$\alpha + \omega$, 2003, Nr. 3, 68–71

Pasirinkti geriausią

Vilius Stakėnas
vilius@ktl.mii.lt

Straipsnyje nagrinėjamas pasirinkimo uždavinys: Kokia tikimybė iš atsitiktinai pasirodančių objektų sekos pasirinkti geriausią?

Kaip mes renkamės?

Retai renkamės nesvarstydami. Dažniausiai renkamės daiktą, kelią ar galimybę po tam tikrų stebėjimų, lyginimų ir samprotavimų tikėdamiesi pasirinkti geriausią. Rinkimosi aplinkybių būna visokių. Kartais galima rinktis palyginus visas galimybes. Pavyzdžiui, dviratį ar kompiuterį parduotuvėje tikriausiai pirksime tik apžvelgę visas siūlomas jų rūšis. O grybams miške rinkti tokia strategija visiškai netinka. Tuo labiau netinka draugui ar draugei pasirinkti...

Kartais rinkdamiesi jau turime išankstinių žinių apie tai, koks yra pats geriausias objektas, kartais tas žinias įgyjame jau rinkdamiesi, t. y. lygindami objektus, kuriuos sutinkame. Kartais turime išankstinę rinkimosi strategiją, o kartais renkamės spontaniškai...

Taigi norėdami matematiškai panagrinėti rinkimosi situaciją, turime tiksliau apibrėžti rinkimosi aplinkybes.

Tegu jos būna tokios:

- *Iš n objektų, tarp kurių nėra lygiaverčių, reikia pasirinkti vieną. Objektai pasirodo mums vienas po kito. Jeigu objekto nepasirinkome, prie jo sugrįžti nebegalima. Pasirodančius objektus galime lyginti ir „pagal kokybę" rikiuoti į eilę.*
- *Prieš prasidedant „veiksmui" pasirenkame skaičių m, $0 \leqslant m \leqslant n$, ir nutariame nė vieno iš pirmųjų m objektų nesirinkti; juos tiesiog rikiuosime į eilę pagal „kokybę". Kai tokią eilę sudarysime, rinksimės pirmą vėliau pasirodžiusį objektą, kuris yra geresnis už geriausiąjį iš pirmųjų m.*

Kokia tikimybė, kad šitaip rinkdamiesi pasirinksime patį geriausią iš n objektų? Kokia tikimybė, kad pasirinksime antrąjį (trečiąjį, ketvirtąjį...) pagal gerumą? Kokia tikimybė, kad liksime be nieko?

Galbūt uždavinį bus įdomiau nagrinėti, jei galvosime ne apie beveidžius abstrakčius objektus, bet apie geriausiojo gyvenimo draugo ar draugės pasirinkimą...

Ribiniais atvejais, t. y. kai $m = 0$ ar $m = n$, atsakymą galima pasakyti iškart.

Išties, jei $m = n$, tai jau prieš pasirodant pirmajam objektui nusprendžiame nieko nesirinkti. Tada tikimybė pasirinkti geriausiąjį objektą lygi nuliui. Šia vertinga strategija aš visada naudojuosi, kai kartais užklystu į prabangias parduotuves.

Jei $m = 0$, tai mūsų strategija — rinktis pirmą pasitaikiusį. Aišku, kad šitaip renkantis tikimybė pasirinkti geriausiąjį (taip pat pagal gerumą antrąjį, trečiąjį...) lygi $\frac{1}{n}$.

Kai $0 < m < n$, uždavinys nėra visiškai paprastas.

Baigtys

Kokia mums rūpimo įvykio tikimybė? Jeigu tai ir tampa pirmuoju mūsų klausimu, tai tik trumpam. Nes tuoj pat turime grįžti prie „tikrojo" pirmojo klausimo: Kas yra mūsų bandymas? Ir kokios šio bandymo baigtys?

Taigi kokios yra mūsų rinkimosi bandymo baigtys?

Dėl paprastumo tarkime, kad mūsų objektai — n skirtingos masės dėžučių. Sunkesnės dėžutės yra geresnės už lengvesnes. Dėžutės mums pasirodo viena po kitos, pasvėrę rikiuojame jas vieną šalia kitos pasirodymo eilės tvarka. Kad ir kiek daug svertų ką tik pasirodžiusi dėžutė, mes nežinome, ar nebus dar sunkesnių.

Tarkime, kad visos dėžutės mums nepastebimose vietose sužymėtos numeriais nuo 1 iki n. Sunkiausios dėžutės numeris yra 1, antros pagal svorį — 2 ir taip toliau. Jeigu tuos numerius matytume, pasirinkimo strategija būtų paprasta: lauktume, kol pasirodys dėžutė Nr. 1 ir tiek.

Tarkime, net tada, kai objektą (dėžutę) pasirinkome, jos ir toliau pasirodo, tik atsisakyti pasirinkimo ir rinktis iš naujo nebegalime. Juk ir parduotuvėje jau nusipirkę daiktą ne visada iš karto skubame namo.

Bandymo pabaigoje turime priešais n dėžučių eilę. Tai ir yra bandymo baigtis. Tas, kas žino, kur užrašyti dėžučių numeriai, galėtų šią dėžučių eilę nusakyti numerių seka

$$k_1, k_2, \ldots, k_n.$$

Primenu, kad pirmoji dėžutė, sunkesnė už sunkiausiąją iš pirmųjų m dėžučių, yra mūsų. Jei tokios nėra, tai likome, vadinasi, be nieko.

Dabar jau galime kai ką apskaičiuoti. Akivaizdu, kad bandymo baigčių yra tiek, kiek yra galimybių n numerių išrikiuoti į eilę, t. y. lygiai $n!$.

Kokia tikimybė, kad liksime be nieko? Taip atsitiks, jeigu sunkiausioji dėžutė, t. y. dėžutė, pažymėta numeriu 1, bus tarp m pirmųjų.

Nesunku apskaičiuoti, kad tokio atvejo tikimybė yra

$$\frac{m \cdot (n-1)!}{n!} = \frac{m}{n}.$$

Tikimybė pasirinkti geriausiąjį

Jeigu objektą renkamės po to, kai „ištyrinėję" m pirmųjų objektų susidarome „kriterijų", tai paties geriausio galime iš viso nepasirinkti, geriausiąjį galime pasirinkti rinkdamiesi $(m+1)$-ąjį, $(m+2)$-ąjį, ..., n-ąjį pasirodžiusį objektą.

Tikimybę likti be nieko jau apskaičiavome.

Pažymėkime p_m tikimybę, kad pasirinksime geriausiąjį, o $p_m(m+1)$, $p_m(m+2)$, ..., $p_m(n)$ — tikimybes, kad tas geriausias bus iš eilės $(m+1)$-asis, $(m+2)$-asis, ..., n-asis. Taigi

$$p_m = p_m(m+1) + p_m(m+2) + \cdots + p_m(n).$$

Nesunku apskaičiuoti $p_m(m+1)$, t. y. įvykio, kad $(m+1)$-asis objektas bus pats geriausias, tikimybę. Šiam įvykiui bus palankios tokios baigtys

$$k_1, k_2, \ldots, k_m, 1, k_{m+2}, \ldots, k_n;$$

čia k_i — skirtingi, bet nelygūs 1 skaičiai. Nesunku suvokti, kad

$$p_m(m+1) = \frac{(n-1)!}{n!} = \frac{1}{n}.$$

Apskaičiuokime $p_m(m+2)$, t. y. įvykio, kad $(m+2)$-asis objektas bus pats geriausias, tikimybę. Šiam įvykiui bus palankios baigtys

$$k_1, k_2, \ldots, k_{m+1}, 1, k_{m+3}, \ldots, k_n, \tag{1}$$

čia k_i — skirtingi, nelygūs 1 skaičiai, be to,

$$k_{m+1} > \max\{k_1, k_2, \ldots, k_m\}.$$

Kiek yra tokių baigčių? Galvokime taip: prieš 1 eilėje yra $m+1$ skaičių, kurie parinkti iš aibės $\{2; 3; \ldots; n\}$. Parinkti jų derinį galime

$$C_{n-1}^{m+1}$$

būdais. Keliais būdais šiuos skaičius galime tinkamai išdėstyti (1) eilės pradžioje? Pats mažiausias iš parinktųjų skaičių turi būtinai būti įrašytas į vieną iš m pirmųjų vietų, taigi turime m galimybių. Kiti gali būti bet kur, taigi jiems išdėstyti turime $m!$ galimybių. Tada skaičiams $k_1, k_2, \ldots, k_{m+1}$ tinkamai išdėstyti (1) eilėje turime iš viso

$$m \cdot C_{n-1}^{m+1} \cdot m!$$

galimybių. Išdėstę $m+1$ pirmųjų (1) eilės skaičių, užrašome 1; liko dar $n-(m+2)$ skaičių, kurie po vieneto gali būti išdėstyti bet kaip. Taigi (1) baigčių iš viso yra $m \cdot C_{n-1}^{m+1} \cdot m! \cdot (n-m-2)!$ ir

$$p_m(m+2) = \frac{m \cdot C_{n-1}^{m+1} \cdot m! \cdot (n-m-2)!}{n!} = \frac{m}{n} \cdot \frac{1}{m+1}.$$

Panašiai apskaičiuotume ir kitas tikimybes:

$$p_m(m+j) = \frac{m}{n} \cdot \frac{1}{m+j-1}.$$

Taigi tikimybė, kad pasirinksime patį geriausią, jeigu kriterijų susidarėme pagal m pirmųjų objektų, lygi

$$p_m = \frac{m}{n}\Big(\frac{1}{m} + \frac{1}{m+1} + \cdots + \frac{1}{n-1}\Big) = \frac{m}{n} \cdot \sum_{j=m}^{n-1} \frac{1}{j}.$$

Praktinis klausimas

Jeigu tenka rinktis, pavyzdžiui, iš $n = 10, 20$ ar kito objektų kiekio, su kokiu m tikimybė pasirinkti geriausią yra didžiausia? Jeigu m bus mažas, tai galime nesulaukti geriausio; jeigu m didelis — galime betyrinėdami geriausiąjį pražiopsoti.

Pažymėkime m_n tą m reikšmę, su kuria tikimybė p_m yra didžiausia. Dauguma žmonių tikriausiai manys, kad $m_n \approx \frac{n}{2}$. Ne, taip nėra. Norėdamas skaitytojus tuo įtikinti, šiek tiek paskaičiavau kompiuteriu.

$n =$	5	10	15	20	25	30
$m_n =$	2	4	5	7	9	11
$p_{m_n} =$	0,433	0,398	0,389	0,384	0,381	0,379

Dabar galėtume spėti, kad $m_n \approx \frac{n}{3}$, $p_{m_n} \approx \frac{1}{3}$. Tačiau tiesa apie geriausią pasirinkimą slypi kur kas giliau — iracionaliųjų skaičių gelmėse! Kai n artėja prie begalybės, tai

$$\frac{m_n}{n} \to \frac{1}{\mathrm{e}}, \qquad p_{m_n} \to \frac{1}{\mathrm{e}};$$

čia $\mathrm{e} = 2{,}718281\ldots$ — tas pats įžymusis skaičius, kukliai vadinamas skaičiumi e.

O kaip kiti, ne patys geriausieji?

Rinkdamiesi pagal aptartąją strategiją, galime pasirinkti patį geriausią objektą, tačiau galime pasirinkti ir antrąjį pagal kokybę, ir trečiąjį...

Kuris įvykis labiau tikėtinas: kad pasirinksime geriausiąjį ar kad antrąjį pagal pagal kokybę?

Siūlau skaitytojui atidėti kelioms minutėms žurnalą į šalį ir pačiam pasvarstyti. Ar spėjimas pasitvirtins, ar ne — emocijos garantuotos.

Pažymėkime $p_m(k)$ tikimybę, kad renkantis pagal aptartąją strategiją, t. y. renkantis pirmą objektą, kuris yra geresnis už m pirmųjų, pasirinksime k-ąjį pagal kokybę. Aišku, kad šitaip rinkdamiesi galime pasirinkti tik patį geriausią, antrąjį, trečiąjį, ..., $(n-m)$-ąjį, t. y. $p_m(k) > 0$ tik su $k = 1, 2, \ldots, n-m$.

Tikimybes $p_m(k)$ galima apskaičiuoti panašiai, kaip skaičiavome geriausio pasirinkimo tikimybę. Tereikia tik kruopščiai pertvarkyti reiškinius su faktorialais ir binominiais koeficientais. Nevarginsiu skaitytojo aiškindamas tuos pertvarkius. Užrašysiu formules ir tiek:

$$p_m(1) = \frac{m}{n} \cdot \sum_{j=m}^{n-1} \frac{1}{j},$$

$$p_m(2) = \frac{m}{n} \cdot \sum_{j=m}^{n-1} \Big(1 - \frac{j}{n-1}\Big)\frac{1}{j},$$

$$p_m(3) = \frac{m}{n} \cdot \sum_{j=m}^{n-1} \Big(1 - \frac{j}{n-2}\Big)\Big(1 - \frac{j}{n-1}\Big)\frac{1}{j},$$

....

$$p_m(k) = \frac{m}{n} \cdot \sum_{j=m}^{n-1} \Big(1 - \frac{j}{n-k+1}\Big)\cdots\Big(1 - \frac{j}{n-2}\Big)\Big(1 - \frac{j}{n-1}\Big)\frac{1}{j},$$

....

$$p_m(n-m) = \frac{m}{n} \cdot \sum_{j=m}^{n-1} \Big(1 - \frac{j}{m+1}\Big)\cdots\Big(1 - \frac{j}{n-2}\Big)\Big(1 - \frac{j}{n-1}\Big)\frac{1}{j},$$

Iš užrašytųjų formulių matyti, kad

$$p_m(1) > p_m(2) > \cdots > p_m(n-m).$$

Taigi pasirinkti geresnį yra labiau tikėtina nei blogesnį.

Pasaulis nėra taip jau blogai sutvarkytas, ar ne?

$\alpha + \omega$