Journal of Mathematical Sciences, Vol. 139, No. 3, 2006

STUDENT'S t-TEST FOR GAUSSIAN SCALE MIXTURES

N. K. Bakirov* **and G. J. Székely†** UDC 519.21

A Student-type test is constructed under a condition weaker than normal. We assume that the errors are scale mixtures of normal random variables and compute the critical values of the suggested s-test. Our s-test is optimal in the sense that if the level is at most α, then the s-test provides the minimum critical values. (The most important critical values are tabulated at the end of the paper.) For $\alpha \le .05$, the two-sided s-test is identical with Student's classical t-test. In general, the s-test is a t-type test, but its degree of freedom should be reduced depending on α. The s-test is applicable for many heavy-tailed errors, including symmetric stable, Laplace, logistic, or exponential power. Our results explain when and why the P-value corresponding to the t-statistic is robust if the underlying distribution is a scale mixture of normal distributions. Bibliography: 24 *titles.*

1. Introduction

Student's classical t-test [22] is particularly vulnerable to long-tailed nonnormality. In this paper, a new statistic is proposed to guard against this situation. The new test is optimal in the sense that it minimizes the critical values in the family of Gaussian scale mixtures when the level is at most a given number. Our theorems are closely related to the problems in [3, Sec. 6.1] and [2].

Let $X_1, X_2, \dots, X_n$ be independent normal random variables with common mean μ and not necessarily equal variances σ_k^2 (at least one of them nonzero), $\overline{X} = \sum_{k=1}^n X_k/n$, $S_X^2 = \sum_{k=1}^n (X_k - \overline{X})^2/(n-1) \neq 0$, and $T_n = \sqrt{n}(\overline{X} - \mu)/S_X$.

If $\sigma_1 = \sigma_2 = \ldots = \sigma_n$ and

$$R = \frac{nx^2}{x^2 + n - 1}, \tag{1}$$

then for $x \ge 0$ and $n \ge 2$,

$$P\{|T_n| > x\} = P\{|t_{n-1}| > x\} = P\left\{ \frac{\left(\sum_{i=1}^n \xi_i\right)^2}{\sum_{i=1}^n \xi_i^2} > R \right\},$$

where $\xi_1, \xi_2, \dots, \xi_n$ are i.i.d. standard normal random variables and t_{n-1} is a t-distributed random variable with degree of freedom $n-1$. (For the idea of this equation, see [6, p. 1279].)

In the nonhomogeneous case, denote the supremum of the double-tail probability by

$$2\overline{s}_{n-1}(x) := \sup_{\substack{\sigma_k \ge 0 \\ k=1,2,\dots,n}} P\{|T_n| > x\}. \tag{2}$$

We also need the notation $s_{n-1}(x) = 1 - \overline{s}_{n-1}(x)$ and the inverse transformation of (1): $x = \sqrt{R(n-1)/(n-R)}$.

Theorem 1. *For arbitrary $x \ge 0$ and $n \ge 2$,*

$$\overline{s}_{n-1}(x) = \max_{R < k \le n} P\left\{ t_{k-1} > \sqrt{\frac{R(k-1)}{k-R}} \right\},$$

where $s_{n-1}(x) = 1/2$ if $0 \le x < 1$, $s_{n-1}(1) = 3/4$, and $s_{n-1}(x) = t_{n-1}(x)$ for $x \ge \sqrt{3(n-1)/(n-3)}$.

Theorem 1 can easily be generalized to arbitrary scale mixtures of Gaussian errors. Their PDF has the form $\int_0^\infty \varphi((x-\mu)/\sigma)\, dF(\sigma)$, where φ is the standard normal PDF and $F(\sigma)$ is an arbitrary CDF on the nonnegative half-line. Concerning scale mixtures of normal distributions, see [7] and [11]. Scale mixtures are important in finance and in many other areas of applications where the errors are heavy-tailed, e.g., symmetric stable. Normal scale mixtures also include Student's t, Laplace, logistic, exponential power distributions, etc. See, e.g., [13] and [11].

Theorem 1 obviously implies the following result.

*Institute of Mathematics, Ufa, Russia, e-mail: bakirovnk@rambler.ru.

†Rényi Institute of Mathematics, Budapest, Hungary, e-mail: gabors@bgnet.bgsu.edu.

Published in *Zapiski Nauchnykh Seminarov POMI*, Vol. 328, 2005, pp. 5–19. Original article submitted October 7, 2005.

1072-3374/06/1393-6497

Theorem 2. *Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sample from a Gaussian scale mixture, and let Y_k be independent normal $(0, \sigma_k^2)$ random variables, $\overline{Y} = \sum_{k=1}^n Y_k/n$, and $S_Y^2 = \sum_{k=1}^n (Y_k - \overline{Y})^2/(n-1)$. Then*

$$P\left\{\sqrt{n}\,\frac{\overline{X}-\mu}{S_X} > x\right\} = \int\limits_{R^n} P\left\{\sqrt{n}\,\frac{\overline{Y}}{S_Y} > x\right\} \prod_{k=1}^{n} d\,F(\sigma_k) \le \overline{s}_{n-1}(x).$$

Let us introduce the notation

$$g_k(R) = P\left\{\frac{\left(\sum_{i=1}^k \xi_i\right)^2}{\sum_{i=1}^k \xi_i^2} > R\right\} = P\left\{|t_{k-1}| > \sqrt{\frac{R(k-1)}{k-R}}\right\}$$

and

$$\Delta_k(R) = g_{k+1}(R) - g_k(R).$$

Proposition.

(i) *For $k = 2, 3, \ldots, n-1$, there exists a unique point $r(k) \in (1, k)$ such that $\Delta_k(R) < 0$ if $R < r(k)$ and $\Delta_k(R) > 0$ if $r(k) < R < k$;*

(ii) *$r(1) := 1 < r(2) < r(3) < \ldots < r(n-1) < r(n)$, i.e., the sequence $r(k)$ is strictly increasing;*

(iii) *$r(k) \underset{k\to\infty}{\longrightarrow} 3$.*

Corollary 1. (i) *For $R \in [r(k-1), r(k)]$, $k = 2, 3, \ldots, n-1$,*

$$s_{n-1}(x) = P\left\{t_{k-1} > \sqrt{\frac{R(k-1)}{k-R}}\right\}.$$

(ii) *For $R \ge r(n-1)$,*

$$s_{n-1}(x) = t_{n-1}(x).$$

According to our Table 1, the one-sided level 0.025 s-critical values coincide with the classical t-critical values. *Splus* can easily compute that $r(2) = 1.726$ and $r(3) = 2.040$; thus, according to Table 1, for the one-sided level $\alpha = 0.125$ critical values, $\overline{s}_{n-1}(x) = \overline{t}_1(R) = 0.125$ and, similarly, $\overline{s}_{n-1}(x) = \overline{t}_2(R) = 0.1$. One can also compute that $\overline{s}_{n-1} = \overline{t}_{\min(n-1,13)} = .05$.

Corollary 2. *For $x \ge 0$, the scale mixture counterpart of the standard normal CDF is*

$$\Phi^*(x) := \lim_{n\to\infty} s_n(x) = \sup_{x^2<k} P(t_{k-1} \le x\sqrt{(k-1)/(k-x^2)}) \qquad (*)$$

($\Phi^*(-x) = 1 - \Phi^*(x)$). For $0 \le x < 1$, $\Phi^*(x) = .5$; $\Phi^*(1) = .75$; for $x \ge \sqrt{3}$, $\Phi^*(x) = \Phi(x)$, where $\Phi(x)$ is the standard normal CDF ($\Phi^*(\sqrt{3}) = \Phi(\sqrt{3}) = 0.958$). For quantiles between .5 and .875, the supremum in $(*)$ is attained at $k = 2$, and thus in this interval $\Phi^*(x) = C(x/\sqrt{(2-x^2)})$, where $C(x)$ is the standard Cauchy CDF. It is interesting to compare some critical values of Φ^* and Φ (when they do not coincide): $0.95 = \Phi(1.645) = \Phi^*(1.650)$, $0.9 = \Phi(1.282) = \Phi^*(1.386)$, $0.875 = \Phi(1.150) = \Phi^*(1.307)$ *(see the last row of Table 1).*

On the robustness of the t-statistic and on substitute t-statistics see, e.g., [24]. In this paper, it was found that the "trimmed" t is distributed approximately as a t-variable with reduced degrees of freedom. This result is similar to ours: s-statistics are t-type statistics with reduced degrees of freedom.

Our approach can also be applied to two-sample tests. In a forthcoming paper, the Behrens–Fisher problem will be discussed for Gaussian scale mixture errors with the help of our function $s_n(x)$.

If the error distribution is not necessarily a scale mixture of normal distributions, but is symmetric and unimodal, then, according to a classical result of Khintchin, the errors are scale mixtures of centered uniform distributions (see, e.g., [9, p. 155]). Thus if random variables $U_1, U_2, \ldots, U_n$ are independent and uniformly distributed on $[-1, 1]$, then our theorems suggest that in this case $\overline{s}_{n-1}(x)$ should be replaced by

$$\overline{u}_{n-1}(x) = \max_{R<k\le n} P\left\{\frac{U_1 + U_2 + \cdots + U_k}{\sqrt{U_1^2 + U_2^2 + \cdots + U_k^2}} > \sqrt{\frac{R(k-1)}{k-R}}\right\}.$$

We plan to return to this problem in another paper. For a related result, see [2].

Finally, on the history of Student's test and on the problem of nonnormal errors see [10, 17, 19, 20, 1, 14, 21, 12, 24, 6, 18, 15, 8, 5, 23].

Table 1. Critical x-values for the one-sided s-test

n-1	0.125	0.100	0.050	0.025
2	1.625	1.886	2.920	4.303
3	1.495	1.664	2.353	3.182
4	1.440	1.579	2.132	2.776
5	1.410	1.534	2.015	2.571
6	1.391	1.506	1.943	2.447
7	1.378	1.487	1.895	2.365
8	1.368	1.473	1.860	2.306
9	1.361	1.462	1.833	2.262
10	1.355	1.454	1.812	2.228
11	1.351	1.448	1.796	2.201
12	1.347	1.442	1.782	2.179
13	1.344	1.437	1.771	2.160
14	1.341	1.434	1.761	2.145
15	1.338	1.430	1.753	2.131
16	1.336	1.427	1.746	2.120
17	1.335	1.425	1.740	2.110
18	1.333	1.422	1.735	2.101
19	1.332	1.420	1.730	2.093
20	1.330	1.419	1.725	2.086
21	1.329	1.417	1.722	2.080
22	1.328	1.416	1.718	2.074
23	1.327	1.414	1.715	2.069
24	1.326	1.413	1.712	2.064
25	1.325	1.412	1.709	2.060
100	1.311	1.392	1.664	1.984
500	1.307	1.387	1.652	1.965
1,000	1.307	1.386	1.651	1.962

2. Proof of Theorem 1

Let $\xi_k = (X_k - \theta)/\sigma_k$ be i.i.d. standard normal random variables. Then $A := \{|T_n| > x\} = \{R\sum_{k=1}^n \sigma_k^2\xi_k^2 - \left(\sum_{k=1}^n \sigma_k\xi_k\right)^2 < 0\}$. If $R < 1$, or, equivalently, $|x| < 1$, then the supremum in (2) is $2\overline{s}_{n-1}(x) = 1$, and it is reached when $\sigma_1 = \sigma_2 = \ldots = \sigma_{n-1} = 0$, $\sigma_n \neq 0$. If $x \geq 1$, or, equivalently, $1 \leq R < n$, then $A = \{(\xi, G\xi) < 0\}$, where $\xi = (\xi_1, \xi_2, \ldots, \xi_n)^T, G = D(RI - E)D$, I is the $n \times n$ unit matrix,

$$D = \begin{pmatrix} \sigma_1 & 0 & \ldots & 0 \\ 0 & \sigma_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_n \end{pmatrix}, \qquad E = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ 1 & 1 & \ldots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \ldots & 1 \end{pmatrix}.$$

We can compute the eigenvalues $\lambda_k, k = 1, 2, \ldots, n$, of the matrix G from its characteristic equation $f(\lambda) := \det(G - \lambda I) = 0$.

The following lemma is proved in the Appendix.

Lemma 1.

$$f(\lambda) = \left(1 - \sum_{k=1}^n \frac{\sigma_k^2}{R\sigma_k^2 - \lambda}\right) \prod_{k=1}^n \left(\lambda - R\sigma_k^2\right) = 0. \qquad (3)$$

This equation has a single negative root, because for $\lambda < 0$ only the first factor in (3) can be zero, and the sum in the first factor decreases monotonically from $n/R > 1$ (when $\lambda = 0$) to 0 as $\lambda \to \infty$. For definiteness, denote the unique negative root of (3) by λ_n; thus all the other roots λ_k are nonnegative ($k = 1, 2, \dots, n-1$). Since the sum of the roots of (3) is equal to the negative of the coefficient of λ^{n-1} in the expansion of (3), we have

$$\sum_{k=1}^{n} \lambda_k = (R-1) \sum_{k=1}^{n} \sigma_k^2. \tag{4}$$

The following lemma is also proved in the Appendix.

Lemma 2. *Let $\xi_i, i=0,1,\dots,n$, be i.i.d. standard normal random variables. Then for every n and arbitrary $\mu_1, \mu_2, \dots, \mu_n \ge 0$,*

$$P\left\{\xi_0^2 \ge \sum_{k=1}^{n} \mu_i \xi_i^2\right\} = \frac{1}{\pi} \int_0^\infty \frac{dt}{\sqrt{t}(1+t)\sqrt{\prod_{k=1}^{n}(1+(1+t)\mu_i)}}.$$

This means that

$$P\{A\} = P\left\{\sum_{k=1}^{n} \lambda_k \xi_k^2 < 0\right\} = P\left\{\frac{\xi_n^2}{\sum\limits_{k=1}^{n-1} \frac{\lambda_k}{|\lambda_n|}\xi_k^2} > 1\right\} = \int_0^\infty \frac{t^{-\frac{1}{2}}(1+t)^{-1}\,dt}{\pi\sqrt{\prod\limits_{k=1}^{n-1}\left(1+\frac{\lambda_k}{|\lambda_n|}(1+t)\right)}}. \tag{5}$$

The event A does not change if we multiply every σ_k by the same positive constant; thus we may assume without loss of generality that

$$\sum_{k=1}^{n} \frac{\sigma_k^2}{R\sigma_k^2 + 1} = 1. \tag{6}$$

This means that $\lambda_n = -1$ and thus

$$\prod_{k=1}^{n-1}\left(1+\frac{\lambda_k}{|\lambda_n|}(1+t)\right) = (1+t)^{n-1}\left(\prod_{k=1}^{n}\left(\frac{1}{1+t}+\lambda_k\right)\right) \Big/ \left(\frac{1}{1+t}-1\right) = \frac{(1+t)^n}{t}\left|f\left(-\frac{1}{1+t}\right)\right|,$$

hence, after the change of variables $s = -1/(1+t)$ in (5), we obtain

$$P\{A\} = \frac{1}{\pi}\int_0^1 \frac{s^{\frac{n}{2}-1}\,ds}{\sqrt{f(-s)}} = \frac{1}{\pi}\int_0^1 \frac{\sqrt{R}\,s^{\frac{n}{2}-1}\,ds}{\sqrt{\left(\sum\limits_{k=1}^{n}\frac{x_k}{x_k+s} - R\right)\prod\limits_{k=1}^{n}(x_k+s)}}, \tag{7}$$

where $x_k = R\sigma_k^2$. Condition (6) now has the form

$$\sum_{k=1}^{n} \frac{x_k}{x_k+1} = R. \tag{8}$$

First we show that $2\overline{s}_n(1) = 1/2$ for all n. By the integral representation (5), we have the inequality

$$P\left\{\xi_n^2 > \sum_{k=1}^{n-1}\lambda_k\xi_k^2\right\} \le P\left\{\xi_n^2 > \xi_1^2\sum_{k=1}^{n-1}\lambda_k\right\}.$$

On the other hand, by (4) and (8),

$$\sum_{k=1}^{n-1}\lambda_k = 1 + \left(1-\frac{1}{R}\right)\sum_{k=1}^{n} x_k \ge 1 + \left(1-\frac{1}{R}\right) n\frac{R}{n-R},$$

therefore

$$s_{n-1}(x) \le P\left\{\xi_n^2 > \xi_1^2\left(1+\frac{n(R-1)}{n-R}\right)\right\} = 1-\frac{2}{\pi}\arctan\sqrt{1+\frac{n(R-1)}{n-R}}.$$

Thus if $x = R = 1$, we obtain $2\overline{s}_n(1) \le 1/2$, and the equality can be reached by choosing $x_1 = x_2 \ne 0$, $x_3 = x_4 = \ldots = x_n = 0$.

Finally, consider the most important case $R > 1$. Assume that the supremum of (7) is attained at some finite point $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. With the notation

$$U(\mathbf{x}, s) = \frac{s^{\frac{n}{2}-1}\sqrt{R}}{\sqrt{P(\mathbf{x}, s)\prod_{k=1}^{n}(x_k+s)}}, \qquad P(\mathbf{x}, s) = \sum_{k=1}^{n}\frac{x_k}{(x_k+1)(x_k+s)},$$

we can rewrite (7) in the following way:

$$P\{A\} = \frac{1}{\pi}\int_0^1 \frac{U(\mathbf{x}, s)}{\sqrt{1-s}}\,d\,s.$$

Now fix all the x_k's except $x_i = y$ and $x_j = z$ and consider $x_j = y$ as a function of z. By (8),

$$\frac{d\,y}{d\,z} = -\frac{(y+1)^2}{(z+1)^2}. \tag{9}$$

Assuming $z > 0$, $y > 0$,

$$\frac{d\,P\{A\}}{d\,z} = -\frac{\Delta}{2\pi}\int_0^1 \frac{h\,(z, y, P(\mathbf{x}, s))}{\sqrt{1-s}}U(\mathbf{x}, s)\,d\,s, \tag{10}$$

where

$$\Delta = \frac{z-y}{(1+z)^2},$$

$$f(z, y, v) = \frac{\alpha+\beta s}{(z+s)(y+s)} - \frac{2s\left(\alpha+\frac{1+s}{2}\beta\right)}{(z+s)^2(y+s)^2 v}, \tag{11}$$

$$\alpha = yz-1, \qquad \beta = y+z+2.$$

Define a functional $L(h)$ as follows:

$$L(h) = \frac{1}{\pi}\int_0^1 \frac{h(s)}{\sqrt{1-s}}U(\mathbf{x}, s)\,d\,s.$$

Then

$$P\{A\} = L(1), \qquad \frac{d\,P\{A\}}{d\,z} = -\frac{\Delta}{2}L(h),$$

$$\frac{d\,L(h)}{d\,z} = -\frac{\Delta}{2}L(h^2) + L\left(\frac{d\,h}{d\,z}\right). \tag{12}$$

The following lemma is proved in the Appendix.

Lemma 3.

$$\frac{d\,h}{d\,z} = \Delta(h-h^2).$$

Finally, if $y > 0$, $z > 0$ at the point of maximum $\mathbf{x}$ and $z \ne y$, then, by the necessary condition for a maximum, $L(h) = 0$, and, by Lemma 3 and (12),

$$\frac{d^2P\{A\}}{d\,z^2} = \frac{3\Delta^2}{4}L(h^2) > 0,$$

which contradicts the maximality of $P\{A\}$. Thus at the point of maximum all nonzero x_k's are equal.

The only claim we have not proved is that $s_{n-1}(x) = t_{n-1}(x)$ for $x \ge \sqrt{3(n-1)/(n-3)}$. It follows from Proposition 1 (iii).

The theorem is proved.

3. Proof of Proposition 1

(i) It is easy to see that $\Delta_k(k) = g_{k+1}(k) > 0$. On the other hand, (7) implies

$$g_k(1) = \frac{1}{\pi} \int_0^1 \frac{s^{-1/2}(1-s)^{-1/2} d\,s}{\sqrt{\left(1 + \frac{1}{s(k-1)}\right)^{k-1}}},$$

where the integrand is strictly decreasing, whence $\Delta_k(1) < 0$ for all positive k. Thus $\Delta_k(R) = 0$ for at least one $R \in (1, k)$.

If we differentiate

$$\Delta_k(R) = \frac{2\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{\pi k}\Gamma\left(\frac{k}{2}\right)} \int_0^{\sqrt{\frac{Rk}{k+1-R}}} \left(1 + \frac{u^2}{k}\right)^{-\frac{k+1}{2}} d\,u - \frac{2\Gamma\left(\frac{k}{2}\right)}{\sqrt{\pi(k-1)}\Gamma\left(\frac{k-1}{2}\right)} \int_0^{\sqrt{\frac{R(k-1)}{k-R}}} \left(1 + \frac{u^2}{k-1}\right)^{-\frac{k}{2}} d\,u$$

with respect to R, we can see that $\Delta_k(R)$ cannot have more than one zero.

(ii) For small k, the monotonicity of the function $r(k)$ can be seen from computing the actual values of $r(k)$. Some approximate r-values are as follows: $r(2) = 1.726$, $r(3) = 2.040$, $r(4) = 2.226$, $r(5) = 2.352$, $r(6) = 2.442$, $r(7) = 2.510$, $r(8) = 2.568$, $r(9) = 2.607$, $r(10) = 2.642, \ldots, r(20) = 2.881$, $r(120) = 2.967$. For general k, let us rewrite the definition of $r(k)$, the equation $\Delta_k(R) = g_{k+1}(R) - g_k(R) = 0$, as follows:

$$\frac{2\Gamma\left(\frac{k}{2}\right)}{\sqrt{\pi(k-1)}\Gamma\left(\frac{k-1}{2}\right)} \int_0^{\sqrt{\frac{R(k-1)}{k-R}}} \left(1 + \frac{u^2}{k-1}\right)^{-\frac{k}{2}} d\,u = \frac{2\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{\pi k}\Gamma\left(\frac{k}{2}\right)} \int_0^{\sqrt{\frac{Rk}{k+1-R}}} \left(1 + \frac{u^2}{k}\right)^{-\frac{k+1}{2}} d\,u. \qquad (13)$$

This equation defines $r(k)$ for all real numbers $k > 1$, and we can show that $r'(k) > 0$. We omit the details of the proof, which is routine but long.

(iii) Let us now prove the most interesting part of Proposition 1. Rewrite Eq. (13) as follows:

$$M_k \int_0^p A_k(u) d\,u = \int_0^q B_k(u) d\,u, \qquad (14)$$

where

$$M_k: \ = \frac{\Gamma^2\left(\frac{k}{2}\right)\sqrt{k}}{\Gamma\left(\frac{k-1}{2}\right)\Gamma\left(\frac{k+1}{2}\right)\sqrt{k-1}},$$

$$A_k(u) := \left(1 + \frac{u^2}{k-1}\right)^{-\frac{k}{2}},$$

$$B_k(u) := \left(1 + \frac{u^2}{k}\right)^{-\frac{k+1}{2}} = A_k(u) \exp\left\{\frac{2u^2 - u^4}{4k^2} + o\left(\frac{1}{k^2}\right)\right\},$$

$$p := \sqrt{\frac{R(k-1)}{k-R}}, \qquad q := \sqrt{\frac{Rk}{k+1-R}}.$$

By Stirling's formula, as $z \to \infty$,

$$\log \Gamma(z) = -z + \left(z - \frac{1}{2}\right) \log z + \log \sqrt{2\pi} + \frac{1}{12z} + \frac{\theta}{360|z|^3}, \quad |\theta| \leq 1;$$

thus

$$M_k = \exp\left\{-\frac{1}{4k^2} + O\left(\frac{1}{k^3}\right)\right\}.$$

One can easily show that the sequence $r(k)$, $k = 1, 2, \ldots$, is bounded; thus it has a finite limit r^*.

If $R = r^* + o(1)$, we have

$$q = p + \frac{\sqrt{r^*}(1 - r^*)}{2k^2} + o\left(\frac{1}{k^2}\right);$$

thus the k^{-2} order asymptotics of (14) is

$$\frac{\sqrt{r^*}(1 - r^*)}{2} e^{-\frac{r^*}{2}} + \int_0^{\sqrt{r^*}} \frac{2u^2 - u^4 + 1}{4} e^{-\frac{u^2}{2}} d\,u = 0.$$

Integration by parts shows that $r^* = 3$.

4. Appendix

Proof of Lemma 1. It is sufficient to consider the case where $\sigma_k \neq 0$ for all k. Then $f(\lambda) = \det(\lambda I - G) = \det D^2 \det\left(\lambda D^{-2} - RI + E\right)$. Let $a_k := \lambda\sigma_k^{-2} - R$. Then

$$\det\left(\lambda D^{-2} - RI + E\right) = \det\begin{pmatrix} a_1 + 1 & 1 & \ldots & 1 \\ 1 & a_2 + 1 & \ldots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \ldots & a_n + 1 \end{pmatrix} = \det\begin{pmatrix} a_1 + 1 & 1 & 1 & \ldots & 1 \\ -a_1 & a_2 & 0 & \ldots & 0 \\ -a_1 & 0 & a_3 & \ldots & 0 \\ -a_1 & 0 & \ldots & \ldots & a_n \end{pmatrix}$$

$$= (a_1 + 1)a_2 a_3 \ldots a_n + \sum_{i=2}^{n} (-1)^{1+i}(-1)^{1+i} \frac{a_2 a_3 \ldots a_n}{a_i} = \prod_{i=1}^{n} a_i \left[1 + \frac{1}{a_1} + + \frac{1}{a_2} + \cdots + \frac{1}{a_n}\right],$$

which proves Lemma 1. $\square$

Proof of Lemma 2 [16]. Denote

$$g(x) = P\{\xi_0^2 > x\} = \frac{1}{\sqrt{2\pi}} \int_0^\infty z^{-\frac{1}{2}} e^{-\frac{z}{2}} d\,z = \frac{1}{\sqrt{2\pi}} x^{\frac{1}{2}} e^{-\frac{x}{2}} \int_0^\infty (1 + t)^{-\frac{1}{2}} e^{-\frac{tx}{2}} d\,t.$$

The latter integral is a degenerate Tricomi hypergeometric function (see [4, 6.5 (2) and (6)]):

$$\Psi(a, c, z) := \frac{1}{\Gamma(a)} \int_0^\infty t^{a-1}(1 + t)^{c-a-1} e^{-zt} d\,t, \qquad \mathrm{Re}\, a > 0, \quad \mathrm{Re}\, z > 0,$$

and

$$\Psi(a, c, z) = z^{1-c}\Psi(1 - c + a, 2 - c, z).$$

Thus

$$\int_0^\infty (1 + t)^{-\frac{1}{2}} e^{-\frac{tx}{2}} d\,t = \Psi\left(1, \frac{3}{2}, \frac{x}{2}\right) = x^{-\frac{1}{2}}\Psi\left(\frac{1}{2}, \frac{1}{2}, \frac{x}{2}\right) = \left(\frac{2}{\pi}\right)^{1/2} x^{-\frac{1}{2}} \int_0^\infty t^{-\frac{1}{2}}(1 + t)^{-1} e^{-\frac{tx}{2}} d\,t,$$

whence

$$g(x) = \frac{1}{\pi} \int_0^\infty t^{-\frac{1}{2}}(1 + t)^{-1} e^{-\frac{x(1+t)}{2}} d\,t\,.$$

For $\tau := \sum_{k=1}^n \mu_i \xi_i^2$,

$$P\left\{\xi_0^2 \geq \sum_{k=1}^n \mu_i \xi_i^2\right\} = Eg(\tau) = \int_0^\infty t^{-\frac{1}{2}}(1 + t)^{-1} E e^{-\frac{\tau(1+t)}{2}} d\,t = \frac{1}{\pi} \int_0^\infty \frac{d\,t}{\sqrt{t}(1 + t)\sqrt{\prod_{k=1}^n (1 + (1 + t)\mu_i)}}.$$

Lemma 2 is proved. $\square$

Proof of Lemma 3. It is easy to check that

$$(y+s)(z+s) = \alpha + \beta s + (1-s)^2. \tag{A1}$$

By (9), we have

$$\frac{d\,\alpha}{d\,z} = y + zy' = \Delta\alpha, \qquad \frac{d\,\beta}{d\,z} = 1 + y' = \Delta\beta,$$

$$\frac{d}{d\,z}\left((y+s)(z+s)\right) = \frac{d}{d\,z}(\alpha+\beta s) = \Delta(\alpha+\beta s);$$

thus, for the first term in h,

$$\frac{d}{d\,z}\left(\frac{\alpha+\beta s}{(y+s)(z+s)}\right) = \frac{\Delta(1-s)^2(\alpha+\beta s)}{(y+s)^2(z+s)^2}. \tag{A2}$$

With $V := P(\mathbf{x}, s)$ and $\gamma := 2s\left(\alpha + (1+s)\beta/2\right)$,

$$\frac{d\,V}{d\,z} = \frac{d}{d\,z}\frac{1}{1-s}\left(\frac{y}{y+s}+\frac{z}{z+s}\right) = \frac{d}{d\,z}\frac{1}{1-s}\frac{2\alpha+\beta s+2(1-s)}{(y+s)(z+s)} = -\frac{\Delta\gamma}{(y+s)^2(z+s)^2};$$

therefore, for the second term in h,

$$\frac{d}{d\,z}\left(\frac{\gamma}{(y+s)^2(z+s)^2V}\right) = \frac{\Delta\gamma}{(y+s)^2(z+s)^2V}\left(\frac{(1-s)^2-(\alpha+\beta s)}{(y+s)(z+s)} + \frac{\gamma}{(y+s)^2(z+s)^2V}\right). \tag{A3}$$

Finally, (A1), (A2), and (A3) imply Lemma 3. $\square$

The work of the first author was supported by the grants RFBR 05-01-97924 and INTAS 04-77-7173.

REFERENCES

1. M. S. Bartlett, "The effect of nonnormality on the t-distribution," *Proc. Camb. Phil. Soc.*, **31**, 223–231 (1935).
2. S. Basu and A. DasGupta, "Robustness of standard confidence intervals for location parameters under departure from normality," *Ann. Statist.*, **23**, 1433–1442 (1995).
3. Y. Benjamini, "Is the t-test really conservative when the parent distribution is long-tailed?," *J. Amer. Statist. Assoc.*, **78**, 645–654 (1983).
4. H. Bateman and A. Erdelyi, *Higher Transcendental Functions*, Vol. 1, McGraw-Hill, New York (1953).
5. N. Cressie, "Relaxing assumption in the one-sided t-test," *Austral. J. Statist.*, **22**, 143–153 (1980).
6. B. Efron, "Student's t-test under symmetry conditions," *J. Amer. Statist. Assoc.*, **64**, 1278–1302 (1969).
7. B. Efron and R. A. Olshen, "How broad is the class of normal scale mixtures?," *Ann. Statist.*, **6**, No. 5, 1159–1164 (1978).
8. C. Eisenhart, "On the transition of "Student's" z to "Student's" t," *Amer. Statist.*, **33**, No. 1, 6–10 (1979).
9. W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. II, Wiley, New York–London–Sydney (1966).
10. R. A. Fisher, "Applications of "Student's" distribution," *Metron*, **5**, 90–104 (1925).
11. T. Gneiting, "Normal scale mixtures and dual probability densities," *J. Stat. Comput. Simul.*, **59**, 375–384 (1997).
12. H. Hotelling, "The behavior of some standard statistical tests under nonstandard conditions," in: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, University of California Press, Berkeley (1961), pp. 319–360.
13. D. Kelker, "Infinite divisibility and variance mixtures of the normal distribution," *Ann. Math. Statist.*, **42**, No. 2, 802–808 (1971).
14. J. Landerman, "The distribution of "Student's" ratio for samples of two items drawn from non-normal universes," *Ann. Math. Statist.*, **10**, 376–379 (1939).
15. A. F. S. Lee and J. Gurland, "One-sample t-test when sampling from a mixture of normal distributions," *Ann. Statist.*, **5**, No. 4, 803–807 (1977).

16. A. V. Makshanov and O. V. Shalaevsky, "Some problems of asymptotic theory of distributions," *Zap. Nauchn. Semin. LOMI*, **74**, 118–138 (1977).
17. E. S. Pearson, "The distribution of frequency constants in small samples from nonnormal symmetrical and skew populations," *Biometrika*, **21**, 259–286 (1929).
18. P. Prescott, "A simple alternative to Student's t," *Appl. Statist.*, **24**, 210–217 (1975).
19. P. R. Rider, "On the distribution of the ratio of mean to standard deviation in small samples from nonnormal universes," *Biometrika*, **21**, 124–143 (1929).
20. P. R. Rider, "On small samples from certain nonnormal universes," *Ann. Math. Statist.*, **2**, 48–65 (1931).
21. P. R. Rietz, "On the distribution of the "Student" ratio for small samples from certain nonnormal populations," *Ann. Math. Statist.*, **10**, 265–274 (1939).
22. Student, "The probable error of a mean," *Biometrika*, **6**, 1–25 (1908).
23. G. J. Székely, *Paradoxes in Probability Theory and Mathematical Statistics*, Reidel, Dordrecht (1986).
24. J. W. Tukey and D. H. McLaughin, "Less vulnerable confidence and significance procedures for location based on a single sample: Trimming/Winsorization," *Sankhyā, Ser. A*, **25**, 331–352 (1963).