Matematičkim očekivanjem do kombinatoričkih identiteta

Vedran Krčadinac[1]

Sažetak

U enumerativnoj kombinatorici obično najprirodnijim smatramo dokaze u kojima se direktno prebrojava elemente nekog konačnog skupa ili uspostavlja bijekcija između dvaju skupova. Pokazat ćemo da zanimljive kombinatoričke identitete možemo izvesti slučajnim biranjem elemenata konačnog skupa i računanjem matematičkog očekivanja odgovarajućih slučajnih varijabli. Prednost ovog pristupa je što primjenom iste ideje u različitim situacijama otkrivamo nove identitete. Možemo ih dokazati i prebrojavanjem, ali za to unaprijed trebamo znati identitete koje dokazujemo.

1. Uvod

U kombinatorici se često pojavljuju identiteti poput ovih:

$$\sum_{i=0}^{n} \binom{n}{i} = 2^n, \tag{1}$$

$$\sum_{i=0}^{n} i \cdot \binom{n}{i} = n \cdot 2^{n-1}. \tag{2}$$

Možemo ih dokazati metodom dvostrukog prebrojavanja. Ideja je interpretirati formule na lijevoj i na desnoj strani kao broj elemenata istog skupa. Na primjer, formula 2^n na desnoj strani identiteta (1) je broj podskupova n-članog skupa $N = \{1, ..., n\}$. Na lijevoj strani imamo binomne koeficijente $\binom{n}{i}$ koji prebrojavaju i-člane podskupove od N. Sumiranjem po $i = 0, ..., n$ dobivamo ukupan broj podskupova i zato je lijeva strana jednaka desnoj.

Za identitet (2) prebrojavamo parove (S, x), pri čemu je $S \subseteq N$, a $x \in S$ jedan njegov istaknuti element. Desnu stranu dobivamo tako da prvo biramo $x \in N$, što možemo na n načina, a zatim ga dopunimo do podskupa S izborom preostalih ele-

[1]Vedran Krčadinac, PMF – Matematički odsjek, Sveučilište u Zagrebu

menata. Za to imamo 2^{n-1} mogućnosti, pa je broj parova (S, x) jednak $n \cdot 2^{n-1}$. Na lijevoj strani prvo biramo i-člani podskup S na $\binom{n}{i}$ načina, a zatim istaknuti element $x \in S$ na i načina. Ukupan broj parova dobivamo množenjem $i \cdot \binom{n}{i}$ i sumiranjem po $i = 0, \ldots, n$.

2. Vjerojatnosna metoda

Druga metoda za dokazati identitet (2) je slučajno biranje podskupa $S \subseteq N$. Pretpostavimo da je izbor svakog od 2^n podskupova jednako vjerojatan (kažemo da S biramo uniformno). Promotrimo koliki je očekivani broj elemenata od S, tj. matematičko očekivanje slučajne varijable $X = |S|$.

Ako slučajna varijabla X poprima konačno mnogo vrijednosti $x_1, \ldots, x_k \in \mathbb{R}$ s vjerojatnostima $p_1, \ldots, p_k$, matematičko očekivanje od X definiramo kao sumu

$$E(X) = \sum_{i=1}^{k} x_i \cdot p_i = \sum_{i=1}^{k} x_i \cdot P(X = x_i).$$

Smisao definicije je „prosječna vrijednost“ slučajne varijable. Prema zakonu velikih brojeva, aritmetička sredina vrijednosti koje poprima X s velikom je vjerojatnosti blizu $E(X)$. Aproksimacija je bolja što više puta ponavljamo pokus.

Naša slučajna varijabla $X = |S|$ poprima vrijednosti $0, \ldots, n$ s vjerojatnostima koje dobijemo dijeljenjem broja povoljnih mogućnosti s ukupnim brojem mogućnosti:

$$P(|S| = i) = \frac{\binom{n}{i}}{2^n}.$$

Matematičko očekivanje je stoga

$$E(X) = \sum_{i=0}^{n} i \cdot \frac{\binom{n}{i}}{2^n} = \frac{1}{2^n} \sum_{i=0}^{n} i \cdot \binom{n}{i}. \tag{3}$$

S druge stane, „prosječni“ slučajno izabrani podskup $S \subseteq N$ ima oko pola od ukupnog broja elemenata. To možemo obrazložiti prikazom X kao zbroja indikatorskih slučajnih varijabli. Za $i \in N$, varijabla X_i poprima vrijednost 1 ako je i sadržan u slučajno izabranom podskupu, a 0 inače:

$$X_i = \begin{cases} 1, & i \in S, \\ 0, & i \notin S. \end{cases}$$

Vrijedi $X = \sum_{i=1}^{n} X_i$, a matematička očekivanja indikatorskih varijabli su

$$E(X_i) = 1 \cdot P(i \in S) + 0 \cdot P(i \notin S) = P(i \in S) = \frac{1}{2}.$$

Zbog linearnosti matematičkog očekivanja vrijedi

$$E(X)=E\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}E(X_i)=\sum_{i=1}^{n}\frac{1}{2}=\frac{n}{2}.$$

Izjednačavanjem izraza (3) s $\frac{n}{2}$ dobivamo identitet (2).

Pionir primjene vjerojatnosti za dokazivanje teorema iz kombinatorike mađarski je matematičar Paul Erdős (1913. – 1996.). U knjizi [1] dani su vjerojatnosni dokazi mnogih važnih teorema. U njima se, osim matematičkog očekivanja, koriste i drugi pojmovi i rezultati iz teorije vjerojatnosti: varijanca, korelacija, uvjetna vjerojatnost i formula potpune vjerojatnosti, entropija i martingali. U ovom članku koristimo matematičko očekivanje za izvođenje kombinatoričkih identiteta na sličan način kao identiteta (2). Mijenjat ćemo objekte koje slučajno biramo i varijable kojima računamo očekivanje. Prikazivanjem kao zbroja indikatorskih varijabli i primjenom linearnosti očekivanja dobit ćemo razne zanimljive identitete.

3. Podskupovi

Pretpostavimo da uniformno biramo podskup od N sa zadanim brojem elemenata. Njegova je veličina zadana, ali ima smisla računati očekivani broj elemenata u presjeku ili uniji dvaju takvih podskupova. Biramo $A \subseteq N$ veličine $|A| = a$ uniformno među svim a-članim podskupovima od N. Na isti način nezavisno biramo $B \subseteq N$ veličine $|B| = b$. Slučajna varijabla $X = |A \cap B|$ poprima vrijednosti od 0 do $\min\{a, b\}$ s vjerojatnostima

$$P\left(|A\cap B|=i\right)=\frac{\binom{a}{i}\binom{n-a}{b-i}}{\binom{n}{b}}.$$

Ako je podskup A izabran (svejedno koji), za podskup B biramo i elemenata iz A koji će biti u presjeku i još $b - i$ elemenata iz $N\backslash A$. U brojniku je broj izbora B za koje je $|A \cap B| = i$, a u nazivniku je ukupan broj izbora B. Matematičko očekivanje je po definiciji

$$E(X)=\sum_{i\geq 0}i\cdot P\left(|A\cap B|=i\right)=\frac{1}{\binom{n}{b}}\cdot\sum_{i\geq 0}i\cdot\binom{a}{i}\binom{n-a}{b-i}. \tag{4}$$

S druge strane, X je suma po svim $i \in N$ indikatorskih slučajnih varijabli

$$X_i=\begin{cases}1, & i\in A\cap B,\\ 0, & i\notin A\cap B\end{cases}$$

s matematičkim očekivanjima

$$E(X_i)=P(i\in A\cap B)=P(i\in A)\cdot P(i\in B)=\frac{a}{n}\cdot\frac{b}{n}=\frac{ab}{n^2}.$$

Stoga je

$$E(X)=E\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}E(X_i)=\sum_{i=1}^{n}\frac{ab}{n^2}=\frac{ab}{n}.$$

Izjednačavanjem sa (4) i korištenjem svojstva apsorpcije $\binom{n}{b}=\frac{n}{b}\binom{n-1}{b-1}$ dobivamo identitet

$$\sum_{i\geq 0}i\cdot\binom{a}{i}\binom{n-a}{b-i}=\frac{ab}{n}\cdot\binom{n}{b}=a\cdot\binom{n-1}{b-1}. \tag{5}$$

I ovaj identitet možemo dokazati dvostrukim prebrojavanjem, slično kao (2). Pretpostavimo da su prvih a brojeva iz skupa N obojeni crveno, a preostalih $n-a$ brojeva plavo. Prebrojavamo parove (B, x) podskupa $B\subseteq N$ veličine b i jednog njegovog istaknutog elementa $x\in B$ koji je crven. Desnu stranu identiteta (5) dobivamo ako prvo biramo x, a lijevu stranu ako prvo biramo B.

Još jedan identitet dobivamo računanjem matematičkog očekivanja slučajne varijable $Y=|A\cup B|$. Ako je A izabran i biramo B tako da bude $|A\cup B|=i$, trebamo izabrati $i-a$ elemenata iz $N\backslash A$, a preostalih $b-(i-a)=a+b-i$ elemenata iz A. Zato je

$$E(Y)=\sum_{i\geq a}i\cdot P(|A\cup B|=i)=\sum_{i\geq a}i\cdot\frac{\binom{n-a}{i-a}\binom{a}{a+b-i}}{\binom{n}{b}}=\frac{1}{\binom{n}{b}}\cdot\sum_{i\geq a}i\cdot\binom{n-a}{i-a}\binom{a}{a+b-i}.$$

Indikatorske slučajne varijable

$$Y_i=\begin{cases}1, & i\in A\cup B,\\ 0, & i\notin A\cup B\end{cases}$$

imaju očekivanja

$$E(Y_i)=P(i\in A\cup B)=P(i\in A)+P(i\in B)-P(i\in A\cap B)=\frac{a}{n}+\frac{b}{n}-\frac{ab}{n^2}.$$

Slijedi

$$E(Y)=E\left(\sum_{i=1}^{n}Y_i\right)=\sum_{i=1}^{n}E(Y_i)=a+b-\frac{ab}{n}.$$

Izjednačavanjem i korištenjem svojstva apsorpcije i Pascalove rekurzije $\binom{n}{b}=\binom{n-1}{b}+\binom{n-1}{b}$ dobivamo identitet

$$\sum_{i\geq a}i\cdot\binom{n-a}{i-a}\binom{a}{a+b-i}=(a+b)\binom{n}{b}-a\binom{n-1}{b-1}=a\binom{n-1}{b}+b\binom{n}{b}. \tag{6}$$

Pokušajte ga dokazati metodom dvostrukog prebrojavanja!

4. Particije

Particija skupa N je rastav na disjunktne neprazne podskupove. Točnije, to je familija $\mathcal{P}=\{S_1,\ldots,S_k\}$ podskupova $S_i\subseteq N$, $S_i\neq\emptyset$ koje zovemo blokovima takva da vrijedi $\bigcup_{i=1}^{k}S_i=N$ i $S_i\cap S_j=\emptyset$ za $i\neq j$. Ukupan broj particija n-članog skupa je Bellov broj B_n, a broj particija s točno k blokova je Stirlingov broj druge vrste $\left\{ n \atop k \right\}$. Na isti način kao (1) slijedi identitet $\sum_{k=1}^{n}\left\{ n \atop k \right\}=B_n$. Stirlingovi brojevi druge vrste zadovoljavaju rekurziju analognu Pascalovoj:

$$\left\{ n \atop k \right\}=\left\{ n-1 \atop k-1 \right\}+k\cdot\left\{ n-1 \atop k \right\}.$$

Bellovi brojevi zadovoljavaju rekurziju

$$B_{n+1}=\sum_{i=0}^{n}\binom{n}{i}B_{n-i}. \tag{7}$$

Dokazi ovih rekurzija raspisani su u skripti [4].

Ako slučajno i uniformno biramo particiju $\mathcal{P}$, vjerojatnost da se sastoji od k blokova je

$$P\left(\left|\mathcal{P}\right|=k\right)=\frac{\left\{ n \atop k \right\}}{B_n}.$$

Matematičko očekivanje slučajne varijable $X=\left|\mathcal{P}\right|$ je

$$E(X)=\sum_{k=1}^{n}k\cdot\frac{\left\{ n \atop k \right\}}{B_n}=\frac{1}{B_n}\cdot\sum_{k=1}^{n}k\cdot\left\{ n \atop k \right\}. \tag{8}$$

S druge strane, za neprazan podskup $S\subseteq N$ promotrimo indikatorsku varijablu

$$X_S=\begin{cases}1, & S \text{ je blok od } \mathcal{P},\\ 0, & \text{inače.}\end{cases}$$

Njezino očekivanje ovisi samo o kardinalitetu $\left|S\right|=i$:

$$E\left(X_S\right)=P\left(S \text{ je blok od } \mathcal{P}\right)=\frac{B_{n-i}}{B_n}.$$

Vrijedi $X=\sum_S X_S$, gdje suma ide po svim nepraznim podskupovima od N. Iz linearnosti očekivanja i grupiranjem po kardinalitetu od S dobivamo

$$E(X)=\sum_S E\left(X_S\right)=\frac{1}{B_n}\cdot\sum_{i=1}^{n}\binom{n}{i}B_{n-i}.$$

Zadnja suma ide od $i = 1$ jer prazan skup ne može biti blok particije. U rekurziji suma ide od $i = 0$, pa uspoređivanjem dobivamo

$$E(X) = \frac{1}{B_n} \cdot \left(B_{n+1} - B_n\right).$$

Izjednačavanjem s slijedi

$$\sum_{k=1}^{n} k \cdot \left\{ {n \atop k} \right\} = B_{n+1} - B_n. \tag{9}$$

Identitet (9) također možemo dokazati dvostrukim prebrojavanjem. Na lijevoj strani brojimo particije od N s jednim istaknutim blokom. Uspostavimo bijekciju između takvih particija i particija skupa $\{1,..., n, n + 1\}$ u kojima element $n + 1$ nije sam u bloku: dodajemo element $n + 1$ u istaknuti blok, a u drugom smjeru istaknemo blok kojim je pokriven $n + 1$ i iz njega obrišemo taj element. Particija u kojima je element $n + 1$ sam u bloku ima B_n, pa je broj particija od N s jednim istaknutim blokom upravo $B_{n+1} - B_n$.

5. Permutacije

Permutacija je bijekcija $\pi : N \to N$. Broj permutacija n-članog skupa je faktorijela $n! = 1 \cdot 2 \cdot \ldots \cdot n$. Svaku permutaciju možemo na jedinstven način prikazati kao kompoziciju disjunktnih ciklusa. To su permutacije oblika $c = \left(i_1\, i_2 \cdots i_k\right)$ koje preslikavaju $i_1 \mapsto i_2$, $i_2 \mapsto i_3$, ..., $i_k \mapsto i_1$, a ostale elemente iz N ostavljaju fiksnim. Broj permutacija s točno i disjunktnih ciklusa označavamo $\left[{n \atop i} \right]$ i zovemo Stirlingovim brojem prve vrste. Identitet analogan (1) je $\sum_{i=1}^{n} \left[{n \atop i} \right] = n!$, a rekurzija analogna Pascalovoj

$$\left[{n \atop i} \right] = \left[{n-1 \atop i-1} \right] + (n-1) \cdot \left[{n-1 \atop i} \right]. \tag{10}$$

Dokazi su raspisani u [4].

Neka je vrijednost slučajne varijable X broj ciklusa permutacije π koju biramo uniformno. Njezino matematičko očekivanje je

$$E(X) = \sum_{i=1}^{n} i \cdot P\left(\pi \text{ ima točno } i \text{ ciklusa}\right) = \sum_{i=1}^{n} i \cdot \frac{\left[{n \atop i} \right]}{n!} = \frac{1}{n!} \cdot \sum_{i=1}^{n} i \cdot \left[{n \atop i} \right]. \tag{11}$$

Indikatorska varijabla ciklusa $c = \left(i_1\, i_2 \cdots i_k\right)$ dana je s

$$X_c = \begin{cases} 1, & c \text{ je ciklus od } \pi, \\ 0, & \text{inače.} \end{cases}$$

Matematičko očekivanje dobivamo tako da podijelimo broj permutacija kojima je c ciklus s ukupnim brojem permutacija:

$$E(X_c) = P(c \text{ je ciklus od } \pi) = \frac{(n-k)!}{n!}.$$

Vidimo da očekivanje ovisi samo o duljini ciklusa k. Slučajna varijabla X je zbroj indikatorskih varijabli po svim mogućim ciklusima: $X = \sum_c X_c$. Primijenimo linearnost očekivanja i grupiramo sumu po duljini ciklusa:

$$E(X) = \sum_c E(X_c) = \sum_{k=1}^{n} \binom{n}{k} \cdot (k-1)! \cdot \frac{(n-k)!}{n!}.$$

Ovdje je $\binom{n}{k} \cdot (k-1)!$ broj ciklusa duljine k (biramo elemente $i_1,\ldots, i_k$ i permutiramo ih do na ciklički pomak). Uvrštavanjem $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ i sređivanjem dobivamo

$$E(X) = \sum_{k=1}^{n} \frac{1}{k} = H_n. \tag{12}$$

Sumu recipročnih vrijednosti prvih n prirodnih brojeva nazivamo n-tim harmonijskim brojem i označavamo H_n. Izjednačavanjem (11) i (12) dobivamo

$$\sum_{i=1}^{n} i \cdot \left[{n \atop i} \right] = n! \cdot H_n.$$

Izraz na desnoj strani zadovoljava $n! \cdot H_n = \left[{n+1 \atop 2} \right]$, što možemo dokazati indukcijom koristeći se rekurzijom (10). Tako dolazimo do identiteta

$$\sum_{i=1}^{n} i \cdot \left[{n \atop i} \right] = \left[{n+1 \atop 2} \right]. \tag{13}$$

Pokušajte i njega dokazati dvostrukim prebrojavanjem! U članku [3] dvostrukim prebrojavanjem dokazani su mnogi slični identiteti s binomnim koeficijentima i Stirlingovim brojevima prve i druge vrste, a [2] je knjiga posvećena toj metodi dokazivanja. U ovom članku „sučelili” smo je s vjerojatnosnom metodom i vidjeli da mnogi kombinatorički identiteti na prirodan način slijede iz linearnosti matematičkog očekivanja.

Literatura

1. N. Alon i J. H. Spencer, *The probabilistic method, fourth edition*, Wiley, 2016.
2. A. T. Benjamin i J. J. Quinn, *Proofs that really count. The art of combinatorial proof*, Mathematical Association of America, 2003.
3. M. Knežević, V. Krčadinac i L. Relić, *Matrix products of binomial coefficients and unsigned Stirling numbers*, u *Proceedings of the 3rd Croatian Combinatorial Days* (urednici T. Došlić i S. Majstorović), Građevinski fakultet, Sveučilište u Zagrebu, 2021., str. 35-44. https://doi.org/10.5592/CO/CCD.2020.04
4. V. Krčadinac, *Kombinatorika, skripta*, Sveučilište u Zagrebu, 2022. https://web.math.pmf.unizg.hr/~krcko/nastava/komb/komb-skripta.pdf