STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

S 405

SP 110

Some remarks concerning the quotient of sample median and sample range for a sample of size 2n+1 from a normal distribution

L. de Haan and J.Th. Runnenburg

Reprinted from

Statistica Neerlandica, 23(1969), p 227-234

1969

Some remarks concerning the quotient of sample median and sample range for a sample of size $2n+1$ from a normal distribution

by L. DE HAAN * and J. TH. RUNNENBURG **

0. Introduction and summary

Consider an ordered sample $\underline{x}_{(1)}, \underline{x}_{(2)}, \ldots, \underline{x}_{(2n+1)}$ of size $2n+1$ from the normal distribution with parameters μ and σ. We then have with probability one

$$\underline{x}_{(1)} < \underline{x}_{(2)} < \ldots < \underline{x}_{(2n+1)}.$$

The random variable

$$\underline{h}_n = \frac{\underline{x}_{(n+1)}}{\underline{x}_{(2n+1)} - \underline{x}_{(1)}}, \tag{1}$$

that can be described as the quotient of the sample median and the sample range, provides us with an estimate for μ/σ, that is easy to calculate. To calculate the distribution of $\underline{h}_n$ is quite a different matter***. The distribution function of $\underline{h}_1$ and the density of $\underline{h}_2$ are given in section 1. Our results seem hardly promising for general $\underline{h}_n$. In section 2 it is shown that $\underline{h}_n$ is asymptotically normal.

In the sequel we suppose $\mu = 0$ and $\sigma = 1$, i.e. we consider only the „central" distribution. Note that $\underline{h}_n$ can be used as a test statistic replacing Student's $\underline{t}$. In that case the central $\underline{h}_n$ is all that is needed.

This research was suggested by Prof. Dr. J. HEMELRIJK.

1. Exact distribution

1.1 *Existence of moments*

We define

$$\underline{t}_n = \frac{\underline{\bar{x}}}{\sqrt{\frac{1}{2n+1} \sum_{i=1}^{2n+1} (\underline{x}_i - \underline{\bar{x}})^2}}$$

with

$$\underline{\bar{x}} = \frac{1}{2n+1} \sum_{i=1}^{2n+1} \underline{x}_i.$$

Report S 405 (SP 110), Statistical Department, Mathematical Centre, Amsterdam.
* Mathematical Centre, Amsterdam.
** University of Amsterdam.
*** A different attack from ours is given in [1].

We first show that for each $\alpha > 0$

$$E|\underline{t}_n|^\alpha < \infty \Longleftrightarrow E|\underline{h}_n|^\alpha < \infty. \tag{2}$$

This is an immediate consequence of the inequalities

$$|\underline{h}_n| - 1 < |\underline{t}_n| < 2\sqrt{2n+1}(|\underline{h}_n|+1), \tag{3}$$

which we shall now prove: From

$$|\underline{x}_k - \underline{\bar{x}}| < \underline{x}_{(2n+1)} - \underline{x}_{(1)} \quad \text{for} \quad k = 1, 2, \ldots, 2n+1 \tag{4}$$

it follows that

$$|\underline{\bar{x}}| - |\underline{x}_{(n+1)}| < \underline{x}_{(2n+1)} - \underline{x}_{(1)}$$

and

$$|\underline{x}_{(n+1)}| - |\underline{\bar{x}}| < \underline{x}_{(2n+1)} - \underline{x}_{(1)},$$

hence

$$|\underline{x}_{(n+1)}| \left\{1 - \frac{1}{\underline{h}_n}\right\} < |\underline{\bar{x}}| < |\underline{x}_{(n+1)}| \left\{1 + \frac{1}{\underline{h}_n}\right\}. \tag{5}$$

Obviously

$$\frac{1}{2n+1} \sum_{i=1}^{2n+1} (\underline{x}_i - \underline{\bar{x}})^2 < \underline{x}_{(2n+1)} - \underline{x}_{(1)}. \tag{6}$$

For $k = 1, 2, \ldots, 2n+1$ we have

$$|\underline{x}_k - \underline{\bar{x}}| < \sqrt{\sum_{i=1}^{2n+1} (\underline{x}_i - \underline{\bar{x}})^2},$$

hence

$$\underline{x}_{(2n+1)} - \underline{x}_{(1)} < |\underline{x}_{(2n+1)} - \underline{\bar{x}}| + |\underline{x}_{(1)} - \underline{\bar{x}}| < 2\sqrt{\sum_{i=1}^{2n+1} (\underline{x}_i - \underline{\bar{x}})^2}. \tag{7}$$

Combining (6) and (7) we obtain

$$\frac{1}{\underline{x}_{(2n+1)} - \underline{x}_{(1)}} < \frac{1}{\sqrt{\dfrac{1}{2n+1} \sum_{i=1}^{2n+1} (\underline{x}_i - \underline{\bar{x}})^2}} < \frac{2\sqrt{2n+1}}{\underline{x}_{(2n+1)} - \underline{x}_{(1)}}. \tag{8}$$

Now (3) is an easy consequence of (5) and (8).

As $\sqrt{2n} \cdot \underline{t}_n$ has Student's distribution with $2n$ degrees of freedom, we obtain from (2)

$$\left.\begin{aligned} &E|\underline{h}_n|^{2n} = \infty \\ &E|\underline{h}_n|^\alpha < \infty \quad \text{for} \quad 0 < \alpha < 2n. \end{aligned}\right\} \tag{9}$$

1.2 *Distribution of* $\underline{h}_1$ *and* $\underline{h}_2$

We shall compute the distribution function of $\underline{h}_1$ and the density of $\underline{h}_2$. As a first step we write the integral representing the distribution function $F_n(t)$ of $\underline{h}_n$ in a convenient form. The details are given for $n = 1$ only, for other values of n an analogous result applies.

$$F_1(t) = \frac{\underset{\substack{x<y<z \\ y<t(z-x)}}{\int\int\int} \exp\{-\tfrac{1}{2}(x^2+y^2+z^2)\}dxdydz}{\underset{x<y<z}{\int\int\int} \exp\{-\tfrac{1}{2}(x^2+y^2+z^2)\}dxdydz} =$$

$$= \frac{\underset{\substack{R_1 \cap R_2 \\ 0<r<\infty}}{\int\int\int} \exp(-\tfrac{1}{2}r^2)r^2 \sin\theta \, drd\phi d\theta}{\underset{\substack{R_2 \\ 0<r<\infty}}{\int\int\int} \exp(-\tfrac{1}{2}r^2)r^2 \sin\theta \, drd\phi d\theta} = \frac{\underset{R_1 \cap R_2}{\int\int} \sin\theta \, d\phi d\theta}{\underset{R_2}{\int\int} \sin\theta \, d\phi d\theta},$$

where we have applied the transformation

$$\begin{cases} x = r\cos\phi \ \sin\theta \\ x = r\sin\phi \ \sin\theta \\ z = r\cos\theta \end{cases}$$

and then integrated over r. The region $R_1(\phi,\theta)$ is determined by

$$\sin\phi \sin\theta < t(\cos\theta - \cos\phi \ \sin\theta)$$

and the region $R_2(\phi,\theta)$ by

$$\begin{cases} \cos\phi < \sin\phi \\ \sin\phi \ \sin\theta < \cos\theta \\ 0 < \phi \leqslant 2\pi \\ 0 < \theta \leqslant \pi \end{cases}$$

This last quotient for $F_1(t)$ equals

$$\frac{\underset{\substack{R_1 \cap R_2 \\ 0<r\leqslant 1}}{\int\int\int} r^2 \sin\theta \, d\phi d\theta dr}{\underset{\substack{R_2 \\ 0<r\leqslant 1}}{\int\int\int} r^2 \sin\theta \, d\phi d\theta dr} = \frac{\underset{\substack{x<y<z \\ y<t(z-x) \\ x^2+y^2+z^2\leqslant 1}}{\int\int\int} dxdydz}{I_3}, \tag{10}$$

where ([2] part II p. 304)

$$I_k = \underset{\substack{x_1 < x_2 < \ldots < x_k \\ x_1^2 + x_2^2 + \ldots + x_k^2 \leqslant 1}}{\int\int\int \ldots \int} dx_1 dx_2 \ldots dx_k = \frac{\pi^{k/2}}{k!\,\Gamma\left(\frac{k+2}{2}\right)}; \quad \left(I_3 = \frac{4.\pi}{3.3!},\ I_5 = \frac{8.\pi^2}{15.5!}\right).$$

To determine $F_n(t)$ for arbitrary n we have to compute the content of a part of the unit hypersphere in R^{2n+1} limited by $2n+1$ hyperplanes. This can also be seen without any calculation. Equivalently we can compute the area of a part of the surface of the unit hypersphere in R^{2n+1} limited by $2n+1$ hyperplanes. We choose the latter formulation for $n = 1$: we need only compute the surface of a spherical triangle. The result is wellknown (see [7] p. 216). In this particular case we have

$$F_1(t) = \frac{\frac{2\pi}{3} - \frac{1}{3}\left\{\arccos(-\frac{1}{2}) + \arccos\left(\frac{t-1}{\sqrt{4t^2+2}}\right) + \arccos\left(\frac{t+1}{\sqrt{4t^2+2}}\right)\right\}}{\frac{4.\pi}{3.3!}},$$

i.e.

$$F_1(t) = 2 - \frac{3}{2\pi}\left\{\arccos\frac{t-1}{\sqrt{3t^2+2}} + \arccos\frac{t+1}{\sqrt{4t^2+2}}\right\}. \tag{11}$$

For $n = 2$ we compute the content of the part of the unit hypersphere in R^5 determined by the linear inequalities

$$\begin{cases} x_i < x_{i+1} & \text{for} \quad i = 1, 2, 3, 4, \\ x_3 < tx_5 - tx_1. \end{cases}$$

To this end we use some results from VAN DER VAART [7]. We only quote those parts of the definitions and two theorems in his paper that we need.

Consider a given set of r independent vectors $b_1, b_2, \ldots, b_r \in R^r$ with $b_i' b_i = 1$ for $i = 1, 2, \ldots, r$ (a prime is used to denote the transpose of a vector or a matrix). We consider the hyperspherical simplex

$$\{x | x \in R^r,\ b_1' x \geqslant 0,\ b_2' x \geqslant 0,\ \ldots,\ b_r' x \geqslant 0,\ x' x \leqslant 1\}. \tag{12}$$

The content of this simplex depends only on the matrix

$$C = BB',$$

where B is the matrix with b_i as i-th row ($i = 1, 2, \ldots$, r) .The content of the simplex (12) is denoted by

$$V_r\,(C; 1, 2, \ldots, r),$$

whereas for $1 \leqslant k < l \leqslant r$

$$V_{r-2}(C; \overline{k,l}; 1,2,\ldots,k-1,k+1,\ldots,l-1,l+1,\ldots,r) =$$
$$= \text{content}\{x|x \in R^r, b_1'x \geqslant 0, \ldots, b_{k-1}'x \geqslant 0, b_k'x = 0, b_{k+1}'x \geqslant 0, \ldots,$$
$$b_{l-1}'x \geqslant 0, b_l'x = 0, b_{l+1}'x \geqslant 0, \ldots, b_r'x \geqslant 0, x'x \leqslant 1\}. \quad (13)$$

The index $r-2$ is used to indicate that the simplex can be imbedded in an $(r-2)$-dimensional space.

Theorem 1 (SCHLÄFLI)

$$\frac{\partial}{\partial c_{kl}} V_r(C; 1,2,\ldots,r) = \frac{1}{r} \cdot \frac{1}{\sqrt{1-c_{kl}^2}} V_{r-2}(C; \overline{k,l}; 1,2,\ldots,k-1,k+1,\ldots,l-1, l+1,\ldots,r),$$

where c_{kl} is the (k,l)-element of C (with $1 \leqslant k < l \leqslant r$).

The set (13) van be imbedded in R^{r-2}; a formula giving the matrix $'C$ of an $(r-2)$-dimensional hyperspherical simplex with the same content as (13) is given in the next theorem.

Theorem 2 (VAN DER VAART)

$$V_{r-2}(C; \overline{k,l}, 1,2,\ldots,k-1,k+1,\ldots,l-1,l+1,\ldots,r) = V_{r-2}('C; 1,2,\ldots,r-2)$$

with for $p, q = 1, 2, \ldots, r-2$

$$'c_{pq} = \frac{\det(c_{k,l,v}^{k,l,u})}{\sqrt{\det(c_{k,l,u}^{k,l,u}) \det(c_{k,l,v}^{k,l,v})}},$$

where the matrix $c_{k,l,v}^{k,l,u}$ is the 3×3-submatrix of C where only the rows k, l and u and the columns k, l and v are maintained; here u is the p^{th} element and v the q^{th} element of the sequence $1, 2, \ldots, k-1, k+1, \ldots, l-1, l+1, \ldots, r$.

According to (10) we have to compute

$$V_5(C; 1,2,\ldots,5) =$$
$$\text{content}\left\{(x_1, x_2, \ldots, x_5)|x_1 < x_2 < \ldots < x_5; x_2 \leqslant t(x_5 - x_1); \sum_{i=1}^{5} x_i^2 \leqslant 1\right\} \quad (14)$$

The matrix C is here

$$
\begin{pmatrix}
1 & -\frac{1}{2} & 0 & 0 & \frac{t}{\sqrt{4t^2+2}} \\
-\frac{1}{2} & 1 & -\frac{1}{2} & 0 & \frac{-1}{\sqrt{4t^2+2}} \\
0 & -\frac{1}{2} & 1 & -\frac{1}{2} & \frac{1}{\sqrt{4t^2+2}} \\
0 & 0 & -\frac{1}{2} & 1 & \frac{t}{\sqrt{4t^2+2}} \\
\frac{t}{\sqrt{4t^2+2}} & \frac{-1}{\sqrt{4t^2+2}} & \frac{1}{\sqrt{4t^2+2}} & \frac{t}{\sqrt{4t^2+2}} & 1
\end{pmatrix}.
$$

It is wellknown ([5] p. 171) that the content of a hyperspherical simplex in R^r with r > 3 cannot be expressed in terms of elementary functions of the elements of C. In our case however it is possible to compute the derivative of V_5 as a function of t with the aid of the two preceding theorems. We have

$$
\frac{dV_5(C(t);1,\ldots,5)}{dt} = \sum_{i=1}^{4} \frac{dc_{i5}(t)}{dt} \cdot \frac{\partial V_5(C(t);1,\ldots,5)}{\partial c_{i5}} =
$$

$$
= \sum_{i=1}^{4} \frac{dc_{i5}(t)}{dt} \cdot \tfrac{1}{5} \frac{1}{\sqrt{1-c_{i5}^2(t)}} \cdot V_3(C;\overline{i,5};1,\ldots,i-1,i+1,\ldots,4) =
$$

$$
= \sum_{i=1}^{4} \frac{dc_{i5}(t)}{dt} \cdot \tfrac{1}{5} \frac{1}{\sqrt{1-c_{i5}^2(t)}} \cdot V_3('C^{(i)};1,2,3).
$$

The last quantity only involves a three-dimensional spherical simplex, the content of which can be computed. This computation gives us for the density of $\underline{h}_2$

$$
\frac{dF_2(t)}{dt} = \frac{15}{\pi^2} \frac{1}{2t^2+1} \cdot
$$

$$
\cdot \left[\frac{1}{\sqrt{3t^2+2}} \left\{ 2\pi - \arccos\left(\frac{-\frac{1}{2}(3t^2+t)}{\sqrt{(2t^2+t+\frac{1}{2})(3t^2+1)}} \right) - \arccos\left(\frac{-\frac{1}{2}(t^2-2t)}{\sqrt{(2t^2+t+\frac{1}{2})(2t^2+2)}} \right) + \right. \right.
$$

$$
\left. - \arccos\left(\frac{-\frac{1}{2}(3t^2+2t+2)}{\sqrt{(3t^2+1)(2t^2+2)}} \right) \right\} + \frac{2t}{\sqrt{4t^2+1}} \left\{ 2\pi - \arccos\left(\frac{-\frac{1}{2}(2t^2+t)}{\sqrt{(2t^2+t+\frac{1}{2})(3t^2+\frac{1}{2})}} \right) + \right.
$$

$$
\left. - \arccos\left(\frac{-\frac{1}{2}(2t^2-t)}{\sqrt{(2t^2+t+\frac{1}{2})(3t^2+1)}} \right) - \arccos\left(\frac{-\frac{1}{2}(4t^2+t+1)}{\sqrt{(3t^2+\frac{1}{2})(3t^2+1)}} \right) \right\} +
$$

$$-\frac{2t}{\sqrt{4t^2+1}}\left\{2\pi-\arccos\left(\frac{-\frac{1}{2}(2t^2-t)}{\sqrt{(3t^2+\frac{1}{2})(2t^2-t+\frac{1}{2})}}\right)-\arccos\left(\frac{-\frac{1}{2}(4t^2-t+1)}{\sqrt{(3t^2+1)(3t^2+\frac{1}{2})}}\right)+\right.$$
$$\left.-\arccos\left(\frac{-\frac{1}{2}(2t\ +t)}{\sqrt{(3t^2+1)(2t^2-t+\frac{1}{2})}}\right)\right\}+\frac{1}{\sqrt{3t^2+2}}\left\{2\pi-\arccos\left(\frac{-\frac{1}{2}(3t_2-2t+2)}{\sqrt{(2t^2+2)(3t^2+1)}}\right)+\right.$$
$$\left.\left.-\arccos\left(\frac{-\frac{1}{2}(t^2+2t)}{\sqrt{(2t^2+2)(2t^2-t+\frac{1}{2})}}\right)-\arccos\left(\frac{-\frac{1}{2}(3t^2-t)}{\sqrt{(3t^2+1)(2t^2-t+\frac{1}{2})}}\right)\right\}\right]. \tag{15}$$

Some percentage points of the distribution of $\underline{h}_2$ obtained by numerical integration are

$$\begin{cases} P\{\underline{h}_2 > 0.560\} = 0.025 \\ P\{\underline{h}_2 > 0.735\} = 0.01 \\ P\{\underline{h}_2 > 0.888\} = 0.005. \end{cases}$$

Theoretically it is possible to do a similar computation for $n = 3$, but the result seems hardly worth the effort.

This method of computation can also be applied to other quotients of linear combinations of order statistics, e.g. the so-called extremal quotient $-\underline{x}_{(2n+1)} \cdot \underline{x}_{(1)}^{-1}$.

2. Asymptotic distribution

GNEDENKO [4] proved that if a_n satisfies

$$\Phi(a_n) = 1 - \frac{1}{n},$$

where Φ is the standard normal distribution function, then

$$\lim_{n\to\infty} P\{a_{2n+1}^{-1}\underline{x}_{(2n+1)} \leqslant t\} = \begin{cases} 0 & \text{when} \quad t < 1 \\ 1 & \text{when} \quad t > 1. \end{cases}$$

The same relation holds for $-\underline{x}_{(1)}$. Hence

$$\lim_{n\to\infty} P\{2^{-1}a_{2n+1}^{-1}(\underline{x}_{(2n+1)}-\underline{x}_{(1)}) \leqslant t\} = \begin{cases} 0 & \text{when} \quad t < 1 \\ 1 & \text{when} \quad t > 1. \end{cases}$$

CRAMÈR [3] proved that

$$\lim_{n\to\infty} P\{2\sqrt{2n+1}\cdot\underline{x}_{(n+1)} \leqslant t\} = \Phi(t) \quad \text{for all} \quad t.$$

As $a_n \sim \sqrt{2 \log n}$ for $n \to \infty$, these results imply

$$\lim_{n\to\infty} P\{8\sqrt{n \log n}\cdot \underline{h}_n \leqslant t\} = \lim_{n\to\infty} P\{4\sqrt{2n+1}\cdot\sqrt{2\log(2n+1)}\cdot \underline{h}_n \leqslant t\} = \Phi(t)$$

for all t. But then $\underline{h}_n$ is asymptotically normal with parameters $\mu_n = 0$ and $\sigma_n = (8\sqrt{n \log n})^{-1}$.

References

[1] N. Bouma en A.Vehmeyer, Percentiles and approximations of the sample median over the sample range for samples of size 3,5, 7 and 9 from a standard normal distribution. Statistica Neerlandica **23** (1969) 235–239.

[2] R. Courant, Integral and differential calculus, I and II. New York 1936.

[3] H. Cramèr, The elements of probability theory and some of its applications. New York 1955.

[4] B. Gnedenko, Sur la distribution limite du terme maximum d'une série aléatoire. Annals of Math. **44** (1943) 423–453.

[5] A. Sarhan and B. Greenberg, Contributions to order statistics. New York 1962.

[6] L. Schläfli, Theorie der Vielfachen Kontinuität. Gesammelte Mathematische Abhandlungen, Band I. Basel 1950.

[7] H. van der Vaart, The content of some classes of non-euclidian polyhedra for any number of dimensions, with several applications, I and II. Indag. Math. **17** (1955) 199–221.