the nature of the approximations we have been using, by considering the random angle between X and an arbitrary fixed vector $c \in S_n^+$, $c \neq e$. Let $\theta_{x,c}$ be this angle, and define

$$S_n(c) = \sqrt{n} \operatorname{Cos} \theta_{x,c} = \sqrt{n} \sum_{i=1}^{n} c_i U_i,$$

where $U = X/\|X\|$ as before (recalling that $S_n = \sqrt{n} \operatorname{Cox} \theta_{x,e}$).

Conditioning on $\xi = (|U_1|, |U_2|, \ldots, |U_n|)$, this can be written as

$$S_\xi(c) = \sqrt{n \sum_{i=1}^{n} c_i^2 \xi_i^2} \sum_{i=1}^{n} \Delta_i \tilde{\xi}_i$$

where

$$\tilde{\xi}_i = \frac{c_i \xi_i}{\sqrt{\sum_1^n c_i^2 \xi_i^2}} \qquad i = 1,2,\ldots,n,$$

and under orthant symmetry the Δ_i equal +1 or -1 independently with probabilities $\frac{1}{2}$. The sum $S_{\tilde{\xi}} = \sum_{i=1}^{n} \Delta_i \tilde{\xi}_i$ is a generalized binomial random variable, as defined in Section 3, and we see that the mixture lemma of that section takes the following form in our present situation:

$$P(S_n(c) < s) = \int_{S_n^+} P\left(S_{\tilde{\xi}} < \frac{s}{\sqrt{n \sum_{i=1}^{n} c_i^2 \xi_i^2}}\right) d\lambda_n^+(\xi) .$$

If n is moderately large and the components $\tilde{\xi}_i$ are not too drastically different from one another, the conditional distribution of $S_\xi(c)$ can be well approximated by a $N(0, n\sum_1^n c_i^2 \xi_i^2)$ distribution. The conditional variance has expected value 1 over all realizations of ξ, but in a testing situation we might prefer to work directly with the conditional value $n\sum_1^n c_i^2 \xi_i^2$, particularly since we will usually be unable to approximate the unconditional distribution of $S_n(c)$, except indirectly via the mixture lemma. Thus we may have asymptotic normality for $S_n(c)$, but this will derive from the more direct limiting normality of the $S_\xi(c)$ and the fact that $\sqrt{n \sum_{i=1}^{n} c_i^2 \xi_i^2}$ approaches a constant as n grows large. The moments theorem,

$$ES_\xi^\nu(c) < (n\sum_1^n c_i^2 \xi_i^2)^{\nu/2} \; EN^\nu(0,1) ,$$

for $\nu = 4,6,8,\ldots$, may not hold for $S_n(c)$.

Let us consider a hypothetical example: suppose we wish to test $H_0 : X_i \overset{\text{ind}}{\sim} N(0,\sigma^2)$ vs $H_1 : X_i \overset{\text{ind}}{\sim} N(\beta i,\sigma^2)$, $\beta > 0$, $i = 1,2,\ldots,n$ (a "regression alternative"), so that the UMP(α) test is to reject for large values of S_n(c , $c = \frac{2}{n(n+1)}(1,2,3,\ldots,n)$, We observe $S_n(c) = 1.75$, which is at about the .04 level of the unconditional distribution. However, we compute $\sqrt{n\sum_1^n c_i^2\xi_i^2} = 1.5$, so the significance level in the conditional distribution is only about .13. If we have a great deal of confidence in our normal model we will probably believe the .04 significance. However, the size of $\sqrt{n\sum_1^n c_i^2\xi_i^2}$ already points to some abnormality in the data, and we will be a good deal safer if we follow the conditional inference*.

There are, of course, ways we can retreat part way from the full normal hypothesis, without going all the way to the test based on orthant symmetry. We could, for instance, take H_0 to be "the X_i are i.i.d. symmetric about the origin", and test conditionally given the order statistic of ξ, $\xi_{[1]} \le \xi_{[2]} \le \cdots \le \xi_{[n]}$. Under H_0, the resulting statistic will be an equally weighted mixture of n! scaled generalized binomials, corresponding to taking all n! permutations of the order statistic to give different ξ vectors. The scaling factors $\sqrt{n\sum_1^n c_i^2\xi_{\pi(i)}^2}$ average to no more than unity,

$$\frac{1}{n!}\sum_{\pi}\sqrt{n\sum_1^n c_i^2\xi_{\pi(i)}^2} \le 1,$$

*Metaphysical statements of preference between the two modes of inference abound in the literature, but no compelling criterion of selection seems to exist at present.

by the concavity of the square root function. Therefore, we might feel that a $N(0,1)$ approximation to the statistic would tend to be conservative. However, in this case it is not necessarily true that the higher moments of the statistic will be bounded by those of a $N(0,1)$ random variable.

Other sampling characteristics of the angular distribution of X can be approximated from the central limit theorem. For example, let H_c a fixed k-dimensional subspace of E^n, determined by the orthonormal spanning vectors $c_1, c_2, \ldots, c_k$. The conditional distribution of $n \operatorname{Cos}^2\theta_{X,c}$ given ξ, where $\theta_{X,c}$ is the angle between X and H_c, is approximated by $\sum_{j=1}^{k} \ell_j(\xi)\chi^2(j)$, where the $\chi^2(j)$ are independent χ_1^2 random variables, and the $\ell_j(\xi)$ are the eigenvalues of

$$nC'\Xi^2C,$$

$C = (c_1, c_2, \ldots, c_k)$, Ξ = the diagonal matrix with $\xi_1, \xi_2, \ldots, \xi_n$ as diagonal elements. These considerations are relevant to the permutation distribution of Hotelling's T^2, which the author will consider in a companion paper.

10. Hotelling's Paper and Other References.

This work was stimulated by Hotelling's 1961 paper "The Behavior of Some Standard Statistical Tests Under Non-standard Conditions" [11]. After setting up the geometry, Hotelling approximates the size of the t-test by $\alpha \frac{d\lambda_n^+(e)}{d\lambda(e)}$ (in our notation). This approximation requires $C_{n,\alpha}$, the rejection set, to be small enough so that the measure λ_n has close to constant density over it, which leads Hotelling to consider very small α levels, $\alpha < \frac{1}{2^n}$.

For this range of α, he shows that the size of the t-test relative to the nominal size may vary from 0 to ∞, even with i.i.d. symmetric bounded components. Since we know that reasonable n and α actually yield very large sets $C_{n,\alpha}$ spreading over a good portion of S_n, it is not surprising that Hotelling's results are quite different from those developed here.

By now it is a matter of some hubris to claim originality for any topic bearing on the t-test. Many of the topics presented here have been discussed by other authors. Hoeffding's 1952 paper [10] is particularly relevant. The case of the double exponential with a translation parameter, discussed in Section 7, has been investigated by Lehmann [13], and others. For an extensive review and bibliography of Student's test under non-normal conditions the reader is referred to [9].

The author is indebted to R. R. Bahadur and M. J. Eaton of Chicago, and J. Hartigan of Princeton for enlightening conversation and correspondence on the subject presented here.

REFERENCES

[1] Chung, K. L., 1946. The approximate distribution of Student's t statistic. Annals Math. Stat., Vol. 17, 447-465.

[2] Cramer, H., 1946. Mathematical Methods of Statistics, Princeton University Press, Princeton, New Jersey.

[3] Darling, D.A., 1952. The influence of the maximum term in the addition of independent random variables. Trans. Amer. Math. Soc., Vol. 73, 95-107.

[4] Dwight, H.B., 1957. Tables of Integrals and Other Mathematical Data. Macmillan, New York.

[5] Feller, W., 1966. An Introduction to Probability Theory and Its Applications, Vol. II, John Wiley, New York.

[6] Fisher, R.A., 1925. Applications of Student's distribution. Metron, Vol. 5, 90-104.

[7] Fraser, D.A.S., 1957. Nonparametric Methods in Statistics, John Wiley, New York.

[8] Hajek, J., 1962. Inequalities for the generalized Student's distribution. Sci. Trans. Math. Stat. Prob., Vol. 2, 63-74

[9] Hatch, L.O., and Posten, H.O., 1967. Robustness of the Student procedure. Research Report No. 24, Department of Statistics, University of Connecticut, Storrs, Connecticut.

[10] Hoeffding, W., 1952. The large sample power of tests based on permutations of observations. Annals of Math. Stat., Vol. 23, No. 2, 169-192.

[11] Hotelling, H., 1961. The behavior of some standard statistical tests under non-standard conditions. Proc. 4th Berkeley Symp., Vol. 1, 319-360.

[12] Huber, P., 1964. Robust estimation of location. Annals of Math. Stat., Vol. 35, No. 1, 73-101.

[13] Lehmann, E., 1955. Ordered families of distributions. Annals of Math. Stat., Vol. 26, No. 3, 399-419.

[14] Logan, B.F., Mallows, C.L., Rice, S.O., and Shepp, L.A., 1968. Asymptotic behavior of t in the stable case. Unpublished.

Appendix of Mathematical Proofs

Section 3: $ES_\xi^\nu < ES_e^\nu$ for all $\xi \epsilon S_n^+$, $\xi \neq e$, for $\nu = 4,6,8,\ldots$.

Proof: Assume we have proven the result for the case n-1, and write $S_\xi = cS_\xi + \xi_n \Delta_n$, where $c = \sqrt{\sum_1^{n-1} \xi_i^2}$ and $\dot{S}_\xi = \sum_{i=1}^{n-1} \Delta_i \xi_i / c$. Using the symmetry about zero and independence of $\dot{S}_\xi$ and Δ_n (remember that these calculations are conditional on ξ), we get

$$ES_\xi^\nu = c^\nu E\dot{S}_\xi^\nu + \binom{\nu}{2} c^{\nu-2} \xi_n^2 E\dot{S}_\xi^{\nu-2} + \ldots + \xi_n^\nu .$$

By the induction hypothesis, this expression will be increased if we change ξ_i to $\sqrt{(1-\xi_n^2)/(n-1)}$ for $i = 1,2,\ldots,n-1$, unless the first n-1 ξ_i were already equal. By applying the same argument to the last n-1 ξ_i, we see that $\xi = e$ is the only possible maximum point for ES_ξ^ν over the compact set S_n^+.

It remains to verify the result for the case n = 2. We have

$ES_\xi^\nu = \frac{1}{2}[(\xi_1+\xi_2)^\nu + (\xi_1-\xi_2)^\nu] = \frac{1}{2}[(\sqrt{\gamma} + \sqrt{1-\gamma})^\nu + (\sqrt{\gamma} - \sqrt{1-\gamma})^\nu]$,

where $\gamma = \xi_1^2$. Thus

$$\frac{dES_\xi^\nu}{d\gamma} = \frac{\nu}{4}[(\sqrt{\gamma} + \sqrt{1-\gamma})^{\nu-1}(\frac{1}{\sqrt{\gamma}} - \frac{1}{\sqrt{1-\gamma}}) + (\sqrt{\gamma} - \sqrt{1-\gamma})^{\nu-1}$$

$$(\frac{1}{\sqrt{\gamma}} + \frac{1}{\sqrt{1-\gamma}})],$$

which can be written as

$$\frac{\nu}{2}\left\{[\binom{\nu-1}{0}-\binom{\nu-1}{0}][\sqrt{\gamma}^{\nu-2} - \sqrt{1-\gamma}^{\nu-2}] + [\binom{\nu-1}{2}-\binom{\nu-1}{3}][\sqrt{\gamma}^{\nu-4}\sqrt{1-\gamma}^{2} - \sqrt{\gamma}^{2}\sqrt{1-\gamma}^{\nu-4}] \right.$$

$$\left. + \ldots\ldots + [\binom{\nu-1}{\frac{\nu}{2}-2}-\binom{\nu-1}{\frac{\nu}{2}-1}][\sqrt{\gamma}^{\nu/2}\sqrt{1-\gamma}^{\nu/2-2} - \sqrt{\gamma}^{\nu/2-2}\sqrt{1-\gamma}^{\nu/2}]\right\} .$$

This is negative for $\gamma > \frac{1}{2}$ and positive for $\gamma < \frac{1}{2}$, showing the ES_ξ^ν attains its maximum for $\gamma = \frac{1}{2}$, or $\xi = e$.

$ES_e^{\nu} < EN^{\nu}(0,1)$ for $\nu = 4,6,8,\ldots$.

Proof: $N(0,1) = \sum_1^n Z_i/\sqrt{n}$, $Z_i \overset{\text{ind}}{\sim} N(0,1)$, and $S_e = \sum_1^n \Delta_i/\sqrt{n}$ where the Δ_i are independent, $\Delta_i = +1$ or -1 with probabilities $\frac{1}{2}$. The result follows immediately from the fact that $EZ_i^j \geq E\Delta_i^j$ for j even, with strict inequality for $j \geq 4$.

Section 4: Edgeworth Expansion for $P(S_n < s)$.

If $S = \sum_1^n V_i/\sqrt{n}$, where the V_i are independent random variables, symmetric about zero, $\sum_{i=1}^n \sigma^2(V_i) = n$, then the Edgeworth expansion for $P(S_n < s)$ (see [2], pp. 221-231) can be written as

$$P(S_n < s) = \Phi(s) + \frac{1}{n}\frac{\bar{\chi}_4}{4!}\,\Phi^{(4)}(s) + \frac{1}{n^2}\left\{\frac{\bar{\chi}_6}{6!}\,\Phi^{(6)}(s) + \frac{1}{2}\left(\frac{\bar{\chi}_4}{4!}\right)\Phi^{(8)}(s)\right\}$$
$$+ \frac{1}{n^3}\left\{\frac{\bar{\chi}_8}{8!}\,\Phi^{(8)}(s) + \frac{\bar{\chi}_4\bar{\chi}_6}{4!6!}\,\Phi^{(10)}(s) + \frac{1}{6}\left(\frac{\bar{\chi}_4}{4!}\right)\Phi^{(12)}(s)\right\}$$
$$+ \ldots .$$

Here $\bar{\chi}_\nu$ is the average ν^{th} cumulant,

$$\bar{\chi}_\nu = \frac{\sum_1^n \chi_\nu(V_i)}{n},$$

where we recall that the characteristic function of V_i defines $\chi_\nu(V_i)$ by

$$\log \phi_{V_i}(t) = \sum_1^\infty \frac{\chi_\nu(V_i)}{\nu!}(it)^\nu.$$

The superscripts on Φ indicate repeated differentiation. The terms are grouped in such a way that the indicated orders of

ıde in n hold for the case of the V_i i.i.d.

In our case we let $V_{\xi,i} = \sqrt{n}\,\xi_i\Delta_i$, and note that the characteristic function is $\phi_{V_i}(t) = \mathrm{Cos}\sqrt{n}\xi_i t$. A standard expansion of log cos now yields $\bar{\chi}_{\xi,\nu} = c_\nu n^{\nu/2-1}\sum_1^n \xi_i^\nu$ for $\nu = 2,4,6,\ldots$, where

$$c_\nu = \frac{(-1)^{\frac{\nu}{2}+1}\,2^\nu(2^\nu-1)}{\nu}\,B_{\nu/2} = (-1)^{\frac{\nu}{2}+1}(\nu-1)!\,2\left(\frac{2}{\pi}\right)^\nu\left[1+\frac{1}{3^\nu}+\frac{1}{5^\nu}+\ldots\right],$$

$B_{\nu/2}$ being the $\nu/2^{\text{th}}$ Bernoulli number (ref [4], #603.3 and #47.3).

If we use the Edgeworth expansion with the values $\bar{\chi}_{\xi,\nu}$, we get an expansion for the generalized binomial probability $P(S_\xi < s)$. From the mixture lemma, $P(S_n < s) = E_\xi(P(S_\xi < s))$, and we can take this expectation term by term in the expansion. The leading correction term to $\Phi(s)$, $\frac{1}{n}\frac{\bar{\chi}_{\xi,4}}{4!}\Phi^{(4)}(s)$, has expectation

$$E_\xi\frac{1}{n}\left(\frac{\bar{\chi}_{\xi,4}}{4!}\right)\Phi^{(4)}(s) = \frac{c_4}{4!}\,\Phi^{(4)}(s)\sum_1^n E\xi_i^4 .$$

If we assume exchangeable components then $\sum_1^n E\xi_i^4 = nE\xi_1^4 = nEU_1^4$. Proceeding in this way yields the expansion of Section 4.

Section 5: Angular Distribution For Vector With Inverted Normal Components.

We calculate the angular distribution of a random vector X with components $X_i = 1/\tilde{X}_i$, where $\tilde{X}_i \overset{\text{ind}}{\sim} N(0,1)$. Rather than work with the coordinates $\xi_1,\xi_2,\ldots,\xi_n$ on S_n^+, which are redundant and must be reduced to an n-1 component set, we calculate our densities with respect to the coordinates $v = (y_2,y_3,\ldots,y_n)$,

$$y_i \equiv \frac{\xi_i}{\xi_1} \qquad i = 2,3,\ldots,n,$$

taking values in $\mathcal{Y}^+ = \{y : y_i > 0,\ i = 2,3,\ldots,n\}$. (As before, it is sufficient to consider only the case of positive observations because of orthant symmetry.)

It is easy to verify that the normally distributed vector $\tilde{X} = (\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n)$ yields a density

$$f_{\tilde{Y}}(\tilde{y}) = c_n(1 + \tilde{y}_2^2 + \ldots + \tilde{y}_n^2)^{-\frac{n}{2}}$$

for $\tilde{y} \in \mathcal{Y}^+$, where $c_n = 2^{n-1}\Gamma(\frac{n}{2})/\pi^{\frac{n}{2}}$ (the "multivariate Cauchy distribution"). The transformation $X_i = 1/\tilde{X}_i$, $i = 1,2,\ldots,n$, induces the transformation $y_i = 1/\tilde{y}_i$, $i = 2,3,\ldots,n$, in $\mathcal{Y}^+$. We see that inverted normal components induce a density

$$f_Y(y) = c_n(\prod_2^n y_i^2)^{-1}(1 + 1/y_2^2 + 1/y_3^2 + \ldots + 1/y_n^2)^{-\frac{n}{2}}$$

on $\mathcal{Y}^+$. The Radon-Nikodyn derivative of this density with respect to the former is

$$\frac{f_Y(y)}{f_{\tilde{Y}}(y)} = (\prod_2^n y_i^2)^{-1}\left(\frac{1+y_2^2+y_3^2+\ldots+y_n^2}{1+1/y_2^2+1/y_3^2+\ldots+1/y_n^2}\right)^{\frac{n}{2}} .$$

In particular the derivative at $y = (1,1,\ldots,1)$, or equivalently at $\xi = e$, is equal to 1 as claimed in Section 5. If we approach the corner $\xi = (1,0,0,\ldots,0)$ of S_n^+ by way of vectors $y = (\varepsilon,\varepsilon,\ldots,\varepsilon)$, ε approaching 0, the derivative goes to infinity as $1/\varepsilon^n$.

<u>Calculation of Kurtosis of S_n for Inverted Normal Components.</u>

We have

$$\begin{aligned} \mathrm{kurt}(S_n) &= -2nEU_1^4 \\ &= -2nE[X_1^2/(X_1^2 + X_2^2 + \ldots + X_n^2)]^2, \end{aligned}$$

where $X_i = 1/\tilde{X}_i$, $\tilde{X}_i \overset{\text{ind}}{\sim} N(0,1)$. Using the fact that X_i^2 is the positive stable law of order $\frac{1}{2}$, this equals

$$-2nE[1 + (n-1)^2 X_2^2/X_1^2]^{-2} = -2nE[1 + (n-1)^2 \tilde{X}_1^2/\tilde{X}_2^2]^{-2}$$

$$= -\frac{4n}{\pi}\int_0^{\pi/2} [1 + (n-1)^2 \tan^2\theta]^{-2}\, d\theta,$$

the last step following from the fact that $\theta = \tan^{-1} \tilde{X}_1/\tilde{X}_2$ is uniformly distributed between 0 and π. The substitution $V = (n-1)\tan\theta$ gives

$$\mathrm{kurt}(S_n) = -\frac{4}{\pi}\left(\frac{n}{n-1}\right)\int_0^{\infty} (1+V^2)^{-2}\left(1 + \left(\frac{V}{n-1}\right)^2\right)^{-1} dV.$$

The approximation $\left(1 + \left(\frac{V}{n-1}\right)^2\right)^{-1} \cong 1 - \left(\frac{V}{n-1}\right)^2$ then gives

$$\mathrm{kurt}(S_n) \cong \frac{n}{n-1}\left[1 - \left(\frac{1}{n-1}\right)^2\right] = -1 - \frac{1}{n-1} + O(n^{-2}) .$$

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) Department of Statistics Harvard University Cambridge, Massachusetts	2a REPORT SECURITY CLASSIFICATION 2b GROUP

3 REPORT TITLE

Student's t-test Under Non-normal Conditions.

4 DESCRIPTIVE NOTES (Type of report and inclusive dates)
Technical Report; May 22, 1968

5 AUTHOR(S) (Last name, first name, initial)

Efron, Bradley

6 REPORT DATE	7a TOTAL NO. OF PAGES	7b NO OF REFS
May 22, 1968	36	14

8a CONTRACT OR GRANT NO. Nonr 1866(37) b PROJECT NO. c MR-042-097 d	9a ORIGINATOR'S REPORT NUMBER(S) Technical Report No. 21 9b OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10 AVAILABILITY/LIMITATION NOTICES

Distribution of this document is unlimited

11 SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Logistics & Mathematical Statistics Branch Office of Naval Research Department of the Navy Washington, D.C.

13 ABSTRACT

The size and power of Student's t-test are discussed under weaker than normal conditions. It is shown that assuming only a symmetry condition for the null hypothesis leads to effective bounds on the dispersion of the t-statistic. (The symmetry condition is weak enough to include all cases of independent but not necessarily identically distributed observations, each symmetric about the origin.) The connection between Student's test and the usual non-parametric tests is examined, as well as power considerations involving Winsorization and permutation tests. Simultaneous use of different one-sample tests is also discussed.

DD FORM 1 JAN 64 1473

Unclassified
Security Classification

Unclassified

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through ________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through ________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through ________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.

DD FORM 1 JAN 64 1473 (BACK)

Unclassified

Security Classification