Pravděpodobnost, náhoda, kostky

Radek Pelánek

IV122

Výhled

- pravděpodobnost
- náhodná čísla
- lineární regrese
- detekce shluků

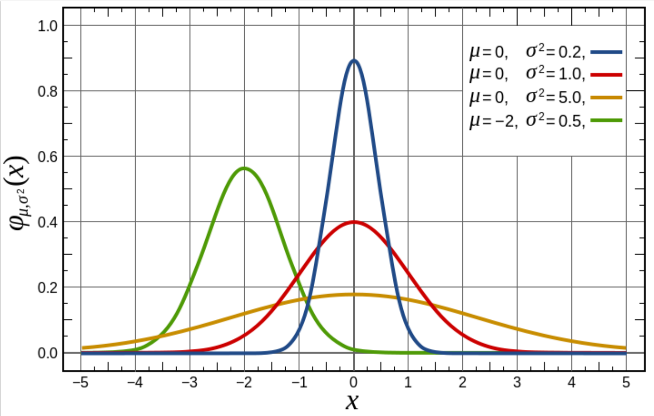
Dnes

- lehce nesourodá směs úloh souvisejících s pravděpodobností
- připomenutí, souvislosti
- krátké programy, realizovatelné i v tabulkovém editoru
- základní myšlenka: využití jednoduchých simulací a analýz pro lepší pochopení abstraktních matematických pojmů
- „kostky“

Pojmy

- pravděpodobnost, podmíněná pravděpodobnost, nezávislost
- střední hodnota, rozptyl, směrodatná odchylka
- distribuční funkce
- normální distribuce

Normální distribuce



Wikipedia

Normální distribuce

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- μ – průměr
- σ – standardní odchylka

Monty Hall Problem

- troje dveře, za jedněmi z nich je poklad, cílem je najít poklad
- vyberete jedny dveře
- já otevřu jedny z nevybraných dveří, za kterými není poklad
- vy nyní můžete zůstat u své volby nebo změnit své rozhodnutí
- co je rozumné udělat?
  - zůstat u své volby
  - změnit rozhodnutí
  - je to úplně jedno (můžeme se rozhodnout náhodně)

Monty Hall Problem: řešení

- je výhodnější změnit rozhodnutí:
  - zůstat u své volby: 33 %
  - změnit rozhodnutí: 66 %
  - rozhodnout se náhodně: 50 %
- problém známý tím, že i mnoho matematiků se v něm snadno splete

Využití simulace

- pro vybudování intuice (lepší pochopení) se hodí simulace
- Monty Hall – velmi jednoduché využití simulace
- užitečný obecný princip

Monty Hall: experimentálně

- implementujte simulátor hry
- vyzkoušejte strategie „zůstat při původním rozhodnutí“, „změnit rozhodnutí“, „náhodně měnit rozhodnutí“
- experimentálně vyhodnoťte úspěšnost strategií v dlouhém běhu

alternativa: jiný, podobný úkol

Náhodná čísla

- aplikace:
  - počítačové hry, loterie
  - kryptografie
  - vědecké výpočty, simulace
- zdroje:
  - „pseudonáhodná čísla“ – běžné `random()`, „deterministické s chaotickým chováním“
  - „opravdová náhoda“ – např. atmosférický tlak, `www.random.org`

Co to jsou náhodná čísla?

„Házení kostkou“ – čísla 1-6

Která z následující posloupností je více pravděpodobná?

- 1 1 2 2 3 3 4 4 5 5 6 6
- 1 5 2 3 4 6 2 3 3 1 2 4

Co to jsou náhodná čísla?

„Házení kostkou“ – čísla 1-6

Která z následující posloupností je více pravděpodobná?

- 1 1 2 2 3 3 4 4 5 5 6 6
- 1 5 2 3 4 6 2 3 3 1 2 4

Obě mají stejnou pravděpodobnost $\left(\frac{1}{6}\right)^{12}$

Úkol: (ne)náhodné posloupnosti

- máte k dispozici několik posloupností čísel „hody kostkou“ $\sim$ celá čísla 1 až 6
- určete, které z nich jsou „nenáhodné“ a proč
- co to znamená, že posloupnost je „náhodná“?

Testování náhodnosti

Testování náhodnosti

- nenáhodná posloupnost:
  - predikovatelná – dokážete předpovědět další číslo (lépe než náhodným tipem)?
  - zdroje nenáhodnosti např. zkreslení, korelace, vzory, periodicita
- existují rozsáhlé sady testů náhodnosti
- vztah statistické testy

Testování náhodnosti: frekvence

Frekvence čísel ve 300 hodech

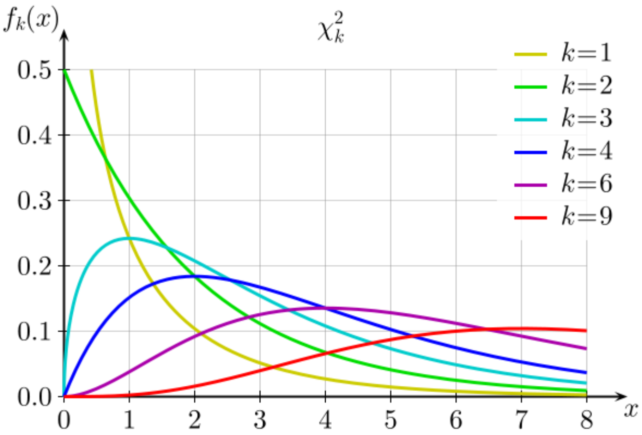
	1	2	3	4	5	6
očekávané	50	50	50	50	50	50
série 1	49	50	48	51	52	50
série 2	56	45	43	62	44	50
série 3	52	71	66	34	30	48

Odpovídá to náhodnému generování?

Testování náhodnosti: Chí kvadrát test

- O_i – očekávaný počet
- P_i – pozorovaný počet
- $S = \sum_{i=1}^{6} \frac{(P_i - O_i)^2}{P_i}$
- S – pro velké n má přibližně χ^2-rozložení o 5 stupních volnosti
- $\chi^2(k) = \sum_{i=1}^{k} Z_i^2$
  kde Z_i má standardní normální rozdělení
- test: určíme p-hodnotu $\chi^2(5)$ pro S, pokud příliš malá – zamítnout

Chí kvadrát



Wikipedia

Centrální limitní věta

Centrální limitní věta

zjednodušeně:

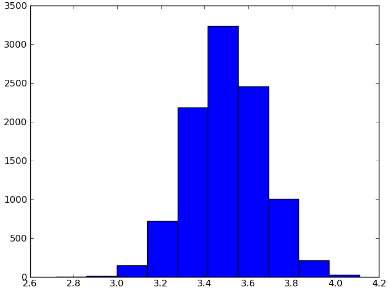
- nezávislé a identicky rozložené proměnné
- vzorky velikosti n
- pro velké n je průměr vzorku přibližně normálně rozložen

Centrální limitní věta: příklad

hody (férovou) kostkou
vzorky velikosti 100
z nich vypočítám průměr
počet vzorků 10000
jak vypadá distribuce průměrů?

Centrální limitní věta: příklad

hody (férovou) kostkou
vzorky velikosti 100
z nich vypočítám průměr
počet vzorků 10000
jak vypadá distribuce průměrů?



Centrální limitní věta: poznámky

- umožňuje modelovat mnoho „neznámých vlivů“ pomocí normální distribuce
- typický příklad – šum v datech (chyba měření):
  - předpokládáme, že šum je výsledkem mnoha dílčích vlivů
  - modelujeme pomocí normální distribuce
- pozor na:
  - předpoklad „nezávislé a identicky rozložené“
  - platí pro aritmetický průměr („aditivní“ veličiny)
  - rychlost konvergence závisí na výchozí distribuci

Centrální limitní věta: příklady kostky

K_a = zatížená kostka, která preferuje vyšší čísla
(pravděpodobnost úměrná počtu teček)
K_b = inverzně zatížená kostka

Jak to dopadne (rozmyslete „teoreticky“, udělejte simulaci):

- hody kostkou K_a
- pro každý hod náhodně vybereme jednu z kostek K_a, K_b
- náhodně vybereme jednu z kostek K_a, K_b a tou házíme všechna čísla ve vzorku

Věnujte pozornost tvaru výsledné distribuce, průměru i směrodatné odchylce.

Bayesova věta

pojmy:

- Bayesova věta
- prior, posterior
- likelihood – věrohodnost
- Bayesovská analýza dat

Bayesova věta

- $P(A|B)$ – podmíněná pravděpodobnost
- Bayesova věta

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayesova věta

- D – pozorovaná data
- H_i – hypotézy o vzniku dat
- $P(H_i)$ – „prior“, odhad pravděpodobnosti H_i předtím, než jsme viděli data
- $P(D|H_i)$ – pravděpodobnost dat při dané hypotéze
- $P(H_i|D)$ – „posterior“, odhad pravděpodobnosti H_i korigovaný daty
- Bayesova věta

$$P(H_i|D) = \frac{P(D|H_i)P(H_i)}{P(D)}$$

- $P(D) = \sum_i P(D|H_i)P(H_i)$ – pravděpodobnost dat

Bayesova věta – klasický příklad

- předpokládejme
  - výskyt AIDS: 6 z 1000
  - spolehlivý test na AIDS:
    - správný výsledek 99,9 % pro ty, co mají AIDS
    - 99 % pro ty, co nemají AIDS
- výsledek testu osoby X je pozitivní
- jaká je pravděpodobnost, že X má AIDS?

Bayesova věta – klasický příklad

hypotézy: A = AIDS, N = nemá AIDS
data: V = pozitivní výsledek

$$
\begin{aligned}
P(A|V) &= \frac{P(V|A)P(A)}{P(V|A)P(A)+P(V|N)P(N)} \\
&= \frac{0.006 \cdot 0.999}{0.006 \cdot 0.999 + 0.994 \cdot 0.01} \sim 0.38
\end{aligned}
$$

Bayesova věta – příklad kostky

- 100 kostek, 1 falešná (samé 6), ostatní poctivé
- náhodně vytáhnu jednu kostku, 3 krát hodím, vždy padne šestka
- jaká je pravděpodobnost, že jde o poctivou kostku?

Vypočítejte:

- vzorcem (Bayes)
- simulací