Liet. matem. rink., T. 43, spec. nr., 677–681

Moments of extremes of normally distributed values

Agnė BURAUSKAITĖ, Algimantas AKSOMAITIS (KTU)
e-mail: agn@delfi.lt, aksoma@fmf.ktu.lt

1. Introduction

In this paper, we look at Gaussian distribution from a point of view of extreme value theory. More concretely, moments of maximum and minimum of normally distributed values are considered.

There are methods to calculate moments of extremes of independent normally distributed values in certain cases. In a case of iid normal values, for n $\leqslant 5$ (n – a number of values), there are formulas to express every order moments of extremes using elementary functions. Expressions for two and three independent normally distributed values with the same mean, but different variances are given in [1].

In this paper, we analyze a case of dependent normal variables. It is known that dependent normal variables could be written as sums of independent variables [4]. However it does not help to transform expressions so that formulas, mentioned before, could be used. To solve a problem we have to look for new methods.

2. Presentation of results

Theorem 1. *A mean value of normal dependent variables $X, Y \sim N(m, m, \sigma_1, \sigma_2, \rho)$ is following:*

$$\mathrm{E}\left(\max(X,Y)\right) = m + \sqrt{\frac{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}{2\pi}}. \tag{1}$$

Theorem 2. *For normal dependent variables $X, Y \sim N(m, m, \sigma_1, \sigma_2, \rho)$ the following holds:*

$$\mathrm{E}\left(\max^{2k}(X,Y)\right) = \frac{1}{2}\,(2k-1)!!\left(\sigma_1^{2k} + \sigma_2^{2k}\right) \quad \forall k \in N. \tag{2}$$

COROLLARY 1. A variance of normal dependent variables $X, Y \sim N(m, m, \sigma_1, \sigma_2, \rho)$ is following:

$$\mathrm{V}ar\ \left(\max(X,Y)\right) = \frac{1}{2}\left(\sigma_1^2 + \sigma_2^2 - \frac{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}{\pi}\right). \tag{3}$$

Theorem 3. *A mean value of normal dependent variables X, Y, Z ~ N(m, m, m, $\sigma_1, \sigma_2, \sigma_3, \rho_{12}, \rho_{23}, \rho_{31}$) is following:*

$$\mathrm{E}\left(\max(X,Y,Z)\right) = m + \frac{1}{2\sqrt{2\pi}} \times\left(\sqrt{\sigma_1^2-2\rho_{12}\sigma_1\sigma_2+\sigma_2^2}+\sqrt{\sigma_2^2-2\rho_{23}\sigma_2\sigma_3+\sigma_3^2}+\sqrt{\sigma_3^2-2\rho_{31}\sigma_3\sigma_1+\sigma_1^2}\right). \quad (4)$$

3. Proofs of Theorems

Proof of Theorem 1. Let us consider random variables $X, Y \sim N(0,0,\sigma_1,\sigma_2,\rho)$.

It is shown in [2] that an average could be expressed in a form $\mathrm{E}X = \mathrm{E}\,(X|\,A)\,\mathrm{P}\,(A)+ \mathrm{E}\,(Y\,|B\,)\,\mathrm{P}(B)$, using conditional averages and probabilities of events A and B, where $\mathrm{P}(AUB) = 1$.

Applying this formula to our problem we get an expression that does not contain a function *max*:

$$\begin{aligned}\mathrm{E}\left(\max(X,Y)\right) &= \mathrm{E}\left(\max(X,Y)|\,X \geqslant Y\right)\mathrm{P}\left(X \geqslant Y\right) \\ &\quad + \mathrm{E}\left(\max(X,Y)\,|X < Y\,\right)\mathrm{P}(X < Y) \\ &= \mathrm{E}\left(X|\,X \geqslant Y\right)\mathrm{P}\left(X \geqslant Y\right) + \mathrm{E}\left(Y\,|X < Y\,\right)\mathrm{P}(X < Y). \quad (5)\end{aligned}$$

After that we replace variables with $Z = X - Y$ and $W = Y - X$. Also formulas for calculating conditional averages [3] are used:

$$\begin{aligned}&\mathrm{E}\left(\max(X,Y)\right) = \mathrm{E}\left(Y|\,Z \leqslant 0\right)\mathrm{P}\left(Z \leqslant 0\right) + \mathrm{E}\left(X\,|W < 0\,\right)\mathrm{P}(W < 0) \\ &= \int_{-\infty}^{0} \mathrm{E}\left(Y|\,Z = z\right)p_Z(z)\,\mathrm{d}z + \int_{-\infty}^{0} \mathrm{E}\left(X|\,W = w\right)p_W(w)\,\mathrm{d}w \\ &= \int_{-\infty}^{0}\int_{-\infty}^{\infty} yp(z,y)dydz + \int_{-\infty}^{0}\int_{-\infty}^{\infty} xp(w,x)\,\mathrm{d}xdw. \quad (6)\end{aligned}$$

To prove the theorem we need to determine two dimensional probability densities $p(z,y),\ p(w,x)$ and perform integration. Finally we get the result:

$$\mathrm{E}\left(\max(X,Y)\right) = \sqrt{\frac{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}{2\pi}}. \quad (7)$$

If $X_c, Y_c \sim N(0,0,\sigma_1,\sigma_2,\rho)$ and $X, Y \sim N(m,m,\sigma_1,\sigma_2,\rho)$, because of a property $\mathrm{E}\left(\max(X,Y)\right) = m + \mathrm{E}\left(\max(X_C,Y_C)\right)$, we get formula (1). Theorem is proved.

Proof of Theorem 2. Let us consider random values $X, Y \sim N(0,0,\sigma_1,\sigma_2,\rho)$.

In this proof we will also use conditional averages:

$$\mathrm{E}\left(\max^{2k}(X,Y)\right) = \mathrm{E}\left(\max^{2k}(X,Y)\,|X < 0, Y < 0\right)\mathrm{P}\left(X < 0, Y < 0\right)$$

$$
\begin{aligned}
&+\mathrm{E}\left(\max^{2k}(X,Y)\,|X<0,Y>0\right)\mathrm{P}\left(X<0,Y>0\right)\\
&+\mathrm{E}\left(\max^{2k}(X,Y)\,|X>0,Y<0\right)\mathrm{P}\left(X>0,Y<0\right)\\
&+\mathrm{E}\left(\max^{2k}(X,Y)\,|X>0,Y>0\right)\mathrm{P}\left(X>0,Y>0\right)\\
=&\ \mathrm{E}\left(Y^{2k}\,|Y<0\right)\mathrm{P}\left(X<0,Y<0,X<Y\right)\\
&+\mathrm{E}\left(X^{2k}\,|X<0\right)\mathrm{P}\left(X<0,Y<0,X>Y\right)\\
&+\mathrm{E}\left(Y^{2k}\,|Y>0\right)\mathrm{P}\left(X<0,Y>0\right)\\
&+\mathrm{E}\left(X^{2k}\,|X>0\right)\mathrm{P}\left(X>0,Y<0\right)\\
&+\mathrm{E}\left(Y^{2k}\,|Y>0\right)\mathrm{P}\left(X>0,Y>0,X<Y\right)\\
&+\mathrm{E}\left(X^{2k}\,|X>0\right)\mathrm{P}\left(X>0,Y>0,X>Y\right). \qquad (8)
\end{aligned}
$$

Probabilities of events related to X and Y have following properties:

$$
\begin{aligned}
&\mathrm{P}\left(X>0,Y>0,X<Y\right)=\mathrm{P}\left(X<0,Y<0,X>Y\right)=P_1,\\
&\mathrm{P}\left(X>0,Y>0,X>Y\right)=\mathrm{P}\left(X<0,Y<0,X<Y\right)=P_2,\\
&\mathrm{P}\left(X>0,Y<0\right)=\mathrm{P}\left(X<0,Y>0\right)=P_3 \qquad (9)
\end{aligned}
$$

$$
\begin{aligned}
&P_1+P_2+P_3+P_3+P_1+P_2\\
&\quad=\mathrm{P}\left(X>0,Y>0,X<Y\right)+\mathrm{P}\left(X>0,Y>0,X>Y\right)\\
&\quad\quad+\mathrm{P}\left(X>0,Y<0\right)+\mathrm{P}\left(X<0,Y>0\right)\\
&\quad\quad+\mathrm{P}\left(X<0,Y<0,X>Y\right)+\mathrm{P}\left(X<0,Y<0,X<Y\right)=1 \qquad (10)
\end{aligned}
$$

Using properties (9) and (10) for expression (8) we get:

$$
\mathrm{E}\left(\max^{2k}(X,Y)\right)=\frac{1}{2}\left(\mathrm{E}\left(X^{2k}\right)+\mathrm{E}\left(Y^{2k}\right)\right). \qquad (11)
$$

It means that an even order moment of maximum of two dependent normal variables is equal to arithmetic average of the same order moments of each of them. Also it shows that this moment does not depend on correlation coefficient and is the same for dependent as well as for independent variables.

Integrating by parts we can easily get:

$$
\mathrm{E}X^{2k}=\int_{-\infty}^{\infty}x^{2k}e^{-\frac{x^2}{2\sigma_x}}\,\mathrm{d}x=(2k-1)!!\sigma_x^{2k}, \qquad (12)
$$

From equalities (11) and (12) we get a final result (2). Theorem is proved.

Proof of Corollary. Let us consider random values $X,Y\sim N(m,m,\sigma_1,\sigma_2,\rho)$ and $X_c,Y_c\sim N(0,0,\sigma_1,\sigma_2,\rho)$. By a property of variance and using results of theorems (1) and (2) we get a proof:

$$
\begin{aligned}
&\mathrm{V}ar\ \left(\max(X,Y)\right)=\mathrm{V}ar\ \left(\max(X_C,Y_C)\right)\\
&\quad=\mathrm{E}\left(\min^2(X_C,Y_C)\right)-\mathrm{E}^2\left(\min(X_C,Y_C)\right)
\end{aligned}
$$

$$= \frac{1}{2}\left(\sigma_1^2 + \sigma_2^2\right) - \left(\sqrt{\frac{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}{2\pi}}\right)^2$$
$$= \frac{1}{2}\left(\sigma_1^{2k} + \sigma_2^{2k} - \frac{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}{\pi}\right). \tag{13}$$

Proof of Theorem 3. To prove this theorem we will need a following lemma.

Lemma 1. *For a real number vector (x, y, z) it is true:*

$$\max(x,y,z) - \min(x,y,z) = \max(x,y) + \max(y,z) + \max(z,x) - (x+y+z). \tag{14}$$

Lemma could be easily proved using following equalities:

$$\max(x,y) = \frac{x+y+|x-y|}{2}, \quad \max(y,z) = \frac{y+z+|y-z|}{2},$$
$$\max(z,x) = \frac{z+x+|z-x|}{2}, \tag{15}$$
$$\max(x,y,z) - \min(x,y,z) = \frac{|x-y| + |y-z| + |z-x|}{2}. \tag{16}$$

Let us consider random variables $X, Y, Z \sim N(0,0,0,\sigma_1,\sigma_2,\sigma_3,\rho_{12},\rho_{23},\rho_{31})$. Variables X, Y, Z are centered, therefore there exist relations:

$$\min(X,Y,Z) = -\max(X,Y,Z),$$
$$\mathrm{E}\left(\max(X,Y,Z) - \min(X,Y,Z)\right) = 2\mathrm{E}\left(\max(X,Y,Z)\right). \tag{17}$$

Applying the lemma and theorem (1) we get a needed expression:

$$\mathrm{E}\left(\max(X,Y,Z)\right) = \frac{1}{2}\mathrm{E}\left(\max(X,Y,Z) - \min(X,Y,Z)\right)$$
$$= \frac{1}{2}\mathrm{E}\left(\max(X,Y) + \max(Y,Z) + \max(Z,X) - (X+Y+Z)\right)$$
$$= \frac{1}{2}\left(\mathrm{E}\left(\max(X,Y)\right) + \mathrm{E}\left(\max(Y,Z)\right) + \mathrm{E}\left(\max(Z,Y)\right) - \mathrm{E}X - \mathrm{E}Y - \mathrm{E}Z\right)$$
$$= \frac{1}{2\sqrt{2\pi}}\left(\sqrt{\sigma_1^2 - 2\rho_{12}\sigma_1\sigma_2 + \sigma_2^2} + \sqrt{\sigma_2^2 - 2\rho_{23}\sigma_2\sigma_3 + \sigma_3^2}\right.$$
$$\left. + \sqrt{\sigma_3^2 - 2\rho_{31}\sigma_3\sigma_1 + y\sigma_1^2}\right). \tag{18}$$

According to a property of average, for not centered variables with mean value m we have expression (4). The proof of theorem is complete.

4. Simulation

A problem with four and more dependent normal variables was studied experimentally. A program, generating dependent normal values and calculating empirical moments of their extremes, was implemented in Matlab. In Fig. 1 we see first moment dependences on $\sqrt{1-\rho}$ (correlation coefficient ρ is the same for each pair of variables). Lowest lines represent theoretical results expressed by formulas (1) and (4) for cases of two and three variables. Other two lines show experimental results for four and one hundred variables. It is clear that the more variables are taken the larger is angle between a line and x-axis.

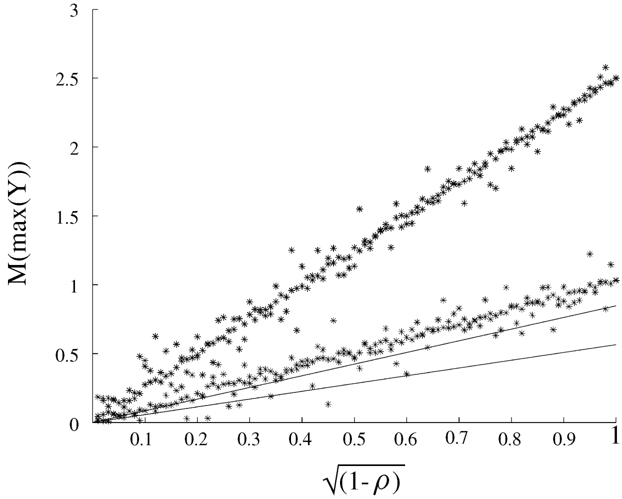

Fig. 1. First moment dependences on $\sqrt{1-\rho}$.

References

[1] И.Н. Коваленко, А.А. Левитская, М.Н. Савчик, *Избранные задачи вероятностной комбинаторики*, Киев (1986).

[2] J. Kubilius, *Tikimybių teorija ir matematinė statistika*, Mokslas, Vilnius (1980).

[3] A. Aksomaitis, *Tikimybių teorija ir statistika*, Kaunas, Technologija (2000).

[4] Я. Галамбош, *Асимптотическая теория экстремальных порядковых статистик*, Наука, Москва (1984).

Normaliųjų dydžių ekstremumų momentų skaičiuotė

A. Burauskaitė, A. Aksomaitis

Nepriklausomų normaliųjų dydžių ekstremumų pradiniai momentai yra išsamiai išnagrinėti. Literatūroje [1] galima rasti formules dviejų ir trijų nepriklausomų normaliųjų dydžių su vienodais vidurkiais, bet skirtingomis dispersijomis pradiniams momentams skaičiuoti. Taip pat yra ištirti ribiniai atvejai.

Šiame darbe yra ieškoma analogiškų išraiškų priklausomų dydžių atveju. Yra gautos ir įrodytos formulės dviejų priklausomų normaliųjų dydžių ekstremumo vidurkiui, dispersijai ir bet kurios lyginės eilės pradiniams momentams skaičiuoti. Taip pat yra pateikta išraiška trijų priklausomų normaliųjų dydžių ekstremumo vidurkiui apskaičiuoti.