ROBUSTNESS OF STEIN'S TWO STAGE PROCEDURE

by

Hyun Sook Oh and Anirban DasGupta
Purdue University

Technical Report #93-22

Department of Statistics
Purdue University

May 1993

ROBUSTNESS OF STEIN'S TWO STAGE PROCEDURE

by
HYUN SOOK OH AND ANIRBAN DASGUPTA
Purdue University

May 14, 1993

1 Introduction

The robustness of Stein's two-stage procedure against possible departures from normality was considered originally by Bhattacharjee(1965). There had been works to show that the procedure is quite robust. For example, Blumenthal and Govindarajulu(1977) claim that the procedure is remarkably robust under a mixture of two normal populations differing in location parameters and having the same unknown variance. Using an Edgeworth series model, Ramkaran(1983) showed that the procedure is quite robust by reviewing Bhattacharjee(1965)'s assertion. In this article, the robustness of the procedure is considered by investigating the expected sample size for appropriate nonnormal distributions such as the bounded uniform, Double exponential, t distributions with 3 and 5 degrees of freedom, logistic and Von Mises extreme value density. Notice the last distribution is asymmetric.

In section 2, Stein's two-stage procedure is briefly summarized. In section 3, a general formula for an upper bound on the ratio of the expected sample sizes for a normal population and a given nonnormal population is derived when the first stage sample size is taken to be 2. Let $X_1, X_2, \ldots$ be observations from a nonnormal population with unknown mean and unknown but finite variance. Then the derived upper bound on the ratio of the expected sample sizes is given by a function of a single variable,

$$R(z) = \frac{z^2 + 2z^2 \int_0^{\sqrt{2}/z} yF(y)dy}{2P(\chi_1^2 < 2/z^2) + z^2 P(\chi_3^2 > 2/z^2)}, \quad z \geq 0,$$

where F is the distribution function of $X_1 - X_2$. In the above, z plays the role of $\sigma/\sqrt{c}$, σ and c being the standard deviation and a preassigned constant respectively (see section 2). In section 4, we specialize to the case when the underlying distribution is absolutely continuous. Then $R(z)$ is demonstrated to be unimodal if the density function of F crosses the half-Normal density function only once from below. This can be regarded as a tail ordering. The implication of this unimodality result is that a local maximum must be a global maximum. The general Theorem is then applied to several specific distributions and in each case it is shown that the loss of efficiency is small, uniformly in σ and c. Throughout the article, we take an initial sample size equal to 2. Admittedly, this restricts the applicability of the results; but most of the analytic results including the general upper bound seemed unprovable for a general initial sample size. Plots of $R(z)$ are provided at the end. The principal achievements of this note are the exact bounds it was possible to obtain; in the process, the illustrative examples give rise to some interesting calculations.

2 The Two-Stage Procedure

Assume that $X_1, X_2, \ldots$ is a sequential sample from $\mathcal{N}(\mu, \sigma^2)$ (μ and σ^2 are unknown). Take an initial sample of size n_0 from this sequence and calculate the unbiased estimate of σ^2,

$$s^2 = \sum_{i=1}^{n_0} (x_i - \overline{x}_0)^2/(n_0 - 1)$$

where $\overline{x}_0$ is the mean of the first n_0 observations. Then $n - n_0$ additional observations are taken from the same sequence. The size of the second sample is such that

$$n = max\{n_0, [\frac{s^2}{c} + 1]\}.$$

The quantity $c > 0$ is a preassigned constant and $[y]$ denotes the largest integer less than or equal to y. Let $\overline{x}_n = \sum_{i=1}^{n} x_i/n$, the mean of n observations obtained by pooling the two samples; then the "pivotal-quanty"

$$t = \sqrt{n}(\overline{x}_n - \mu)/s$$

can be used for inference about μ. Stein(1945) has shown that the sampling distribution of t is Student's t with $n_0 - 1$ d.f. and the coverage probability of the interval $(\overline{x}_n \pm l)$, a function of σ^2, is always greater than or equal to $1-\alpha$, irrespective of the values of μ and σ^2, provided $c = l^2/b^2$. The quantity b is the upper $100\alpha/2$ % point of t distribution with $n_0 - 1$ d.f. A martingale argument can be given to show that the infimum coverage is exactly $1 - \alpha$. This fact does not appear to be formally known. For conditional coverages of the Stein two stage interval, see Casella(1988).

3 Upper Bound on the Ratio of the Expected Sample Size for Normal Data and a Given Nonnormal Data

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random observations from a population. Assume that $Var(X_i) = \sigma^2 < \infty$. Then, the expected sample size of the Stein procedure for an initial sample size n_0 is

$$E_{\sigma^2}(n) = n_0 + \sum_{m=n_0}^{\infty} P_{\sigma^2}(s^2 \geq cm)$$

where s^2 is the sample variance from the first sample of size n_0. Stein(1945) derived the following lower and upper bound for the expected sample size for a Normal population.

Lemma 1. *(Stein) Suppose $X_1, X_2, \ldots \sim \mathcal{N}(\mu, \sigma^2)$. Let $E_{\Phi,\sigma^2}(n)$ denote the expected sample size. Then*

$$\begin{aligned} & n_0 P(\chi^2_{n_0-1} < n_0(n_0-1)c/\sigma^2) + \tfrac{\sigma^2}{c} P(\chi^2_{n_0+1} > n_0(n_0-1)c/\sigma^2) \\ < \; & E_{\Phi,\sigma^2}(n) \\ < \; & n_0 P(\chi^2_{n_0-1} < n_0(n_0-1)c/\sigma^2) + \tfrac{\sigma^2}{c} P(\chi^2_{n_0+1} > n_0(n_0-1)c/\sigma^2) \\ & \qquad + P(\chi^2_{n_0-1} > n_0(n_0-1)c/\sigma^2). \end{aligned}$$

Note that the difference between the upper and lower bound of $E_{\Phi,\sigma^2}(n)$ is less than 1. So $E_{\Phi,\sigma^2}(n)$ can be approximately evaluated from tables of the incomplete gamma function by using one of the bounds.

From now on, throughout this paper, let us consider the following set up: Take $n_0 = 2$. Let $Y = (X_1 - X_2)/\sigma$ and let F be the distribution function of Y. Assume that F is absolutely continuous with the density function f. Then

$$E_{\sigma^2}(n) = 2 + \sum_{m=2}^{\infty} P(Y^2 > 2cm/\sigma^2).$$

An upper bound for $\frac{E_{\sigma^2}(n)}{E_{\Phi,\sigma^2}(n)}$ can be evaluated by taking an appropriate upper bound of $E_{\sigma^2}(n)$ and the lower bound of $E_{\Phi,\sigma^2}(n)$ as in Lemma 1. First we give an illustrative example.

Example 1. Let $X_1, X_2, \ldots$ be i.i.d. random variables from $Uniform(\mu - \sqrt{3}\sigma, \mu + \sqrt{3}\sigma)$. This parametrization keeps the interpretation of σ^2 as variance intact. Then $f(y) = \frac{1}{12}(2\sqrt{3} - |y|), |y| \leq 2\sqrt{3}$. So

$$\begin{aligned} E_{\sigma^2}(n) &= 2 + \sum_{m=2}^{\infty}(1 - \frac{\sqrt{2cm}}{\sqrt{3}\sigma} + \frac{2cm}{12\sigma^2})I_{(2cm \leq 12\sigma^2)} \\ &= \begin{cases} 2 & \text{if } \sigma^2/c < 1/3 \\ 2 + \sum_{m=2}^{[6\sigma^2/c]}(1 - \frac{\sqrt{2cm}}{\sqrt{3}\sigma} + \frac{2cm}{12\sigma^2}) & \text{if } \sigma^2/c \geq 1/3. \end{cases} \end{aligned}$$

But

$$\begin{aligned} \sum_{m=2}^{[6\sigma^2/c]}(1 - \frac{\sqrt{2cm}}{\sqrt{3}\sigma} + \frac{2cm}{12\sigma^2}) &< \int_1^{6\sigma^2/c}(1 - \frac{\sqrt{2cx}}{\sqrt{3}\sigma} + \frac{cx}{6\sigma^2})dx \\ &= -1 + \sigma^2/c + \frac{2\sqrt{2}}{3\sqrt{3}}\sqrt{c}/\sigma - \frac{1}{12}c/\sigma^2. \end{aligned}$$

Thus if $\sigma^2/c < 1/3$, $\frac{E_{\sigma^2}(n)}{E_{\Phi,\sigma^2}(n)}$ is less than

$$\frac{2}{2P(\chi_1^2 < 2c/\sigma^2) + \sigma^2/cP(\chi_3^2 > 2c/\sigma^2)}$$

and if $\sigma^2/c \geq 1/3$, it is less than

$$\frac{1 + \sigma^2/c + \frac{2\sqrt{2}}{3\sqrt{3}}\sqrt{c}/\sigma - \frac{1}{12}c/\sigma^2}{2P(\chi_1^2 < 2c/\sigma^2) + \sigma^2/cP(\chi_3^2 > 2c/\sigma^2)}.$$

Combined with Lemma 1, this gives an upper bound on the ratio of expected sample sizes.

Theorem 1. *Suppose the support of the underlying distribution is the whole Real line. Then $R(z)$ as defined below is an upper bound on the ratio of the expected sample sizes for Normal data and the given data, where*

$$R(z) = \frac{z^2 + 2z^2 \int_0^{\sqrt{2}/z} yF(y)dy}{2P(\chi_1^2 < 2/z^2) + z^2 P(\chi_3^2 > 2/z^2)}, \quad z = \sigma/\sqrt{c}.$$

Proof:

$$\begin{aligned}
\sum_{m=2}^{\infty} P(Y^2 > 2cm/\sigma^2) &\leq 2\int_1^{\infty}\int_{\sqrt{2cx}/\sigma}^{\infty} f(y)dydx \\
&= 2\int_{\sqrt{2c}/\sigma}^{\infty} f(y)(\frac{\sigma^2 y^2}{2c} - 1)dy \qquad (by\ Fubini) \\
&= \frac{\sigma^2}{c}(1 + 2\int_0^{\sqrt{2c}/\sigma} yF(y)dy) - 2.
\end{aligned}$$

The inequality holds because $P(Y^2 > 2cm/\sigma^2)$ is a monotone function of m. Thus

$$E_{\sigma^2}(n) < \frac{\sigma^2}{c} + \frac{2\sigma^2}{c}\int_0^{\sqrt{2c}/\sigma} yF(y)dy.$$

By using the lower bound of $E_{\Phi,\sigma^2}(n)$ given in Lemma 1, the result now follows. □

Discussion: Of course, the interesting question is what is a global upper bound on the ratio of expected sample sizes, uniformly in σ and c. A numerical maximization of $R(z)$ cannot be rigorously asserted as a tool for achieving this per se due to the unbounded nature of z. In the next section, we provide an analytic justification for the numerical method by establishing that $R(\cdot)$ is unimodal. This result implies that a numerical maximum is indeed a global maximum.

4 A Global Upper Bound

Example 1.(continued) Let $z = \sigma/\sqrt{c}$. Then, it is easy to see that

$$\sup_{z<1/\sqrt{3}} \frac{2}{2P(\chi_1^2 < 2/z^2) + z^2 P(\chi_3^2 > 2/z^2)} = 1.$$

And

$$\sup_{z\geq 1/\sqrt{3}} \frac{1 + z^2 + \frac{2\sqrt{2}}{3\sqrt{3}}\frac{1}{z} - \frac{1}{12}\frac{1}{z^2}}{2P(\chi_1^2 < 2/z^2) + z^2 P(\chi_3^2 > 2/z^2)} = 1.12744,$$

(by numerical maximization) which is attained at $z = 2.6167$ i.e. $\sigma = 2.6167\sqrt{c}$. Hence $\frac{E_{\sigma^2}(n)}{E_{\Phi,\sigma^2}(n)} < 1.12744$ for all σ and for all c.

Before proving the next theorem, we define a tail ordering of two random variables.

Definition 1. Let V and W have the density function g and h, respectively. The tail of V is thicker than the tail of $W (V >_t W)$ if $\exists K > 0$ such that $g(x) > h(x)$ for all $x > K$.

Theorem 2. *Let $U \sim \mathcal{N}(0, 2)$. With $Y = (X_1 - X_2)/\sigma$, let $Y >_t U$. Suppose that f crosses 2ϕ only once in the positive half line where ϕ is the density function of the standard Normal distribution. Then $R(z)$ is unimodal with a unique maximum.*

Proof: see Appendix.

Next, several distributions are considered as Examples. All distributions considered in the following Examples satisfy the conditions of Theorem 2. Thus any local maximum of $R(z)$ is the global maximum of $R(z)$, which is a global upper bound on the ratio of the expected sample sizes. Without loss of generality, assume that the mean is zero for the given distribution.

Example 2. Let $X_1, X_2, \ldots$ be i.i.d. observations from a Double Exponential distribution with variance σ^2. Then the density function of $Y = (X_1 - X_2)/\sigma$, is given by

$$f(y) = \frac{\sqrt{2}}{4} e^{-\sqrt{2}|y|}(1 + \sqrt{2}|y|), \quad -\infty < y < \infty.$$

Thus

$$\begin{aligned} E_{\sigma^2}(n) &< 2 + 2\int_1^\infty \int_{\sqrt{2cx}/\sigma}^\infty f(y)dydx \\ &= 2 + \frac{2\sigma^2}{c} \int_{\sqrt{c}/\sigma}^\infty y(1+y)e^{-2y}dy \\ &= 2 + (\frac{2\sigma}{\sqrt{c}} + \frac{\sigma^2}{c} + 1)e^{-2\sqrt{c}/\sigma}. \end{aligned}$$

Hence

$$R(z) = \frac{2 + (z+1)^2 e^{-2/z}}{2P(\chi_1^2 < 2/z^2) + z^2 P(\chi_3^2 > 2/z^2)},$$

with

$$\sup_{z>0} R(z) = 1.1514 \quad at \quad z = \sigma/\sqrt{c} = 1.38839.$$

That is, the ratio of the expected sample sizes for Normal data and Double exponential data having the same variance is at most 1.1514 for all σ and c.

Example 3. Let $X_1, X_2, \ldots$ be i.i.d. observations from the Student's t distribution with 3 degrees of freedom, i.e., the density function is given by $\frac{2}{\sigma\pi}\frac{1}{(1+x^2/\sigma^2)^2}$. Then

$$f(y) = \frac{4(y^2+20)}{(\pi(y^2+4)^3}, \quad -\infty < y < \infty.$$

Thus

$$E_{\sigma^2}(n) \;<\; 2 + 2\int_1^\infty \int_{\sqrt{2cx}/\sigma}^\infty \frac{4(y^2+20)}{\pi(y^2+4)^3} dydx$$

$$
\begin{aligned}
&= 2+2\int_{\sqrt{2c}/\sigma}^{\infty}\frac{4(y^2+20)}{\pi(y^2+4)^3}(\frac{\sigma^2}{2c}y^2-1)dy \\
&= 1+z^2+\frac{16\sqrt{2}z}{\pi(4+2/z^2)^3}-\frac{2z^2}{\pi}tan^{-1}(\frac{\sqrt{2}}{2z})+\frac{8\sqrt{2}}{\pi z(4+2/z^2)^2} \\
&\quad +\frac{4\sqrt{2}}{\pi z(4+2/z^2)}+\frac{2}{\pi}\tan^{-1}(\frac{\sqrt{2}}{2z}), \qquad z=\sigma/\sqrt{c}.
\end{aligned}
$$

It follows that

$$\sup_{z>0} R(z) = 1.17023 \ \ \textit{which is attained at} \ \ z=\sigma/\sqrt{c}=1.3108.$$

Example 4. Let $X_1, X_2, \ldots$ be i.i.d. observations from the Student's t distribution with 5 degrees of freedom. i.e., the density function of the distribution is given by $\frac{8}{3\sqrt{3}}\frac{1}{\pi\sigma}\frac{1}{(1+\frac{x}{3\sigma^2})^3}$. Then

$$f(y)=\frac{16}{3\sqrt{3}\pi}\frac{\frac{1}{9}y^4+8y^2+336}{(\frac{1}{3}y^2+4)^5}, \quad -\infty<y<\infty.$$

Thus

$$
\begin{aligned}
E_{\sigma^2}(n) &< 2+2\int_{\sqrt{2c}/\sigma}^{\infty}\frac{16}{3\sqrt{3}\pi}\frac{\frac{1}{9}y^4+8y^2+336}{(\frac{1}{3}y^2+4)^5}(\frac{\sigma^2y^2}{2c}-1)dy \\
&= (8960\pi-8575tan^{-1}(\sqrt{2}z))(1+24z62+216z^4+864z^6+1296z^8) \\
&\quad +8575\sqrt{6}z(1+22z^2+\frac{876}{5}z^4+\frac{20088}{35}z^6) \\
&\quad +35\sqrt{24^9}z^5H_z(3/2,5/2)(54z^2+6z^4+3/2z^6+z^8) \\
&\quad +7\sqrt{2^{19}}\sqrt{27}z^5H_z(5/2,7/2)(-1+12z^2+24z^4+ \\
&\quad 6912z^6+29808z^8+5184z^{10})+5\sqrt{2^{23}}\sqrt{3^7}z^7H_z(7/2,9/2)(-1-3z^2+ \\
&\quad 288z^4+3672z^6+16848z^8+27216z^{10}),
\end{aligned}
$$

where

$$H_z(a,b)=\frac{\Gamma(b)}{\Gamma(a)\Gamma(b-a)}\int_0^1 t^{a-1}(1-t)^{b-a-1}(1-6tz^2)^{-5}dt$$

$$\textit{and} \quad z=\sigma/\sqrt{c}.$$

Hence

$$\sup_{z>0} R(z) = 1.14503 \ \ \textit{which is attained at} \ \ z=\sigma/\sqrt{c}=1.40092.$$

Example 5. Let $X_1, X_2, \ldots$ be i.i.d. observations from a Logistic distribution with variance σ^2, i.e., the density function is given by $\frac{\pi}{\sqrt{3}\sigma}\frac{e^{-\pi x/\sqrt{3}\sigma}}{(1+e^{-\pi x/\sqrt{3}\sigma})^2}$. Let $W = |Y| = |X_1 - X_2|/\sigma$. Then

$$W = (X_{(2)} - X_{(1)})/\sigma,$$

where $X_{(i)}$ denotes the *ith* order statistic. *Gupta and Shah*[1965] showed that the distribution of W has the density

$$f(w) = \frac{2\pi e^{\pi w/\sqrt{3}}(6+\sqrt{3}\pi w + (\sqrt{3}\pi w - 6)e^{\pi w/\sqrt{3}})}{3\sqrt{3}(-1+e^{\pi w/\sqrt{3}})^3}.$$

Thus

$$\begin{aligned}
E_\sigma(n) &< 2 + \int_{\sqrt{2c}/\sigma}^{\infty} hs(\frac{\sigma^2 w^2}{2c} - 1) f(w)dw \\
&= 2 + \frac{\sigma^2}{2c}(E(W^2) - \int_0^{\sqrt{2c}/\sigma} w^2 f(w)dw) - \int_{\sqrt{2c}/\sigma}^{\infty} f(w)dw \\
&= 2 + \frac{z^2}{2}(2 - \int_0^{\sqrt{2}/z} w^2 f(w)dw) - \int_{\sqrt{2}/z}^{\infty} f(w)dw, \quad z = \sigma/\sqrt{c} \\
&= 2 - \frac{\sqrt{2}\pi e^{\sqrt{2}\pi/(\sqrt{3}z)} + \sqrt{3}z - \sqrt{3}z e^{\sqrt{2}\pi/(\sqrt{3}z)}}{\sqrt{3}z(-1+e^{\sqrt{2}\pi/(\sqrt{3}z)})^2} \\
&\quad + \frac{z^2}{2}(4 + \frac{\sqrt{32/3}\pi}{z^3(-1+e^{\sqrt{2}\pi/(3z)})^2} + \frac{4}{z^2} - \frac{4(\sqrt{2}\pi + \sqrt{3}z)}{\sqrt{3}z^3(1-e^{\sqrt{2}\pi/(3z)})} \\
&\quad - \frac{\sqrt{96}}{\pi z}\ln(1 - e^{\sqrt{2}\pi/(3z)}) - \frac{12}{\pi^2} Polylog[2, e^{\sqrt{2}\pi/(3z)}]),
\end{aligned}$$

where

$$Polylog[2,t] = \int_t^0 \frac{\ln(1-x)}{x}dx.$$

Hence

$$\sup_{z>0} R(z) = 1.13843 \quad \textit{which is attained at } z = \sigma/\sqrt{c} = 1.43368.$$

Example 6. Let $X_1, X_2, \ldots$ be i.i.d. observations from the Extreme value distribution with the density function as

$$\frac{1}{\theta}e^{(x-\mu)/\theta}e^{-e^{-(x-\mu)/\theta}}, \quad \theta = \frac{\sqrt{6}\sigma}{\pi}.$$

Notice that this ditribution is not symmetric. It is known that the distribution of $Y = X_1 - X_2$ is a Logistic distribution (*Gumbel*(1961)). That is,

$$f(y) = \frac{1/\beta e^{-y/\beta}}{(1+e^{-y/\beta})^2}, \quad \beta = \frac{\sqrt{6}}{\pi}.$$

Thus

$$E_\sigma(n) < z^2 + 2z^2 \int_0^{\sqrt{2}/z} \frac{y}{1+e^{-\pi y}} dy$$

Numerical calculation now gives that

$$\sup_{z>0} R(z) = 1.14426 \quad \textit{which is attained at } z = \sigma/\sqrt{c} = 1.40703.$$

5 Discussion

Our exact results show that Stein's Two-Stage Procedure is quite robust with respect to the expected sample size under the distributions considered (Bounded Uniform, Double Exponential, t distributions with 3 and 5 degrees of freedom, Logistic and Von Mises Extreme Value distribution) when the initial sample size is 2. Specially, an asymmetric distribution, Von mises extreme value distribution, was considered and it worked as well as others. Except the uniform, each distribution considered as an example here belongs to the family of scale mixtures of Normal distributions. Most distributions in that family satisfy the condition of Theorem 2. Robustness with respect to the coverage probability will be treated elsewhere.

References

[1] BHATTACHARJEE, G. P. (1965). Effect of non-normality on Stein's two sample test. *Ann. Math. Statist.* **36** *651-663.*

[2] BLUMENTHAL, S. and GOVINDARAJULU, Z. (1977). Robustness of Stein's two-stage procedure for mixtures of normal population. *J. Amer. Statist. Assoc.* **72** *192-196.*

[3] CASELLA, G. (1988). Conditionally Acceptable Frequentist Solutions. *Proceedings of Purdue Symposium.* **1** *73-84.*

[4] GUMBEL, E. J. (1961). Sommes et différences de valeurs extrêmes indépendentes. *Compts Rendus de l'Académie des Sciences, Paris,* **253** *2838-2839.*

[5] GUPTA, S. S. and SHAH, B. K. (1965). Exact moments and percentage points of the order statistics and the distribution of the range from the logistic distribution. *Ann. Math. Statist.* **36** *907-920.*

[6] JOHNSON, N. L. and KOTZ, S. (1970). *Continuous univariate distributions - 1.* John Wiley & Sons, Inc.

[7] JOHNSON, N. L. and KOTZ, S. (1970). *Continuous univariate distributions - 2.* John Wiley & Sons, Inc.

[8] RAMKARAN (1983). The robustness of Stein's two-stage procedure. *Ann. Stat.* **11** *1251-1256.*

[9] STEIN, C. (1945). A two sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.* **16** *243-258.*

Appendix

Proof of Theorem 2:

Let

$$Q(z) = 1 + 2\int_0^{\sqrt{2}/z} xF(x)dx$$

and let

$$P(Z) = 2/z^2 P(\chi_1^2 < 2/z^2) + P(\chi_3^2 > 2/z^2).$$

Then $R(z) = \frac{Q(z)}{P(z)}$.
Let us look at the sign of the second derivative at the zeros of the first derivative, which is

$$sgn(Q''(z)P(z) - Q(z)P''(z)) = sgn(Q''(z) - \frac{Q'(z)}{P'(z)}P''(z). \qquad (1)$$

Now,

$$Q'(z) = -\frac{4}{z^3}F(\sqrt{2}/z), P'(z) = -\frac{4}{z^3}P(\chi_1^2 < 2/z^2).$$

and

$$\begin{aligned} Q''(z) &= \frac{12}{z^4}F(\sqrt{2}/z) + \frac{4\sqrt{2}}{z^5}f(\sqrt{2}/z), \\ P''(z) &= \frac{12}{z^4}P(\chi_1^2 < 2/z^2) + \frac{8}{\sqrt{\pi}z^5}e^{-1/z^2}. \end{aligned}$$

Let $h(x) = 2\phi(x) = \sqrt{\frac{2}{\pi}}e^{-x^2/2}, \quad x > 0$.
Then

$$P(\chi_1^2 < 2/z^2) = \int_0^{2/z^2} \frac{1}{2\pi}\frac{1}{\sqrt{x}}e^{-x/2}dx = \int_0^{\sqrt{2}/z} h(x)dx \equiv H(\sqrt{2}/z), \quad say.$$

Thus

$$(1) = sgn(\frac{f(\sqrt{2}/z)}{F(\sqrt{2}/z)} - \frac{h(\sqrt{2}/z)}{H(\sqrt{2}/z)}).$$

By the assumptions, the tail of F is thicker than that of H and so h cuts f exactly once (and from below).
Let x_0 be the crossing point. Then clearly $x_0 > m_h$, where m_h is the median of h. Because if $x_0 \leq m_h$, $F(x_0) < H(x_0) < H(m_h) = \frac{1}{2}$, which is impossible since f is symmetric about 0.
Define $t = \inf\{x > 0 : F(x) \leq H(x)\}$. Then $m_h < t \leq x_0$.
We will now prove $R(z)$ is unimodal. The proof consists of two steps: (a) To show that R has at most one mode on $(\sqrt{2}/t, \infty)$; (b) R has no extrema on $(0, \sqrt{2}/t]$.
(a) Consider $z > \sqrt{2}/t \ (\Leftrightarrow \sqrt{2}/z < t)$:
Then $f(\sqrt{2}/z) < h(\sqrt{2}/z)$ and $F(\sqrt{2}/z) > H(\sqrt{2}/z)$.
So if $\exists y \ni R'(y) = 0$, then

$$sgn(R''(y)) < 0.$$

Hence every extrema on $(\sqrt{2}/t, \infty)$ is a maximum. Thus there exists at most one maximum on $(\sqrt{2}/t, \infty)$.
(b) Consider $z \leq \sqrt{2}/z$ $(\Leftrightarrow \sqrt{2}/z \geq t)$:
Suppose that $\exists v \ni R'(v) = 0$.
Then

$$R(v) = \frac{F(\sqrt{2}/v)}{H(\sqrt{2}/v)} \leq 1. \tag{2}$$

But

$$\begin{aligned} E_{\sigma^2}(n) &= 2 + \sum_{m=2}^{\infty} P(S^2 > cm) \\ &= 2 + 2\sum_{m=2}^{\infty} P(Y > \sqrt{2cm}/\sigma) \\ &= 2 + 2\sum_{m=2}^{\infty}(1 - F(\sqrt{2m}/z)), \quad z = \sigma/\sqrt{c}. \end{aligned}$$

Let G denote the distribution function of $U \sim \mathcal{N}(0, 2)$.
By the assumption of $Y >_t U$, for z near 0,

$$1 - F(\sqrt{2m}/z) \geq 1 - G(\sqrt{2m}/z) \quad for \;\; each \;\; m = 2, 3, \ldots.$$

Now,

$$E_{\Phi,\sigma^2}(n) = 2 + 2\sum_{m=2}^{\infty}(1 - G(\sqrt{2m}/z)).$$

So $E_{\sigma^2}(n) \geq E_{\Phi,\sigma^2}(n)$.
Thus $R(z) > \frac{E_{\sigma^2}(n)}{E_{\Phi,\sigma^2}(n)} \geq 1$ for z near 0.
Hence if there is v such that $R'(v) = 0$, then there must be an extrema above 1. This contradicts (2). That is, there is no extrema on $(0, \sqrt{2}/t]$.

Consequently, there is at most one mode for $z > 0$.
We will now finally demonstrate that there is exactly one mode.

First, $\lim_{z\to\infty} R(z) = 1$. And since

$$\begin{aligned} 2 \leq E_{\sigma^2}(n) &\leq z^2 + 2z^2 \int_0^{\sqrt{2}/z} xF(x) \\ &\leq z^2 + 2F(\sqrt{2}/z) \end{aligned}$$

and $$\lim_{z\to 0}(z^2 + 2F(\sqrt{2}/z)) = 2,$$

one also has that

$$\lim_{z\to 0} R(z) = 1.$$

Also, $R(z) > 1$ for z near zero. Hence there must be one mode and $R(z) \geq 1$ for all $z > 0$. □

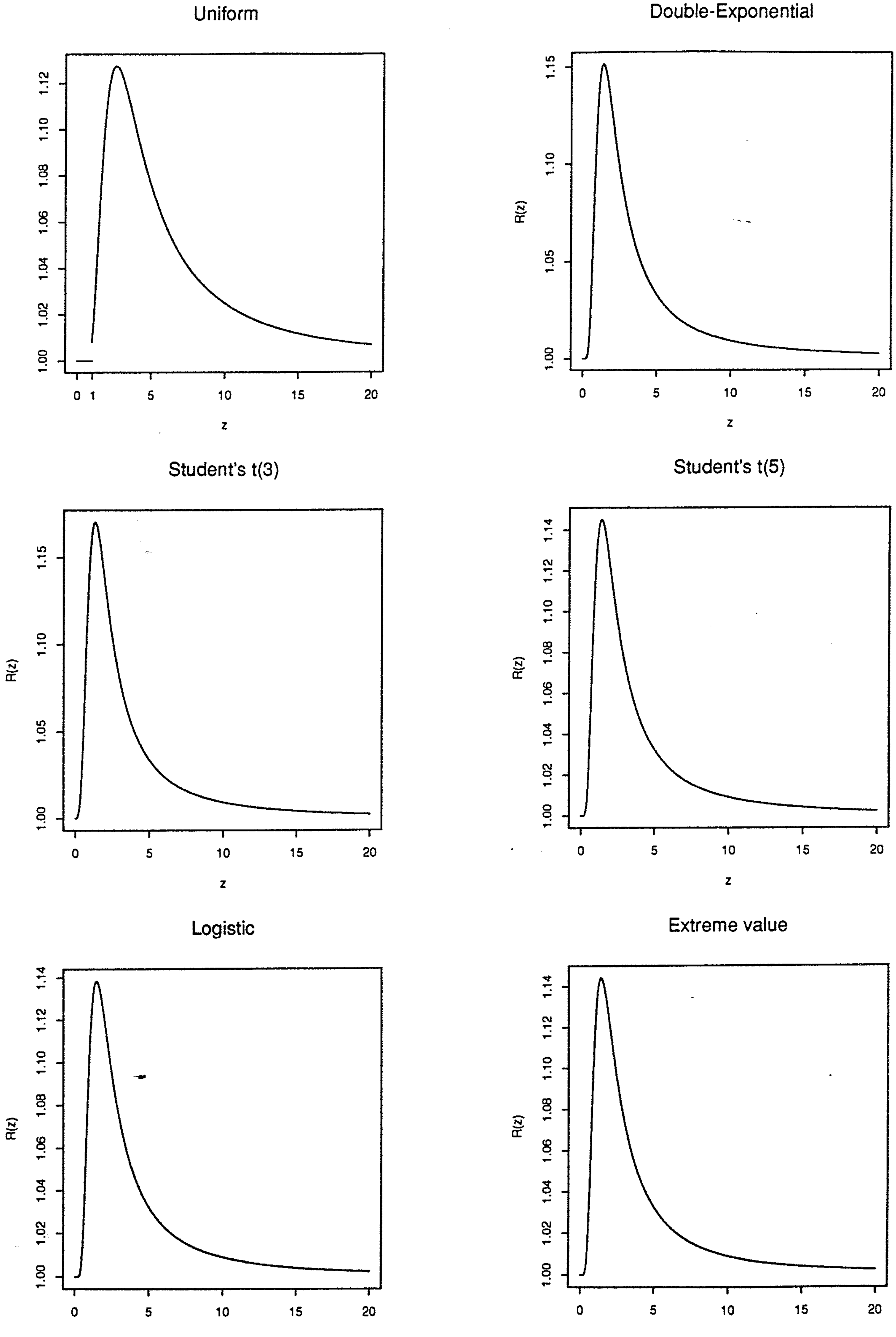

Figure 1: Plots of Upper bound on the ratio of the expected sample size for normal distribution and each distribution where $z = \sigma/\sqrt{c}$