ŠTATISTICKÁ DEFINÍCIA PRAVDEPODOBNOSTI V RÁMCI SIMULÁCIE HODU KOCKOU V PROSTREDÍ MICROSOFT EXCEL

Abstrakt

Pri realizácii dostatočne veľkého počtu opakovaní náhodného pokusu relatívna početnosť výskytu danej udalosti A konverguje k určitému číslu, ktoré nazývame pravdepodobnosť nastatia udalosti A. V prípade určenia tejto pravdepodobnosti napríklad pomocou klasickej definície, môžeme túto informáciu využiť na predikciu početnosti výskytu udalosti A v predpokladanej sérii nezávislých pokusov. O správnosti týchto tvrdení sa môžeme presvedčiť napríklad pri simulácii náhodného pokusu, ktorým bude v tomto príspevku hod kockou. Simulovaný hod kockou budeme realizovať v prostredí Microsoft Excel pomocou rovnomerného rozdelenia na základe inverznej transformačnej metódy. Na generovanie hodnôt 1, 2, 3, 4, 5, 6 využijeme dostupné excelovské funkcie RAND() a INT(). Na základe štatistickej definície pravdepodobnosti je možná aj percentuálna predikcia počtu výskytu udalosti A za predpokladu dostatočne veľkého počtu opakovaní pokusu. V tomto príspevku sa nebudeme zaoberať určením počtu pokusov potrebného na dosiahnutie požadovanej presnosti týkajúcej sa relatívnej početnosti a pravdepodobnosti $P(A)$.

Kľúčové slová

relatívna početnosť, počet pokusov, rovnomerné rozdelenie, simulácia, pravdepodobnosť, predikcia,

1 ŠTATISTICKÁ DEFINÍCIA PRAVDEPODOBNOSTI

Uvažujeme o realizácii náhodného pokusu, ktorý vykonáme n- krát. Počet nastatí udalosti A v tejto sérii pokusov označíme k_n a nazývame ho ***početnosťou*** nastatia udalosti A. Potom číslo f_n, ktoré určíme na základe vzťahu

$$f_n = \frac{k_n}{n}, \tag{1}$$

nazývame ***relatívnou početnosťou*** nastatia udalosti A v sérii n pokusov.

Náhodný pokus budeme opakovať dostatočný počet krát. Zvyšovaním tohto počtu opakovaní a sledovaním relatívnej početnosti nastatia náhodnej udalosti A by sme na základe empirických skúseností mohli konštatovať, že sa s rastúcim počtom pokusov relatívna početnosť f_n približuje k určitému číslu. Toto číslo reprezentuje **pravdepodobnosť nastatia náhodnej udalosti** A v danom pokuse, označujeme $P(A)$. Predchádzajúce tvrdenie môžeme vyjadriť nasledujúcim vzťahom

$$\lim_{n\to\infty} f_n = P(A) \tag{2}$$

Keďže relatívna početnosť je číslo z intervalu $\langle 0;\ 1\rangle$, tak aj $P(A)$ je z intervalu $\langle 0;\ 1\rangle$. Na základe vzťahu

$$f_n = \frac{k_n}{n} \doteq P(A), \tag{3}$$

ktorý platí pre $n \to \infty$, dostaneme pre percentuálny výskyt w udalosti A v rámci realizácie n opakovaných pokusov vzťah

$$w = \frac{k_n}{n} \cdot 100\,\% \doteq P(A) \cdot 100\,\%, \tag{4}$$

kde $P(A)$ je pravdepodobnosť nastatia udalosti A určená napríklad pomocou klasickej definície pravdepodobnosti. Tento vzťah môžeme interpretovať tak, že v rámci $n \to \infty$ nezávislých opakovaní náhodného pokusu sa udalosť A v nich vyskytne v $P(A) \cdot 100\,\%$ pokusov. Pre jednoduchosť uvedieme, že ak je počet pokusov rádovo v státisícoch, môžeme uvedenú podmienku považovať za splnenú.

2 SIMULÁCIA HODU KOCKOU POMOCOU ROVNOMERNÉHO ROZDELENIA V PROSTREDÍ MICROSOFT EXCEL

Rovnomerné rozdelenie $X \sim Ro(a;b)$

Náhodná premenná X, ktorá nadobúda hodnoty z intervalu $(a;b)$, kde $a, b \in R, a < b,$ má **rovnomerné rozdelenie** pravdepodobnosti práve vtedy, ak jej hustota pravdepodobnosti má tvar

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a,b) \\ 0 & x \notin (a,b) \end{cases} \tag{5}$$

Distribučná funkcia $F(x)$ náhodnej premennej $X \sim Ro(a;b)$ má tvar

$$F(x) = \begin{cases} 0 & x \le a \\ \frac{x-a}{b-a} & x \in (a,b) \\ 1 & x \ge b \end{cases} \tag{6}$$

Na základe inverznej transformačnej metódy je možné generovať hodnoty náhodnej premennej $X \sim Ro(a;b)$ z intervalu (a,b) podľa vzťahu

$$x = p \cdot (b-a) + a\text{, kde } p \in (0;1). \tag{7}$$

Na generovanie hodnôt $p \in (0;1)$ použijeme v prostredí Microsoft Excel funkciu RAND( ). (Ak chceme, aby funkcia RAND vygenerovala náhodné číslo, ale nechceme, aby sa čísla menili pri každom prepočte bunky v hárku, môžeme do riadka vzorcov zadať výraz = RAND( ) a potom stlačením klávesu F9 zameniť vzorec za náhodné číslo.) V prípade simulácie hodu kockou t. j. náhodného generovania hodnôt 1, 2, 3, 4, 5, 6 je potrebné nastaviť parametre rovnomerného rozdelenia a, b nasledovne

$$X \sim Ro(1;7),$$

pričom musíme zabezpečiť, aby sa vygenerovaná hodnota zaokrúhlila na celé číslo. V prostredí Microsoft Excel využijeme pre tento účel funkciu INT( ).

Na základe uvedeného potom pre náhodné generovanie počtu bodov na kocke dostaneme vzťah

$$x = INT\left(RAND() \cdot 6 + 1\right).$$

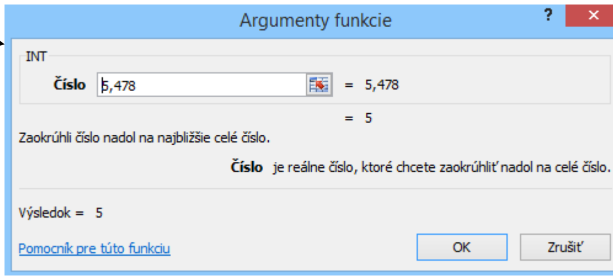
Pravdepodobnosť, že padne na kocke 6 bodov pri hode kockou je na základe klasickej definície pravdepodobnosti rovná $\frac{1}{6}$. Realizáciou simulácie hodu kockou si ukážeme, že relatívna početnosť padnutia 6 bodov so zvyšujúcim sa počtom hodov konverguje k číslu $\frac{1}{6} = 0{,}1\overline{6}$. Pri dodržaní podmienky o počte opakovaní toho náhodného pokusu ($n \to \infty$) je možná predikcia o percentuálnom výskyte padnutia 6 bodov v rámci daného počtu pokusov n. Na základe vzťahu (4) dostaneme

$$P(A) \cdot 100\,\% = 0{,}1\overline{6} \cdot 100\,\% = 16{,}\overline{6}\,\%,$$

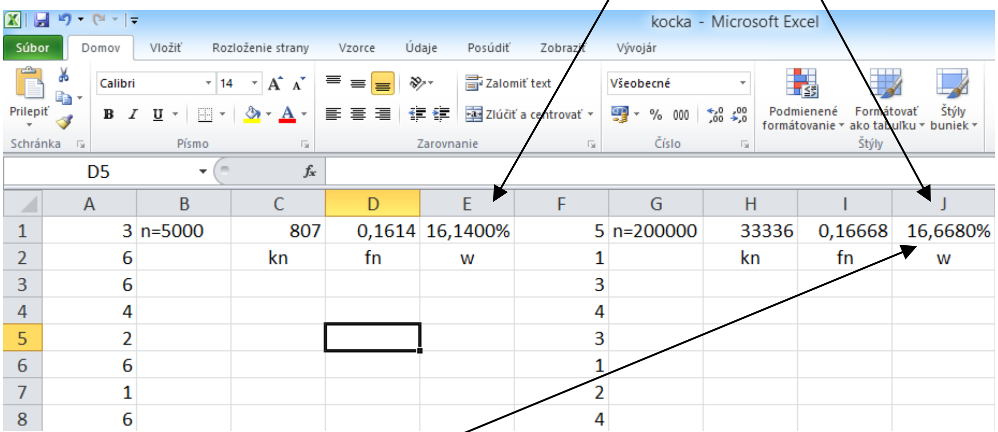
čo znamená, že v dostatočne veľkom počte opakovaní hodu kockou padnutie 6 bodov nastane v $16{,}\overline{6}\,\%$ pokusov. Túto predikciu si overíme v praktickej realizácii.

3 PRAKTICKÁ REALIZÁCIA V PROSTREDÍ MICROSOFT EXCEL

V prostredí Microsoft Excel budeme simulovať náhodne hod kockou pre počet pokusov $n = 5000$ a $n = 200\,000$. Do bunky A1 vložíme vzorec $= INT\left(RAND(\;)*6+1\right)$, ktorý skopírujeme až po bunku A5000, resp. A200000 vrátane. Súčasťou vzorca je spomínaná funkcia INT( ).

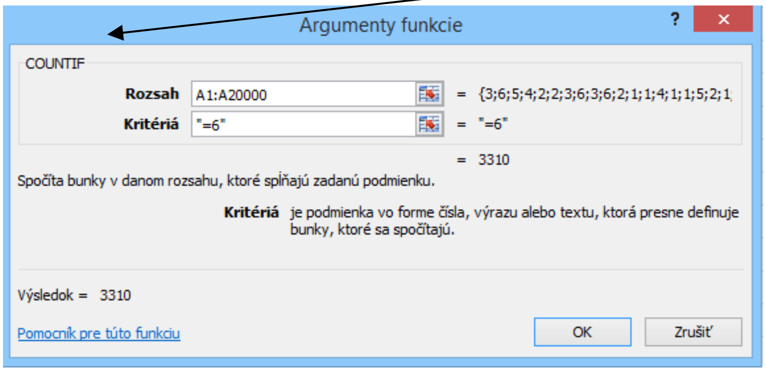


Po nasimulovaní daného počtu pokusov $n = 5000$ do stĺpca A a $n = 200\,000$ do stĺpca F sme do bunky C1, resp. H1 určili početnosť k_n padnutia 6 bodov na kocke. Do bunky D1, resp. I1 sme vypočítali relatívnu početnosť f_n. V bunke E1, resp. J1 sme určili počet padnutia 6 bodov v % v rámci n pokusov.



Pri počte pokusov $n = 200\,000$ sa potvrdil predikovaný percentuálny počet padnutia 6 bodov na kocke ($16{,}\overline{6}\,\%$), ktorý sme určili na základe informácie o pravdepodobnosti ($\frac{1}{6} = 0{,}1\overline{6}$).

Pri určovaní početnosti k_n padnutia 6 bodov na kocke sme využili funkciu COUNTIF.



Záver

Pri realizácii dostatočne veľkého počtu opakovaní náhodného pokusu relatívna početnosť f_n výskytu danej udalosti A konverguje k číslu $0{,}1\overline{6}$. Prezentovaná metodológia by mala prispieť k lepšiemu pochopeniu štatistickej definície pravdepodobnosti a interpretácii pravdepodobnosti nastatia náhodnej udalosti.

Použitá literatúra

1 HORÁKOVÁ, G.- HUŤKA,V. 2010. *Teória pravdepodobnosti.* Bratislava: EKONÓM, 2010. 238 s. ISBN 978-80-225-2888-7

2 WALKENBACH, J. 2010. *Excel® 2010. Power Programming with VBA*. New Jersey. John Wiley & Sons,Inc. 2010.1052s. ISBN 978-0-470-47535-5

Kontaktné údaje

Mgr. Vladimír Mucha, PhD., Katedra matematiky a aktuárstva, Fakulta hospodárskej informatiky, Ekonomická univerzita v Bratislave, Dolnozemská cesta 1, 852 35 Bratislava, tel. +421 2/672 95 810, e-mail: vladimir.mucha@euba.sk