01 Štefan išče Jano

January 28, 2024

Tile, prvi, zapiski bodo bolj matematični. Kasnejši bodo zelo redko matematični, pač pa bodo bolj programerski.

1 Štefan išče Jano

Štefan gre na žurko, na kateri je n deklet. Rekli so mu, da mora nujno spoznati Jano. Ta je seveda ena sama. Nesrečnik hodi od ene do druge in jih sprašuje "A si ti mogoče Jana?" Koliko jih bo vprašal, preden jo najde?

V najboljšem primeru eno samo. V najslabšem vseh n. Pa v poprečnem?

1.1 Kako računamo poprečja

Najprej moramo ponoviti nekaj iz verjetnosti in statistike. Predstavljajmo si, da nekje delijo denar. Pridemo tja in vržemo kocko. Če pade 6, dobimo 20 evrov, sicer 10. Koliko dobimo v poprečju?

Odgovor ni 15, ker večkrat dobimo 10 kot 20. Najpreprosteje si predstavljamo, da bomo tja prišli 60-krat, od tega bomo (tako je pričakovati) desetkrat vrgli 6 in petdesetkrat kaj drugega, torej bomo dobili

$$10 \times 20 + 50 \times 10 = 700 \text{ evrov.}$$

V poprečju torej dobimo

$$700 \ / \ 60 = 11,66 \text{ evrov.}$$

Seveda dobimo enak rezultat tudi z n poskusi, ko v poprečju dobimo

$$(p_{20}n \times 20 + p_{10}n \times 10)/n \text{ evrov,}$$

kjer smo s p_{20} in p_{10} označili verjenosti, da dobimo 20 oziroma 10 evrov.

n seveda okrajšamo in dobimo

$$p_{20} \times 20 + p_{10} \times 10.$$

V splošnem: če imamo neke možne izide (dobičke, cene, število potrebnih vprašanj) x_1, x_2, $...x_n$ in so verjetnosti teh izidov enake p_1, p_2, ..., p_n, je poprečje, oziroma, bolj pravilno, pričakovana vrednost, enaka

$$p_1 x_1 + p_2 x_2 + ... + p_n x_n$$

ali, bolj učeno,

$$\sum_{i=1}^{n} p_i x_i$$

1.2 Nazaj k Štefanu

Koliko deklet bo nadlegoval Štefan? Že vemo: če bo imel srečo (oziroma: če bodo imele srečo) bo težil samo eni, če smolo vsem, sicer pa nekje vmes.

Kakšna je verjetnost, da bo težil i-tim? Takšna: najprej mora $i-1$-krat zgrešiti, nato pa zadeti. Verjetnost, da zgreši prvič, je enaka $\frac{n-1}{n}$, saj je na žurki n deklet in med njimi je $n-1$ ne-Jan. Verjetnost, da zgreši drugič, je enaka $\frac{n-2}{n-1}$: kandidatk je namreč še $n-1$ in med njimi je $n-2$ takšnih, ki niso Jane. Verjetnost, da zgreši v poskusu številka 3, je $\frac{n-3}{n-2}$. ... In tako naprej. Verjetnost, da zgreši v poskusu $i-1$, je, po zgornjem vzorcu $\frac{n-(i-1)}{n-(i-2)}$. In verjetnost, da v i-tem poskusu zadane, je $\frac{1}{n-(i-1)}$, saj je $i-1$ deklet že izločil in Jana je ena od preostalih $n-(i-1)$.

Namnožimo torej te verjetnosti: verjetnost, da zgreši prvič, drugič, tretjič ... $i-1$-vič, potem pa v i-tem poskusu zadane, je

$$p_i = \frac{n-1}{n}\frac{n-2}{n-1}\frac{n-3}{n-2}\frac{n-4}{n-3}...\frac{1}{n-i-1}$$

Vsak števec se krajša z naslednjim imenovalcem, pa imamo (oh, kakšna neslana šala!)

$$p_i = 1/n$$

Vse verjetnosti so enake kar 1 / n. To je seveda popolnoma pričakovano, tako bi rekli že na prvi pogled. Vendar se bo kdaj pa kdaj zgodilo tudi kaj nepričakovanega, zato je pametno, da smo previdni. Pa malo vaje iz računanja in, predvsem, razmišljanja, nam tudi ni škodilo.

Koliko poskusov bo torej potreboval v poprečju?

$$\sum_{i=1}^{n} p_i i = \sum_{i=1}^{n} \frac{1}{n} i = \frac{1}{n} \sum_{i=1}^{n} i$$

Vsoto števil od 1 do n bomo pri tem predmetu srečali skoraj vsak teden, zato se jo enkrat za vselej naučimo izračunati.

1.3 Vsota aritmetične vrste

Koliko zvezdic je na sliki?

```
1 *
2 **
3 ***
4 ****
5 *****
6 ******
7 *******
8 ********
  12345678
```

Težko reči? Potem pa povejte, koliko zvezdic in plusov (skupaj) je na tej sliki?

```
1 *++++++++
2 **+++++++
3 ***++++++
4 ****+++++
5 *****++++
6 ******+++
7 *******++
8 ********+
  123456789
```

Imamo osem vrstic in devet stolpcev (v zadnji vrstici je namreč osem zvezdic in še en plus), torej je zvezdic in plusov skupaj $8 \times 9 = 72$. Vidimo pa tudi, da je obojih enako, torej je zvezdic 36.

V splošnem: če bi narisal 1 + 2 + 3 + 4 + 5 + 6 + ... + n zvezdic, bi dodal še natančno toliko plusov ter imel $n(n+1)$ zvezdic in plusov. Torej $\frac{n(n+1)}{2}$ zvezdic. Zatorej

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

1.4 Spet nazaj k Štefanu

Spet se vrnimo k Štefanu. Ugotovili smo, da bo v poprečju težil $\frac{1}{n}\sum_{i=1}^{n} i$ puncam (če ga prej ne vržejo ven).

$$\frac{1}{n}\sum_{i=1}^{n} i = \frac{1}{n}\frac{n(n+1)}{2} = \frac{n+1}{2}$$

To je intuitivno in logično in tako naprej. Najprej: zdelo se nam je, da jih bo moral vprašati nekako pol. Vidimo, da za 0.5 več kot pol. Če je tam samo ena, jih bo vprašal $(1+1)/2 = 1$. Njo pač. Če sta dve, jih bo vprašal $(1+2)/2 = 1.5$ – pač eno ali dve. Če so tri, jih v poprečju vpraša $(3+1)/2 = 2$ – to je, eno, dve ali tri.

Poleg tega je stvar smiselna tudi zato, ker je linearna: če pride na žurko dvakrat več deklet, je pričakovati, da bo imel dvakrat toliko dela.

2 Pijani Štefan

Kot da to še ne bi bilo dovolj: na drugo žurko je prišel pijan. Gre k prvi in jo vpraša, če je Jana. Gre k drugi. Gre k tretji. Potem gre k četrti - ki pa je pravzaprav ista kot druga, saj je, kot rečeno, pijan in je pozabil, da je tisto že vprašal.

Nova naloga je torej takšna: hodi od ene do druge, pri čemer sproti pozablja, katero je že vprašal. Kolikokrat bo vprašal, preden najde Jano?

V najboljšem primeru enkrat. V najslabšem bo prej konec žurke. Oziroma - poljubno velikokrat. Matematiki nas bodo grdo gledali, ampak recimo, da neskončnokrat. Pa poprečje?

Najprej: kaj pravi intuicija? Gotovo več kot prej. Jih bo moral vprašati n? Ali še več?

Če kdo meni, da n, naj razmisli tole: recimo, da na naslednjo žurko pride dvakrat toliko punc. Bo moral Štefan postaviti dvakrat toliko vprašanj? Ali več? Kako narašča količina Štefanovega dela s številom deklet? Linearno? Počasneje? Najbrž ne. Hitreje? Zdi se, da. Če je punc več je delo treznega Štefana pač težje, delo pijanega Štefana pa *bolj težje* (*sorry, slavisti*). Če jih namesto 100 pride 1000, bo imel trezni (v poprečju) desetkrat več dela. Pijani pa … pri 100 še gre, pri 1000 pa nima šans, ne?

Naloga je ekvivalentna tej: imamo vse pikove karte. Trinajst jih je. Pomešamo in preverimo, ali je na vrhu as. Če ni, premešamo ponovno. Če ni, ponovno. Kolikokrat moramo to ponoviti, da bo na vrhu as?

Spet je potrebno sešteti vsoto $\sum_{i=1}^{\infty} p_i i$, pri čemer pa je p zdaj malo drugačna. Kakšna je verjetnost, da bo Jano odkril z natančno i vprašanji?

Verjetnost je pravzaprav preprosteje izračunati. Ker sproti pozablja, je v vsakem poskusu verjetnost, da bo zgrešil, enaka $\frac{n-1}{n} = 1 - \frac{1}{n}$, verjetnost, da zadane, pa $\frac{1}{n}$. Verjetnost, da najprej $i-1$-krat zgreši, potem pa zadane, je torej

$$p_i = \left(1 - \frac{1}{n}\right)^{i-1} \frac{1}{n}$$

Vsota, ki nas zanima, je

$$\sum_{i=1}^{n} p_i i = \sum_{i=1}^{n} \left(1 - \frac{1}{n}\right)^{i-1} \frac{1}{n} i$$

ali, če izpostavimo $\frac{1}{n}$ in postavimo i na začetek,

$$\frac{1}{n} \sum_{i=1}^{n} i \left(1 - \frac{1}{n}\right)^{i-1}$$

Znamo?

Najprej zamenjajmo

$$q = 1 - \frac{1}{n}$$

da dobimo

$$\sum_{i=1}^{\infty} iq^i$$

Nato ugotovimo še, da je vseeno, če seštevamo kar od 0 naprej, saj je $0q^0$ enako 0. Torej bomo seštevali

$$\sum_{i=0}^{\infty} iq^i$$

Znamo sešteti to?

2.1 Vsota geometrijske vrste

Ne. Pač pa bi morali znati sešteti tole:

$$1 + q + q^2 + q^3 + ... q^n$$

To pa zato, ker vemo, od enkrat iz srednje šole, da je

$$(1 + q + q^2 + q^3 + ... q^n)(1 - q)$$

enako

$$(1 - q + q - q^2 + q^2 - q^3 + q^3 - q^4 ... + q^{n+1}$$

kar je, ko se vsi vmesni členi paroma pokoljejo, enako

$$1 - q^{n+1}$$

Torej

$$(1 + q + q^2 + q^3 + ... q^n)(1 - q) = 1 - q^{n+1}$$

Se pravi

$$1 + q + q^2 + q^3 + ... q^n = \frac{1 - q^{n+1}}{1 - q}$$

(Jasno, ob predpostavki, da je q različen od 1. Če je enak 1, pa tega ulomka ne smemo napisati, vendar nam ga na srečo tudi ni treba, saj je tedaj vstoa očitno enaka n.)

Pa vsota neskočne geometrijske vrste?

$$1 + q + q^2 + q^3 + ... = \lim_{n \to \infty} \frac{1 - q^n}{1 - q}$$

Če je q večji ali enak 1, tole zbezlja v neskončnost. Če je q (oziroma $|q|$ manjši od 1, pa je to enako

$$1 + q + q^2 + q^3 + ... = \frac{1}{1 - q}$$

Torej

$$\sum_{i=0}^{\infty} q^i = \frac{1}{1 - q} \quad \text{(če } |q| < 1\text{)}$$

2.2 Nazaj k pijanemu Štefanu

Vse lepo in prav, samo mi moramo sešteti

$$\sum_{i=0}^{\infty} iq^{i-1}$$

Pa opazimo, da je iq^{i-1} ravno odvod q^i po q in napišemo

$$\sum_{i=0}^{\infty} iq^{i-1} = \sum_{i=0}^{\infty} \frac{dq^i}{dq}$$

Potem pa se spomnimo, da je odvod linearna zadeva in je odvod vsote enak vsoti odvodov. Ali, kot nam pride prav tule: vsota odvodov je enaka odvodu vsote. Vseeno je, ali najprej odvajaš in potem seštevaš ali obratno.

Pravzaprav ni vseeno: tu nam bo prišlo prav, da najprej seštejemo. Torej obrnimo, postavimo vsoto znotraj odvoda.

$$\sum_{i=0}^{\infty} iq^{i-1} = \frac{d \sum_{i=0}^{\infty} q^i}{dq}$$

To vsoto namreč znamo izračunati!

$$\sum_{i=0}^{\infty} iq^{i-1} = \frac{d \frac{1}{1-q}}{dq}$$

Napnemo svoje matematične mišice: tale odvod je enak

$$\sum_{i=0}^{\infty} iq^{i-1} = \frac{1}{(1 - q)^2}$$

Zdaj pa se vrnimo čisto nazaj: poprečno število vprašanj pijanega Štefana je

$$\frac{1}{n}\sum_{i=1}^{n} iq^{i-1}$$

kjer je q enak $1 - \frac{1}{n}$.

Vsoto smo pravkar ugnali, torej imamo

$$\frac{1}{n}\frac{1}{(1-q)^2}$$

Če vemo, da je $q = 1 - \frac{1}{n}$, je $1 - q = \frac{1}{n}$.

Pa imamo

$$\frac{1}{n}\frac{1}{\left(\frac{1}{n}\right)^2} = \frac{1}{n}\,n^2 = n$$

Pijani Štefan bo v poprečju vprašal n deklet.

Rezultat je zanimiv. Kot prvo: če imamo le pikove karte, bo potrebno v poprečju premešati 13-krat, da bo na vrhu as. Zanimivo. Predvsem pa je zanimivo - in za večino nepričakovano - tole: če pride na žurko dvakrat več deklet, bo imel pijani Štefan natančno dvakrat toliko dela. Ne več.

Pijani Štefan sploh ni tako neučinkovit, kot bi si mislili.

3 Bisekcija deklet

Štefan torej pride še na tretjo žurko. Tokrat: trezen. In z načrtom.

Dekleta poprosi, da se postavijo v vrsto po abecedi. Vejo, da je tip zatežen, vendar ocenijo, da bo mogoče boljše za vse, če storijo, kot hoče. Češ, mogoče bo hitreje mimo.

Postavijo se torej v vrsto, Štefan (ki še vedno išče Jano) stopi k srednji in jo vpraša, kako ji je ime. Izve, da je Fanči. Ker so urejene po abecedi, se prijazno zahvali levi polovici, češ lahko se usedejo, grejo po pijačo ali karkoli ali, zaradi njega, tudi domov, saj ve, da Jane ni med njimi.

Usmeri se na desno polovico. Napoti se k srednji, jo vpraša po imenu, izve da je Micka. Tokrat odslovi vse desno od Micke (vključno z Micko samo).

In tako naprej.

Koliko deklet bo moral vprašati po imenu, preden dobi Jano?

V najboljšem primeru (zanj, predvsem pa za vse prisotne punce, razen za Jano), stoji Jana na sredi in je torej vprašal le njo.

Pa v najslabšem?

Takole razmislimo. Recimo, da je na žurki 1024 punc. V prvem koraku jih odslovi polovico, torej jih ostane še 512. V drugem še polovico in jih ostane 256. Po tretjem koraku 128 … Kolikokrat je potrebno razpoloviti 1024, da dobimo 1?

Matematiki imajo posebno funkcijo, ki računa to vrednost. Reče se ji: dvojiški logaritem.

$$\log_2 1024 = 10$$

V najslabšem primeru bo Štefan ponadlegoval 10 deklet. Točneje 11, saj bo vprašal tudi tisto, ki je ostala sama.

(Točen izračun je malenkost drugačen. Najprej se spomnimo, da Štefan števila deklet ne razpolavlja, temveč odslovi tudi srednjo. Boljše začetno število deklet je 1023. V prvem poskusu vpraša srednjo od njih in potem odslovi vse, ki so levo ali desno, torej mu jih ostane še (1023 - 1) / 2 = 511. Potem vpraša srednjo, odslovi njo in polovico ostalih in ostane mu jih (511 - 1) / 2 = 255. Potem (255 - 1) / 2 = 127. Nato 63, 31, 15, 7, 3, 1. Pri 1023 dekletih imamo točno 10 razpolavljanj. V splošnem: če razlagamo ta primer, se splača nastaviti n na število, ki je za 1 manjše od neke potence dvojke. Če imamo $n = 2^q - 1$, bomo potrebovali q vprašanj. Pod pogojem, da odstrani srednjo in polovico ostalih.)

Kaj pa poprečni primer?

Na prvih predavanjih smo eksperimentalno ugotovili naslednje: v poprečju potrebujemo le eno vprašanje manj kot v najslabšem primeru. To je seveda nenavadno in nepričakovano. No, nenavadno in nepričakovano je za tiste, ki so pravkar poslušali šele prvo predavanje iz Algoritmov in podatkovnih struktur. V resnici pa je to, da sta si poprečna in najslabša situacija podobni, zelo običajno in bo posledično sčasom postalo tudi pričakovano.

Zakaj?

Kje bi morala stati Jana, da bi Štefan vprašal le eno punco? Točno na sredi. V vsej vrsti je natančno eno mesto, do katerega Štefan pride z enim vprašanjem.

Če Jane ni tam, bo šel Štefan na levo ali na desno. Torej obstajata dve mesti, do katerih Štefan pride z dvema vprašanjema. Drugače povedano: če ne bi iskal ravno Jane, temveč koga drugega, obstajata natančno dve dekleti, ki sta "dve vprašanji daleč". Od vsake od teh dveh gre lahko levo ali desno. Torej obstajajo štiri dekleta na razdalji 3. In osem deklet na razdalji 4. In šestnajst deklet na razdalji 5. In 32 na razdalji 6 …

In tako naprej. Na vsakem nivoju je dvakrat toliko deklet kot na prejšnjem.

Kot smo ugotovili na predavanjih, je vsako drugo dekle - se pravi pol deklet! - na zadnjem nivoju, torej tam, do koder potrebuje Štefan največje število vprašanj. Najslabši scenarij se zgodi kar v polovici primerov!

Toliko o intuciji. Kako pa se to izračuna?

Nadaljevanje prihodnjič