Liet. matem. rink., **44**, *spec. nr.*, 2004, 241–244

Priklausomų normaliųjų dydžių maksimumo vidurkis

Agnė BURAUSKAITĖ, Algimantas AKSOMAITIS (KTU)
el. paštas: agn@delfi.lt, aksoma@fmf.ktu.lt

Šiame darbe yra tiriamas normaliųjų atsitiktinių dydžių sekos maksimumo vidurkis. Neprarasdami bendrumo toliau nagrinėsime tik standartinių normaliųjų dydžių seką $X_i \sim N(0,1)$, $i = 1, \ldots, n$, ir jos maksimumo vidurkį $\mathrm{E}(\max(X_1, X_2, \ldots, X_n))$ (žymėsime μ_n). Yra daug būdų šią charakteristiką apskaičiuoti apytiksliai, tačiau ieškant matematinės išraiškos elementariosiomis funkcijomis iškyla nemažai problemų. Uždavinys tampa dar sudėtingesnis, jei atsitiktiniai dydžiai yra priklausomi. Mes maksimumo vidurkį įvairiais atvejais tyrėme modeliuodami kompiuteriu.

Tuomet, kai standartiniai normalieji dydžiai yra nepriklausomi, mažiems n ($n < 6$) yra žinomos ekstremumų vidurkių išraiškos elementariosiomis funkcijomis [2]. Kai $n = 6$ ir $n = 7$, vidurkiai užrašomi papildomai įvedant specialios formos integralą. Didesniems n išraiškos nėra gautos.

Darbe [1], tiriant priklausomus normaliuosius dydžius su vienodais koreliacijos koeficientais ρ, įvairiems n buvo eksperimentiškai nustatyta, kad maksimumo vidurkis tiesiškai priklauso nuo $\sqrt{1-\rho}$. Be to, pagal tame pačiame darbe pateiktas dviejų ir trijų priklausomų normaliųjų dydžių ekstremumų vidurkių skaičiavimo išraiškas pastebime, kad priklausomų ir nepriklausomų dydžių vidurkiai turi panašią struktūrą. Priklausomiems standartiniams normaliesiems dydžiams su vienodais koreliacijos koeficientais gauname:

$$\mu_2{*} = \frac{\sqrt{1-\rho}}{\sqrt{\pi}}, \quad \mu_3{*} = \frac{3\sqrt{1-\rho}}{2\sqrt{\pi}}.$$

Atitinkami nepriklausomų dydžių vidurkiai:

$$\mu_2 = \frac{1}{\sqrt{\pi}}, \quad \mu_3 = \frac{3}{2\sqrt{\pi}}.$$

Remiantis šiais pastebėjimais, yra pagrindo manyti, kad galioja sąryšis

$$\mu_n{*} = \mu_n\sqrt{1-\rho}, \quad n \geqslant 1, \tag{1}$$

čia $\mu_n{*}$ – priklausomų, o μ_n – nepriklausomų normaliųjų dydžių maksimumo vidurkiai.

Šį pastebėjimą taip pat paremia ir matematinis modeliavimas. Modeliavimas buvo atliktas Matlab programinėje aplinkoje. Trumpai aptarsime modeliavimo metodiką.

Nepriklausomų normaliųjų dydžių maksimumo vidurkiai, kai $n < 8$ yra žinomi [2]. Didesniems n vidurkius skaičiuosime tokiu būdu: generuojamos standartinių normaliųjų dydžių sekos, išrenkamos sekų maksimalios reikšmės, apskaičiuojamas reikšmių aritmetinis vidurkis.

Priklausomų normaliųjų dydžių maksimumo vidurkių modeliavimo algoritmas:

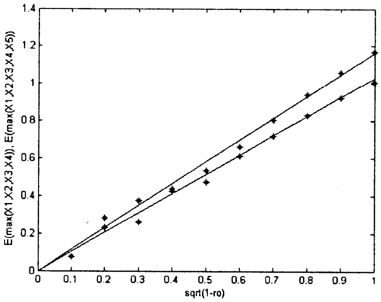
1) generuojamos standartinių normaliųjų dydžių sekos $X_j = (X_1, X_2, \ldots, X_n)$;
2) sekos transformuojamos, gaunamos priklausomų dydžių sekos Y_j:

$$Y_j = V^T X_j^T, \quad \text{čia } V^T V = C = \begin{pmatrix} 1 & \rho & \ldots & \rho \\ \rho & 1 & \ldots & \rho \\ \ldots & \ldots & \ldots & \ldots \\ \rho & \rho & \ldots & 1 \end{pmatrix}$$

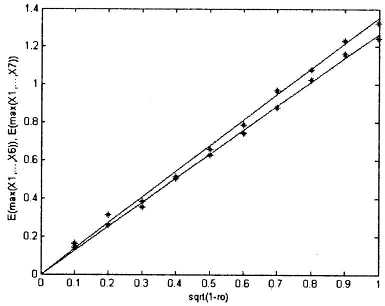
(transformacijų matrica V apskaičiuojama standartine Matlab funkcija $V = chol(C)$);

3) išrenkamos sekų Y_j maksimalios reikšmės;
4) apskaičiuojamas maksimumų aritmetinis vidurkis.

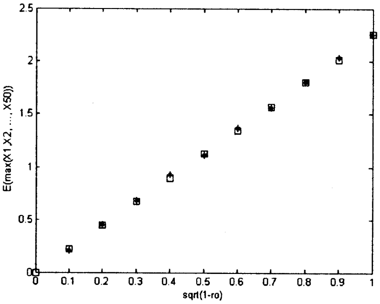
Toliau pateikiami modeliavimo rezultatai, kai $n = 4, 5, 6, 7$ (1 pav., 2 pav.), tai yra tais atvejais, kai nepriklausomų dydžių maksimumo vidurkiai yra žinomi, o taip pat, kai $n = 50$ (3 pav., 4 pav.).



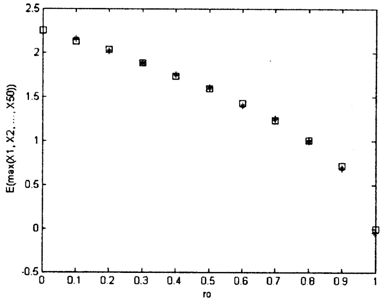
1 pav. μ_n^*, $\mu_n\sqrt{1-\rho}$, kai $n = 4$, $n = 5$.



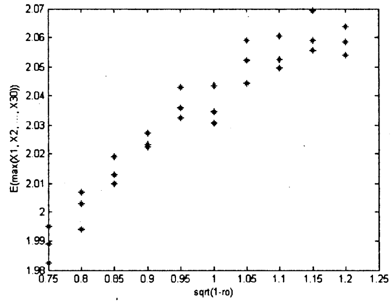
2 pav. μ_n^*, $\mu_n\sqrt{1-\rho}$, kai $n = 6$, $n = 7$.



3 pav. μ_n^*, $\mu_n\sqrt{1-\rho}$, kai $n = 50$.



4 pav. μ_n^*, $\mu_n\sqrt{1-\rho}$, kai $n = 50$.



5 pav. μ_n^*, kai $n = 30$.

Gauti rezultatai iliustruoja prielaidos teisingumą, tačiau šis sąryšis yra įrodytas tik dviejų ir trijų atsitiktiniams dydžių atveju.

Panaudojant transformacijų matricą, kuri nepriklausomų normaliųjų dydžių seką pakeičia į priklausomų, nagrinėjamąjį sąryšį (1) galima užrašyti kita forma:

$$\mathrm{E}\bigl(\max(V^T(X_1, X_2, \dots, X_n)^T)\bigr) = \mathrm{E}\bigl(\max(\sqrt{1-\rho}(X_1, X_2, \dots, X_n)^T)\bigr). \quad (2)$$

Taigi, modeliavimo rezultatai rodo, kad atsitiktiniai dydžiai $\max(V^T(X_1, X_2, \dots, X_n)^T)$ ir $\max(\sqrt{1-\rho}(X_1, X_2, \dots, X_n)^T)$ yra ekvivalentūs vidurkio prasme.

Taip pat tyrimai buvo atlikti su kitokios formos koreliacine matrica. Praktikoje dažnai pasitaiko atvejis, kai ryšys egzistuoja tik tarp kaimyninių atsitiktinių dydžių, o likusieji yra nepriklausomi. Buvo modeliuojama pagal prieš tai aprašytą schemą, tačiau su kita koreliacine matrica, kurios elementai yra apibrėžiami taip:

$$c_{ij} = \begin{cases} 1, & \text{kai } i = j; \\ \rho, & \text{kai } i = j \pm 1; \\ 0, & \text{kitais atvejais}. \end{cases}$$

Tikrinome, ar, taip pat kaip ir normaliųjų dydžių sekoms su vienodais koreliacijos koeficientais, galioja maksimumo vidurkio tiesinė priklausomybė nuo $\sqrt{1-\rho}$. Deja, pagal grafikus matyti, kad labai tikėtina, jog ši priklausomybė nėra tiesinė. Kaip pavyzdį pateikiame 30-ties kintamųjų sekos tyrimo rezultatus (5 pav.). Taip pat nėra aišku ar šiuo atveju galimas tiesioginis sąryšis su nepriklausomų normaliųjų dydžių maksimumo vidurkiu.

Literatūra

1. A. Burauskaitė, A. Aksomaitis, Moments of extremes of normally distributed values, *Liet. matem. rink.*, **43** (spec. nr.), 677–681 (2003).
2. Mathsoft resaurses, extreme value constants.
   `http://www.mathsoft.com/mathresources/constants/diskrete structures /article/0,,2202,00.html.`

SUMMARY

A. Burauskaitė, A. Aksomaitis. A mean of dependent normal variables maximum

A mean value of normal sequence maximum is analyzed. In a case of standard normal variables, for $n \leqslant 5$ (n – a length of sequence), there are formulas to express every order moments of extremes using elementary functions. For longer sequences it is not possible. Also there are analogous formulas for a mean value of two and three dependent normal variables [1]. In this work we study a relation between mean values of dependent and independent variables maxima. It is shown that there is a possibility to calculate a mean value of dependent normal variables maximum using the result of independent case. To test the relation we use computer simulation.

Keywords: extreme value, mean maximum, normal sequence.