Пет задатака из вероватноће

др Звонимир Шикић, Загреб

1. Колико пута, у просеку, морате бацити коцку да бисте добили 6?

2. У свакој кутији детерџента DEX налази се један од купона означених бројевима 1, 2, 3, 4 или 5. Када сакупите свих пет различитих купона, добијате награду. Колико кутија DEX-а, у просеку, треба да купите да бисте добили награду?

3. Узастопним бацањем кованица добијате блокове глава и блокове писама различитих дужина. Колика је просечна дужина блока?

4. Самјуел Пипс замолио је Исака Њутна да му израчуна који је од следећа три догађаја вероватнији: барем једна шестица у 6 бацања, барем две шестице у 12 бацања, барем три шестице у 18 бацања коцке. Шта је ваш одговор?

5. Колики је најмањи број људи у групи таквој да је вероватноћа да две особе рођендан славе истог датума (године се не морају поклапати) већа од 50%?

Саветујемо читаоцу да пре него што настави са читањем размисли о сваком задатку и покуша самостално да га реши. У наставку нудимо решења у која смо уплели доста занимљиве математике (геометријски ред, Ојлерову константу γ, развој за e^{-x} у ред степена итд.).

Решења

1. Већина људи некако осећа да је одговор 6. То је један од ретких случајева у којима нас вероватносни осећај не вара.

Ми ћемо доказати нешто општији резултат. Претпоставимо да је вероватноћа жељеног исхода неког експеримента p. Тада нежељени исход има вероватноћу $q = 1 - p$. Вероватноћа да се жељени исход догоди тек у n-том понављању међусобно независних експеримената је $P_n = pq^{n-1}$ ($n-1$ пута се мора догодити нежељени исход чија је вероватноћа q, а затим се једанпут мора догодити исход чија је вероватноћа p). Очекивана вредност потребног броја покушаја зато је:

$$E = 1P_1 + 2P_2 + \cdots = \sum_{n=1}^{\infty} nP_n = p + 2pq + 3pq^2 + \cdots = \sum_{n=1}^{\infty} npq^{n-1}.$$

Очекивану вредност E можемо израчунати методом којом се обично рачуна сума геометријског реда.

$$\begin{aligned} E &= p + 2pq + 3pq^2 + 4pq^3 + 5pq^4 + \cdots \\ qE &= \phantom{p + {}} pq + 2pq^2 + 3pq^3 + 4pq^4 + \cdots \\ E - qE &= p + \ pq + \ pq^2 + \ pq^3 + \ pq^4 + \cdots \end{aligned}$$

Применом формуле за суму геометријског реда сада лако налазимо:

$$E(1-q) = p(1+q+q^2+q^3+q^4+\cdots) = \frac{p}{1-q} = \frac{p}{p} = 1.$$

Дакле, $E = \frac{1}{1-q} = \frac{1}{p}$. То значи да, у просеку, експеримент морамо поновити $\frac{1}{p}$ пута да бисмо дошли до жељеног исхода. Ако је $p = \frac{1}{6}$, као у нашем задатку, експеримент треба поновити просечно 6 пута.

Исти проблем можемо решити и без рачунања бесконачних сума. Уочимо да први у низу експеримената може дати нежељени исход (са вероватноћом q) или жељени исход (са вероватноћом p). У првом случају крећемо испочетка, тј. очекивани број експеримената до жељеног исхода сада је $1+E$. У другом случају дошли смо до жељеног исхода, тј. очекивани број експеримената до жељеног исхода је 1. Сада је

$$E = (1+E)q + 1p,$$

одакле одмах следи

$$E(1-q) = q+p = 1, \quad E = \frac{1}{1-q} = \frac{1}{p}.$$

2. У првој кутији сигурно ћете наћи један број. Вероватноћа да у следећиј кутији нађете нови број износи $\frac{4}{5}$. Према решењу претходног задатка, просечно треба да купите $\frac{1}{4/5} = \frac{5}{4}$ кутија да би се то десило. Вероватноћа да у следећој кутији нађете нови број је $\frac{3}{5}$ и у просеку треба купити $\frac{1}{3/5} = \frac{5}{3}$ кутија да би се то десило. Задња два броја, у просеку захтевају куповину $\frac{1}{2/5} = \frac{5}{2}$ и $\frac{1}{1/5} = 5$ кутија. Укупан (просечни) број кутија које треба купити је:

$$1 + \frac{5}{4} + \frac{5}{3} + \frac{5}{2} + 5 = 5\left(\frac{1}{5} + \frac{1}{4} + \frac{1}{3} + \frac{1}{2} + \frac{1}{1}\right) \approx 11,4.$$

Када би DEX у својој наградној игри користио n бројева, просечан број кутија који би требало купити је:

$$n\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right).$$

Ту вредност можемо лако израчунати користећи се апроксимацијом

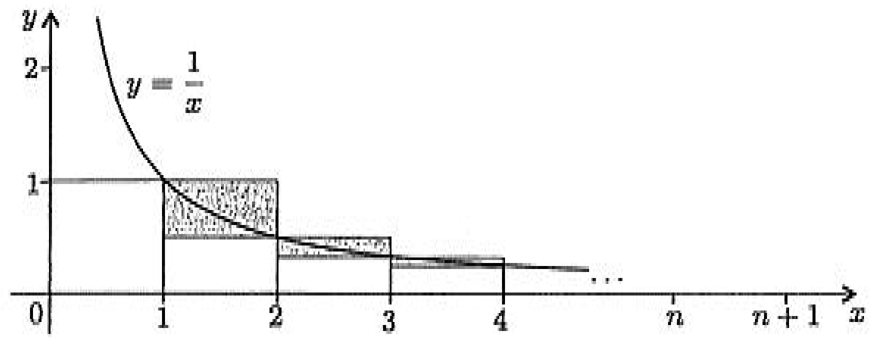
$$H_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \approx \ln n + \gamma + \frac{1}{2n},$$

где је γ славна Ојлерова константа, $\gamma \approx 0,577218\ldots$ На пример, за $n=5$ налазимо

$$5\left(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5}\right) \approx 5\ln 5 + 5\gamma + \frac{1}{2} \approx 11,4,$$

што је, на једну децималу, иста вредност коју добијамо и директним рачунањем вредности $5\left(1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}\right)$.

Објаснићемо како можемо доћи до те апроксимације, до саме константе γ. Кренимо од одређивања површина на следећој слици.



Слика 1.

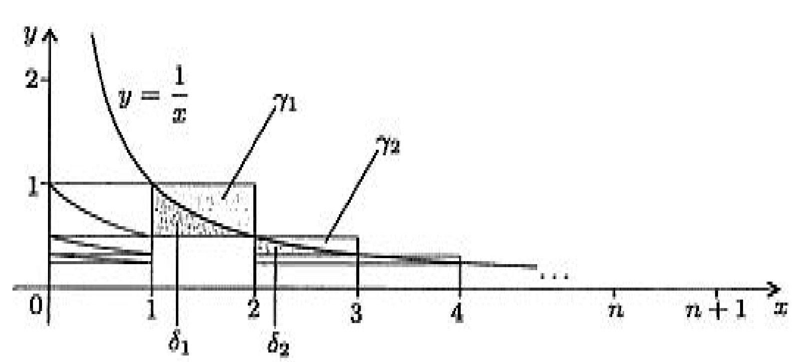
Укупна површина незатамљених правоугаоника, чије се основице протежу од 0 до n, једнака је

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

Иста је и укупна површина затамљених и незатамљених правоугаоника чије се основице протежу од 1 до $n+1$. Површина између x-осе и графика $y=\frac{1}{x}$ над интервалом $[1,n]$ једнака је $\ln n$. Одавде, узимајући у обзир збирове $\overline{\delta}_n=\delta_1+\delta_2+\cdots+\delta_n$ и $\overline{\gamma}_n=\gamma_1+\gamma_2+\cdots+\gamma_{n-1}$ (видети слику 2), лако следи:

(1) $H_n = 1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n} = \ln(n+1)-(\delta_1+\delta_2+\cdots+\delta_{n-1}) = \ln(n+1)-\overline{\delta}_n$,

(2) $H_n = 1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n} = \ln(n+1)-(\gamma_1+\gamma_2+\cdots+\gamma_n) = \ln(n+1)-\overline{\gamma}_{n+1}$.



Слика 2.

Прикажемо ли све површине $\gamma_1, \gamma_2, \ldots; \delta_1, \delta_2, \ldots$ у квадрату 1×1 (види слику 2) лако ћемо се уверити да важе следећа тврђења:

$$\sum_{n=1}^{\infty} \gamma_n = \lim_{n\to\infty} \overline{\gamma}_n = \gamma > \frac{1}{2}, \quad \sum_{n=1}^{\infty} \delta_n = \lim_{n\to\infty} \overline{\delta}_n = \delta < \frac{1}{2}, \quad \gamma + \delta = 1, \tag{3}$$

$$(\gamma - \overline{\gamma}_n) + (\delta - \overline{\delta}_n) = \frac{1}{n}, \quad \delta - \overline{\delta}_n \approx \frac{1}{2n} \approx \gamma - \overline{\gamma}_n. \tag{4}$$

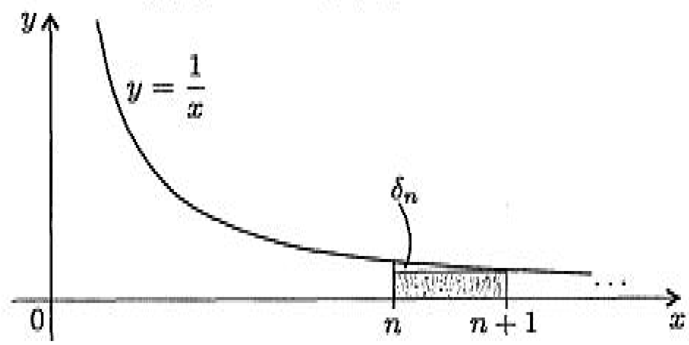
Из (1) и (4) сада лако следи:

$$H_n = 1 + \ln n - \overline{\delta}_n = 1 + \ln n + (\delta - \delta_n) - \delta = \ln n + \gamma + (\delta - \delta_n) \approx \ln n + \gamma + \frac{1}{2n},$$

што је наша апроксимација. Осим тога, из (2) и (4) следи:

$$\begin{aligned} H_n &= \ln(n+1) + \overline{\gamma}_{n+1} = 1 + \ln(n+1) + \overline{\gamma}_{n+1} - \gamma + \gamma \\ &\approx \ln(n+1) + \gamma - \frac{1}{2(n+1)} \approx \ln n + \frac{1}{n+1} + \gamma - \frac{1}{2(n+1)} \\ &= \ln n + \gamma - \frac{1}{2(n+1)}. \end{aligned}$$

Наиме, $\ln(n+1) - \ln n = \frac{1}{n+1} + \delta_n \approx \frac{1}{n+1}$, што се лако види са наредне слике.



Слика 3.

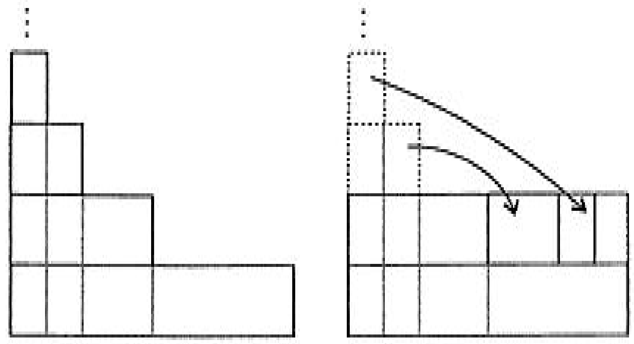
Све у свему, имамо следећу процену:

$$\ln n + \gamma - \frac{1}{2(n+1)} < H_n < \ln n + \gamma + \frac{1}{2n}.$$

3. Вероватноћа писма је $\frac{1}{2}$. Вероватноћа два писма у низу је $\frac{1}{4}$. Вероватноћа три писма у низу је $\frac{1}{8}$. Вероватноћа четири писма у низу је $\frac{1}{16}$, итд. Дакле, вероватноћа блока дужине n је $\frac{1}{2^n}$. Следи да је просечна дужина блока

$$D = \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 3 + \cdots + \frac{1}{2^n} \cdot n + \cdots$$

Овај бесконачни ред једнак је 2, што можемо израчунати на разне начине. Показаћемо како га је (у сасвим другом контексту) израчунао Никола из Орезме у XIV веку. Он је користио следећу слику којом је представио ову суму.



Слика 4.

И овде можемо заобићи бесконачна сабирања. Запитајмо се на колико промена (из главе у писмо или из писма у главу) просечно наилазимо у низу од n бацања новчића. Будући да је вероватноћа промене $\frac{1}{2}$, просечно наилазимо на $\frac{n}{2}$ промена. То значи да је просечна дужина блока 2.

4. Већина људи некако осећа да је вероватноћа иста у свим трима случајевима. (Просечан број шестица на шест бацања коцки је 1, на 12 бацања је 2, на 18 бацања је 3, а $1:6=2:12=3:18$.) Нажалост, то је један од честих случајева када нас вероватносни осећај вара. (Просеци и вероватноће нису исто.) Заборавимо зато осећаје и кренимо рачунати. Вероватноћа појављивања x шестица у n бацања је

$$\binom{n}{x}\left(\frac{1}{6}\right)^{x}\left(\frac{5}{6}\right)^{n-x}, \quad x=0,1,\ldots,n.$$

Када бацамо $6n$ пута, вероватноћа да се појави бар n шестица једнака је

$$\sum_{x=n}^{6n}\binom{n}{x}\left(\frac{1}{6}\right)^{x}\left(\frac{5}{6}\right)^{n-x}.$$

Њутн је израчунао ове вредности „ручно", а ми се данас можемо послужити таблицама кумулативне биномне дистрибуције, или рачунати пакетима који их садрже, и добићемо следеће вредности:

n	$6n$	$P(\geqslant n$ шестица$)$
1	6	0,665
2	12	0,619
3	18	0,597
4	24	0,584
5	30	0,576
$\vdots$	$\vdots$	$\vdots$
100	600	0,517

Дакле, највероватнија је барем једна шестица у 6 бацања, а најмање вероватне су барем три шестице у 18 бацања. То је и Њутново решење. Самјуел Пипс је сазнао на шта се треба кладити.

5. Већина људи има осећај да је у скупу људи величине школског разреда вероватноћа подударног рођендана пуно мања од 50%. И то је један од честих случајева лошег осећаја за вероватноће. Наиме, у скупу те величине подударност је око 70%. Окренимо се опет рачуну. Нека је $N = 365$ број једнако вероватних дана (занемарићемо преступне године), а s број људи у посматраном скупу. Израчунајмо вероватноћу да свих s људи има међусобно различите рођендане.

Рођендан првог од s људи може бити било којег од N дана. Други мора бити различит, па зато може бити било који од преосталих $N-1$ дана. Трећи може бити било који од преосталих $N-2$ дана, итд. до преосталих $N-s+1$ дана. Дакле, укупан број начина на који се могу реализовати сви међусобно различити рођендани је

$$N(N-1)(N-2)\cdots(N-s+1) = \frac{N!}{(N-s)!}.$$

С друге стране, укупан број начина на који се могу реализовати сви могући рођендани је N^s. Одавде следи да вероватноћа свих међусобно различитих рођендана износи:

$$\frac{N}{N}\cdot\frac{N-1}{N}\cdot\frac{N-2}{N}\cdot\ldots\cdot\frac{N-s+1}{N} = \frac{N!}{(N-s)!N^s}.$$

Комплементарна вероватноћа,

$$P_s = 1 - \frac{N!}{(N-s)!N^s},$$

јесте вероватноћа да скуп од s људи има барем један подударан рођендан. У следећој таблици дате су вредности те вероватноће за $N = 365$ и неке s.

s	5	10	20	23	30	40	60
P_s	0,027	0,117	0,411	0,507	0,706	0,891	0,994

Видимо да је вероватноћа подударног рођендана већа од 50% већу у скупу од 23 човека. У скупу од 60 људи подударан рођендан је готово сигуран.

Најтежи проблем у горњем разматрању јесте рачунање вредности P_s. Наиме, 365!, $(365-23)!$ и 365^{23} огромни су бројеви који захтевају огромне рачуне. Зато те рачуне овде нисмо ни увели, него смо (у таблици) дали само коначне резултате.

Пуно је лакше спровести приближне рачуне, попут следећег у којем се користи добро позната апроксимација:

$$e^{-x} = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \frac{x^4}{4!} - \cdots \approx 1 - x.$$

Ова апроксимација је добра за мале вредности x и обично се користи тако да десном страном $1-x$ апроксимирамо леву e^{-x}. Ми ћемо је користити у обрнутом смеру.

$$\begin{aligned}\frac{N!}{(N-s)!N^s} &= \frac{N}{N}\cdot\frac{N-1}{N}\cdot\frac{N-2}{N}\cdot\ldots\cdot\frac{N-s+1}{N}\\ &= \left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\cdots\left(1-\frac{s-1}{N}\right)\\ &\approx e^{-\frac{1}{N}}e^{-\frac{2}{N}}\cdots e^{-\frac{s-1}{N}} = e^{-\frac{1+2+\cdots+(s-1)}{N}}\\ &= e^{-\frac{s(s-1)}{2N}}\end{aligned}$$

Уочимо да је та апроксимација (због $e^{-x} > 1-x$) већа од тражене вредности. Сада лако закључујемо да је

$$P_s = 1-\frac{N!}{(N-s)!N^s} \approx 1-e^{-\frac{s(s-1)}{2N}}$$

и да је та апроксимација мања од тражене вредности. Желимо ли наћи s за који је $P_s > 0,5$, треба решити једначину

$$1-e^{-\frac{s(s-1)}{2\cdot 365}} = 0,5, \quad e^{-\frac{s(s-1)}{830}} = 0,5, \quad -\frac{s(s-1)}{830} = \ln 0,5,$$

$$s^2 - s + 830\ln 0,5 = 0.$$

Ова квадратна једначина има позитивно решење $24 < s < 25$. Дакле, за $s = 25$ сигурно важи $P_s > 0,5$. С обзиром на то да су наше апроксимације биле мање од тражене вредности, можда је $P_s > 0,5$ и за $s = 23$ или чак $s = 22$. То овај рачун с e^{-x} не може одлучити. Међутим, тачнији резултати показују да је $s = 23$ први s за који је $P_s > 0,5$.

Још пет задатака

1. Колика је вероватноћа да у два бацања новчића не добијете главу, а колика да у 6 бацања коцке не добијете 6? Колика је вероватноћа да се у n поновљених експеримента не деси догађај чија је вероватноћа $\frac{1}{n}$? Колико је e^{-1}?

2. Ако купите пет кутија DEX-а, колико различитих купона можете очекивати?

3. Ако падне глава ви мени дајете 1 динар, ако падне писмо ја вама дајем један динар. Да ли је могуће да сам ја од неког корака надаље стално на добитку? (Можда ће вас занимати, иако за решење задатка није битно, да након n корака вероватноћа да смо оба на нули износи $(n\pi)^{-1/2}$.)

4. Витез де Мере уочио је (коцкајући се) да је бар једна шестица у 4 бацања коцке вероватнија од $1/2$, док су бар две шестице у 24 бацања мање вероватне од $1/2$ (иако је вероватноћа једне шестице $1/6$, две $1/36$ и $4:6 = 24:36$). Замолио је Паскала да то математички докаже. Паскал је успео. Можете ли и ви?

5. Колики је најмањи број људи у групи таквој да је вероватноћа да бар један има исти рођендан као и ви већа од $1/2$?

Статијата прв пат е публикувана во списанието Тангента на ДМ на Србија во 2013/14 година