THE JOINT DISTRIBUTION OF THE STANDARDIZED MAXIMUM AND STANDARDIZED MINIMUM FOR A NORMAL SAMPLE

L.K. Shiryaeva
Samara State Economic University
Samara, RUSSIA
e-mail: Shiryeva_LK@mail.ru

Abstract

We find the joint distribution of Grubbs test statistics for a normal sample. Those statistics are standardized maximum and standardized minimum. We note some properties of the joint distribution function which can be applied for possible application of this function.

1 Introduction

Let X_1, X_2, ..., X_{n-1}, X_n be a random sample from a normal $N(a, \sigma^2)$ distribution with mean a and variance σ^2. Grubbs proposed the standardized maximum and standardized minimum [1]:

$$T_n^{(1)} = \frac{\max\limits_{1 \le i \le n}\{X_i\} - \overline{X}}{S}; \qquad T_n^{(1)} = \frac{\overline{X} - \min\limits_{1 \le i \le n}\{X_i\}}{S},$$

where $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the sample mean and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ is the sample variance.

It is known that distributions of statistics $T_n^{(1)}$ and $T_{n,(1)}$ coincide, i.e.

$$P(T_n^{(1)} < t) = P(T_{n,(1)} < t). \tag{1}$$

Let $F_n^{(1)}(t)$ be the distribution function of $T_n^{(1)}$, then [2]:

$$F_n^{(1)}(t) = P(T_n^{(1)} < t) = \begin{cases} 0, \ t \le \frac{1}{\sqrt{n}}, \ n \ge 2; \\ n \int\limits_{\frac{1}{\sqrt{n}}}^{t} F_{n-1}^{(1)}(g_n(x)) f_{T_n}(x)dx, \ \frac{1}{\sqrt{n}} < t \le \frac{n-1}{\sqrt{n}}, \ n \ge 3; \\ 1, \ t > \frac{n-1}{\sqrt{n}}, n \ge 2; \end{cases} \tag{2}$$

where

$$f_{T_n}(x) = \frac{1}{n-1}\sqrt{\frac{n}{\pi}}\Gamma\left(\frac{n-1}{2}\right)/\Gamma\left(\frac{n-2}{2}\right)\left(1 - \frac{n}{(n-1)^2}x^2\right)^{\frac{n-4}{2}}, \ |x| < \frac{n-1}{\sqrt{n}}; \tag{3}$$

$$g_n(x) = \frac{n}{n-1}x \Big/ \sqrt{\frac{n-1}{n-2}\left(1 - \frac{n}{(n-1)^2}x^2\right)}, \ |x| < \frac{n-1}{\sqrt{n}}, \ n \ge 3. \tag{4}$$

We find the joint distribution of Grubbs test statistics $T_{n,(1)}$ and $T_n^{(1)}$ for a normal sample.

2 The joint distribution of Grubbs test statistics for a normal sample

Define

$$T_i = \frac{X_i - \overline{X}}{S}, \ (i = 1, 2, \ldots, n)\,.$$

Random variables $T_1,\ T_2, \ldots, T_n$ are identically distribused. Let $f_{i;n}(x)$ be density function of studentized deviation T_i, $i = \overline{1,\ n}$, then [3]:

$$f_{i;n}(x) = f_{T_n}(x), (i = 1,\ 2,\ \ldots, n),$$

where $f_{T_n}(x)$ can be calculated with using (3).

We have the following equalities for statistics $T_n^{(1)}$ and $T_{n,(1)}$:

$$T_{n,(1)} = -\min_{1\le i\le n}\{T_i\}, \tag{5}$$

$$T_n^{(1)} = \max_{1\le i\le n}\{T_i\}. \tag{6}$$

It follows from relationships (1) and (2) that

$$\frac{1}{\sqrt{n}} \le T_{n,(1)} \le \frac{n-1}{\sqrt{n}},\ \frac{1}{\sqrt{n}} \le T_n^{(1)} \le \frac{n-1}{\sqrt{n}}. \tag{7}$$

Let $\Lambda_n(t_1, t_2) = P(T_{n,(1)} < t_1, T_n^{(1)} < t_2)$ be the joint distribution function of Grubbs test statistics $T_n^{(1)}$ and $T_{n,(1)}$. The following theorem describes our main result.

Theorem 1. *If $X_1,\ X_2,\ \ldots,\ X_n$ is a random sample from a normal $N(a, \sigma^2)$ distribution, then the joint distribution function of Grubbs test statistics $T_n^{(1)}$ and $T_{n,(1)}$ for the case $n = 2$ is given by*

$$\Lambda_2(t_1, t_2) = \begin{cases} 1, & (t_1, t_2) \in \Delta_2,\ \Delta_2 = [\frac{\sqrt{2}}{2} < t_1 < \infty; \frac{\sqrt{2}}{2} < t_2 < \infty]; \\ 0, & (t_1, t_2) \notin \Delta_2, \end{cases} \tag{8}$$

and for the case $n > 2$

$$\Lambda_n(t_1, t_2) = \begin{cases} F_n^{(1)}(t_2),\ \ t_1 \ge \frac{n-1}{\sqrt{n}}; \\ F_n^{(1)}(t_1),\ \ t_2 \ge \frac{n-1}{\sqrt{n}}; \\ n \int\limits_{\frac{1}{\sqrt{n}}}^{t_2} \Lambda_{n-1}\left(\rho_n(t_1, -x), g_n(x)\right) f_{T_n}(x) dx, \ \ (t_1, t_2) \in \Delta_n; \\ 0,\ \ (t_1, t_2) \notin \Delta_n,\ t_1 < \frac{n-1}{\sqrt{n}},\ t_2 < \frac{n-1}{\sqrt{n}}, \end{cases} \tag{9}$$

where distribution function $F_n^{(1)}(t)$ can be calculated with using (2);

$$\rho_n(u, v) = \left(u + \frac{v}{n-1}\right) \Big/ \sqrt{\frac{n-1}{n-2}\left(1 - \frac{n}{(n-1)^2} v^2\right)},\ |v| < \frac{n-1}{\sqrt{n}}; \tag{10}$$

functions $g_n(x)$ and $f_{T_n}(x)$ can be calculated with using (4) and (3) correspodingly; $\Delta_n = [1/\sqrt{n} < t_1 < (n-1)/\sqrt{n};\ 1/\sqrt{n} < t_2 < (n-1)/\sqrt{n}\,]$, if $n > 2$.

Proof. Suppose that $n = 2$. In that case $T_{2,(1)} = T_2^{(1)} = \frac{1}{\sqrt{2}}$. We have $\Lambda_2(t_1, t_2) = P\left\{(1/\sqrt{2} < t_1) \cap (1/\sqrt{2} < t_2)\right\}$, that proves the formula (8).

Suppose that $n > 2$. Then for $t_1 \geq \frac{n-1}{\sqrt{n}}$ by (7) we have

$$\Lambda_n(t_1, t_2) = P\{(T_{n,(1)} < (n-1)/\sqrt{n}) \cap (T_n^{(1)} < t_2)\} = P\left(T_n^{(1)} < t_2\right).$$

Therefore,

$$\Lambda_n(t_1, t_2) = F_n^{(1)}(t_2),\ t_1 \geq \frac{n-1}{\sqrt{n}}, \tag{11}$$

where $F_n^{(1)}(t_2)$ is defined by (2).

Similarly, in the case $t_2 \geq (n-1)/\sqrt{n}$ we have $\Lambda_n(t_1, t_2) = P\left(T_{n,(1)} < t_1\right)$. It follows from (1) that

$$\Lambda_n(t_1, t_2) = F_n^{(1)}(t_1), \quad t_2 \geq (n-1)/\sqrt{n}. \tag{12}$$

Next, it follows from conditions (7), that

$$\Lambda_n(t_1, t_2) = 0,\ \text{ if } (t_1 \leq 1/\sqrt{n}) \text{ or } (t_2 \leq 1/\sqrt{n}). \tag{13}$$

Let be $\Delta_n = [1/\sqrt{n} < t_1 < (n-1)/\sqrt{n};\ 1/\sqrt{n} < t_2 < (n-1)/\sqrt{n}\,]$. We have with use (5) and (6) that $\Lambda_n(t_1, t_2) = P\{(\min\limits_{1\leq j\leq n}\{T_j\} > -t_1) \cap (\max\limits_{1\leq j\leq n}\{T_j\} < t_2)\}$. Hence, the joint distribution function of statistics $T_{n,(1)}$ and $T_n^{(1)}$ is defined by

$$\Lambda_n(t_1, t_2) = P\{\cap_{i=1}^{n}(-t_1 < T_i < t_2)\}. \tag{14}$$

Next, using formulas (14), (6) and (7), we have for $(t_1, t_2) \in \Delta_n$

$$\Lambda_n(t_1, t_2) = P\left\{\cap_{i=1}^{n}\left(-t_1 < T_i \leq \max_{1\leq j\leq n}\{T_j\}\right) \cap \left(\frac{1}{\sqrt{n}} < \max_{1\leq j\leq n}\{T_j\} < t_2\right)\right\}.$$

Then

$$\Lambda_n(t_1, t_2) = nP\left\{\cap_{i=1}^{n-1}(-t_1 < T_i < T_n) \cap \left(\frac{1}{\sqrt{n}} < T_n < t_2\right)\right\}. \tag{15}$$

Define

$$T_i^* = \frac{X_i - \overline{X}^*}{S^*},\ (i = 1, 2, \ldots, n-1),$$

where $\overline{X}^* = \frac{1}{n-1}\sum\limits_{k\neq n} X_k,\ S^{*2} = \frac{1}{n-2}\sum\limits_{k\neq n}(X_k - \overline{X}^*)^2$.

It is valid that [3]

$$T_i = T_i^*\sqrt{\frac{n-1}{n-2}\left(1 - \frac{n}{(n-1)^2}T_n^2\right)} - \frac{1}{n-1}T_n,\ i = \overline{1, n-1},\ |T_n| < \frac{n-1}{\sqrt{n}}.$$

Therefore the relationship (15) can be written as:

$$\Lambda_n(t_1, t_2) = nP\left\{\cap_{i=1}^{n-1}(\rho_n(-t_1, T_n) < T_i^* < g_n(T_n)) \cap (1/\sqrt{n} < T_n < t_2)\right\}, \tag{16}$$

where the function $g_n(x)$ is defined by (4) and $\rho_n(u, v)$ — by (10).

It is well known that random variable T_n is the function of $\overline{X}^*$, S^* and X_n [3]. Besides, random variables $\left\{T_1^*, T_2^*, \ldots, T_{n-1}^*\right\}$ is independent of $\{\overline{X}^*, S^*, X_n\}$. Hence, the random variable T_n is independent of $\left\{T_1^*, T_2^*, \ldots, T_{n-1}^*\right\}$ [3]. Therefore the relationship (16) can be written as

$$\Lambda_n(t_1, t_2) = n \int\limits_{1/\sqrt{n}}^{t_2} P\{\cap_{i=1}^{n-1}\{\rho_n(-t_1, x)) < T_i^* < g_n(x)\} f_{T_n}(x)dx.$$

It follows from (14) that

$$\Lambda_n(t_1, t_2) = n \int\limits_{1/\sqrt{n}}^{t_2} \Lambda_{n-1}\{\rho_n(t_1, -x)), g_n(x)\} f_{T_n}(x)dx, \quad (t_1, t_2) \in \Delta_n. \tag{17}$$

If we combine relationships (11), (12), (13) and (17), then we recieve (9), that proves the theorem.

Note some properties of the joint distribution function $\Lambda_n(t_1, t_2)$ which can be derived from (9) .

1. Function $\Lambda_n(t_1, t_2)$ is symmetrical, i.e. $\Lambda_n(t_1, t_2) = \Lambda_n(t_2, t_1)$.
2. It is valid for $(t_1,\, t_2) \in D_n$ and $n > 2$:

$$\Lambda_n(t_1, t_2) = F_n^{(1)}(t_1) - n \int\limits_{t_2}^{\frac{n-1}{\sqrt{n}}} F_{n-1}^{(1)}(\rho_n(t_1, -x)) f_{T_n}(x)dx,$$

where $D_n = [\frac{1}{\sqrt{n}} < t_1 < \frac{n-1}{\sqrt{n}}; \tau_n^* \le t_2 < \frac{n-1}{\sqrt{n}}]$; $\tau_n^* = \sqrt{\frac{(n-1)(n-2)}{2n}}$.

3. It is valid for $(t_1,\, t_2) \in \Xi_n$ and $n > 2$:

$$\Lambda_n(t_1, t_2) = F_n^{(1)}(t_1) + F_n^{(1)}(t_2) - 1, \tag{18}$$

where $\Xi_n = [t_n^* \le t_1 < \frac{n-1}{\sqrt{n}}; t_n^* \le t_2 < \frac{n-1}{\sqrt{n}}]$; $t_n^* = \sqrt{\frac{n-1}{2}}$;

$F_n^{(1)}(t) = 1 - n \int\limits_{t}^{\frac{n-1}{\sqrt{n}}} f_{T_n}(x)dx, \quad t \ge t_n^*.$

References

[1] Grubbs F. (1950). Sample Criteria for Testing Outlying observations *Ann. Math. Statist.* Vol. **21**, pp. 27–58.

[2] Barnett V., Lewis T. (1984).*Outliers in statistical data.* John Wiley Sons, Chichester.

[3] Zhang J., Keming Y. (2006). The null distribution of the likelihood-ratio test for one or two outliers in a normal sample.*An Official Journal of the Spanish Society of Statistics and Operations Research.* Vol. **15**, pp. 141–150.