|  |
| --- |
|  |
| Problém zberateľa kartičiek |
|  |

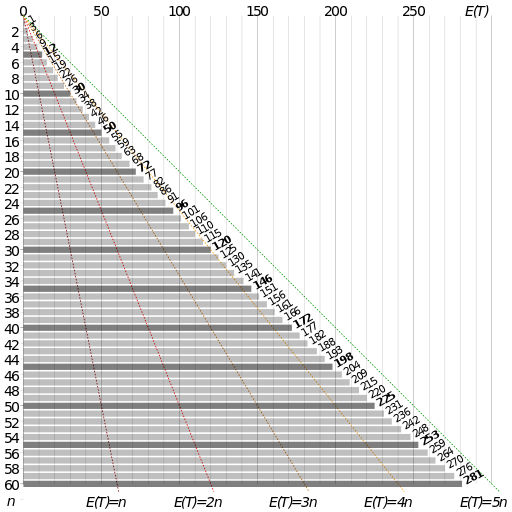
|  |
| --- |
| Petra Kochaniková |



**1.Úvod**  
V dnešnej dobe plnej marketingu sa čoraz častejšie stretávame s nápadom predajcov pridávať k jednotlivým výrobkom rôzne kartičky alebo nálepky, ktoré si potom zberatelia lepia do svojich zberateľských albumov a po získaní celej sady nálepiek si môžu album vymeniť za nejakú vecnú cenu. Preto sa mnohým ľuďom, najčastejšie rodičom malých ratolestí, natíska otázka či je výhodné začínať so zbierkou alebo radšej zainvestovať a svojmu dieťaťu kúpiť nejakú hračku podľa jeho výberu. Okrem zisťovania informácie, koľko výrobkov potrebujeme kúpiť aby sme získali úplnú sadu kartičiek a teda zaplnili cely album sa na tento problém kupujúci môžu pozerať z rôznych uhlov. Môžeme chcieť zistiť koľko rôznych kartičiek získame pri kúpe určitého poctu výrobkov alebo koľko kusov výrobku potrebujeme kúpiť, aby sme získali vybrane typy kartičiek. Každý zákazník by mal mat vedomosť o tom, koľko výrobkov potrebuje nakúpiť, aby získal ponúkanú vecnú cenu a až potom sa rozhodnúť či sa mu oplatí do tejto zberateľskej činnosti vôbec púšťať.  
  
V úvode sme spomínali, že tento problém sa týka hlavne dnešnej marketingovej doby, no ak sa pozrieme do minulosti zistíme, že prvá zmienka tohto problému pochádza už z roku 1907   
v publikácii De Mensura Sortis, ktorú napísal A. De Moivre. O tento problém sa tiež zaujímali matematici ako Euler, Laplace, Willam Feller a v literatúre, na rôznych prednáškach či na známom kanále videí youtube sa pravidelne objavuje ešte aj v dnešnej dobe.  
  
**2.Matematický model**  
Ako vidíme tento problém spadá do odvetvia pravdepodobnosti, a preto je vhodne preformulovať problém zbierania kartičiek do matematického modelu . Typickým zadaním úlohy z teórie pravdepodobnosti, ktorá korešponduje s vyššie uvedeným popisom reálneho problému je nasledujúca formulácia. Mame 2 hráčov, ktorí niekoľkokrát hádžu viacstennou kockou a majú pevne určený počet rôznych strán, ktoré musia hráčovi padnúť. Aká je pravdepodobnosť, že pri danom počte hodov kockou padne určitý počet rôznych strán? Resp. Koľkokrát musí hráč hodiť kockou, aby padol zvolený počet strán?  
  
Po úlohách typu hodu kockou sú v pravdepodobnosti obľúbené priehradkové úlohy. Tie spočívajú v žrebovaní čísel, následnom vhode guľôčky s rovnakým číslom do priehradky a vrátení čísla do osudia, z ktorého sa žrebuje. Potom môžeme zisťovať koľko čísel potrebujem vyžrebovať aby sme mali zaplnený vybraný počet resp. všetky priehrady. Taktiež je možnosť skúmať aká je pravdepodobnosť zaplnenia všetkých priehradok ak vyžrebujeme iba určitý počet čísel.  
  
  
Teraz sa pozrieme na matematické riešenie danej úlohy. Predpokladáme, že kartičky sú k výrobkom rovnomerne rozdelene, teda sú pribaľované s rovnakou pravdepodobnosťou. Náhodná veličina určuje počet výrobkov, ktorý si zákazník musí kúpiť pokiaľ chce získať celu sadu kartičiek a my potrebujeme zistiť strednú hodnotu, rozptyl a pravdepodobnostne rozdelenie tejto náhodnej veličiny, aby sme získali odpovede na vyššie kladené otázky. Je dôležité uvedomiť si, že ide o postupnosť nezávislých nákupov s rovnakou pravdepodobnosťou obdŕžania každej kartičky, to znamená, že sa neznižuje pravdepodobnosť obdŕžania kartičky ktorú sme už vo výrobku našli.  
Budeme sa zaoberať riešením problému v diskrétnom prípade a k výpočtom budeme používať metódy klasickej pravdepodobnosti.  
Najprv budeme chcieť získať jednu úplnú sadu nálepiek.

**3.Zisk úplnej sady kartičiek**V nasledujúcej časti uvediem výpočet, koľko výrobkov si musí človek v obchode kúpiť, aby získal jednu úplnú sadu kartičiek. Predpokladajme, že počet rôznych typov kartičiek v sade je pevne daný, označme ho *n*. Predpokladajme tiež, že kartičky sú do výrobkov pridávané rovnomerne a náhodné. Aby sme zistili koľko výrobkov mame nakúpiť, aby sme získali úplnú sadu, potrebujeme vypočítať strednú hodnotu poctu nákupov .  
Označím výrazom ** počet zakúpených výrobkov k získaniu celej zbierky.  
Predpokladajme , že už mam zakúpených *i* rôznych kartičiek a  je počet výrobkov, ktoré ešte potrebujem kúpiť, aby som získala novu kartičku, teda po tomto počte nákupov získame i+1 rôznych kartičiek.   
Potom celkový počet výrobkov k zakúpeniu celej zbierky * vypočítam ako:*  
  
Všimneme si, že náhodná veličina  je určená geometrickým rozdelením s parametrom . Geometrické rozdelenie je čakanie na prvý úspech. Potom už vieme, že pravdepodobnosť, zisku výrobku, ktorý ešte v zbierke nemáme je určená vzťahom:  
  
  
  
  
  
  
Kde pravdepodobnosť, že získame výrobok, ktorý v zbierke nemáme je vyjadrená základným pravdepodobnostným vzťahom, kde v menovateli sú všetky možnosti a v čitateli je počet priaznivých možností:  
 pre   
Vidíme že pravdepodobnosti  sú rôzne, preto aj rozdelenie náhodných veličín  je navzájom rôzne a taktiež majú rôznu strednú hodnotu, ktorú vyjadríme vzťahom:



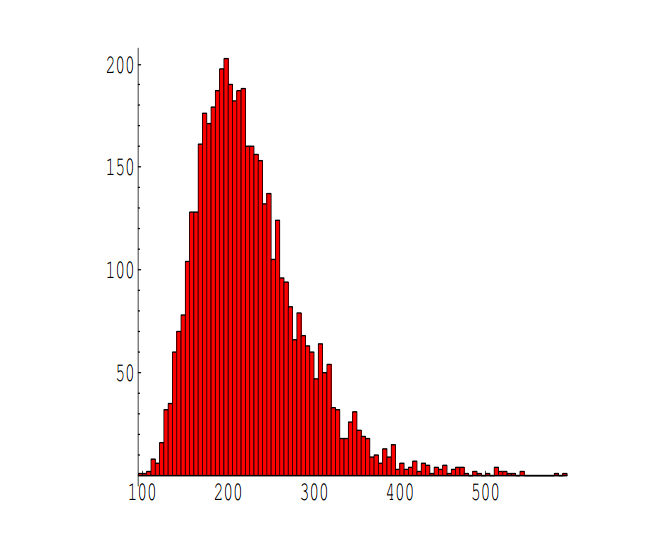
Takže, konečne pre strednú hodnotu počtu zakúpených výrobkov dostávame hodnotu:  
  
  
Aby sme zistili ako sa správa stredná hodnota počtu zakúpených výrobkov potrebných k nazbieraniu celej sady kartičiek pri zväčšujúcom sa poctu kartičiek *n*, budeme   
skúmať  pre .   
 Dostávam vzťah:  
, kde  je Eulerova konštanta  
 Teda teraz si už jednoduchým dosadením do vzorca vieme vypočítať, koľko nákupov musíme uskutočniť aby sme získali sadu všetkých *n* kartičiek.  
  
  
  
  
  
Na obrázku vidíme vypočítané stredné hodnoty pre n={0,…,60}.  
Všimneme si, že napr. pre n=50 potrebujeme nakúpiť približne 225 výrobkov.  
  
  
  
  
  
Ďalej uvedieme vzťahy na výpočet rozptylu. Aby sme mohli použiť vzorec na výpočet rozptylu potrebujeme si určiť hodnoty  a .  
Najprv určíme strednú hodnotu z druhých mocnín náhodnej veličiny  pomocou sumy:



ktorú zderivujem:



 dostanem vzťah: 

A konečne môžeme vypočítať hľadaný rozptyl:  
Keďže vieme, že náhodné veličiny  sú nezávislé hľadaný rozptyl  vyjadríme pomocou vzorca:  
  
Všimneme si, že tento vzťah sa dá zjednodušiť pomocou nasledujúcich úprav a odhadov:A po dosadení dostávam výslednú hodnotu: ****Teraz si uvedomíme čo sme vlastne počítali. Pre n=50 nám vyjde vysoká číselná hodnota to znamená, že môžeme očakávať vysoké odchýlky od strednej hodnoty teda počtu nákupov k získaniu celej sady nálepiek. Demonštrujeme to na nasledujúcom grafe, kde sme nasimulovali 5000 pokusov nazbierania celej sady nálepiek a pozorovali sme koľko nákupov musíme uskutočniť. ****  
Najväčší počet nákupov bol 592 a najmenšia počet bol iba 99. Hodnota všetkých počtov nákupov, ktorá sa opakovala najviac krát bola 225, 161 čo je približne hodnota, ktorú sme očakávali na základe nášho teoretického výpočtu :224,96 .  
  
Práve sme si predstavili základný problém zberateľa kartičiek a v ďalšej časti textu sa pozrieme na jeho drobné modifikácie a rozšírenia.  
  
  
  
   
 **4. Zisk predom daných l kartičiek z celkového počtu n kartičiek**



V tejto časti budeme chcieť získať predom daných l kartičiek z celkového počtu n kartičiek.  
Potrebný počet zakúpených výrobkov k získaniu l kartičiek zapíšeme pomocou podobného zápisu ako v 3. časti tohto textu:  
  
kde náhodná veličina  značí počet výrobkov, ktoré potrebujeme kúpiť, aby sme   
dostali (i+1)-vú kartičku z danej l-tice vybraných kartičiek.  ma geometrické rozdelenie s parametrom  pre   
  
Teraz podľa rovnakého postupu určíme strednú hodnotu počtu výrobkov, ktoré musíme zakúpiť aby sme dostali z danej l-tice typ kartičky, ktorú ešte nemáme:  
  
  
A následne stredná hodnota počtu nákupov, ktoré musíme urobiť, aby sme získali danú l-ticu kartičiek je daná vzorcom:  
  
  
Tento vzťah môžeme v realite uplatniť ak napríklad výrobca pribaľuje do výrobku rôzne živočíchy no my budeme mať záujem iba o nálepky cicavcov.



**5. Pravdepodobnosť získania celej sady kartičiek po r nákupoch.**Posledná časť tejto práce sa zaoberá počítaním pravdepodobností, čo znamená, že túto aplikáciu už zberateľ kartičiek v praxi nevyužije, ale nachádzame uplatnenie v biológii a rôznych laboratóriách a nemocniciach. Majme nejaké kontaminované bunky v tele pacientov rôznymi vírusmi môže nás zaujímať aká je pravdepodobnosť, že po n testoch sa nijaký vírus v tele pacienta nezopakoval viac ako jeden krát. Túto otázku môžeme rôzne obmieňať čo by malo za následok iba jednoduchú zmenu vzorcov, ktoré budú ďalej uvedené. V ďalšej časti naďalej používam už zaužívané označenia z karičkového problému, no čitateľ si to jednoducho môže previesť na termíny z biológie.   
  
Budeme predpokladať, že naraz zakúpime r výrobkov (r>n). Bude nás zaujímať pravdepodobnosť, že sme obdŕžali kompletnú sadu n kartičiek. Tuto pravdepodobnosť označím ako .   
  
Označme  situáciu, že po nákupe r kartičiek nám chyba i-tá kartička. Vyberanie kartičiek sú variácie s opakovaním, preto všetkých možnosti je a čitateľ vyjadruje počet priaznivých možností teda v každom ďalšom nákupe môžeme zakúpiť akúkoľvek kartičku okrem práve tej jednej i-tej. Potom pravdepodobnosť tejto situácie určená vzťahom:  
  
  
Ďalej si premyslíme ako vyjadríme pravdepodobnosť, že nám chyba k typov kartičiek k<n, dostávame:



  
A konečne pravdepodobnosť, že pri kúpe r výrobkov dostaneme úplnú sadu kartičiek vypočítame pomocou doplnkovej pravdepodobnosti a využitia princípu inklúzie a exkluzie:  
=  
  
Teda po poprehadzovaní členov dostávame hľadanú pravdepodobnosť:  
  
  
  
Na obrázku vidíme pravdepodobnosť získania celej sady kartičiek pre n=50  
  
  
  
  
  
  
  
Na záver ešte využijeme práve získanú pravdepodobnosť k zisteniu rozdelenia .   
  
  
  
  
  
  
  
Všimneme si, že očakávame malé hodnoty, pretože ak túto pravdepodobnosť preformulujeme z matematického textu do slovnej interpretácie dostávame otázku. Aká je pravdepodobnosť, že k zakúpeniu celej zbierky potrebujem práve r kartičiek. Pre n=50 si túto úvahu môžeme overiť na priloženom grafe:







**6.Zhrnutie**Na prvý pohľad sa môže zdať, že ide iba o akési jednoduché a zábavné cvičenie pravdepodobnostnej aplikácie, ktoré môže spočítať každý človek, ktorý ovláda metódy klasickej pravdepodobnosti a základné operácie matematickej analýzy. Pravdou však je, že rôzne, na výpočet náročné modifikácie tohto problému sa využívajú vo fyzike, biológii, elektrotechnike a optimalizácii.  
 **7.Zdroje**

1. Matematika náhody, J. Anděl, 2000, matfyzpress
2. Pravdepodobnost a matematicka statistika, V. Dupač, M. Hušková, 2001, Nakladatelství Karolinum
3. History of Probability and Statistics and Their Applications before 1750, A. Hald, Wiley-Interscience,2003
4. http://sarielhp.org/teach/notes/rand\_alg/lec/04\_occupancy.pdf
5. <http://en.wikipedia.org/wiki/Coupon_collector's_problem>
6. <https://www.math.ucdavis.edu/~tracy/courses/math135A/UsefullCourseMaterial/couponProblem.pdf>
7. <http://web.cs.wpi.edu/~hofri/CCP.pdf>
8. <http://www.d.umn.edu/math/Technical%20Reports/Technical%20Reports%202007-/TR%202012/dai.pdf>
9. <http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1806&context=cstec>