Ann. Inst. Statist. Math.
Vol. 51, No. 2, 217–230 (1999)

CONSERVATISM OF THE z CONFIDENCE INTERVAL UNDER SYMMETRIC AND ASYMMETRIC DEPARTURES FROM NORMALITY

SANJIB BASU*

Department of Mathematical Sciences, Fulbright College of Arts and Sciences, University of Arkansas, 301 Science-Engineering Building, Fayetteville, AR 72701, U.S.A.

(Received April 22, 1996; revised September 11, 1997)

Abstract. The robustness of the z-interval under nonnormality is investigated by finding its infimum coverage probability over suitably chosen broad class of distributions. In the case of $n = 1$, the infimum coverage probabilities over the normal scale mixture and the symmetric unimodal families of distributions are obtained analytically. For general n, the infimum problem is theoretically reduced to a finite dimensional minimization which is then obtained numerically. The obtained minimum coverage probabilities are very close to the nominal probabilities. These exact minimum coverage probabilities are often notably sharper than the lower bounds given by the Camp-Meidell-Gauss inequality. The family of general unimodal distributions is considered next to investigate the possible effect of asymmetry. The obtained infimum coverage probabilities over this family are found to be the same as the ones over the symmetric unimodal class.

Key words and phrases: Camp-Meidell inequality, coverage probability, normal scale mixture, robustness, unimodal, z-interval.

1. Introduction

Consider a standard location-scale set up: $Y_i = \theta + \sigma Z_i$, $i = 1, \dots, n$ where $\{Z_1, \dots, Z_n\}$ 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 the unknown location parameter θ is given by the *z-interval*: $\bar{Y} \pm z_{\alpha/2}\sigma/\sqrt{n}$, where $\bar{Y}$ is the sample mean 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*: $\bar{Y} \pm t_{\alpha/2}s/\sqrt{n}$, where s is the sample standard deviation and $t_{\alpha/2}$ is the $(1-\alpha/2)$ percentile of the "Student's t" distribution with $(n-1)$ degrees of freedom.

Suppose now 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

---

* Now at Division of Statistics, Northern Illinois University, DeKalb, IL 60115, U.S.A.

use the above mentioned confidence intervals (as practitioners often do). Let $\rho_z(F)$ and $\rho_t(F)$ denote respectively the coverage of the 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 *et al.* (1973) showed that the limiting coverage probability $\rho_t(F)$ of the t-interval can be quite strange.

Basu and DasGupta (1995) recently investigated the small sample properties and robustness of the t-interval under nonnormality. Our main focus in this article will be directed towards the known variance case, i.e., towards investigating the small sample behavior and sensitivity of the z-interval. Without loss of generality, we will assume that the known $\sigma^2 = 1$. We will model departure from normality by requiring the parent distribution F to belong to a suitably chosen class of distributions $\mathcal{F}$. The sensitivity of the z-interval to nonnormality will be examined by considering the worst coverage of the z-interval over the class $\mathcal{F}$, i.e., $\underline{\rho} = \inf_{F \in \mathcal{F}} \rho(F)$.

It is generally believed that the z-interval is conservative for samples from a heavy-tailed (heavier tailed than normal) symmetric distribution, though the foundation behind this belief is mostly experimental results and sketchy theoretical arguments. A simple and common way to obtain heavy-tailed distributions is through mixtures. In Section 2, we examine the robustness of the z-interval over the class of all normal scale mixture distributions with a fixed variance,

$$\mathcal{F}_N = \left\{ F : F(z) = \int_0^\infty \Phi(z/\sqrt{s})dG(s) \text{ and } \mathrm{Var}_F(Z) = 1 \right\}. \tag{1.1}$$

Here $\Phi(\cdot)$ is the cumulative distribution function (cdf) of $N(0,1)$ and G is an arbitrary mixing distribution with support on $[0,\infty)$. Normal scale mixture distributions provide a broad class of continuous, symmetric, unimodal distributions on the real line which are heavier tailed than normal, but otherwise have a wide variety of tail structures. A large body of literature exists on scale mixture of normal distributions including Andrews and Mallows (1974), Efron and Olshen (1978), Dey (1990), Dey and Birmiwal (1996), Basu and DasGupta (1995) and Basu (1996). We use geometrical arguments and techniques from moment theory to find the worst coverage probability $\underline{\rho}$ of the z-interval over the class $\mathcal{F}_N$. Our findings suggest that the z-interval is, indeed, quite robust over the class of normal scale mixtures.

In Section 3, we broaden our examination to the class of symmetric unimodal distributions

$$\mathcal{F}_{SU} = \{F : F \text{ is symmetric and unimodal about zero and } \mathrm{Var}_F(Z) = 1\}. \tag{1.2}$$

The class of symmetric unimodal distributions includes heavy tailed ($\mathcal{F}_N$ is a subclass of $\mathcal{F}_{SU}$) as well as distributions which are shorter-tailed than normal (for example, symmetric uniforms and triangular distributions). Moment theory techniques and the Khintchine representation of expressing symmetric unimodal distributions as mixtures of symmetric uniforms are used to find $\underline{\rho}$ over the class $\mathcal{F}_{SU}$. Our results indicate that, even in this broader class, the infimum coverage of the z-interval is very close to the nominal coverage, thus implying conservatism.

The effect of asymmetry of the parent distribution F on the coverage probability of the z-interval is studied in Section 4. Here we remove the symmetry assumption and consider the class of symmetric and asymmetric unimodal distributions

$$\mathcal{F}_A = \{F : F \text{ is unimodal (not necessarily at zero) and } E_F(Z) = 0,\ \mathrm{Var}_F(Z) = 1\}. \tag{1.3}$$

Most surprisingly, the obtained results indicate that the infimum coverage $\underline{\rho}$ over the class $\mathcal{F}_A$ of symmetric and asymmetric distributions is typically attained at a symmetric F. This implies that the infimum coverage over the broader class $\mathcal{F}_A$ is typically the same as $\inf \rho$ over $\mathcal{F}_{SU}$.

The question of sensitivity of the z and t intervals to nonnormality have an extensive literature, see Basu and DasGupta (1995), Basu (1991), Lehmann and Loh (1990) the review article by Herrendörfer *et al.* (1983) and the references therein. The principal achievement of this article is that we formally establish the conservatism of the z-interval under symmetric and asymmetric departures from normality. In contrast, Basu and DasGupta (1995) show that the t-interval can be extremely sensitive to departures from normality.

In the context of test of hypothesis, a different type of robustness investigation can be formulated in terms of invariance. See Sinha and Kariya (1989), Das and Sinha (1990) for exposition to this idea.

2. Normal scale mixture distributions

In this section, we investigate the robustness of the z-interval when the parent distribution F comes from the normal scale mixture family $\mathcal{F}_N$ described in (1.1). This family contains a broad class of symmetric, unimodal distributions with different tail behaviors including the t distributions with different degrees of freedom, Double Exponential distribution, Logistic distribution and many others. We examine the sensitivity of the z-interval by finding its infimum coverage over the class $\mathcal{F}_N$. Thus our sensitivity criterion is $\inf_{F\in\mathcal{F}_N} P_F(|\bar{Z}| \le c)$ where $c = z_{\alpha/2}/\sqrt{n}$. Our first result here consider the simplest case when we have a sample of size $n = 1$.

THEOREM 1. *For $n = 1$, $\inf_{F\in\mathcal{F}_N} P_F(|Z_1| \le c) = P_{F_0}(|Z_1| \le c)$ where*

$$F_0 = \begin{cases} N(0,1) & \text{if } c \le 1.1906 \\ (1 - 1/s_0)\delta_{\{0\}} + 1/s_0 N(0, s_0) & \text{if } c > 1.1906 \\ & \quad \text{with } s_0 = (c/1.1906)^2. \end{cases}$$

PROOF. The proof is based on the geometric approach to moment problems outlined in Kemperman (1968). For $F \in \mathcal{F}_N$, $P_F(|Z_1| \le c) = \int_0^\infty (2\Phi(c/\sqrt{s}) - 1)dG(s)$. Also, $\text{Var}_F(Z_1) = 1 \Leftrightarrow \int_0^\infty s dG(s) = 1$. The problem thus reduces to finding $\inf_G \int_0^\infty \Phi(c/\sqrt{s})dG(s)$, subject to $\int_0^\infty s dG(s) = 1$. It is easily seen that $\Phi(c/\sqrt{s})$, for fixed c and $s \ge 0$, is concave in s up to a certain point, and then decreases convexly. Now, suppose $\exists$ a straight line $bs+1$ such that $\Phi(c/\sqrt{s}) \ge bs+1$ $\forall s \ge 0$ with equality attaining at $s = 0$ and at $s = s_0 \ge 1$. We then have $\int_0^\infty \Phi(c/\sqrt{s})dG(s) \ge \int_0^\infty (bs+1)dG(s) = b+1$ $\forall G$ satisfying the given conditions, whereas for $G_0 = (1-1/s_0)\delta_{\{0\}} + 1/s_0\delta_{\{s_0\}}$, $\int_0^\infty \Phi(c/\sqrt{s})dG_0(s) = b+1$, thus giving G_0 to be the minimal distribution. If $bs+1$ is moreover tangent to $\Phi(c/\sqrt{s})$ at s_0, we can solve for b and s_0 from $\Phi(c/\sqrt{s_0}) = bs_0 + 1$ and $\frac{d}{ds}\Phi(c/\sqrt{s})|_{s=s_0} = b$. Eliminating b, and putting $c/\sqrt{s_0} = t$, we get $\Phi(t) = 1 - t\phi(t)/2$ where $\phi(t) = \frac{d}{dt}\Phi(t)$ is the $N(0,1)$ density function. A numerical solution to the above equation yields $t = c/\sqrt{s_0} = 1.1906$. Thus, for $c \ge 1.1906$, there does exist such a straight line $bs+1$ satisfying all requirements. For $c < 1.1906$, the tangent to $\Phi(c/\sqrt{s})$ at $s = 1$ always lies below the curve. Thus, $G_0 = \delta_{\{1\}}$ attains the minimum in this case (through the same argument). This completes the proof. □

For general $n > 1$, we have $Z_1, \ldots, Z_n$ i.i.d. from a distribution $F \in \mathcal{F}_N$. Using the mixture representation of F, we can write the coverage probability $P_F(|\bar{Z}| \le c)$ as $-1 + 2\int \Phi(c/\sqrt{s_1 + \cdots + s_n})dG(s_1)\cdots dG(s_n)$. The problem of finding the infimum of the coverage probability over the class $\mathcal{F}_N$ thus reduces to finding $\inf_G \int \Phi(c/\sqrt{s_1 + \cdots + s_n})dG(s_1)\cdots dG(s_n)$ subject to the condition $\int_0^\infty s dG(s) = 1$. This is a very hard optimization problem, mainly because of the fact that in the minimization process, we need to consider all n coordinates simultaneously. Hoeffding and Srikhande (1955) consider a similar problem in a simpler set-up. They obtain exact results only for $n = 2$ and bounds for larger n's.

We propose a novel line of attack to tackle this hard optimization problem. Suppose, instead of assuming $Z_1, \ldots, Z_n$ are i.i.d. from a distribution $F \in \mathcal{F}_N$ we assume that $Z_1, \ldots, Z_n$ are independent with $Z_i \sim F_i \in \mathcal{F}_N$, $i = 1, \ldots, n$. In this setup, the infimum problem reduces to $\inf_{F_1,\ldots,F_n \in \mathcal{F}_N} P(|\bar{Z}| \le c \mid Z_i \sim F_i$ and independent). Thus, we have embedded the original infimum problem over an "i.i.d. class" to a new infimum problem over the larger "independent class". Notice, however that the "i.i.d. class" is a subclass of the "independent class". For latter reference, we will refer to this embedding as the "*independent embedding*". Basu and DasGupta (1995) also use this *independent embedding* in optimization problems related to the t-interval.

The infimum of a functional over the "independent class", in general, may not be the same as the infimum over the "i.i.d. class". However, these two infima are same if the infimum over the larger "independent class" is attained in the "i.i.d. subclass". This will be the case in many of the problems we consider.

Our next result shows that the infimum over the "independent class" can be reduced significantly. In fact, the infimum can be obtained by considering only those $F_i \in \mathcal{F}_N$ which are mixtures of a point mass at 0 and a $N(0,s)$ distribution. Notice that this is a direct generalization of Theorem 1 to the $n > 1$ case.

THEOREM 2. *Let* $F_s = (1 - 1/s)\delta_{\{0\}} + 1/sN(0, s)$ *be a mixture of a point mass at* 0 *and a* $N(0, s)$ *distribution. Notice that* $\mathrm{Var}_{F_s}(Z) = 1$ *and hence* $F_s \in \mathcal{F}_N$ *for any* $s > 1$. *We have* $\inf_{F_1,\ldots,F_n \in \mathcal{F}_N} P(|\bar{Z}| \leq c \mid Z_i \sim F_i$ *and independent*$) = \inf_{s_1,\ldots,s_n \geq 1} P(|\bar{Z}| \leq c \mid Z_i \sim F_{s_i}$ *and independent*$)$.

PROOF. Note that if $F_i \in \mathcal{F}_N$, $i = 1, \ldots, n$, then $P(|\bar{Z}| \leq c \mid Z_i \sim F_i$ and independent$) = 2P(Z_1 + \cdots + Z_n \leq nc \mid Z_i \sim F_i$ and independent$) - 1$ since the distribution of $Z_1 + \cdots + Z_n$ is then symmetric about 0. Let $c_0 = nc$.

We will prove the result sequentially, i.e., we will sequentially show that each coordinate F_i can be reduced $F_{s_i} = (1-1/s_i)\delta_{\{0\}} + 1/s_i N(0, s_i)$, $i = 1, \ldots, n$. The argument for each coordinate has similarities with the proof for the $n = 1$ case of Theorem 1.

Fix a coordinate i between 1 and n. Let $F_j \in \mathcal{F}_N$ denote the distribution of Z_j, $j = 1, \ldots, i-1, i+1, \ldots, n$. Let F_{-i} denote the distribution of $Z_1 + \cdots + Z_{i-1} + Z_{i+1} + \cdots + Z_n$. We will assume that F_j, $j = 1, \ldots, i-1, i+1, \ldots, n$ and hence F_{-i} are fixed and we will minimize over the distribution of the i-th coordinate F_i. Notice that each F_j is symmetric unimodal about 0. The distribution F_{-i}, which is a convolution of the F_j's, $j = 1, \ldots, i-1, i+1, \ldots, n$, is then also symmetric and unimodal about 0 (see Dharmadhikari and Joag-dev (1988), p. 13). This further implies that F_{-i} is absolutely continuous except possibly for a point mass at 0. Let f_{-i} be the Radon-Nikodyn derivative of the absolutely continuous component of F_{-i} (w.r.t. Lebesgue measure). Then f_{-i} is symmetric about 0 and is nonincreasing on $(0, \infty)$. We will use this fact later.

We next consider the fixed coordinate i on which we are minimizing. Since $F_i \in \mathcal{F}_N$, we have $F_i(x) = \int \Phi(x/\sqrt{s})dG_i(s)$ where the mixing distribution G_i has support on $[0, \infty)$ with $\int s dG_i(s) = 1$. For our subsequent calculations, here we use a different parametrization of the normal scale mixture distribution F_i. Let $\tau_i = 1/\sqrt{s_i}$ and $H(\tau_i) = 2G(1/\tau_i^2)/\tau_i^3$. We can then write $F_i(x) = \int \Phi(x\tau_i)dH(\tau_i)$ where the mixing distribution H_i has support on $[0, \infty]$ with $\int (1/\tau^2)dH_i(\tau) = 1$. Notice that $s_i = 0$ results in $\tau_i = \infty$. For this reason, we will use the extended nonnegative real line $\bar{\Re}^+ = [0, \infty]$ as the support of H_i. We then have

$$(2.1) \qquad \begin{aligned} &P(Z_1 + \cdots + Z_n \leq c_0 \mid Z_j \sim F_j \text{ and independent}) \\ &\quad = \int F_i(c_0 - x)dF_{-i}(x) = \iint \Phi((c_0 - x)\tau)dF_{-i}(x)dH_i(\tau) \end{aligned}$$

where the last step holds since interchanging the order of integrals is valid here. Let $g(c_0, \tau) = \int \Phi((c_0 - x)\tau)dF_{-i}(x)$.

Claim. For every fixed c_0, the function $g(c_0, \tau)$ is nondecreasing and concave in τ.

PROOF OF CLAIM. We have $\frac{d}{d\tau}g(c_0, \tau) = \int \frac{d}{d\tau}\Phi((c_0 - x)\tau)dF_{-i}(x)$ (since the conditions for differentiation under the integral sign are met) $= \int (c_0 - x)\phi((c_0 - x)\tau)dF_{-i}(x) = \int_{(0,\infty)} y\phi(y\tau)\{dF_{-i}(c_0 - y) - dF_{-i}(c_0 + y)\} = \int_0^\infty y\phi(y\tau)\{f_{-i}(c_0 - y) - f_{-i}(c_0 + y)\}dy + c_0\phi(c_0\tau)\{F_{-i}(0) - F_{-i}(0-)\}$ where we make the change of variable $y = c_0 - x$ and then note that F_{-i} is absolutely continuous except possibly at 0. Since $c_0 > 0$ and the function $f_{-i}(\cdot)$ is symmetric and unimodal about 0, we have $f_{-i}(c_0 - y) \geq f_{-i}(c_0 + y)$ for every $y > 0$. Thus $\frac{d}{d\tau} g(c_0, \tau) \geq 0$ or $g(c_0, \tau)$ is nondecreasing in $\tau \geq 0$.

By a similar argument, we have $\frac{d^2}{d\tau^2} g(c_0,\tau) = \int \frac{d^2}{d\tau^2}\Phi((c_0 - x)\tau)dF_{-i}(x) = -\int \tau(c_0-x)^3\phi((c_0-x)\tau)dF_{-i}(x) = -\int_{(0,\infty)} \tau y^3\phi(y\tau)\{dF_{-i}(c_0-y) - dF_{-i}(c_0+y)\} = -\int_0^\infty \tau y^3 \phi(y\tau)\{f_{-i}(c_0-y) - f_{-i}(c_0+y)\}dy - \tau c_0^3\phi(c_0\tau)\{F_{-i}(0) - F_{-i}(0-)\} \leq 0$. This shows that $g(c_0, \tau)$ is a concave function of τ and proves the claim.

Thus, $g(c_0, \tau)$ is a nondecreasing concave function with $g(c_0, 0) = 1/2$ and $g(c_0, \infty) = F_{-i}(c_0) > 1/2$.

Consider the function $h(\overline{b_0}, b_1, \tau) = \overline{b_0} - b_1/\tau^2$ where $\overline{b_0} = g(c_0, \infty)$ and $b_1 \geq 0$. h is clearly a nondecreasing concave function. Also, $h(\overline{b_0}, b_1, \tau) = g(c_0, \tau)$ at $\tau = \infty$. As $b_1 \uparrow \infty$, we have $h(\overline{b_0}, b_1, \tau) \leq g(c_0, \tau)$ for all $\tau \in \Re^+$. Moreover, informally, as $b_1 \downarrow 0$, we will have $h(\overline{b_0}, b_1, \tau) \geq g(c_0, \tau)$ for all τ. From this geometric picture, it is easy to see that $\exists$ a value of $b_1^* > 0$ such that $h(\overline{b_0}, b_1^*, \tau) \leq g(c_0, \tau)$ for all $\tau \in \bar{\Re}^+$ with $h(\overline{b_0}, b_1^*, \tau) = g(c_0, \tau)$ at $\tau = \infty$ and at least at another value of τ. Let τ_0 be the smallest of the τ values at which the two functions are equal.

For our specific c_0 value, if we find $\tau_0 \leq 1$, then we are done. Define $H_i^{(0)} = \tau_0^2\delta_{\{\tau_0\}} + (1-\tau_0^2)\delta_{\{\infty\}}$. We have $\int (1/\tau^2)dH_i^{(0)}(\tau) = 1$. Also, $\int g(c_0, \tau)dH_i^{(0)}(\tau) = \int h(\overline{b_0}, b_1^*, \tau)dH_i^{(0)}(\tau)$ (since the two functions are equal on the support of $H_i^{(0)}$) $= \overline{b_0} - b_1^*$ (since $\int(1/\tau^2)dH_i^{(0)}(\tau) = 1$). For any other mixing distribution H with $\int(1/\tau^2)dH(\tau) = 1$, we have $\int g(c_0,\tau)dH(\tau) \geq \int h(\overline{b_0}, b_1^*, \tau)dH(\tau) = \overline{b_0} - b_1^*$. This establishes $H_i^{(0)}(\tau)$ as the minimal mixing distribution in this case.

Now suppose we are in the case where, for our specific c_0 value, we get $\tau_0 > 1$. Select a value of b_0 between $\overline{b_0} = g(c_0, \infty)$ and $\underline{b}_0 = g(c_0, 0)$. As before, consider the function $h(b_0, b_1, \tau) = b_0 - b_1/\tau^2$. As $b_1 \uparrow \infty$, we have $h(b_0, b_1, \tau) \leq g(c_0, \tau)$ for all $\tau \geq 0$. Also, informally, as $b_1 \downarrow 0$, we have $h(b_0, b_1, \tau) \geq g(c_0, \tau)$ for all $\tau \in \{\tau : g(c_0, \tau) < b_0\}$. Arguing as before, $\exists$ a value of $b_1^*(b_0) > 0$ such that $h(b_0, b_1^*(b_0), \tau) \leq g(c_0, \tau)$ for all $\tau \in \bar{\Re}^+$ with $h(b_0, b_1^*(b_0), \tau) = g(c_0, \tau)$ for at least one $\tau < \infty$. Let $\tau^*(b_0)$ be the smallest of these τ values at which the two functions are equal.

For $b_0 = \overline{b_0} = g(c_0, \infty)$, we have $\tau^*(\overline{b_0}) = \tau_0 > 1$. For $b_0 < g(c_0, 1)$, surely $\tau^*(b_0) < 1$. Since $g(c_0, \tau)$ is a smooth, infinitely differentiable function, it can be argued that $\tau^*(b_0)$ is a continuous function of b_0, and hence there must exist b_0^* for which $\tau^*(b_0^*) = 1$. For this b_0^*, we have $h(b_0^*, b_1^*(b_0^*), \tau) \leq g(c_0, \tau)$ for all $\tau \in \bar{\Re}^+$ with $h(b_0^*, b_1^*(b_0^*), 1) = g(c_0, 1)$. Following our earlier argument, it can be easily seen that $H_i(\tau) = \delta_{\{1\}}$ is the minimal mixing distribution in this case.

We now combine the two cases ($\tau_0 \leq 1$ and $\tau_0 > 1$) and return back to our earlier parametrization of $s_i = 1/\tau_i^2$. We find that the minimal mixing distribution in terms of s_i is $G_i^0(s_i) = (1 - 1/s_0)\delta_{\{0\}}(s_i) + 1/s_0\delta_{\{s_0\}}(s_i)$ where $s_0 \geq 1$.

We have thus proved that for a fixed $1 \leq i \leq n$, $\inf_{F_i \in \mathcal{F}_N} P(|\bar{Z}| \leq c \mid Z_j \sim F_j, j \neq i, Z_i \sim F_i$ and independent$) = \inf_{F_{s_i} \in \mathcal{F}_N} P(|\bar{Z}| \leq c \mid Z_j \sim F_j, j \neq i, Z_i \sim F_{s_i}$ and independent$)$ where $F_{s_i} = (1 - 1/s_i)\delta_{\{0\}} + 1/s_i N(0, s_i)$. The proof of Theorem 2 is completed by applying this result sequentially to each coordinate. □

Theorem 2 reduces the problem of finding $\inf_{F_1,\ldots,F_n \in \mathcal{F}_N} P(|\bar{Z}| \leq c \mid Z_i \sim$

Table 1. Minimum coverage of the z-interval over the normal scale mixture family $\mathcal{F}_N$ and the symmetric unimodal family $\mathcal{F}_{SU}$ and the lower bound on the coverage probability from the Camp-Meidell-Gauss (C-M-G) inequality.

n	$z_{\alpha/2}$	Nominal coverage	C-M-G lower bound	Normal scale mixtures	Symmetric unimodal
1	1.04	0.70	0.586	0.70	0.598
2				0.70	0.667
8				0.70	0.694
1	1.28	0.80	0.729	0.798	0.729
2				0.80	0.766
8				0.80	0.784
1	1.65	0.90	0.836	0.878	0.836
2				0.882	0.849
3				0.883	0.853
5				0.884	0.856
8				0.885	0.857
1	1.96	0.95	0.884	0.914	0.884
2				0.916	0.891
3				0.917	0.893
5				0.917	0.894
8				0.918	0.895
1	2.58	0.99	0.933	0.950	0.933
2				0.951	0.935
3				0.951	0.936
5				0.951	0.937
8				0.951	0.937

F_i and independent) to only an n dimensional minimization over $s_1, \ldots, s_n \geq 1$. Notice that this is a significant reduction of dimensions in the minimization process. In the next section, we present a powerful moment theory result due to Mulholland and Rogers (1958) (to be referred as MR result from now on). A direct application of the MR result would have given $F_i = (s_{i,2}-1)/(s_{i,2}-s_{i,1})N(0, s_{i1}) + (1-s_{i,1})/(s_{i,2}-s_{i,1})N(0, s_{i2})$ as the form of the minimal distributions. This would have resulted in a $2n$ dimensional minimization over $0 \leq s_{i1} \leq 1 \leq s_{i2} < \infty$, $i = 1, \ldots, n$. The analytical argument of Theorem 2 makes a further substantial reduction to only an n dimensional minimization and establishes $s_{i1} = 0$, $i = 1, \ldots, n$. □

The final n dimensional minimization to obtain $\min_{s_1,\ldots,s_n \geq 1} P(|\bar{Z}| \leq c \mid Z_i \sim F_{s_i}$ and independent) is carried out numerically. The results are surprising. In all the attempted cases, the minimum is attained when all the s_i's are equal, i.e., $s_1 = \cdots = s_n = s_*$ (say). Since the minimum over the "independent class" is actually obtained at $F_{s_1} = \cdots = F_{s_n} = F_{s_*}$, i.e., in the "i.i.d. class", the obtained minimum in the attempted cases equals $\min_{F\in\mathcal{F}_N} P(|\bar{Z}| \le c \mid Z_i \sim F$ and i.i.d.).

The obtained values of the minimum coverage probability are shown in Table 1. If the sample size $n \ge 2$ and we consider intervals which have nominal coverage $\le$80%, then the value of s_* where the minimum is attained is, in fact, 1. This implies that $N(0,1)$ yields the smallest coverage probability over the normal scale mixture class $\mathcal{F}_N$. For intervals with higher nominal coverages, the minimum is no longer attained at $N(0,1)$, but the values in Table 1 show that the minimum coverages are close to the nominal ones. We are thus assured of a high degree of conservatism of the z-interval over the normal scale mixture family. We did not go beyond $n = 8$ because of the high dimensions of the numerical minimization involved. We note in passing that Benjamini (1983) studies the conservatism of the t-interval over the normal scale mixture family.

3. Symmetric unimodal distributions

The normal scale mixture family only allows distributions which are heavier tailed than normal. The larger family of symmetric unimodal distributions includes heavy tailed distributions as well as distributions which are shorter tailed than normal. The latter family thus allows us to do our robustness investigation in a broader set up. Unimodal distributions arise naturally in statistics. They have an extensive literature, see Basu and DasGupta (1996) and the book by Dharmadhikari and Joag-dev (1988). Our objective in this section will be to examine the robustness of the z-interval when the parent distribution F comes from the family $\mathcal{F}_{SU}$ as described in (1.2). Specifically, we consider the worst coverage of the z-interval, $\inf_{F\in\mathcal{F}_{SU}} P_F(|\bar{Z}| \le c)$, as our robustness criterion (here $c = z_{\alpha/2}/\sqrt{n}$).

The Khintchine mixture representation (see, for example, Dharmadhikari and Joag-dev (1988), p. 10) states that any symmetric unimodal cdf F can be written as $F(z) = \int U_s(z)dG(s)$ where $U_s(z)$ is the cdf of the Uniform$[-s,s]$ distributions and G is a distribution on $[0,\infty)$. We will use $U[-s,s]$ to denote a Uniform$[-s,s]$ distribution and U_s to denote a random variable having $U[-s,s]$ distribution.

As in Section 2, we first consider the case of $n = 1$

THEOREM 3. *For $n = 1$, $\inf_{F\in\mathcal{F}_{SU}} P_F(|Z_1| \le c) = P_{F_*}(|Z_1| \le c)$ where*

$$F_* = \begin{cases} U[-\sqrt{3},\sqrt{3}] & \text{if } c \le 2/\sqrt{3} \\ (1-3/s_*)\delta_{\{0\}} + 3/s_* U[-\sqrt{s_*},\sqrt{s_*}] & \text{if } c > 2/\sqrt{3} \\ & \text{where } s_* = 9c^2/4. \end{cases}$$

PROOF. The argument is very similar to the proof of Theorem 1. By an equivalent Khintchine representation, we have $\{F$ is a symmetric unimodal (at 0) distribution on $\Re\} = \{F(z) = \int U_{\sqrt{s}}(z)dG(s)$, G is a distribution on $[0,\infty)\}$. For $F \in \mathcal{F}_{SU}$, we additionally have $\mathrm{Var}_F(Z) = 1$ which translates to $\int s dG(s) = 3$ in terms of the mixing distribution G. We then have $P_F(|Z_1| \le c) = \int P(|U_{\sqrt{s}}| \le c)dG(s) = \int g_c(s)dG(s)$ where $g_c(s) = 1$ if $s \le c^2$ and $= c/\sqrt{s}$ if $s > c^2$. The function $g_c(s)$ thus equals 1 and is parallel to the horizontal axis up to $s \le c^2$, then it decreases convexly. Let $bs + 1$ be the straight line which is tangent to

$g_c(s)$ at $s_* > c^2$ and also meets $g_c(s)$ at $s = 0$. It follows that such a tangent will always lie below the curve $g_c(s)$, i.e., $g_c(s) \geq bs + 1 \ \forall s \geq 0$. Solving the two equations $b = \frac{d}{ds} g_c(s) \mid_{s=s_*}$ and $bs_* + 1 = g(s_*)$ as in Theorem 1, we get $s_* = 9c^2/4$. To satisfy the moment restriction $\int s dG(s) = 3$, we further need $s_* \geq 3 \Leftrightarrow c > 2/\sqrt{3}$. Arguing as in Theorem 1, we find that the mixing distribution $G_* = (1 - 3/s_*)\delta_{\{0\}} + (3/s_*)\delta_{\{s_*\}}$ attains $\min \int P(|U_{\sqrt{s}} \leq c) dG(s)$ over all mixing distributions G on $[0, \infty)$ which satisfy $\int s dG(s) = 3$.

If $c \leq 2/\sqrt{3}$, the tangent to the curve $g_c(s)$ at $s = 3$ passes below the point $(0, g_c(0))$. This implies that the tangent always lies below the curve $g_c(s)$. It follows that $G_* = \delta_{\{3\}}$ attains the minimum in this case and completes the proof of Theorem 3. □

Remark. For the minimizing distribution F_* of Theorem 3, we have $P_{F_*}(|Z_1| \leq c) = c/\sqrt{3}$ if $c \leq 2/\sqrt{3}$ and $= 1 - 4/(9c^2)$ if $c \geq 2/\sqrt{3}$. On the other hand, for any unimodal (about 0) distribution F which satisfies $E_F(Z) = 0$ and $\text{Var}_F(Z) = 1$ (not necessarily symmetric though), the well known Camp-Meidell inequality or the Gauss inequality (see Dharmadhikari and Joag-dev (1988), pp. 24–25) gives the lower bound $P_F(|Z| \leq c) \geq 1 - 4/(9c^2)$. Theorem 3 thus establishes that in the case of $c \geq 2/\sqrt{3}$, the lower bound of the Camp-Meidell-Gauss inequality is actually attained in the smaller class of only symmetric unimodal distributions, and in particular, it is attained when $F = F_*$. For the case of $c < 2/\sqrt{3}$, it is easy to establish that the exact infimum probability of $c/\sqrt{3}$ obtained in Theorem 3 is strictly sharper than the lower bound of the Camp-Meidell-Gauss inequality.

We next consider the case of general $n > 1$. We have $Z_1, \ldots, Z_n$ i.i.d. from a symmetric unimodal distribution $F \in \mathcal{F}_{SU}$ and we want to find the infimum coverage of the z-interval, $\inf_{F \in \mathcal{F}_{SU}} P_F(|\bar{Z}| \leq c \mid Z_i \text{ i.i.d.} \sim F)$. Since this "i.i.d. problem" is hard to tackle, we use our previous *independent embedding* technique and aim to find $\inf_{F_1,\ldots,F_n \in \mathcal{F}_{SU}} P(|\bar{Z}| \leq c \mid Z_i \sim F_i \text{ and independent})$.

The last infimum over the "independent class" is very similar to the infimum problem considered in Theorem 2. In Theorem 2, the infimum is over the normal scale mixture family whereas here it is over the class of scale mixtures of symmetric uniform distributions. It was thus expected that the infimum over the symmetric unimodal class can be obtained by using arguments similar to Theorem 2. We, however, found this path hard to follow. Theorem 2 uses two features: (infinite) differentiability of the normal cdf and differentiation under the integral sign. These features do not immediately generalize in the symmetric unimodal case due to the bounded support of the uniform distribution.

We, instead, follow an alternate line of argument and obtain the following result.

THEOREM 4. *Let* $F_{s_1,s_2} = \frac{s_2^2-3}{s_2^2-s_1^2} U[-s_1, s_1] + \frac{3-s_1^2}{s_2^2-s_1^2} U[-s_2, s_2]$. *Then* $\inf_{F_1,\ldots,F_n \in \mathcal{F}_{SU}} P(|\bar{Z}| \leq c \mid Z_i \sim F_i$ *and independent*$) = \inf_{s_{i1}, s_{i2} \geq 0} P(|\bar{Z}| \leq c \mid Z_i \sim F_{s_{i1}, s_{i2}}$ *and independent*$)$.

PROOF. By the Khintchine representation, $\inf_{F_1,\ldots,F_n\in\mathcal{F}_{SU}} P(|\bar{Z}| \leq c \mid Z_i \sim F_i$ and independent) $= \inf_{G_1,\ldots,G_n} \int h(s_1,\ldots,s_n)dG_1(s_1)\cdots dG_n(s_n)$ where the G_i's are distributions on $[0,\infty)$ satisfying $\int s^2dG(s) = 3$, the function $h(s_1,\ldots,s_n) = P(|U_{s_1}+\cdots+U_{s_n}|/n \leq c)$ and the U_{s_i}'s are independent $U[-s_i,s_i]$ random variables. Theorem 4 will be proved if we show that the latter infimum can be reduced to considerations of only two point mixing distributions $G_i = (1-p)\delta_{\{s_{i1}\}} + p\delta_{\{s_{i2}\}}$ which satisfy $\int s^2dG_i(s) = (1-p)s_{i1}^2 + ps_{i2}^2 = 3$. This last statement follows from the following result due to Mulholland and Rogers (1958). □

Result (*MR*). For each integer u with $1 \leq u \leq m$, let $f_{1u}(x),\ldots,f_{ku}(x)$ be Borel-measurable functions and let $\mathcal{F}_u$ be the set of distribution functions $F(x)$ satisfying $\int f_{ru}(x)dF(x) = 0$, $r = 1,\ldots,k$.

Let $g(x_1,\ldots,x_m)$ be Borel-measurable as a function of $(x_1,\ldots,x_m)$. Let $\mathcal{E}_u$ be the set of extreme functions of $\mathcal{F}_u$, for $u = 1,\ldots,m$. Then

$$\inf_{F_u\in\mathcal{F}_u} \int\cdots\int g(x_1,\ldots,x_m)dF_1(x_1)\cdots dF_m(x_m)$$
$$= \inf_{H_u\in\mathcal{E}_u} \int\cdots\int g(x_1,\ldots,x_m)dH_1(x_1)\cdots dH_m(x_m),$$

provided the left-hand side is finite.

The extreme functions of $\mathcal{F}_u$ are the step-functions of $\mathcal{F}_u$ having γ jumps at suitable points $x_u^1,\ldots,x_u^\gamma$, where $1 \leq \gamma \leq k+1$, and where the vectors $(1, f_{1u}(x_u^j),\ldots,f_{ku}(x_u^j))$, $j = 1,\ldots,\gamma$ are linearly independent.

Theorem 4 thus reduces the infimum problem over the "independent class" to an $2n$ dimensional minimization over $0 \leq s_{i1} \leq 1 \leq s_{i2}$, $i = 1,\ldots,n$. This finite dimensional minimization is obtained numerically. At each step of the minimization process, we need to compute $P(|\bar{Z}| \leq c \mid Z_i \sim F_{s_{i1},s_{i2}}$ and independent) (see Theorem 4). This probability, on the other hand, is a convex combination of $P(|U_{s_1}+\cdots+U_{s_n}|/n \leq c)$ where each s_i is either s_{i1} or s_{i2}.

The distribution of the mean of n independent $U[0,s_i]$ random variables can be found in Roach (1963). The density consists of at most 2^n-1 distinct connected arcs, making point contact. This result can easily be adapted to our case of the mean of n independent $U[-s_i,s_i]$ random variables.

The results of the $2n$ dimensional numerical minimization shows that the minimum, in every attempted case, is again attained at the "i.i.d. subclass", and in particular, when $s_{i1} = 0$, $i = 1,\ldots,n$ and $s_{12} = \cdots = s_{n2}$. We thus obtain the minimum over the "i.i.d. subclass" that we are seeking. These minimum coverage probabilities are listed in Table 1.

Remark. If $Z_1,\ldots,Z_n$ are i.i.d. $\sim F$ where F is a symmetric unimodal distribution about 0, then the convolution distribution F_+ of $Z_1+\cdots+Z_n$ is again symmetric unimodal at 0 by a result due to Winter (1938) (see Dharmadhikari and Joag-dev (1988), p. 13). We can then use the Camp-Meidell-Gauss inequality on the convoluted distribution F_+ to obtain the lower bound

$P(|\bar{Z}| \leq z_{\alpha/2}/\sqrt{n} \mid Z_i \text{ i.i.d.} \sim F) \leq 1-4/(9z_{\alpha/2}^2)$ for any $F \in \mathcal{F}_{SU}$. Our technique of Theorem 4 and the subsequent numerical work, on the other hand, obtain the exact infimum. The lower bounds from the Camp-Meidell-Gauss inequality are shown along with the exact infima in Table 1. The exact infimum values are always sharper. Moreover, the differences between the exact values and the lower bounds are notably significant in smaller critical values.

The minimum coverage probabilities in Table 1 suggest that not only the z-interval is conservative over the normal scale mixture family $\mathcal{F}_N$, it is indeed robust over the larger family $\mathcal{F}_{SU}$ of symmetric unimodal distributions as well. We are thus assured of a high degree of conservatism of the z-interval in finite samples, in presence of heavier as well as lighter tails, as long as the symmetry and unimodality of the underlying distribution is preserved. We note here that robustness investigations on the t-interval yield strikingly different behavior, Basu and DasGupta (1995) consider the infimum coverage of the t-interval over the symmetric unimodal family and find that the t-interval can be quite sensitive for small critical values.

4. Asymmetric unimodal distributions

In Sections 2 and 3, we have found that the coverage probability of the z-interval is, indeed, robust under symmetry and unimodality preserving departures from normality. But practical data often project distributions which are heavier tailed and also slightly asymmetric. In this section, we incorporate asymmetry as well as heavier and lighter (than normal) tail structures into our model by considering the class $\mathcal{F}_A$ as described in (1.3). We investigate the possible effect of asymmetry on the robustness of the z-interval by considering its worst coverage probability $\inf_F P(|\bar{Z}| \leq c \mid Z_i \text{ i.i.d.} \sim F)$ over the class $F \in \mathcal{F}_A$.

The mode M of a distribution $F \in \mathcal{F}_A$ is kept arbitrary in the description of the class $\mathcal{F}_A$ in (1.3). However, it is well known that for a unimodal distribution, $(\text{mean} - \text{mode})^2 \leq 3\text{variance}$ (see Dharmadhikari and Joag-dev (1988), p. 9. Basu and DasGupta (1996) obtain useful extensions of this inequality in univariate and multivariate situations). It follows that the mode $M \in [-\sqrt{3}, \sqrt{3}]$ for any $F \in \mathcal{F}_A$.

Since the infimum problem over the "i.i.d." class is hard to tackle, we again use our *independent embedding* technique to embed the problem over the larger "independent class". We obtain the following result.

THEOREM 5.

$$\inf_{F_1,\ldots,F_n \in \mathcal{F}_A} P(|\bar{Z}| \leq c \mid Z_i \sim F_i \textit{ and independent})$$
$$= \inf_{H_1,\ldots,H_n} \int P\left(-nc - nM \leq \sum_{i=1}^{n} U_i w_i \leq nc - nM\right) dH_1(w_1)\cdots dH_n(w_n)$$

where the U_i*'s are independent* $U[0,1]$ *random variables, and the* H_i*'s are 3-point distributions satisfying* $E_{H_i}(W_i) = -2M$*, and* $E_{H_i}(W_i^2) = 3(1+M^2)$ *with* $|M| \leq \sqrt{3}$.

PROOF. Let $Z_i \sim F_i \in \mathcal{F}_A$ be a unimodal random variable having mode at M. By the Khintchine representation of unimodal random variables, we have $Z_i - M = U_i W_i$ where $U_i \sim$ Uniform$[0,1]$, $W_i \sim H$, and U_i and W_i are independent. The moment restrictions $E_{F_i}(Z_i) = 0$ and $\text{Var}_{F_i}(Z_i) = 1$ translate to $E_{H_i}(W_i) = -2M$ and $E_{H_i}(W_i^2) = 3(1+M^2)$. The fact that it is enough to consider only 3-point mixing distribution H_i's now follows by a simple application of the MR result. □

Theorem 5 reduces the infimum problem over the "independent class" to a minimization over the 3-point H_i's and over $M \in (-\sqrt{3}, \sqrt{3})$. We obtain this minimization numerically. For the specific cases we tried, the numerical optimization always obtains the minimum at identical H_i's, thus establishing that the obtained minimum is actually the minimum over the "i.i.d. subclass". Moreover, most surprisingly, the minimum in all attempted cases is attained at $M = 0$ and, in fact, at a symmetric F. This implies that in the attempted cases $\inf_{F \in \mathcal{F}_A} P_F(|\bar{Z}| \leq c) = \inf_{F \in \mathcal{F}_{SU}} P_F(|\bar{Z}| \leq c)$ even though $\mathcal{F}_{SU}$ is only a subclass of $\mathcal{F}_A$. Thus, even if we allow asymmetric distributions, they do not affect the minimum coverage, a fact which we find rather intriguing. For other ways to formulate departure from symmetry, see MacGillivray (1986). We also note that Basu and DasGupta (1995) consider the robustness of the t-interval over asymmetric families. They find that asymmetry of the parent distribution can have drastic effect on the coverage probability of the t-interval.

5. Comments

It has long been a part of statistical folklore that z and t intervals are conservative for a sample from a long-tailed distribution. In this article, we establish some explicit results about the robustness of the z-interval under unimodal departures from normality. Our theoretical results show that the z-interval is indeed quite robust for broad general symmetric unimodal populations with heavier or lighter tails. We also find that inclusion of asymmetric unimodal distributions into our model does not have any effect on the infimum coverage of the z-interval. The main achievement of this article is thus to establish the common belief about the robustness of the z-interval on a firm theoretical ground.

Basu and DasGupta (1995), in contrast, consider the robustness of the t-interval over symmetric and asymmetric unimodal families. A comparison of their t-interval results with the results obtained here shows that there are real qualitative differences between the cases of the z and t intervals. We give a qualitative summary of this comparison below.

- t-interval

(i) *Normal scale mixture distributions*: The t-interval is conservative if the critical value is greater than 1.8 (Benjamini (1983)).

(ii) *Symmetric unimodal distributions*: The t-interval is conservative for high nominal coverage intervals (when the critical value is greater than a threshold value). However, when the critical value is <1, the t-interval can be very sensitive to nonnormality.

(iii) *Asymmetry*: The coverage probability of the t-interval is often sensitive to asymmetry of the parent distribution.

- z-interval

The z-interval is conservative for the *normal scale mixture distributions* as well as for the *symmetric unimodal distributions*. Inclusion of *asymmetric* distributions does not affect the infimum coverage of the z-interval.

Acknowledgements

Some part of this work was done when I was a Ph.D. student under the supervision of Anirban DasGupta. He first suggested this problem to me. He also has significant contributions in many of the results, specifically in the proof of Theorem 1. I am deeply indebted to him for his support and advice.

REFERENCES

Andrews, D. F. and Mallows, C. (1974). Scale mixtures of normal distributions, *J. Roy. Statist. Soc. Ser. B*, **36**, 99–102.

Basu, S. (1991). Robustness of Bayesian and classical inference under distribution bands and shape restricted families, Ph.D. Thesis, Purdue University.

Basu, S. (1996). Existence of a normal scale mixture with a given variance and a percentile, *Statist. Probab. Lett.*, **28**, 115–120.

Basu, S. and DasGupta, A. (1995). Robustness of standard confidence intervals for location parameters under departure from normality, *Ann. Statist.*, **23**, 1433–1442.

Basu, S. and DasGupta, A. (1996). The mean, median and mode of unimodal distributions: a characterization, *Theory Probab. Appl.*, **41**, 336–352.

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

Das, R. and Sinha, B. K. (1990). Robust optimum invariant tests of covariance structures useful in linear models, *Sankhyā Ser. A*, **52**, 244–258.

Dey, D. K. (1990). Estimation of scale parameters in mixture distributions, *Canad. J. Statist.*, **18**, 171–178.

Dey, D. K. and Birmiwal, L. R. (1996). On identifying mixing density of scale mixtures of normal distributions (preprint).

Dharmadhikari, S. and Joag-dev, K. (1988). *Unimodality, Convexity, and Applications*, Academic Press, San Diego.

Efron, B. and Olshen, R. A. (1978). How broad is the class of normal scale mixtures?, *Ann. Statist.*, **6**, 1159–1164.

Herrendörfer, G., Rasch, D. and Feige, K. D. (1983). Robustness of statistical methods II. Methods for the one-sample problem, *Biometrical J.*, **4**, 327–343.

Hoeffding, W. and Shrikhande, S. S. (1955). Bounds for the distribution function of a sum of independent, identically distributed random variables, *Ann. Math. Statist.*, **25**, 439–449.

Kemperman, J. H. B. (1968). The general moment problem, a geometric approach, *Ann. Math. Statist.*, **39**, 93–122.

Lehmann, E. L. and Loh, W. Y. (1990). Pointwise versus uniform robustness of some large-sample tests and confidence intervals, *Scand. J. Statist.*, **17**, 177–187.

Logan, B., Mallows, C., Rice, S. and Shepp, L. (1973). Limit distributions of self-normalized sums, *Ann. Probab.*, **1**, 788–807.

MacGillivray, H. L. (1986). Skewness and asymmetry: measures and orderings, *Ann. Statist.*, **14**, 994–1011.

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.
Sinha, B. K. and Kariya, T. (1989). *Robustness of Statistical Tests*, Academic Press, Boston.
Winter, A. (1938). *Asymptotic Distributions and Infinite Convolutions*, Edwards Brothers, Ann Arbor, Michigan.