The Annals of Statistics
1995, Vol. 23, No. 4, 1433–1442

ROBUSTNESS OF STANDARD CONFIDENCE INTERVALS FOR LOCATION PARAMETERS UNDER DEPARTURE FROM NORMALITY[1]

BY SANJIB BASU[2] AND ANIRBAN DASGUPTA

University of Arkansas and Purdue University

Let $X_i = \theta + \sigma Z_i$, where Z_i are i.i.d. from a distribution F, and $-\infty < \theta < \infty$ and $\sigma > 0$ are unknown parameters. If F is $N(0,1)$, a standard confidence interval for the unknown mean θ is the *t-interval* $\bar{X} \pm t_{\alpha/2} s/\sqrt{n}$. The question of conservatism of this interval under nonnormality is considered by evaluating the infimum of its coverage probability when F belongs to a suitably chosen class of distributions $\mathscr{F}$. Some rather surprising phenomena show up. For $\mathscr{F} =$ {all symmetric unimodal distributions} it is found that, for high nominal coverage intervals, the minimum coverage is attained at $U[-1,1]$ distribution, and the *t-interval* is quite conservative. However, for intervals with low or moderate nominal coverages ($t_{\alpha/2} < 1$), it is proved that the infimum coverage is zero, thus indicating drastic sensitivity to nonnormality. This phenomenon carries over to more general families of distributions. Our results also relate to robustness of the *P-value* corresponding to the *t-statistic* when the underlying distribution is nonnormal.

1. Introduction. Consider a standard location-scale setup: $X_i = \theta + \sigma Z_i$, $i = 1, \dots, n$; here $(Z_1, \dots, Z_n)$ is a random sample from a distribution F. If F is $N(0,1)$ and σ^2 is known, a standard frequentist confidence interval for unknown θ is given by the *z-interval*: $\bar{X} \pm z_{\alpha/2}\sigma/\sqrt{n}$, where $\bar{X} = (1/n)\sum_{i=1}^n X_i$ and $z_{\alpha/2}$ is the $(1-\alpha/2)$ percentile of the $N(0,1)$ distribution. For unknown σ, the relevant interval is the *t-interval*: $\bar{X} \pm t_{\alpha/2} s/\sqrt{n}$, where

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^n \left(X_i - \bar{X}\right)^2}$$

and $t_{\alpha/2}$ is the $(1-\alpha/2)$ percentile of the Student's t distribution with $(n-1)$ degrees of freedom. For later reference, $T_n(\mathbf{X}) = \sqrt{n}\,\bar{X}/s$ will denote the *t-statistic*.

Suppose now that the assumption "F is normal" is not justified. A natural question would be to seek how much we would lose in terms of coverage probability if we still use the abovementioned confidence intervals (as practitioners often do). Let $\rho_z(F)$ and $\rho_t(F)$ denote, respectively, the coverage of the

Received January 1992; revised December 1994.
[1]Research supported by NSF Grants DMS-89-23071 and DMS-93-07727.
[2]This work was part of the Ph.D. dissertation at Purdue University.
AMS 1991 *subject classifications*. 62F25, 62G35.
Key words and phrases. t-interval, coverage probability, P-value, infimum, normal scale mixture, symmetric, unimodal, ε-contamination, orthant symmetric

z-interval and the t-interval when the underlying distribution is F. If F has finite second moment, an easy application of the *central limit theorem* shows that both $\rho_z(F)$ and $\rho_t(F) \to 1 - \alpha$ as $n \to \infty$. This is often described by saying that these intervals are *asymptotically robust* against nonnormality. However, the central limit theorem per se does not give any indications about the sample size required for the normal approximations to be approximately valid for any specified F. Further, if F does not have a finite second moment, Logan, Mallows, Rice and Shepp (1973) showed that the limiting behavior of $\rho_t(F)$ can be quite strange.

Our main focus will be directed towards the small sample behavior of the t interval under nonnormality. Departure from normality will be modeled by requiring F to belong to a suitably chosen class of distributions $\mathscr{F}$, and we will judge the sensitivity of the intervals by considering the natural quantity $\rho = \inf_{F \in \mathscr{F}} \rho(F)$. In Section 2, we consider $\mathscr{F} = \mathscr{F}_1 = \{F: F$ is symmetric and unimodal (about 0)$\}$ and examine the performance of the t-interval. We first show that, for n (sample size) equal to 2, the minimum coverage $\inf_{F \in \mathscr{F}_1} \rho_t(F)$ is attained at symmetric uniform distributions if the critical value $t = t_{\alpha/2} \geq 1$, whereas for critical values less than 1, the minimum coverage is, in fact, zero. Next, we prove the surprising result that if $t < 1$, the infimum coverage over $\mathscr{F}_1$ is zero for all n. For intervals with higher nominal coverages (and general $n > 2$), we prove that if the cutoff $t \geq n - 1$, the infimum coverage over the class $\mathscr{F}_1$ is attained at symmetric uniform distributions. By a combination of moment theory techniques and numerical methods, we next demonstrate that the threshold beyond which the infimum is attained at a symmetric uniform is, in fact, much lower than $n - 1$ (though dependent on n). Generally, for the intervals with nominal coverage greater than or equal to 0.95, the minimal coverage is quite close to the nominal coverage, thus implying conservatism. Some discussions follow on the nature of the infimum coverage for t between 1 and the threshold. The section ends with exploration of the class of symmetric contaminations of $N(0,1)$. We again find that the t-interval is conservative for high critical values, whereas for intervals with nominal coverage less than or equal to 60%, the infimum coverage over this class gets significantly worse.

Section 3 explores the class of arbitrary contaminations of $N(0,1)$. We show that in small samples the t-interval behaves well, but its coverage can be significantly worse than the nominal coverage for moderate sample sizes. Section 4 looks at the robustness of the *P-value* of the *t-test*. Due to the obvious duality, our findings on the t-interval apply directly to considerations of the *P-value*. For arbitrary contaminations of $N(0,1)$, we show that surprisingly the maximum P-value differs more drastically from the nominal value as the sample size increases.

The principal achievements of this article are that we establish the surprising result that even if only symmetric unimodal populations are allowed, the minimum coverage of the t-interval is zero for all sample sizes whenever the cutoff t is less than 1, but for intervals with high nominal coverages the nominal and minimal coverages are close. The case of known σ^2, in which case the standard interval is the z-interval, is briefly commented on in the concluding section, Section 5.

2. Symmetric unimodal distributions.

2.1. *Preliminaries.* Let $X_i = \theta + \sigma Z_i$, $i = 1, \dots, n$, where $Z_1, \dots, Z_n$ are i.i.d. from a distribution $F \in \mathscr{F}_1 = \{F: F \text{ is symmetric and unimodal about zero}\}$. The class $\mathscr{F}_1$ contains heavy-tailed (Cauchy, double exponential, t) as well as short-tailed (symmetric uniforms and triangular) distributions. Because of the location structure, θ can be ignored, and we will consider $\inf_{F \in \mathscr{F}_1} P_F(|T_n(\mathbf{Z})| \le t)$ as our criterion, where $t > 0$ is arbitrary but fixed. Sometimes we will switch to the equivalent statistic $S_n(\mathbf{Z}) = \{\sum_{i=1}^n Z_i\}/\{\sum_{i=1}^n Z_i^2\}^{1/2}$; T_n and S_n are one–one, increasing functions of one another.

We need the following definition and the subsequent result from Benjamini (1983).

Definition 1. Let F and G be two symmetric distributions on $\mathfrak{R}$. The distribution F is stretched with respect to $G(F >_{\text{st}} G)$ if $\{F^{-1}(p) - F^{-1}(0.5)\}/\{G^{-1}(p) - G^{-1}(0.5)\}$ is increasing in p for $0.5 < p < 1$, where $F^{-1}(p) = \inf\{x: F(x) \ge p\}$.

In particular, a symmetric distribution G, obtained as a scale mixture of another fixed symmetric distribution H $[G(z) = \int_0^\infty H(z/\sigma)\, d\nu(\sigma)$ for some $\nu]$, is *stretched* with respect to H. Also, note that *stretching* allows a comparison between F and G even when G is discrete. This was explicitly used by Benjamini.

Theorem 1 (Benjamini). *Let F and G be two distributions on $\mathfrak{R}$, symmetric around* 0. *If $F >_{\text{st}} G$ and $t \ge n - 1$, then $P_F(T_n(\mathbf{Z}) > t) \le P_G(T_n(\mathbf{Z}) > t)$.*

For a proof, see Benjamini (1983); also see Basu (1991). Using this result and some numerical evidence, Benjamini showed that for $\mathscr{F}$ equal the normal scale mixture family, $\inf_{F \in \mathscr{F}} P_F(|T_n(\mathbf{Z})| \le t)$ is attained at $N(0,1)$ for all $n \ge 2$, if $t > 1.8$.

Theorem 2. *For any $n \ge 2$, $\inf_{F \in \mathscr{F}_1} P_F(|T_n(\mathbf{Z})| \le t) = P_{U[-1,1]}(|T_n(\mathbf{Z})| \le t)$ for $t \ge n - 1$.*

Proof. Any $F \in \mathscr{F}_1$ can be written as $F(z) = \int_0^\infty U_s(z)\, dG(s)$, where $U_s(z)$ is the c.d.f. of the $U[-s, s]$ distribution and G is a distribution on $[0, \infty)$ (Khintchine representation). Thus, $F >_{\text{st}} U[-s, s]$. The conclusion follows from Theorem 1 and scale invariance of $T_n(\mathbf{Z})$. $\square$

2.2. *Sample of size* 2. The following corollary follows immediately from Theorem 2.

COROLLARY 1. *For* $n = 2$, $\inf_{F \in \mathcal{F}_1} P_F(|T_2(\mathbf{Z})| \le t) = t/(t+1)$ *for* $t \ge 1$.

The "$t = 1$" corresponds to a nominal coverage of 50% for $n = 2$. We next examine the sensitivity of the t-interval when $t < 1$.

THEOREM 3. *For* $n = 2$, $\inf_{F \in \mathcal{F}_1} P_F(|T_2(\mathbf{Z})| \le t) = 0 \ \forall t < 1$.

PROOF. An anonymous referee has graciously provided us with the following constructive proof. Let U_1 and U_2 be i.i.d. $\sim U[-1,1]$. For $\alpha > 0$, define $X_i = \operatorname{sign}(U_i)|U_i|^{1/\alpha}$, $i = 1,2$. Clearly, X_1 and X_2 are i.i.d. from a distribution F which is symmetric and unimodal about zero. Notice that $T_2(X_1, X_2) = (X_1 + X_2)/|X_1 - X_2|$. For $0 < t < 1$, simple calculations show that $\{(x_1, x_2): |T_2(x_1, x_2)| \le t\}$ is simply the region within the square $[-1,1] \otimes [-1,1]$ bounded by the two straight lines $(1-t)x_1 + (1+t)x_2 = 0$ and $(1+t)x_1 + (1-t)x_2 = 0$. Integration yields

$$P(|T_2(X_1, X_2)| \le t) = \frac{1}{2}\left[1 - \left(\frac{1-t}{1+t}\right)^{\alpha}\right],$$

which goes to 0 as $\alpha \to 0$. This completes the proof. □

Thus, there is a surprising discontinuity at the point $t = 1$, below which the infimum coverage is zero, whereas for $t = 1$ the minimum equals 0.5, and then it continuously increases.

2.3. *General n.* We first prove that the assertion of Theorem 3 holds for all $n \ge 2$. In fact, it holds for a very general class of distributions $\mathcal{F}$.

THEOREM 4. *Let* $\mathcal{F}$ *be any nonempty family of distributions on* $\Re$ *satisfying the following*:

(i) $F \in \mathcal{F} \Rightarrow F$ *is continuous at* 0 *and symmetric*;
(ii) $F \in \mathcal{F} \Rightarrow$ *for any* $\sigma > 0$, *the distribution G defined by* $G(y) = F(y/\sigma)$ *also belongs to* $\mathcal{F}$;
(iii) $F_1, F_2 \in \mathcal{F} \Rightarrow \alpha F_1 + (1-\alpha)F_2 \in \mathcal{F}$ *for any* $0 \le \alpha \le 1$.

Then $\inf_{F \in \mathcal{F}} P_F(|T_n| \le t) = 0 \ \forall n \ge 2$ *if* $t < 1$.

The idea behind Theorem 4 is simple. Suppose, for fixed $t < 1$, the above infimum $= \gamma$ (say) is strictly positive and is attained at $F \in \mathcal{F}$. Consider $G(x) = (1-\varepsilon)F(x/\sigma) + \varepsilon F(x)$, where ε and σ are both positive and small. Then

$$\begin{aligned} P_G(|T_n| \le t) \le{} & (1-\varepsilon)^n P\left(|T_n| \le t \,|\, X_i \text{ i.i.d. } F\left(\frac{\cdot}{\sigma}\right)\right) \\ & + n\varepsilon(1-\varepsilon)^{n-1} P\left(|T_n| \le t \,|\, X_n \sim F(\cdot),\, X_1, \dots, X_{n-1} \text{ i.i.d. } F\left(\frac{\cdot}{\sigma}\right)\right) \\ & + \sum_{i=2}^{n} \binom{n}{i} \varepsilon^i (1-\varepsilon)^{n-i}. \end{aligned}$$

The first term in the above expansion equals γ by scale invariance of T_n, whereas the third term is small for small ε. For the second term, since X_n is much larger than the other X_i's with high probability, $\sum_{i=1}^n X_i \approx X_n$ and $\sum_{i=1}^n X_i^2 \approx X_n^2$, making $|T_n| \approx 1$. Hence, the second term is approximately equal to 0 since $t < 1$. Thus, $P_G(|T_n| \le t) \le (1-\varepsilon)^n\gamma +$ (a small term) $< \gamma$, leading to a contradiction. The formal proof is available in Basu and DasGupta (1992).

The subclass of continuous distributions in the symmetric unimodal family $\mathscr{F}_1$ satisfies conditions (i)–(iii); so does the normal scale mixture family. Hence, Theorem 4 holds for both of these families; "$t = 1$" corresponds to nominal coverages of 50, 62 and 68%, for $n = 2$, 5 and ∞, respectively. Thus, for $n \ge 5$ and for an interval with nominal coverage less than 62%, the infimum coverage is zero. This is striking, and it shows that the t-interval becomes very sensitive to nonnormality for critical values $t < 1$.

We next proceed to examine the case when $t \ge 1$. Notice that by Theorem 2, for general n, $\inf_{F \in \mathscr{F}_1} P_F(|T_n(\mathbf{Z})| \le t)$ is attained at $U[-1,1]$ if $t \ge n-1$. This result, however, is not of much practical use. For example, if $n = 6$, it does not cover even the 99% nominal interval. We propose a different line of attack in the following:

1. We can write $\inf_{F \in \mathscr{F}_1} P_F(|T_n(\mathbf{Z})| \le t) = \inf_G \int h(s_1, \dots, s_n)\, dG(s_1) \cdots dG(s_n)$, where $h(s_1, \dots, s_n) = P(|T_n(\mathbf{U})| \le t)$ with $\mathbf{U} = (U_1, \dots, U_n)$, the U_i's are mutually independent $U[-s_i, s_i]$ r.v.'s and G is any distribution on $[0, \infty)$ (Khintchine representation).
2. Suppose, we assume that Z_i's are *independent* and $Z_i \sim F_i \in \mathscr{F}_1$, $i = 1, \dots, n$ (instead of $Z_1, \dots, Z_n$ i.i.d. from $F \in \mathscr{F}_1$), thus embedding the "infimum problem" into a larger class (from now on, we will refer to it as *independent embedding*). In this setup, the infimum problem reduces to $\inf_{G_1, \dots, G_n} \int h(s_1, \dots, s_n)\, dG_1(s_1) \cdots dG_n(s_n)$.

Clearly, point 2 is not the same as point 1. However, they are the same if the infimum in point 2 is attained at $G_1 = \cdots = G_n$.

3. It follows that point 2 is equivalent to $\inf_{s_i \ge 0} P(|T_n(\mathbf{U})| \le t)$, U_i's independent $\sim U[-s_i, s_i]$.

The problem is thus reduced to evaluating $P(|T_n(\mathbf{U})| \le t)$ (as in 3) and a finite-dimensional minimization. These are obtained numerically and a numerical minimization is the only feasible method of attack, as far as we can see. The numerical results *indicate* that for each $n \le 10$ there exists a *threshold* t_n (dependent on n) such that for critical values $t \ge t_n$ the infimum is attained when $s_1 = \cdots = s_n = 1$. Thus, for $t \ge t_n$, our numerical results indicate that $\inf_{F \in \mathscr{F}_1} P_F(|T_n| \le t)$ equals $P_{U[-1,1]}(|T_n| \le t)$. This was not attempted for $n > 10$ because of the intensive nature of the computations. The threshold t_n's are shown in Table 1.

A comparison of nominal and numerically obtained minimal coverages for critical values above t_n is given in Table 2. Examining these values, we infer

TABLE 1
Threshold t_n

Sample size (n)	2	3	4	5	7	10
$t^* = n - 1$	1.0	2.0	3.0	4.0	6.0	9.0
t_n	1.00	1.73	1.91	1.92	2.00	2.25

that, for moderate sample sizes and high nominal coverages, the t-interval is conservative for symmetric unimodal distributions.

REMARK. For the family $\mathscr{F}_1^{ST} = \{F: F$ is symmetric and strongly unimodal about zero$\}$, we have $\inf_{F \in \mathscr{F}_1^{ST}} P_F(|T_n(\mathbf{Z})| \le t) \ge \inf_{F \in \mathscr{F}_1} P_F(|T_n(\mathbf{Z})| \le t)$. However, since $U[-1, 1] \in \mathscr{F}_1^{ST}$, our numerical results indicate that $\inf_{F \in \mathscr{F}_1^{ST}} P_F(|T_n| \le t) = P_{U[-1,1]}(|T_n| \le t)$ for $t \ge t_n$ and $2 \le n \le 10$.

The results described so far settle our problem whenever $t < 1$ or $t \ge t_n$. For $1 \le t < t_n$, our answers are not complete; but we do give useful bounds. Suppose we broaden our consideration to all "orthant symmetric" random vectors $\mathbf{Z}$. The random vector $\mathbf{Z} = (Z_1, \dots, Z_n)$ is said to be *orthant symmetric* (o.s.) if it has the same distribution as $\mathbf{Z}_\delta = (\delta_1 Z_1, \dots, \delta_n Z_n)$ for every choice of $\delta_i = \pm 1$, $i = 1, \dots, n$. If $\mathbf{Z}$ has an o.s. distribution, then $P(|S_n(\mathbf{Z})| \le s) = \int_{\Omega^+} P(|W_\xi| \le s) d\lambda(\xi)$, where $\Omega^+ = \{\xi = (\xi_1, \dots, \xi_n): \xi_i \ge 0, \ \sum_{i=1}^n \ \xi_i^2 = 1\}$, $W_\xi = \sum_{i=1}^n \xi_i \Delta_i$, where $\Delta_i = \pm 1$ with probability $\frac{1}{2}$, Δ_i's are independent and λ is a (suitable) measure on Ω^+ [see Efron (1969)]. It follows that

$$\inf_{(\text{o.s.})\mathbf{Z}} P(|S_n(\mathbf{Z})| \le s) = \inf_{\xi \varepsilon \Omega^+} P(|W_\xi| \le s).$$

Numerical minimization of the latter over $\xi \in \Omega^+$ leads us to the following conjecture.

CONJECTURE 1. *For all $k = 1, \dots, n - 1$, if $k \le s < k + 1$, then*

$$\inf_{\mathbf{Z}\ o.s.} P(S_n^2(\mathbf{Z}) \le s) = 1 - \frac{1}{2^k}.$$

We have checked this conjecture numerically up to $n = 10$. A complete analytic proof was not possible. Now, since i.i.d. samples from symmetric

TABLE 2
Minimum coverage of the t-interval in symmetric unimodal family

Sample size (n)	2	3	5	7	10
$P_\Phi(\lvert T_n\rvert \le t) = 0.80$	0.75	0.77			
$P_\Phi(\lvert T_n\rvert \le t) = 0.90$	0.86	0.87	0.89		
$P_\Phi(\lvert T_n\rvert \le t) = 0.95$	0.92	0.92	0.93	0.94	0.945
$P_\Phi(\lvert T_n\rvert \le t) = 0.99$	0.98	0.98	0.98	0.983	0.986

unimodal distributions constitute a subclass of the class of all orthant symmetric random vectors, the bound in the conjecture provides a lower bound for $\inf_{F \in \mathscr{F}_1} P_F(S_n^2(\mathbf{Z}) \leq s)$. We also obtain upper bounds by computing the probability for a fixed, suitably chosen, mixture of two symmetric uniforms. The reader is referred to Basu and DasGupta [(1992), Table 3], where numerical values of these bounds are reported. This table shows that the conjectured lower bounds are sometimes indeed useful. Also, the discontinuity at the point $s = 1$ comes out sharply; for $s < 1$ the infimum over $\mathscr{F}_1$ is zero (by Theorem 4), whereas for $s \geq 1$ the infimum over $\mathscr{F}_1$ is ≥ 0.5 (if Conjecture 1 is valid).

2.4. *Symmetric contaminations.* Instead of all symmetric unimodal distributions, in this section we look at the restricted class of *symmetric contaminations of normal* $\mathscr{F}_S^\varepsilon = \{F: F = (1-\varepsilon)N(0,1) + \varepsilon G, G \in \mathscr{C}_s\}$, where $\mathscr{C}_s = \{$all symmetric (about 0) distributions G on $\mathfrak{R}\}$.

We want to find $\inf_{F \in \mathscr{F}_s^\varepsilon} P_F(|T_n(\mathbf{X})| \leq t)$. Using *independent embedding* (see Section 2.3), we embed it into the problem of finding

$$\inf_{F_1, \dots, F_n \in \mathscr{F}_S^\varepsilon} P_{F_1, \dots, F_n}(|T_n(\mathbf{X})| \leq t),$$

where the X_i's are now independent and $X_i \sim F_i \in \mathscr{F}_S^\varepsilon$, $i = 1, \dots, n$.

THEOREM 5.

$$\inf_{F_1, \dots, F_n \in \mathscr{F}_S^\varepsilon} P(|T_n(\mathbf{X})| \leq t | X_i \sim F_i \textit{ and independent})$$
$$= \inf_{F_1^*, \dots, F_n^* \in \mathscr{G}_S^\varepsilon} P(|T_n(\mathbf{X})| \leq t | X_i \sim F_i^* \textit{ and independent}),$$

where

$$\mathscr{G}_S^\varepsilon = \{F^*: F^* = (1-\varepsilon)N(0,1) + \varepsilon G^* \textit{ and}$$
$$G^* \in \mathscr{C}_s^0 = \{G_v = 0.5\delta_{\{-v\}} + 0.5\delta_{\{v\}}, v \geq 0\}\}.$$

PROOF. Let $F_i = (1-\varepsilon)\Phi + \varepsilon G_i$, $i = 1, \dots, n$. It can be shown that $P_{F_1, \dots, F_n}(|T_n(\mathbf{X})| \leq t)$ can be written as the integral of a suitable Borel measurable function $h(z_1, \dots, z_n)$ with respect to the product measure $G_1 \otimes G_2 \otimes \cdots \otimes G_n$ [for details, see Basu (1991)]. Notice that (i) $\int h(z_1, \dots, z_n) \, dG_1(z_1) \cdots dG_n(z_n)$ is a linear function of each of $G_1, \dots, G_n$, and (ii) $\mathscr{C}_s$ is the closed (in weak topology) convex hull of $\mathscr{C}_s^0$ (follows from Theorem A.5 of Dharmadhikari and JoagDev (1988), page 255). It follows that

$$\inf_{G_1, \dots, G_n \in \mathscr{C}_s} \int h(z_1, \dots, z_n) \, dG_1(z_1) \cdots dG_n(z_n)$$
$$= \inf_{G_1^*, \dots, G_n^* \in \mathscr{C}_s^0} \int h(z_1, \dots, z_n) \, dG_1^*(z_1) \cdots dG_n^*(z_n),$$

which completes the proof. □

Theorem 5 reduces the problem to (n-dimensional) minimization of an integral, and this integral can be exactly evaluated [see Basu (1991)]. Numerical minimization for some cases shows that the infimum is attained when the v_i's are all identical. Combining Theorem 5 and the subsequent numerical work, we thus get (for selected n and t) $\inf_{F \in \mathscr{F}_S^\varepsilon} P_F(|T_n| \le t) = \inf_{F^* \in \mathscr{G}_S^\varepsilon} P_{F^*}(|T_n| \le t)$. These minimum coverages are listed in Basu and DasGupta [(1992), Table 4]. When nominal coverage is greater than or equal to 70%, (nominal − minimal) coverage is small, but a drastic difference appears when nominal coverage is less than or equal to 60%. Thus, the phenomenon we observed for the symmetric unimodal family is partially present for symmetric contaminations as well.

3. Arbitrary contaminations. In the previous section, we considered symmetric contaminations of normal. Here, we look at the general *ε-contamination class* $\mathscr{F}^\varepsilon = \{F = (1-\varepsilon)N(0,1) + \varepsilon G, G \in Q\}$, where $Q = \{$all distributions G on $\Re$ with $E_G(Z) = 0\}$. Such a class arises when we are $100(1-\varepsilon)\%$ certain that the underlying distribution is normal and $100\varepsilon\%$ uncertain about the form of the distribution. Note that $\mathscr{F}^\varepsilon$ is not a superclass of $\mathscr{F}_S^\varepsilon$, since here we are assuming finite first moment. As before, we use the technique of *independent embedding*.

THEOREM 6.

$$\inf_{F_1,\dots,F_n \in \mathscr{F}^\varepsilon} P(|T_n(\mathbf{X})| \le t | X_i \sim F_i \text{ and independent})$$
$$= \inf_{F_1^*,\dots,F_n^* \in \mathscr{G}^\varepsilon} P(|T_n(\mathbf{X})| \le t | X_i \sim F_i^* \text{ and independent}),$$

where

$$\mathscr{G}^\varepsilon = \{F^*: F^* = (1-\varepsilon)N(0,1) + \varepsilon G^*, G^* = p\delta_{\{v\}} + (1-p)\delta_{\{w\}}, \text{ with } pv + (1-p)w = 0\}.$$

PROOF. Following arguments similar to Theorem 5, $P(|T_n(\mathbf{X})| \le t | X_i \sim F_i \in \mathscr{F}^\varepsilon$ and independent) can be written as $\int h(z_1,\dots,z_n)\, dG_1(z_1) \cdots dG_n(z_n)$, where $F_i = (1-\varepsilon)\Phi + \varepsilon G_i$ with the G_i's satisfying $\int x\, dG_i(x) = 0\ \forall\ i$. The fact that it is enough to consider only two point G_i's for the minimization now follows from Mulholland and Rogers [(1958), Theorems 2 and 8]. □

By Theorem 6, the infimum problem in the independent (not i.i.d.) case is reduced again to ($2n$-dimensional) minimization of an integral (this integral can be exactly calculated). Numerical minimization shows that the minimum, again, is attained when the F_i^*'s are identical, thus giving us the required infimum for the i.i.d. case. Table 5 in Basu and DasGupta (1992), where these minima are listed, shows that the t-interval is conservative for small n, but as n increases the minimum coverages differ more from the nominal ones.

4. Robustness of *P*-values: tyranny of large samples. In our setup, if we desire to test the null hypothesis $H_0: \theta = \theta_0$ against the alternative hypothesis $H_1: \theta \neq \theta_0$, the *p-value* (or observed significance level) for the observed data $\mathbf{x} = (x_1, \dots, x_n)$ is $p_v = P_F^{\theta=\theta_0}(|T_n(\mathbf{X})| > t) = P_F(|T_n(\mathbf{Z})| > t)$, where $t = T_n(\mathbf{x})$ is the observed value of the *t-statistic*. Thus, our findings on the infimum coverage of the interval carry over directly to the problem of finding the supremum of the p-value over the respective family of distributions. For example, if we only know that the unknown distribution F belongs to a family $\mathscr{F}$ satisfying the conditions of Theorem 4, and if we observe $|T_n(\mathbf{z})| < 1$, $n \geq 2$, then Theorem 4 says that in the worst case we have $p_v = 1$, that is, we are always going to accept H_0. Reinterpretation of other results, in terms of robustness of p-values, follows similarly.

The preceding section suggests that, as the sample size (n) increases, the maximum p-value over the class $\mathscr{F}^\varepsilon$ differs more from the nominal p-value. However, the numerically intensive method cannot be carried out for higher n's. Instead, suppose, for large n, we consider $F_* = (1-\varepsilon)N(0,1) + \varepsilon G$, where ε is small, and $G = p\delta_{\{-a\}} + (1-p)\delta_{\{b\}}$, with a small and b large; $E_G(X) = 0$ implies $p = b/(a+b)$, making p close to 1. For n large and $X_1, \dots, X_n$ i.i.d. from F, it can then be shown that [see Basu (1991)]

$$\begin{aligned}
P_F&\big(|T_n(\mathbf{X})| \leq t\big)\\
&\leq \sum_{k=n/2}^{n} \binom{n}{k} (1-\varepsilon)^k \varepsilon^{n-k} p^{n-k}\\
&\quad \times P\big[|T_n(\mathbf{X})| \leq t \,|\, X_1, \dots, X_k \sim N(0,1), X_{k+1}, \dots, X_n \sim \delta_{\{-a\}}\big]\\
&\quad + \sum_{k=0}^{n/2} \binom{n}{k} (1-\varepsilon)^k \varepsilon^{n-k} + \sum_{k=n/2+1}^{n} \binom{n}{k} (1-\varepsilon)^k \varepsilon^{n-k} (1-p^{n-k}).
\end{aligned}$$

Lower bounds on the p-value of F_* computed according to above are listed in Table 3. Consider the case when the nominal p-value is 0.10. If we allow 10% contamination ($\varepsilon = 0.1$), the p-value of F_* is greater than or equal to 0.21 (0.35 and 0.54) for $n = 20$ ($n = 50$ and $n = 100$). Notice that $F_* \in \mathscr{F}^\varepsilon$. Hence, the supremum p-value over $\mathscr{F}^\varepsilon$ is also greater than or equal to 0.21 (0.35 and 0.54) for $n = 20$ ($n = 50$ and $n = 100$). Thus, over the contamination class, the p-value of the t-statistic is not very sensitive for small n, but the lack of robustness becomes more pronounced as the sample size n increases.

TABLE 3
Lower bounds on the p-value ($\varepsilon = 0.1$)

	Nominal = 0.30			**Nominal = 0.10**			**Nominal = 0.05**			**Nominal = 0.01**		
n	20	50	100	20	50	100	20	50	100	20	50	100
	0.45	0.60	0.76	0.21	0.35	0.54	0.14	0.24	0.42	0.04	0.09	0.20

5. Comments. In this article, we establish some explicit and surprising results about the robustness of the widely used t confidence interval under departure from normality. When the population variance σ^2 is known, a frequentist optimal confidence interval, under normality, for the unknown mean θ is the *z-interval* $\bar{X} \pm z_{\alpha/2}\sigma/\sqrt{n}$. Without loss of generality, we can assume $\sigma^2 = 1$. Thus, $X_i = \theta + Z_i$, Z_i's i.i.d. from F with $E_F(Z_i) = 0$, $E_F(Z_i^2) = 1$. To examine robustness against nonnormality, one can take $F \in \mathscr{F}$, a suitably chosen family, and consider $\inf_{F \in \mathscr{F}} P_F\ (|\bar{Z}| \leq c)$ as a sensitivity criterion (here $c = z_{\alpha/2}/\sqrt{n}$). The detailed technical results will appear separately. It turns out that, consistent with common belief based on simulation studies, the theorems show that the z-interval is more robust in comparison under skewness and very robust for broad general symmetric unimodal populations.

Acknowledgments. We have benefitted from valuable discussions with Peter Bickel, Arup Bose, Larry Brown, Morris Eaton, Jayanta Ghosh, Jiunn Hwang, Y. L. Tong and Jim Zidek. The Editor and a referee made numerous constructive suggestions. They all have our sincerest appreciation.

REFERENCES

BASU, S. (1991). Robustness of Bayesian and classical inference under distribution bands and shape restricted families. Ph.D. dissertation, Purdue Univ.

BASU, S. and DASGUPTA, A. (1992). Robustness of standard confidence intervals for location parameters under departure from normality. Technical Report 210, Dept. Statistics and Applied Probability, Univ. California, Santa Barbara.

BENJAMINI, Y. (1983). Is the t-test really conservative when the parent distribution is long-tailed? *J. Amer. Statist. Assoc.* **78** 645–654.

BICKEL, P. J. and KRIEGER, A. M. (1989). Extensions of Chebychev's inequality with applications. Unpublished manuscript.

CRESSIE, N. (1980). Relaxing assumptions in the one sample t-test. *Austral. J. Statist.* **22** 143–153.

DHARMADHIKARI, S. and JOAG-DEV, K. (1988). *Unimodality, Convexity, and Applications*. Academic Press, New York.

EFRON, B. (1969). Student's t-test under symmetry conditions. *J. Amer. Statist. Assoc.* **64** 1278–1302.

HATCH, L. O. and POSTEN, H. O. (1966). Robustness of the Student-procedure: a survey. Research Report 24, Dept. Statistics, Univ. Connecticut, Storrs.

HERRENDORFER, G., RASCH, D. and FEIGE, K. D. (1983). Robustness of statistical methods II. Methods for the one-sample problem. *Biometrical J.* **25** 327–343.

LOGAN, B., MALLOWS, C., RICE, S. and SHEPP, L. (1973). Limit distributions of self-normalized sums. *Ann. Probab.* **1** 788–807.

MULHOLLAND, H. P. and ROGERS, C. A. (1958). Representation theorems for distribution functions. *Proc. London Math. Soc.* **8** 177–223.

ROACH, S. A. (1963). The frequency distribution of the sample mean where each member of the sample is drawn from a different rectangular distribution. *Biometrika* **50** 508–513.

UNIVERSITY OF ARKANSAS
301 SCIENCE-ENGINEERING BLDG.
FAYETTEVILLE, ARKANSAS 72701

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907