Rušenje rekorda:
o jednom problemu iz teorije vjerojatnosti

Dario Lovreković,
Goran Trupčević
i Anđa Valent, Zagreb

1. Uvod

Problem kojim se bavimo u ovom članku ukratko je opisan pitanjem: *Koliko se često ruše rekordi?* Pretpostavimo da vršimo n nezavisnih mjerenja neke pojave (npr. prosječne godišnje temperature ili količine padalina na nekom području) ili da jednostavno n puta biramo slučajan broj. Rekordnom vrijednošću smatramo onu koja je veća od svih dotad izmjerenih vrijednosti (naravno da na isti način možemo promatrati i rekordno male vrijednosti). Koliki broj rekorda možemo očekivati?

Naglasak rada nije na samom rješenju problema, već na različitim pristupima i stupnjevima rješavanja. Tako ćemo problem najprije riješiti na intuitivan, matematički neformalan način. U idućem koraku želimo potaknuti učenike da svoje rezultate eksperimentalno provjere, a za što su problemi iz područja vjerojatnosti osobito pogodni. I konačno, ostaje nam pitanje dokaza rješenja. Svoj ćemo rezultat dokazati na dva načina. Prvi dokaz je kombinatoran i smatramo ga razumljivim učenicima upoznatim sa osnovnim kombinatornim principima i tehnikom deriviranja. Drugi dokaz u potpunosti slijedi naše intuitivno rješenje i daje mu matematičku pozadinu, ali naglašavamo da nadilazi okvire srednjoškolskog znanja. U prvom čitanju može se i preskočiti i treba ga gledati kao motivaciju za neke daljnje teme iz područja vjerojatnosti.

Pa započnimo! Pretpostavimo da na slučajan način n puta biramo realan broj iz nekog zadanog intervala. Smatrat ćemo razumnim za pretpostaviti da je vjerojatnost višestrukog pojavljivanja iste vrijednosti jednaka 0. Što možemo unaprijed, prije nego što odaberemo ijedan broj, zaključiti o vjerojatnosti da u nekom trenutku dobijemo rekord? Najprije, u prvom ćemo odabiru bez sumnje dobiti rekord, pa ako biramo samo jedan broj, očekivani broj rekorda je 1. Za drugi broj koji ćemo birati unaprijed znamo da će biti rekord s vjerojatnosti $\frac{1}{2}$, jer je jednako vjerojatno da će biti veći ili manji od prvog. Dakle, ako biramo dva broja, očekivani broj rekorda je $1+\frac{1}{2}$. Isto tako, za treći broj koji ćemo birati možemo zaključiti da će biti rekord s vjerojatnosti $\frac{1}{3}$ jer ima jednaku šansu biti najveći kao i prva dva. Dakle, nakon tri odabira, očekivani broj rekorda je $1+\frac{1}{2}+\frac{1}{3}$. Nastavljajući na isti način zaključujemo da očekivani broj rekorda nakon n odabranih brojeva iznosi

$$H_n = 1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}.$$

2. Eksperimentalna provjera

Naše početno rješenje je intuitivno jasno, ali bez matematičke strogoće. Može li onda biti i točno? Prije pokušaja dokaza rezultat smo odlučili eksperimentalno provjeriti. U prvoj smo provjeri koristili program Microsoft Excel. S pomoću funkcije RAND koja generira slučajan realan broj u intervalu [0,1) dobili smo 100 takvih brojeva i zapisali broj rekorda, pri čemu smo zasebno mjerili broj rekordno velikih i broj rekordno malih vrijednosti. Pokus smo ponovili 1000 puta. Za prosječnu vrijednost za rekordno velike vrijednosti dobili smo 5.1879, a za rekordno male vrijednosti 5.1989. Kako je $H_{100} = 1+\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{100}\approx 5.1873$, prvu provjeru svog rezultata smatramo uspješnom.

U drugoj smo se provjeri odlučili na nešto drukčiji pristup, opisan u [4]. Za svaki od 12 mjeseci u godini promatrali smo prosječnu izmjerenu mjesečnu temperaturu u gradu Zagrebu, na meteorološkoj postaji Zagreb-Grič, za posljednjih $n = 50$ godina, kao što je i prikazano u tablici 1. Prebrojavali smo izmjerene rekordne vrijednosti i pokus ponovili 24 puta, jer smo za svaki mjesec u godini gledali broj rekordno velikih i broj rekordno malih vrijednosti. Tako primjerice za mjesec veljaču imamo 3 rekordno izmjerene najveće prosječne temperature (za godine 1963., 1964. i 1966.) i 1 izmjerenu rekordno malu prosječnu temperaturu (za godinu 1963.) Prosječan broj ta 24 rekorda iznosi 4.4583 što je blizu našem izračunanom očekivanom broju rekorda $H_{50}\approx 4.4992$.

Naravno da je upitno koliko se mjerenja klimatskih veličina mogu smatrati nezavisnim mjerenjima, pogotovo za vremenska razdoblja duža od 100 godina. Primijetimo i da se ponekad ponavljaju iste vrijednosti, a razlog je ponajprije taj što se prema uobičajenoj metodologiji prosječne temperature zaokružuju na jednu decimalu. Želimo napomenuti da su naši podatci dobiveni korištenjem ECA&D baze podataka [1], iz koje je moguće dobiti razne dnevno izmjerene podatke o klimatskim veličinama za mnogobrojne klimatske postaje u Europi. Tako su primjerice za klimatsku postaju Zagreb-Grič dostupni podatci od 1861. godine. Moguće je rezultat iz ovog rada, ali naravno i neke druge rezultate iz vjerojatnosti, testirati i na drugim izmjerenim klimatskim veličinama ili drugim meteorološkim postajama.

3. Kombinatorni izračun očekivanog broja rekorda

U ovom ćemo odjeljku problemu određivanja očekivanog broja rekorda pristupiti na drukčiji način. Polazimo od pretpostavke da ćemo u n mjerenja dobiti sve različite rezultate, pa će broj rekorda zapravo samo ovisiti o poretku dobivenih veličina. Zbog toga svaki skup od n mjerenja identificiramo s permutacijom skupa $\{1,2,\ldots,n\}$. Ukupan broj takvih permutacija je $n!$. Cilj nam je odrediti broj takvih permutacija s točno k rekorda, za $k = 1,2,\ldots,n$. Zašto? Zato što očekivani broj rekorda nije ništa drugo nego prosječan broj rekorda koji ćemo onda izračunati formulom

$$\begin{aligned}\overline{R}(n) &= \frac{R(n,1)\cdot 1 + R(n,2)\cdot 2+\ldots+R(n,n)\cdot n}{n!}\\ &= \frac{1}{n!}\sum_{k=1}^{n} k\cdot R(n,k)\end{aligned}$$

pri čemu smo uveli oznaku $R(n,k)$ za broj permutacija skupa $\{1,2,\ldots,n\}$ s točno k rekorda.

Primjer. Za $n = 3$ postoji točno $3! = 6$ permutacija skupa $\{1,2,3\}$. Među njima permutacije (312) i (321) imaju točno jedan rekord, pa je $R(3,1) = 2$. Permutacije (132), (213) i (231) imaju točno dva rekorda, pa je $R(3,2) = 3$, a permutacija (123) ima točno tri rekorda, pa je $R(3,3) = 1$. Prosječan broj rekorda sada je

$$\overline{R}(3) = \frac{2\cdot 1+3\cdot 2+1\cdot 3}{6} = \frac{11}{6}.$$

S druge je strane $H_3 = 1+\frac{1}{2}+\frac{1}{3} = \frac{11}{6} = \overline{R}(3)$.

Tablica 1. Prosječne mjesečne temperature Zagreb–Grič, 1963. – 2012.

	1	2	3	4	5	6	7	8	9	10	11	12
1963.	−4.7	−2.1	5.5	13	16.1	20.6	23	21.5	17.9	11.5	12	−2.5
1964.	−4.8	2	4.6	12.8	16.2	21.4	21.6	19.6	16.7	11.4	8.6	1.4
1965.	2.5	−0.4	7.4	10.5	14.9	19.6	20.9	18.8	17.1	11.5	4.7	4
1966.	−1.7	8.9	6.5	14	16.2	20.5	19.9	19.5	17.5	15.5	4.8	2.8
1967.	−0.2	4.3	8.7	10.9	17.1	18.8	23.1	21.1	17.9	13.3	7	0.9
1968.	−0.2	3.3	8.2	14.2	16.7	19.5	21.2	19.4	16.2	12.7	7	−0.9
1969.	−0.9	1.2	5	11.5	18.1	18.7	21.1	19.2	17.3	11.5	9.8	−1.8
1970.	0.2	2.2	5.5	10.7	14.1	20.7	20.9	20.8	16.9	10.9	9.5	1.3
1971.	0.8	4.6	3.8	13	17.6	18.7	22.2	22	14.8	11.4	6.2	2
1972.	−0.6	5	9.7	11.6	15.4	20	21.1	19.8	14.1	10.5	6.5	2
1973.	0.3	3.4	7	9.6	17.5	19.7	21.3	21	17.4	10.2	4.8	2.6
1974.	2.2	7.2	9.3	11.3	15	19	21	22.1	16.8	7.7	7.3	4.6
1975.	5.6	3.2	8.4	11.9	17.6	18.4	21.3	20	19.3	10.9	5.2	2.1
1976.	2.5	1.7	4.1	11.8	16.5	19.2	22.1	17.7	15.5	11.5	8.4	1.8
1977.	3.2	6.3	11	10.4	16.4	20	20.7	20.2	14.7	12.5	6.9	0.9
1978.	1.4	2.1	8.4	10.6	14	18.8	19.5	19	15.8	11.1	2.2	2.1
1979.	−0.3	3.1	9.1	10.9	17.5	21.5	19.9	19.7	16.8	10.6	7	5.4
1980.	−0.7	4.2	6.9	9.4	13.5	18.9	20.2	20.7	16.5	11.4	5.1	0.6
1981.	−0.5	1.8	10.3	12.2	16.1	19.9	21.3	20.9	17.2	13.6	6.2	2.1
1982.	−0.9	0.7	6.9	9.4	17	20.6	22.2	20.7	19.5	12.5	7.1	5.1
1983.	4.3	0.7	8.7	14.1	17.8	19.3	23.8	21.4	17.3	12.1	3.9	2.3
1984.	1.5	1.3	6.2	11.3	14.5	18.4	19.5	19.9	16.8	12.8	7	1.9
1985.	−3.4	−1.3	6.1	11.7	17.5	17.6	22	21.5	17.8	11.5	4.2	6.4
1986.	2.2	−2.1	4.5	12.8	19.2	18.3	20.5	21.9	16.1	11.6	7.1	0.7
1987.	−1.8	3	2.7	12.6	14.8	19.5	23.2	19.8	20.6	12.8	6.1	2.7
1988.	5	5	6.8	11.5	16.8	18.8	23.3	22	17.2	11.5	2.6	2.6
1989.	0.2	5.4	11	12.9	15.7	17.7	21.8	20.8	16.9	12.3	5.5	4.3
1990.	1.1	8.6	11.3	11.2	17.9	19.1	21.6	22.2	15.8	12.5	6.8	2.2
1991.	2.8	0.2	10.3	10.5	13.3	19.8	23	21.4	18.3	10.2	7.1	0.2
1992.	2.9	5.8	8.3	12.8	17.3	20.2	22.5	25.8	18.7	11.2	8	2.3
1993.	3.1	2.3	7.1	13	19.3	20.6	21.7	22.4	16.5	12.9	2.9	3.3
1994.	5.2	4	11.7	12.3	17	20.4	23.9	23.4	19.3	10.2	9	3.4
1995.	1.9	7.4	6.7	13.1	16.2	18.6	23.8	20.4	15.6	13.3	5.8	2
1996.	0	0.7	4.6	12.1	17.6	21.1	20.2	20.8	13.8	12.6	8.9	−0.3
1997.	−0.3	6	8.3	9.3	17.8	20.6	21.4	21.6	17.9	10.4	6.7	4
1998.	5.1	7.7	6.5	13.6	17.1	21.3	22.4	22.5	16.7	12.6	4.8	−1.5
1999.	1.7	3.2	9.6	13.5	17.4	20.7	22.4	21.6	19.7	13	4.6	2.9
2000.	−0.2	6.5	8.9	15.4	18.5	22.3	21.5	24.4	17.8	14.2	10.5	5.9
2001.	4.6	5.9	10.9	11.8	18.7	19.1	22.9	23.6	15	15.4	4.7	−0.7
2002.	2.2	7.6	10.4	11.4	18.9	21.9	22.5	21.6	16.1	12.3	10.8	2.7
2003.	0.5	−0.1	8.9	11.9	20.4	24.5	23.6	25.8	17	10.1	9.4	3
2004.	0.9	4	6.7	12.5	15.6	19.6	21.7	21.7	17	13.9	7.7	2.9
2005.	2	−0.1	6.8	12.7	17.5	20.5	22.1	19.2	17.5	13.1	6	2.6
2006.	−0.1	2.8	6.4	13.5	16.7	21.2	24.7	19.3	18.8	14.6	9.9	5.2
2007.	7.5	8.1	10.2	15.9	18.8	22.7	23.8	22.1	15.5	11.2	6.1	1.4
2008.	3.8	6.8	8.3	13	18.2	21.6	22.8	22.5	16.5	14	8.8	4.3
2009.	0	4.1	8.4	15.8	19.2	20.2	23.2	23.6	19.7	12.7	9.3	4.1
2010.	0.3	3.6	7.8	13.3	16.7	21	24.1	21.7	15.9	10.6	9.9	1.7
2011.	3.1	2.9	8.7	14.9	18.1	21.6	22.9	24.3	21.4	11.7	4.3	4.8
2012.	3.8	−0.2	11.5	13.3	17.5	22.9	25	25.4	18.9	12.9	10	2.8

Kako bismo odredili brojeve $R(n,k)$, polazimo od sljedeće rekurzivne relacije.

Lema. Za $k > 0$ vrijedi

$$R(n,k) = R(n-1,k-1) + (n-1) \cdot R(n-1,k)$$

Dokaz. Skup permutacija skupa $\{1,2,\ldots,n\}$ s točno k rekorda rastavljamo na dva podskupa, ovisno o poziciji jedinice. Broj takvih permutacija s 1 na prvom mjestu jednak je $R(n-1,k-1)$ jer je tada 1 rekord, pa u permutaciji preostalih $n-1$ elemenata trebamo još $k-1$ rekorda. Drugi podskup čine permutacije u kojima je 1 na nekom drugom mjestu, njih ukupno $n-1$. Tada 1 nije rekord, pa za svako takvo mjesto, u permutaciji preostalih $n-1$ elemenata trebamo točno k rekorda. Broj takvih permutacija jednak je dakle $(n-1) \cdot R(n-1,k)$ iz čega slijedi tvrdnja leme. ■

U nastavku izračuna brojeva $R(n,k)$ uvodimo funkciju $f_n(x)$ formulom

$$f_n(x) = \sum_{k=1}^{n} R(n,k) \cdot x^k.$$

Tako je primjerice $f_1(x) = R(1,1) \cdot x = x$.

Zašto tražimo takvu funkciju? Deriviranjem funkcije $f_n(x)$ dobivamo

$$f_n'(x) = \sum_{k=1}^{n} k \cdot R(n,k) \cdot x^{k-1}$$

to jest

$$f_n'(1) = \sum_{k=1}^{n} k \cdot R(n,k)$$

pa će očekivani broj rekorda biti jednak

$$\overline{R}(n) = \frac{1}{n!} \sum_{k=1}^{n} k \cdot R(n,k) = \frac{1}{n!} f_n'(1).$$

Sada ćemo odrediti eksplicitnu formulu za funkciju $f_n(x)$. Iz definicije funkcije $f_n(x)$ i rekurzije iz leme dobivamo da vrijedi

$$f_n(x) = x \cdot f_{n-1}(x) + (n-1) \cdot f_{n-1}(x),$$

a odavde

$$f_n(x) = (x+n-1) f_{n-1}(x).$$

Višestrukim korištenjem posljednje formule dobivamo

$$f_n(x) = (x+n-1) \cdot (x+n-2) \cdot \ldots \cdot (x+1) \cdot f_1(x)$$

to jest

$$f_n(x) = (x+n-2) \cdot \ldots \cdot (x+1) \cdot x.$$

Radi lakšeg deriviranja, gornji izraz logaritmiramo i dobivamo

$$\ln f_n(x) = \ln x + \ln(x+1) + \ldots + \ln(x+n-1).$$

Deriviranjem slijedi

$$\frac{1}{f_n(x)} \cdot f_n'(x) = \frac{1}{x} + \frac{1}{x+1} + \ldots + \frac{1}{x+n-1}.$$

Uvrstimo li $x = 1$ dobivamo

$$\frac{1}{n!} f_n'(1) = 1 + \frac{1}{2} + \ldots + \frac{1}{n} = H_n$$

pa konačno dobivamo

$$\overline{R}(n) = \frac{1}{n!} f_n'(1) = H_n,$$

što smo i željeli pokazati.

Osvrnimo se kratko na metodu izračuna. Problem smo riješili standardnim kombinatornim metodama – s pomoću rekurzivne relacije i, iako to nije posebno napomenuto, s pomoću funkcije izvodnice. Za obje metode, više detalja i puno zadataka prikladnih srednjoškolskoj razini može se pronaći u [2].

Napomenimo i da se brojevi $(-1)^{n-k} R(n,k)$ nazivaju Stirlingovi brojevi prve vrste. O njima se više može pročitati u [6].

4. Osvrt na početno rješenje problema

U prethodnom smo odjeljku uspješno dokazali naše početno neformalno rješenje, ali bitno drukčijim pristupom. Kao što smo najavili u uvodu, rad ćemo završiti dokazom koji prati naš prvi način razmišljanja. Nije nam cilj objasniti sve detalje, već zaokružiti problem i motivirati neke nove teme iz područja vjerojatnosti ([3], [5]) Radi jednostavnosti, promatrat ćemo slučaj $n = 3$ i pretpostaviti da biramo tri slučajna broja X_1, X_2, X_3 iz intervala $[0, 1]$. Mi dakle vršimo pokus čiji ishod nije unaprijed određen. Skup svih mogućih ishoda je skup $\Omega = [0,1] \times [0,1] \times [0,1] \subset \mathbf{R}^3$. Za $k = 1, 2, 3$ uvodimo funkcije Y_k čije vrijednosti ovise o ishodu našeg pokusa

$$Y_k = \begin{cases} 1, & \text{ako je } X_k \text{ rekord} \\ 0, & \text{ako } X_k \text{ nije rekord.} \end{cases}$$

Ovakve funkcije koje slučajnom događaju pridružuju neki realni broj, nazivaju se diskretne slučajne varijable. Ako je Y neka diskretna slučajna varijabla koja poprima vrijednosti $y_1, y_2, \ldots, y_n$ i ako je p_i vjerojatnost da Y poprimi vrijednost y_i, onda se očekivana vrijednost od Y označava s $E(Y)$ i računa kao $E(Y) = p_1 y_1 + \ldots + p_n y_n$. Prema tomu, očekivane vrijednosti slučajnih varijabli Y_k iznose

$$\begin{aligned} E(Y_k) &= p(X_k \text{ rekord}) \cdot 1 + p(X_k \text{ nije rekord}) \cdot 0 \\ &= p(X_k \text{ rekord}). \end{aligned}$$

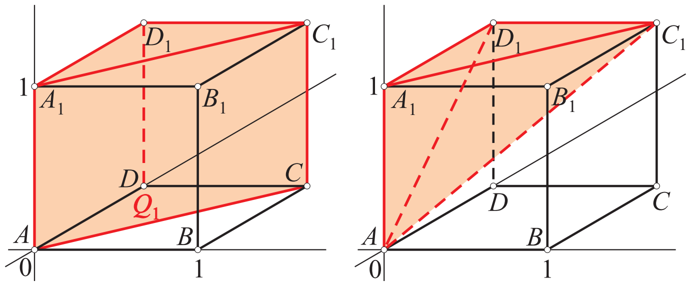
Kako nas zanima ukupan broj rekorda, uvodimo novu slučajnu varijablu $Y = Y_1 + Y_2 + Y_3$. Matematičko očekivanje ima svojstvo linearnosti, pa očekivani broj rekorda iznosi

$$E(Y) = E(Y_1) + E(Y_2) + E(Y_3).$$

Sada, baš kao i u uvodu najprije očito vrijedi $E(Y_1) = p(X_1 \text{ rekord}) = 1$. Vjerojatnosti da su X_2 i X_3 rekordi, izračunat ćemo s pomoću geometrijske vjerojatnosti. Skup svih mogućih ishoda $\Omega = [0,1] \times [0,1] \times [0,1]$, te skup svih ishoda takvih da je X_2, odnosno X_3, rekord, možemo prikazati u koordinatnom sustavu, kako je prikazano na slici 1. Tražene vjerojatnosti računamo s pomoću formula

$$p(X_i \text{ rekord}) = \frac{\text{obujam}(X_i \text{ rekord})}{\text{obujam}(\Omega)}.$$

Skupu $\{X_2 \text{ rekord}\}$ odgovara prizma $ABCA_1B_1C_1$, a skupu $\{X_3 \text{ rekord}\}$ piramida $A_1C_1D_1A$.



Slika 1. Skup Ω svih mogućih ishoda u koordinatnom sustavu prikazan je kockom $ABCDA_1B_1C_1D_1$.

Primijetimo da je Ω jedinična kocka volumena 1, skup $\{X_2 \text{ rekord}\}$ je prizma volumena $\frac{1}{2}$, a skup $\{X_3 \text{ rekord}\}$ je piramida volumena $\frac{1}{3}$. Stoga je $E(Y_2) = p(X_2 \text{ rekord}) = \frac{1}{2}$, odnosno $E(Y_3) = p(X_3 \text{ rekord}) = \frac{1}{3}$. Sada je

$$E(Y) = E(Y_1) + E(Y_2) + E(Y_3) = 1 + \frac{1}{2} + \frac{1}{3} = H_3,$$

upravo što smo željeli i pokazati.

Za kraj, prisjetimo se da je red $1 + \frac{1}{2} + \ldots + \frac{1}{n}$ divergentan što znači da kad $n \to \infty$, broj rekorda također neograničeno raste. Štoviše, za velike vrijednosti od n vrijedi da je $H_n \approx \ln n + \gamma$, gdje je $\gamma \approx 0.5772$ Euler-Mascheronijeva konstanta. To znači da za npr. $n = 1000$ bez izračuna H_{1000} odmah možemo procijeniti da će očekivani broj rekorda iznositi $\approx \ln 1000 + 0.5722 \approx 7.485$.

LITERATURA

1/ www.ecad.eu/

2/ M. Cvitković, *Kombinatorika*, Element, Zagreb, 2007.

3/ N. Elezović, *Diskretna vjerojatnost*, Element, Zagreb, 2011.

4/ N. Glick, *Breaking Records and Breaking Boards*, American Mathematical Monthly 85(1), 2–26, 1978.

5/ N. Sarapa, *Teorija vjerojatnosti*, Školska knjiga, Zagreb, 1992.

6/ D. Veljan, *Kombinatorika s teorijom grafova*, Školska knjiga, Zagreb, 1989.