Pravděpodobnost a statistika

Tomáš Masopust

Aktualizováno: 24. března 2022

Obsah

Obsah

- Pascal a Fermat vytvořili pojednání o hazardních hrách dvou hráčů. Výsledek jejich diskuse vedl k základům teorie pravděpodobnosti.

  Je pozoruhodné, že věda, jenž se započala uvažováním o hazardních hrách, by se měla stát nejvýznamějším objektem lidského vědění.

  (Pierre-Simon de Laplace)

- Pravděpodobnost je plná překvapivých výsledků a paradoxů, více něž jakákoliv jiná matematická disciplína.

  Asi největším paradoxem ze všech je to, že existují paradoxy v matematice.

  (Edward Kasner)

Americký matematik, zavedl pojem "googol"$=10^{100}$ (navrhl jeho devítiletý synovec Milton Sirotta)

- Dalo by se očekávat, že na základě našeho smyslu riskovat a životních zkušeností bychom měli mít dobře vyvinutý instinkt.
- Vyzkoušejme si.

- Vsadili byste si na to, že ve třídě jsou alespoň dva lidé narození ve stejný den?
- Pokud je ve třídě alespoň 23 lidí, tak je taková šance alespoň 51 %. (Zkusme prakticky ověřit...)

- Pokud je ve čtyřčlenné rodině jedno dítě chlapec, jaká je pravděpodobnost, že je i druhé dítě chlapec?
- Možnosti jsou: (Ch,Ch), (Ch,D), (D,Ch), (D,D), tj. $\frac{1}{3}$.

- Je pravděpodobnost toho, že při hodu dvěmi hracími kostkami padne součet 12 stejná jako pravděpodobnost toho, že padne součet 11?
- Spočítejme si...
- Leibnitz veřil, že tyto pravděpodobnosti jsou stejné.

- Je pravděpodobnost toho, že třikrát po sobě padne "orel" stejná jako pravděpodobnost toho, že na třech mincích hozených současně padne na všech "orel"?
- D'Alembert, francouzský matematik 18. století, o tom nebyl přesvědčen.

- Je pravda, že když dlouho padá "orel", tak je pak pravděpodobnější, že v dalším hodu padne "orel"?
- D'Alembert tomu věřil.

- Uvažujme následující hru.
- Máme tři krabice.
- V jedné jsou klíče od nového vozu, ve druhých není nic.
- Já vím, kde co je. Vy nevíte.
- Já vás nechám vybrat jednu krabici.
- Ze zbývajících dvou pak otevřu prázdnou krabici.
- Nyní jsou dvě zavřené krabice – ta vámi vybraná a ta druhá.
- Když vám nyní dám na vybranou nechat si tu vámi vybranou krabici, nebo si vzít tu druhou, co uděláte?
- Paul Erdös, jeden z nejvýznamějších matematiků posledních let nevěřil správnému řešení, dokud mu to jeho přítel Ron Graham trpělivě nevysvětlil.

Literatura

1. První tištěnou práci o pravděpodobnosti sepsal holandský matematik Christiaan Huygens (1629-1695).
2. M. Gardner, The Colossal Book of Mathematics
3. M. S. Petković, Famous Puzzles of Great Mathematicians
4. J. Likeš, J. Machek, Počet pravděpodobnosti

Studijní literatura:

1. L. Wasserman, *All of Statistics–A Concise Course in Statistical Inference*
2. S. Ross, *A First Course in Probability*
3. M. Mitzenmacher a E. Upfal, *Probability and Computing*

Pravděpodobnost

- Pravděpodobnost je matematický jazyk pro určení nejistoty
- Zavedeme základní koncepty teorie pravděpodobnosti

Výběrové prostory a jevy

Prostor elementárních jevů

- Výběrový prostor či prostor elementárních jevů Ω je množina možných výsledků náhodného pokusu
- Prvky $\omega \in \Omega$ se nazývají elementární jevy
- Podmnožiny Ω se nazývají (náhodné) jevy

Příklad 1.

- Náhodný pokus = hod dvakrát mincí
- Pak $\Omega = \{OO, OP, PO, PP\}$[1]
- Jev, že "na první hod padne orel" je $A = \{OO, OP\}$

[1] "O"=orel, "P"=panna

Příklad 2.

- Nechť ω je výsledek měření nějaké fyzikální veličiny, např. teploty
- Pak $\Omega = \mathbb{R} = (-\infty, \infty)$
- $\Omega = \mathbb{R}$ není přesné, protože teplota má dolní hranici
- Obvykle není na škodu vzít výběrový prostor větší než potřeba
- Jev "teplota je větší než 10 a menší nebo rovna 23" je $A = (10, 23]$

Příklad 3.

Jestliže házíme mincí do nekonečna, pak výběrový prostor je nekonečná množina

$$\Omega = \{\omega = (\omega_1, \omega_2, \dots) \mid \omega_i \in \{O, P\}\}$$

Nechť E je jev "první orel se objeví na třetí hod". Pak

$$E = \{(\omega_1, \omega_2, \dots) \mid \omega_1 = P, \omega_2 = P, \omega_3 = O, \omega_i \in \{O, P\} \text{ pro } i > 3\}$$

Operace na jevech

- Pro jev A je $A^c = \{\omega \in \Omega \mid \omega \notin A\}$ jev.
  Komplement Ω je prázdná množina $\emptyset$.
- Sjednocení jevů A a B je jev
  $$A \cup B = \{\omega \in \Omega \mid \omega \in A \text{ nebo } \omega \in B\}$$
  Pro jevy $A_1, A_2, \ldots$ je $\bigcup_{i=1}^{\infty} A_i = \{\omega \in \Omega \mid \exists i \text{ t.ž. } \omega \in A_i\}$ jev.
- Průnik jevů A a B je jev
  $$A \cap B = \{\omega \in \Omega \mid \omega \in A \text{ a } \omega \in B\}$$
  $A \cap B$ budeme též zapisovat jako AB nebo (A, B).
  Pro jevy $A_1, A_2, \ldots$ je $\bigcap_{i=1}^{\infty} A_i = \{\omega \in \Omega \mid \omega \in A_i \text{ pro všechna } i\}$ jev.

- Rozdíl jevů je jev $A - B = \{\omega \mid \omega \in A, \omega \notin B\}$
- Jestliže kadý prvek A je také v B, píšeme $A \subset B$ či $B \supset A$.
- Jestliže A je konečná množina, značí $|A|$ počet prvků A

Shrnutí

Ω	výběrový prostor
ω	elementární jev (bod nebo prvek)
A	jev (podmnožina Ω)
A^c	komplement A
$A \cup B$	sjednocení
$A \cap B$ nebo AB	průnik
$A - B$	množinový rozdíl
$A \subset B$	množinová inkluze
$\emptyset$	nemožný jev (vždy false)
Ω	jistý jev (vždy true)

Disjunktní jevy, rozklad

- Jevy $A_1, A_2, \ldots$ jsou disjunktní nebo vzájemně neslučitelné, jestliže
$$A_i \cap A_j = \emptyset$$
kdykoliv $i \neq j$.
- Např. $A_1 = [0, 1), A_2 = [1, 2), A_3 = [2, 3), \ldots$ jsou disjunktní
- Rozklad Ω je posloupnost disjunktních množin $A_1, A_2, \ldots$ t.ž.
$$\bigcup_i A_i = \Omega$$

Indikátory

- Pro jev A definujeme indikátor či charakteristickou funkci

$$I_A(\omega) = \begin{cases} 1 & \text{pro } \omega \in A \\ 0 & \text{pro } \omega \notin A \end{cases}$$

Rostoucí a klesající posloupnosti množin

- Posloupnost množin $A_1, A_2, \ldots$ je monotóně rostoucí, jestliže

$$A_1 \subset A_2 \subset \cdots$$

  Definujme

$$\lim_{n\to\infty} A_n = \bigcup_{i=1}^{\infty} A_i$$

- Posloupnost množin $A_1, A_2, \ldots$ je monotóně klesající, jestliže

$$A_1 \supset A_2 \supset \cdots$$

  Definujme

$$\lim_{n\to\infty} A_n = \bigcap_{i=1}^{\infty} A_i$$

- V obou případech budeme psát $A_n \to A$, kde $A = \lim_{n\to\infty} A_n$

Příklad 4.

- Nechť $\Omega = \mathbb{R}$ a nechť $A_i = [0, \frac{1}{i})$ pro $i = 1, 2, \ldots$. Pak
  $\cup_{i=1}^{\infty} A_i = [0, 1)$ a
  $\cap_{i=1}^{\infty} A_i = \{0\}$

- Jestliže $A_i = (0, \frac{1}{i})$, pak
  $\cup_{i=1}^{\infty} A_i = (0, 1)$
  $\cap_{i=1}^{\infty} A_i = \emptyset$

Pravděpodobnost

Definice 5.

Funkce $\mathbb{P}$ přiřazující reálné číslo $\mathbb{P}(A)$ každému jevu A je pravděpodobnostní míra, jestliže splňuje následující tři axiomy:

Axiom 1: $\mathbb{P}(A) \geq 0$ pro každý jev A

Axiom 2: $\mathbb{P}(\Omega) = 1$

Axiom 3: Jestliže $A_1, A_2, \ldots$ jsou disjunktní jevy, pak

$$\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

Algebra jevů

- Obecně není možné přiřadit pravděpodobnost všem podmnožinám Ω (pokud je Ω nespočetná)
- Omezíme se proto na množinu jevů nazývanou σ-algebra
- σ-algebra je třída $\mathcal{A}$ splňující
  1. $\emptyset \in \mathcal{A}$
  2. Jestliže $A_1, A_2, \ldots \in \mathcal{A}$, pak $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$
  3. Jestliže $A \in \mathcal{A}$, pak též $A^c \in \mathcal{A}$
- Množiny v $\mathcal{A}$ se nazývají měřitelné
- $(\Omega, \mathcal{A})$ je měřitelný prostor
- Jestliže $\mathbb{P}$ je pravděpodobnostní míra na $\mathcal{A}$, je $(\Omega, \mathcal{A}, \mathbb{P})$ pravděpodobnostní prostor
- Pokud je Ω reálná osa, vezmeme $\mathcal{A}$ jako nejmenší σ-algebru, která obsahuje všechny otevřené podmnožiny, tzv. Borelovská σ-algebra[2]

[2] Omezíme se na případ, kdy otevřené množiny "znamená" intervaly.

- Existuje mnoho interpretací $\mathbb{P}(A)$. Dvě základní jsou frekvence a stupeň důvěry.
- Frekvence: $\mathbb{P}(A)$ vyjadřuje poměr, kolikrát je A slpňeno při dlouhodobém opakování pokusu.
  - Např. "pravděpodobnost padnutí orla je $1/2$" znamená, že s rostoucím počtem hodů mincí jde poměr hozených orlů vzhledem ke všem pokusům k $1/2$.
- Stupeň důvěry: $\mathbb{P}(A)$ určuje pozorovatelovu intenzitu důvěry, že A je splňeno.
- V obou interpretacích vyžadujeme platnost Axiomů 1 až 3.
- Rozdíl mezi interpretacemi hraje roli až ve statistické inferenci:
  - frekvencionistická vs. Bayesovská škola

Vlastnosti pravděpodobnosti

Z axiomů lze odvodit mnoho vlastností $\mathbb{P}$:

- $\mathbb{P}(\emptyset) = 0$
  - $1 =_{A2} \mathbb{P}(\Omega) = \mathbb{P}(\Omega \cup \emptyset) =_{A3} \mathbb{P}(\Omega) + \mathbb{P}(\emptyset) = 1 + \mathbb{P}(\emptyset)$
- $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(\mathbb{B})$
  - $B = A \cup (B - A) \Rightarrow \mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B - A) \Rightarrow_{A1} \mathbb{P}(A) \leq \mathbb{P}(B)$
- $0 \leq \mathbb{P}(A) \leq 1$
  - použij předchozí pro $A \subseteq \Omega$
- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
  - $\Omega = A \cup A^c \Rightarrow 1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c)$
- $A \cap B = \emptyset \Rightarrow \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$
  - z Axiomu 3 dosazením $A_1 = A$, $A_2 = B$, $A_3 = A_4 = \cdots = \emptyset$

Vlastnosti $\mathbb{P}$

Lemma 6.

Pro libovolné jevy A a B je $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB)$.

Důkaz.

$A \cup B = (AB^c) \cup (AB) \cup (A^cB)$ a uvedené jevy jsou disjunktní. Opakovaným použitím Axiomu 3 dostáváme

$$
\begin{aligned}
\mathbb{P}(A \cup B) &= \mathbb{P}((AB^c) \cup (AB) \cup (A^cB)) \\
&= \mathbb{P}(AB^c) + \mathbb{P}(AB) + \mathbb{P}(A^cB) \\
&= \mathbb{P}(AB^c) + \mathbb{P}(AB) + \mathbb{P}(A^cB) + \mathbb{P}(AB) - \mathbb{P}(AB) \\
&= \mathbb{P}((AB^c) \cup (AB)) + \mathbb{P}((A^cB) \cup (AB)) - \mathbb{P}(AB) \\
&= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB)
\end{aligned}
$$

$\square$

Příklad 7.

- Dva hody mincí
- H_1 je jev “orel padne v prvním hodu”
- H_2 =“orel padne v druhém hodu”
- Pokud jsou všechny výsledky stejně pravděpodobné, pak

$$\begin{aligned}\mathbb{P}(H_1 \cup H_2) &= \mathbb{P}(H_1) + \mathbb{P}(H_2) - \mathbb{P}(H_1 H_2) \\ &= \frac{1}{2} + \frac{1}{2} - \frac{1}{4} \\ &= \frac{3}{4}\end{aligned}$$

Věta 8.

Jestliže $A_n \to A$, *pak* $\mathbb{P}(A_n) \to \mathbb{P}(A)$ *pro* $n \to \infty$.

Důkaz.

- Předpokládejme, že A_n je rostoucí, tj. $A_1 \subset A_2 \subset \cdots$
- Nechť $A = \lim_{n\to\infty} A_n = \cup_{i=1}^{\infty} A_i$
- Definujme $B_1 = A_1$, $B_2 = \{\omega \in \Omega \mid \omega \in A_2, \omega \notin A_1\}$, $B_3 = \{\omega \in \Omega \mid \omega \in A_3, \omega \notin A_2, \omega \notin A_1\}, \ldots$
- $B_1, B_2, \ldots$ jsou disjunktní, $A_n = \cup_{i=1}^{n} A_i = \cup_{i=1}^{n} B_i$ pro každé n a $\cup_{i=1}^{\infty} B_i = \cup_{i=1}^{\infty} A_i$
- Z Axiomu 3 máme $\mathbb{P}(A_n) = \mathbb{P}(\cup_{i=1}^{n} B_i) = \sum_{i=1}^{n} \mathbb{P}(B_i)$, tj.

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(B_i) = \sum_{i=1}^{\infty} \mathbb{P}(B_i) = \mathbb{P}(\cup_{i=1}^{\infty} B_i) = \mathbb{P}(A)$$

□

Konečný výběrový prostor

Konečné výběrové prostory

- $\Omega = \{\omega_1, \ldots, \omega_n\}$ konečný
- Pro "hod dvakrát kostkou" je $|\Omega| = 36$:
  - $\Omega = \{(i, j) \mid i, j \in \{1, \ldots, 6\}\}$
- Pokud je každý výsledek stejně pravděpodobný, je

$$\mathbb{P}(A) = \frac{|A|}{36}$$

  kde $|A|$ je počet prvků jevu A
- Pravděpodobnost, že padne součet 11 je $\frac{2}{36}$, protože existují dva výsledky se sumou 11
  - Které?

Pravděpodobnost na konečném výběrovém prostoru

- Pokud je Ω konečný a každý výsledek je stejně pravděpodobný, pak

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}$$

je rovnoměrná (uniformní) pravděpodobnostní míra

- K určení pravděpodobnosti tedy potřebujem znát počet výsledků příznivých jevu A
  - k tomu slouží kombinatorické metody

Základní princip počítání

Věta 9 (Základní princip počítání).

Mějme dva experimenty. Pokud má první experiment m možných výsledků a pro každý výsledek existuje n možných výsledků druhého experimentu, pak počet možných výsledků obou experimentů společně je mn.

Důkaz.

Vyčísleme všechny možné výsledky obou experimentů:

$$\begin{gathered}(1,1),(1,2),\ldots,(1,n)\\(2,1),(2,2),\ldots,(2,n)\\\vdots\\(m,1),(m,2),\ldots,(m,n)\end{gathered}$$

kde (i,j) značí i-tý možný výsledek prvního experimentu a j-tý možný výsledek druhého. Množina všech možných výsledků tedy sestává z m řádků, kde každý má n prvů. $\square$

Příklad 10.

- Mějme skupinu 10 žen, kde každá žena má 3 děti.
- Pokud bychom měli zvolit jednu ženu a jedno její dítě za matku a dítě roku, kolik možností máme?
- **Řešení:**
- První experiment: volba ženy
- Druhý experiment: volba jednoho z jejich dětí
- Základní princip počítání dává $10 \cdot 3 = 30$ možností.

Zobecnění základního principu počítání

Věta 11 (Zobecněný základní princip počítání).

Jestliže máme provést r experimentů, kde první má n_1 možných výsledků a pro každý z možných výsledků má druhý n_2 možných výsledků a pro každý z možných výsledků prvních dvou experimentů má třetí experiment n_3 možných výsledků atd., pak celkový počet možných výsledků těchto r experimentů je

$$n_1 \cdot n_2 \cdot n_3 \cdot \dots \cdot n_r \,.$$

□

Příklad 12.

- Středoškolská komise se skládá ze
  - 3 prváků
  - 4 druháků
  - 5 třeťáků a
  - 2 čtvrťáků.
- Volíme 4 zastupitele tak, aby z každého ročníku byl přítomen jeden člen komise.
- Kolik různých zastupitelstev můžeme sestavit?

- **Řešení:** $3 \cdot 4 \cdot 5 \cdot 2 = 120$

◂ Narozeninový příklad

Příklad 13.

- Kolik různých 7-místných SPZ lze sestavit, jestliže první 3 místa jsou písmena a další 4 čísla?
  - $26 \cdot 26 \cdot 26 \cdot 10 \cdot 10 \cdot 10 \cdot 10 = 175\,760\,000$

Příklad 14.

- Kolik existuje funkcí na n-prvkové množině, když hodnoty funkcí jsou 0 nebo 1?
  - $\{1, 2, \dots, n\}$. Jelikož $f(i) \in \{0, 1\}$ pro všechna $i = 1, 2, \dots, n$, existuje 2^n funkcí.

Příklad 15.

- Kolik SPZ by bylo možno sestavit, pokud se písmena ani čísla nesmí opakovat?
  - $26 \cdot 25 \cdot 24 \cdot 10 \cdot 9 \cdot 8 \cdot 7 = 78\,624\,000$

Permutace

- Kolik je různých uspořádání písmen a, b, a c?
  - Přímým výpočtem dostaneme 6: abc, acb, bac, bca, cab a cba.
- Každé z těchto 6 uspořádání je permutace.
- Základní princip počítání dává, že první prvek permutace může být libovolný ze 3, druhý libovolný ze 2 a třetí ten jeden zbývající
- Tedy máme $3 \cdot 2 \cdot 1 = 6$ možných permutací

Definice 16 (Permutace).

Mějme n prvků, pak existuje

$$n \cdot (n-1) \cdot (n-2) \cdots 3 \cdot 2 \cdot 1 = n!$$

různých permutací n prvků.

- Zatímco $n!$ („n faktoriál“) je definován jako $1 \cdot 2 \cdots n$ pro celé $n \geq 1$, je vhodné dodefinovat $0! = 1$.

Příklad 17.

- Kolik možných uspořádání existuje v baseballovém týmu 9 hráčů?
  - $9! = 362\,880$ uspořádání

Příklad 18.

- Zkoušku skládá 6 mužů a 4 ženy. Žádní dva nemají stejný počet bodů.

(a) Kolik různých pořadí podle počtu bodů existuje?

(b) Kolik různých pořadí existuje, pokud jsou muži a ženy řazeny zvlášť?

(a) 10 lidí lze uspořádat $10! = 3\,628\,800$ způsoby

(b) Jelikož 6 mužů lze uspořádat 6! způsoby a 4 ženy 4! způsoby, základní princip počítání dává celkem $(6!)(4!) = (720)(24) = 17\,280$ možných pořadí.

Příklad 19.

- Paní Nováková má 10 knih, které chce dát do knihovny. Z toho jsou
  - 4 o matematice
  - 3 o chemii
  - 2 o historii a
  - 1 jazyková učebnice.
- Paní Nováková chce knihy uspořádat tak, aby stejné předměty byly pohromadě.
- Kolika způsoby může knihy uspořádat?

- Paní Nováková má $4!3!2!1!$ možností, přičemž knihy o matematice jsou první a následují knihy o chemii, o historii a nakonec jazyková učebnice.
- Čtyři témata knih lze uspořádat $4!$ způsoby
- Proto má paní Novákové celkem $4!(4!3!2!1!) = 6\,912$ možností.

Příklad 20.

- Jaká je pravděpodobnost, že mezi 25 lidmi jsou alespoň dva, kteří mají narozeniny ve stejný den v roce?
  - Požadovaná pravděpodobnost je rovna $1 - \mathbb{P}(\text{“všichni v různý den”})$
  - výběrový prosto jsou funkce z $\{1, \ldots, 25\}$ do $\{1, \ldots, 365\}$ (ignorujeme přestupné roky)
  - $|\Omega| = 365^{25}$
  - “všichni v různý den” odpovídá injektivním zobrazením, těch je $365 \cdot 364 \cdots 341$
  - Tedy
  $$\mathbb{P}(\text{“všichni v různý den”}) = \frac{365 \cdot 364 \cdots 341}{365^{25}} \approx 0.4$$
  a tedy hledaná pravděpodobnost je přibližně 0.6.

Kombinatorika přehled

Příklad 21.

- Kolik různých přesmyček slov PEPPER lze sestavit?

- Existuje $6!$ permutací písmen $P_1E_1P_2P_3E_2R$
  - pokud jsou ta tři písmenka P a dvě E od sebe navzájem rozlišitelná.
  - Uvažme libovolnou z těchto permutací, např. $P_1P_2E_1P_3E_2R$.
- Pokud prohazujeme P-čka mezi sebou a E-čka mezi sebou, výsledek stále je $PPEPER$.
- Tedy $3!2!$ permutací jsou tvaru $PPEPER$:

$$\begin{array}{lll}
P_1P_2E_1P_3E_2R & P_3P_1E_1P_2E_2R & P_2P_1E_2P_3E_1R \\
P_1P_3E_1P_2E_2R & P_3P_2E_1P_1E_2R & P_2P_3E_2P_1E_1R \\
P_2P_1E_1P_3E_2R & P_1P_2E_2P_3E_1R & P_3P_1E_2P_2E_1R \\
P_2P_3E_1P_1E_2R & P_1P_3E_2P_2E_1R & P_3P_2E_2P_1E_1R
\end{array}$$

- Celkem tedy máme $\dfrac{6!}{3!2!} = 60$ přesmyček písmen slova $PEPPER$.

Permutace s opakováním

Definice 22 (Permutace s opakováním).

Počet permutací n prvků, kde první prvek se vyskytuje k_1-krát, druhý k_2-krát, až n-tý k_n-krát je

$$\frac{n!}{k_1!k_2!\cdots k_n!}$$

Příklad 23.

- Šachový turnaj má 10 účastníků: 4 z Ruska, 3 z USA, 2 z UK a 1 z Brazílie.
- Pokud výsledková listina udává pouze národnost hráče na daném pořadí, kolik je možných výsledků?
  - $\frac{10!}{4!3!2!1!} = 12\,600$

Příklad 24.

- Kolik různých signálů lze vytvořit ze 4 bílých, 3 červených a 2 modrých vlajek, jestliže se každý symbol skládá z 9 vlajek zavěšených vedle sebe?
  - Vlajky stejné barvy jsou identické.
  - $\frac{9!}{4!3!2!} = 1\,260$

Kombinace

- Často nás zajímá počet různých skupin (podmnožin) r prvků z n.
  - Např. kolik různých skupin 3 prvků lze vybrat z prvků A, B, C, D, E?
  - 5 způsobů jak vybrat první, 4 jak vybrat druhý a 3 jak vybrat poslední, proto $5 \cdot 4 \cdot 3$ způsobů, ale...
  - každá skupina 3 prvků (např. A, B, C) bude počítána 6-krát (počítáme všechny permutace ABC, ACB, BAC, BCA, CAB a CBA).
  - Celkový počet skupin tedy je
$$\frac{5 \cdot 4 \cdot 3}{3 \cdot 2 \cdot 1} = 10$$
- Obecně máme $n(n-1) \cdots (n-r+1)$ různých způsobů, jak vybrat r-prvků z n prvků, kde záleží na pořadí prvků, a každá r-tice je počítána $r!$-krát, tedy máme
$$\frac{n(n-1) \cdots (n-r+1)}{r!} = \frac{n!}{r!(n-r)!}$$
  r-prvkových podmnožin z n prvků.

Definice 25.

Definujeme číslo $\binom{n}{r}$ pro $r \leq n$ jako

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

a nazýváme je kombinační číslo; čteme "n nad r"

- Z definice $0! = 1$ dostaneme, že $\binom{n}{n} = \binom{n}{0} = \frac{n!}{0!n!} = 1$
- Pro $r > n$ nebo $r < 0$ definujeme $\binom{n}{r} = 0$

Příklad 26.

- Z 20 lidí má být vytvořena tříčlenná komise. Kolik různých komisí lze vytvořit?
  - $\binom{20}{3} = 1\,140$

Příklad 27.

- Z 5 žen a 7 mužů máme vytvořit komisi 2 žen a 3 mužů. Kolik máme možností?
  - $\binom{5}{2}$ možností jak vybrat 2 ženy z 5 a
  - $\binom{7}{3}$ možností jak vybrat 3 muže ze 7
  - dává $\binom{5}{2}\binom{7}{3} = 350$ možností.
- Co když se 2 muži nemají rádi a odmítají být spolu v komisi?
  - Počet skupin tří mužů, kde jsou oba, kteří se nemají rádi, je $\binom{2}{2}\binom{5}{1} = 5$,
  - proto máme celkem $(\binom{7}{3} - 5)\binom{5}{2} = 30 \cdot 10 = 300$ možností.

Příklad 28.

- Mějme n antén, z nichž je m vadných a $n-m$ funkčních. Antény jsou nerozlišitelné. Kolika způsoby je můžeme uspořádat tak, aby žádné dvě vadné nebyly vedle sebe?
  - Představme si, že $n-m$ funkčních antén jsou seřazeny vedle sebe.
  - Pokud žádné dvě vadné nemají být vedle sebe, obsahuje každé místo mezi dvěmi funkčními anténami nejvýše jednu vadnou (včetně krajních míst).
  - Tedy máme $n-m+1$ pozic mezi $n-m$ funkčními anténami, kam umístit m vadných antén.
  - To dává $\binom{n-m+1}{m}$ možných uspořádání.

- Užitečná identita je

$$\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r} \qquad 1 \leq r \leq n \tag{1}$$

Důkaz.

- Mějme n prvků a fixujme jeden z nich, řekněme x.
- Dostaneme $\binom{n-1}{r-1}$ skupin r prvků obsahující x
- ... a $\binom{n-1}{r}$ skupin r prvků neobsahující x.

□

Věta 29 (Binomická věta).

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} \tag{2}$$

Důkaz.

- Pro $n = 1$ máme $(x+y) = \binom{1}{0}x^0y^1 + \binom{1}{1}x^1y^0 = y + x$ ✓
- Nechť (2) platí pro $n-1$, pak $(x+y)^n$

$$\begin{aligned}
&= (x+y)(x+y)^{n-1} \\
&= (x+y)\textstyle\sum_{k=0}^{n-1} \binom{n-1}{k} x^k y^{n-1-k} \\
&= \textstyle\sum_{k=0}^{n-1} \binom{n-1}{k} x^{k+1} y^{n-1-k} + \sum_{k=0}^{n-1} \binom{n-1}{k} x^k y^{n-k} \\
&= \textstyle\sum_{i=1}^{n} \binom{n-1}{i-1} x^i y^{n-i} + \sum_{i=0}^{n-1} \binom{n-1}{i} x^i y^{n-i} \\
&= x^n + \textstyle\sum_{i=1}^{n-1} \left[\binom{n-1}{i-1} + \binom{n-1}{i}\right] x^i y^{n-i} + y^n \\
&= x^n + \textstyle\sum_{i=1}^{n-1} \binom{n}{i} x^i y^{n-i} + y^n \\
&= \textstyle\sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i} \quad \checkmark
\end{aligned}$$

Příklad 30.

- Rozviňte $(x+y)^3$
  - $(x+y)^3 = \binom{3}{0}x^0y^3 + \binom{3}{1}xy^2 + \binom{3}{2}x^2y + \binom{3}{3}x^3y^0 = y^3 + 3xy^2 + 3x^2y + x^3$.

Příklad 31.

- Kolik podmnožin má n prvková množina?
  - k-prvkových podmnožin je $\binom{n}{k}$, proto všech je $\sum_{k=0}^{n}\binom{n}{k} = (1+1)^n = 2^n$.

Multinomiální koeficienty

- Množina n různých prvků má být rozdělena do r různých košů tak, že první koš bude obsahovat n_1 prvků, druhý koš n_2 prvků, až r-tý koš bude obsahovat n_r prvků.
- Platí $\sum_{i=1}^{r} n_i = n$.
- Kolik máme možností?
- Máme $\binom{n}{n_1}$ možností pro první koš
- Pro každou volbu pro první koš máme $\binom{n-n_1}{n_2}$ možností pro druhý koš, atd.
- Celkem $\binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\cdots\binom{n-n_1-n_2-\cdots-n_{r-1}}{n_r}$

$$= \frac{n!}{(n-n_1)!n_1!}\frac{(n-n_1)!}{(n-n_1-n_2)!n_2!}\cdots\frac{(n-n_1-n_2-\cdots-n_{r-1})!}{0!n_r!} = \frac{n!}{n_1!n_2!\cdots n_r!}$$

Definice 32 (Permutace s opakováním).

Jestliže $n_1 + n_2 + \cdots + n_r = n$, definujeme

$$\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1! n_2! \cdots n_r!}$$

Příklad 33.

- Policejní stanice má 10 strážníků.
  - 5 strážníků musí hlídat v ulicích
  - 2 musí pracovat na stanici
  - 3 musí být v záloze na stanici
- Kolik různých rozdělení 10 strážníků do těchto tří skupin existuje?
  - $\dfrac{10!}{5!3!2!} = 2\,520$

Příklad 34.

- 10 dětí se má rozdělit do dvou týmů A a B o 5 členech
- Tým A bude hrát jednu ligu, tým B druhou
- Kolik týmů lze vytvořit?
  - $\dfrac{10!}{5!5!} = 252$

Příklad 35.

- 10 dětí se má rozdělit do dvou týmů po 5 členech. Kolik týmů lze vytvořit?
- V čem je rozdíl oproti předchozímu příkladu?
  - Na pořadí týmů nezáleží, jde o rozdělení do dvou skupin po 5.
  - $\frac{10!/(5!5!)}{2!} = 126$

Věta 36 (Multinomiální věta).

$$(x_1 + x_2 + \cdots + x_r)^n = \sum_{\substack{(n_1, n_2, \ldots, n_r) \\ n_1 + n_2 + \cdots + n_r = n}} \binom{n}{n_1, n_2, \ldots, n_r} x_1^{n_1} x_2^{n_2} \cdots x_r^{n_r}$$

Všechna $n_i \geq 0$.

Příklad 37.

$$
\begin{aligned}
(x_1+x_2+x_3)^2 &= \binom{2}{2,0,0} x_i^2 x_2^0 x_3^0 + \binom{2}{0,2,0} x_i^0 x_2^2 x_3^0 \\
&+ \binom{2}{0,0,2} x_i^0 x_2^0 x_3^2 + \binom{2}{1,1,0} x_i^1 x_2^1 x_3^0 \\
&+ \binom{2}{1,0,1} x_i^1 x_2^0 x_3^1 + \binom{2}{0,1,1} x_i^0 x_2^1 x_3^1 \\
&= x_1^2 + x_2^2 + x_3^2 + 2x_1x_2 + 2x_1x_3 + 2x_2x_3
\end{aligned}
$$

Shrnutí

Uspořádaný výběr		
Bez opakování	Variace bez opakování	$\dfrac{n!}{(n-k)!}$
	Permutace bez opakování	$n!$
S opakováním	Variace s opakováním	n^k
	Permutace s opakováním	$\dfrac{n!}{n_1! \cdot n_2! \cdots n_k!}$

Neuspořádaný výběr		
Bez opakování	Kombinace bez opakování	$\dbinom{n}{k}$
S opakováním	Kombinace s opakováním	$\dbinom{n+k-1}{k}$

Nezávislé jevy

Nezávislé jevy

- Pokud dvakrát hodíme férovou mincí, pravděpodobnost dvou orlů bude $\frac{1}{2} \times \frac{1}{2}$
- Pravděpodobnosti násobíme, protože bereme hody za nezávislé

Definice 38 (Nezávislé jevy).

Jevy A a B jsou nezávislé, jestliže

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)\,.$$

Množina jevů $\{A_i \mid i \in I\}$ je nezávislá, pokud

$$\mathbb{P}(\bigcap_{i \in J} A_i) = \prod_{i \in J} \mathbb{P}(A_i)$$

pro každou konečnou podmnožinu $J \subseteq I$.

Nezávislost vzniká ve dvou případech:

1. Předpokládáme, že jevy jsou nezávislé
   - např. při hodu mince dvakrát po sobe často předpokládáme, že hody jsou nezávislé, což odráží fakt, že mince nemá paměť na první hod

2. Nezávislost odvodíme (ověřením $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$)
   - např. při hodu férovou kostkou pro $A = \{2, 4, 6\}$ a $B = \{1, 2, 3, 4\}$ je $A \cap B = \{2, 4\}$ a $\mathbb{P}(AB) = 2/6 = \mathbb{P}(A)\mathbb{P}(B) = (1/2) \times (2/3)$, tj. A a B jsou nezávislé

- Předpokládejme, že A a B jsou disjunktní jevy s nenulovou pravděpodobností.
- Mohou být nezávislé?
- **Ne**: $\mathbb{P}(A)\mathbb{P}(B) > 0$, ale $\mathbb{P}(AB) = \mathbb{P}(\emptyset) = 0$
- Až na tento speciální případ neexistuje způsob, jak zjistit nezávislost pouze z Vennova diagramu

Příklad 39.

- Házíme férovou mincí 10 krát
- $A =$"padnul alespoň jednou orel"
- T_j je jev, že orel nepadnul v j-tém hodu

$$\begin{aligned}
\mathbb{P}(A) &= 1 - \mathbb{P}(A^c) \\
&= 1 - \mathbb{P}(\text{"žádný orel"}) \\
&= 1 - \mathbb{P}(T_1 T_2 \cdots T_{10}) \\
&= 1 - \mathbb{P}(T_1)\mathbb{P}(T_2) \cdots \mathbb{P}(T_{10}) \qquad \text{(z nezávislosti)} \\
&= 1 - (1/2)^{10} \approx 0.999.
\end{aligned}$$

Příklad 40.

- Dva hráči hází míč do koše:
  - první s úspěšností $1/3$
  - druhý s úspěšností $1/4$.
- Jaká je pravděpodobnost, že se první trefí dříve než druhý? (Označme jako jev E)
- Nechť A_j je jev, že se první trefí jako první, a to v j-tém hodu.
  - Pak $A_1, A_2, \dots$ jsou disjunktní a $E = \cup_{j=1}^{\infty} A_j$, tj. $\mathbb{P}(E) = \sum_{j=1}^{\infty} \mathbb{P}(A_j)$
  - $\mathbb{P}(A_1) = 1/3$
  - A_2 nastane, pro “první míjí, druhý míjí, první trefuje”
    - $\mathbb{P}(A_2) = (2/3)(3/4)(1/3) = (1/2)(1/3)$
  - Obecně $\mathbb{P}(A_j) = (1/2)^{j-1}(1/3)$ a

$$\mathbb{P}(E) = \sum_{j=1}^{\infty}(1/3)(1/2)^{j-1} = (1/3)\sum_{j=1}^{\infty}(1/2)^{j-1} = 2/3$$

    - Pozn. platí $\sum_{j=k}^{\infty} r^j = r^k/(1-r)$ pro $0 < r < 1$

Shrnutí

- Jevy A a B jsou nezávislé právě když $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$
- Nezávislost se někdy předpokládá, někdy odvozuje
- Disjunktní jevy s nenulovou pravděpodobností nejsou nezávislé

Podmíněná pravděpodobnost

Podmíněná pravděpodobnost

Definice 41.

Jestliže $\mathbb{P}(B) > 0$, pak podmíněná pravděpodobnost jevu A za předpokladu, že nastal jev B je

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

- $\mathbb{P}(A|B)$ vyjadřuje kolikrát nastal jev A mezi jevy, kde nastal jev B
- Pro pevné B s $\mathbb{P}(B) > 0$ je $\mathbb{P}(\cdot|B)$ pravděpodobnost (splňuje axiomy)
  - $\mathbb{P}(A|B) \geq 0$
  - $\mathbb{P}(\Omega|B) = 1$ a
  - pokud jsou $A_1, A_2, \ldots$ disjunktní, tak $\mathbb{P}(\cup_{i=1}^{\infty} A_i|B) = \sum_{i=1}^{\infty} \mathbb{P}(A_i|B)$
- Obecně neplatí $\mathbb{P}(A|B \cup C) = \mathbb{P}(A|B) + \mathbb{P}(A|C)$
- Obecně neplatí $\mathbb{P}(A|B) = \mathbb{P}(B|A)$
  - např. pravděpodobnost vyrážky u spalniček je 1, ale pravděpodobnost spalniček při vyrážce není 1

Příklad 42 (Testování).

Test na nemoc n dává výsledky $+$ a $-$. Otestování 1000 vzorků (10 s virem, 990 bez) dopadlo:

	n	n^c
$+$	.009	.099
$-$	.001	.891

Z definice podmíněné pravděpodobnosti máme

$$\mathbb{P}(+|n) = \frac{\mathbb{P}(+\cap n)}{\mathbb{P}(n)} = \frac{.009}{.009+.001} = .9 \quad \text{a} \quad \mathbb{P}(-|n^c) = \frac{.891}{.099+.891} = .9$$

Nemocný má pozitivní test (senzitivita) v 90 % a zdravý má negativní test (specificita) v 90 %. Když si nechám udělat test a ten bude pozitivní, jaká je pravděpodobnost, že jsem nemocný?

$$\mathbb{P}(n|+) = \frac{\mathbb{P}(+\cap n)}{\mathbb{P}(+)} = \frac{.009}{.009+.099} \approx .08 \quad \text{asi 8 \%}$$

Lemma 43.

1. *Jestliže A a B jsou nezávislé jevy, pak $\mathbb{P}(A|B) = \mathbb{P}(A)$.*
2. *Pro libovolnou dvojici jevů A a B je*

$$\mathbb{P}(AB) = \mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A)$$

Důkaz.

Přímo z definice. □

- Jiná interpretace nezávislosti tedy je, že znalost B nemění pravděpodobnost A
- Rovnost $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B|A)$ se hodí

Příklad 44.

- Bez opakování vybereme z balíčku dvě karty
  - Nechť A je jev, že první karta je křížové eso
  - Nechť B je jev, že druhá karta je kárová dáma
- Pak $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B|A) = (1/52) \times (1/51)$

Shrnutí

- Jestliže $\mathbb{P}(B) > 0$, pak $\mathbb{P}(A|B) = \dfrac{\mathbb{P}(AB)}{\mathbb{P}(B)}$
- $\mathbb{P}(\cdot|B)$ splňuje axiomy pravděpodobnosti pro fixní B
- Obecně $\mathbb{P}(A|\cdot)$ nesplňuje axiomy pravděpodobnosti pro fixní A
- Obecně $\mathbb{P}(A|B) \neq \mathbb{P}(B|A)$
- A a B jsou nezávislé právě když $\mathbb{P}(A|B) = \mathbb{P}(A)$

Bayesova věta

Bayesova věta

Bayesova věta je základem expertních systémů a Beyesovských sítí

Věta 45 (Věta o úplné pravděpodobnosti).

Nechť $A_1, \ldots, A_k$ je rozklad Ω. Pak pro libovolný jev B platí

$$\mathbb{P}(B) = \sum_{i=1}^{k} \mathbb{P}(B|A_i)\mathbb{P}(A_i).$$

Důkaz.

Definujme $C_j = B \cap A_j$, pak $C_1, \ldots, C_k$ jsou disjunktní a $B = \cup_{j=1}^{k} C_j$, tedy

$$\mathbb{P}(B) = \sum_{j=1}^{k} \mathbb{P}(C_j) = \sum_{j=1}^{k} \mathbb{P}(B \cap A_j) = \sum_{j=1}^{k} \mathbb{P}(B|A_j)\mathbb{P}(A_j)$$

protože $\mathbb{P}(B \cap A_j) = \mathbb{P}(B|A_j)\mathbb{P}(A_j)$ z definice podmíněné pravděpodobnosti. $\square$

Věta 46 (Bayesova věta).

Nechť $A_1, \ldots, A_k$ *je rozklad* Ω *t.ž.* $\mathbb{P}(A_i) > 0$ *pro všechna* i*. Jestliže* $\mathbb{P}(B) > 0$*, pak pro každé* $i = 1, \ldots, k$ *platí*

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{j=1}^{k} \mathbb{P}(B|A_j)\mathbb{P}(A_j)}$$

Důkaz.

Z dvojího použití definice podmíněné pravděpodobnosti a věty o úplné pravděpodobnosti máme

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_i B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_j \mathbb{P}(B|A_j)\mathbb{P}(A_j)}$$

$\square$

Příklad 47.

- Rozdělíme emaily do tří kategorií:
  - A_1 = "spam"
  - A_2 = "nedůležité" a
  - A_3 = "důležité"
- Ze zkušenosti víme, že $\mathbb{P}(A_1) = .7$, $\mathbb{P}(A_2) = .2$ a $\mathbb{P}(A_3) = .1$.
  - Zajisté platí $.7 + .2 + .1 = 1$.
- Nechť B je jev, že email obsahuje slovo "free".
  - Ze zkušenosti víme, že $\mathbb{P}(B|A_1) = .9$, $\mathbb{P}(B|A_2) = .01$, $\mathbb{P}(B|A_3) = .01$.
    - Pozn. $.9 + .01 + .01 \neq 1$.
- Pokud obdržíme email se slovem "free", jaká je pravděpodobnost, že jde o spam?
  - Bayesova věta dává

$$\mathbb{P}(A_1|B) = \frac{.9 \times .7}{(.9 \times .7) + (.01 \times .2) + (.01 \times .1)} \approx .995$$

Jednoduchá aplikace

- Máme tři mince a víme, že dvě jsou férové a jedna je cinknutá – orel padá s pravděp. $2/3$.
- Nevíme, která je cinknutá.
- Náhodně zamícháme mince a pak je postupně hodíme.
- Na první a druhé padne orel, na třetí panna.
- Jaká je pravděpodobnost, že první mince je ta cinknutá?

- Mince jsou v náhodném pořadí, a proto každá z nich má stejnou šanci být ta cinknutá.
- Nechť E_i je jev, že i-tá hozená mince je ta cinknutá.
- Nechť B je jev, že na mincích postupně padl orel, orel a panna.

- Před hodem mincí máme, že $\mathbb{P}(E_i) = 1/3$, pro všechna i.
- Též lze určit pravděpodobnost jevu B za předpokladu jevu E_i:

$$\mathbb{P}(B|E_1) = \mathbb{P}(B|E_2) = \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{6}$$

  a

$$\mathbb{P}(B|E_3) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{12}$$

- Použitím Bayesovy věty dostaneme, že

$$\mathbb{P}(E_1|B) = \frac{\mathbb{P}(B|E_1)\mathbb{P}(E_1)}{\sum_{i=1}^{3} \mathbb{P}(B|E_i)\mathbb{P}(E_i)} = \frac{2}{5}$$

Aplikace v informatice

První aplikace: rovnost polynomů

Rovnost polynomů

- Počítače dělají chyby (programátoři, hardware, zaokrouhlování,. . .)
- Pro některé problémy můžeme k ověření použít náhodnost
- Jak např. ověřit, zda

$$(x+1)(x-2)(x+3)(x-4)(x+5)(x-6) \equiv x^6 - 7x^3 + 25?$$

- Jak ověřit, že $F(x) \equiv G(x)$ pro dva polynomy $F(x)$ a $G(x)$?
  - Lze je převést na kanonickou formu a porovnat koeficienty. . .
  - Pokud je $F(x)$ ve tvaru $F(x) = \prod_{i=1}^{d}(x - a_i)$, převod na kanonickou formu postupným roznásobování vyžaduje $\Theta(d^2)$ násobení
    - Předpoklad: násobení je konstantní operace

Náhodnostní algoritmus

- Nechť d je maximální stupeň polynomů $F(x)$ a $G(x)$
- Algoritmus zvolí celé číslo $r \in \{1, \dots, 100d\}$ s rovnoměrnou pravděpodobností
- Spočítá hodnoty $F(r)$ a $G(r)$
  - Pokud je $F(r) \neq G(r)$, vrátí "neekvivalentní"
  - Pokud je $F(r) = G(r)$, vrátí "ekvivalentní"

- Složitost:
  - Nechť stačí jeden krok na vygenerování $r \in \{1, \dots, 100d\}$
  - Výpočet $F(r)$ a $G(r)$ vezme $O(d)$ $\rightsquigarrow$ rychlejší než výpočet kanonické formy

Korektnost

- Algoritmus: zvol $r \in \{1, \dots, 100d\}$ rovnoměrně a spočítej $F(r)$ a $G(r)$:
  - Pokud je $F(r) \neq G(r)$, vrať "neekvivalentní"
  - Pokud je $F(r) = G(r)$, vrať "ekvivalentní"

- Náhodnostní algoritmus může dát chybnou odpověď
  - Pokud $F(x) \equiv G(x)$, odpověď je správná
  - Pokud $F(x) \not\equiv G(x)$ a $F(r) \neq G(r)$, odpověď je opět správná
  - Pokud $F(x) \not\equiv G(x)$ a $F(r) = G(r)$, odpověď je špatná
    - To nastane, pokud r je řešení rovnice $F(x) - G(x) = 0$
- Jaká je šance udělat chybu?
  - Stupeň polynomu $F(x) - G(x)$ je nejvýše d
  - Polynom stupně nejvýše d má nejvýše d kořenů
  - Pro $F(x) \not\equiv G(x)$ dává nejvýše d hodnot z $\{1, \dots, 100d\}$ rovnost $F(r) = G(r)$, tj.

$$\mathbb{P}(\text{špatné odpovědi}) = \frac{d}{100d} = \frac{1}{100}$$

Druhá aplikace: násobení matic

Násobení matic

- Mějme tři binární matice (tj. prvky jsou modulo 2) A, B, C typu $n \times n$.
- Je $AB = C$?

- Standardní násobení matic vyžaduje $\Theta(n^3)$ operací,
  - sofistikovanější metody $\Theta(n^{2.37})$ operací.

Náhodnostní algoritmus:

- Zvolme $r = (r_1, r_2, \ldots, r_n) \in \{0, 1\}^n$ a spočítejme $A(Br)$ a Cr.
  - Pokud je $A(Br) \neq Cr$, vrať $AB \neq C$; jinak vrať $AB = C$.
- Složitost: $\Theta(n^2)$ operací.

Věta 48.

Pokud je $AB \neq C$ *a* $r \in \{0,1\}^n$ *je zvolen náhodně s rovnoměrnou pravděpodobností, tak*

$$\mathbb{P}(ABr = Cr) \leq \frac{1}{2}$$

Důkaz:

- Prostor elementárních jevů vektoru r je $\Omega = \{0,1\}^n$.
- Uvažujeme jev $\{r \in \{0,1\}^n \mid ABr = Cr\} \subseteq \Omega$.
- Volba $r = (r_1, \ldots, r_n) \in \{0,1\}^n$ s rovnoměrnou pravděpodobností je ekvivalentní s volbami r_i nezávisle s rovnoměrnou pravděpodobností z $\{0,1\}$.
  - Každé r_i je zvoleno nezávisle, rovnoměrně, proto je každý vektor z 2^n možných vektorů zvolen s pravděpodobností 2^{-n}.
- Nechť $D = AB - C \neq 0$, pak $ABr = Cr$ dává $Dr = 0$.
- Ale $D \neq 0$, proto má nenulový prvek, řekněme d_{11}.

Důkaz pokračování

- Pro $Dr = 0$ musí platit, že $\sum_{j=1}^{n} d_{1j} r_j = 0$, tj.

$$r_1 = -\frac{\sum_{j=2}^{n} d_{1j} r_j}{d_{11}} \tag{3}$$

- Místo vektoru r budeme volit r_k nezávisle a rovnoměrně z $\{0, 1\}$ v pořadí od r_n k r_1.
- Uvažme situaci těsně před volbou r_1:
  - v tomto okamžiku máme na pravé straně (3) číslo
  - a nejvýše jednu volbu pro r_1 splňující (3).
- Protože existují dvě možnosti pro volbu $r_1 \in \{0, 1\}$, rovnost nastane s pravděp. $\leq 1/2$.

⇝ Této technice se říká princip odloženého rozhodování.

- Pokud existuje několik náhodných veličin, jako naše r_i vektoru r, často pomůže uvažovat o některých jako by byly definovány v nějakém kroku algoritmu, zatímco ostatní jsou voleny náhodně – odloženy – do nějakého budoucího bodu v analýze.
- Formálně to odpovídá podmiňování odhalených hodnot: když jsou nějaké náhodné veličiny odhaleny, musíme podmínit odhalené hodnoty ve zbytku analýzy.

Dokončení důkazu – použijeme větu o úplné pravděpodobnosti

$$
\begin{aligned}
\mathbb{P}(ABr = Cr) &= \sum_{(x_2,\dots,x_n)\in\{0,1\}^{n-1}} \mathbb{P}((ABr = Cr) \cap ((r_2,\dots,r_n) = (x_2,\dots,x_n))) \\
&\le \sum_{(x_2,\dots,x_n)\in\{0,1\}^{n-1}} \mathbb{P}((r_1 = -\frac{\sum_{j=2}^n d_{1j}r_j}{d_{11}}) \cap ((r_2,\dots,r_n) = (x_2,\dots,x_n))) \\
&= \sum_{(x_2,\dots,x_n)\in\{0,1\}^{n-1}} \mathbb{P}(r_1 = -\frac{\sum_{j=2}^n d_{1j}r_j}{d_{11}})\mathbb{P}((r_2,\dots,r_n) = (x_2,\dots,x_n)) \\
&\le \sum_{(x_2,\dots,x_n)\in\{0,1\}^{n-1}} \frac{1}{2}\mathbb{P}((r_2,\dots,r_n) = (x_2,\dots,x_n)) \\
&= \frac{1}{2}
\end{aligned}
$$

Nezávislost r_1 a $(r_2,\dots,r_n)$ je použita na čtvrtém řádku. $\square$

Jak snížit chybu?

- K vylepšení chyby můžeme algoritmus několikrát opakovat.
  - Pokud najdeme r t.ž. $ABr \neq Cr$, algoritmus správně vrátí $AB \neq C$.
  - Pokud při každém opakování dostaneme, že $ABr = Cr$, pak algoritmus vrátí $AB = C$
    - ale je zde jistá pravděpodobnost chyby.
- Opakování algoritmu k-krát zvýší složitost na $\Theta(kn^2)$.
  - Např. pro 100 opakování je složitost stále $\Theta(n^2)$.
- Volba $r \in \{0,1\}^n$ pro k opakování dává pravděpodobnost chyby 2^{-k}.
  - Pravděpodobnost chyby algoritmu při 100 opakováních je nejvýše 2^{-100}.
  - V praxi je pak mnohem pravděpodobnější, že počítač “klekne” během provádění algoritmu než to, že vrátí špatnou odpověď.

Jak přispívá opakování algoritmu ke zvýšení důvěry?

- Pokud nemáme další informace o procesu, který vygeneroval testovanou identitu, je rozumné začít s tím, že identita platí s pravděpodobností 1/2.
  - Pokud spustíme test jednou a on vrátí, že identita platí, jak to ovlivní naši jistotu ve správnost identity?
- Mějme jevy $E =$ "rovnost platí" a $B =$ "test vrací, že rovnost platí".
- Pak $\mathbb{P}(E) = \mathbb{P}(E^c) = 1/2$ a protože má test chybu $1/2$, dostaneme
  - $\mathbb{P}(B|E) = 1$ a
  - $\mathbb{P}(B|E^c) \leq 1/2$.
- Bayesova věta pak dává, že

$$\mathbb{P}(E|B) = \frac{\mathbb{P}(B|E)\mathbb{P}(E)}{\mathbb{P}(B|E)\mathbb{P}(E) + \mathbb{P}(B|E^c)\mathbb{P}(E^c)} \geq \frac{\frac{1}{2}}{\frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2}} = \frac{2}{3}$$

- Opakujme test a předpokládejme, že znovu vrátí, že rovnost platí.
  - Po prvním testu můžeme zrevidovali model a dostat $\mathbb{P}(E) \geq 2/3$ a $\mathbb{P}(E^c) \leq 1/3$.
- Nechť $B =$ "nový test vrací, že rovnost platí";
  - protože testy jsou nezávislé, máme $\mathbb{P}(B|E) = 1$ a $\mathbb{P}(B|E^c) \leq 1/2$.
- Bayesova věta dává

$$\mathbb{P}(E|B) \geq \frac{2/3}{2/3 + 1/3 \cdot 1/2} = \frac{4}{5}$$

- Obecně: pokud náš model před testem je že $\mathbb{P}(E) \geq 2^i/(2^i + 1)$ a pokud "test vrátí, že rovnost platí" (jev B), pak

$$\mathbb{P}(E|B) \geq \frac{\frac{2^i}{2^i+1}}{\frac{2^i}{2^i+1} + \frac{1}{2}\frac{1}{2^i+1}} = \frac{2^{i+1}}{2^{i+1}+1} = 1 - \frac{1}{2^{i+1}+1}$$

  - Pokud po 100 opakováních test vrací, že rovnost platí, je naše víra ve správnost rovnosti alesoň $1 - 1/(2^{101} + 1) > 0.99999$.

Aplikace: Naivní bayesovský klasifikátor

Naivní bayesovský klasifikátor

- Algoritmus učení s učitelem
  - Klasifikuje objekty odhadováním pravděpodobnosti pomocí Bayesovy věty v jednoduchém (naivním) modelu.
  - Velmi dobrý v aplikacích: klasifikace textových dokumentů či emailový spam filter
  - Deterministický algoritmus postavený na konceptu podmíněné pravděpodobnosti

- Mějme n tréninkových dat $\{(D_1, c(D_1)), \ldots, (D_n, c(D_n))\}$
  - D_i je vektor tvaru $x^i = (x^i_1, \ldots, x^i_m)$
    - D_i je objekt s vlastnostmi $(X_1, \ldots, X_m)$
    - $x^i = (x^i_1, \ldots, x^i_m)$ znamená, že pro D_i je $X_1 = x^i_1$, ..., $X_m = x^i_m$.
    - Např. pokud je D_i textový dokument a vlastnosti jsou klíčová slova, X_j může být t.ž.

      $$x^i_j = \begin{cases} 1 & \text{pokud je } j\text{-té klíčové slovo v } D_i \\ 0 & \text{jinak} \end{cases}$$
  - Mějme množinu $C = \{c_1, \ldots, c_t\}$ možných klasifikací objektu
    - C může být např. množina labelů $\{``spam'', ``no\text{-}spam''\}$
    - $c(D_i)$ je klasifikace D_i

- Pro klasifikaci předpokládáme, že tréninková množina je vzorek z nějakého neznámého rozdělení pravděpodobnosti.
  - Cílem je pro daný nový dokument najít přesnou klasifikaci.

- Obecněji můžeme hledat vektor $(z_1, \ldots, z_t)$, kde z_j je odhad pravděp., že $c(D_i) = c_j$.
  - Pokud bychom hledali nejpravděpodobnější klasifikaci, vrátíme c_j s největší hodnotou z_j.

- Mějme velkou tréninkovou množinu $D = \{(D_1, c(D_1)), \ldots, (D_n, c(D_n))\}$. Z ní určíme empirickou podmíněnou pravděpodobnost toho, že objekt s vlastnostmi $y = (y_1, \ldots, y_m)$ je klasifikován jako c_j:

$$p_{y,j} = \frac{|\{i \mid D_i = y, c(D_i) = c_j\}|}{|\{i \mid D_i = y)\}|}$$

$$p_{y,j} = \frac{|\{i \mid D_i = y, c(D_i) = c_j\}|}{|\{i \mid D_i = y)\}|}$$

- Nechť D^* je nový objekt s vlastnostmi x^* a stejným rozdělením, pak $p_{x^*,j}$ je empirický odhad podmíněné pravděpodobnosti

$$\mathbb{P}(c(D^*) = c_j \mid x^* = (x_1^*, \dots, x_m^*))$$

  - Tyto hodnoty by šlo předpočítat a pro x^* vrátit vektor $(z_1, \dots, z_t) = (p_{x^*,1}, \dots, p_{x^*,t})$.
  - Museli bychom ale uvažovat všechny možné kombinace hodnot m vlastností
    - pro vlastnosti se dvěmi hodnotami potřebujeme 2^m podmíněných pravděp. pro každou třídu, tj. celkem $\Omega(|C|2^m)$ vzorků.

- Trénovací proces je rychlejší a pokud uvážíme naivní model, kde vlastnosti jsou nezávislé, vyžaduje výrazně méně příkladů.
  - Pak máme

$$\begin{aligned}\mathbb{P}(c(D^*) = c_j \mid x^*) &= \frac{\mathbb{P}(x^* \mid c(D^*) = c_j) \cdot \mathbb{P}(c(D^*) = c_j)}{\mathbb{P}(x^*)} \\ &= \frac{\prod_{k=1}^{m} \mathbb{P}(x_k^* = x_k \mid c(D^*) = c_j) \cdot \mathbb{P}(c(D^*) = c_j)}{\mathbb{P}(x^*)}\end{aligned}$$

  - Zde x_k^* reprezentuje k-tou komponentu vektoru x^* objektu D^*.

- S konstantním počtem možných hodnot pro každou vlastnost potřebujeme odhady pouze pro $O(m|C|)$ pravděpodobností.

- Označme $\hat{\mathbb{P}}$ empirickou pravděp., tj. relativní frekvenci jevů z trénovací množiny.
  - notace zdůrazňuje, že bereme odhady pravděpodobností jak byly určeny z trénovací množiny.
    - v praxi se často dělá drobná modifikace, jako např. přidání $1/2$ k čitateli každého zlomku, aby empirická pravděp. nebyla 0.

- Trénovací process je následující:
  1. Pro každou klasifikaci c_j spočítáme
     $$\hat{\mathbb{P}}(c(D^*) = c_j) = \frac{|\{i \mid c(D_i) = c_j\}|}{|D|}$$
     kde $|D|$ je počet objektů v trénovací množině.
  2. Pro každou vlastnost X_k a každou její hodnotu x_k spočítáme
     $$\hat{\mathbb{P}}(x^*_k = x_k \mid c(D^*) = c_j) = \frac{|\{i \mid x^i_k = x_k, c(D_i) = c_j\}|}{|\{i \mid c(D_i) = c_j\}|}$$

Klasifikace nového objektu D^*

- Určení nejpravděpodobnější klasifikace $x^* = x = (x_1, \ldots, x_m)$

$$c(D^*) = \arg\max_{c_j \in C} \left\{ \left(\prod_{k=1}^{m} \hat{\mathbb{P}}(x_k^* = x_k \mid c(D^*) = c_j \right) \hat{\mathbb{P}}(c(D^*) = c_j) \right\}$$

  - tj., po natrénování klasifikátoru je clasifikace nového objektu D^* s vektorem vlastností $x^* = (x_1^*, \ldots, x_m^*)$ určena pomocí

$$\left(\prod_{k=1}^{m} \hat{\mathbb{P}}(x_k^* = x_k \mid c(D^*) = c_j) \right) \cdot \hat{\mathbb{P}}(c(D^*) = c_j)$$

    pro každé c_j a vzetím klasifikace s nejvyšší hodnotou.

- Naivní bayesovský klasifikátor je efektivní a jednoduchý na implementaci
  - kvůli naivnímu modelu
- Může však vést k chybným výsledkům, pokud klasifikace závisí na kombinaci vlastností:
  - Mějme položky charakterizované booleovskými vlastnostmi X a Y.
  - Pokuk je $X = Y$, je položka klasifikována jako patřící do třídy A, jinak do třídy B.
  - Pokud má trénovací množina stejný počet položek v každé třídě pro každou hodnotu X a Y, pak jsou všechny podmíněné pravděpodobnosti určené klasifikátorem rovny 0.5
    - náš klasifikátor tak není lepší než házení mincí
  - V praxi se takové úkazy vyskytují zřídka a naivní bayesovský klasifikátor je často velmi efektivní.

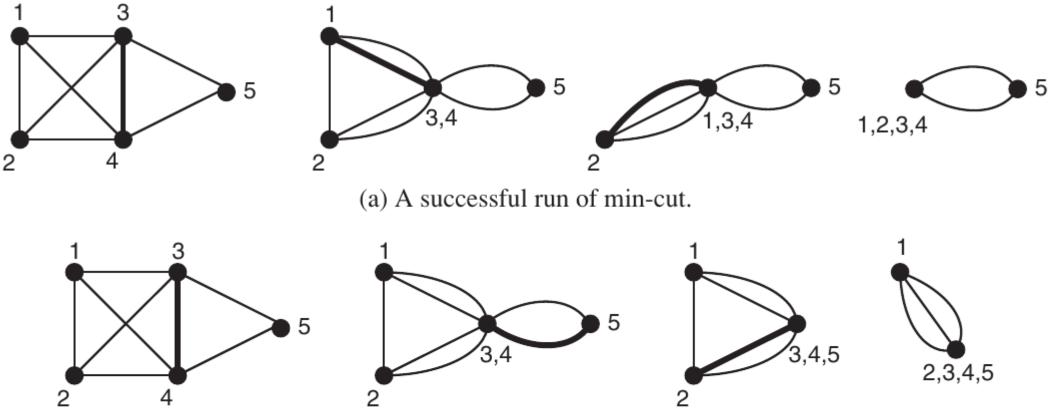
Aplikace: Náhodnostní Min-Cut Algoritmus

Náhodnostní Min-Cut Algoritmus

- Řez v grafu je množina hran, jejíž odstranění rozdělí graf na dvě či více souvislých komponent.
- Min-cut problém $=$ najdi v grafu $G = (V, E)$ s n vrcholy řez s minimální kardinalitou.
  - Problém má mnoho použití, např. ve spolehlivosti sítí.
    - Když uzly reprezentují počítače a hrany jejich propojení v síti, tak min-cut je nejmenší počet spojení, které se musí přerušit, aby některé dva počítače nemohly komunikovat.
  - Další použití např. ve shlukové analýze.
    - Pokud jsou uzly webové stránky (či lib. hypertextové dokumenty) a dva uzly mají hranu když mezi sebou mají hyperlink, pak malé řezy rozdělují graf do shluků dokumentů mezi nimiž je málo linků.
    - Mezi dokumenty v různých shlucích není pravděpodobně žádný vztah.

Aplikace: Náhodnostní Min-Cut Algoritmus

- Algoritmus obsahuje $n-2$ iterací.
  - V každé iteraci vybere hranu a kontrahuje ji.
    - Kontrakce hrany $(u,v) =$ sloučení u a v do jednoho uzlu a odstranění všech hran mezi u a v.
    - Výsledný graf může mít paralelní hrany, ale ne smyčky.
  - Algoritmus vybírá hranu náhodně s uniformním rozdělením z aktuální množiny hran.
- Každá iterace redukuje počet uzlů o jeden, tj. po $n-2$ iteracích má graf pouze dva uzly.
  - Algoritmus vrátí množinu hran spojující tyto dva uzly.
- Ověřte, že lib. řez grafu po iteraci je též řez grafu před iterací.
  - Opak neplatí, tj. ne každý řez grafu před iterací je též řez grafu po iteraci – některé hrany řezu mohly být kontrahovány v předchozích iteracích.
  - Nicméně, výsledek algoritmu je vždy řez, ne však nutně minimální kardinality, viz obrázek.



(a) A successful run of min-cut.

(b) An unsuccessful run of min-cut.

Figure 1.1: An example of two executions of min-cut in a graph with minimum cut-set of size 2.

Credit: *M. Mitzenmacher, E. Upfal: Probability and Computing, Cambridge University Press, 2017.*

Věta 49.

Algoritmus vrátí minimální řez s pravděpodobností alespoň $2/(n(n-1))$.

Důkaz

- Nechť k je velikost min. řezů v G a C je nějaký min. řez.
  - Odstranění hran z C rozloží uzly G na S a $V-S$ t.ž. neexistuje hrana z S do $V-S$.
- Pokud algoritmus v každém kroku kontrahuje hranu spojující dva uzly z S či dva z $V-S$, tj. ne z C, pak po $n-2$ iteracích vrátí graf se dvěma uzly spojené hranami z C.
  - Pokud tedy nikdy nevybere hranu z C ve svých $n-2$ iteracích, vrátí C jako minimální řez.
- Nechť E_i je jev, že "kontrahovaná hrana vybraná v iteraci i není z C"
- Nechť $F_i = \bigcap_{j=1}^{i} E_j$ je jev, že "v prvních i iteracích nebyla kontrahována žádná hrana z C".
  - Určíme $\mathbb{P}(F_{n-2})$.

$E_i =$ "kontrahovaná hrana vybraná v iteraci i není z C"; $F_i = \bigcap_{j=1}^{i} E_j$.

Důkaz pokračování

- Začneme s $\mathbb{P}(E_1) = \mathbb{P}(F_1)$.
- Jelikož má min řez k hran, jsou všechny uzly grafu stupně alespoň k
  - tj. graf musí mít alesoň $nk/2$ hran
- První kontrahovaná hrana je vybrána náhodně (rovnoměrně) z množiny všech hran.
- Jelikož máme alespoň $nk/2$ hran a C má k hran, pravděpodobnost toho, že vybereme hranu z C v první iteraci je

$$\mathbb{P}(E_1) = \mathbb{P}(F_1) \geq 1 - \frac{2k}{nk} = 1 - \frac{2}{n}$$

$E_i =$ "kontrahovaná hrana vybraná v iteraci i není z C"; $F_i = \bigcap_{j=1}^{i} E_j$.

Důkaz pokračování

- Nechť první kontrakce neodstranila hranu z C, tj. nastal jev F_1.
  - Po první iteraci tak máme $(n-1)$-uzlový graf s min řezem velikosti k.
  - Stupeň jeho uzlů je alesoň k a graf má alesoň $k(n-1)/2$ hran.
  - Pak
$$\mathbb{P}(E_2 \mid F_1) \geq 1 - \frac{k}{k(n-1)/2} = 1 - \frac{2}{n-1}$$
  - a obecně
$$\mathbb{P}(E_i \mid F_{i-1}) \geq 1 - \frac{k}{k(n-i+1)/2} = 1 - \frac{2}{n-i+1}$$

Důkaz pokračování

- Nyní máme

$$
\begin{aligned}
\mathbb{P}(F_{n-2}) &= \mathbb{P}(E_{n-2} \cap F_{n-3}) \\
&= \mathbb{P}(E_{n-2} \mid F_{n-3}) \cdot \mathbb{P}(F_{n-3}) \\
&= \mathbb{P}(E_{n-2} \mid F_{n-3}) \cdot \mathbb{P}(E_{n-3} \mid F_{n-4}) \cdots \mathbb{P}(E_2 \mid F_1) \cdot \mathbb{P}(F_1) \\
&\geq \prod_{i=1}^{n-2} (1 - \frac{2}{n-i+1}) \\
&= \prod_{i=1}^{n-2} (\frac{n-i-1}{n-i+1}) \\
&= \frac{n-2}{n} \cdot \frac{n-3}{n-1} \cdot \frac{n-4}{n-2} \cdots \frac{2}{4} \cdot \frac{1}{3} \\
&= \frac{2}{n(n-1)}
\end{aligned}
$$

- Jelikož má algoritmus jednostrannou chybu, můžeme pravděpodobnost chyby algoritmu zredukovat jeho opakováním.
- Pokud bychom algoritmus opakovali $n(n-1)\ln n$ krát a vrátili řez s minimální velikostí nalezený během všech opakování, pak pravděpodobnost toho, že výsledek není min řez je nejvýše
$$\left(1-\frac{2}{n(n-1)}\right)^{n(n-1)\ln n} \leq e^{-2\ln n} = \frac{1}{n^2}$$
    - Zde se použilo to, že $1-x \leq e^{-x}$.

Náhodná veličina

Náhodná veličina

- Statistika a data mining se zabývají daty
- Jak spojit výběrové prostory a jevy s daty?
  - Pomocí konceptu náhodné veličiny

Definice 50.

Náhodná veličina je funkce $X\colon \Omega \to \mathbb{R}$, která přiřazuje reálné číslo $X(\omega)$ každému výsledku $\omega \in \Omega$.

- Pravděpodobnostní míra $\mathbb{P}$ je definována na σ-algebře $\mathcal{A}$ prostoru Ω.
- Náhodná veličina X je měřitelná funkce $X\colon \Omega \to \mathbb{R}$.
  - Měřitelná znamená, že pro každé x je množina $\{\omega \in \Omega \mid X(\omega) \leq x\}$ jev
    - tj. $\{\omega \mid X(\omega) \leq x\} \in \mathcal{A}$

Poznámka 51.

Ač budeme pracovat přímo s náhodnými veličinami bez uvádění výběrového prostoru, je třeba si uvědomit, že výběrový prostor tam někde vždy je!

Příklad 52.

Hod mincí desetkrát po sobě. Nechť $X(\omega)$ be počet orlů v posloupnosti ω, např. pro $\omega = OOPOOPOOPP$ je $X(\omega) = 6$.

Příklad 53.

Nechť $\Omega = \{(x, y) \in \mathbb{R} \times \mathbb{R} \mid x + y \leq 1\}$. Náhodně zvolme bod z Ω. Typický výsledek je tvaru $\omega = (x, y)$. Příklady náhodných veličin:

- $X(\omega) = x$
- $Y(\omega) = y$
- $Z(\omega) = x + y$
- $W(\omega) = \sqrt{x^2 + y^2}$

Pro danou náhodnou veličinu X a podmnožinu A reálné osy definujeme

$$X^{-1}(A) = \{\omega \in \Omega \mid X(\omega) \in A\}$$

a dále

$$\mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in A\})$$

$$\mathbb{P}(X = x) = \mathbb{P}(X^{-1}(x)) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) = x\})$$

Všimněme si, že X značí náhodnou veličinu a x značí konkrétní hodnotu X.

Příklad 54.

Hod mincí dvakrát po sobě. Nechť X je počet orlů. Pak

- $\mathbb{P}(X = 0) = \mathbb{P}(\{PP\}) = 1/4$
- $\mathbb{P}(X = 1) = \mathbb{P}(\{OP, PO\}) = 1/2$
- $\mathbb{P}(X = 2) = \mathbb{P}(\{OO\}) = 1/4.$

Náhodnou veličinu a její distribuci lze zapsat tabulkou

ω	$\mathbb{P}(\{\omega\})$	$X(\omega)$
PP	1/4	0
PO	1/4	1
OP	1/4	1
OO	1/4	2

x	$\mathbb{P}(X = x)$
0	1/4
1	1/2
2	1/4

Distribuční a pravděpodobnostní funkce

Distribuční a pravděpodobnostní funkce

Definice 55.

(Kumulativní) distribuční funkce je funkce $F_X\colon \mathbb{R} \to [0,1]$ definovaná jako

$$F_X(x) = \mathbb{P}(X \leq x)\,.$$

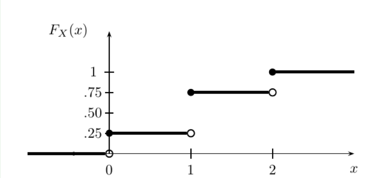
Distribuční funkce obsahuje veškerou informaci o náhodné veličině.

Občas píšeme distribuční funkce jako F místo F_X.

Příklad 56.

Hod férovou mincí dvakrát, X je počet orlů. Pak $\mathbb{P}(X = 0) = \mathbb{P}(X = 2) = 1/4$ a $\mathbb{P}(X = 1) = 1/2$. Distribuční funkce je

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1/4 & 0 \leq x < 1 \\ 3/4 & 1 \leq x < 2 \\ 1 & 2 \leq x \end{cases}$$



Důkladně prostudujte, distribuční funkce mohou být komplikované. Funkce je zprava spojitá, neklesající a definovaná pro všechna x. Proč je $F_X(1.4) = .75$?

Distribuční funkce plně určuje rozložení náhodné veličiny.

Věta 57.

Nechť X má distribuční funkci F a Y má distribuční funkci G. Jestliže

$$F(x) = G(x)$$

pro všechna x, pak

$$\mathbb{P}(X \in A) = \mathbb{P}(Y \in A)$$

pro všechna A.[3]

[3] Přesněji máme, že $\mathbb{P}(X \in A) = \mathbb{P}(Y \in A)$ pro každý měřitelný jev A

Věta 58.

Funkce $F\colon \mathbb{R} \to [0,1]$ je distribuční funkce pro nějakou pravděpodobnostní míru $\mathbb{P}$ právě tehdy, když F splňuje následující tři podmínky:

1. *F je neklesající:* $x_1 < x_2$ *implikuje, že* $F(x_1) \leq F(x_2)$
2. *F je normalizovaná:* $\lim_{x\to-\infty} F(x) = 0$ *a* $\lim_{x\to\infty} F(x) = 1$
3. *F je zprava spojitá:* $F(x) = F(x^+)$ *pro všechna x, kde* $F(x^+) = \lim\limits_{y\to x,\, x<y} F(y)$

Důkaz.

Ukážeme 3. Nechť x je reálné číslo a $y_1, y_2, \ldots$ posloupnost reálných čísel t.ž. $y_1 > y_2 > \cdots$ a $\lim_i y_i = x$. Nechť $A_i = (-\infty, y_i]$ a $A = (-\infty, x]$. Pak $A = \cap_{i=1}^{\infty} A_i$ a $A_1 \supset A_2 \supset \cdots$. Protože jsou jevy monotónní, je $\lim_i \mathbb{P}(A_i) = \mathbb{P}(\cap_i A_i)$, tj.

$$F(x) = \mathbb{P}(A) = \mathbb{P}(\cap_i A_i) = \lim_i \mathbb{P}(A_i) = \lim_i F(y_i) = F(x^+).$$

Důkaz 1 a 2 je podobný. Opačná implikace je komplikovanější. $\square$

Diskrétní náhodná veličina

Diskrétní náhodná veličina

Definice 59.

Náhodná veličina X je diskrétní, jestliže nabývá spočetně mnoho hodnot $\{x_1, x_2, \dots\}$.
Definujme pravděpodobnostní funkci pro X jako

$$f_X(x) = \mathbb{P}(X = x)$$

Pak $f_X(x) \geq 0$ pro všechna $x \in \mathbb{R}$ a $\sum_i f_X(x_i) = 1$.

Někdy píšeme f místo f_X.

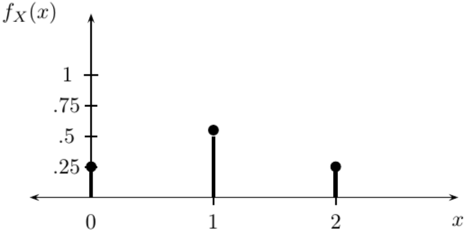
Distribuční funkce X souvisí s pravděp. funkcí f_X následovně

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{x_i \leq x} f_X(x_i)$$

Příklad 60.

Hod férovou mincí dvakrát, X je počet orlů. Pak $\mathbb{P}(X = 0) = \mathbb{P}(X = 2) = 1/4$ a $\mathbb{P}(X = 1) = 1/2$. Pravděpodobnostní funkce je

$$f_X(x) = \begin{cases} 1/4 & x = 0 \\ 1/2 & x = 1 \\ 1/4 & x = 2 \\ 0 & \text{jinak} \end{cases}$$



Vybrané diskrétní náhodné veličiny

Vybrané diskrétní náhodné veličiny

- $X \sim F$ značí, že X má rozdělení F.
- $X \sim F$ tedy čteme jako "X má rozdělení F", nikoli "X je přibližně F".

Bodové rozdělení

X má bodové rozdělení v a, $X \sim \delta_a$, jestliže $\mathbb{P}(X = a) = 1$, přičemž

$$F(x) = \begin{cases} 0 & x < a \\ 1 & x \geq a \end{cases}$$

Pravděpodobnostní funkce je $f(x) = 1$ pro $x = a$ a 0 jinak.

Diskrétní rovnoměrné rozdělení

Nechť $k > 1$ a nechť X má pravděpodobnostní funkci

$$f(x) = \begin{cases} \frac{1}{k} & \text{pro } x = 1, \ldots, k \\ 0 & \text{jinak} \end{cases}$$

Pak X má rovnoměrné (uniformní) rozdělení na $\{1, \ldots, k\}$.

Bernoulliho rozdělení

Nechť X představuje hod mincí. Pak

$$\mathbb{P}(X = 1) = p$$

a

$$\mathbb{P}(X = 0) = 1 - p$$

pro $p \in [0, 1]$.
Řekneme, že X má Bernoulliho rozdělení, $X \sim Bernoulli(p)$.
Pravděpodobnostní funcke je

$$f(x) = p^x (1-p)^{1-x} \text{ pro } x \in \{0, 1\}$$

Binomické rozdělení

Mějme minci na níž padá orel s pravděpodobností p pro nějaké $0 \leq p \leq 1$. Házíme n krát a X je počet hodů kdy padl orel. Předpokládejme, že hody jsou nezávislé. Pak pravděpodobnostní funkce je

$$f(x) = \mathbb{P}(X = x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, \dots, n \\ 0 & \text{jinak} \end{cases}$$

Taková NV se nazývá binomická, $X \sim Binomial(n, p)$.

Jestliže $X_1 \sim Binomial(n_1, p)$ a $X_2 \sim Binomial(n_2, p)$, pak $X_1 + X_2 \sim Binomial(n_1 + n_2, p)$.

Příklad 61.

Házíme pětkrát férovou mincí a předpokládáme, že hody jsou nezávislé. Jaká je pravděpodobnostní funkce počtu padlých orlů? Nechť X je NV vyjadřující počet úspěchů, tj. počet padlých orlů. Pak X má binomické rozdělení s parametry $n = 5$ a $p = 1/2$:

$$\mathbb{P}(X = 0) = \binom{5}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^5 = \frac{1}{32}$$

$$\mathbb{P}(X = 1) = \binom{5}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^4 = \frac{5}{32}$$

$$\mathbb{P}(X = 2) = \binom{5}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^3 = \frac{10}{32}$$

$$\mathbb{P}(X = 3) = \binom{5}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2 = \frac{10}{32}$$

$$\mathbb{P}(X = 4) = \binom{5}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^1 = \frac{5}{32}$$

$$\mathbb{P}(X = 5) = \binom{5}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^0 = \frac{1}{32}$$

Příklad

Příklad 62.

Jistý výrobce vyrábí produkt, o němž je známo, že je špatný s pravděpodobností 0.01, nezávisle na ostatních. Tento produkt výrobce prodává v balení po 10 kusech a nabízí výměnu balení, pokud je více než jeden produkt špatný. Jaké množství prodaných balení musí výrobce vyměnit?

Řešení: Pokud X značí počet špatných produktů, pak X je binomická NV s parametry $(10, .01)$. Pravděpodobnost, že balení bude vyměněno je tedy

$$1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) = 1 - \binom{10}{0}(0.01)^0(0.99)^{10} - \binom{10}{1}(0.01)^1(0.99)^9 \approx 0.004$$

Výrobce tedy bude muset vyměnit 0.4 procenta balení.

Poznámka

- X je náhodná veličina
- x je konkrétní hodnota náhodné veličiny
- n a p jsou parametry, tj. fixní reálná čísla
- Parametr p je obvykle neznámý a musí být odhadnut z dat
  - o tom je statisická inference
- V mnoha statistických modelech jsou NV a parametry – nezaměňovat!

Geometrické rozdělení

NV X má geometrické rozdělení s parametrem $p \in (0,1)$, $X \sim Geom(p)$, jestliže

$$f(x) = \mathbb{P}(X = x) = p(1-p)^{x-1}, \qquad x \ \geq 1.$$

Pak platí

$$\sum_{x=1}^{\infty} \mathbb{P}(X = x) = p \sum_{x=0}^{\infty} (1-p)^x = p \frac{1}{1-(1-p)} = 1$$

Příklad 63.

X vyjadřuje počet hodů mincí než poprvé padne orel.

Příklad 64.

Urna obsahuje N bílých a M černých míčků. Míčky jsou náhodně vybírány jeden po druhém, dokud není vybrán černý. Pokud je každý míček nahrazen jiným před tím, než se vybírá další, jaká je pravděpodobnost, že (a) je potřeba přesně n tahů? (b) je potřeba alespoň k tahů?

Řešení: Nechť X značí počet tahů, které jsou potřeba k vybrání černého míčku. Pak X je geometrická NV s $p = M/(M+N)$ a máme:

(a)

$$\mathbb{P}(X = n) = \left(\frac{N}{N+M}\right)^{n-1} \frac{M}{N+M} = \frac{MN^{n-1}}{(M+N)^n}$$

(b)

$$\mathbb{P}(X \geq k) = \frac{M}{M+N} \sum_{n=k}^{\infty} \left(\frac{N}{N+M}\right)^{n-1} = \left(\frac{N}{M+N}\right)^{k-1} = (1-p)^{k-1}$$

Poissonovo rozdělení

NV X má Poissonovo rozdělení s parametrem λ, $X \sim Poisson(\lambda)$, jestliže

$$f(x) = \mathbb{P}(X = x) = e^{-\lambda}\frac{\lambda^x}{x!} \qquad x \geq 0$$

Platí

$$\sum_{x=0}^{\infty} f(x) = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = e^{-\lambda}e^{\lambda} = 1$$

Poznámka 65.

Poissonovo rozdělení se často používá jako model pro počítání vyjímečných jevů (radioaktivní rozklad, dopravní nehody, atd.).

Jestliže $X_1 \sim Poisson(\lambda_1)$ a $X_2 \sim Poisson(\lambda_2)$, pak $X_1 + X_2 \sim Poisson(\lambda_1 + \lambda_2)$.

Aproximace binomického rozdělení

Poissonovo rozdělení má mnoho aplikací – lze jej použít jako aproximaci binomického rozdělení s parametry (n, p), kde n je velké a p je dostatečně malé, aby np bylo přiměřené.

Nechť X je binomická NV s parametry (n, p) a $\lambda = np$. Pak

$$\begin{aligned}\mathbb{P}(X = i) &= \frac{n!}{(n-i)!i!} p^i (1-p)^{n-i} = \frac{n!}{(n-i)!i!} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i} \\ &= \frac{n(n-1)\cdots(n-i+1)\lambda^i (1-\lambda/n)^n}{n^i i! (1-\lambda/n)^i}\end{aligned}$$

Pro velké n a přiměřené λ dostáváme

$$\left(1 - \frac{\lambda}{n}\right)^n \approx e^{-\lambda} \qquad \frac{n(n-1)\cdots(n-i+1)}{n^i} \approx 1 \qquad \left(1 - \frac{\lambda}{n}\right)^i \approx 1$$

a tedy

$$\mathbb{P}(X = i) \approx e^{-\lambda} \frac{\lambda^i}{i!}$$

Příklady použití

Pokud provedeme n nezávislých pokusů, každý s pravděpodobností úspěchu p, a pokud n je velké a p dostatečně malé, aby np bylo přiměřené, tak počet výskytu úspěchů je přibližně Poissonova NV s parametrem $\lambda = np$.

Hodnota λ se obvykle zjistí empiricky.

Příklady NV, které mají Poissonovo rozdělení:

- Počet překlepů na stránce knihy
- Počet lidí v komunitě, kteří se dožijí 100 let
- Počet špatně vytočených telefonních čísel denně
- Počet balíků psích sucharů prodaných v daném obchodě za den
- Počet zákazníků, kteří přijdou daný den na poštu
- Počet α-částic uvolněných z radioaktivního materiálu během fixního časového období

Příklad 66.

Nechť počet typografických chyb na jedné stránce knihy má Poissonovo rozdělení s parametrem $\lambda = 1/2$. Určete pravděpodobnost, že na dané stránce je chyba.

Řešení: Nechť X značí počet chyb na tomto slajdu. Pak máme

$$\mathbb{P}(X \geq 1) = 1 - P(X = 0) = 1 - e^{-1/2} \approx .393$$

Příklad 67.

Nechť pravděpodobnost, že produkt vyrobený jistým strojem bude vadný je .1. Určete pravděpodobnost toho, že vzorek 10 produktů bude obsahovat nejvýše jeden vadný.

Řešení: Hledaná pravděpodobnost je

$$\binom{10}{0}(.1)^0(.9)^{10} + \binom{10}{1}(.1)^1(.9)^9 = .7365$$

kdežto aproximace pomocí Poissonova rozdělení dává hodnotu

$$e^{-1} + e^{-1} \approx .7358$$

Poznámka

- NV jsou funkce z výběrového prostoru Ω do $\mathbb{R}$, ale v rozděleních nezmiňujeme výběrový prostor Ω. Ten tam vždy je a lze ho zkonstruovat.
- Zkonstruujme výběrový prostor např. pro Bernoulliho NV.
- Nechť $\Omega = [0,1]$ a definujme $\mathbb{P}([a,b]) = b - a$ pro $0 \leq a \leq b \leq 1$.
- Fixujme $p \in [0,1]$ a definujme

$$X(\omega) = \begin{cases} 1 & \omega \leq p \\ 0 & \omega > p \end{cases}$$

- Pak $\mathbb{P}(X = 1) = \mathbb{P}(\omega \leq p) = \mathbb{P}([0,p]) = p$ a $\mathbb{P}(X = 0) = 1 - p$.
- Tedy $X \sim Bernoulli(p)$.

Spojitá náhodná veličina

Spojitá náhodná veličina

Definice 68.

Náhodná veličina X je spojitá, jestliže existuje funkce f_X t.ž.

1. $f_X(x) \geq 0$ pro všechna x
2. $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$
3. pro každé $a \leq b$ platí

$$\mathbb{P}(a < X < b) = \int_a^b f_X(x)\,dx$$

Funkce f_X se nazývá hustota. Platí

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$$

a $f_X(x) = F_X'(x)$ ve všech bodech x, ve kterých je F_X diferencovatelná.

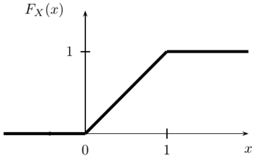
Příklad 69.

Nechť X má hustotu

$$f_X(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{jinak} \end{cases}$$

Pak $f_X(x) \ge 0$ a $\int f_X(x)\,dx = 1$. NV s touto hustotou má uniformní (rovnoměrné) rozdělení na intervalu $(0,1)$, tj. náhodný výběr bodu mezi 0 a 1. Distribuční funkce je pak dána jako

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x & 0 \le x \le 1 \\ 1 & x > 1 \end{cases}$$



Příklad 70.

Nechť X má hustotu

$$f_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{(1+x)^2} & \text{jinak} \end{cases}$$

Jelikož $\int f_X(x)\,dx = 1$, jde skutečně o hustotu.

Pozor!

- Spojité náhodné veličiny mohou být záludné!

- Pokud je X spojitá, tak $\mathbb{P}(X = x) = 0$ pro každé x.
  - Neuvažujte tedy o $f(x)$ jako o $\mathbb{P}(X = x)$, to platí pouze pro diskrétní NV.
  - Pravděpodobnosti získáme z hustoty integrací.

- Hustota může být větší než 1 (narozdíl od pravděpodobnosti)
  - např. pro
$$f(x) = \begin{cases} 5 & \text{pro } x \in [0, 1/5] \\ 0 & \text{jinak} \end{cases}$$
je $f(x) \geq 0$ a $\int f(x)\,dx = 1$, tudíž jde o hustotu, ač $f(x) = 5$.

- Hustota může být i neomezená
  - např. pro
$$f(x) = \begin{cases} (2/3)x^{-1/3} & \text{pro } 0 < x < 1 \\ 0 & \text{jinak} \end{cases}$$

Příklad 71.

Nechť

$$f(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{1+x} & \text{jinak} \end{cases}$$

Pak nejde o hustotu, protože $\int f(x)\,dx = \int_0^\infty \frac{dx}{1+x} = \int_1^\infty \frac{du}{u} = \ln\infty = \infty$.

Lemma 72.

Nechť F je distribuční funkce náhodné veličiny X. Pak

1. $\mathbb{P}(X = x) = F(x) - F(x^-)$, *kde* $F(x^-) = \lim_{y \to x^-} F(y)$*;*
2. $\mathbb{P}(x < X \leq y) = F(y) - F(x)$*;*
3. $\mathbb{P}(X > x) = 1 - F(x)$*;*
4. *Pokud je X spojitá, tak*

$$\begin{aligned} F(b) - F(a) &= \mathbb{P}(a < X < b) = \mathbb{P}(a \leq X < b) \\ &= \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X \leq b). \end{aligned}$$

Důkaz.

Intuice pro bod 2:
$\{X \leq y\} = \{X \leq x\} \cup \{x < X \leq y\}$, a tedy $\mathbb{P}(X \leq y) = \mathbb{P}(X \leq x) + \mathbb{P}(x < X \leq y)$ $\square$

Vybrané spojité náhodné veličiny

Rovnoměrné rozdělení

NV X má rovnoměrná rozdělení na intervalu (a, b), $X \sim Uniform(a, b)$, jestliže

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & \text{jinak} \end{cases}$$

kde $a < b$. Distribuční funkce je

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & x \in [a, b] \\ 1 & x > b \end{cases}$$

Příklad

Příklad 73.

Nechť X je NV s rovnoměrným rozdělením na intervalu $(0, 10)$. Určete pravděpodobnost, že

- X $<$ 3
- X $>$ 6
- 3 $<$ X $<$ 8

Řešení:

$$\mathbb{P}(X < 3) = \int_0^3 \frac{1}{10} dx = \frac{3}{10}$$

$$\mathbb{P}(X > 6) = \int_6^{10} \frac{1}{10} dx = \frac{4}{10}$$

$$\mathbb{P}(3 < X < 8) = \int_3^8 \frac{1}{10} dx = \frac{1}{2}$$

Normalní (Gaussovo) rozdělení

NV X má normální (Gaussovo) rozdělení s parametry μ a σ, $X \sim N(\mu, \sigma^2)$, jestliže

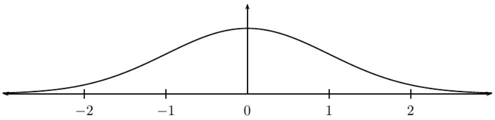
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad x \in \mathbb{R}$$

kde $\mu \in \mathbb{R}$ a $\sigma > 0$.

- Parametr μ je střední hodnota rozdělení
- σ je "rozložení" (směrodatná odchylka) rozdělení
  - střední hodnotu a směrodatnou odchylku definujeme později.
- Normální rozdělení hraje důležitou roli v pravděpodobnosti a statistice.
  - Mnoho přírodních fenoménů má přibližně normální rozdělení.
- Později budeme studovat centrální limitní větu, která říká, že rozdělení sumy náhodných veličin lze aproximovat normálním rozdělením.

Standardní normální rozdělení

- NV X má standardní normální rozdělení pokud je $\mu = 0$ a $\sigma = 1$
- Tradičně se standardní normální NV značí Z
- Hustota a distribuční funkce standardní NV se značí $\varphi(z)$ a $\Phi(z)$



Vlastnosti standardní NV

1. Jestliže $X \sim N(\mu, \sigma^2)$, pak $Z = \dfrac{(X - \mu)}{\sigma} \sim N(0, 1)$.
2. Jestliže $Z \sim N(0, 1)$, pak $X = \mu + \sigma Z \sim N(\mu, \sigma^2)$.
3. Jestliže $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, n$ jsou nezávislé, pak

$$\sum_{i=1}^{n} X_i \sim N\Big(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\Big)$$

Pokud je $X \sim N(\mu, \sigma^2)$, pak

$$\mathbb{P}(a < X < b) = \mathbb{P}\left(\frac{a - \mu}{\sigma} < Z < \frac{b - \mu}{\sigma}\right) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

- $\Phi(-x) = 1 - \Phi(x)$

Příklad 74.

Nechť $X \sim N(3,5)$. Určete $\mathbb{P}(X > 1)$.

$$\mathbb{P}(X > 1) = 1 - \mathbb{P}(X < 1) = 1 - \mathbb{P}(Z < \frac{1-3}{\sqrt{5}}) = 1 - \Phi(-0.8944) = 0.81$$

Určete $q = \Phi^{-1}(0.2)$.
Musíme najít q t.ž. $\mathbb{P}(X < q) = 0.2$.

$$0.2 = \mathbb{P}(X < q) = \mathbb{P}\left(Z < \frac{q-\mu}{\sigma}\right) = \Phi\left(\frac{q-\mu}{\sigma}\right)$$

Z tabulek $\Phi(-0.8416) = 0.2$
Proto $-0.8416 = \dfrac{q-3}{\sqrt{5}}$, tj. $q = 1.1181$.

Příklad 75.

Nechť $X \sim N(3,9)$. Určete $\mathbb{P}(2 < X < 5)$, $\mathbb{P}(X > 0)$ a $\mathbb{P}(|X-3| > 6)$.

Řešení:

$$\begin{aligned}\mathbb{P}(2 < X < 5) &= \mathbb{P}\left(\frac{2-3}{3} < \frac{X-3}{3} < \frac{5-3}{3}\right) = \mathbb{P}\left(-\frac{1}{3} < Z < \frac{2}{3}\right) = \Phi\left(\frac{2}{3}\right) - \Phi\left(-\frac{1}{3}\right) \\ &= \Phi\left(\frac{2}{3}\right) - \left(1 - \Phi\left(\frac{1}{3}\right)\right) \approx .3779\end{aligned}$$

$$\mathbb{P}(X > 0) = \mathbb{P}\left(\frac{X-3}{3} > \frac{0-3}{3}\right) = \mathbb{P}\left(Z > -1\right) = 1 - \Phi\left(-1\right) = \Phi\left(1\right) \approx .8413$$

$$\begin{aligned}\mathbb{P}(|X-3| > 6) &= \mathbb{P}\left(X > 9\right) + \mathbb{P}\left(X < -3\right) = \mathbb{P}\left(Z > 2\right) + \mathbb{P}\left(Z < -2\right) \\ &= 1 - \Phi\left(2\right) + \Phi\left(-2\right) = 2(1 - \Phi\left(2\right)) \approx .0456\end{aligned}$$

Exponenciální rozdělení

NV X má exponenciální rozdělení s parametrem β, $X \sim Exp(\beta)$, jestliže

$$f(x) = \begin{cases} \frac{1}{\beta} e^{-\frac{x}{\beta}} & x > 0 \end{cases}$$

kde $\beta > 0$.

Exponenciální rozdělení se používá na modelévání životnosti elektronických počítačů či doby čekání mezi vzácnými jevy (životnost, doba do poruchy):

- doba mezi nehodami na jisté křižovatce
- doba čekání ve frontě

Příklad 76.

Předpokládejme, že délka telefonního hovoru v minutách je exponencíální NV s parametrem $\beta = 10$. Pokud někdo přijde do veřejné telefonní budky těsně před vámi, jaká je pravděpodobnost, že budete čekat
(a) více jak 10 minut
(b) mezi 10 a 20 minutami.

Řešení: Nechť X značí délku hovoru člověka v budce před vámi. Pak

$$\mathbb{P}(X > 10) = 1 - \mathbb{P}(X \leq 10) = 1 - \int_0^{10} \frac{1}{10} e^{-\frac{x}{10}} dx = 1 - (1 - e^{-1}) = e^{-1} \approx .368$$

$$\mathbb{P}(10 < X < 20) = \text{Dú} = e^{-1} - e^{-2} \approx .233$$

Gamma rozdělení

Pro $\alpha > 0$ je Gamma funkce definována jako

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y}\, dy$$

NV X má Gamma rozdělení s parametry α a β, $X \sim Gamma(\alpha, \beta)$, jestliže

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{\frac{-x}{\beta}} \quad x > 0$$

kde $\alpha, \beta > 0$.

Exponenciální rozdělení je $Gamma(1, \beta)$.

Pro $X_i \sim Gamma(\alpha_i, \beta)$ nezávislé je $\sum_{i=1}^n X_i \sim Gamma(\sum_{i=1}^n \alpha_i, \beta)$.

Gamma rozdělení se často používá k modelování doby trvání, např. při testování životnosti výrobku jde o dobu do "smrti" výrobku.

Beta rozdělení

NV X má Beta rozdělení s parametry $\alpha > 0$ a $\beta > 0$, $X \sim Beta(\alpha, \beta)$, jestliže

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \qquad 0 < x < 1.$$

Beta rozdělení se používá k modelování chování NV omezených na intervaly konečné délky.

Beta rozdělení je vhodný model pro náhodné chování procent a podílů.

Studentovo t rozdělení a Cauchyho rozdělení

NV X má t rozdělení s ν stupni volnosti, $X \sim t_\nu$, jestliže

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{(1 + \frac{x^2}{\nu})^{\frac{1+\nu}{2}}}$$

Normální rozdělení odpovídá t rozdělení s $\nu = \infty$.

Cauchyho rozdělení je speciální případ t rozdělení pro $\nu = 1$. Hustota je

$$f(x) = \frac{1}{\pi(1 + x)^2}$$

χ^2 rozdělení

NV X má χ^2 rozdělení s p stupni volnosti, $X \sim \chi^2_p$, jestliže

$$f(x) = \frac{1}{\Gamma(\frac{p}{2})2^{\frac{p}{2}}} x^{\frac{p}{2}-1} e^{-\frac{x}{2}} \qquad x > 0$$

Jestliže $Z_1, \ldots, Z_p$ jsou nezávislé std. normální NV, pak $\sum_{i=1}^{p} Z_i^2 \sim \chi^2_p$.

Též Pearsonovo rozdělení. Využíváno ve statistice. Má velký význam pro určování, zda množina dat vyhovuje dané distribuční funkci.

Kvantilová funkce

Kvantily

Kvantily jsou užitečný nástroj. Používají se k sumarizace skupiny čísel.
Intuitivně, kvantil znamená, že vzorek je rozdělen na několik stejně velkých částí.

Definice 77 (Kvantil řádu n).

Nechť $y_1 \leq v_2 \leq \cdots \leq y_N$ je uspořádaný statistický soubor. Definujme číslo

$$r = \left\lfloor j\frac{N}{n} + 1 \right\rfloor$$

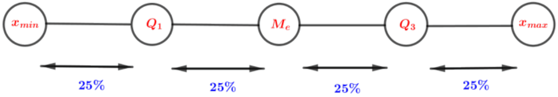
Kvantily řádu n jsou hodnoty $K_1, \ldots, K_n$ počítané následovně:

$$K_j = \begin{cases} y_r & \text{pokud } j\frac{N}{n} \notin \mathbb{N} \\ \frac{y_{r-1}+y_2}{2} & \text{pokud } j\frac{N}{n} \in \mathbb{N} \end{cases}$$

Kvantily řádu n definují n intervalů $(y_1, K_1], [K_1, K_2], \ldots, [K_{n-1}, y_N)$, kde v každém intervalu je nejvýše $\frac{100}{n}$ % hodnot souboru.

Speciální typy kvantilů

- Kvantil řádu 2 je medián
- Kvantily řádu 4 jsou kvartily
- Kvantily řádu 10 jsou decily
- Kvantily řádu 100 jsou percentily



Jinými slovy, kvartily jsou 3 čísla, která dělí soubor na 4 stejně velké částí, decily 9 čísel dělící soubor na 10 stejně velkých částí a percentily 99 čísel dělící soubor na 100 stejně velkých částí.
Q_1 je dolní kvartil a Q_3 je horní kvartil.
Kvartily jsou speciálním případem percentilů. 25-tý percentil je první kvartil, 50-tý percentil je medián a 75-tý percentil je třetí kvartil.
Např. 60-tý percentil znamená, že číslo je větší než 60 % ostatních čísel v souboru.

Příklady

Příklad 78 (Příklad 1).

Najděte kvartily následujícího souboru: $-1, -3, 0, -1, -1, 5, 0, -3, 1, 2, 3, 3.$

Řešení: Nejprve uspořádáme: $-3, -3, -1, -1, -1, 0, 0, 1, 2, 3, 3, 5.$
Pak $12/4 = 2$, $2 \cdot 12/4 = 6$ a $3 \cdot 12/4 = 9$, z čehož dostáváme

- $Q_1 = \frac{y_3+y_4}{2} = -1$
- $Q_2 = \frac{y_6+y_7}{2} = 0$
- $Q_3 = \frac{y_9+y_{10}}{2} = 2.5$

Příklady

Příklad 79 (Příklad 2).

Najděte decily D_1, D_3 a D_8 souboru $22, 20, 24, 30, 32, 28, 35$.

Řešení: Nejprve uspořádáme: $20, 22, 24, 28, 30, 32, 35$.
Pak

- $7/10 = 0,7$, tj. $r = \lfloor 0,7 \rfloor + 1 = 1$ a $D_1 = y_1 = 20$
- $3 \cdot 7/10 = 2,1$, tj. $r = 3$ a $D_3 = y_3 = 24$
- $8 \cdot 7/10 = 5,6$, tj. $r = 6$ a $D_8 = y_6 = 32$

Inverzní distribuční funkce či kvantilová funkce

Definice 80.

Nechť $X\colon \Omega \to \mathbb{R}$ je NV s distribuční funkcí $F\colon \mathbb{R} \to [0,1]$. Inverzní distribuční funkce či kvantilová funkce je definována jako

$$F^{-1}(q) = \inf\{x \in \mathbb{R} \mid F(x) > q\}$$

pro $q \in [0,1]$.

Pokud je F striktně rostoucí a spojitá, pak je $F^{-1}(q)$ jediné reálné číslo x t.ž. $F(x) = q$.

Některé speciální hodnoty:

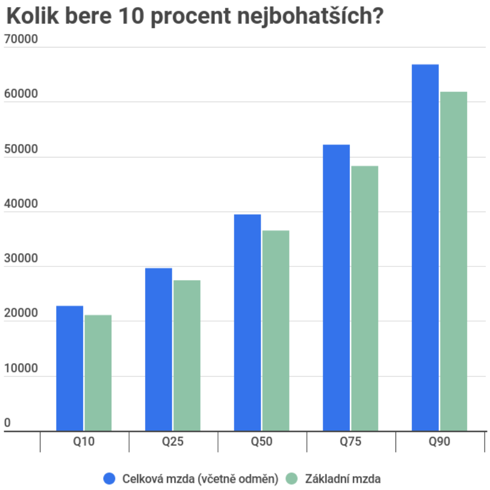
- $Q25 = F^{-1}(1/4)$ se nazývá první kvartil
- $Q50 = F^{-1}(1/2)$ je medián či druhý kvartil
- $Q75 = F^{-1}(3/4)$ je třetí kvartil

Příklad

Salary (in €)	Percent p_i100%	Cumulative percentage	Width of the class
499.5 - 700.5	0.1	0.1	200
700.5 - 900.5	0.2	0.3	200
900.5 - 1100.5	2.6	2.9	200
1100.5 - 1300.5	6.5	9.4	200
1300.5 - 1500.5	12.3	21.7	200
1500.5 - 1700.5	16.5	38.2	200
1700.5 - 1900.5	23.8	62.0	200
1900.5 - 2100.5	14.9	76.9	200
2100.5 - 2300.5	11.1	88.0	200
2300.5 - 2500.5	7.0	95.0	200
2500.5 - 3000.5	4.2	99.2	500
3000.5 - 4000.5	0.8	100.00	1000

- $Q25 = F^{-1}(1/4) = 1500.5$
- $Q50 = F^{-1}(1/2) = 1700.5$
- $Q75 = F^{-1}(3/4) = 1900.5$

Příklad (převzato z Aktuálně.cz)



- Decily a kvartily (Q10, Q25, Q50, Q75, Q90)
  - Decil a kvartil dělí statistický soubor na desetiny a čtvrtiny.
  - Q25 znamená, že 75 procent Čechů má vyšší mzdu, než je číslo uvedené v grafu.
  - Např. modře označený údaj Q10 uvádí, že 90 procent Čechů má vyšší celkovou hrubou mzdu (včetně odměn) než 22.762,- Kč.
- Databáze obsahuje údaje od 47.000 lidí.

- Dvě náhodné veličiny X a Y jsou si rovné v rozdělení,

$$X \stackrel{d}{=} Y$$

pokud je

$$F_X(x) = F_Y(x)$$

pro všechna x.

- To neznamená, že X a Y si jsou rovny, ale že pravděpodobnosti tvrzení o X a Y jsou stejné!

Příklad 81.

- Nechť $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = 1/2$ a nechť $Y = -X$.
- Pak $\mathbb{P}(Y = 1) = \mathbb{P}(Y = -1) = 1/2$, a tedy $X \stackrel{d}{=} Y$
- Očividně se X a Y nerovnají, $\mathbb{P}(X = Y) = 0$

Sdružená rozdělení

Sdružená rozdělení

- Pro dvě diskrétní NV X a Y definujeme sdruženou pravděpodobnostní funkci

$$f(x, y) = \mathbb{P}(X = x \text{ a } Y = y)$$

- $\mathbb{P}(X = x \text{ a } Y = y)$ budeme stručně zapisovat $\mathbb{P}(X = x, Y = y)$
- Pokud budeme chtít specifikovat NV, budeme psát $f_{X,Y}$

Příklad 82.

Mějme sdružené rozdělení NV X a Y, kde každá NV nabývá hodnot 0 nebo 1:

	$Y=0$	$Y=1$	
$X=0$	1/9	2/9	1/3
$X=1$	2/9	4/9	2/3
	1/3	2/3	1

Pak např. $f(1,1) = \mathbb{P}(X=1, Y=1) = 4/9$.

Sdružená funkce hustoty

Definice 83.

Ve spojitém případě je $f(x, y)$ hustota sdružené NV (X, Y), jestliže

1. $f(x, y) \geq 0$ pro všechna (x, y)
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dx \, dy = 1$
3. pro lib. $A \subset \mathbb{R} \times \mathbb{R}$ je $\mathbb{P}((X, Y) \in A) = \iint_A f(x, y) \, dx \, dy$

V diskrétním i spojitém případě je sdružená distribuční fce definována jako

$$F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y)$$

Příklad 84.

Nechť sdružená NV (X, Y) má rovnoměrné rozdělení na jednotkovém čtverci. Pak

$$f(x,y) = \begin{cases} 1 & 0 \leq x, y \leq 1 \\ 0 & \text{jinak} \end{cases}$$

Určete $\mathbb{P}(X < 1/2, Y < 1/2)$.

Řešení: Jev $A = \{X < 1/2, Y < 1/2\}$ odpovídá podmnožině jednotkového čtverce.

Integrace fce f přes A odpovídá obsahu A, který je $1/4$, tj.

$$\mathbb{P}(X < 1/2, Y < 1/2) = 1/4$$

Příklad 85.

Nechť (X, Y) má hustotu

$$f(x,y) = \begin{cases} x+y & 0 \le x, y \le 1 \\ 0 & \text{jinak} \end{cases}$$

Pak

$$\int_0^1 \int_0^1 (x+y)\,dx\,dy = \int_0^1 \frac{1}{2}\,dy + \int_0^1 \frac{1}{2}\,dx = 1$$

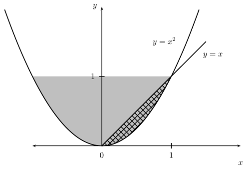
tj, f je skutečně hustota.

Pozn.: $\iint\limits_I (f(x,y) + g(x,y))\,dx\,dy = \iint\limits_I f(x,y)\,dx\,dy + \iint\limits_I g(x,y)\,dx\,dy$

Příklad 86.

Nechť (X, Y) má hustotu $f(x, y) = \begin{cases} cx^2y & x^2 \le y \le 1 \\ 0 & \text{jinak} \end{cases}$, tj. $-1 \le x \le 1$. Určete c.

Řešení: Nechme x běhat přes definiční obor a pro každou hodnotu x nechme y běhat přes svůj definiční obor, tj. $x^2 \le y \le 1$, viz obrázek.



Pak $1 = \iint f(x, y)\, dy\, dx = c \int_{-1}^{1} \int_{x^2}^{1} x^2 y\, dy\, dx = \dfrac{4c}{21}$, a tedy $c = 21/4$.

Určeme $\mathbb{P}(X \ge Y)$, tj. $A = \{(x, y) \mid 0 \le x \le 1, x^2 \le y \le x\}$.

$$\mathbb{P}(X \ge Y) = \frac{21}{4} \int_0^1 \int_{x^2}^{x} x^2 y\, dy\, dx = \frac{3}{20}$$

Marginální rozdělení

Marginální rozdělení

Definice 87.

Jestliže sdružená NV (X, Y) má sdružené rozdělení s pravděpodobnostní funkcí $f_{X,Y}$, pak marginální pravděpodobnost X je

$$f_X(x) = \mathbb{P}(X = x) = \sum_y \mathbb{P}(X = x, Y = y) = \sum_y f(x, y)$$

a podobně marginální pravděpodobnost Y je

$$f_Y(y) = \mathbb{P}(Y = y) = \sum_x \mathbb{P}(X = x, Y = y) = \sum_x f(x, y)$$

Příklad 88.

Nechť $f_{X,Y}$ je dána tabulkou

	$Y = 0$	$Y = 1$	
$X = 0$	1/10	2/10	3/10
$X = 1$	3/10	4/10	7/10
	4/10	6/10	1

Pak

- marginální rozdělení X odpovídá sumě řádků a
- marginální rozdělení Y sumě sloupců.

Např., $f_X(0) = 3/10$ a $f_X(1) = 7/10$.

Definice 89.

Pro spojité NV je marginální hustota

$$f_X(x) = \int f(x,y)\,dy \qquad \text{a} \qquad f_Y(y) = \int f(x,y)\,dx$$

Marginální distribuční funkce se značí F_X a F_Y.

Příklad 90.

Nechť

$$f_{X,Y}(x,y) = e^{-(x+y)}$$

$x, y \geq 0$. Pak

$$f_X(x) = e^{-x} \int_0^\infty e^{-y}\,dy = e^{-x}$$

Příklad 91.

Nechť

$$f(x,y) = \begin{cases} x+y & 0 \leq x, y \leq 1 \\ 0 & \text{jinak} \end{cases}$$

Pak

$$f_Y(y) = \int_0^1 (x+y)\,dx = \frac{1}{2} + y$$

Příklad 92.

Nechť (X, Y) má hustotu

$$f(x,y) = \begin{cases} \frac{21}{4}x^2y & x^2 \leq y \leq 1 \\ 0 & \text{jinak} \end{cases}$$

Pak

$$f_X(x) = \begin{cases} \int f(x,y)\,dy = \frac{21}{4}x^2 \int_{x^2}^{1} y\,dy = \frac{21}{8}x^2(1-x^4) & \text{pro } -1 \leq x \leq 1 \\ \\ 0 & \text{jinak} \end{cases}$$

Nezávislé náhodné veličiny

Nezávislé náhodné veličiny

Definice 93.

Dvě náhodné veličiny X a Y jsou nezávislé, jestliže pro každé A a B platí

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \cdot \mathbb{P}(Y \in B)$$

Ověřit, zda X a Y jsou nezávislé, znamená ověřit podmínku pro všechny podmnožiny A a B. Následující trzení dává zjednodušení.

Věta 94.

Nechť X a Y mají sdruženou pravděpodobnost (hustotu) $f_{X,Y}$. Pak X a Y jsou nezávislé, jestliže

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$$

pro všechny hodnoty x a y.[4]

[4] Tvrzení není úplně přesné, hustota je definována pouze pro množiny nenulové míry.

Příklad 95.

Nechť X a Y mají následující rozdělení

	$Y = 0$	$Y = 1$	
$X = 0$	1/4	1/4	1/2
$X = 1$	1/4	1/4	1/2
	1/2	1/2	1

Pak $f_X(0) = f_X(1) = 1/2$ a $f_Y(0) = f_Y(1) = 1/2$.

X a Y jsou nezávislé, protože

- $f_X(0)f_Y(0) = f(0,0)$
- $f_X(0)f_Y(1) = f(0,1)$
- $f_X(1)f_Y(0) = f(1,0)$
- $f_X(1)f_Y(1) = f(1,1)$

Příklad 96.

Pokud by X a Y měly následující rozdělení

	$Y=0$	$Y=1$	
$X=0$	1/2	0	1/2
$X=1$	0	1/2	1/2
	1/2	1/2	1

tak by nebyly nezávislé, protože

$$f_X(0)f_Y(1) = (1/2)(1/2) = 1/4$$

ale

$$f(0,1) = 0\,.$$

Příklad 97.

Nechť X a Y jsou nezávislé a mají stejnou hustotu

$$f(x) = \begin{cases} 2x & 0 \leq x \leq 1 \\ 0 & \text{jinak} \end{cases}$$

Určete $\mathbb{P}(X + Y \leq 1)$.

Řešení: Z nezávislosti máme, že sdružená hustota je

$$f(x,y) = f_X(x) f_Y(y) = \begin{cases} 4xy & 0 \leq x, y \leq 1 \\ 0 & \text{jinak} \end{cases}$$

Pak

$$\mathbb{P}(X + Y \leq 1) = \iint_{x+y \leq 1} f(x,y)\, dy\, dx = 4 \int_0^1 x \left(\int_0^{1-x} y\, dy \right) dx = \frac{1}{6}$$

Věta 98.

Nechť obor hodnot NV X a Y je (nekonečný) obdélník. Pokud $f(x,y) = g(x)h(y)$ pro nějaké funkce g a h (ne nutně hustoty), pak X a Y jsou nezávislé.

Příklad 99.

Nechť (X,Y) má sdruženou hustotu

$$f(x,y) = \begin{cases} 2e^{-(x+2y)} & 0 < x, y \\ 0 & \text{jinak} \end{cases}$$

Obor hodnot X a Y je obdélník $(0,\infty) \times (0,\infty)$ a $f(x,y) = g(x)h(y)$ pro $g(x) = 2e^{-x}$ a $h(y) = e^{-2y}$. Proto jsou X a Y nezávislé.

Podmíněná rozdělení

Podmíněná rozdělení

Jestliže X a Y jsou diskrétní NV, pak můžeme spočítat podmíněné rozdělení X za předpokladu $Y = y$.

$$\mathbb{P}(X = x|Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}$$

Definice 100.

Podmíněná pravděpodobnostní funkce je fce

$$f_{X|Y}(x|y) = \mathbb{P}(X = x|Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

pro $f_Y(y) > 0$.

Definice platí i pro spojitá rozdělení, interpretace se však liší.
V diskrétním případě je $f_{X|Y}(x|y) = \mathbb{P}(X = x|Y = y)$.
Ve spojitém musíme integrovat, abychom dostali pravděpodobnost.[5]

Definice 101.

Pro spojitou NV je podmíněná hustata definována jako

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

pro $f_Y(y) > 0$, tj.

$$\mathbb{P}(X \in A|Y = y) = \int_A f_{X|Y}(x|y)\,dx$$

[5]Ve spojitém případě podmiňujeme $\mathbb{P}(X \in A|Y = y)$ jevem $\{Y = y\}$, který má nulovou pravděpodobnost. Tomu se vyhneme pomocí hustoty. To, že jde o dobře definovanou teorii viz R.B. Ash, Basic Probability Theory

Příklad 102.

Nechť X a Y mají sdružené rovnoměrné rozdělení na jednotkovém čtverci. Pak

$$f_{X|Y}(x|y) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{jinak.} \end{cases}$$

Pro $Y = y$ tak má X rovnoměrné rozdělení na $(0, 1)$.

To lze zapsat jako $(X|Y = y) \sim Uniform(0, 1)$.

Příklad 103.

Nechť

$$f(x,y) = \begin{cases} x + y & 0 \leq x, y \leq 1 \\ 0 & \text{jinak} \end{cases}$$

Určeme $\mathbb{P}(X < 1/4 | Y = 1/3)$.

Řešení: Již jsme si ukázali, že $f_Y(y) = y + 1/2$, a proto

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{x+y}{y+\frac{1}{2}}$$

Tedy

$$\mathbb{P}(x < 1/4 | Y = 1/3) = \int_0^{1/4} f_{X|Y}(x|1/3)\, dx = \int_0^{1/4} \frac{x + \frac{1}{3}}{\frac{1}{3} + \frac{1}{2}}\, dx = \frac{11}{80}$$

Příklad 104.

Nechť $X \sim Uniform(0,1)$ a nechť po obdržení hodnoty X dostaneme $(Y|X = x) \sim Uniform(x,1)$. Jaké je marginální rozdělení Y?

Řešení: Předně $f_X(x) = 1$ pro $0 \leq x \leq 1$, 0 jinak, a

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{1-x} & 0 < x < y < 1 \\ 0 & \text{jinak} \end{cases}$$

a tedy $f_{X,Y}(x,y) = f_{Y|X}(y|x) f_X(x)$.

Marginální rozdělení Y je

$$f_Y(y) = \int_0^y f_{X,Y}(x,y)\,dx = \int_0^y \frac{dx}{1-x} = -\ln(1-y)$$

pro $0 < y < 1$.

Příklad 105.

Nechť (X, Y) má hustotu

$$f(x,y) = \begin{cases} \frac{21}{4}x^2 y & x^2 \leq y \leq 1 \\ 0 & \text{jinak} \end{cases}$$

Určete $f_{Y|X}(y|x)$ a $\mathbb{P}(Y \geq 3/4 | X = 1/2)$.

Řešení: Pro $X = x$ musí y splňovat $x^2 \leq y \leq 1$. Víme, že $f_X(x) = (21/8)x^2(1 - x^4)$, a tedy pro $x^2 \leq y \leq 1$ platí

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{\frac{21}{4}x^2 y}{\frac{21}{8}x^2(1 - x^4)} = \frac{2y}{1 - x^4}$$

Pak

$$\mathbb{P}(Y \geq 3/4 | X = 1/2) = \int_{3/4}^{1} f_{Y|X}(y|1/2)\, dy = \int_{3/4}^{1} \frac{32y}{15}\, dy = \frac{7}{15}$$

Náhodné vektory

Náhodné vektory

Nechť $X = (X_1, \dots, X_n)$, kde $X_1, \dots, X_n$ jsou náhodné veličiny. Pak X je náhodný vektor.

Nechť $f(x_1, \dots, x_n)$ je hustota náhodného vektoru X. Pak je možné definovat marginální a podmíněné pravděpodobnosti podobně jako ve sdruženém případě.

Řekneme, že $X_1, \dots, X_n$ jsou nezávislé, jestliže pro každé $A_1, \dots, A_n$ je

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i)$$

Opět stačí ověřit, že $f(x_1, \dots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$.

Definice 106.

Jestliže $X_1, \ldots, X_n$ jsou nezávislé NV a každá má stejné marginální rozdělení s distribuční funkcí F, pak říkáme, že $X_1, \ldots, X_n$ jsou IID (independent a identically distributed) a píšeme

$$X_1, \ldots, X_n \sim F$$

Jestliže F má hustotu f, píšeme též $X_1, \ldots, X_n \sim f$.

$X_1, \ldots, X_n$ též nazýváme náhodný výběr velikosti n z F.

Dvě důležitá rozdělení náhodných vektorů

Multinomiální rozdělení

Uvažme losování míčků z urny, která obsahuje míčky k různých barev $(c_1, \ldots, c_k)$. Nechť $p = (p_1, \ldots, p_k)$, $p_j \geq 0$ a $\sum_{j=1}^{n} p_j = 1$, kde p_j je pravděpodobnost vytažení míčku s barvou c_j. Opakujme losování n krát (nezávislé tahy s opakováním) a označme $X = (X_1, \ldots, X_k)$, kde X_j je počet výskytu barvy c_j, tj. $n = \sum_{j=1}^{k} X_j$. Pak X má multinomiální rozdělení s parametry (n, p), psáno $X \sim Multinomial(n, p)$.

Pravděpodobnostní funkce je

$$f(x) = \binom{n}{x_1, \ldots, x_k} p_1^{x_1} \cdots p_k^{x_k} \qquad \text{kde } \binom{n}{x_1, \ldots, x_k} = \frac{n!}{x_1! \cdots x_k!}$$

Lemma 107.

Nechť $X \sim Multinomial(n, p)$, *kde* $X = (X_1, \ldots, X_k)$ *a* $p = (p_1, \ldots, p_k)$.
Pak marginální rozdělení X_j *je* $Binomial(n, p_j)$.

Vícerozměrné normální rozdělení

Normální rozdělení má dva parametry, μ a σ. Ve vícerozměrné verzi je μ vektor a σ matice Σ. Nechť

$$Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_k \end{pmatrix}$$

kde $Z_1, \ldots, Z_k \sim N(0,1)$ jsou nezávislé. Hustota Z pak je

$$f(z) = \prod_{i=1}^{k} f(z_i) = \frac{1}{(2\pi)^{k/2}} e^{-\frac{1}{2}\sum_{j=1}^{k} z_j^2} = \frac{1}{(2\pi)^{k/2}} e^{-\frac{1}{2} z^T z}$$

a Z má standardní vícerozměrné normální rozdělení, $Z \sim N(0, I)$, kde 0 reprezentuje nulový vektor a I jednotkovou matici.

Obecně, vektor X má vícerozměrné normální rozdělení, $X \sim N(\mu, \Sigma)$, pokud má hustotu

$$f(x; \mu, \Sigma) = \frac{1}{(2\pi)^{k/2}} |\Sigma|^{1/2} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}$$

kde

- $|\Sigma|$ značí determinant Σ
- μ je vektor délky k
- Σ je symetrická pozitivně definitní matice typu $k \times k$.[6]

Pro $\mu = 0$ a $\Sigma = I$ dostáváme standarní normální rozdělení.

[6] Σ je pozitivně definitní, pokud pro všechny nenulové vektory x je $x^T \Sigma x > 0$.

Protože je Σ symetrická a pozitivně definitní, existuje matice $\Sigma^{1/2}$ tzv. odmocnina Σ t.ž.

1. $\Sigma^{1/2}$ je symetrická
2. $\Sigma = \Sigma^{1/2}\Sigma^{1/2}$
3. $\Sigma^{1/2}\Sigma^{-1/2} = \Sigma^{-1/2}\Sigma^{1/2} = I$

Věta 108.

Jestliže $Z \sim N(0, I)$ *a* $X = \mu + \Sigma^{1/2}Z$, *pak* $X \sim N(\mu, \Sigma)$.
Naopak, jestliže $X \sim N(\mu, \Sigma)$, *pak* $\Sigma^{-1/2}(X - \mu) \sim N(0, I)$.

Nechť náhodný normální vektor X lze rozdělit jako $X = (X_a, X_b)$ a podobně $\mu = (\mu_a, \mu_b)$ a

$$\Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}$$

Věta 109.

Nechť $X \sim N(\mu, \Sigma)$. Pak

1. *marginální rozdělení X_a je $N(\mu_a, \Sigma_{aa})$.*
2. *podmíněné rozdělení X_b za předpokladu $X_a = x_a$ je*

$$X_b | X_a = x_a \sim N(\mu_b + \Sigma_{ba}\Sigma_{aa}^{-1}(x_a - \mu_a), \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})$$

3. *jestliže a je vektor, pak $a^T X \sim N(a^T \mu, a^T \Sigma a)$*
4. $V = (X - \mu)^T \Sigma^{-1} (X - \mu) \sim \chi_k^2$

Transformace náhodných veličin

Transformace náhodných veličin

- Nechť X je náhodná veličina s hustotou f_X a distribuční funkcí F_X.
- Nechť $Y = r(X)$ je funkce X, pak $Y = r(X)$ se nazývá transformace X.
  - např. $Y = X^2$ nebo $Y = e^X$
- Jak určit hustotu a distribuční funkci Y?
- V diskrétním případě je pravděpodobnost Y dána jako

$$f_Y(y) = \mathbb{P}(Y = y) = \mathbb{P}(r(X) = y) = \mathbb{P}(\{x \mid r(x) = y\}) = \mathbb{P}(X \in r^{-1}(y)).$$

Příklad 110.

Nechť $\mathbb{P}(X = -1) = \mathbb{P}(X = 1) = 1/4$ a $\mathbb{P}(X = 0) = 1/2$. Nechť $Y = X^2$. Pak $\mathbb{P}(Y = 0) = \mathbb{P}(X = 0) = 1/2$ a $\mathbb{P}(Y = 1) = \mathbb{P}(X = 1) + \mathbb{P}(X = -1) = 1/2$. Dohromady

x	$f_X(x)$
-1	1/4
0	1/2
1	1/4

$\rightsquigarrow$

$y = x^2$	$f_X(x)$
1	1/4
0	1/2
1	1/4

$\rightsquigarrow$

y	$f_Y(y)$
0	1/2
1	1/2

Transformace náhodných veličin

Příklad 111.

Nechť X je diskrétní NV zadaná tabulkou

x	0	$\pi/4$	$\pi/2$	$3\pi/4$	π
$f_X(x)$	0.1	0.3	0.2	0.1	0.3

Určete rozdělení NV $Y = \sin(X)$.

x	0	$\pi/4$	$\pi/2$	$3\pi/4$	π
$sin(x)$	0	$\sqrt{2}/2$	1	$\sqrt{2}/2$	0
$f_X(x)$	0.1	0.3	0.2	0.1	0.3

A tedy

$y = sin(x)$	0	$\sqrt{2}/2$	1
$f_X(x)$	0.4	0.4	0.2

Kroky transformace pro spojitý případ:

1. Pro každé y určíme množinu $A_y = \{x \mid r(x) \leq y\}$.
2. Určíme distribuční funkci

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(r(X) \leq y) = \mathbb{P}(\{x \mid r(x) \leq y\}) = \int_{A_y} f_X(x)\,dx$$

3. Hustota pak je $f_Y(y) = F'_Y(y)$.

Pozn. Po transformaci se může rozdělení změnit (i ze spojité na diskrétní, ne naopak).

Příklad 112.

Nechť $f_X(x) = e^{-x}$ pro $x > 0$. Pak $F_X(x) = \int_0^x f_X(s)\,ds = 1 - e^{-x}$.

Nechť $Y = r(X) = \ln(X)$. Pak $A_y = \{x \mid x \leq e^y\}$ a

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(\ln(X) \leq y) = \mathbb{P}(X \leq e^y) = F_X(e^y) = 1 - e^{-e^y}$$

a tedy $f_Y(y) = e^y e^{-e^y}$ pro $y \in \mathbb{R}$.

Příklad 113.

Nechť $X \sim Uniform(-1,3)$, tj. $f_X(x) = \begin{cases} 1/4 & -1 \leq x \leq 3 \\ 0 & \text{jinak} \end{cases}$

Určeme hustotu NV $Y = X^2$; Y nabývá hodnot $(0,9)$.

- Pro $0 < y < 1$, $A_y = [-\sqrt{y}, \sqrt{y}]$ a $F_Y(y) = \int_{A_y} f_X(x)\,dx = (1/2)\sqrt{y}$.
- $1 \leq y < 9$, $A_y = [-1, \sqrt{y}]$ a $F_Y(y) = \int_{A_y} f_X(x)\,dx = (1/4)(\sqrt{y}+1)$.

Derivováním F dostaneme

$$f_Y(y) = \begin{cases} \frac{1}{4\sqrt{y}} & 0 < y < 1 \\ \frac{1}{8\sqrt{y}} & 1 < y < 9 \\ 0 & \text{jinak} \end{cases}$$

Pokud je r striktně rostoucí nebo striktně klesající, pak má inverzi a platí

$$f_Y(y) = f_X(r^{-1}(y)) \left| \frac{dr^{-1}(y)}{dy} \right|$$

Příklad 114.

Nechť X má hustotu $f_X(x) = \begin{cases} \geq 0 & x > 0 \\ 0 & \text{jinak} \end{cases}$

Určeme hustotu NV $Y = a \cdot \ln(X)$, $a \neq 0$.

$$r(x) = a\ln(x) \rightsquigarrow r^{-1}(y) = e^{\frac{y}{a}} \rightsquigarrow f_Y(y) = f_X(e^{\frac{y}{a}})e^{\frac{y}{a}}\frac{1}{|a|}$$

Pak pro $c > 0$ a $f_X(x) = 1/c$, $0 < x < c$, má NV $Y = -\ln(X)$ hustotu

$$f_Y(y) = \begin{cases} \frac{1}{c}e^{-y} & y > -\ln(c) \\ 0 & \text{jinak} \end{cases}$$

Transformace více náhodných veličin

Pokud jsou X a Y NV, jaké je rozdělení X/Y, $X+Y$, $\max\{X,Y\}$ či $\min\{X,Y\}$?

Nechť $Z = r(X,Y)$.

1. Pro každé z určíme $A_z = \{(x,y) \mid r(x,y) \leq z\}$
2. Určíme distribuční funkci

$$F_Z(z) = \mathbb{P}(Z \leq z) = \mathbb{P}(r(X,Y) \leq z) = \mathbb{P}(A_z) = \iint_{A_z} f_{X,Y}(x,y)\, dx\, dy$$

3. $f_Z(z) = F'_Z(z)$

Příklad I

Nechť $X_1, X_2 \sim Uniform(0,1)$ jsou nezávislé. Určeme hustotu $Y = X_1 + X_2$.

Řešení: Sdružená hustotu (X_1, X_2) je

$$f(x_1, x_2) = \begin{cases} 1 & 0 < x_1, x_2 < 1 \\ 0 & \text{jinak} \end{cases}$$

Označme $r(x_1, x_2) = x_1 + x_2$, pak

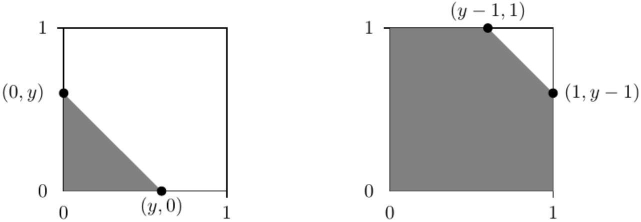
$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) = \mathbb{P}(r(X_1, X_2) \leq y) \\ &= \mathbb{P}(\{(x_1, x_2) \mid r(x_1, x_2) \leq y\}) = \iint_{A_y} f(x_1, x_2)\, dx_1\, dx_2 \end{aligned}$$

Jak najít A_y?

Příklad II

Pro $0 \leq y < 1$ je A_y množina na prvním obrázku o ploše $y^2/2$.

Pro $1 \leq y \leq 2$ je A_y množina na druhém obrázku o ploše $1 - (2-y)^2/2$.



Příklad III

Tedy

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ \frac{y^2}{2} & 0 \leq y < 1 \\ 1 - \frac{(2-y)^2}{2} & 1 \leq y < 2 \\ 1 & y \geq 2 \end{cases}$$

a hustota je

$$f_Y(y) = \begin{cases} y & 0 \leq y \leq 1 \\ 2 - y & 1 \leq y \leq 2 \\ 0 & \text{jinak} \end{cases}$$

Střední hodnota

Definice 115.

Očekávaná či střední hodnota (či první moment) NV X je definována jako

$$\mathbb{E}[X] = \int x\,dF(x) = \begin{cases} \sum_x x f(x) & X \text{ je diskrétní} \\ \int x f(x)\,dx & X \text{ je spojitá} \end{cases}$$

za předpokladu, že daná suma/integrál existuje.

Notace:

$$\mathbb{E}[X] = \mathbb{E}(X) = \mathbb{E}X = \int x\,dF(x) = \mu = \mu_X$$

Pozn.: $\int x\,dF(x)$ je zde pouze notace pro zjednodušení, abychom nemuseli rozlišovat diskrétní a spojitý případ. (V analýze má vlastní význam!)

- Střední hodnota je jednočíselný souhrn rozdělení.
- Uvažujme o $\mathbb{E}(X)$ jako o průměru $\sum_{i=1}^{n} X_i/n$ velkého počtu IID výběrů $X_1, \ldots, X_n$.
- Fakt, že $\mathbb{E}(X) \approx \sum_{i=1}^{n} X_i/n$ je ve skutečnosti věta nazývaná zákon velkých čísel (později).
- $\mathbb{E}(X)$ existuje, jestliže $\int |x| dF_X(x) < \infty$; jinak neexistuje.

Příklad 116.

Nechť $X \sim Bernoulli(p)$, pak

$$\mathbb{E}(X) = \sum_{x=0}^{1} x f(x) = 0(1-p) + 1p = p$$

Příklad 117.

Hod férovou mincí dvakrát. Nechť X je počet orlů. Pak

$$\begin{aligned}\mathbb{E}(X) &= \sum_{x} x f(x) = 0f(0) + 1f(1) + 2f(2) \\ &= 0(1/4) + 1(1/2) + 2(1/4) = 1\end{aligned}$$

Příklad 118.

Nechť $X \sim Uniform(-1, 3)$, pak

$$\mathbb{E}(X) = \int x f(x)\, dx = \frac{1}{4} \int_{-1}^{3} x\, dx = 1$$

Příklad 119.

Náhodná veličina má Cauchyho rozdělení, jestliže má hustotu $f_X(x) = (\pi(1 + x^2))^{-1}$. Integrací per partes ($u = x$ a $v = tan^{-1}x$) dostaneme, že

$$\int |x| dF(x) = \frac{2}{\pi} \int_0^\infty \frac{x\, dx}{1 + x^2} = [x tan^{-1}(x)]_0^\infty - \int_0^\infty tan^{-1}x\, dx = \infty$$

tj. střední hodnota neexistuje.

Kdykoliv dále mluvíme o střední hodnotě, tak předpokládáme, že exituje.

Nechť $Y = r(X)$. Jak určit $\mathbb{E}(Y)$? Lze najít $f_Y(y)$ a spočítat $\mathbb{E}(Y) = \int y f_Y(y)\,dy$. Existuje však jednodušší způsob.

Věta 120 (Pravidlo líného statistika).

Nechť $Y = r(X)$. *Pak*

$$\mathbb{E}(Y) = \mathbb{E}(r(X)) = \int r(x)\,dF_X(x).$$

Intuice: Uvažme hru, kde vybíráme X náhodně a dostanete $Y = r(X)$. Průměrná výhra je $r(x)$ krát pravděpodobnost, že $X = x$, sečteno či zintegrováno přes všechny hodnoty x.

Speciální případ:

Nechť A je jev a $r(x) = I_A(x)$, kde $I_A(x) = 1$ pro $x \in A$ a $I_A(x) = 0$ pro $x \notin A$. Pak

$$\mathbb{E}(I_A(X)) = \int I_A(x) f_X(x)\,dx = \int_A f_X(x)\,dx = \mathbb{P}(X \in A).$$

Tj., pravděpodobnost je speciálním případem střední hodnoty.

Příklad 121.

Nechť $X \sim Uniform(0,1)$. Nechť $Y = r(X) = e^X$. Pak

$$\mathbb{E}(Y) = \int_0^1 e^x f(x)\,dx = \int_0^1 e^x\,dx = e - 1$$

Dú: Vyjádřete $f_Y(y)$ a s její pomocí určete $\mathbb{E}(Y)$ (dostanete $f_Y(y) = 1/y$ pro $1 < y < e$)

Příklad 122.

Mějme jehlu jednotkové délky a zlomme ji v náhodném místě. Nechť Y je délka delší části. Jaká je střední hodnota Y?

Řešení: Pokud X značí bod zlomu, pak $X \sim Uniform(0,1)$ a

$$Y = r(X) = \max\{X, 1 - X\}$$

tj. $r(x) = 1 - x$ pro $0 < x < 1/2$ a $r(x) = x$ pro $1/2 \leq x < 1$. Pak

$$\mathbb{E}(Y) = \int r(x)\, dF(x) = \int_0^{1/2} (1 - x)\, dx + \int_{1/2}^{1} x\, dx = \frac{3}{4}$$

Pro funkce více proměnných to funguje podobně. Pokud $Z = r(X,Y)$, pak

$$\mathbb{E}(Z) = \mathbb{E}(r(X,Y)) = \iint r(x,y)\, dF(x,y)$$

Příklad 123.

Nechť náhodný vektor (X,Y) má sdružené rovnoměrné rozdělení na jednotkovém čtverci a nechť $Z = r(X,Y) = X^2 + Y^2$. Pak

$$\begin{aligned}\mathbb{E}(Z) &= \iint f(x,y)\, dF(x,y) = \int_0^1 \int_0^1 (x^2 + y^2)\, dx\, dy \\ &= \int_0^1 x^2\, dx + \int_0^1 y^2\, dy = \frac{2}{3}\end{aligned}$$

Definice 124.

k-tý moment X je definován jako $\mathbb{E}(X^k)$, pokud $\mathbb{E}(|X|^k) < \infty$.

Věta 125.

Jestliže k-tý moment existuje a $j < k$, pak existuje i j-tý moment.

Důkaz.

$$\begin{aligned}\mathbb{E}(|X|^j) &= \int_{-\infty}^{\infty} |x|^j f_X(x)\,dx = \int_{|x|\le 1} |x|^j f_X(x)\,dx + \int_{|x|>1} |x|^j f_X(x)\,dx \\ &\le \int_{|x|\le 1} f_X(x)\,dx + \int_{|x|>1} |x|^k f_X(x)\,dx \le 1 + \mathbb{E}(|X|^k) < \infty\end{aligned}$$

□

Definice 126.

k-tý centrální moment je definován jako $\mathbb{E}((X - \mathbb{E}(X))^k)$.

Vlastnosti střední hodnoty

Věta 127.

Jestliže $X_1, \ldots, X_n$ *jsou náhodné veličiny a* $a_1, \ldots, a_n$ *jsou konstanty, pak*

$$\mathbb{E}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i \mathbb{E}(X_i)$$

Příklad 128.

Nechť $X \sim Binomial(n, p)$. Jaká je střední hodnota X?

Řešení: Z definice je $\mathbb{E}(X) = \sum_x x f_X(x) = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x}$, což není snadné určit. Místo toho si všimněme, že $X = \sum_{i=1}^{n} X_i$ pro $X_i = 1$ když na i-tý hod padl orel a $X_i = 0$ jinak. Pak

$$\mathbb{E}(X_i) = p \cdot 1 + (1-p) \cdot 0 = p$$

a proto je $\mathbb{E}(X) = \mathbb{E}(\sum_i X_i) = \sum_i \mathbb{E}(X_i) = np$

Věta 129.

Nechť $X_1, \ldots, X_n$ *jsou* *nezávislé* *náhodné veličiny. Pak*

$$\mathbb{E}\left(\prod_{i=1}^{n} X_i\right) = \prod_{i_1}^{n} \mathbb{E}(X_i)$$

Pozn. Pravidlo součtu nevyžaduje nezávislost, pravidlo součinu ano!

Aplikace: Průměrná časová složitost Quicksortu

Algoritmus 1: Quicksort

Input: Seznam n různých čísel $S = \{x_1, \ldots, x_n\}$

Output: Seřazený seznam S

if $|S| \leq 1$ **then return** S

$p \leftarrow$ náhodně (uniformě, rovnoměrně) zvolený prvek S

$S_1 = \{x \in S \mid x < p\}$

$S_2 = \{x \in S \mid x > p\}$

Zavolej Quicksort na S_1 a S_2

return S_1, p, S_2

Pozn. Toto je tzv. Randomized Quicksort; při volbě pivota jako např. prvního prvku seznamu je analýza průměrné časové složitosti v podstatě stejná

Věta 130.

Nechť je pivot p vybírán nezávisle a rovnoměrně ze všech možností. Pak očekávaný počet porovnání dvou čísel pro libovolný vstup je $2n \ln n + O(n)$.

Důkaz

- Nechť $y_1, \ldots, y_n$ jsou hodnoty vstupů $x_1, \ldots, x_n$ seřazené vzestupně.
- Pro $i < j$ je NV X_{ij} rovna 1 pokud jsou y_i a y_j porovnány algoritmem; 0 jinak.
- Celkový počet X porovnání dvou čísel splňuje

$$X = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij}$$

a

$$\mathbb{E}(X) = \mathbb{E}\left(\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij} \right) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbb{E}(X_{ij})$$

Důkaz pokračování

- Jelikož je X_{ij} indikátor, $\mathbb{E}(X_{ij}) = \mathbb{P}(X_{ij} = 1)$.
- Potřebujeme tedy určit pravděpodobnost, že prvky y_i a y_j budou porovnány.
- Prvky y_i a y_j budou porovnány právě tehdy, když y_i či y_j je pivot vybraný z množiny $Y^{ij} = \{y_i, y_{i+1}, \ldots, y_{j-1}, y_j\}$:
  - Pokud je y_i (či y_j) pivot z Y^{ij}, pak y_i a y_j musí být ve stejném podseznamu, a tedy budou porovnány.
  - Pokud ani jedno není vybráno jako pivot, pak y_i a y_j budou rozděleny do dvou různých podseznamů, a tedy nebudou nikdy porovnány.
- Jelikož vybíráme pivoty nezávisle a rovnoměrně z každého podseznamu, má každý prvek Y^{ij} stejnou pravděpodobnost, že bude vybrán jako pivot.
- Tedy pravděpodobnost, že y_i či y_j je vybráno jako pivot z Y^{ij}, tj. pravděpodobnost, že $X_{ij} = 1$, je $2/(j - i + 1)$.

Důkaz pokračování.

- Za použití substituce $k = j - i + 1$ dostáváme

$$\mathbb{E}(X) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1} = \sum_{i=1}^{n-1} \sum_{k=2}^{n-i+1} \frac{2}{k} = \sum_{k=2}^{n} \sum_{i=1}^{n+1-k} \frac{2}{k}$$

$$= \sum_{k=2}^{n} (n+1-k)\frac{2}{k} = \sum_{k=2}^{n} (n+1)\frac{2}{k} - \sum_{k=2}^{n} k\frac{2}{k}$$

$$= \left((n+1)\sum_{k=2}^{n} \frac{2}{k}\right) - 2(n-1)$$

$$= \left(2(n+1)\sum_{k=2}^{n} \frac{1}{k}\right) - 2n + 2 + 2(n+1) - 2(n+1) = (2n+2)\sum_{k=1}^{n} \frac{1}{k} - 4n$$

- Jelikož platí, že $\sum_{k=1}^{n} \frac{1}{k} = \ln n + \Theta(1)$, dostáváme $\mathbb{E}(X) = 2n \ln n + \Theta(n)$.

$\square$

Variance a Kovariance

Variance

Variance (též rozptyl) je charakteristika variability rozdělení pravděpodobnosti náhodné veličiny. Vyjadřuje variabilitu rozdělení souboru náhodných hodnot kolem své střední hodnoty.

Definice 131.

Nechť X je náhodná veličina se střední hodnotou $\mu = \mathbb{E}[X]$.
Variance X (značeno σ^2, σ_X^2 či $Var(X)$) je definována jako

$$Var(X) = \sigma^2 = \mathbb{E}[(X-\mu)^2] = \int (x-\mu)^2 dF(x)$$

za předpokladu, že střední hodnota existuje.

Pozn.: Všimněte si, že nelze použít $\mathbb{E}(X-\mu)$ jako míru rozptylu, neboť $\mathbb{E}(X-\mu) = \mathbb{E}(X) - \mu = \mu - \mu = 0$. Občas se používá $\mathbb{E}|X-\mu|$, častěji však variance.

Směrodatná odchylka

Definice 132.

Směrodatná odchylka náhodné veličiny X se značí jako σ, σ_X či $sd(X)$ a je definovaná jako

$$\sqrt{Var(X)}\,.$$

Pozn.: Směrodatná odchylka vypovídá o tom, nakolik se od sebe navzájem typicky liší jednotlivé případy v souboru zkoumaných hodnot.

- Je-li malá, jsou si prvky souboru většinou navzájem podobné
- Je-li velká, signalizuje to velké vzájemné odlišnosti

Věta 133.

Variance má následující vlastnosti:

1. $Var(X) = \mathbb{E}(X^2) - \mu^2$
2. *Pokud jsou* a *a* b *konstanty, pak* $Var(aX + b) = a^2 Var(X)$
3. *Pokud jsou* $X_1, \ldots, X_n$ *nezávislé a* $a_1, \ldots, a_n$ *konstanty, pak*

$$Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 Var(X_i)$$

Důkaz.

1. $Var(X) = \mathbb{E}[(X-\mu)^2] = \sum_x (x-\mu)^2 f(x) = \sum_x (x^2 - 2\mu x + \mu^2) f(x) = \sum_x x^2 f(x) - 2\mu \sum_x x f(x) + \mu^2 \sum_x f(x) = \mathbb{E}[X^2] - 2\mu \mathbb{E}[X] + \mu^2 = \mathbb{E}[X^2] - \mu^2$.

2. $Var(aX + b) = \mathbb{E}[(aX + b - a\mu - b)^2] = \mathbb{E}[a^2 (X-\mu)^2] = a^2 Var(X)$. □

Příklad 134.

Nechť $X \sim Binomial(n,p)$ a nechť $X = \sum_i X_i$, kde $X_i = 1$ pokud na i-tý hod padne orel, jinak $X_i = 0$. NV X_i jsou nezávislé a $\mathbb{P}(X_i = 1) = p$ a $\mathbb{P}(X_i = 0) = 1 - p$. Již jsme ukázali, že $\mathbb{E}(X_i) = p$. Proto

$$\mathbb{E}(X_i^2) = p \cdot 1^2 + (1-p) \cdot 0^2 = p$$

$$Var(X_i) = \mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2 = p - p^2 = p(1-p)$$

a tedy

$$Var(X) = Var\left(\sum_i X_i\right) = \sum_i Var(X_i) = np(1-p)$$

Pozn.: Všimněte si, že $Var(X) = 0$, pokud $p = 1$ nebo $p = 0$. (Dú: Vidíte proč?)

Jestliže $X_1, \ldots, X_n$ jsou NV, definujeme výběrovou střední hodnotu jako

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

a výběrovou varianci jako

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X_n})^2$$

Věta 135.

Nechť $X_1, \ldots, X_n$ jsou IID a nechť $\mu = \mathbb{E}(X_i)$ a $\sigma^2 = Var(X_i)$. Pak

$$\mathbb{E}(\overline{X_n}) = \mu, \qquad Var(\overline{X_n}) = \frac{\sigma^2}{n} \qquad \text{a} \qquad \mathbb{E}(S_n^2) = \sigma^2$$

Kovariance a korelace

Pokud jsou X a Y náhodné veličiny, pak kovariance a korelace mezi X a Y určuje, jak silná je lineární závislost mezi X a Y.

Definice 136.

Nechť X a Y jsou náhodné veličiny se střední hodnotou μ_X a μ_Y a směrodatnými odchylkami σ_X a σ_Y. Definujme kovarianci mezi X a Y jako

$$Cov(X,Y) = \mathbb{E}((X-\mu_X)(Y-\mu_Y))$$

a korelaci jako

$$\rho = \rho_{X,Y} = \rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

Věta 137.

Kovariance splňuje

$$Cov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

a korelace splňuje

$$-1 \leq \rho(X,Y) \leq 1$$

Pro $Y = aX + b$, kde a, b konstanty, je

$$\rho(X,Y) = \begin{cases} 1 & \text{pro } a > 0 \\ -1 & \text{pro } a < 0 \end{cases}$$

(vidíte, jak se chová korelace při lineární závislosti Y na X?)

Pro X a Y nezávislé je $Cov(X,Y) = \rho = 0$ (srovnejte následující větu s bodem 3 Věty 133). Opak obecně neplatí.

Věta 138.

Nechť X a Y jsou náhodné veličiny, pak

1. $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$
2. $Var(X-Y) = Var(X) + Var(Y) - 2Cov(X,Y)$
3. *Obecně, pro náhodné veličiny $X_1, \dots, X_n$,*

$$Var\left(\sum_i a_i X_i\right) = \sum_i a_i^2 Var(X_i) + 2\sum\sum_{i<j} a_i a_j Cov(X_i, X_j)$$

Střední hodnota a variance důležitých NV

rozdělení	st. hodnota	variance
Bodové v a	a	0
$Bernoulli(p)$	p	$p(1-p)$
$Binomial(n,p)$	np	$np(1-p)$
$Geometric(p)$	$1/p$	$(1-p)/p^2$
$Poisson(\lambda)$	λ	λ
$Uniform(a,b)$	$(a+b)/2$	$(b-a)^2/12$
$Normal(\mu,\sigma^2)$	μ	σ^2
$Exponencial(\beta)$	β	β^2
$Gamma(\alpha,\beta)$	$\alpha\beta$	$\alpha\beta^2$
$Beta(\alpha,\beta)$	$\alpha/(\alpha+\beta)$	$\alpha\beta/((\alpha+\beta)^2(\alpha+\beta+1))$
t_ν	0 pro $\nu>1$	$\nu/(\nu-2)$ pro $\nu>2$
χ^2_p	p	$2p$
$Multinomial(n,p)$	np	viz dále
$Multivariate\,Normal(\mu,\Sigma)$	μ	Σ

Nechť $X = (X_1, \ldots, X_k)^T$ je náhodný vektor.

Střední hodnota náhodného vektoru X je definována jako

$$\mu = (\mu_1 \ \ldots, \mu_k)^T = \begin{pmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_k) \end{pmatrix}$$

Variančně-kovarianční matice Σ je definována jako

$$Var(X) = \begin{pmatrix} Var(X_1) & Cov(X_1, X_2) & \cdots & Cov(X_1, X_k) \\ Cov(X_2, X_1) & Var(X_2) & \cdots & Cov(X_2, X_k) \\ \vdots & \vdots & \vdots & \vdots \\ Cov(X_k, X_1) & Cov(X_k, X_2) & \cdots & Var(X_k) \end{pmatrix}$$

Příklad 139.

Pro $X \sim Multinomial(n,p)$ je

$$Var(X) = \begin{pmatrix} np_1(1-p_1) & -np_1p_2 & \cdots & -np_1p_k \\ -np_2p_1 & np_2(1-p_2) & \cdots & -np_2p_k \\ \vdots & \vdots & \vdots & \vdots \\ -np_kp_1 & -np_kp_2 & \cdots & np_k(1-p_k) \end{pmatrix}$$

- Máme $X_i \sim Binomial(n, p_i)$, tj. $\mathbb{E}(X_i) = np_i$ a $Var(X_i) = np_i(1-p_i)$.
- Z $X_i + X_j \sim Binomial(n, p_i + p_j)$ máme
  - $Var(X_i + X_j) = n(p_i + p_j)(1 - [p_i + p_j])$
- Z $Var(X_i + X_j) = Var(X_i) + Var(X_j) + 2Cov(X_i, X_j)$ máme
  - $Var(X_i + X_j) = np_i(1-p_i) + np_j(1-p_j) + 2Cov(X_i, X_j)$
- Z výše uvedeného dostáváme $Cov(X_i, X_j) = -np_ip_j$

Lemma 140.

Jestliže a je vektor a X je náhodný vektor se střední hodnotou μ a variančně-kovarianční maticí Σ, tak

$$\mathbb{E}(a^T X) = a^T \mu \qquad \text{a} \qquad Var(a^T X) = a^T \Sigma a$$

Jestliže A je matice, tak

$$\mathbb{E}(AX) = A\mu \qquad \text{a} \qquad Var(AX) = A\Sigma A^T$$

Podmíněná střední hodnota

Podmíněná střední hodnota

Definice 141.

Podmíněná střední hodnota X za předpokladu $Y = y$ je

$$\mathbb{E}(X|Y = y) = \begin{cases} \sum x f_{X|Y}(x, y) & \text{v diskrétním případě} \\ \int x f_{X|Y}(x, y)\, dx & \text{ve spojitém případě} \end{cases}$$

Jestliže $r(x, y)$ je funkce x a y, pak

$$\mathbb{E}(r(X, Y)|Y = y) = \begin{cases} \sum r(x, y) f_{X|Y}(x, y) & \text{v diskrétním případě} \\ \int r(x, y) f_{X|Y}(x, y)\, dx & \text{ve spojitém případě} \end{cases}$$

Poznámka

- Zatímco $\mathbb{E}(X)$ je číslo, $\mathbb{E}(X|Y=y)$ je funkce y
  - dokud nepozorujeme hodnotu Y, neznáme hodnotu $\mathbb{E}(X|Y=y)$
  - jde o náhodnou veličinu značenou $\mathbb{E}(X|Y)$
- Jinak řečeno, $\mathbb{E}(X|Y)$ je NV jejíž hodnota je $\mathbb{E}(X|Y=y)$ pro $Y=y$.
- $\mathbb{E}(r(X,Y)|Y)$ je NV s hodnotou $\mathbb{E}(r(X,Y)|Y=y)$ pro $Y=y$.

Příklad 142.

- Zvolme $X \sim Uniform(0,1)$.
- Po té, co pozorujeme, že $X = x$, zvolme $Y|X = x \sim Uniform(x,1)$.
- Intuitivně očekáváme, že $\mathbb{E}(Y|X = x) = (1+x)/2$.
- Ve skutečnosti je $f_{Y|X}(y|x) = 1/(1-x)$ pro $x < y < 1$, a tedy

  $$\mathbb{E}(Y|X = x) = \int_x^1 y f_{Y|X}(y|x)\, dy = \frac{1+x}{2}$$

  jak jsme očekávali.
- Tedy $\mathbb{E}(Y|X) = (1+X)/2$.
- Všimněme si, že $\mathbb{E}(Y|X) = (1+X)/2$ je náhodná veličina jejíž hodnota je číslo $\mathbb{E}(Y|X = x) = (1+x)/2$ jakmile víme, že $X = x$.

Věta 143.

Pro náhodné veličiny X a Y, které mají střední hodnoty, platí

$$\mathbb{E}[\mathbb{E}(Y|X)] = \mathbb{E}(Y) \quad \text{a} \quad \mathbb{E}[\mathbb{E}(X|Y)] = \mathbb{E}(X)$$

Obecně, pro libovolnou funkci $r(x,y)$ platí $\mathbb{E}[\mathbb{E}(r(X,Y)|X)] = \mathbb{E}(r(X,Y))$.

Důkaz.

Ukážeme první rovnost. Z definice podmíněné střední hodnoty a $f(x,y) = f(x)f(y|x)$ máme

$$\begin{aligned}
\mathbb{E}[\mathbb{E}(Y|X)] &= \int \mathbb{E}(Y|X=x) f(x)\,dx = \iint y f(y|x)\,dy\, f(x)\,dx \\
&= \iint y f(y|x) f(x)\,dx\,dy = \iint y f(x,y)\,dx\,dy \\
&= \int y \left(\int f(x,y)\,dx \right) dy = \int y f_Y(y)\,dy = \mathbb{E}(Y)
\end{aligned}$$

$\square$

Příklad 144.

- Uvažme Příklad 142, kde $X \sim Uniform(0,1)$.
- Jak určíme $\mathbb{E}(Y)$?
- Můžeme najít sdruženou hustotu $f(x,y)$ a spočítat $\mathbb{E}(Y) = \iint y f(x,y)\, dx\, dy$.
- Jednodušší způsob je následující:
- My už víme, že $\mathbb{E}(Y|X) = (1+X)/2$, a proto

$$\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}\left(\frac{1+X}{2}\right) = \frac{1+\mathbb{E}(X)}{2} = \frac{1+\frac{1}{2}}{2} = \frac{3}{4}.$$

Definice 145.

Podmíněná variance je definována jako

$$Var(Y|X = x) = \int (y - \mu(x))^2 f(y|x)\, dy$$

kde $\mu(x) = \mathbb{E}(Y|X = x)$.

Věta 146.

Pro náhodné veličiny X a Y platí

$$Var(Y) = \mathbb{E}(Var(Y|X)) + Var(\mathbb{E}(Y|X)).$$

Příklad 147.

- Zvolme náhodně okres a z něj n lidí. Nechť X je počet lidí, kteří mají jistou nemoc.
  - Jestliže Q je poměr lidí majících onu nemoc v daném okrese, pak Q je NV, neboť se liší od okresu k okresu.
  - Pro $Q = q$ má $X \sim Binomial(n, q)$, tj. $\mathbb{E}(X|Q = q) = nq$ a $Var(X|Q = q) = nq(1 - q)$
- Předpokládejme, že Q má rovnoměrné rozdělení na $(0, 1)$, tj. máme tzv. hierarchistický model

$$Q \sim Uniform(0, 1) \quad \text{a} \quad X|Q = q \sim Binomial(n, q)$$

  - Pak $\mathbb{E}(X) = \mathbb{E}\mathbb{E}(X|Q) = \mathbb{E}(nQ) = n\mathbb{E}(Q) = n/2$ a
  - $Var(X) = \mathbb{E}(Var(X|Q)) + Var(\mathbb{E}(X|Q)) = (n/6) + (n^2/12)$, protože
    - $\mathbb{E}Var(X|Q) = \mathbb{E}[nQ(1 - Q)] = n\mathbb{E}(Q(1 - Q)) = n \int q(1 - q)f(q)\,dq = n \int_0^1 q(1 - q)\,dq = n/6$
    - $Var\mathbb{E}(X|Q) = Var(nQ) = n^2 Var(Q) = n^2 \int (q - (1/2))^2\,dq = n^2/12$

Momentové vytvořující funkce

Definice 148.

Momentová vytvořující funkce či Laplaceova transformace NV X je definována jako

$$\psi_X(t) = \mathbb{E}(e^{tX}) = \int e^{tx} dF(x) \qquad t \in \mathbb{R}$$

Předpokládáme, že je definována pro všechna t v nějakém otevřeném intervalu kolem $t = 0$.[7]

Pokud je fce dobře definovaná, platí

$$\psi'(0) = \left[\frac{d}{dt}\mathbb{E}(e^{tX})\right]_{t=0} = \mathbb{E}\left[\frac{d}{dt}e^{tX}\right]_{t=0} = \mathbb{E}[Xe^{tX}]_{t=0} = \mathbb{E}[X]$$

k-tá derivace $\psi^{(k)}(0) = \mathbb{E}(X^k)$, což dává způsob, jak počítat momenty rozdělení.

[7] Příbuzná fce je charakteristická fce definovaná jako $\mathbb{E}(e^{itX})$, která je dobře definovaná pro všechna t.

Příklad 149.

Nechť $X \sim Exp(1)$, tj. $f_X(x) = e^{-x}$. Pro pro $t < 1$ máme

$$\psi_X(t) = \mathbb{E}(e^{tX}) = \int_0^\infty e^{tx} e^{-x}\, dx = \int_0^\infty e^{(t-1)x}\, dx = \frac{1}{1-t}$$

a pro $t \geq 1$ integrál diverguje, tj. $\psi_X(t) = 1/(1-t)$ pro $t < 1$.

Protože $\psi'(0) = 1$ a $\psi''(0) = 2$, máme

$$\mathbb{E}(X) = 1$$

a

$$Var(X) = \mathbb{E}(X^2) - \mu^2 = 2 - 1 = 1$$

Lemma 150.

Vlastnosti momentové vytvořující funkce.

- *Jestliže* $Y = aX + b$, *pak* $\psi_Y(t) = e^{bt}\psi_X(at)$
- *Jestliže* $X_1, \ldots, X_n$ *jsou nezávislé a* $Y = \sum_i X_i$, *pak* $\psi_Y(t) = \prod_i \psi_i(t)$
  - *zde* ψ_i *je momentová vytvořující funkce* X_i

Příklad 151.

- Nechť $X \sim Binomial(n, p)$.
- Víme, že $X = \sum_{i=1}^{n} X_i$, kde $\mathbb{P}(X_i = 1) = p$ a $\mathbb{P}(X_i = 0) = 1 - p$.
- Pak $\psi_i(t) = \mathbb{E}(e^{tX_i}) = (p \cdot e^{t \cdot 1}) + (1 - p)e^{t \cdot 0} = pe^t + q$, kde $q = 1 - p$.
- Tedy
$$\psi_X(t) = \prod_i \psi_i(t) = (pe^t + q)^n$$

Zopakujme, že X a Y jsou rovné v rozdělení, $X \stackrel{d}{=} Y$, pokud mají stejné distribuční funkce.

Věta 152.

Nechť X a Y jsou NV. Jestliže $\psi_X(t) = \psi_Y(t)$ pro všechna t v nějakém otevřeném intervalu kolem 0, pak $X \stackrel{d}{=} Y$.

Příklad 153.

- Nechť $X_1 \sim Binomial(n_1, p)$ a $X_2 \sim Binomial(n_2, p)$ jsou nezávislé.
- Nechť $Y = X_1 + X_2$.
- Pak
$$\psi_Y(t) = \psi_1(t)\psi_2(t) = (pe^t + q)^{n_1}(pe^t + q)^{n_2} = (pe^t + q)^{n_1+n_2}$$
kde poslední je momentová vytvořující fce rozdělení $Binom(n_1 + n_2, p)$.
- Protože momentová vytvořující fce charakterizuje rozdělení (tj. neexistuje jiná NV se stejnou momentovou vytvořující fcí) dostáváme, že $Y \sim Binomial(n_1 + n_2, p)$.

Momentové vytvořující fce pro některá rozdělení

rozdělení	momentové vytvořující funkce $\psi(t)$
$Bernoulli(p)$	$pe^t + (1-p)$
$Binomial(n,p)$	$(pe^t + (1-p))^n$
$Poisson(\lambda)$	$e^{\lambda(e^t-1)}$
$Normal(\mu,\sigma)$	$e^{\mu t + \frac{\sigma^2 t^2}{2}}$
$Gamma(\alpha,\beta)$	$(\frac{1}{1-\beta t})^\alpha$ pro $t < 1/\beta$

Příklad 154.

- Nechť $Y_1 \sim Poisson(\lambda_1)$ a $Y_2 \sim Poisson(\lambda_2)$ jsou nezávislé.
- Momentová generající fce $Y = Y_1 + Y_2$ je

$$\psi_Y(t) = \psi_{Y_1}(t)\psi_{Y_2}(t) = e^{\lambda_1(e^t-1)}e^{\lambda_2(e^t-1)} = e^{(\lambda_1+\lambda_2)(e^t-1)}$$

  což je momentová generující fce $Poisson(\lambda_1 + \lambda_2)$.
- Tedy suma nezávislých Poissonových NV má Poissonovo rozdělení.

Nerovnosti

Nerovnosti jsou užitečné k ohraničení hodnot, které je těžké spočítat.

Věta 155 (Markovova nerovnost).

Nechť X je nezáporná náhodná veličina a nechť $\mathbb{E}(X)$ existuje. Pak pro libovolné $t > 0$ platí

$$\mathbb{P}(X > t) \leq \frac{\mathbb{E}(X)}{t}$$

Důkaz.

Protože $X > 0$, máme

$$\begin{aligned}\mathbb{E}(X) &= \int_0^\infty xf(x)\,dx = \int_0^t xf(x)\,dx + \int_t^\infty xf(x)\,dx \\ &\geq \int_t^\infty xf(x)\,dx \geq t\int_t^\infty f(x)\,dx = t\mathbb{P}(X > t)\end{aligned}$$

□

Věta 156 (Čebyševova nerovnost).

Nechť $\mu = \mathbb{E}(X)$ *a* $\sigma^2 = Var(X)$. *Pak*

$$\mathbb{P}(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2} \qquad a \qquad \mathbb{P}(|Z| \geq k) \leq \frac{1}{k^2}$$

kde $Z = (X - \mu)/\sigma$. *Zejména tedy platí*

- $\mathbb{P}(|Z| > 2) \leq 1/4$
- $\mathbb{P}(|Z| > 3) \leq 1/9$

Důkaz.

Použijeme Markovovu nerovnost a dostaneme

$$\mathbb{P}(|X - \mu| \geq t) = \mathbb{P}(|X - \mu|^2 \geq t^2) \leq \frac{\mathbb{E}(X - \mu)^2}{t^2} = \frac{\sigma^2}{t^2}$$

Druhá část plyne z toho, že položíme $t = k\sigma$. $\square$

Příklad 157.

- Testujeme predikční metodu, např. neuronovou síť, na n případech.
  - $X_i = \begin{cases} 1 & \text{pokud se prediktor mýlí} \\ 0 & \text{pokud ne} \end{cases}$
- Pak $\overline{X_n} = \frac{1}{n} \sum_{i=1}^{n} X_i$ je pozorovaná míra chybovosti.
  - Každé X_i lze považovat za Bernoulliho rozdělení s neznámou střední hodnotou p.
  - Rádi bychom znali správnou, avšak neznámou míru chybovosti p.
  - Intuitivně očekáváme, že $\overline{X_n}$ by mělo být blízko p.
- Jak je pravděpodobné, že $\overline{X_n}$ není v ϵ okolí p?
  - Máme $Var(\overline{X_n}) = Var(X_1)/n = p(1-p)/n$ a

$$\mathbb{P}(|\overline{X_n} - p| > \epsilon) \leq \frac{Var(\overline{X_n})}{\epsilon^2} = \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}$$

    neboť $p(1-p) \leq 1/4$ pro všechna p.
  - Pro $\epsilon = 0.2$ a $n = 100$ je hranice 0.0625.

Věta 158 (Hoeffdingova nerovnost).

Nechť $Y_1, \ldots, Y_n$ jsou nezávislá pozorování t.ž. $\mathbb{E}(Y_i) = 0$ a $a_i \leq Y_i \leq b_i$. Nechť $\epsilon > 0$. Pak pro libovolné $t > 0$ je

$$\mathbb{P}\left(\sum_{i=1}^{n} Y_i \geq \epsilon\right) \leq e^{-t\epsilon} \prod_{i=1}^{n} e^{t^2(b_i - a_i)^2/8}.$$

Důkaz

Pro libovolné $t > 0$ dává Markovova nerovnost

$$\mathbb{P}\left(\sum_{i=1}^{n} Y_i \geq \epsilon\right) = \mathbb{P}\left(t\sum_{i=1}^{n} Y_i \geq t\epsilon\right) = \mathbb{P}\left(e^{t\sum_{i=1}^{n} Y_i} \geq e^{t\epsilon}\right)$$

$$\leq e^{-t\epsilon}\mathbb{E}\left(e^{t\sum_{i=1}^{n} Y_i}\right) = e^{-t\epsilon}\prod_{i}\mathbb{E}\left(e^{tY_i}\right)$$

Důkaz pokračování.

Protože $a_i \leq Y_i \leq b_i$, lze je psát $Y_i = \alpha b_i + (1-\alpha)a_i$, kde $\alpha = (Y_i - a_i)/(b_i - a_i)$, a tedy

$$e^{tY_i} \leq \frac{Y_i - a_i}{b_i - a_i} e^{tb_i} + \frac{b_i - Y_i}{b_i - a_i} e^{ta_i}$$

Vzitím středních hodnot obou stran a faktu, že $\mathbb{E}(Y_i) = 0$, dostáváme

$$\mathbb{E}(e^{tY_i}) \leq -\frac{a_i}{b_i - a_i} e^{tb_i} + \frac{b_i}{b_i - a_i} e^{ta_i} = e^{g(u)}$$

kde $u = t(b_i - a_i)$, $g(u) = -\gamma u + \log(1 - \gamma + \gamma e^u)$ a $\gamma = -a_i/(b_i - a_i)$.

Platí $g(0) = g'(0) = 0$ a $g''(u) \leq 1/4$ pro všechna $u > 0$. Taylorova věta dává, že existuje $\xi \in (0, u)$ t.ž. $g(u) = g(0) + ug'(0) + \frac{u^2}{2} g''(\xi)$, tj.

$$g(u) = \frac{u^2}{2} g''(\xi) \leq \frac{u^2}{8} = \frac{t^2 (b_i - a_i)^2}{8}$$

Tedy $\mathbb{E}(e^{tY_i}) \leq e^{g(u)} \leq e^{t^2(b_i - a_i)^2/8}$, což dokazuje tvrzení.

Věta 159.

Nechť $X_1, \ldots, X_n \sim Bernoulli(p)$. *Pak pro libovolné* $\epsilon > 0$ *je*

$$\mathbb{P}(|\overline{X_n} - p| > \epsilon) \leq 2e^{-2n\epsilon^2}$$

kde $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Důkaz.

Nechť $Y_i = (1/n)(X_i - p)$. Pak $\mathbb{E}(Y_i) = 0$ a $a \leq Y_i \leq b$, kde $a = -p/n$ a $b = (1-p)/n$. Rovněž $(b-a)^2 = 1/n^2$. Z Věty 158 máme, že

$$\mathbb{P}(\overline{X_n} - p > \epsilon) = \mathbb{P}\left(\sum_i Y_i > \epsilon\right) \leq e^{-t\epsilon}e^{t^2/(8n)}$$

což platí pro libovolné $t > 0$. Pak $t = 4n\epsilon$ dává $\mathbb{P}(\overline{X_n} - p > \epsilon) \leq e^{-2n\epsilon^2}$. Podobně $\mathbb{P}(\overline{X_n} - p < -\epsilon) \leq e^{-2n\epsilon^2}$, tj. $\mathbb{P}(|\overline{X_n} - p| > \epsilon) \leq 2e^{-2n\epsilon^2}$. $\square$

Příklad 160.

Nechť $X_1, \ldots, X_n \sim Bernoulli(p)$. Nechť $n = 100$ a $\epsilon = 0.2$.
Čebyševova nerovnost dává

$$\mathbb{P}(|\overline{X_n} - p| > \epsilon) \leq .0625.$$

Hoeffdingova nerovnost dává

$$\mathbb{P}(|\overline{X_n} - p| > 0.2) \leq 2e^{-2(100)(0.2)^2} = 0.00067$$

což je mnohem menší než 0.0625.

Hoeffdingova nerovnost dává způsob, jak vytvořit interval spolehlivosti pro binomický parametr p (více později). Fixujme $\alpha > 0$ a nechť

$$\epsilon_n = \sqrt{\frac{1}{2n} \log\left(\frac{2}{\alpha}\right)}$$

Hoeffdingova nerovnost říká, že $\mathbb{P}(|\overline{X_n} - p| > \epsilon_n) \leq 2e^{-2n\epsilon_n^2} = \alpha$. Nechť

$$C = (\overline{X_n} - \epsilon_n, \overline{X_n} + \epsilon_n)$$

Pak

$$\mathbb{P}(p \notin C) = \mathbb{P}(|\overline{X_n} - p| > \epsilon_n) \leq \alpha \quad \text{tj.} \quad \mathbb{P}(p \in C) \geq 1 - \alpha$$

Tedy náhodně zvolený interval C obsahuje hodnotu parametru p s pravděpodobností $1 - \alpha$; budeme nazývat C interval spolehlivosti s koeficientem spolehlivosti α.

Následující nerovnost je užitečná pro ohraničování pravděpodobnostních tvrzení o normálních náhodných veličinách.

Věta 161 (Millova nerovnost).

Nechť $Z \sim N(0,1)$. *Pak*

$$\mathbb{P}(|Z| > t) \leq \sqrt{\frac{2}{\pi}} \frac{e^{-\frac{t^2}{2}}}{t}$$

Nerovnosti pro střední hodnoty

Věta 162 (Cauchyho-Schwartzova nerovnost).

Jestliže X a Y mají konečné variance, pak

$$\mathbb{E}(|XY|) \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}$$

Funkce g je konvexní, jestliže pro každé x, y a každé $\alpha \in [0, 1]$ platí

$$g(\alpha x + (1 - \alpha)y) \leq \alpha g(x) + (1 - \alpha)g(y).$$

- Pokud je $g''(x) \geq 0$ pro všechna x, pak g je konvexní.
- Pokud je g konvexní, pak g leží nad tečnou.
- Funkce g je konkávní, jestliže $-g$ je konvexní.
    - příklady konvexních fcí jsou $g(x) = x^2$ a $g(x) = e^x$
    - konkávních $g(x) = -x^2$ a $g(x) = \log x$

Věta 163 (Jensenova nerovnost).

- *Jestliže g je konvexní, pak $\mathbb{E}(g(X)) \geq g(\mathbb{E}X)$.*
- *Jestliže g je konkávní, pak $\mathbb{E}(g(X)) \leq g(\mathbb{E}X)$.*

Důkaz.

Nechť $L(x) = a + bx$ je tečna fce $g(x)$ v bodě $\mathbb{E}(X)$. Protože g je konvexní, leží g nad tečnou $L(x)$, tj. $\mathbb{E}(g(X)) \geq \mathbb{E}(L(X)) = \mathbb{E}(a + bX) = a + b\mathbb{E}(X) = L(\mathbb{E}(X)) = g(E(X))$. □

Důsledek 164.

- $\mathbb{E}(X^2) \geq (\mathbb{E}X)^2$
- *pokud je X pozitivní, pak $\mathbb{E}(1/X) \geq 1/\mathbb{E}(X)$*
- *protože $\log$ je konkávní, $\mathbb{E}(\log X) \leq \log \mathbb{E}(X)$*

Konvergence náhodných veličin

- Důležitá část teorie pravděpodobnosti se věnuje chování sekvence náhodných veličin.
- Říká se jí large sample theory (limitní teorie, asymptotická teorie).
  - Otázkou je, co můžeme říct o limitním chování posloupnosti náhodných veličin $X_1, X_2, X_3, \ldots$?

- Statistika a data mining jsou o získávání dat.
- Co se děje, když máme více a více dat?

- V analýze posloupnost reálných čísel x_n konverguje k limitě x, jestliže pro každé $\epsilon > 0$ je $|x_n - x| < \epsilon$ pro všechna dostatečně velká n.
  - Mějme na chvíli $x_n = x$ pro všechna n, pak $\lim_{n\to\infty} x_n = x$.
- V pravděpodobnosti je situace trochu komplikovanější.
- Nechť $X_1, X_2, \ldots$ jsou nezávislé NV každá s rozdělením $N(0,1)$.
  - Jelikož všechny NV mají stejné rozdělení, zdálo by se, že X_n "konverguje" k $X \sim N(0,1)$.
- To ale není v pořádku, neboť $\mathbb{P}(X_n = X) = 0$ pro všechna n.
  - Dvě spojité NV jsou rovné s pravděpodobností nula.
  - $\mathbb{P}(X = Y) = \mathbb{P}(X - Y = 0) = 0$
- Jiný příklad.
  - Nechť $X_1, X_2, \ldots$, kde $X_i \sim N(0, 1/n)$
  - X_n je koncentrováno kolem 0 pro velká n
  - Chtěli bychom proto říct, že X_n konverguje k 0.
    - Ale $\mathbb{P}(X_n = 0) = 0$ pro všechna n.
- Potřebujeme nějaký nástroje k definici konvergence.

Zde se podíváme na následující dva:

- Zákon velkých čísel
  říká, že výběrový průměr $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ konverguje v pravděpodobnosti ke střední hodnotě $\mu = \mathbb{E}(X_i)$
  - tj. $\overline{X}_n$ je blízko μ s velkou pravděpodobností
- Centrální limitní věta
  říká, že $\sqrt{n}(\overline{X}_n - \mu)$ konverguje v rozdělení k normálnímu rozdělení
  - tj. výběrový průměr má přibližně normální rozdělení pro velká n

Typy konvergence

Definice 165.

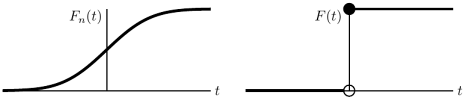
Nechť $X_1, X_2, \ldots$ je posloupnost NV a X je další NV. Nechť F_n značí distribuční funkci X_n a F distribuční funkci X.

1. X_n konverguje k X v pravděpodobnosti, $X_n \xrightarrow{P} X$, jestliže pro každé $\epsilon > 0$ platí
$$\mathbb{P}(|X_n - X| > \epsilon) \to 0$$
pro $n \to \infty$.
2. X_n konverguje k X v rozdělení, $X_n \rightsquigarrow X$, jestliže
$$\lim_{n\to\infty} F_n(t) = F(t)$$
pro všechna t, kde je F spojitá.

Pokud je limitní NV bodová, $\mathbb{P}(X = c) = 1$ a $X_n \xrightarrow{P} X$, píšeme $X_n \xrightarrow{P} c$.
Podobně pro $X_n \rightsquigarrow X$ píšeme $X_n \rightsquigarrow c$.

Příklad 166 (Konvergence v rozdělení).

- Nechť $X_n \sim N(0, 1/n)$. Konverguje X_n k 0? Platí$\sqrt{n}X_n \sim N(0, 1)$?
- Nechť F je distribuční funkce s bodovým rozdělením v 0.
- Nechť Z je standardní normální NV.
  - Pro $t < 0$ je $F_n(t) = \mathbb{P}(X_n < t) = \mathbb{P}(\sqrt{n}X_n < \sqrt{n}t) = \mathbb{P}(Z < \sqrt{n}t) \to 0$, neboť $\sqrt{n}t \to -\infty$.
  - Pro $t > 0$ je $F_n(t) = \mathbb{P}(X_n < t) = \mathbb{P}(\sqrt{n}X_n < \sqrt{n}t) = \mathbb{P}(Z < \sqrt{n}t) \to 1$, neboť $\sqrt{n}t \to \infty$.
  - x Tedy $F_n(t) \to F(t)$ pro všechna $t \neq 0$, tj. $X_n \rightsquigarrow 0$.
- Avšak $F_n(0) = 1/2 \neq F(1/2) = 1$, tj. konvergence neplatí v $t = 0$. Na tom ale nezáleží, protože v $t = 0$ není F spojitá a definice konvergence v rozdělení vyžaduje konvergenci v bodech, kde je fce spojitá.



Příklad 167 (Konvergence v pravděpodobnosti).

- Nechť $X_n \sim N(0, 1/n)$. Konverguje X_n k 0?
- Opět je F distribuční funkce s bodovým rozdělením v 0.
- Jak je to s konvergencí v pravděpodobnosti?
- Pro libovolné $\epsilon > 0$ dává Markovova nerovnost

$$\mathbb{P}(|X_n| > \epsilon) = \mathbb{P}(|X_n|^2 > \epsilon^2) \leq \frac{\mathbb{E}(X_n^2)}{\epsilon^2} = \frac{\frac{1}{n}}{\epsilon^2} \to 0$$

  pro $n \to \infty$
- Tedy $X_n \xrightarrow{P} 0$.

Následující věta popisuje vztah mezi typy konvergence.

Věta 168.

Následující vztah platí:

$$X_n \xrightarrow{P} X \text{ implikuje, že } X_n \rightsquigarrow X$$

Opačná implikace neplatí, až na následující speciální případ:

Jestliže $X_n \rightsquigarrow X$ *a* $\mathbb{P}(X = c) = 1$ *pro nějaké reálné* c*, pak* $X_n \xrightarrow{P} X$.

Důkaz. TODO

Věta 169.

Nechť X_n, X, Y_n, Y jsou náhodné veličiny. Nechť g je spojitá funkce.

1. *Jestliže $X_n \xrightarrow{P} X$ a $Y_n \xrightarrow{P} Y$, pak $X_n + Y_n \xrightarrow{P} X + Y$.*
2. *Jestliže $X_n \rightsquigarrow X$ a $Y_n \rightsquigarrow c$, pak $X_n + Y_n \rightsquigarrow X + c$.*
3. *Jestliže $X_n \xrightarrow{P} X$ a $Y_n \xrightarrow{P} Y$, pak $X_n Y_n \xrightarrow{P} XY$.*
4. *Jestliže $X_n \rightsquigarrow X$ a $Y_n \rightsquigarrow c$, pak $X_n Y_n \rightsquigarrow cX$.*
5. *Jestliže $X_n \xrightarrow{P} X$, pak $g(X_n) \xrightarrow{P} g(X)$.*
6. *Jestliže $X_n \rightsquigarrow X$, pak $g(X_n) \rightsquigarrow g(X)$.*

Obecně $X_n \rightsquigarrow X$ a $Y_n \rightsquigarrow Y$ neimplikuje $X_n + Y_n \rightsquigarrow X + Y$.

Zákon velkých čísel

Zákon velkých čísel je jeden z hlavních výsledků teorie pravděpodobnosti. Říká, že střední hodnota velkého výběru se blíží střední hodnotě rozdělení; např. při velkém počtu hodů padne orel kolem poloviny případů.

Nechť $X_1, X_2, \ldots$ jsou IID NV. Nechť $\mu = \mathbb{E}(X_i)$ a $\sigma^2 = Var(X_i)$. Zopakujme, že výběrový průměr $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ a že $\mathbb{E}(\overline{X}_n) = \mu$ a $Var(\overline{X}_n) = \sigma^2/n$.

Věta 170 (Slabý zákon velkých čísel).

Jestliže $X_1, \ldots, X_n$ jsou IID, pak $\overline{X}_n \xrightarrow{P} \mu$.

Tj., rozdělení $\overline{X}_n$ se začíná koncentrovat kolem μ s rostoucím n.

Důkaz.

Předpokládejme, že $\sigma < \infty$. Tento předpoklad není nutný, ale zjednodušuje důkaz. Z Čebyševovy nerovnosti máme

$$\mathbb{P}(|\overline{X}_n - \mu| > \epsilon) \leq \frac{Var(X_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

což jde k 0 pro $n \to \infty$. $\square$

Příklad 171.

Hod mincí s pravděp. $p =$ padne orel. Nechť X_i je výsledek jednoho hodu (0 nebo 1), tj.

$$p = \mathbb{P}(X_i = 1) = \mathbb{E}(X_i)$$

Poměr orlů po n hodech je $\overline{X}_n$. Zákona velkých čísel říká, že $\overline{X}_n$ konverguje v pravděpodobnosti k p. To však neznamená, že $\overline{X}_n$ se bude rovnat p, ale že pro velká n bude rozdělení $\overline{X}_n$ koncentrované těsně kolem p.

Nechť $p = 1/2$. Jak velké musí být n, aby $\mathbb{P}(.4 \leq \overline{X}_n \leq .6) \geq .7$?

Máme $\mathbb{E}(\overline{X}_n) = p = 1/2$ a $Var(\overline{X}_n) = \sigma^2/n = p(1-p)/n = 1/(4n)$. Čebyševova nerovnost pak dává

$$\begin{aligned}\mathbb{P}(.4 \leq \overline{X}_n \leq .6) = \mathbb{P}(|\overline{X}_n - \mu| \leq .1) = 1 - \mathbb{P}(|\overline{X}_n - \mu| > .1)\\ \geq 1 - \frac{1}{4n(.1)^2} = 1 - \frac{25}{n}\end{aligned}$$

což je větší než .7 pro $n = 84$.

Silný zákon velkých čísel

Zatímco slabý zákon velkých čísel říká, že $\overline{X}_n$ konverguje v pravděpodobnosti ke střední hodnotě $\mathbb{E}(X_1)$, silný zákon velkých čísel říká, že skoro jistě konveguje ke střední hodnotě.

Věta 172 (Silný zákon velkých čísel).

Nechť $X_1, X_2, \ldots$ *jsou IID. Jesliže* $\mu = \mathbb{E}(|X_1|) < \infty$, *pak*

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = \mu\right) = 1$$

Slabý vs. silný zákon velkých čísel

Slabý zákon velkých čísel říká, že pro lib. specifikovanou velkou hodnotu n^* je $(X_1 + \cdots + X_{n^*})/n^*$ blízko μ. Neříká však, že $(X_1 + \cdots + X_n)/n$ musí být blízko μ pro všechny hodnoty $n > n^*$, tj. připouští možnost, že se velké hodnoty

$$\left| \frac{X_1 + \cdots + X_n}{n} - \mu \right|$$

mohou vyskytnout nekonečně často.

Silný zákon říká, že toto nemůže nastat, tj. s pravděpodobností 1 bude

$$\left| \frac{X_1 + \cdots + X_n}{n} - \mu \right|$$

větší než jakékoliv $\epsilon > 0$ pouze konečně mnohokrát.

Centrální limitní věta

- Zákon velkých čísel říká, že rozdělení $\overline{X}_n$ jde k μ.
- To nám nepomůže aproximovat tvrzení o $\overline{X}_n$.
- K tomu potřebujeme centrální limitní větu.

- Nechť $X_1, \ldots, X_n$ jsou IID se střední hodnotou μ a variancí σ^2.
- Centrální limitní věta říká, že $\overline{X}_n = \frac{1}{n} \sum_i X_i$ má rozdělení, které je přibližně rovno normálnímu rozdělení se střední hodnotou μ a variancí σ^2/n.
- Toto je pozoruhodné, neboť o rozdělení X_i nic nepředpokládáme, mimo toho, že střední hodnoty a variance existují.

Věta 173 (Centrální limitní věta).

Nechť $X_1, \ldots, X_n$ *jsou IID se střední hodnotou* μ *a variancí* σ^2. *Nechť* $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. *Pak tzv. normalizovaná či standardizovaná veličina*

$$Z_n \equiv \frac{\overline{X}_n - \mu}{\sqrt{Var(\overline{X}_n)}} = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \rightsquigarrow Z$$

kde $Z \sim N(0,1)$. *Jinak řečeno,*

$$\lim_{n\to\infty} \mathbb{P}(Z_n \leq z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

Pozn. Pravděpodobnostní tvrzení o $\overline{X}_n$ lze aproximovat pomocí normálního rozdělení. Aproximujeme pravděpodobnostní tvrzení, nikoliv samotnou náhodnou veličinu.

Příklad 174.

- Předpokládejme, že počet chyb na jeden počítačový program má Poissonovo rozdělení se střední hodnotou 5.
- Dostaneme 125 programů.
- Nechť $X_1, \ldots, X_{125}$ jsou počty chyb v programech.
- Budeme aproximovat $\mathbb{P}(\overline{X}_n < 5.5)$.
- Nechť $\mu = \mathbb{E}(X_1) = \lambda = 5$ a $\sigma^2 = Var(X_1) = \lambda = 5$.
- Pak $\mathbb{P}(\overline{X}_n < 5.5) = \mathbb{P}\left(\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} < \frac{\sqrt{n}(5.5 - \mu)}{\sigma}\right) \approx \mathbb{P}(Z < 2.5) = .9938$.

- Centrální limitní věta říká, že

$$Z_n = \sqrt{n}(\overline{X}_n - \mu)/\sigma$$

  je přibližně $N(0,1)$.
- My však neznáme σ.
- Později uvidíme, že σ^2 lze odhadnout z $X_1, \ldots, X_n$ pomocí

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

Tím vzniká otázka, platí centrální limitní věta, pokud nahradíme σ za S_n? Odpověď je ano.

Věta 175.

Předpokládejme stejné podmínky jako u centrální limitní věty. Pak

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n} \rightsquigarrow N(0,1)$$

Jak přesná je tato normální aproximace?

Věta 176 (Berry-Essèenova nerovnost).

Předpokládejme, že $\mathbb{E}|X_1|^3 < \infty$. Pak

$$\sup_z |\mathbb{P}(Z_n \leq z) - \Phi(z)| \leq \frac{33}{4} \frac{\mathbb{E}|X_1 - \mu|^3}{\sqrt{n}\sigma^3}$$

Delta Method

Jestliže má Y_n limitní normální rozdělení, pak delta metoda nám umožňuje najít limitní rozdělení $g(Y_n)$, kde g je libovolná hladká funkce.

Věta 177 (Delta Metoda).

Předpokládejme, že

$$\frac{\sqrt{n}(Y_n - \mu)}{\sigma} \rightsquigarrow N(0,1)$$

a že g je diferencovatelná funkce t.ž. $g'(\mu) \neq 0$. Pak

$$\frac{\sqrt{n}(g(Y_n) - g(\mu))}{|g'(\mu)|\sigma} \rightsquigarrow N(0,1)$$

Jinak řečeno, $Y_n \approx N(\mu, \frac{\sigma^2}{n})$ implikuje, že $g(Y_n) \approx N(g(\mu), (g'(\mu))^2 \frac{\sigma^2}{n})$.

Příklad 178.

Nechť $X_1, \ldots, X_n$ jsou IID s konečnou střední hodnotou μ a konečnou variancí σ^2. Podle centrální limitní věty máme

$$\sqrt{n}(\overline{Xn} - \mu)/\sigma \rightsquigarrow N(0,1)$$

Nechť $W_n = e^{\overline{X}_n}$, a tedy $W_n = g(\overline{X}_n)$ kde $g(s) = e^s$. Protože $g'(s) = e^s$, delta metoda implikuje, že

$$W_n \approx N(e^{\mu}, e^{2\mu}\sigma^2/n)$$