Václav Kůs; Tomáš Hobza; Jitka Hanousková

Jak se testují pravděpodobnostní modely

Rozhledy matematicko-fyzikální, Vol. 87 (2012), No. 1, 5–13

Persistent URL: http://dml.cz/dmlcz/146452

Jak se testují pravděpodobnostní modely

Václav Kůs, Tomáš Hobza a Jitka Hanousková, FJFI ČVUT Praha

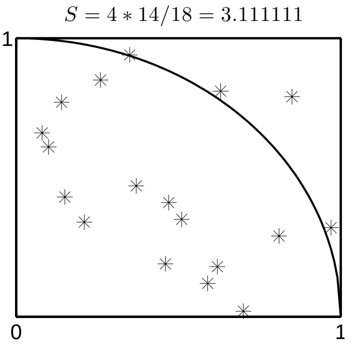
Abstract. Several physical random number generators are described and tested in frame of standard probability models. Applying the Pearson's χ^2 goodness-of-fit test statistics we explore the regularity of common playing dice and we test the distribution of bombarding scheme of V2 rockets during the Second World War. Further corresponding tasks for practical home applications are proposed.

Jedna zajímavá turistická úloha

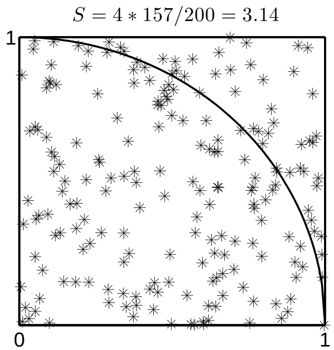
Máme dán kruh se středem v počátku o poloměru 1 a naším úkolem je spočítat jeho obsah. Problém je, že právě ležíme na pláži a nemůžeme si vzpomenout na hodnotu čísla π a k dispozici nemáme nic víc než vlastní hlavu. Jsme schopni úlohu uspokojivě vyřešit? Ne? Zkuste vymyslet nějaký trik, a pokud se vám to nepovede, podívejte se na naše navržené řešení.

Navrhneme následující algoritmus:

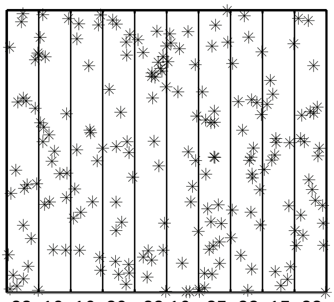
(a) Uzavřeme čtvrtkruh do čtverce s délkou strany 1, jak je nakresleno na obrázcích 1(a)(b).



(a)



(b)



(c)

Obr. 1: Monte Carlo výpočet obsahu kruhu S a testování rovnoměrnosti generátoru

(b) Cyklus: Obejdeme postupně v našem okolí několik bdících turistů (dejme tomu 18, resp. 200) a požádáme každého z nich, aby nám řekl nějaké libovolné přirozené číslo mezi 1 a 99. Získaná čísla si buď zapamatujeme, nebo zapíšeme pro jistotu do písku. Předpokládejme, že jsme tímto postupem získali čísla uvedená v prvním řádku následující tabulky 1 (hodnoty pro 200 opakování nejsou tabelovány, ale jsou vyneseny do obr. 1(b)).

(c) Cyklus: Obejdeme postupně dalších (18, resp. 200) turistů, a získáme tak druhou sadu náhodně volených čísel mezi 1 a 99 odpovídající druhému řádku tabulky 1.

řádek 1: 63, 21, 8, 81, 10, 37, 97, 46, 14, 70, 35, 15, 51, 59, 26, 85, 62, 47
řádek 2: 81, 34, 66, 29, 61, 47, 32, 19, 77, 2, 94, 43, 35, 12, 85, 79, 18, 41
$x^2 + y^2 < 1$: N, A, A, A, A, A, N, A, A, A, N, A, A, A, A, N, A, A

Tab. 1: Obsah kruhu S

(d) Všechna získaná čísla vydělíme 100 a vyneseme ve formě dvourozměrných bodů do grafu na obr. 1(c) tak, že x-ová souřadnice bodu je vždy číslo z prvního řádku tab. 1 a y-ová souřadnice odpovídá číslu z druhého řádku tab. 1, tzn. vyneseme body v desetinných číslech mezi 0,01 a 0,99.

(e) Dále následuje nejkomplikovanější část algoritmu. Je totiž třeba zjistit, zda daný vynesený bod (x, y) leží pod nebo nad čtvrtkružnicí.

To vyžaduje umocnit zpaměti obě souřadnice z každé vynesené dvojice čísel (x, y) a zjistit, zda jejich součet $x^2 + y^2$ je menší než jedna (rovno jedné neuvažujeme, jedná se totiž o jev prakticky nemožný. Z matematické analýzy víme, že hranice kruhu je množina míry nula vzhledem k plošné míře $\mathbb{R}^2$).

(f) Spočítáme, kolik vynesených bodů splňuje podmínku z předchozího bodu (e). Tedy následuje cyklus přes počet bodů (v našem případě 18 bodů z tab. 1): IF $x^2 + y^2 < 1$ THEN $P := P + 1$ (P nastaveno předem na nulu).

(g) Obsah kruhu o poloměru 1 vypočítáme jako čtyřnásobek počtu bodů splňujících podmínku z bodu (e) dělený celkovým počtem uvažovaných bodů, tzn. $S := 4 * P/18$ (resp. $S := 4 * P/200$). V našem případě pomocí třetího řádku tab. 1 zjistíme, že obsah jednotkového kruhu je $S = 4 * 14/18 = 3{,}111111$. Pro 200 bodů získáme obsah $S = 4 * 157/200 = 3{,}14$, který je přesnějším přiblížením.

Princip algoritmu: Pokryjeme čtverec $(0, 1) \times (0, 1)$ v prvním kvadrantu rovnoměrně rozloženými body a poměr obsahu čtvrtkruhu ku obsahu jednotkového čtverce porovnáme s poměrem počtu bodů, které padly dovnitř čtvrtkruhu, ku celkovému počtu bodů ve čtverci. Postup má jisté předpoklady a též možnosti zobecnění. Zkuste některé navrhnout.

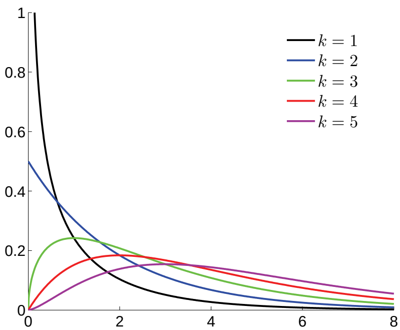
Může se stát, že na pláži zrovna vůbec žádní turisté nejsou, nebo je jich velmi málo. Potom jako generátor náhodných čísel můžeme použít např. přesýpání písku nebo budeme počítat počet mušlí vyplavených mořem za určitý čas apod. Bod (e) předpokládá, že umíme zpaměti mocnit čísla mezi 1 až 99 na druhou. Pokud ne, snížíme interval např. na čísla pouze mezi 1 a 20. Tím však můžeme ztratit dostatečnou přesnost výsledku. Z teorie pravděpodobnosti máme zajištěnu pouze asymptotickou přesnost našeho počínání, tzn. náš výpočet bude tím přesnější, čím více dvourozměrných bodů rozložíme a čím rovnoměrnější pokrytí celého čtverce uděláme. Se všemi problémy se ale dokážeme vypořádat; dnes můžeme snadno získat na počítači velmi velký počet bodů pomocí různých generátorů náhodných čísel. Jediný závažný problém, na kterém náš výsledek podstatně závisí, je kvalita rovnoměrného rozložení takto vygenerovaných dvojrozměrných bodů. Teorie pravděpodobnosti a matematická statistika vyvinuly za účelem testování kvality generátorů řadu postupů. V následujícím odstavci se seznámíme s jedním z nich a následně otestujeme náš „turistický generátor“ náhodných čísel z obr. 1.

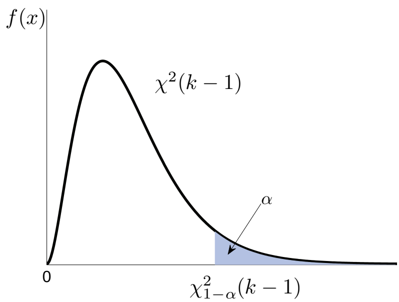
Statistický model

Mějme nějaké rozložení pravděpodobnosti P na množině A možných výsledků experimentu. Oblast A rozdělme na $k > 1$ podoblastí A_i, $i = 1, 2, \ldots, k$. Pravděpodobnost výskytu nějakého výsledku experimentu v oblasti A_i označíme $p_i = P(A_i)$ pro každé i a zavedeme pravděpodobnostní model ve formě vektoru tzv. teoretických pravděpodobností $p = (p_1, p_2, \ldots, p_k)$. Nyní mějme experiment spočívající v provedení n pokusů a sledujme, kolik z celkového počtu pokusů padlo do dané podoblasti A_i. Tyto pozorované hodnoty označme m_i; jejich velikosti nabývají hodnot čísel z množiny $\{0, \ldots, n\}$. Ptáme se, zda napozorovaný vektor hodnot $m = (m_1, m_2, \ldots, m_k)$ odpovídá předpokládanému rozdělení P na oblasti A. Aby tomu tak bylo, musely by být vektory teoretických četností np a napozorovaných četností m blízko sebe. Pro účel tohoto srovnání britský matematik K. Pearson (1857–1936) navrhl použít vzdálenostní míru $\chi^2(m, p)$ (tzv chí-kvadrát) a odvodil její asymptotické rozdělení pravděpodobnosti, které se nazývá Pearsonovo. V teorii pravděpodobnosti tento fakt vyjadřujeme vztahem

$$\chi^2(m, p) = \sum_{i=1}^{k} \frac{(m_i - np_i)^2}{np_i} \sim \chi^2(k-1)$$

a tvar rozdělení $\chi^2(k-1)$, tzn. odpovídající hustota pravděpodobnosti $\chi^2(k-1)$ je vidět na obr. 2. Tvar rozdělení závisí na hodnotě parametru k, respektive hodnotě $k-1$, kterou nazýváme *počtem stupňů volnosti*. Čím více se liší experiment (m) od předpokládané skutečnosti (np), tím větší je hodnota statistiky $\chi^2(m, p)$, čímž se dostáváme do stále vzdálenějších oblastí asymptotického rozdělení $\chi^2(k-1)$ (obr. 2).





Obr. 2: Tvar rozdělení $\chi^2(k)$ a kritická hodnota testu $\chi^2_{1-\alpha}(k-1)$

Statistiku $\chi^2(m,p)$ nazýváme testovací statistika a ta počítá odchylku dvou vektorů m a np jako váženou euklidovskou vzdálenost, tzn. součet vážených kvadratických odchylek jednotlivých složek obou vektorů m a np. Testování předpokládaného modelu na základě statistiky $\chi^2(m,p)$ pak provádíme tak, že pokud příslušná testovací statistika překročí kritickou hodnotu rozdělení,

$$\chi^2(m,p) > \chi^2_{1-\alpha}(k-1),$$

hypotézu o shodě experimentu a pravděpodobnostního modelu zamítáme. Význam čísla $\chi^2_{1-\alpha}(k-1)$ je opět znázorněn na obr. 2. Jde o hraniční hodnotu, za kterou již považujeme napozorovanou realizaci za značně nepravděpodobnou, tzn. vyšrafovaná plocha na obrázku pod křivkou je již velmi malá a předpokládáme, že je menší než pevně zvolené číslo $\alpha \in (0,1)$. Číslo α volíme před testováním modelu a má význam spolehlivosti testu, tzn. vyjadřuje maximální povolenou pravděpodobnost toho, že zamítneme předpokládaný model, i když je ve skutečnosti pravdivý. Volme všude v tomto článku $\alpha = 0{,}05$; následně kritická hodnota pro zamítnutí hypotézy o modelu je značena $\chi^2_{0,95}(k-1)$ a je závislá pouze na počtu podoblastí A_i, na které je rozdělena oblast A. Kritické hodnoty v závislosti na α jsou tabelovány ve statistických tabulkách nebo jsou dostupné prostřednictvím počítačových procedur a knihoven, např. [4], [5]. Další podrobnosti k uvedenému statistickému modelu testování hypotéz je možné nalézt v [1].

Praktické aplikace a úlohy

1. Kvalita turistického generátoru: Otestujeme rovnoměrnost našeho generátoru pro 200 náhodných bodů. Na obr. 1(c) jsme interval $(0,1)$ rozdělili na 10 dílků ($k = 10$). Určíme počet vygenerovaných bodů, jež padly do jednotlivých subintervalů, a kolik z celkového počtu 200 bodů by do jednotlivých subintervalů mělo padnout za předpokladu rovnoměrného rozdělení, tzn. $m = (23,19,16,20,22,16,25,22,15,22)$, $np = (20,20,20,20,20,20,20,20,20,20)$. Spočteme hodnotu testovací statistiky $\chi^2(m,p) = 5{,}2$ a porovnáme tuto hodnotu s kritickou hodnotou testu pro $k = 10$, která je rovna $\chi^2_{0,95}(9) = 16{,}92$. Tuto kritickou hodnotu jsme tedy nepřesáhli, a náš turistický generátor rovnoměrného rozdělení náhodných čísel na intervalu $(0,100)$ tak můžeme považovat za kvalitní. V [2] je zkonstruováno a kvalitativně porovnáno mnoho dalších kongruenčních a jiných typů generátorů s praktickými ukázkami softwaru na CD včetně aplikací při Monte Carlo simulacích.

Úloha 1: Vymyslete svůj vlastní fyzikální generátor náhodných čísel, otestujte ho pomocí χ^2 testu a použijte vaše vlastní realizace pro výpočet obsahu kruhu z obr. 1. Byl váš výsledek lepší než ten pořízený pomocí turistického generátoru?

2. Pravidelnost používané hrací kostky: Bylo provedeno 180 hodů standardní hrací kostkou zakoupenou v obchodě od neznámého výrobce. V tabulce 2 jsou uvedeny počty jednotlivých výsledků našich hodů. Je tato kostka pravidelná, má správně vyvážené těžiště a je materiál, ze kterého je vyrobena, homogenní? Odpovědět na tuto otázku můžeme opět pomocí χ^2 testu. Snadno spočítáme testovací statistiku χ^2 a porovnáme ji s kritickou hodnotou pro $k = 6$, jak je uvedeno v posledním řádku tabulky 2, z čehož vidíme, že naše hrací kostka se prokázala jako nepravidelná, preferující číslo '4' a podceňující číslo '3' na opačné straně kostky. Tedy těžiště této hrací kostky je zřejmě vychýlené.

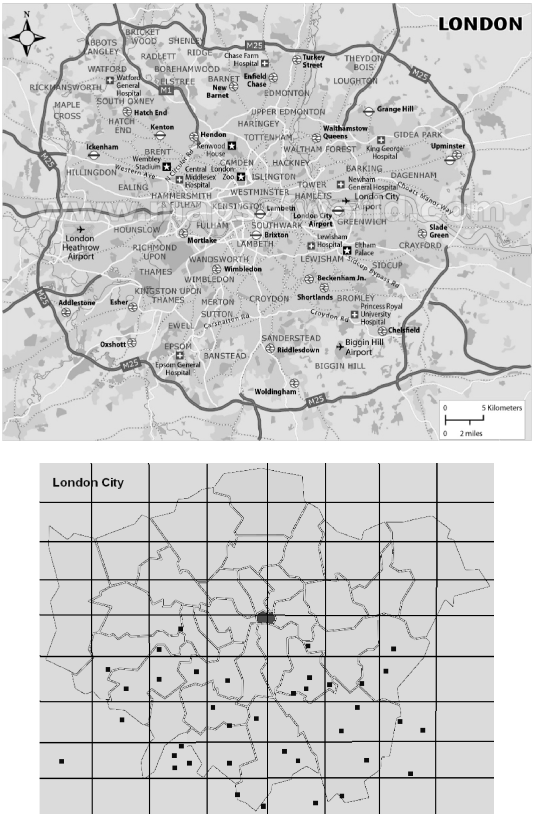
Výsledek na kostce	'1'	'2'	'3'	'4'	'5'	'6'
Četnost hodů (vektor m)	28	31	19	43	24	35
Teoretická četnost (vektor np)	30	30	30	30	30	30
Hodnota $(m_i - np_i)^2/np_i$	0,133	0,033	4,033	5,633	1,200	0,833
Testovací statistika $\chi^2(m,p) = 11{,}87 > 11{,}07 = \chi^2_{0,95}(6-1)$						

Tab. 2: Hrací kostka

Úloha 2: Vezměte všechny domácí hrací kostky ve vašem okolí, s každou z nich proveďte sadu vlastních hodů a otestujte pomocí χ^2 testu jejich pravidelnost. Našli jste nějakou špatně vyrobenou a nevyváženou kostku, která preferuje některé určité výsledky hodů před jinými?

3. Rozložení řízených střel na Londýn: Během druhé světové války byl bombardován Londýn německými raketami V2, které dopadaly do různých částí jižního Londýna. Nebylo však zřejmé, zda bylo mířeno s větším či menším úspěchem na nějaké konkrétní cíle, aby mohla být učiněna vhodná protiopatření, či zda střely dopadaly víceméně náhodně. Jižní část Londýna byla tedy rozdělena na sektory – čtverce o ploše 0,25 km^2 – a v tabulce 3 jsou uvedeny pozorované četnosti – do kolika oblastí dopadlo kolik řízených střel (viz [3]).

Úkolem je zjistit, zda rozdělení dat v tabulce 3 lze považovat za Poissonovo rozdělení s frekvenční funkcí popsanou teoretickými pravděpodobnostmi $p_i = \lambda^i e^{-\lambda}/i!$, $i = 0, 1, 2, \ldots$, které závisí na neznámém parametru $\lambda > 0$ (viz [1], [3]). To by pak v souladu s pravděpodobnostním modelem řídkých jevů znamenalo, že nebylo mířeno na konkrétní cíle.



Obr. 3: Mapa Londýna s ilustrativním rozdělením na sektrory

Počet střel v oblasti	0	1	2	3	4 a více	$\sum$
Počet takových oblastí (m)	227	211	92	35	8	573
Teoretické četnosti pro λ (np)	225,17	210,32	98,22	30,58	8,71	573
Hodnota $(m_i - np_i)^2/np_i$	0,015	0,002	0,394	0,639	0,058	1,108
Testovací statistika $\chi^2(m,p) = 1{,}108 < 7{,}81 = \chi^2_{0,95}(5-1-1)$						

Tab. 3: Řízené střely na Londýn

Postupujeme tak, že nejdříve neznámý parametr λ, vyjadřující průměrný počet střel v oblasti, odhadneme z dostupných napozorovaných hodnot ve formě aritmetického průměru

$$\lambda = \frac{\sum x_i}{n} = 0{,}934,$$

kde $\sum x_i$ je celkový počet střel a n je celkový počet uvažovaných sektorů. Ze statistického hlediska jde o maximálně věrohodný odhad parametru λ. Pomocí tohoto odhadu napočteme z výše uvedeného vzorce pro Poissonovo rozdělení jednotlivé diskrétní teoretické četnosti vektoru np do třetího řádku tabulky 3. Z hodnot jednotlivých vážených kvadratických odchylek obou vektorů m a np (čtvrtý řádek tabulky 3) je viďet, že experimentální a teoretické četnosti si vzájemně velmi dobře odpovídají, o čemž svědčí i nízká hodnota naší testovací statistiky $\chi^2(m,p) = 1{,}108$ nepřekračující kritickou hodnotu testu $\chi^2_{0{,}95}(5-1-1) = 7{,}81$. Tedy rozdělení lze považovat za Poissonovo a to znamená, že tímto statistickým testem nebylo prokázáno, že by raketami bylo mířeno na nějaké konkrétní cíle jižního Londýna (obr. 3).

Všimněte si, že jsme v případě tohoto testu s jedním neznámým parametrem λ snížili počet stupňů volnosti o další jedničku. To je způsobeno právě potřebou odhadování parametru λ ze zaznamenaných dat, což přináší menší počet stupňů volnosti této úlohy.

Úloha 3: Vezměme si kopii mapy nějaké oblasti, města, podmořského dna apod., nakreslete na ni čtvercovou nebo obdélníkovou síť a pomocí házené mince se trefujte na vodorovně položenou mapu; zaznamenávejte počty dopadů mincí do jednotlivých sektorů mapy podobně jako v předchozím příkladu. Následně testujte shodu Poissonova modelu s daty pomocí χ^2 testu. Jednou zkuste „mířit“ na nějaký konkrétní sektor a podruhé „palte“ zcela náhodně. Uvidíte, zda použitý test bude odpovídat vaší cílené strategii.

Další modely a oblasti použití

Podobné statistické modely jsou vyvinuty pro testování dvou různých souborů dat a pro jejich statisticky signifikantní porovnávání. Tyto obecnější a složitější úlohy se nazývají metody analýzy rozptylu a vedou na testovací statistiku s Fisher–Snedecorovým rozdělením $F(m,n)$, které je jistým zobecněním Pearsonova $\chi^2(k)$ rozdělení. Tyto statistické postupy jsou prakticky použitelné, např. každý výrobce si může sledováním a zaznamenáváním prodejnosti svých různých druhů výrobků otestovat, který z jeho produktů je dlouhodobě úspěšnější.

Existují další modely pro testování a odhady hustot pravděpodobnosti, jako např. regresní aproximace, metrické i nemetrické přístupy stochastického modelování. Jde zejména o aplikace v oblastech monitorování dopravních toků, řízení dopravy a regulace dopravních kolon, předpovědí chování chodců ve stresových situacích, sledování populačních indexů a jejich předpovědi pro jednotlivé kraje a oblasti, lokalizace a detekce mikrotrhlin v materiálech a sledování spolehlivosti průmyslových objektů, jako jsou chemické kulové zásobníky, zařízení vodních i jaderných elektráren, přívodní potrubní systémy, mosty i železobetonové konstrukce budov, křídla letadel apod.

Literatura

[1] Anděl, J.: *Statistické metody.* 2. přeprac. vyd., Matfyzpress, UK Praha, 1998.

[2] Moeschlin, O., et al.: *Experimental Stochastics.* Sec. edit. with CD ROM, Springer, Berlin, 2003.

[3] Rogalewicz, V.: *Pravděpodobnost a statistika pro inženýry.* ČVUT, Praha, 2000.

[4] `www.itl.nist.gov/div898/handbook/eda/section3/eda3674.htm`

[5] `http://www.fourmilab.ch/rpkp/experiments/analysis/chiCalc.html` (on-line kalkulátor).

* * * * *

Kombinační čísla

Zklamala vás kombinační
čísla,
když úloha nesprávně vám
vyšla?
Poslední naději máte
ještě:
Vemte na ni kombinační
kleště!

Emil Calda*)

*) *Úvod do obecné teorie prostoru,* Karolinum, Praha, 2003