

Domáca úloha: Diskrétne náhodné vektory

Beáta Stehlíková, FMFI UK, 2020

Metódy riešenia úloh z pravdepodobnosti a štatistiky

Dve verzie domácej úlohy v maximálnej hodnote 100 bodov (ako 100 percent sa berie 60 bodov, tak ako pri úlohách v škole) - môžete si vybrať:

- Príklad 1 - každá úloha v tomto príklade je za 10 bodov
- Príklady 2 a 3 - každá úloha (4 úlohy v príklade 2 a úloha 3) je za 20 bodov

1 Príklad 1: Hlasovanie komisie

Majme trojčlennú komisiu, ktorá rozhoduje na základe hlasovania väčšiny jej členov. Otázky, ktoré rieši, často nie sú jednoduché a tak sa niekedy stane, že sa po čase názor člena komisie (a teda aj jeho hlasovanie) ako nesprávne. Na základe toho, že komisia funguje už dlhšie, sú známe pravdepodobnosti toho, že jednotliví členovia komisie rozhodnú pri riešení určitého problému správne. Označme členov komisie A, B, C a pravdepodobnosti ich správneho rozhodnutia p_A, p_B a p_C . Budeme predpokladať, že $p_A \geq p_B \geq p_C$. Všetky pravdepodobnosti sú aspoň $1/2$ (ak niekto nechce vyslovene škodiť, v prípade pravdepodobnosti menšej ako polovica by radšej jednoducho hlasoval naopak, aby pravdepodobnosť správnej odpovede zvýšil). Takisto, ak nebude povedané inak, členovia ko rozhodujú nezávisle.

Rozhodovanie komisie charakterizujeme náhodným vektorom (X_1, X_2, X_3) . Jeho zložky nadobúdajú hodnoty 0 a 1 podľa toho, či sa jednotliví členovia komisie (v poradí A, B, C) rozhodli nesprávne (0) alebo správne (1). Bude nás zaujímať pravdepodobnosť, že komisia v rôznych situáciách rozhodne správne.

1. Určte pravdepodobnostné rozdelenie nášho náhodného vektora - teda aké hodnoty môže nadobúdať a s akými pravdepodobnosťami.
2. Pomocou predchádzajúceho bodu ukážte, že pravdepodobnosť správneho rozhodnutia komisie je $p_A p_B + p_A p_C + p_B p_C - 2p_A p_B p_C$.
3. Nech $p_A = 0.8, p_B = 0.7, p_C = 0.6$. Ukážte, že pravdepodobnosť správneho rozhodnutia komisie je menšia, ako keby sme nechali rozhodovať iba člena A.

4. Teraz nech $p_A = 0.9, p_B = 0.8, p_C = 0.7$. Ukážte, že v tomto prípade je komisia užitočná - pravdepodobnosť jej správneho rozhodnutia je vyššia ako pravdepodobnosť správneho rozhodnutia jej ľubovoľného člena.
5. Uvažujme situáciu z bodu 4. C zistil, že sa najčastejšie mýli a preto sa namiesto samostatného hlasovania rozhodol odpísať vždy hlasovanie člena B, ktorý sedí vedľa neho. Ako sa tým zmení pravdepodobnosť správneho rozhodnutia komisie?
6. Predpokladajme, že v predchádzajúcom bode by sa člen B rozhodol radšej hádzať mincou. Minca je pravidelná, takže pravdepodobnosť jeho správneho rozhodnutia je $1/2$. Čo je z hľadiska správneho rozhodnutia komisie lepšie - ak bude odpisovať alebo ak bude hádzať mincou?
7. Ponechajme hodnoty $p_B = 0.8, p_C = 0.75$. Pri akej pravdepodobnosti p_A by bolo správne rozhodnutie komisie menej pravdepodobné ako p_A ? Teda: Ako múdry by musel byť A, aby bolo lepšie nechať rozhodnutie na neho, ako zvolávať komisiu?
8. Zistíme, nakoľko môže rozhodovanie ovplyvniť jej najmenej múdry člen. Dokážte (pri všeobecných hodnotách $1/2 \leq p_C \leq p_B \leq p_A$), že pravdepodobnosť správneho rozhodnutia komisie nemôže byť menšia ako priemer hodnôt p_A a p_B .
9. Predpokladajme teraz, že všetci traja členovia sú rovnako múdri a správne sa rozhodujú s rovnakou pravdepodobnosťou $p \in [1/2, 1]$. Označme P pravdepodobnosť správneho rozhodnutia komisie. Zistite, pre ktoré p nastávajú možnosti $P > p, P = p, P < p$.
10. V situácii z predchádzajúceho bodu určte p , pre ktoré je rozdiel $P - p$ maximálny. Teda, pri akom p je prínos komisie v porovnaní s rozhodovaním len jedného jej člena najväčší?

2 Príklad 2: Zastavovanie výťahu a počítačové simulovanie náhodných javov

V niektorých prípadoch môže byť užitočné naprogramovať si situáciu, ktorú modelujeme. Buď preto, že je príliš komplikovaná na analytický výpočet alebo preto, že si chceme spraviť kontrolu našich výpočtov. Ukážeme si to na príklade riešenia úlohy o výťahu:

V dome, ktorý má prízemie a n poschodí, je výťah. Na prízemí doňho nastúpi k ľudí. Každý z nich vystúpi nezávisle na ostatných s rovnakou pravdepodobnosťou $1/n$ na ktoromkoľvek poschodí. Aká je stredná hodnota počtu poschodí, na ktorých výťah zastaví, aby z neho niekto vystúpil?

Simuláciu tejto situácie (v jazyku R, ktorý sa budete v škole učiť budúci školský rok) si môžete vyskúšať spúšťaním kódu na stránke

Pozrime sa najskôr na funkciu `vytah`:

- Ako vstup dostáva počet poschodí `n` a počet ľudí `k`.
- V prvom riadku sa použije funkcia `sample`. Tá vygeneruje náhodný výber. Ako jej prvý argument treba zadať, z čoho sa má vyberať. V našom prípade je to vektor `1:n`, teda vektor s hodnotami `1, 2, ..., n`. Druhý parameter `size` zase hovorí, koľko hodnôt sa má vyberať. V našom prípade je to `k`. Napokon, `replace = TRUE` alebo `FALSE` určí, či má ísť o výber s návratom alebo bez návratu.
- Vzhľadom na zadanie príkladu sa nám teda do premennej `vystupovanie` uloží **realizácia náhodného vektora** $(X_1, X_2, \dots, X_k)$, pričom X_i je číslo poschodia, na ktorom vystúpi i -ty človek. Každé X_i nadobúda hodnoty $1, 2, \dots, n$, každú s rovnakou pravdepodobnosťou. Zložky tohto náhodného vektora sú nezávislé.
- Funkcia `unique` vynechá z vektora opakujúce sa hodnoty. Ak napríklad dostane ako vstup vektor s hodnotami `1, 4, 1, 3, 3, 1`, výstupom bude vektor s hodnotami `1, 4, 3`. V našom prípade dostaneme zoznam poschodí, na ktorých niekto vystúpil.
- Keďže funkcia `length` vracia dĺžku vektora a `return` určí, čo má naša funkcia vracať, výstupom bude dĺžka vektora so zoznamom poschodí. Inak povedané, bude to počet poschodí, na ktorých niekto vystúpil.

Čo môžeme s touto funkciou robiť:

- Prirodzene, môžeme ju zavolať s nejakými parametrami, napr. `vytah(6, 6)`. Dostaneme jedno číslo, zodpovedajúce jednej situácii so 6 ľuďmi v 6-poschodovom dome. Napríklad 3, čo znamená, že ľudia vystupovali na troch rôznych poschodiach.
- Kvôli náhodnosti dostaneme pri opätovnom za volaní funkcie nový výsledok. Chceli by sme funkciu spustiť veľa krát a zaznamenať si výsledky. Na to slúži funkcia `replicate`. Treba. jej povedať, koľko opakovaní má spraviť (my ich budeme mať 100 000) a čo má opakovať.
- Pri simulácií sa generujú náhodné čísla. Funkciou `set.seed` dosiahneme to, že pri opätovnom spustení programu sa vygenerujú tie isté náhodné čísla.

Čo spraviť s výsledkami stotisíc simulácií, ktoré máme vo vektor `vysledky`?

- Pomocou `table` si môžeme pozrieť, koľkokrát bola výsledkom jednotka, dvojka, atď. V našom prípade sa 7 krát stalo, že všetci vystúpili na tom istom poschodí. 1995 krát vystupovali na dvoch poschodiach. Už častejšie, 23 296 krát, to boli tri poschodia. Atď.

- Ak túto tabuľku vložíme do funkcie `prop.table`, dostaneme namiesto počtu prípadov ich podiel na celkovom počte simulácií. (0.25 by potom napríklad znamenalo 25 percent.)
- Pomocou funkcie `barplot` si môžeme nakresliť graf a namiesto tabuľky sa na výsledky pozrieť vizuálne.
- Podľa zadania príkladu nás zaujíma stredná hodnota počtu poschodí, na ktorých niekto vystúpil. Ak teda spravíme priemer našich hodnôt vo vektore `výsledky`, dostaneme odhad tejto strednej hodnoty. V našom prípade je to 3.98732.

Úlohy pre vás:

1. Ponechajte si v programe definíciu funkcie `vytah` a nastavenie `set.seed`. Potom niekoľkokrát pod seba skopírujte generovanie simulácií a výpočet priemeru (pre $n = 6, k = 6$, ako vo vzorovom kóde). Aké hodnoty priemerov ste dostali?
2. Zopakujte predchádzajúci bod pre prípad, že chcete analyzovať situáciu, v ktorej má budova 10 poschodí a do výťahu nastupuje 7 ľudí.
3. Odvodte strednú hodnotu analyticky pre všeobecnom prípade, pre ľubovoľné n a k . *Návod: Uvažujme indikátory Y_i , ktoré vyjadrujú, či na i -tom poschodí niekto vystúpil. Jednoducho vieme vypočítať pravdepodobnosť, že sa takýto indikátor rovná nule, teda že na danom konkrétnom poschodí nevystúpil nikto. Počet poschodí, na ktorých niekto vystúpil, je potom súčet týchto indikátorov.*
4. Vypočítajte presné stredné hodnoty pre situácie z prvých dvoch prípadov. Mali by sa podobáť na výsledky zo simulácií, ktoré ste robili.

3 Príklad 3: Príklad zo skúšok americkej *Society of actuaries*

A car dealership sells 0, 1, or 2 luxury cars on any day. When selling a car, the dealer also tries to persuade the customer to buy an extended warranty for the car. Let X denote the number of luxury cars sold in a given day, and let Y denote the number of extended warranties sold, and suppose that

$$P(X = x, Y = y) = \begin{cases} 1/6 & \text{for } (x, y) = (0, 0), \\ 1/12 & \text{for } (x, y) = (1, 0), \\ 1/6 & \text{for } (x, y) = (1, 1), \\ 1/12 & \text{for } (x, y) = (2, 0), \\ 1/3 & \text{for } (x, y) = (2, 1), \\ 1/6 & \text{for } (x, y) = (2, 2). \end{cases}$$

What is the variance of X ?