Bernoulli **17**(1), 2011, 276–289
DOI: 10.3150/10-BEJ262

On the heavy-tailedness of Student's t-statistic

FREDRIK JONSSON

Department of Mathematics, Uppsala University, Box 480, SE-751 06 Uppsala, Sweden.
E-mail: jonsson@math.uu.se

Let $\{X_i\}_{i\geq 1}$ be an i.i.d. sequence of random variables and define, for $n \geq 2$,

$$T_n = \begin{cases} n^{-1/2}\hat{\sigma}_n^{-1} S_n, & \hat{\sigma}_n > 0, \\ 0, & \hat{\sigma}_n = 0, \end{cases} \qquad \text{with } S_n = \sum_{i=1}^{n} X_i,\ \hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - n^{-1}S_n)^2.$$

We investigate the connection between the distribution of an observation X_i and finiteness of $\mathrm{E}|T_n|^r$ for $(n,r) \in \mathbb{N}_{\geq 2} \times \mathbb{R}^+$. Moreover, assuming $T_n \xrightarrow{d} T$, we prove that for any $r > 0$, $\lim_{n\to\infty} \mathrm{E}|T_n|^r = \mathrm{E}|T|^r < \infty$, provided there is an integer n_0 such that $\mathrm{E}|T_{n_0}|^r$ is finite.

Keywords: finiteness of moments; robustness; Student's t-statistic; t-distributions; t-test

1. Introduction

Assume, in the following, that $\{X_i\}_{i\geq 1}$ is a sequence of independent random variables, each with distribution F. Then, for $n \geq 2$, define the *t-statistic* random variables

$$T_n = \begin{cases} n^{-1/2}\hat{\sigma}_n^{-1} S_n, & \hat{\sigma}_n > 0, \\ 0, & \hat{\sigma}_n = 0, \end{cases} \qquad \text{with } S_n = \sum_{i=1}^{n} X_i,\ \hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - n^{-1}S_n)^2.$$

In the case where F is a normal distribution with mean zero, the distribution of T_n is the well-known *t-distribution with $n-1$ degrees of freedom*. The effect of non-normality of F on the distribution of T_n has received considerable attention in the statistical literature. For a review, see [7]. t-distributions do not only occur in the inference of means, but also sometimes in models of data in the economic sciences; see [6]. There seem to be two characteristic properties which, in comparison with the normal distribution, make these distributions convenient in certain modeling situations: a higher degree of heavy-tailedness (moments are finite only below the degree of freedom) and a higher degree of so-called kurtosis.

This paper investigates the tail behaviour of T_n and the related issue of the existence of moments $\mathrm{E}|T_n|^r$, for a parameter $r > 0$, under more general conditions than the normal assumption. Motivating questions were the following: Is it generally true that $\mathrm{E}|T_n|^r$ can only be finite for $r < n-1$? For which kinds of distributions is the converse implication false? Assuming the often encountered $T_n \xrightarrow{d} T$, is it then generally true that $\mathrm{E}|T_n|^r \to \mathrm{E}|T|^r$?

1350-7265

2. Summary

The fundamental result is Theorem 3.1, which presents two conditions, each equivalent to finiteness of $\mathrm{E}|T_n|^r$. The result is based on a connection between the tail behaviour of T_n and probabilities of having almost identical observations $X_1, \dots, X_n$. Theorem 4.1 states that finiteness of $\mathrm{E}|T_n|^r$ implies finiteness of $\mathrm{E}|T_{n+1}|^r$, and is followed by Theorem 4.2 which states that t-statistic random variables never possess moments above the degree of freedom unless F is discrete. It is established in Section 5, under the assumption that F is continuous, that *regularity*, referring to the degree of heavy-tailedness of t-statistic random variables, is measurable in terms of the behaviour of certain *concentration functions* related to F. Theorem 6.2 states that $\lim_{n\to\infty} \mathrm{E}|T_n|^r = \mathrm{E}|T|^r$ whenever there is an integer n_0 such that $\mathrm{E}|T_{n_0}|^r$ is finite and $\{T_n\}$ converges in distribution.

Remark. This paper is an abridged version of [5]. The results found in Section 5 here are there generalized beyond the continuity assumption. We also refer to [5] for a discussion of related results previously obtained by H. Hotelling.

3. Characterizing $\mathrm{E}|T_n|^r < \infty$ through bounds on $\mathrm{P}(|T_n| > x)$

A close connection exists between T_n and the *self-normalized sum* S_n/V_n; see Lemma 3.1 (whose elementary proof we omit). The connection allows $\mathrm{E}|T_n|^r$ to be expressed with probabilities relating to S_n/V_n, as in Lemma 3.2, revealing that finiteness of $\mathrm{E}|T_n|^r$ depends on the magnitude of the probabilities of having S_n/V_n close to $\pm\sqrt{n}$. Some geometric relations between S_n/V_n close to $\pm\sqrt{n}$ and almost identical observations $X_1, \dots, X_n$ are then given in Lemmas 3.3 and 3.4.

Lemma 3.1. *Define*

$$V_n = \left(\sum_{i=1}^n X_i^2\right)^{1/2}, \qquad U_n^* = \begin{cases} 0, & S_n/V_n = n \text{ or } V_n = 0, \\ (S_n/V_n)^2, & \text{otherwise.} \end{cases}$$

It then holds, for any $x \ge 0$, that $T_n^2 > x$ if and only if $U_n^* > nx/(n + x - 1)$.

Lemma 3.2. *For $r > 0$ and U_n^* as in Lemma 3.1,*

$$\mathrm{E}|T_n|^r = \frac{r}{2} n(n-1)^{r/2} \int_0^n z^{r/2-1} \mathrm{P}(U_n^* > z)(n-z)^{-(r/2+1)}\,\mathrm{d}z.$$

Lemma 3.3. *Let $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ and $h \in (0, 1)$ be given such that $x_1 \neq 0$ and $n - u_n < h^2$ with $u_n = (\sum_{i=1}^n x_i)^2 / \sum_{i=1}^n x_i^2$. Then, with $C_1 = \sqrt{5}$,*

$$|x_i - x_1| < hC_1|x_1| \qquad \text{for all } i \neq 1.$$

Moreover, $C_1 = C_1(n, h) = \sqrt{2 + 2h + h^2}$ is optimal for the conclusion to be valid for all $\mathbf{x}$.

Lemma 3.4. *Let* $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ *and* $h \in (0, 1)$ *be given such that, with* $C_2 = 1$,

$$|x_i - x_1| < C_2 h |x_1| / \sqrt{n-1} \qquad \textit{for all } i \neq 1.$$

Then $n - u_n < h^2$ *with* $u_n = (\sum_{i=1}^n x_i)^2 / \sum_{i=1}^n x_i^2$. *Moreover, in the case where n is odd,* $C_2 = C_2(n, h)$ *must satisfy* $C_2 \le \sqrt{n/(n-h^2)}$ *for the conclusion to be valid for all* $\mathbf{x}$.

Theorem 3.1. *The following three quantities are either all finite or all infinite*:

(i) $\mathrm{E}|T_n|^r$;

(ii) $\mathrm{E}\left(|X_1|^r \bigwedge_{i=2}^{n} |X_i - X_1|^{-r} I\{|X_i - X_1| > 0, \text{ some } i \le n\} \right)$;

(iii) $\int_{x \neq 0} \int_0^1 h^{-(r+1)} \big(\big(\mathrm{P}(|X - x| < h|x|)\big)^{n-1} - p_x^{n-1} \big) \, \mathrm{d}h \, \mathrm{d}F(x) \qquad$ *with* $p_x = \mathrm{P}(X = x)$.

Proof of Lemma 3.2. By [4], Theorem 12.1, Chapter 2, together with Lemma 3.1 and a change of variables, we have

$$\begin{aligned}
\mathrm{E}|T_n|^r &= \frac{r}{2} \int_0^\infty y^{r/2-1} \mathrm{P}(T_n^2 > y) \, \mathrm{d}y \\
&= \frac{r}{2} \int_0^\infty y^{r/2-1} \mathrm{P}\big(U_n^* > ny/(n+y-1)\big) \, \mathrm{d}y \\
&= \frac{r}{2} n (n-1)^{r/2} \int_0^n z^{r/2-1} \mathrm{P}(U_n^* > z)(n-z)^{-(r/2+1)} \, \mathrm{d}z.
\end{aligned}$$

□

Proof of Lemma 3.3. We argue by contraposition. Due to the invariance with respect to scaling of $\mathbf{x}$ and permutation of the coordinates $x_2, \dots, x_n$, it suffices to prove that

$$|x_2 - x_1| \ge h|x_1| \quad \Longrightarrow \quad n - u_n \ge h^2 / C_1^2$$

with $C_1 = \sqrt{2 + 2h + h^2}$ and that equalities are simultaneously attained. Set $x_2 = x_1 + \varepsilon$ and $\underline{x} = (x_3, \dots, x_n)$. We then minimize $n - u_n$ with respect to $\underline{x}$ and ε. Note that

$$\frac{\partial (n - u_n)}{\partial x_j} = \frac{-2 \sum_{i=1}^n x_i (\sum_{i=1}^n x_i^2 - x_j \sum_{i=1}^n x_i)}{(\sum_{i=1}^n x_i^2)^2}. \tag{1}$$

First, set (1) to zero for $j = 3, \dots, n$. Since $\sum x_i = 0$ corresponds to $u_n = 0$, which is non-interesting with respect to the minimization of $n - u_n$, these equations reduce to

$$\sum_{i=3}^n x_i^2 - x_j \sum_{i=3}^n x_i = x_j(x_1 + x_2) - (x_1^2 + x_2^2) \qquad \text{for } j = 3, \dots, n. \tag{2}$$

We claim that (2) has the unique solution

$$x_j = (x_1^2 + x_2^2)/(x_1 + x_2) = (2x_1^2 + 2x_1\varepsilon + \varepsilon^2)/(2x_1 + \varepsilon) \qquad \text{for } j = 3, \ldots, n. \tag{3}$$

To verify this, assume that $\underline{x}$ is a solution of (2). Since $\sum_{i=3}^n x_i^2$ and $\sum_{i=3}^n x_i$ do not vary with j, $\underline{x}$ must be of the form $x_j = const.$, $j = 3, \ldots, n$. However, the left-hand side of (2) then vanishes for all j, which gives (3) as the unique solution. Inserting the solution into $n - u_n$ gives

$$(n - u_n)_{\min}(\varepsilon) = \varepsilon^2/(x_1^2 + x_2^2) = \varepsilon^2/(2x_1^2 + 2x_1\varepsilon + \varepsilon^2). \tag{4}$$

It remains to minimize with respect to ε with $\varepsilon \notin (-h|x_1|, h|x_1|)$. The equation

$$\frac{\partial}{\partial\varepsilon}\left(\frac{\varepsilon^2}{2x_1^2 + 2x_1\varepsilon + \varepsilon^2}\right) = 0$$

has the unique solution $\varepsilon = -2x_1$ which cannot be a minimum since a minimum must satisfy $\operatorname{sign}(\varepsilon) = \operatorname{sign}(x_1)$, by the representation (4). The solution is hence obtained for $\varepsilon = \operatorname{sign}(x_1)h|x_1|$,

$$(n - u_n)_{\min} = (hx_1)^2/\big(x_1^2(2 + 2h + h^2)\big) = h^2/(2 + 2h + h^2).$$

It follows that $C_1 = C_1(h) = \sqrt{2 + 2h + h^2} \le \sqrt{5}$ is an optimal constant, as claimed. □

Proof of Lemma 3.4. Assume that

$$|x_i - x_1| < C_2 h|x_1|/\sqrt{n-1} \qquad \text{for all } i = 2, \ldots, n. \tag{5}$$

The aim is to verify that $n - u_n < h^2$ with $C_2 = C_2(n, h)$ optimally large. We therefore maximize $n - u_n$ over the rectangular region (5) with $x_1 \neq 0$, C_2 and h fixed. It suffices to consider the restriction of $n - u_n$ to the corners of the region (5) since the maximum attained at a point $y = (y_1, \ldots, y_n)$ in the interior of the region, or in the interior of an edge, would mean that, for some $j = 2, \ldots, n$ and some $\eta > 0$,

$$\frac{\partial(n - u_n)}{\partial x_j}(y) = 0, \tag{6}$$

$$\frac{\partial(n - u_n)}{\partial x_j}(y_1, \ldots, y_{j-1}, y_j - h, y_{j+1}, \ldots, y_n) \ge 0 \qquad \text{for all } 0 < h < \eta, \tag{7}$$

$$\frac{\partial(n - u_n)}{\partial x_j}(y_1, \ldots, y_{j-1}, y_j + h, y_{j+1}, \ldots, y_n) \le 0 \qquad \text{for all } 0 < h < \eta. \tag{8}$$

Recall, from the proof of Lemma 3.3, that

$$\frac{\partial(n - u_n)}{\partial x_j} = \frac{-2\sum_{i=1}^n x_i(\sum_{i \neq j} x_i^2 - x_j \sum_{i \neq j} x_i)}{(\sum_{i=1}^n x_i^2)^2}.$$

We may assume that $C_2 h < \sqrt{n-1}$ since the point $x_i \equiv 0$ would otherwise belong to the region yielding $u_n = 1$, in which case $n - u_n < h^2$ cannot hold. This implies that $\operatorname{sign}(x_i) = \operatorname{sign}(x_1)$ for all $i = 2, \dots, n$ so that neither $\sum x_i$ nor $\sum_{i \neq j} x_i$ change sign within the region. Assume, due to invariance with respect to scaling, that $x_1 > 0$. Conditions (6)–(8) may then be reformulated as

$$\sum_{i \neq j} y_i^2 - y_j \sum_{i \neq j} y_i = 0, \qquad \sum_{i \neq j} y_i^2 - (y_j - h) \sum_{i \neq j} y_i < 0, \qquad \sum_{i \neq j} y_i^2 - (y_j + h) \sum_{i \neq j} y_i > 0,$$

which is contradictory since $h > 0$ and $\sum_{i \neq j} y_i > 0$.

Now, consider the restriction of $n - u_n$ to the corners of the region (5). Set $k := |\{i : x_i = x_1 + \varepsilon\}| - |\{i : x_i = x_1 - \varepsilon\}|$ so that

$$\begin{aligned} n - u_n &= \frac{n(nx_1^2 + (n-1)\varepsilon^2 + 2k\varepsilon x_1) - (nx_1 + k\varepsilon)^2}{nx_1^2 + (n-1)\varepsilon^2 + 2k\varepsilon x_1} \\ &= \frac{\varepsilon^2(n(n-1) - k^2)}{nx_1^2 + (n-1)\varepsilon^2 + 2k\varepsilon x_1} = \frac{h^2 C_2^2 (n - k^2/(n-1))}{n + C_2^2 h^2 + 2kC_2 h/\sqrt{n-1}}. \end{aligned} \tag{9}$$

Take $C_2 = 1$ in (9) and $z = k(n-1)^{-1/2}$. Algebraic manipulations yield

$$\frac{n - k^2/(n-1)}{n + h^2 + 2kh/\sqrt{n-1}} \leq 1 \quad \Longleftrightarrow \quad (h + z)^2 \geq 0$$

so that $C_2 = 1$ is sufficiently small for the desired bound $n - u_n < h^2$. We find, by taking $k = 0$ in (9) (which is possible when n is odd) that

$$C_2^2 n/(n + C_2^2 h^2) \leq 1 \quad \Longleftrightarrow \quad C_2^2 \leq n/(n - h^2)$$

so that $C_2 \leq \sqrt{n/(n-h^2)}$ is then necessary for $n - u_n < h^2$ to hold. □

Proof of Theorem 3.1. We first deduce the equivalence between (i) and (iii). By Lemma 3.2, we find that $E|T_n|^r < \infty$ is equivalent to, for some $\delta < 1$,

$$\int_{n-\delta}^{n} z^{r/2-1} \mathrm{P}(U_n^* > z)(n-z)^{-(r/2+1)} \, \mathrm{d}z < \infty \quad \Longleftrightarrow \quad \int_0^{\delta} h^{-(r+1)} \mathrm{P}(n - U_n^* < h^2) \, \mathrm{d}h < \infty,$$

which, in turn, is equivalent to

$$\int \int_0^{\delta} h^{-(r+1)} \mathrm{P}(0 < n - U_n < h^2 \mid X_1 = x) \, \mathrm{d}h \, \mathrm{d}F(x) < \infty. \tag{10}$$

The event $X_1 = 0$ implies $U_n \leq n - 1$ by the Cauchy–Schwarz inequality so that (10) reduces to

$$\int_{x \neq 0} \int_0^{\delta} h^{-(r+1)} \mathrm{P}(0 < n - U_n < h^2 \mid X_1 = x) \, \mathrm{d}h \, \mathrm{d}F(x) < \infty,$$

which is equivalent to

$$\int_{x\neq 0}\int_0^{\delta} h^{-(r+1)}\mathrm{P}(n-U_n<h^2 \mid X_1=x)-p_x^{n-1}\,\mathrm{d}h\,\mathrm{d}F(x)<\infty$$

since $U_n=n$ corresponds to $X_i=X_1$ with $p_x=\mathrm{P}(X=x)$. Finally, apply Lemmas 3.3 and 3.4, and set $\delta=1$ to arrive at condition (iii).

For the equivalence between (ii) and (iii), define $A_n=\{|X_i-X_1|>0, \text{ some } i\le n\}$. Condition on X_1 and convert expectation into integration of tail probabilities (cf. [4], Theorem 12.1, Chapter 2):

$$\begin{aligned}\mathrm{E}\left(|X_1|^r\bigwedge_{i=2}^{n}|X_i-X_1|^{-r}I_{A_n}\right)&=\int_{x\neq 0}\mathrm{E}\left(\bigwedge_{i=2}^{n}(|X_i-x||x|^{-1})^{-r}I_{A_n}\right)\mathrm{d}F(x)\\&=r\int_{x\neq 0}\int_0^{\infty}h^{-(r+1)}\big(\mathrm{P}(|X-x|<h|x|)\big)^{n-1}-p_x^{n-1}\,\mathrm{d}h\,\mathrm{d}F(x).\end{aligned}$$

The equivalence between (ii) and (iii) then follows from the fact that

$$\begin{aligned}&\int_{x\neq 0}\int_1^{\infty}h^{-(r+1)}\big(\mathrm{P}(|X-x|<h|x|)\big)^{n-1}\,\mathrm{d}h\,\mathrm{d}F(x)\\&\quad\le\int_{x\neq 0}\int_1^{\infty}h^{-(r+1)}\,\mathrm{d}h\,\mathrm{d}F(x)<\infty.\end{aligned}$$

□

4. Two general facts regarding finiteness of $\mathrm{E}|T_n|^r$

Theorem 4.1. *For any couple* $(n,r)\in\mathbb{N}_{\ge 2}\times\mathbb{R}^+$, *if* $\mathrm{E}|T_n|^r$ *is finite, then so is* $\mathrm{E}|T_{n+1}|^r$.

Proof. Due to Theorem 3.1, it suffices to show that

$$\mathrm{E}\left[|X_1|^r\bigwedge_{i=2}^{n}|X_i-X_1|^{-r}I_{A_n}\right]<\infty\quad\Longrightarrow\quad\mathrm{E}\left[|X_1|^r\bigwedge_{i=2}^{n+1}|X_i-X_1|^{-r}I_{A_{n+1}}\right]<\infty,\qquad(11)$$

where $A_k:=\{|X_i-X_1|>0, \text{ some } i\le k\}$. Define $A_n'=\{|X_i-X_1|>0, \text{ some } 3\le i\le n+1\}$. It follows that $A_{n+1}=A_n\cup A_n'$ so that $I_{A_{n+1}}\le I_{A_n}+I_{A_n'}$, which gives

$$\begin{aligned}&\mathrm{E}\left[|X_1|^r\bigwedge_{i=2}^{n+1}|X_i-X_1|^{-r}I_{A_{n+1}}\right]\\&\quad\le\mathrm{E}\left[|X_1|^r\bigwedge_{i=2}^{n+1}|X_i-X_1|^{-r}I_{A_n}\right]+\mathrm{E}\left[|X_1|^r\bigwedge_{i=2}^{n+1}|X_i-X_1|^{-r}I_{A_n'}\right]\end{aligned}$$

$$\leq \mathrm{E}\left[|X_1|^r \bigwedge_{i=2}^{n} |X_i - X_1|^{-r} I_{A_n}\right] + \mathrm{E}\left[|X_1|^r \bigwedge_{i=3}^{n+1} |X_i - X_1|^{-r} I_{A'_n}\right]$$

$$= 2\mathrm{E}\left[|X_1|^r \bigwedge_{i=2}^{n} |X_i - X_1|^{-r} I_{A_n}\right].$$

The conclusion follows. □

Theorem 4.2. *Assume that F decomposes into $F_d + F_c$, with discrete and continuous measures F_d and F_c, respectively, and that $F_c \not\equiv 0$. It is then necessary that $r < n - 1$ for $\mathrm{E}|T_n|^r$ to be finite.*

Proof. Let F_c have total mass $\varepsilon > 0$. It suffices to verify that $\mathrm{E}|T_n|^{n-1}$ is infinite, which, by Theorem 3.1, is equivalent to

$$\int_{x \neq 0} \int_0^1 h^{-n} \left(\left(\mathrm{P}(|X - x| < h|x|) \right)^{n-1} - p_x^{n-1} \right) \mathrm{d}h \, \mathrm{d}F(x) = \infty.$$

The last identity is a consequence of

$$\int \int_0^1 h^{-n} \left(\mathrm{P}(|X - x| < h|x|) \right)^{n-1} \mathrm{d}h \, \mathrm{d}F_c(x) = \infty. \tag{12}$$

To verify (12), consider the restriction of F_c to a set $[-C, -1/C] \cup [1/C, C]$ with C sufficiently large so that the restricted measure still has positive mass. It then suffices to establish the condition

$$\int \left(\mathrm{P}(|X - x| < h) h^{-1} \right)^{n-1} \mathrm{d}F_c(x) > \eta_n \qquad \text{for all } h \text{ and some constant } \eta_n = \eta_n(F_c, n). \tag{13}$$

First, consider $n = 2$. Discretize $[-C, C]$ uniformly with interval length h, that is, put $x_k = hk$ for $k \in [-N, N]$ and $N = \lceil Ch^{-1} \rceil$. Then

$$\begin{aligned}
\int \mathrm{P}(|X_c - x| < h) \, \mathrm{d}F_c(x) &= \sum_{k=-N}^{k=N} \int_{x_{k-1}}^{x_k} \mathrm{P}(|X_c - x| < h) \, \mathrm{d}F_c(x) \\
&\geq \sum_{k=-N}^{k=N} \int_{x_{k-1}}^{x_k} \mathrm{P}\left(X_c \in (x_{k-1}, x_k]\right) \mathrm{d}F_c(x) \\
&= \sum_{k=-N}^{k=N} \left(\mathrm{P}\left(X_c \in (x_{k-1}, x_k]\right) \right)^2.
\end{aligned}$$

Applying the Cauchy–Schwarz inequality, we obtain

$$\sum_{k=-N}^{k=N} \big(\mathrm{P}(X_c \in (x_{k-1}, x_k])\big)^2 \ge \left(\sum_{k=-N}^{k=N} \mathrm{P}\big(X_c \in (x_{k-1}, x_k]\big)\right)^2 (2N)^{-1} = \varepsilon^2 (2N)^{-1} \ge C^{-1}\varepsilon^2 h.$$

Conclusion (13) follows with $\eta_2 = C^{-1}\varepsilon^2$. For $n > 2$, an application of the Hölder inequality yields

$$\eta_2^{n-1} \le \left(\int \mathrm{P}(|X_c - x| < h)h^{-1}\,\mathrm{d}F_c(x)\right)^{n-1} \le \varepsilon^{n-2}\int \big(\mathrm{P}(|X_c - x| < h)h^{-1}\big)^{n-1}\,\mathrm{d}F_c(x).$$

The desired conclusion (13) follows with $\eta_n = \eta_2^{n-1}\varepsilon^{2-n}$. □

5. Regularity and concentration functions

Definition 5.1. *Given the distribution of a random variable X, define the concentration functions q and Q, for real-valued arguments $h \ge 0$, by*

$$Q(h) = \sup_x \mathrm{P}(|X - x| \le h), \qquad q(h) = \sup_x \mathrm{P}(|X - x| \le |x|h).$$

Q is known as the *Lévy concentration function*. Theorem 5.1 below characterizes finiteness of $\mathrm{E}|T_n|^r$ in terms of the limiting behaviour of $q(h)$ as h tends to zero. Note that a statement of the kind "$Q(h) = \mathcal{O}(h^\lambda)$" (for some $\lambda \le 1$) refers to the *local* behaviour of the distribution. The most regular behaviour in this respect is that of an absolutely continuous distribution with bounded density function, in which case $Q(h) = \mathcal{O}(h)$, while $\lambda < 1$ typically corresponds to one or several "explosions" of the density function. The Cantor distributions also form fundamental examples of such irregularity (cf. [5], pages 29–31). The parameter λ has, in this sense, a meaning of "degree of irregularity" concerning the distribution, with smaller values of λ indicating higher degrees of irregularity. A statement $q(h) = \mathcal{O}(h^\lambda)$, on the other hand, also has a *global* component. It requires more regularity of the distribution "at infinity" compared with $Q(h) = \mathcal{O}(h^\lambda)$, while, at the same time, being less restrictive regarding the local behaviour of the distribution at the origin.

Theorem 5.1. *The following two implications hold for any continuous probability measure F:*

(i) $q(h) = \mathcal{O}(h^\lambda)$ *for some* $\lambda > r/(n-1) \Longrightarrow \mathrm{E}|T_n|^r < \infty$;

(ii) $\mathrm{E}|T_n|^r < \infty \Longrightarrow q(h) = \mathcal{O}(h^\lambda)$ *with* $\lambda = r/n$.

A simple criterion guaranteeing the optimal $q(h) = \mathcal{O}(h)$ is given by the following proposition.

Proposition 5.1. *The property $q(h) = \mathcal{O}(h)$ is obtained for any absolutely continuous distribution F with bounded density function f satisfying the assumption of a positive constant N such that*

$$f(x_2) \le f(x_1) \qquad \text{for any } x_1, x_2 \text{ such that } N \le x_1 \le x_2 \text{ or } -N \ge x_1 \ge x_2. \tag{14}$$

Proof of Theorem 5.1. For (i), condition (iii) of Theorem 3.1 reads, by continuity,

$$\int_{x\neq 0}\int_0^1 h^{-(r+1)}\big(\mathrm{P}(|X-x|<h|x|)\big)^{n-1}\,\mathrm{d}h\,\mathrm{d}F(x)<\infty. \tag{15}$$

Applying the assumption on q to the integrand yields

$$\begin{aligned}
&\int_{x\neq 0}\int_0^1 h^{-(r+1)}\big(\mathrm{P}(|X-x|<h|x|)\big)^{n-1}\,\mathrm{d}h\,\mathrm{d}F(x)\\
&\le C\int_{x\neq 0}\int_0^1 h^{-(r+1)}h^{\lambda(n-1)}\,\mathrm{d}h\,\mathrm{d}F(x)=C\int_0^1 h^{-(r+1)}h^{\lambda(n-1)}\,\mathrm{d}h\\
&=C/\big(\lambda(n-1)-r\big),
\end{aligned}$$

which proves (15). To verify the second implication, we argue by contraposition. Assume that

$$q(h)\neq\mathcal{O}(h^{\lambda})\qquad\text{with }\lambda=r/n. \tag{16}$$

It suffices, by condition (ii) of Theorem 3.1 and the assumption of continuity, to prove that

$$\mathrm{E}\Bigg(|X_1|^r\bigwedge_{i=2}^{n}|X_i-X_1|^{-r}\Bigg)=\infty. \tag{17}$$

Statement (16) is equivalent to the existence of sequences $\{x_k\}_{k\ge 1}$ and $\{h_k\}_{k\ge 1}$ such that

$$1/2>h_k>0,\qquad \lim_{k\to\infty}h_k=0,\qquad \lim_{k\to\infty}h_k^{-r/n}\mathrm{P}(|X-x_k|\le|x_k|h_k)=\infty. \tag{18}$$

Define intervals $I_k=(x_k-|x_k|h_k,x_k+|x_k|h_k)$. It then follows that for some K and all $k\ge K$,

$$\begin{aligned}
\mathrm{E}\Bigg(|X_1|^r\bigwedge_{i=2}^{n}|X_i-X_1|^{-r}\Bigg)&\ge \mathrm{E}\Bigg(|X_1|^r\bigwedge_{i=2}^{n}|X_i-X_1|^{-r}I\{X_i\in I_k,\ \text{all } i\}\Bigg)\\
&\ge 2^{-1}|x_k|^r\mathrm{E}\Bigg(\bigwedge_{i=2}^{n}|X_i-X_1|^{-r}I\{X_i\in I_k,\ \text{all } i\}\Bigg)\\
&\ge 2^{-(r+1)}|x_k|^r h_k^{-r}|x_k|^{-r}\mathrm{E}(I\{X_i\in I_k,\ \text{all } i\})\\
&=2^{-(r+1)}h_k^{-r}\big(\mathrm{P}(|X-x_k|\le|x_k|h_k)\big)^n.
\end{aligned}$$

We conclude from (18) that (17) holds. □

Proof of Proposition 5.1. It follows that, for $x>N$,

$$f(x)(x-N)\le\int_N^x f(y)\,\mathrm{d}y\le 1,\qquad f(-x)(x-N)\le\int_{-x}^{-N}f(y)\,\mathrm{d}y\le 1,$$

so that $f(x)|x| \le C$. Consequently, assuming that $x > 2N$ and $h \le 1/2$, we have

$$\mathrm{P}(|X - x| \le |x|h) = \int_{|x|(1-h)}^{|x|(1+h)} f(y)\,\mathrm{d}y \le \frac{2C}{|x|} \int_{|x|(1-h)}^{|x|(1+h)} \mathrm{d}y = 4Ch. \tag{19}$$

Regarding $0 \le x \le 2N$, we use the fact that f is bounded, $f \le M$, so that

$$\mathrm{P}(|X - x| \le |x|h) = \int_{|x|(1-h)}^{|x|(1+h)} f(y)\,\mathrm{d}y \le M \int_{2N(1-h)}^{2N(1+h)} \mathrm{d}y = 4MNh. \tag{20}$$

Bounds analogous to (19) and (20) follow for negative x, which proves that $q(h) = \mathcal{O}(h)$. □

6. Convergence

Convergence in distribution of $\{T_n\}$ to a random variable T (e.g., standard normally distributed) is, due to Lemma 3.2, equivalent to convergence of $\{S_n/V_n\}$ to T. A complete classification in terms of possible limit distributions with corresponding conditions on F was given recently by Chistyakov and Götze (see [1]). The following interesting property was derived somewhat earlier by Giné, Götze and Mason in [3].

Theorem 6.1. *Let a distribution F be given such that $S_n/V_n \to^d T$. The sequence $\{S_n/V_n\}$ is then sub-Gaussian, in the sense that, for some constant C,* $\sup_n \mathrm{E}[\exp(tS_n/V_n)] \le 2\exp(Ct^2)$.

Corollary 6.1. *For any F satisfying the condition of Theorem 6.1 with respect to a random variable T and any $r > 0$,* $\lim_{n\to\infty} \mathrm{E}|S_n/V_n|^r = \mathrm{E}|T|^r < \infty$.

Proof. The result follows from Theorem 6.1 and general properties of integration; see, for example, [4], Theorem 5.9, Chapter 5, or [4], Corollary 4.1, Chapter 5. □

We are now ready for the main result of this section.

Theorem 6.2. *Let F, T and r be given as in Corollary 6.1. If $\mathrm{E}|T_{n_0}|^r$ is finite for some $n_0 \ge 2$, then* $\lim_{n\to\infty} \mathrm{E}|T_n|^r = \mathrm{E}|T|^r$.

Proof. The case "$X = constant$", which leads to $T_n \equiv 0$, is degenerate and is henceforth excluded. Recall, from Lemma 3.2, that

$$\mathrm{E}|T_n|^r = \frac{r}{2} n(n-1)^{r/2} \int_0^n z^{r/2-1} \mathrm{P}(U_n^* > z)(n-z)^{-(r/2+1)}\,\mathrm{d}z.$$

We split the desired conclusion $\lim_{n\to\infty} \mathrm{E}|T_n|^r = \mathrm{E}|T|^r$ into the two conditions

$$\lim_{n\to\infty} \frac{r}{2} n^{r/2+1} \int_0^{n-\delta} z^{r/2-1} \mathrm{P}(U_n^* > z)(n-z)^{-(r/2+1)}\,\mathrm{d}z = \mathrm{E}|T|^r \qquad \text{for any } 0 < \delta < 1, \tag{21}$$

$$\lim_{n\to\infty} n^r \int_{n-\delta}^n \mathrm{P}(U_n^* > z)(n-z)^{-(r/2+1)}\,\mathrm{d}z = 0 \qquad \text{for some } 0 < \delta < 1. \tag{22}$$

Replace (22), via a change of variables $n - z = h^2$, by the condition

$$\lim_{n\to\infty} n^r \int_0^\delta h^{-(r+1)} \mathrm{P}(n - U_n^* < h^2)\,\mathrm{d}z = 0 \qquad \text{for some } 0 < \delta < 1,$$

which, in turn, by the same steps as in the proof of Theorem 3.1, we find to be equivalent to

$$\begin{aligned}&\lim_{n\to\infty} R_{n,\delta} = 0,\\ &R_{n,\delta} := \int_{x\neq 0}\int_0^\delta n^r h^{-(r+1)}\big(\big(\mathrm{P}(|X-x| < h|x|)\big)^{n-1} - p_x^{n-1}\big)\,\mathrm{d}h\,\mathrm{d}F(x)\end{aligned} \tag{23}$$

for some $0 < \delta < 1$ (with $p_x = \mathrm{P}(X = x)$). We separate the verifications of (21) and (23) into Lemmas 6.2 and 6.1, respectively. Note that the assumption $\mathrm{E}|T_{n_0}|^r < \infty$, via Theorems 3.1 and 4.1, implies that $R_{n,\varepsilon} < \infty$ for all $(n, \varepsilon) \in \mathbb{N}_{\geq n_0} \times \mathbb{R}^+$. The proof of Theorem 6.2 is hence completed by applying Lemmas 6.1 and 6.2. □

Lemma 6.1. *Assume that there exists $n_0 \geq 2$ such that $R_{n,\varepsilon} < \infty$ for all $(n, \varepsilon) \in \mathbb{N}_{\geq n_0} \times \mathbb{R}^+$. There then also exists $\delta > 0$ such that* $\lim_{n\to\infty} R_{n,\delta} = 0$.

Lemma 6.2. *Statement* (21) *is a consequence of Corollary* 6.1.

Proof of Lemma 6.1. We arrive at the conclusion from Lebesgue's dominated convergence theorem, [2], Theorem 2.4.4, page 72, by establishing that the integrand

$$n^r h^{-(r+1)}\big(\big(\mathrm{P}(|X-x| < h|x|)\big)^{n-1} - p_x^{n-1}\big) \tag{24}$$

for some choice of δ and all $h \leq \delta$, is pointwise decreasing in n for sufficiently large n and pointwise converging to 0 as n tends to infinity. To this end, define $\pi_x = \mathrm{P}(|X - x| < h|x|)$, $g_x(y) = y^r(\pi_x^y - p_x^y)$, $\lambda_1 = -\log \pi_x$, $\lambda_2 = -\log p_x$. To see that pointwise convergence to 0 holds, note that for some δ and some $\eta > 0$,

$$\pi_x < 1 - \eta \qquad \text{for all } x \text{ and all } h < \delta. \tag{25}$$

Condition (25) indeed prevails, except in the case where F is degenerate with total mass at a single point. Given δ sufficiently small, $\pi_x^{n-1} - p_x^{n-1}$ therefore decays exponentially in n, which yields pointwise convergence to 0 of (24). The decreasing behaviour is equivalent to the existence of $y_0 \geq 0$ such that

$$g_x(y_1) \geq g_x(y_2) \qquad \text{for all } y_1, y_2 \text{ such that } y_0 \leq y_1 \leq y_2. \tag{26}$$

To verify (26), note that

$$g_x'(y) = -y^r(\lambda_1 \mathrm{e}^{-\lambda_1 y} - \lambda_2 \mathrm{e}^{-\lambda_2 y}) + r y^{r-1}(\mathrm{e}^{-\lambda_1 y} - \mathrm{e}^{-\lambda_2 y}) = f_y(\lambda_2) - f_y(\lambda_1) \tag{27}$$

with $f_y(\lambda) := \mathrm{e}^{-\lambda y}(\lambda y^r - r y^{r-1})$ and furthermore that

$$f_y'(\lambda) = \mathrm{e}^{-\lambda y}(y^r - \lambda y^{r+1} + r y^r) = \mathrm{e}^{-\lambda y}\big((r+1)y^r - \lambda y^{r+1}\big). \tag{28}$$

We verify (26) using the fact that $f_y'(\lambda) < 0$ for $\lambda_1 \le \lambda \le \lambda_2$, which, by (28), is satisfied for $y > y_0$, provided $\lambda_1 > \eta$ for some $\eta > 0$. The latter condition is equivalent to (25). □

Proof of Lemma 6.2. It follows from Corollary 6.1 with $U_n = S_n^2/V_n^2$ that

$$\lim_{n\to\infty} \frac{r}{2}\int_0^n z^{r/2-1}\mathrm{P}(U_n > z)\,\mathrm{d}z = \mathrm{E}|T|^r \qquad \text{for all } r > 0. \tag{29}$$

Define $E_n = \{X_1 = X_2 = \cdots = X_n \ne 0\}$ so that $\mathrm{P}(U_n > z) = \mathrm{P}(U_n^* > z) + \mathrm{P}(E_n)$ for $0 < z < n$. The desired conclusion is hence established by showing that for all $r > 0$,

$$\lim_{n\to\infty} n^{r/2+1}\int_0^{n-\delta} z^{r/2-1}\mathrm{P}(E_n)(n-z)^{-(r/2+1)}\,\mathrm{d}z = 0, \tag{30}$$

$$\lim_{n\to\infty}\int_0^{n-\delta} z^{r/2-1}\mathrm{P}(U_n > z)\big(n^{r/2+1}(n-z)^{-(r/2+1)} - 1\big)\,\mathrm{d}z = 0, \tag{31}$$

$$\lim_{n\to\infty}\int_{n-\delta}^{n} z^{r/2-1}\mathrm{P}(U_n > z)\,\mathrm{d}z = 0. \tag{32}$$

Starting with (30), let $\{a_k\}_{k\ge1}$ be a denumeration of all non-zero points attributed mass by F and define $p_k = \mathrm{P}(X = a_k)$, $p = \sup_{k\ge1} p_k$. It follows that $p < 1$ since X is not constant. Moreover,

$$\mathrm{P}(E_n) = \sum_{k\ge1} p_k^n \le p^{n-1}\sum_{k\ge1} p_k \le p^{n-1}.$$

This shows that $\mathrm{P}(E_n)$ decays exponentially in n. However, the quantities

$$n(n-1)^{r/2}\int_0^{n-\delta} z^{(r-2)/2}(n-z)^{-(r+2)/2}\,\mathrm{d}z$$

are all finite and grow with polynomial rate as n grows. Conclusion (30) follows. Statement (32) may be deduced from (29) in the following way:

$$\int_{n-\delta}^{n} z^{r/2-1}\mathrm{P}(U_n > z)\,\mathrm{d}z \le (n-\delta)^{-1}\int_{n-\delta}^{n} z^{r/2}\mathrm{P}(U_n > z)\,\mathrm{d}z \le (n-\delta)^{-1}C_{r+2},$$

where the constant C_{r+2} stems from the identity in (29) with r replaced by $r+2$. It remains to prove (31), which we split into

$$\lim_{n\to\infty}\int_0^1 z^{(r/2-1)}\mathrm{P}(U_n > z)\big(n(n-1)^{r/2}(n-z)^{-(r/2+1)} - 1\big)\,\mathrm{d}z = 0, \tag{33}$$

$$\lim_{n\to\infty}\int_1^{n-\delta} z^{r/2-1}\mathrm{P}(U_n > z)\big(n(n-1)^{r/2}(n-z)^{-(r/2+1)} - 1\big)\,\mathrm{d}z = 0. \tag{34}$$

Statement (33) follows from Lebesgue's dominated convergence theorem, [2], Theorem 2.4.4, page 72. To verify (34), we introduce the notation

$$f_n(z) = z^{r/2-1}\mathrm{P}(U_n > z)\big(n(n-1)^{r/2}(n-z)^{-(r/2+1)} - 1\big)I_{D_n},$$
$$D_n = \{z : 1 \le z \le (n-\delta)\}, \qquad g_n(z) = z^r\mathrm{P}(U_n > z)I_{D_n}, \qquad g(z) = z^r\mathrm{P}(T^2 > z)I_{D_n}.$$

The desired conclusion (34) is now written as (36), while (37) follows from the assumptions, (29) and the elementary inequalities (35):

$$(n-1)/\big(z(n-z)\big) \le (n-1)/\big(\delta(n-\delta)\big) \le C \qquad \text{when } z \in D_n, \tag{35}$$

$$\lim_{n\to\infty}\int f_n = 0, \tag{36}$$

$$\int g_n \to \int g, \qquad g_n \to g, \qquad f_n \to 0, \qquad |f_n| \le C_1 g_n. \tag{37}$$

By a technique called *Pratt's lemma*, Fatou's lemma, [2], Theorem 2.4.3, page 72, and (37) then give

$$C_1\int g = \int \liminf_n (C_1 g_n - f_n) \le \liminf_n \int (C_1 g_n - f_n) = C_1\int g - \limsup_n \int f_n, \tag{38}$$

$$C_1\int g = \int \liminf_n (C_1 g_n + f_n) \le \liminf_n \int (C_1 g_n + f_n) = C_1\int g + \liminf_n \int f_n. \tag{39}$$

Statement (36) follows from (38) and (39). □

Acknowledgements

I would like to thank my Ph.D. supervisor Allan Gut for guidance, encouragement and persistent reading of drafts. I also wish to express my gratitude to Professor Lennart Bondesson for offering valuable comments and criticism regarding [5] in connection with the defense of my Licentiate thesis.

References

[1] Chistyakov, G.P. and Götze, F. (2004). Limit distributions of studentized means. *Ann. Probab.* **34** 28–71. MR2040775

[2] Cohn, D.L. (1980). *Measure Theory.* Boston, MA: Birkhäuser. MR0578344

[3] Giné, E., Götze, F. and Mason, D.M. (1997). When is the Student t-statistic asymptotically standard normal? *Ann. Probab.* **25** 1514–1531. MR1457629

[4] Gut, A. (2007). *Probability: A Graduate Course*, Corr. 2nd printing. New York: Springer. MR2125120

[5] Jonsson, F. (2008). Existence and convergence of moments of Student's t-statistic. Licentiate thesis, U.U.D.M. Report 2008:18.

[6] Praetz, P.D. (1972). The distribution of share price changes. *J. Business* **45** 49–55.
[7] Zabell, S.L. (2008). On Student's 1908 article "The probable error of a mean". With comments and a rejoinder by the author. *J. Amer. Statist. Assoc.* **481** 1–20. MR2394634

Received August 2009 and revised January 2010