Náhodné procházky

. . . kam jdete?

(verze 1.2 z 2021-03-18)

0. Bude procházka!

Je čtvrtek kolem deváté večer a banda matfyzáků se vrací z IPSka na kolej. Aby to bylo zábavnější, mají s sebou kostku s vhodným počtem stěn. Jakmile potkají křižovatku nebo odbočku, hodí si kostkou a podle toho se rozhodnou, kam půjdou. Dojdou tímto způsobem na kolej včas alespoň na to, aby fyzici ráno stihli tělocvik na opačné straně Prahy? Jak moc je budou bolet nohy?

Mapu Prahy si můžeme představit jako neorientovaný graf a křižovatky (nebo odbočky, které vnímáme jako křižovatky dvou cest) jako jeho vrcholy. Potom se pohyb naší poIPSkové skupinky dá popsat jako náhodná procházka na tomto grafu, kde skupinka v každém vrcholu rovnoměrně náhodně zvolí jednu hranu a vydá se po ní (může se i vracet). Nás zajímá průměrný čas, kdy dojdou na kolej. Pojďme to popsat ještě trochu formálněji.

1. Střední hodnoty

Zopakujme si některé pojmy z teorie pravděpodobnosti. Pokud vám toho mnoho neříkají, nakoukněte do úvodního textu od prof. Matouška na *http://mj.ucw.cz/vyuka/1920/dm/dm-prob-2pp.pdf*.

- Náhodný experiment můžeme popisovat *diskrétním pravděpodobnostním prostorem,* což je konečná nebo spočetná množina Ω *elementárních jevů* – to jsou různé možné výsledky pokusu. (Příklad: házíme-li kostkou, elementární jevy jsou čísla od 1 do 6.)
- Každému elementárnímu jevu $\omega \in \Omega$ je přiřazena nějaká *pravděpodobnost* $P(\omega) \in [0, 1]$, přičemž pravděpodobnosti všech elementárních jevů se sečtou na jedničku. (Příklad s kostkou: všechny elementární jevy mají pravděpodobnost 1/6.)
- Pravděpodobnost můžeme přiřadit i *složeným jevům* – to jsou množiny elementárních jevů $A \subset \Omega$ a $P(A)$ definujeme jako $\sum_{\omega \in A} P(\omega)$. (Příklad: Jev „na kostce padne sudé číslo“ je vlastně množina elementárních jevů $\{2, 4, 6\}$ a má pravděpodobnost $3/6 = 1/2$.)
- *Náhodná veličina* (nebo také *náhodná proměnná*) říkáme libovolné funkci, která elementární jevy ohodnocuje čísly, tedy typicky funkci $X : \Omega \to \mathbb{R}$. (Příklad: Pokud házím dvojicí kostek, elementární

jevy jsou $\{1,\ldots,6\}^2$, náhodná veličina může být třeba součet čísel na obou kostkách.)

- *Booleovské podmínky* týkající se náhodných veličin se chovají jako složené jevy. Třeba $[X = 5]$ je množina všech elementárních jevů ω, pro které platí, že $X(\omega) = 5$. Tím pádem můžeme podmínkám i přiřadit pravděpodobnost a psát třeba $P[X = 5]$ (od běžné pravděpodobnosti budeme rozlišovat hranatými závorkami).
- *Střední hodnota* $\mathbb{E}[X]$ náhodné veličiny X je průměr hodnot vážený pravděpodobnostmi, tedy $\mathbb{E}[X] = \sum_{\omega\in\Omega} X(\omega)\cdot P(\omega)$. (Pozor na to, že pokud je elementárních jevů nekonečně, jedná se o nekonečnou sumu, která nemusí konvergovat!)
- Alternativně můžeme střední hodnotu napsat jako $\mathbb{E}[X] = \sum_{a\in\mathbb{R}} a\cdot P[X = a]$. To je sice nekonečná (dokonce nespočetná!) suma, ale nenulových je nejvýše spočetně členů.
- Důležitá vlastnost střední hodnoty je *linearita:* platí $\mathbb{E}[\alpha X+\beta Y] = \alpha\mathbb{E}[X] + \beta\mathbb{E}[Y]$ pro jakékoliv dvě náhodné veličiny X a Y a konstanty $\alpha, \beta \in \mathbb{R}$. (Tedy samozřejmě pokud všechny tyto střední hodnoty existují.)

Často budeme studovat, jak dlouho musíme něco zkoušet, než se nám to povede. Představte si třeba, že házíme kostkou tak dlouho, než nám padne šestka.*

Počet hodů do první šestky je nějaká celočíselná náhodná veličina X. Jedničce je rovna, pokud hned první hod je šestka, což má pravděpodobnost 1/6. Dvojce, pokud poprvé šestka nepadne a podruhé padne, tedy $5/6 \cdot 1/6$. Obecně $P[X = k] = (5/6)^{k-1} \cdot 1/6$.

Pokud nebudeme čekat zrovna na šestku, ale na nějaký jev, který pokaždé nastane s pravděpodobností p (třeba jdeme se džbánem pro vodu a zajímá nás, jestli se utrhne ucho), dostaneme $P[X = k] = (1 - p)^{k-1} \cdot p$.

Střední počet hodů do první šestky by se tedy podle definice střední hodnoty dal napsat takto:

$$\mathbb{E}[X] = \sum_{k\geq 1} k \cdot P[X = k] = \sum_{k\geq 1} k \cdot (1 - p)^{k-1} \cdot p.$$

Jenže komu by se chtělo sčítat nekonečnou řadu? Pomůžeme si tedy následující úvahou založenou na linearitě střední hodnoty: první hod provedeme vždycky a pak

* S tím, v jakém pravděpodobnostním prostoru se pohybujeme, musíme trochu opatrně, protože všech nekonečných posloupností hodů je nespočetně mnoho, takže by prostor nebyl diskrétní, čímž by sumy zdivočely a staly by se z nich integrály se všemi svízeli, které to přináší. Tak se omezíme jen na ty posloupnosti, které skončí první šestkou, a ty zbývající bez šestek zanedbáme, protože stejně mají nulovou pravděpodobnost. Věřte nám prosím, že víme, co děláme, a případně se ptejte.

s pravděpodobností $1-p$ zkoušíme dál. Dostaneme tedy:

$$\mathbb{E}[X] = 1 + (1-p) \cdot \mathbb{E}[X].$$

Ejhle, to je lineární rovnice s neznámou $\mathbb{E}[X]$, jejímž řešením je evidentně $\mathbb{E}[X] = 1/p$.[†] Platí tedy následující tvrzení:

Lemma: *(o džbánu)* Čekání na událost, která nastane s pravděpodobností p, trvá ve střední hodnotě $1/p$ kroků.

Úkol 1.1: Naše úvaha s lineární rovnicí je hezká, ale trochu neformální (byť by se dala provést zcela korektně pomocí podmíněných středních hodnot, což je analogie podmíněných pravděpodobností). Můžete zkusit vymyslet nějaký trik, jak řadu pro $\mathbb{E}[X]$ sečíst čistě prostředky matematické analýzy.

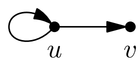
2. Bloudíme po grafech

Zpátky k procházkám po grafech. Mějme nějaký graf G, vyrazíme z vrcholu u a budeme náhodně bloudit tak dlouho, než dorazíme do vrcholu v. Počet kroků, které přitom uděláme, je nějaká náhodná veličina. Její střední hodnotě se říká *hitting time* (říkejme třeba bloudící čas) z u do v a značí se $h(u,v)$, nebo někdy přesněji $h_G(u,v)$.

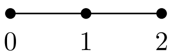
Podobně se definuje *cover time* (pokrývací čas) $c_G(u)$: ten říká, jak dlouho v průměru trvá procházka z u, než navštívíme úplně všechny vrcholy grafu.

Tyto veličiny budeme občas studovat i pro grafy s orientovanými hranami, případně se smyčkami. Začněme jednoduše…

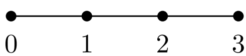
Úkol 2.1: Spočítejte hitting time $h(u,v)$ pro „ušatou hranu“:



Úkol 2.2: Spočítejte hitting time $h(0,2)$ pro cestu délky 2:



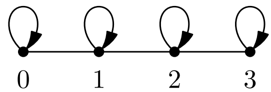
Úkol 2.3: Spočítejte hitting time $h(0,3)$ pro cestu délky 3:



[†] Máte pravdu, zase trochu švindlujeme. Tady drze předpokládáme, že $\mathbb{E}[X]$ existuje a je konečná, jinak by rovnice neměla jednoznačné řešení. Ale to si můžeme ověřit aplikováním nějakého kriteria konvergence na tu předchozí nekonečnou sumu. Mlčky budeme o všech středních hodnotách předpokládat, že existují, a slibujeme vám, že to bude vždy pravda.

Úkol 2.4: Platí v neorientovaných grafech $h(u,v) = h(v,u)$?

Úkol 2.5: Spočítejte hitting time $h(0,3)$ pro ušatou cestu délky 3. Všimněte si, že nezáleží na tom, jestli ve vrcholu 3 je ucho, protože do něj nikdy nemůžeme vstoupit.



Úkol 2.6: Spočítejte hitting time $h(0,n)$ pro cestu délky n.

Úkol 2.7: Spočítejte hitting time $h(0,n)$ pro cestu délky n s uchem ve vrcholu 0.

Úkol 2.8: Spočítejte hitting time mezi dvěma protilehlými vrcholy kružnice délky $2n$.

Úkol 2.9: Spočítejte hitting time mezi dvěma (pevně zvolenými) nejvzdálenějšími vrcholy kružnice délky $2n+1$.

Úkol 2.10: Spočítejte hitting time mezi dvěma různými vrcholy úplného grafu na n vrcholech.

Úkol 2.11: Spočítejte cover time úplného grafu na n vrcholech. Díky symetrii nezáleží na tom, ve kterém vrcholu začneme.

3. Nejbloudivější ze všech bludišť

Teď si položíme hrozně obecnou otázku: Jaké bludiště je pro náhodného bloudícího nejbloudivější? Jinými slovy chceme najít pro libovolně velké n (souvislý neorientovaný) graf na n vrcholech takový, že hitting time mezi nějakými dvěma jeho vrcholy bude největší možný. Nechť $f(n)$ je tento maximální hitting time. Co umíme říci o funkci f? Jak rychle může růst?

Úkol 3.1: Najděte co nejbloudivější bludiště na n vrcholech.

Dál pokračujte až poté, co vyřešíte většinu úloh z předchozích kapitol.

4. Kouzelná věta o odporech

Náhodné procházky v grafech mají nečekané fyzikální souvislosti. Představte si, že graf budeme číst jako schéma elektrického obvodu: hrany budou rezistory o odporu $1\,\Omega$, propojené ve vrcholech. Pak se můžeme ptát na odpor $R(u,v)$ mezi libovolnými dvěma vrcholy u a v (přesněji řečeno $R(u,v)$ bude hodnota tohoto odporu v ohmech).

Příklad: Pro cestu z 0 do n je $R(u,v) = n$ – jedná se o n jednotkových rezistorů zapojených sériově a při sériovém zapojení se odpory sčítají. Snadno spočítáme i odpor mezi protilehlými vrcholy kružnice délky $2n$: jedná se o dvě cesty délky n zapojené paralelně. Při paralelním zapojení se sčítají vodivosti (převrácené hodnoty odporů)* Takže vyjde $1/(1/n + 1/n) = n/2$.

Z čarodějného klobouku vytáhneme následující větu:

Věta: $h(u,v) + h(v,u) = 2mR(u,v)$, kde m je celkový počet hran grafu.†

Větu dokazovat nebudeme, ale dělá se to zhruba tak, že napíšete soustavu lineárních rovnic pro proudy podle Kirchhoffových zákonů, druhou soustavu rovnic pro hitting times (jako jsme dělali v předchozí kapitole) a všimnete si, že obě soustavy jsou regulární a navíc stejné, takže musí mít stejné řešení. Pořádný důkaz najdete třeba v knížce Randomized algorithms od Motwaniho a Raghavana.

Na větě je pěkné, že se dá používat oběma směry: někdy můžeme pomocí matematických úvah o náhodných procházkách vyřešit fyzikální problém, jindy zase úvahami o odporech spočítat hitting time v nějakém záludném grafu.

Příklad: Například je triviální spočítat hitting time mezi konci cesty délky n. Podle věty platí $h(0,n) + h(n,0) = 2n^2$. Navíc ze symetrie víme, že $h(0,n) = h(n,0)$. Proto $h(0,n) = n^2$. Ale můžeme získat i mnohem obecnější výsledek:

Úkol 4.1: Spočítejte $h(u,v) + h(v,u)$ mezi dvěma vrcholy stromu. Jaký je tedy nejbloudivější ze všech stromů na n vrcholech?

Úkol 4.2: Dokažte, že pokud u a v jsou spojené hranou, pak $h(u,v) + h(v,u) \leq 2m$.

Úkol 4.3: Dokažte, že cover time jakéhokoliv souvislého grafu je nejvýš $2nm$. Návod: vyberte si nějakou kostru a postupně ji obcházejte – vždy, když chcete přejít po nějaké hraně, nechte náhodnou procházku běžet tak dlouho, než se dostane na druhý konec hrany. Sestrojená procházka samozřejmě může celý graf projít dřív, než obejde kostru, ale na nerovnost to stačí.

Jelikož počet hran vždy splňuje $m \leq \binom{n}{2} \leq n^2/2$, můžeme cover time dokonce omezit funkcí n^3. Tím pádem ani žádný hitting time nepřekročí n^3. Dokážete sestrojit graf s hitting time řádově n^3 a tím vyřešit úkol 3.1?

---

* Pro úplnost dodejme, že jednotka vodivosti je siemens (S). Ale neoficiálně se používá spíš mho se značkou ℧ :)

† Všimněte si, jak elegantně jsme vybruslili z toho, že hitting time je asymetrický, zatímco odpory symetrické.

5. Univerzální procházecí posloupnosti

Bonbónek na závěr, spíš na ukázku toho, jak se dá s náhodnými procházkami čarovat.

Představte si bludiště ve čtvercové síti s určeným čtverečkem, kde začínáme bloudit. Když nám někdo dá nějakou posloupnost příkazů $\{$*vlevo*, *vpravo*, *nahoru*, *dolů*$\}$, můžeme podle ní chodit po bludišti (kdyby po nás posloupnost chtěla, abychom narazili do zdi, budeme prostě příkaz ignorovat).

O posloupnosti řekneme, že je *procházecí,* pokud pomocí ní projdeme celé bludiště. Posloupnost je *univerzální procházecí,* pokud je procházecí současně pro všechna bludiště zadané velikosti a všechny počáteční čtverečky. Čili něco jako knížka „Jak pohodlně projít každé bludiště (průvodce pro úplné začátečníky)". Tedy pohodlně... jak dlouhá taková posloupnost vlastně může být (vzhledem k velikosti bludiště)? Exponenciálně dlouhou určitě můžeme vyrobit nějakým poslepováním procházecích posloupností pro všechna bludiště. Ale překvapivě existuje i polynomiálně dlouhá!

Nejdřív úlohu trochu zobecníme. Uvažujme nějaký d-regulární graf G (to znamená, že všechny vrcholy mají stejný stupeň d). V každém vrcholu si pořídíme nějaké očíslování jeho hran čísly od 1 do d; konce jedné hrany mohou mít různá čísla. Když nám někdo zadá počáteční vrchol v_0 a nějakou posloupnost $x_1, \ldots, x_t \in \{1, \ldots, d\}$, můžeme podle této posloupnosti chodit po grafu: kdykoliv přijdeme do nějakého vrcholu, další prvek posloupnosti nám řekne, kterou hranou máme pokračovat.

Opět můžeme nadefinovat procházecí posloupnost pro daný graf a univerzální procházecí pro všechny grafy na n vrcholech. A pokud prvky posloupnosti volíme rovnoměrně náhodně, dostáváme naši známou náhodnou procházku.

Úkol 5.1: Nechť G je d-regulární graf na n vrcholech a v_0 nějaký počáteční vrchol. Dokažte, že náhodná posloupnost délky $2n^3$ je procházecí s pravděpodobností aspoň 1/2. Návod: použijte výsledky z předchozí kapitoly a *Markovovu nerovnost:* Pokud X je nějaká nezáporná náhodná veličina, tak $P[X \geq t \cdot \mathbb{E}[X]] \leq 1/t$.

Úkol 5.2: Pokračujme v předchozím úkolu: pokud bude posloupnost dlouhá $2n^5$, pak pravděpodobnost toho, že pro daný graf *není* procházecí, je nejvýš $1/2^{n^2}$.

Úkol 5.3: Spočítejte, kolik dvojic (graf, počáteční vrchol) existuje na n vrcholech. Dokažte, že pravděpodobnost toho, že náhodná posloupnost délky $2n^5$ není procházecí pro aspoň jeden graf, je ostře menší než 1.

Z toho víme, že náhodná posloupnost délky $2n^5$ je univerzální procházecí s pravděpodobností ostře větší než 0. Proto musí existovat aspoň jedna univerzální procházecí posloupnost této délky!

Náš důkaz je nicméně nekonstruktivní: vůbec jsme se z něj nedozvěděli, jak takovou posloupnost najít rychleji než vyzkoušením exponenciálně mnoha možností. Pokud víme, zatím to neumí nikdo. Budete první? :)

Úkol 5.4: Pro malé k (čtverečkové bludiště má $k = 4$) se dá posloupnost výrazně zkrátit. Zkuste to.