Pravdepodobnosť a štatistika 1

Lukáš Laffērs

2025-02-13

Contents

O tomto kurze

Tento kurz je pomalým a zábavným úvodom do pravdepodobnosti a štatistiky. Cieľom tohoto kurzu je naučiť sa pracovať s jazykom a s pojmami, ktoré sa budú používať v situáciách, v ktorých čelíme *neistote*.

Táto neistota je rôzneho pôvodu. Jej zdroj však nebude pre nás podstatný, dôležité bude, že sa naučíme *kvantifikovať* túto neistotu a pracovať s ňou. Na to, aby sme toto vedeli efektívne robiť, sa potrebujeme oboznámiť s mnohými novými pojmami. Aby sme nabrali náležitú plynulosť vo vyjadrovaní sa v jazyku pravdepodobnosti je potrebné

- pracovať s týmito novými pojmami,
- pýtať sa zaujímavé otázky a odpovedať na ne, ale aj
- prepočítať veľa príkladov.

Toto bude zahŕňať veľa práce. Spôsob uchopenia týchto základných stavebných prvkov však do veľkej miery ovplyvní to, ako ľahko/ťažko sa vám neskôr bude pracovať vo vaších dátových, vo finančných, či v poisťovacích analýzach. Má preto zmysel investovať do vytvárania predstáv čas a energiu. Zároveň sa však nádejam, že aspoň v niektorých z vás vzbudím nadšenie a oceníte estetiku matematickej teórie pravdepodobnosti a štatistiky.

Plán je prebrať nasledujúce témy:

- Interpretácia pravdepodobnosti.
- Pravdepodobnostný priestor.
- Podmienená pravdepodobnosť, nezávislosť udalostí a paradoxy.
- Diskrétne rozdelenia náhodných premenných - popis náhodnosti pomocou pravdepodobnostnej funkcie a kumulatívnej distribučnej funkcie.
- Spojité rozdelenia náhodných premenných - popis nahodnosti pomocou funkcie hustoty a kumulatívnej distribučnej funkcie.
- Charakteristiky náhodných premenných - stredná hodnota, variancia.
- Nezávislosť náhodných premenných a súvis náhodných premenných - kovariancia a korelácia.
- Zákon veľkých čísel.
- Centrálna limitná veta.

Na tento kurz nadväzuje praktickejší kurz vo výpočtovom prostredí R. Tento nadväzujúci kurz je na https://lukaslaffers.github.io/pas2/

1 Interpretácia pravdepodobnosti

Teória pravdepodobnosti je jazykom neistoty.

V živote často čelíme neistote, rôznym typom neistoty. Pýtame sa nasledujúce otázky:

- Bude zajtra pršať?
- Bude prebiehať výučba tento semester konečne prezenčne?
- Prekoná cena ropy hranicu 100$ za barel do konca roka?
- Vyhrá politická strana XY voľby?
- Vybuchne tento rok elektráreň v Mochovciach?
- Vezmeš si ma?

Férová a úprimná odpoveď na všetky tieto otázky je: *"Neviem."* (No dobre, okrem tej poslednej.)

Prečo? Nuž, lebo naozaj nevieme. V skutočnosti je však obrovský rozdiel v týchto otázkach: to "Nevieme." zďaleka nie je rovnaké v týchto situáciách. Teda táto odpoveď vôbec nie je uspokojivá a užitočná tiež nie je. Chceli by sme čosi viacej. Chceli by sme nejakým spôsobom kvantifikovať neistotu spojenú s týmito otázkami a tam aj teraz smerujeme. Práve za týmto účelom používame slovo pravdepodobnosť a teraz ideme budovať matematickú teóriu, aby sme neistote lepšie porozumeli a vedeli s ňou pracovať, spoznali jej štruktúru.

Chceme vedieť: **Koľko neistoty** je spojenej s týmito otázkami? Chceme to nejak odmerať.

Iná, tiež férová a tiež úprimná odpoveď na všetky tieto otázky by bola: *"Nevieme. Lebo je to náhodné."*

No dobre, ale rôzne udalosti sú rôzne náhodné. Niektoré nastávajú často, iné menej často. U iných zasa nedáva moc zmysel pýtať sa, ako často nastávajú, nakoľko ide len o jednorázovú udalosť. U iných si veľmi ľahko vieme predstaviť opakovanie a tam sa môžeme pýtať ako často udalosť nastáva.

1.1 Frekventisitická interpretácia

Môžeme sa pýtať aj iné otázky:

- Padne pri 5 nezávislých hodoch férovou kockou aspoň jedenkrát číslo 6?
- Padne pri 10 nezávislých hodoch férovou mincou najviac trikrát znak?
- Vyberiem z kartového balíčka srdcovú postupku?

Asi uznáte, že tieto otázky sú *umelohmotné.* Vyskytujú sa najmä v učebniciach pravdepodobnosti a v reálnom živote už pomenej, ba až vôbec. To ale neznamená, že nie sú užitočné. Sú totiž užitočné na to, aby sme abstrahovali od zložitostí reálneho sveta, ktoré zneprehľadňujú naše rozmýšlanie o pravdepodobnosti a o neistote a sústredili sa len na podstatu. Pochopili koncept pravdepodobnosti. My sa na takýchto umelohmotných prípadoch musíme naučiť rozmýšľať a potom bude oveľa ľahšie vedieť správne priradiť tieto koncepty k reálnym aplikáciám.

Vráťme sa naspäť k tým otázkach. Chceli by sme náhodu *kvantifikovať.*

- Aká je šanca, že pri 5 nezávislých hodoch férovou kockou padne aspoň jedenkrát číslo 6?
- Aká je šanca, že pri 10 nezávislých hodoch férovou mincou padne najviac trikrát znak?
- Aká je šanca, že vyberiem z kartového balíčka srdcovú postupku?

Tu tak potichu predpokladáme, že máme akési *opakovateľné* experimenty v kontrolovanom prostredí. A že pri dostatočne veľkom počte opakovaní sa frekvencia nastatia danej udalosti (napr. pád aspoň 1 šestky z 5 hodov) bude blížiť k nejakému číslu. Toto číslo myslíme pod tým pojmom "šanca". Áno, vidíte kam smerujem; tomuto číslu zvykneme hovoriť aj **pravdepodobnosť**.

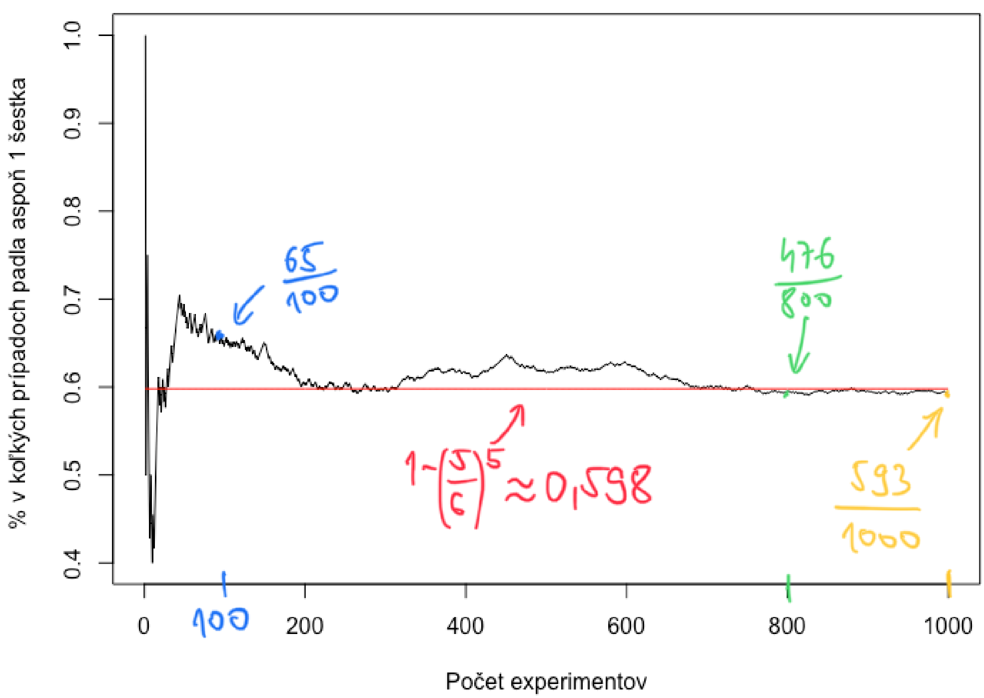


Figure 1: Počet "úspešných" experimentov sa blíži k akejsi hodnote. Z prvých 800 experimentov padla šesťka aspon v jednom prípade z piatich 476 krát. Červenou je skutočná pravdepodobnosť. (Zdroj: vlastná simulácia v R.)

Tento obrázok vyjadruje, ako sa postupne vyvíja frekvencia nastatia udalosti "pri 5 nezávislých hodoch férovou kockou padne aspoň jedenkrát číslo 6". Blíži sa k akémusi číslu a toto nie je zhoda náhod. Súvisí to s jedným hlbokým výsledkom (nazýva sa Zákon o Veľkých Číslach), ktorému sa budeme venovať neskôr. Takejto interpretácii pravdepodobnosti, keď si vieme predstaviť opakujúci sa experiment hovoríme aj **frekventistická**. Táto interpretácia pravdepodobnosti je založená na tom, že jednotlivé experimenty sú rovnocenné.

1.2 Subjektivistická interpretácia

Ale svet nie sú kocky, mince a karty. Niektoré udalosti nástavajú len a nie je zmysluplné uvažovať o ich opakovaní. Skrátka a dobre, frekventistická interpretácia pravdepodobnosti v tomto prípade nesedí.

- Aká je pravdepodobnosť, že cena ropy prekoná hranicu 100$ za barel do konca roka?
- Aká je šanca, že strana XY zvládne zostaviť koalíciu v najbližších parlamentných voľbách?

- Aká je pravdepodobnosť zemetrasenia v Banskej Bystrici počas najbližších 5 rokov?

Asi sa zhodneme na tom, že toto sú zaujímavé otázky, takže by sme konieckoncov chceli vedieť na ne nejakú odpoveď. Každopádne nám tu chýba niečo, čo pri kockách, minciach a kartách máme. Chýba nám objektívny rámec, ktorým by sme vedeli vyhodnotiť úspešnosť danej odpovede. Predtým sme vedeli experimenty opakovať, teraz to však robiť nevieme. Môžeme vytvoriť nejaký matematický **model**, ktorý bude modelovať dynamiku ceny ropy. Toto je samozrejme ohromne náročné. Takýto model je akási zjednodušenina sveta. Robí mnoho zjenodušujúcich predpokladov ale na niektoré otázky môže dávať čiastočne uspokojivé predpovede. Každopádne, či chceme, či nie, takéto vyhodnotenie pravdepodobnosti bude vždy **subjektívne**. Matematický model môže byť validovaný na predošlých dátach a vykazovať úspešné pravdepodobnostné predpovede ale nikdy nebudeme mať garanciu toho, že bude fungovať aj v budúcnosti. Čaro modelovania tkvie v tom urobiť jednoduchý model, ale nie príliš.

Pred prezidentskými voľbami v USA v roku 2016 bola značnou favoritkou Hillary Clinton pred Donaldom Trumpom. Väčšina think-tankov prisudzovala víťazstvu Clinton vyše 80%, tá však voľby nevyhrala.

Figure 2: Toto je vrchná časť grafiky webu fivethirtyeight.com, ktorý odhadoval pravdepodobnosť víťazstva demokratickej kandidátky v prezidentských voľbách USA v roku 2016. (Zdroj: https://projects.fivethirtyeight.com/2016-election-forecast/, prístup 7.2.2022)

Čo presne znamená toto číslo 71.4%?[1] V tomto prípade šlo o to, že v 14180 prípadoch z 20000 model fivethirtyeight.com nasimuloval víťazstvo Clinton. Takže aj keď nemáme objektívny rámec, môžeme si pomôcť tým, že si *simulačný rámec vytvoríme pomocou matematického modelu.* Ich model zahŕňal mnoho expertnej skúsenosti sociológov, volebných expertov, ekonómov, demografov a podobne. Kontinuálne upravovali odhady podľa najčerstvejších prieskumov v jednotlivých štátov, do úvahy tiež brali vierohodnosť jednotlivých prieskumných agentúr (tie ktorých prieskumy boli v minulosti bližšie realite dostali väčšiu váhy). Ale tiež šlo len o subjektívne vyhodnotenie pravdepodobnosti. Subjektívnosť tkvie v tom, ako tento volebný model skonštruovali.[2]

Predikovali 71.4%, a to je viacej ako 50%. Znamená to, že sa mýlili?

Nie je to také jednoduché. Ak by sa takéto voľby konali 1000 krát, tak by sme vedeli vyhodnotiť ako úspešné toto číslo bolo. Nuž ale toto sa nestalo. Voľby boli unikátna udalosť, ktorá nastala presne jeden raz. No a teraz to nevyšlo.[3]

"V najbližsiu hodinu bude na 90% snežiť." Čo presne znamená tých 90%?

Nuž, podobne ako pri voľbách, aj tu je to pravdepodobnostný odhad založený na základe modelu, konkrétne meterologického. Alebo priemeru mnohých meteorologických modelov.

Intuitívne to vnímame, že ako veľmi môžeme veriť tomu, že bude snežiť. Pri 10% veríme málo, pri 90% je zasa veľa. Kedy by sme chápali, že tých 90% je dobrý odhad? Ak by sme mali takú skúsenosť, že mnohokrát ako sme túto aplikáciu kontrolovali, tak keď nám telefón ukazoval 90%, tak snežilo v 9 prípadoch z 10. Takže, ak práve teraz nebude snežiť, tak to nutne neznamená, že táto predpoveď je zlá.

[1] Pozor predpoveď 71.4% neznamená, že Clinton mala získať v priemere 71.4% hlasov. Ide o úplne iné čísla.

[2] New York Times oslovil prominentných odhadovačov, dal im **tie isté surové dáta** z prieskumu v štáte Florida a požiadal ich o predikciu. Tieto predikcie boli výrazné iné. https://www.nytimes.com/interactive/2016/09/20/upshot/the-error-the-polling-world-rarely-talks-about.html

[3] Toto je samozrejme prudké zjednodušenie. Každá volebná predikcia zahŕňala nielen celkový výsledok ale aj výsledky v jednotlivých štátoch. Takže vyhodnotiť úspešnosť predikcie nie je úplne nemožné. Ale nedá sa to urobiť len pomocou jednej vyšlo/nevyšlo otázky.

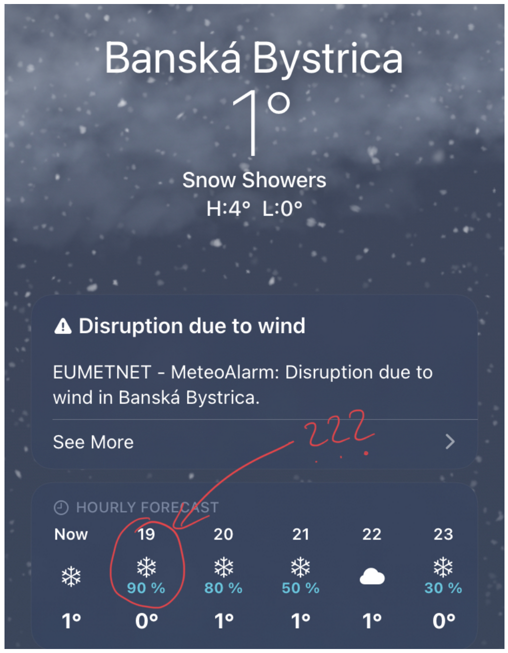


Figure 3: Čo znamená (Zdroj: Weather aplikácia, prístup 7.2.2022)

Dva základné referenčné rámce pre pravdepodobnosť, ktorým veľmi dobre rozumieme sú

- nikdy (0%),
- určite (100%).

Ďalej vieme, že väčšia pravdepodobnosť znamená, že niečo nastáva častejšie (pri opakovaných pokusoch - frekventistická interpretácia) alebo si myslíme viacej, že nastane (subjektivistická interpretácia).

Nielen to. Má zmysel uvažovať aj o tom, že niektorá udalosť môže byť nastávať dvakrát tak často. Ale teraz už predbiehame, o tomto čoskoro.

1.3 Zhrnutie

Pravdepodobnosť sa dá interpretovať rôzne. Napríklad *frekventisticky* alebo *subjektivisticky*. Povzbudením nám môže byť to, že nezávisle od typu interprácie je jazykom neistoty stále matematická teória pravdepodobnosti. Teórii pravdepodnosti je totiž tak trochu jedno, ako si vy čo interpretujete.[4]

1.4 Cvičenia

Cvičenie 1.1. Uveďte príklad na reálnu situáciu (žiadne karty, mince ani kocky), kedy je frekventistická interpretácia pravdepodobnosti adekvátna.

Cvičenie 1.2. Zamyslite sa, ako by ste overili, či mobilná aplikácia s počasím ukazuje presné predpovede.

2 Pravdepodobnosť

Na začiatku si potrebujeme zopakovať množinovú symboliku. Táto predstavuje absolútny základ, bez ktorého sa nedá zmysluplne pokračovať. Plynulosť v používaní matematických symbolov uľahčuje dôvodenie a argumentáciu. Zápis je potom kompaktný a univerzálny, nezávislý od nuáns špecifického jazyka.

[4] Viacej si prečítajte tu: Stark, P. B., et al. "What is the chance of an earthquake." NATO Science Series IV: Earth and Environmental Sciences 32 (2003): 201-213.

2.1 Množinové značenie

Množinou budeme nazývať súhrn objektov/prvkov.

Množina je **konečná**, ak má konečne veľa prvkov. Napríklad $M = \{1, 2, 8\}$ má tri prvky.

Množina je **spočitateľná**, ak má nanajvýš spočitateľne veľa prvkov. Napríklad $M_1 = \{1, 22, -2\}$ alebo $M_2 = \{2, 4, 6, 8, \dots\}$. Množina M je teda spočitateľná, ak existuje nejaké injektívne zobrazenie z množiny M do množiny prirodzených čísel $\mathbb{N}$. Inými slovami, prvkov v množine M nie je "príliš veľa".

Množina M_1 je **podmnožinou** M_2, označujeme $M_1 \subseteq M_2$ ak platí, že každý element M_1 je zároveň aj elementom v M_2, teda ak platí $\forall m : m \in M_1 \implies m \in M_2$.

Množinu M, ktorá nemá žiadne prvky nazývame **prázdna** a označujeme $M = \emptyset$.

Zjednotením dvoch množín M_1 a M_2 nazývame množinu, ktorej prvky sa nachádzajú buď v množine M_1 alebo v množine M_2. Zjednotenie dvoch množín označujeme ako $M_1 \cup M_2$.

Zjednotením viacerých (nanajvýš spočitateľne veľa) množín $M_1, M_2, M_3, \dots$ nazývame množinu, ktorej prvky sa nachádzajú aspoň v jednej z množín $M_1, M_2, M_3, \dots$. Zjednotenie viacerých množín označujeme ako $M_1 \cup M_2 \cup M_3 \cup \dots = \cup_{i=1}^{\infty} M_i$ pre zjednotenie nekonečne veľa množín alebo ako $M_1 \cup M_2 \cup M_3 \cup \dots \cup M_n = \cup_{i=1}^{n} M_i$ pre zjednotenie konečného počtu množín.

Prienikom dvoch množín M_1 a M_2 nazývame množinu, ktorej prvky sa nachádzajú aj v M_1 aj v M_2. Prienik dvoch množín označujeme ako $M_1 \cap M_2$.

Dve množiny M_1 a M_2 nazývame **disjunktné**, ak ich spoločný prienik je prázdna množina, teda $M_1 \cap M_2 = \emptyset$.

Prienikom viacerých (nanajvýš spočitateľne veľa) množín $M_1, M_2, M_3, \dots$ nazývame množinu, ktorej prvky sa nachádzajú v každej jednej z množín $M_1, M_2, M_3, \dots$. Prienik viacerých množín označujeme ako $M_1 \cap M_2 \cap M_3 \cap \dots = \cap_{i=1}^{\infty} M_i$ alebo ako $M_1 \cap M_2 \cap M_3 \cap \dots \cap M_n = \cap_{i=1}^{n} M_i$ pre prienik konečného počtu množín.

Mnohokrát dáva zmysel uvažovať o všetkých možných objektov/prvkov. **Priestor** označuje súhrn všetkých možných objektov. Častokrát ako notáciu používame veľké zakrútené písmená, napr. $\mathcal{M}$. Napríklad ak hovoríme o tom, aké číslo padne na kocke, tak priestor je $\{1, 2, 3, 4, 5, 6\}$.

Nech $\mathcal{M}$ je priestor a nech $M \subseteq \mathcal{M}$. Potom **komplementom** množiny M (vzhľadom na priestor $\mathcal{M}$), označovaným ako M^C, nazývame množinu tých objektov z priestoru $\mathcal{M}$, ktoré nie sú v M.

Nejaké príklady:

- Pre každú množinu M platí:
  - $M \cup M = M$,
  - $M \cap M = M$,
  - $M \cap \emptyset = \emptyset$ a
  - $M \cup \emptyset = M$.
- Ak $M_1 \subseteq M_2$ potom
  - $M_1 \cup M_2 = M_2$ a
  - $M_1 \cap M_2 = M_1$
- Ak $M_i = \{x : 0 < x < \frac{1}{i}\},\ k = 1, 2, 3, \dots$ potom $\cap_{i=1}^{\infty} M_i = \emptyset$
- Ak $M_i = \{x : \frac{1}{i+1} < x \leq \frac{1}{i}\},\ k = 1, 2, 3, \dots$ potom $\cup_{i=1}^{\infty} M_i = (0, 1]$
- Ak $M_1 = \{(a, b), (a, a)\}$ a $M_2 = \{(a, b), (b, b)\}$ potom
  - $M_1 \cap M_2 = \{(a, b)\}$ a
  - $M_1 \cup M_2 = \{(a, b), (a, a), (b, b)\}$
- Nech x je počet hláv pri 3 hodoch mincou. Potom priestor je množina $\mathcal{M} = \{0, 1, 2, 3\}$.
- $\mathcal{M}^C = \emptyset$.
- Nech $M \subseteq \mathcal{M}$. Potom
  - $M \cup M^C = \mathcal{M}$,
  - $M \cap M^C = \emptyset$,
  - $M \cup \mathcal{M} = \mathcal{M}$,

- $M \cap \mathcal{M} = M$,
- $(M^C)^C = M$.

Užitočným nástrojom sú **DeMorganove zákony**. (Ukážte ich formálne ako cvičenie, pomôžte si Vennovými diagramami.)

- $(M_1 \cap M_2)^C = M_1^C \cup M_2^C$
- $(M_1 \cup M_2)^C = M_1^C \cap M_2^C$

Pre operácie prieniku $\cap$ a zjednotenia $\cup$ platia **distributívne zákony**

- $M_1 \cap (M_2 \cup M_3) = (M_1 \cap M_2) \cup (M_1 \cap M_3)$,
- $M_1 \cup (M_2 \cap M_3) = (M_1 \cup M_2) \cap (M_1 \cup M_3)$.

2.2 Pravdepodobnostný priestor

Teraz zadefinujeme dôležitý pojem - pravdepodobnostný priestor.

Uvažujme nasledovnú trojicu:

$$(\Omega, \mathcal{F}, P)$$

s týmito vlastnosťami:

- Ω je **priestor**, teda nejaká neprázdna množina. Bude to množina všetkých možných prípadov, ako môže dopadnúť experiment.
- $\mathcal{F}$ je množina podmnožín Ω. Teda prvky nachádzajúce sa v $\mathcal{F}$ sú množiny. $\mathcal{F}$ bude označovať množinu **udalostí**. No a udalostiam budeme chciet priraďovať pravdepodobnosť. [5]
- P je **pravdepodobnostná funkcia** alebo skrátene len **pravdepodobnosť**. Je to funkcia $\mathcal{F} \to [0, 1]$. [6] Každej udalosti priradí číslo medzi 0 a 1. Funkcia P musí spĺňať nasledovné vlastnosti:
  1. $P(A) \geq 0$ pre všetky udalosti A. Pravdepodobnosť je *nezáporná* funkcia.
  2. $P(\Omega) = 1$. Pravdepodobnosť je *zhora ohraničená funkcia číslom 1.*
  3. $P\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ pre akékoľvek disjunktné udalosti $A_1, A_2, A_3, \ldots$ Tejto vlastnosti sa hovorí aj *spočítateľná aditivita.* V prípade dvoch udalostí A a B máme, že pravdepodobnosť, že nastane udalosť A alebo udalosť B, teda $P(A \cup B)$ je rovná súčtu pravdepodobností týchto udalostí, teda $P(A) + P(B)$.

Túto trojicu $(\Omega, \mathcal{F}, P)$ nazývame **pravdepodobnostný priestor**.

Príklad 2.1 (Férová kocka)**.** Ako prvý príklad hádžme férovou kockou.

- $\Omega = \{1, 2, 3, 4, 5, 6\}$,
- $\mathcal{F} = 2^{\Omega}$, teda všetky možné podmnožiny Ω,
- $P(A) = \frac{|A|}{6}$, kde $|A|$ označuje počet prvkov množiny A.

Teraz

- Aká je pravdepodobnosť, že padne šestka? $P(\{6\}) = \frac{|\{6\}|}{6} = \frac{1}{6}$
- Aká je pravdepodobnosť, že párne číslo? $P(\{2, 4, 6\}) = \frac{|\{2,4,6\}|}{6} = \frac{3}{6} = \frac{1}{2}$

[5] Z formálneho hľadiska $\mathcal{F}$ musí byť $\sigma-$algebra podmnožín Ω. To znamená, že priraďovanie pravdepodobnosti musí byť vnútorne konzistentné. Technicky musia byť splnené tieto tri podmienky: (1) Ak $A \in \mathcal{F}$, teda ak viem priradiť udalosti A nejakú pravdepodobnosť, tak budem musieť vedieť priradiť pravdepodobosť aj udalosti A^C, teda, že A *nenastala.* (2) Ak $A_1, A_2, A_3, \dots \in \mathcal{F}$, potom aj $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$. Vysvetlíme si to na príklade dvoch množín: ak $A, B \in \mathcal{F}$, potom aj $A \cup B \in \mathcal{F}$, teda ak viem priradiť pravdepodobnosť udalostiam A a B, potom musím vedieť priradiť pravdepodobnosť aj udalosti $A \cup B$, teda že nastala udalosť A *alebo* udalosť B. (3) $\Omega \in \mathcal{F}$, teda viem priradiť pravdepodobnosť udalosti, že *niečo* nastalo. V rámci tohoto kurzu sa nebudeme podrobne venovať $\mathcal{F}$. Budeme predpokladať, že ide o korektnú $\sigma-$algebru. Do väčších podrobností sa zahĺbite v rámci kurzov *Teória miery a integrálu* a *Teória pravdepodobnosti.*

[6] Každý jeden krát sa nájde mnoho ľudí, ktorí nesprávne označujú, že pravdepodobnosť je funkcia z Ω do $[0, 1]$. Nie. Nie je. Je to funkcia $\mathcal{F} \to [0, 1]$.

Príklad 2.2 (Neférová kocka)**.** Teraz poďme hádzať pre zmenu neférovou kockou.

- $\Omega = \{1, 2, 3, 4, 5, 6\}$,
- $\mathcal{F} = 2^\Omega$, teda všetky možné podmnožiny Ω,
- $P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = \frac{1}{7}$, a $P(\{6\}) = \frac{2}{7}$.

Všimnime si, že teraz nám netreba zadefinovať pravdepodobnosť P pre úplne všetky možné udalosti. Využítím vlastnosti funkcie P vieme napríklad, že $P(\{1, 2\}) = P(\{1\} \cup \{2\}) = P(\{1\}) + P(\{2\}) = \frac{1}{7} + \frac{1}{7} = \frac{2}{7}$, kvôli tomu, že udalosti $\{1\}$ a $\{2\}$ sú *disjunktné*. Preto sme nepotrebovali zadefinovať samotnú $P(\{1, 2\})$. Poľahky sme si ju dopočítali.

Ďalej si môžeme všimnúť, že

$$\begin{aligned} P(\Omega) &= P(\{1\} \cup \{2\} \cup \{3\} \cup \{4\} \cup \{5\} \cup \{6\}) \\ &= P(\{1\}) + P(\{2\}) + P(\{3\}) + P(\{4\}) + P(\{5\}) + P(\{6\}) \end{aligned}$$

Preto musí platiť, že $P(\{1\}) + P(\{2\}) + P(\{3\}) + P(\{4\}) + P(\{5\}) + P(\{6\}) = 1$ inak by nebola splnená vlastnosť $P(\Omega) = 1$ a teda by funkcia P nemohla byť pravdepodobnosťou.

- Aká je pravdepodobnosť, že padne šestka? $P(\{6\}) = \frac{2}{7}$
- Aká je pravdepodobnosť, že párne číslo? $P(\{2, 4, 6\}) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{1}{7} + \frac{1}{7} + \frac{2}{7} = \frac{4}{7}$

Odteraz aj potom neskôr, pri akejkoľvek otázke/úlohe/cvičení budeme potichu predpokladať, že pravdepodobnostný priestor v rámci ktorého sa dá odpovedať na danú otázku/úlohu/cvičenie *existuje*.[7] [8]

2.3 Vlastnosti pravdepodobnosti

Dôkaz každého z týchto tvrdení skoro okamžite uvidíte, ak si nakreslíte Vennov diagram. Alebo keď si len nahlas prečítate, čo tieto tvrdenia hovoria:

- Pre každú udalosť A platí
  - $P(A^C) = 1 - P(A)$ - Ak je pravdepodobnosť, že dostanem chorobu 0.05, potom je pravdepodobnosť, že nedostanem chorobu $1 - 0.05 = 0.95$.
  - $0 \leq P(A) \leq 1$ - Pravdepodobnosť nutne musí ležať medzi nula a jeden.
- Ak $A \subseteq B$ potom platí $P(A) \leq P(B)$ - Pravdepodobnosť je monotónna.
- Pre každé dve udalosti A a B platí
  - $P(A \cap B^C) = P(A) - P(A \cap B)$ - Pravdepodobnosť, že nastane udalosť A a súčasne nenastane udalosť B je rovná pravdepodobnosti, že nastane A zníženej o pravdepodobnosť toho, že nastanú naraz obe udalosti A aj B.
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ - Pravdepodobnosť, že nastane udalosť A alebo udalosť B je súčet týchto pravdepodobností znížený o ich pravdepodobnosť prieniku (lebo tento zarátame dva razy).

Je dôležité vedieť tieto tvrdenia dokázať len z vlastností (1), (2), (3) pravdepodobnosti.

2.4 Kombinatorika

V mnohých stredoškolských príkladoch týkajúcich sa pravdepodobnosti sa spomínajú pojmy ako permutácie, kombinácie, s opakovaním/bez opakovania a podobne. Toto je častokrát zdrojom nepochopenia a chýb, lebo si študent/ka pre daný príklad *nejaký* z týchto pojmov vyberie a potom použije naučenú formulku. Nečudo,

[7] V matematike je dobrým zvykom pracovať s objektami o ktorých existencii niet pochýb.

[8] Najmä pri komplikovanejších prípadoch toto nie je vždy priamočiare. Pre potreby nášho kurzu to však bude postačujúce. Ak napríklad budeme uvažovať o pravdepodobnostnom priestore, ktorý zodpovedá tomu, že rovnomerne náhodne vyberieme nejaké číslo z intervalu $[0, 1]$, *nebudeme môcť uvažovať* $\mathcal{F} = 2^{[0,1]}$. Takáto voľba $\mathcal{F}$ by spôsobila, že nemôže existovať žiadna funkcia P, ktorá by spĺňala všetky 3 vlastnosti funkcie pravdepodobnosti a zároveň by každému intervalu priradila jeho dĺžku, akoby sme intuitívne očakávali. Konštrukcia vhodnej $\mathcal{F}$ je v tomto prípade veľmi prácna (8 krokový dôkaz) a pravdepodobnosť zadefinovaná na takomto pravdepodobnostnom priestore sa nazýva *Lebesgueova miera* na $[0, 1]$.

že mnohokrát nesprávne. Toto je nebezpečné, lebo si potom napríklad v praxi nebudú vedieť zrátať šancu, že budú mať čistú postupku alebo že si vytiahnu 3 modré guličky.

Vo väčšine prípadov ide len o aplikáciu nasledujúceho vzťahu:

$$P(A) = \frac{\text{počet úspešných pokusov (teda udalosť } A \text{ nastane)}}{\text{počet všetkých možných pokusov}}.$$

Takže stačí nám ak budeme vedieť efektívne počítať počet možností. Potom ich dáme do pomeru a máme pravdepodobnosť.

Využívajú sa nasledovné zákonistosti:

- Koľkými možnými spôsobmi môžeme usporiadať všetky objekty množiny $\{a_1, a_2, a_3, \dots a_n\}$ do postupnosti *keď záleží* na poradí?
  Na prvé miesto môžeme dať n rôznych objektov, na druhé miesto už len $n-1$ rôznych objektov (lebo sme si už jeden objekt minuli) a tak ďalej a na posledné miesto nám už zostal len 1 objekt. Dokopy je teda možností $n \cdot (n-1) \cdot \ \dots \ \cdot 1 = n!$.

- Koľkými možnými spôsobmi môžeme usporiadať k rôznych objektov z množiny $\{a_1, a_2, a_3, \dots a_n\}$ do postupnosti *keď záleží* na poradí?
  Na prvé miesto môžeme dať n rôznych objektov, na druhé miesto už len $n-1$ rôznych objektov (lebo sme si už jeden objekt minule) a tak ďalej a na k-te miesto nám už zostalo len $n-k+1$ rôznych objektov. Dokopy je teda možností $n \cdot (n-1) \cdot \ \dots \ \cdot (n-k+1) = \frac{n!}{(n-k)!}$.

- Koľkými možnými spôsobmi môžeme vybrať k rôznych objektov z množiny $\{a_1, a_2, a_3, \dots a_n\}$ *keď nezáleží* na poradí?
  Ak by bolo záležalo na poradí, bolo by ich $\frac{n!}{(n-k)!}$. Tu sme len použili predošlú úvahu. Ale v našom prípade nezáleží na poradí. A preto sú tam niektoré výbery viackrát. Konkrétne: rôznych usporiadaní k objektov je $k!$, teda toľkokrát ich je tam viacej. Odpoveď na našu otázku je preto $\frac{\frac{n!}{(n-k)!}}{k!} = \frac{n!}{(n-k)!k!} = \binom{n}{k}$.
  Zaujimavým dôsledkom je

$$(a+b)^n = (a+b) \cdot (a+b) \cdots (a+b) = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k},$$

nakoľko $\binom{n}{k}$ je počet možností ako možno vybrať k objektov z množiny n objektov.

Pri všetkých príkladoch typu karty/mince/kocky robíme, často implicitne, nejaké rozumné alebo polorozumné predpoklady. Keď napríklad povieme "hádžeme 3 krát kockou", tak tým v skutočnosti myslíme "*nezávisle* hádžeme 3 krát *férovou* kockou". Sada predpokladov, ktorá umožňuje prepísať danú úlohu do matematického jazyka a vypočítať je väčšinou zrejmá. V realite ale nikdy nehádžeme úplne nezávisle a žiadna kocka nie je úplne férová. Zjednodušenia pri kockách sú prirodzené a nekontroverzné. Pri reálnych príkladoch je však potreba explicitne vymenovať sadu predpokladov výrazne dôležitejšia. Matematický model, teda sada predpokladov, je totiž len teoreticky rámec úplne uzavretého systému, v ktorom sa veľmi dobre pracuje. Adekvátnosť modelu, teda to, ako dobre mapujú zjednodušenia realitu je aspekt, ktorý sa nedá charakterizovať matematickými prostriedkami.

Príklad 2.3. Raketoplán Challenger mal 2 raketové moduly, ktoré ho mali dostať na obežnú dráhu. Každý modul mal 3 gumové tesnenia, ktorých pravdepodobnosť zlyhania pri určitej teplote bola 0.1. Vypočítajte pravdepodobnosť úspešného letu.

Asi by ste hneď vypočítali: $(1-0.1)^6 \approx 0.531$. Je toto správne riešenie?

Samotné zadanie príkladu robí nejaké zjednodušenia, ktoré môžu byť neadekvátne.

- "Pravdepodobnosť zlyhania tesnenia je fixná, rovnakých 10% pre všetky tesnenia." - tesnenie bližšie k palivovému agregátu môže byť náchylnejšie na zlyhanie.

Zároveň na to, aby sme vypočítali tento príklad, musíme urobiť veľmi vážne zjednodušenia.

- "Úspešný let je vtedy ked nezlyhá žiadne tesnenie." - Let však môže zlyhať kvôli mnohým iným dôvodom.
- "Zlyhania jednotlivých tesnení sú nezávislé." - Možno zlyhanie jedného tesnenia ovplyvní zlyhanie iného. Napríklad ak sú blízko seba.

To či sú alebo nie sú tieto predpoklady rozumné matematik sám nevie posúdiť, treba na to pohľad experta na raketové motory.

Príklad 2.4 (Mnoho kombinátorických príkladov.)**.**

- Organizačný výbor plesu pozostával z 12 žiakov/čiek. Koľkými spôsobmi sa dá z nich vybrať moderátor, DJ, výzdobca a účtovník?
- Na plese pri stole s 10 stoličkami si štrngol pohárom jablčného muštu každý s každým. Koľko štrngov bolo počuť?
- Na tombole bolo predaných 350 lístkov. Kúpil som si 10, aká je pravdepodobnosť, že vyhrám aspoň jednu z 40 rôznych cien?
- Na šachovom turnaji s 12 účastníkmi hrá každý s každým 2 zápasy. Koľko zápasov sa odohrá?
- Hádžeme dvomi kockami - modrou a zelenou. Podrobne popíšte pravdepodobnostný priestor. Čo je väčšia šanca: že padne súčet 7 alebo že padne súčet jedno z čísel 2,3,11,12?
- Koľkými spôsobmi je možné hodiť troma kockami súčet 12?
- Aká je šanca, že pri dvoch nezávislých hodoch férovou mincou padne najprv hlava a potom znak?
- Dvanásť študentov/iek sa má rozdeliť na 4 skupiny po 3. Koľkými možnými spôsobmi to ide?
- Majme náhodne umiestnených 8 rôznych bodov v rovine. Koľko rôznych úsečiek (ktorých krajné body patria do množiny pôvodných 8 bodov) existuje, ktoré získame pospájaním týchto bodov? (Explicitne formulujte akékoľvek predpoklady navyše, ktoré robíte.)
- Koľko uhlopriečok má pravidelný 17 uholník?
- Koľko rôznych prirodzených čísel vieme reprezentovať práve piatimi ciframi?
- Majme vo vrecúšku písmenká $\{A, H, P, P, Y\}$. Aká je pravdepodobnosť, že si ich vytiahnem v poradí HAPPY?
- Majme vo vrecúšku písmenká $\{C, C, E, S, S, S, U\}$. Aká je pravdepodobnosť, že si ich vytiahnem v poradí SUCCESS?
- Organizačný výbor plesu pozostával z 12 žiakov/čiek. Koľkými spôsobmi sa dá z nich vybrať 2 moderátori, DJ, 3 výzdobcovia a účtovník?
- V obchode majú 3 druhy cukríkov: modré, zelené a bez cukru. Máme finančné prostriedky na nákup 6 cukríkov. Koľko možností máme, ak minieme všetky peniaze?
- Koľko riešení má $a + b + c = 100$, kde a, b, c sú prirodzené čísla?
- Koľkými spôsobmi (čo sa týka len samotného počtu kariet) je možné rozdať 54 hracích kariet 5 hráčom?
- Aká je šanca, že z perfektne zamiešaného balíčka kariet vyberieme práve 2 esá?

V rôznych učebniciach pravdepodobnosti sa môžete stretnúť s tým, že rôzne typy prípadov sú systematizované. Hovorí sa o permutáciách, kombináciách (s opakovaním alebo bez opakovanie), variáciách a podobne. Môj názor je, že s využívaním tohoto názvoslovia je spojené riziko, že študenti prestanú rozmýšľať a budú sa len snažiť kategorizovať daný príklad do schémy v ktorej už potom mechanicky poznajú vzťah/vzorec, ktorý treba použiť. Názor na to, ako aj spôsob implementácie pri praktickej výučbe nechám na Vašom posúdení.

Zopár (možno užitočných) poznámok:

- Ak neviete pohnúť s riešením, skúste začať mechanicky vypisovať to ako môže dopadnúť experiment.

- Pravdepodobnosti sčítavame len vtedy keď počítame pravdepodobnosť, že nastane jedna alebo druhá udalosť, ktoré sú vzájomne sa vylučujúce (mutually exclusive), teda keď nemôžu nastať naraz. Vo všetkých ostatných prípadoch sa pri sčitavaní pravdepodobností dopustíme chyby, že zarátame tú istú možnosť viacejkrát. (napr. Aká je šanca, že padne na kocke číslo 1 alebo číslo 2 ?) Toto je len aplikácia spočitateľnej aditivity pravdepodobnosti, teda vlastnosti (3).
- Pravdepodobnosti násobíme vtedy, keď ide o nezávislé udalosti. Teda keď jedna udalosť nijakovsky neovplyvní nastatie inej. Viď ďalšia kapitola. (napr. Aká je pravdepodobnosť, že pri prvom hode padne 6ka a pri druhom (nezávislom) hode padne 6ka?)
- Náhodné násobenie a sčítavanie pravdepodobností vedia k správnemu výsledku len s malou pravdepodobnosťou (áno, toto je vtip). Tipovaním sa ďaleko nedostaneme.
- Niekedy je ľahšie počítanie pravdepodobnosti, že udalosť nenastane, ako že nastane. Vtedy použijeme $P(A) = 1 - P(A^C)$.
- Vennove diagramy sú užitočné. Pri počítaní pravdepodobnosti pre viac ako dve udalosti veľmi.

2.5 Zhrnutie

Základným objektom s ktorým budeme pracovať je **pravdepodobnostný priestor**. Je to trojica: Ω hovorí ako môže dopadnúť experiment, $\mathcal{F}$ je množina všetkých podmnožín, ktorým priraďujeme pravdepodobnosť a P je samotná pravdepodobnostná funkcia, ktorá musí byť spočitateľne aditívna. Keď máme konkrétny príklad, je dobré si uvedomiť, čo vlastne príslušný pravdepodobnostný priestor je. Vnesie to do nášho rozmýšlania poriadok. Toto je dôležité, aby sme vedeli správne zrátať počet "úspešných" a počet "všetkých" spôsobov ako môže dopadnúť experiment a tým pádom aj pravdepodobnosť. Rovnaký problém môže byť uchopený cez rôznu formuláciu pravdepodobnostného priestoru.

2.6 Cvičenia

Cvičenie 2.1. Aký je rozdiel medzi $\{a, b\}$ a (a, b)?

Cvičenie 2.2. Aký je rozdiel medzi $A \in \mathcal{M}$ a $A \subseteq \mathcal{M}$?

Cvičenie 2.3. Nájdite komplement množiny M vzhľadom na priestor $\mathcal{M}$.

(a) $M = \{1, 2, 3, 4\}$ a $\mathcal{M} = \{1, 2, 3, 4, 5, 6\}$

(b) $M = \{(x, y) : x^2 + y^2 < 2\}$ a $\mathcal{M} = \{(x, y) : x^2 + y^2 \leq 4\}$

(c) $M = \{x : 0 < x < 1\}$ a $\mathcal{M} = [0, 1]$

Cvičenie 2.4. Uvažujme dve udalosti A a B, také, že $P(A) = 0.3$ a $P(B) = 0.6$.

Vypočítajte $P(A^C \cap B)$ ak viete, že platí jedna z nasledovných podmienok:

(a) Udalosti A a B sú disjunktné;

(b) A je podmnožinou B;

(c) pravdepodobnosť toho, že naraz nastanú udalosti A aj B je 0.1.

Cvičenie 2.5. Počítač bol naprogramovaný na to aby vypočítal rôzne pravdepodobnosti. Pripojte k týmto numerickým odpovediam správny slovný opis (môže byť viacero správnych)

Výstup z programu:

(a) -50%

(b) 0%

(c) 10%

(d) 50%

(e) 90%
(f) 100%
(g) 200%

Slovný popis:

(i) Rovnaká šanca, že to nastane ako, že to nenastane.
(ii) Je veľmi pravdepodobné, že to nastane ale nie isté.
(iii) Toto nenastane.
(iv) Toto môže nastať ale šanca je malá.
(v) Toto určite nastane.
(vi) V mojom programe je nejaká chyba.

Cvičenie 2.6. V triede je 60% žien a 40% mužov. Vieme, že v tejto triede 70% ľudí študuje matematiku.

Percento žien študujúcich matematiku môže byť najmenej _________% a najviac _________%.

Cvičenie 2.7. Hádžeme 6 krát mincou. Tu sú dve možné situácie (H - hlava, Z - znak):

(i) H Z Z H Z H
(ii) H H H H H H

Ktoré z týchto tvrdení je správne? Vysvetlite.

(a) Udalosť (i) je pravdepodobnejšia.
(b) Udalosť (ii) je pravdepodobnejšia.
(c) Udalosti (i) a (ii) sú rovnako pravdepodobné.

Cvičenie 2.8. Nech je silne zamorenom prostredí pravdepodobnosť bakteriálnej infekcie 0.4 a pravdepodobnosť nákazy spôsobej vírusom nech je 0.8. Aká je najväčšia a najmenšia pravdepodobnosť, že človek sa nakazí aj baktériou aj vírusom?

Cvičenie 2.9. Hodíme kocku 6 krát. Balíček kariet perfektne zamiešame.

(a) Šanca, že na prvom hode padne 6ka alebo na poslednom hode padne 6ka je _________.
(b) Šanca, že na prvom hode padne 6ka a na poslednom hode padne 6ka je _________.
(c) Šanca, že na vrchu balíčka je červené eso alebo na spodu balíčka je červené eso je _________.
(d) Šanca, že na vrchu balíčka je červené eso a na spodu balíčka je červené eso je _________.

Cvičenie 2.10. Uvažujme nasledovnú situáciu: jedenkrát hodíme férovou mincou a nezávisle od toho jedenkrát hodíme férovou kockou.

(a) Zostrojte pravdepodobnostný priestor, ktorý zodpovedá tejto situácii.
(b) Aká je pravdepodobnosť, že padne hlava a zároveň nepárne číslo?

Cvičenie 2.11. Uvažujme nasledovnú situáciu: trikrát za sebou hodíme férovou kockou. Jednotlivé hody sú nezávislé.

(a) Zostrojte pravdepodobnostný priestor, ktorý zodpovedá tejto situácii.
(b) Aká je pravdepodobnosť, že padne súčet väčší ako 5?

Cvičenie 2.12. Máme 20 kariet. Desať z nich je modrých, očíslovaných od 1 po 10, desať z nich je červených a očíslovaných od 1 po 10. Náhodne si jednu vyberieme. Vezmime si nasledovné udalosti:

- A - vyberieme si kartu s párnym číslom.
- B - vyberieme si modrú kartu.
- C - vyberieme si kartu s číslom menším ako 5.

Zostrojte pravdepodobnostný priestor zodpovedajúci tejto situácii a opíšte nasledovné udalosti:

(a) $A \cap B \cap C$
(b) $B \cap C^C$
(c) $A \cup B \cup C$
(d) $A \cap (B \cup C)$
(e) $A^C \cap B^C \cap C^C$

Cvičenie 2.13. Hodíme 3 férové kocky. Vezmime si nasledovné udalosti:

- A - na prvej kocke hodíme párne číslo.
- B - na druhej kocke hodíme párne číslo.
- C - na tretej kocke hodíme párne číslo.

Pomocou týchto udalostí vyjadrite:

(a) Udalosť, že na všetkých troch kockách je párne číslo.
(b) Udalosť, že na žiadnej kocke "je" párne číslo.
(c) Udalosť, že aspoň na jednej kocke je nepárne číslo.
(d) Udalosť, že najviac na dvoch kockách je nepárne číslo.

Cvičenie 2.14. Náhodne vyberieme tri rôzne prirodzené čísla z množiny $\{1, 2, 3, \cdots, 20\}$.

(a) Aká je pravdepodobnosť, že ich súčet bude párne číslo?
(b) Aká je pravdepodobnosť, že ich súčin bude párne číslo?

Cvičenie 2.15. Vo vreci je 50 súčiastok, z nich dve sú chybné. Náhodne načrieme rukou vyberieme 5 z nich.

(a) Aká je pravdepodobnosť, že aspoň jedna z nich bude chybná?
(b) Koľko súčiastok musíme vybrať aby šanca, že nájdeme aspoň jednu chybnú súčiastku bola väčšia ako 50% ?

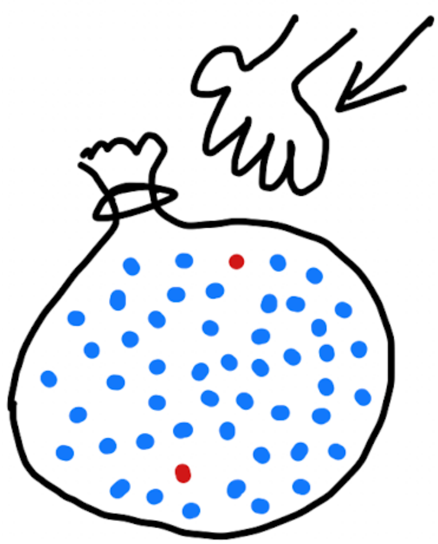


Figure 4: Ilustrácia načierania rukou do vreca s 50 súčiastkami.

Cvičenie 2.16. Ukážte, že zo *spočitateľnej aditivity* (vlastnosť (3) pravdepodobnosti) vyplýva *konečná aditivita*, teda, že platí $P\left(\cup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$ pre akékoľvek disjunktné udalosti $A_1, A_2, A_3, \ldots, A_n$

Cvičenie 2.17. Formálne ukážte vlastnosti pravdepodobnosti z časti 2.3.

Cvičenie 2.18. Ukážte, že pre hocijaké dve udalosti A a B je pravdepodobnosť, že nastane práve jedna z udalostí A a B je daná nasledovným výrazom

$$P(A) + P(B) - 2P(A \cap B).$$

Cvičenie 2.19. Nech je $A_1, A_2, \dots$ akákoľvek postupnosť udalostí a nech $B_1, B_2, \dots$ je postupnosť definovaná nasledovne: $B_1 = A_1$, $B_2 = A_1^C \cap A_2$, $B_3 = A_1^C \cap A_2^C \cap A_3$, $B_4 = A_1^C \cap A_2^C \cap A_3^C \cap A_4$ a tak ďalej. Ukážte, že pre všetky $n = 1, 2, 3, \dots$ platí

$$P\left(\cup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(B_i).$$

Cvičenie 2.20. Nech A, B a D sú nejaké udalosti, také, že $P(A \cup B \cup D) = 0.3$. Aká je hodnota $P(A^C \cap B^C \cap D^C)$?

Cvičenie 2.21 (Bonferroniho nerovnosť)**.** Ukážte, že pre udalosti $A_1, A_2, \dots, A_n$ platí

$$P\left(\cup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i)$$

a

$$P\left(\cap_{i=1}^{n} A_i\right) \geq 1 - \sum_{i=1}^{n} P(A_i^C).$$

Cvičenie 2.22. Nech A, B a C sú nejaké udalosti. Použitím Vennoveho diagramu ukážte, že

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &\quad - P(A \cap B) - P(B \cap C) - P(A \cap C) \\ &\quad + P(A \cap B \cap C). \end{aligned}$$

Cvičenie 2.23. Majme vrecúško s farebnymi guličkami. Konkrétne sa v ňom nachádza č červených guličiek, b bielych guličiek a m modrých guličiek. Postupne vyberáme po jednej guličke z vrecúška tak, že ich nedávame naspäť. Aká je pravdepodobnosť, že vytiahneme všetky červené guličky ešte predtým, než vytiahneme nejakú bielu?

Cvičenie 2.24 (*Princíp zapojenia a vypojenia)**.** Nech $A_1, A_2, \dots, A_n$ sú nejaké udalosti. Ukážte, že platí

$$\begin{aligned} P\left(\cup_{i=1}^{n} A_i\right) &= \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) \\ &\quad + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \sum_{i<j<k<l} P(A_i \cap A_j \cup A_k \cap A_l) + \dots \\ &\quad + (-1)^{n+1} P(A_1 \cap A_2 \cap A_3 \cap \dots \cap A_n). \end{aligned}$$

–> –>

–> –> –>

3 Podmienená pravdepodobnosť

Podmienená pravdepodobnosť je kľúčovým pojmom v teórii pravdepodobnosti. Vďaka nej môžeme skúmať súvzťažnosť rôznych udalostí - ako nastatie jednej udalosti ovplyvní pravdepodobnosť nastatia inej. Je neprekvapivé, že rôzne udalosti nenastávajú vo vákuu, ale navzájom spolu môžu súvisieť. A o tom to teraz bude. Ako tento súvis vyjadriť, porozumieť mu a matematicky ho formalizovať.

3.1 Čo je to podmienená pravdepodobnosť

Nech A a B sú udalosti a nech naviac $P(B) > 0$. **Podmienenou pravdepodobnosťou** udalosti A za podmienky, že nastala udalosť B, nazývame nasledovnú kvantitu:

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

Potrebovali sme uvažovať, aby platilo $P(B) > 0$, nakoľko v inom prípade by podmienená pravdepodobnosť nebola dobre definovaná.[9]

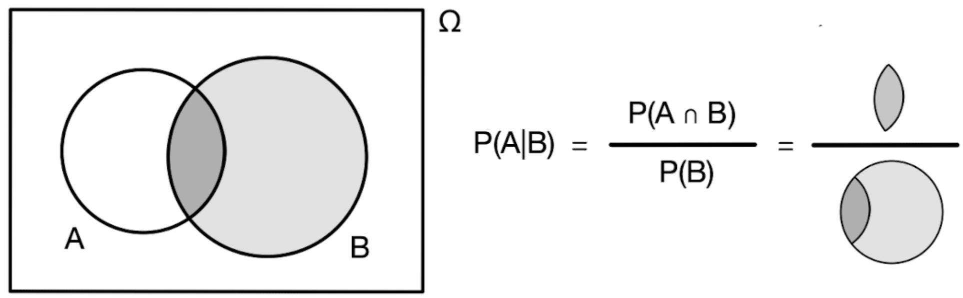


Figure 5: Ilustrácia podmienenej pravdepodobnosti. Pravdepodobnosť jednotlivých množín je reprezentovaná jej plochou.

Podmienená pravdepodobnosť je ohromne užitočný koncept. Hovorí nám napríklad o tom, ako v zmysle poznania novej informácie (dozvedeli sme sa, že nastala udalosť B) meníme to, čo si myslíme o pravdepodobnosti toho, že nastane udalosť A. Na začiatku by sme si boli mysleli, že šanca, že Slovensko vyhrá na MS v ľadovom hokeji je malá ($P(A)$). Avšak vidiac náš fantastický výkon v základnej skupine B sa teraz už nádejame, že zlato nie je až také márne ($P(A|B) > P(A)$).[10]

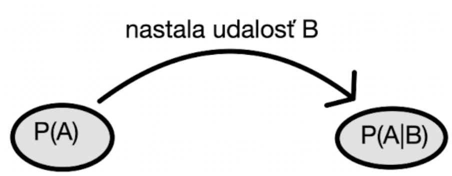


Figure 6: Vo svetle novej informácie (udalosť B nastala) zmeníme pravdepodobnosť, že udalosť A nastane.

Príklad 3.1. Hádžeme dvoma férovými kockami. Pravdepodobnosť, že na nich padne súčet 4, $P(A) = \frac{3}{36} = \frac{1}{12}$. Aká je pravdepodobnosť, že padne súčet 4, ak vieme, že na prvej kocke padlo číslo 3?

Označme si rôzne udalosti:[11]

- A - na dvoch kockách padne súčet 4

[9] Neskôr, ale v inom kurze sa dozvieme, že sa dá podmieňovať aj udalosťami, ktorých pravdepodobnosť je priamo nula. Toto je zaujímavé uvažovať vo finančných aplikáciách. Kým pravdepodobnosť, že v nejakom konkrétnom okamihu bude cena ropy *presne* 100\$ za barel je nula, neznamená to, že je neužitočné uvažovať o tom, aká by bola pravdepodobnosť nejakej inej udalosti, ak by cena ropy naozaj bola *práve* 100\$ za barel. A presne na to nám bude treba poznať trochu viacej štruktúry ako sa učíme teraz.

[10] Update 18.2.2022: Bronz nie je až taký márny.

[11] Akonáhle máme udalosti poriadne označené, polovicu riešenia máme za sebou. Tento prechod od písaného textu do matematického zápisu je kľúčový ale našťastie sa dá do veľkej miery natrénovať.

- B - na prvej kocke padne číslo 3

Použitím vzťahu pre podmienenú pravdepodobnosť dostávame

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6}.$$

Príklad 3.2. Majme triedu, v ktorej je 60% dievčat a 40% chlapcov. Nech 30% dievčat má dlhé vlasy a nech 20% chlapcov má dlhé vlasy. Náhodne vyberieme jednu osobu z tejto triedy. Aká je pravdepodobnosť, že má dlhé vlasy?

Uvažujme nasledovné značenie

- Ž - vybraná osoba je dievča
- M - vybraná osoba je chlapec
- D - vybraná osoba má dlhé vlasy

Zaujíma nás $P(D)$, ale k dispozícii máme len $P(\check{Z}) = 0.6, P(M) = 0.4$, ale aj $P(D|\check{Z}) = 0.3$ a $P(D|M) = 0.2$.

Hneď vieme, $P(D \cap \check{Z}) = P(D|\check{Z})P(\check{Z}) = 0.3 \cdot 0.6 = 0.18$ a $P(D \cap M) = P(D|M)P(M) = 0.2 \cdot 0.4 = 0.08$.

Nakoľko udalosti M a Ž sú disjunktné ($\check{Z} = M^C$), musia byť aj udalosti $D \cap \check{Z}$ a $D \cap M$ tiež disjunktné. Podľa spočitateľnej aditivity pravdpodobnosti teda máme:

$$P(D) = P(\{D \cap \check{Z}\} \cup \{D \cap M\}) = P(D \cap \check{Z}) + P(D \cap M) = 0.18 + 0.08 = 0.26.$$

3.2 Bayesova veta

Bayesova veta je jedným z najdôležitejších výsledkov v teórii pravdepodobnosti.

Uvažujme A_1 a A_2 disjunktné udalosti také, že $\Omega = A_1 \cup A_2$. To znamená, že nastane A_1 alebo $A_2 = A_1^C$, teda $P(A_1) + P(A_2) = 1$. Pre takéto množiny platí

$$A = (A \cap A_1) \cup (A \cap A_2).$$

Nakoľko aj množiny $A \cap A_1$ a $A \cap A_2$ sú tiež disjunktné, vďaka *spočitateľnej aditivite* pravdepodobnosti dostávame

$$P(A) = P(A \cap A_1) + P(A \cap A_2).$$

Využitím definície podmienenej pravdepodobnosti dospejeme k

$$P(A) = P(A \cap A_1) + P(A \cap A_2) = P(A|A_1)P(A_1) + P(A|A_2)P(A_2).$$

Aplikovovaním podmienenej pravdepodobnosti do pôvodného vzťahu máme

$$P(A_1|A) = \frac{P(A_1 \cap A)}{P(A)} = \frac{P(A|A_1)P(A_1)}{P(A|A_1)P(A_1) + P(A|A_2)P(A_2)}.$$

Tomuto výsledku hovoríme **Bayesova veta**. V prípade, že máme viacej disjunktných udalostí $A_1, A_2, \cdots, A_n$, takých, že $\cup_{i=1}^n A_i = \Omega$ (takéto množiny nazývame **rozklad** množiny Ω), podobným dôvodením dostávame:

$$P(A_1|A) = \frac{P(A_1 \cap A)}{P(A)} = \frac{P(A|A_1)P(A_1)}{\sum_{i=1}^n P(A|A_i)P(A_i)}.$$

Prečo je podmienená pravdepodobnosť a teda aj Bayesova veta dôležitá?

Príklad 3.3. Aby sme boli na tepe dňa, pozrime sa na príklad prudko aktuálny (začiatkom 2022). Uvažujme skríningové testovanie na nejakú chorobu a použime nasledovné značenie:

- $\oplus$ - nech označuje pozitívny test,

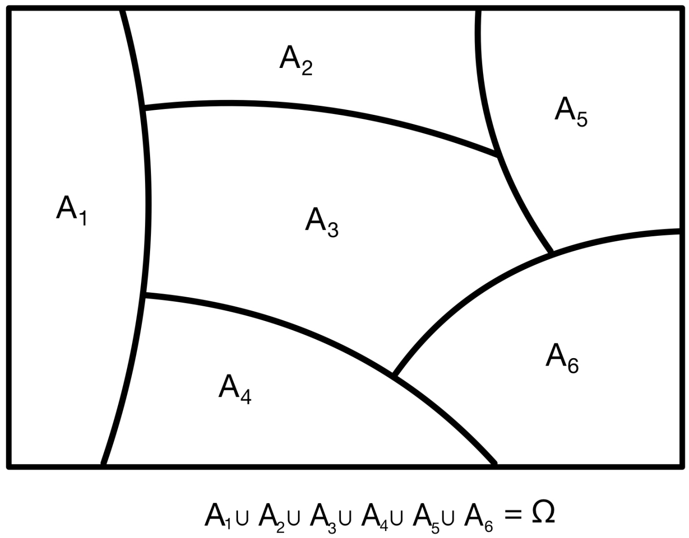


Figure 7: Ilustrácia rozkladu množiny.

- $\ominus$ - nech označuje negatívny test,
- Z - nech označuje, že človek je zdravý,
- CH - nech označuje, že človek je chorý.

Na základe skúmania odpadových vôd vieme, že prítomnosť choroby v populácii je 1%, teda $P(CH) = 0.01 = 1 - P(Z)$. Majme test o ktorom vieme, že pre chorého človeka dá pozitívny výsledok v 90% prípadov, tomuto sa hovorí *senzitivita testu*, matematicky to zapíšeme ako $P(\oplus|CH) = 0.9$. Tento test dá naviac správny negatívny výsledok v prípade zdravého človeka v 99% prípadov (*špecificita testu*), takže $P(\ominus|Z) = 0.99 = 1 - P(\oplus|Z)$. Teraz praktický príklad: náhodne vybraného človeka budeme testovať, test vyjde pozitívny. Aká je pravdepodobnosť, že tento človek je skutočne pozitívny? Ešte predtým, než sa pozrieme nižšie, predstavme si túto situáciu a tipnime si výsledok.

Pozrime sa na to bližšie:

$$
\begin{aligned}
P(CH|\oplus) &= \frac{P(CH \cap \oplus)}{P(\oplus)} = \frac{P(\oplus|CH)P(CH)}{P(\oplus|CH)P(CH) + P(\oplus|Z)P(Z)} \\
&= \frac{0.9 \cdot 0.01}{0.9 \cdot 0.01 + 0.01 \cdot 0.99} = 0.476.
\end{aligned}
$$

Teda je približne 48%. Menej než polovica. Toto je výrazne menej ako väčšina ľudí čaká (tip: vyskúšajte to aj na lekárov). Tuto je nesmierne dôležité zdôrazniť, že išlo o *skríningové* testovanie.[12]

Ak by sme napríklad testovali len ľudí s príznakmi, už by neplatilo, že $P(CH) = 0.01$, ale bolo by to výrazne viacej. Ľudia s príznakmi totiž nie sú rovnakí ako tí náhodne vybraní z populácie.[13]

Vráťme sa naspäť k tomu, prečo je tých 48% tak prekvapivo veľa. Je to člen $P(\oplus|Z)P(Z)$ v menovateli v zlomku na výpočet $P(CH|\oplus)$. Síce je pravdepodobnosť $P(\oplus|Z)$, teda falošnej pozitivity, veľmi malá, ale zdravej populácie je veľká väčšina $P(Z) = 0.99$.

[12] Tieto čísla sú veľmi blízke tým zo skríningového testovania na školách na COVID19 v ČR v januári 2022.

[13] Skríningové testovanie sa odporúča aj pri niektorých typoch rakoviny, kde je skoré zachytenie dôležité pre dobrú prognózu. Vždy však treba brať do úvahy náklady spojené s množstvom falošne pozitívnych prípadov alebo s množstvom prípadov, kde pacient podstúpil náročnú liečbu napriek tomu, že by sa rakovina nemusela počas celého života pretaviť do problémov (keď človek umrie na inú diagnózu).

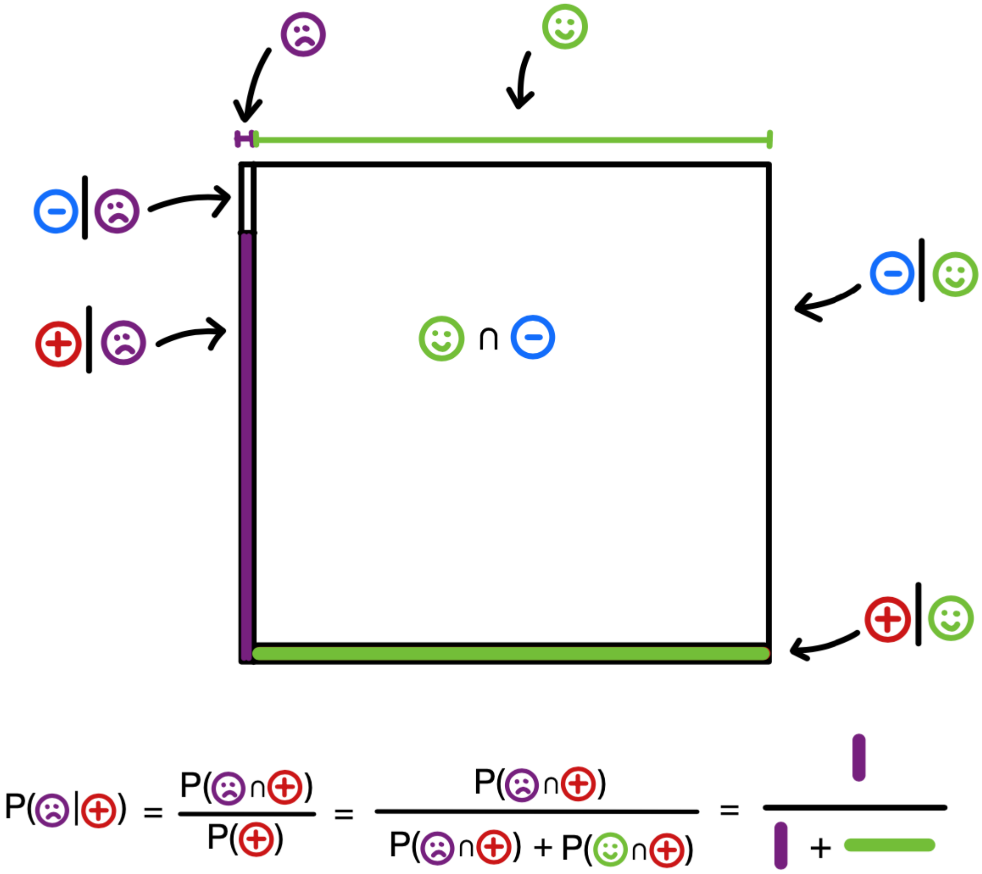


Figure 8: Ilustrácia Bayesovej vety na príklade s testovaním.

Príklad 3.4. Majme dve vrecká s guličkami. V prvom je 7 červených a 2 modré guličky. V druhom vrecku je 5 červených a 9 modrých guličiek. Náhodne vyberieme jedno z vreciek (s pravdepodobnosťou 1/2) a z neho jednu z guličiek. Ukáže sa, že je červená. Aká je pravdepodobnosť, že táto gulička pochádza z prvého vrecka?

Označme si udalosť, že vyberáme z vrecka 1 ako V_1 a že z vrecka 2 ako V_2. Na základe predpokladu vieme, že $P(V_1) = P(V_2) = \frac{1}{2}$. Ďalej označme udalosť, že vyberieme červenú guličku ako Č. Priamočiarym použitím Bayesovej vety dostávame:

$$P(V_1|\text{Č}) = \frac{P(\text{Č}|V_1)P(V_1)}{P(\text{Č}|V_1)P(V_1) + P(\text{Č}|V_2)P(V_2)} = \frac{\frac{7}{9}\frac{1}{2}}{\frac{7}{9}\frac{1}{2} + \frac{5}{14}\frac{1}{2}} \approx 0.685.$$

3.3 Nezávislosť

Nezávislosť dvoch udalostí nám hovorí čosi o tom, ako tieto udalosti nastávajú. To, či nastane jedna udalosť vôbec nijakovsky nesúvisí s tým, či nastane druhá udalosť. Teda inými slovami, nenastávajú naraz ani nenastávajú opačne, skrátka a dobre, obe udalosti si žijú svoj vlastný život a ich nástávanie/nenastávanie sa deje úplne oddelene. Tu treba dať pozor. Nehovoríme o žiadnej príčinnosti. V období, keď je najviac útokov žralokov na človeka, sú aj rekordné predaje zmrzliny. Tieto udalosti sú závislé. Ale to neznamená, že nejak spolu kauzálne súvisia, vôbec nie. Skôr ide o to, že tieto udalosti sú prepojené cez nejakú inú udalosť. V lete, keď je pekne, ľudia sú vonku a kúpu sa, je prirodzene väčšia šanca útoku žraloka ako v zime, keď sú turisti doma.

Dve udalosti A a B nazývame **nezávislé** ak platí

$$P(A \cap B) = P(A)P(B).$$

Súbor udalostí $A_1, A_2, \cdots$ nazývame **nezávislý** ak platí

$$P(A_{i_1} \cap \cdots \cap A_{i_j}) = P(A_{i_1})P(A_{i_2}) \cdots \cdot P(A_{i_j}).$$

pre akýkoľvek konečný podsúbor udalostí $A_{i_1}, A_{i_2} \cdots A_{i_j}$.

Pre *nezávislé* udalosti A, B platí, že vedomosť o tom, že udalosť B nastala nijakovsky neovplyvní pravdepodobnosť, že nastane udalosť A.

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

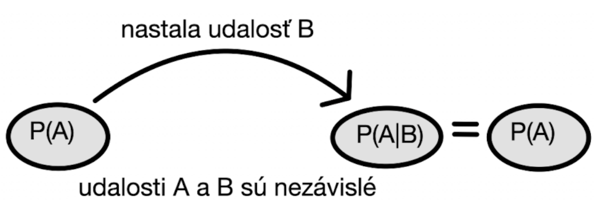


Figure 9: Vo svetle novej informácie (udalosť B nastala) sa pravdepodobnosť, že udalosť A nastane, nezmení, ak sú tieto udalosti nezávislé.

Nezávislosť udalostí A a B sa dá preto alternatívne zadefinovať aj pomocou vzťahu

$$P(A|B) = P(A).$$

To, že dve udalosti sú závislé, znamená, že nastatie jednej udalosti nesie nejakú informáciu o tej druhej udalosti. Môže ju zvýšiť alebo znížiť.

Koncept nezávislosti sme doteraz neformálne používali pri rôznych príkladoch s kockami, s mincami, či s kartami. Teraz sme si tento pojem formalizovali.

Príklad 3.5. Uvažujme pravdepodobnostný priestor $([0,1], \mathcal{F}, P)$, kde P priradí každému intervalu jeho dĺžku a $\mathcal{F}$ je množina všetkých udalostí, ktorým vieme priradiť pravdepodobnosť.[14] Majme nasledovné dve udalosti $A = \left(\frac{1}{2}, \frac{3}{4}\right)$ a $B = \left(0, \frac{2}{3}\right)$. Sú tieto udalosti nezávislé?

Na prvý pohľad by sa možno zdalo, že nie. Lebo nie sú disjunktné. Alebo? Pozrime sa na to bližšie.

$$
\begin{aligned}
P(A \cap B) &= P\left(\left(\frac{1}{2}, \frac{3}{4}\right) \cap \left(0, \frac{2}{3}\right)\right) = P\left(\left(\frac{1}{2}, \frac{2}{3}\right)\right) = \frac{1}{6} \\
P(A)P(B) &= P\left(\left(\frac{1}{2}, \frac{3}{4}\right)\right) P\left(\left(0, \frac{2}{3}\right)\right) = \frac{1}{4}\frac{2}{3} = \frac{1}{6}
\end{aligned}
$$

Takže tieto udalosti A a B vskutku *sú nezávislé*. No lebo spĺňajú definíciu nezávislosti. Pre disjunktné množiny platí, že pravdepodobnosť zjednotenia množín je súčet pravdepodobností množín. Toto je však niečo iné ako nezávislosť.

3.4 Podmienená pravdepodobnosť je tiež pravdepodobnosť

Na podmienenú pravdepodobnosť sa môžeme pozerať ako na pravdepodobnostnú funkciu na tom istom pravdepodobnostom priestore.

$$
P^*(A) = P(A|B) = \frac{P(A \cap B)}{P(B)}.
$$

Poľahky môžeme overiť tri vlastnosti, ktoré musí spĺňať každá pravdepodobnostná funkcia:

(1) $P^*(A) \geq 0$ pre všetky udalosti A platí triviálne, lebo $P(A \cap B) \leq P(B)$ kvôli tomu, že $A \cap B \subset B$
(2) $P^*(\Omega) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$.
(3) $P^*\left(\cup_{i=1}^{\infty} A_i\right) = \frac{P(\cup_{i=1}^{\infty} A_i \cap B)}{P(B)} = \sum_{i=1}^{\infty} \frac{P(\cup_{i=1}^{\infty} A_i \cap B)}{P(B)} = \sum_{i=1}^{\infty} P^*(A_i)$ pre akékoľvek disjunktné udalosti $A_1, A_2, A_3, \ldots$. Prostredná nerovnosť platí, lebo aj množiny $A_1 \cap B, A_2 \cap B, A_3 \cap B, \cdots$ sú tiež disjunktné a len sme využili vlastnosť spočitateľnej aditivity pôvodnej pravdepodobnostnej funkcie P.

Pre fixnú udalosť, pre ktorú platí $P(B) > 0$, sa teda podmienená pravdepodobnosť správa ako každá iná pravdepodobnostná funkcia. Všetky vlastnosti, ktoré sme odvodili pre pôvodnú pravdepodobnostú funkciu z prvotných princípov, budú preto platné aj pre podmienenú pravdepodobnosť.[15]

3.5 Zhrnutie

Podmienená pravdepodobnosť je nástroj, ktorý nám umožnuje kvantifikovať, ako meníme pravdepodobnosť nejakej udalosti vo svetle novej informácie. **Bayesova veta** nám dáva vzťah na výpočet podmienenej pravdepodobnosti. Častokrát vedie k výsledkom, ktoré sú na prvý pohľad prekvapivé.

3.6 Cvičenia

Cvičenie 3.1. Aký má vplyv

- zvýšenie senzitivity $P(P|CH)$,
- zvýšenie špecificitu $P(N|Z)$,
- zvýšenie incidencie choroby $P(CH)$

na $P(CH|P)$? Vysvetlite aj na intuitívnej úrovni.

Cvičenie 3.2. Hádžeme tromi nezávislými férovými kockami. Vypočítajte, o koľko percent sa zväčší/zmenší pravdepodobnosť toho, že padne súčet 12 po tom, ako sa dozvieme, že na jednej z nich padlo číslo 4.

[14] Pravdepodobnosť na takomto pravdepodobnostnom priestore sa nazýva *Lebesgueova miera* na $[0,1]$ a v tomto prípade platí $\mathcal{F} \neq 2^{\Omega}$ a $\mathcal{F}$ je množina Lebesgueovsky merateľných množín. Do väčších podrobností sa zahĺbite v rámci kurzov *Teória miery a integrálu* a *Teória pravdepodobnosti*.

[15] Neskôr sa tiež dozvieme, že množina podmnožín Ω, teda $\mathcal{F}$, môže byť pri podmieňovaní vo všeobecnosti menšia ako celá množina všetkých podmnožín 2^{Ω}.

Cvičenie 3.3. Ukážte, že ak sú udalosti A a B nezávislé, potom sú aj udalosti A a B^C nezávislé.

Cvičenie 3.4 (*). Nájdite taký príklad pravdepodobnostného priestoru a udalostí A_1, A_2, A_3, kde súčasne platí

- A_1 a A_2 sú nezávislé,
- A_2 a A_3 sú nezávislé,
- A_1 a A_3 sú nezávislé,
- A_1, A_2, A_3 nie sú nezávislé.

Cvičenie 3.5. Uveďte príklad pravdepodododobnostného priestoru a nejakých udalostí A_1, A_2, A_3, A_4 tak, aby platilo súčasne

- $P(A_1|A_2) > P(A_1)$,
- $P(A_2|A_3) > P(A_2)$,
- $P(A_3|A_4) > P(A_4)$.

Ak sa to nedá, tak to dokážte.

Cvičenie 3.6. Nech je

- pravdepodobnosť, že sneží 20%,
- pravdepodobnosť nehody 10%,
- podmienená pravdepodobnosť nehody, ak sneží, 40%.

Aká je pravdepodobnosť, že sneží, ak vieme, že nastala nehoda?

Cvičenie 3.7. Máme 3 bezpečnostné systémy, tri rôzne vrstvy ochrany nášho počítačového programu. Ak zlyhá prvý (s pravdepodobnosťou 0.03), aktivuje sa druhý, ktorý zlyhá s pravdepodobnosťou 0.01. V prípade zlyhania druhého je pravdepodobnosť zlyhania posledného 0.05. Aká je pravdepodobnosť, že zlyhajú všetky tri bezpečnostné systémy?

Cvičenie 3.8. Dve výrobné linky produkujú ten istý typ súčiastky. Prvá linka vyrobí 10000 súčiastok týždenne a 1000 je chybných. Druhá linka vyrobí 20000 súčiastok a z nich je 1500 chybných. Aká je šanca, že náhodne vybratá súčiastka bude chybná? Vybrali sme chybnú súčiastku. Aká je pravdepodobnosť, že pochádza z prvej výrobnej linky?

Cvičenie 3.9. Na západnom Slovensku (2.5mil. obyvateľov) má 33% ľudí VŠ vzdelanie, na stredom Slovensku (1.5mil. obyvateľov) má 20% ľudí VŠ vzdelanie a na východnom Slovensku (1.8mil. obyvateľov) je to 22%. Náhodne vybraný človek nemá VŠ vzdelanie. Aká je pravdepodobnosť, že pochádza zo západného Slovenska?

Cvičenie 3.10. Uvažuje test na tuberkulózu, ktorého senzitivita aj špecificita sú 99%. Skutočná prevalencia tuberkulózy v populácii nech je 0.00004. Aká je pravdepodobosť, že náhodne vybraný človek má tuberkulózu, ak má pozitívny test?

Cvičenie 3.11. Malá firma vyhodnotila, že so 75% svojich zamestnancov je spokojná. Zistilo sa, že z týchto malo až 80% predošlú pracovnú skúsenosť. Z tých ostatných zamestnancov, s ktorými firma nebola spokojná to bolo len 15%. Firma práve zamestnala nového pracovníka, ktorý mal predošlú pracovnú skúsenosť. Aká je pravdepodobnosť, že s ním firma nebude spokojná?

Cvičenie 3.12. Desať percent pacientov vstupujúcich na kliniku má poškodenú pečeň. Päť percent týchto pacientov je alkoholikov. Z tých pacientov, ktorí majú poškodenú pečeň, je 7% alkoholikov. Aká je pravdepdoobnosť, že pacient, ktorý je alkoholik má poškodenú pečeň?

Cvičenie 3.13. Nech je pravdepodobnosť, že náš vlak bude meškať 60%. Aká je šanca, že prídeme tam aj späť načas, bez meškania?

Cvičenie 3.14. Na pošte je dlhý rad. Pravdepodobnosť, že požiadavka každého čakajúceho bude vyriešená do 1 minúty je 10%. Aká je pravdepodobnosť, že každý zo 17 ľudí v rade predo mnou bude obslúžený do jednej minúty?

Cvičenie 3.15 (Narodeninový paradox)**.** Na predpandemickej párty (offline) sa stretne n ľudí. Aké je najmenšie n také, aby bola šanca, že aspoň jedna dvojica má narodeniny v ten istý deň väčšia ako 50? (Pomôcka: samozrejme prekvapivo malá, inak by sa to nevolalo *paradox.*) Predpokladajme, že každý jeden človek má pravdepodobnosť práve $\frac{1}{365}$, že sa narodí v nejaký konkrétny deň, ignorujúc prestupné roky.

Cvičenie 3.16 (Monty Hall Problem)**.** Kedysi išla v televízii nasledovná súťaž o auto. Boli troje dverí, A, B, C. Za dvoma z týchto troch dverí bola koza, za jednými dverami bolo auto. Súťažiaci si vybral jednu z týchto dvoch dverí. Moderátor potom otvoril jedny z tých dvoch dverí, ktoré si súťažiaci nevybral a za nimi bola vždy koza. Zostali tak už iba dvoje dverí. Súťažiaci si v druhom kole mohol vybrať z dvoch možností: (1) zostane na svojom pôvodnom tipe, (2) zmení svoje rozhodnutie na tie druhé dvere, ktoré ešte ostali neotvorené.

Ktorá z možností (1) alebo (2) je lepšia? Alebo sú rovnocenné?

(Zaujímavosťou je, že v tejto štúdii výskumníci ukazujú, že holuby rozumejú Monty Hall problému lepšie ako ľudia. . .)

Cvičenie 3.17 (*Mená detí)**.** V rodine sú dve deti. Vieme, že jedno z nich je dievča. Aká je pravdepodobnosť, že obe sú dievčatá?

V rodine sú dve deti. Jedno dieťa sa volá Gertrúda. Aká je pravdepodobnosť, že obe sú dievčatá?

4 Náhodná premenná

Doteraz sme si zadefinovali pravdepodobnostný priestor. Bola to trojica $(\Omega, \mathcal{F}, P)$. Kým Ω bola akákoľvek neprázdna množina, $\mathcal{F}$ aj P museli spĺňať nejaké podmienky, napríklad P musela byť spočitateľne aditívna.

Prvky v množine Ω mohli byť úplne akékoľvek objekty, mohlo to byť číslo 5, mohol to byť symbol ω_1, slovo "zemiak", obrázok domčeka alebo veta "na kocke padlo číslo 6". Bolo to čokoľvek, čo opisovalo nejakú konkrétnu udalosť.

Teraz sa ideme posunúť konceptuálne ďalej. Našou ambíciou bude kvantifikovať náhodnosť ešte podrobnejšie.

V mnohých situáciách je užitočné nielen uvažovať, či nejaká udalosť nastane alebo nie.

- Koľko bude zajtra stupňov?
- Aká bude koncom roka cena ropy?
- Aký padne súčet čísiel na dvoch kockách?

Aj na tieto otázky je férová a úprimná odpoveď: *"Neviem."* Tieto otázky totiž so sebou nesú nejakú neistotu. Odpoveď *"Neviem."* je pravdivá ale neužitočná. Chceli by sme nejakým spôsobom kvantifikovať náhodnosť, teda reprezentovať to, čo vieme o týchto otázkach povedať. Lebo nie je "náhodne" ako "náhodne", sú rôzne typy náhodnosti. Keď hodím 10 krát mincou, tak zhruba v polovici prípadov padne hlava, ale je to neisté. Keď hodím mincou 1000 krát, proporcia padnutých hláv bude tiež okolo jednej polovice. Každopádne tej neistoty v tom bude ovela menej. Kým pri desiatich hodoch padne 7 hláv hocikedy, tak pri tisíc hodoch je 700 hláv extrémne prekvapivá udalosť.

Na to aby sme s ňou vedeli pracovať, musíme zaviesť nový pojem - náhodná premenná. Náhodná premenná premietne tú našu veľmi abstraktnú množinu udalostí Ω na číselnú os. Toto je ohromne užitočné. S číslami totiž vieme veľmi šikovne narábať. Vieme sa pýtať, či je niečo väčšie alebo menšie, čísla vieme sčítavať, násobiť, deliť, transformovať (aplikovať na ne funkcie). Celý matematický aparát, všetko, čo vieme o funkciách, limitách, deriváciách, zrazu všetko toto budeme vedieť používať na to, aby sme spresňovali vyjadrenie náhodnosti. Aby sme vedeli kvalitnejšie odpovedať na otázky, na ktoré je odpoveď *"Neviem."* síce pravdivá ale neužitočná.

4.1 Čo je to náhodná premenná.

Majme pravdepodobnostný priestor $(\Omega, \mathcal{F}, P)$.

Funkciu $X : \Omega \to \mathbb{R}$ nazývame **náhodná premenná** ak pre všetky $x \in \mathbb{R}$ platí

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}.$$

Toto je len technický zápis toho, že náš pravdepodobnostný priestor musí mať dostatočne bohatú množinu udalostí $\mathcal{F}$, ktorým vie priradiť pravdepodobnosť. Takú bohatú, aby sme vedeli priradiť pravdepodobnosť udalosti "náhodná premenná X je menšia alebo rovná ako x".[16]

Náhodná premenná je užitočný koncept, keď je výsledkom experimentu nejaké číslo. No a to je veľmi často. Teplota zajtra, cena ropy, súčet na dvoch kockách. Odteraz budeme pre množinu $\{\omega \in \Omega : X(\omega) \leq x\}$ používať skrátený zápis $\{X \leq x\}$ a pre množinu $\{\omega \in \Omega : X(\omega) \in A\}$ takýto zápis $\{X \in A\}$.

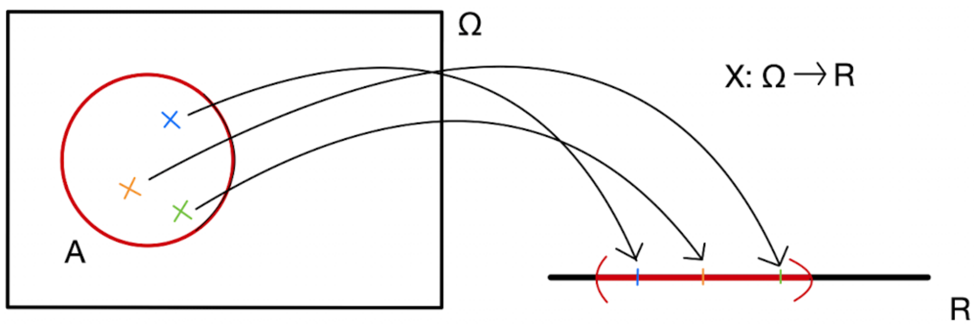


Figure 10: Náhodná premenná.

Príklad 4.1. Hádžeme férovou kockou. Označme $\Omega = \{1, 2, 3, 4, 5, 6\}$, kde napr. 4 značí, že na kocke padlo číslo 4. Nech $\mathcal{F} = 2^\Omega$ a nech $P(A) = \frac{|A|}{6}$. Označme písmenom X náhodnú premennú, ktorá bude označovať číslo, ktoré padne na kocke. Náhodná premenná X priradí každému elementu k z Ω reálne číslo nasledovne: $X(k) = k$. Množina $\{X \leq x\} = \{\omega \in \Omega : X(\omega) \leq x\}$ musí byť v $\mathcal{F}$ pre všetky $x \in \mathbb{R}$. Napríklad

- $\{X \leq -2\} = \emptyset \in \mathcal{F}$,
- $\{X \leq 0.5\} = \emptyset \in \mathcal{F}$,
- $\{X \leq 2.3\} = \{1, 2\} \in \mathcal{F}$,
- $\{X \leq 11.9\} = \Omega \in \mathcal{F}$.

Množina 2^Ω je preto dostatočne "bohatá" na to, aby na nej bola táto funkcia X náhodnou premennou.

Príklad 4.2. Znovu hádžeme férovou kockou. Označme $\Omega = \{1, 2, 3, 4, 5, 6\}$, kde napr. 4 značí, že na kocke padlo číslo 4. Nech množina udalostí $\mathcal{F}^* = \{\emptyset, \Omega, \{1, 3, 5\}, \{2, 4, 6\}\}$ je tentokrát "chudobnejšia" a nech $P(A) = \frac{|A|}{6}$. Označme písmenom Y funkciu, ktorej hodnota bude 1, ak bude číslo párne a 0 ak bude nepárne.

Množina $\{Y \leq y\} = \{\omega \in \Omega : Y(\omega) \leq y\}$ musí byť v $\mathcal{F}^*$ pre všetky $y \in \mathbb{R}$. Napríklad

- $\{Y \leq -2\} = \emptyset \in \mathcal{F}^*$,
- $\{Y \leq 0.5\} = \{1, 3, 5\} \in \mathcal{F}^*$,
- $\{Y \leq 2.3\} = \Omega \in \mathcal{F}^*$,
- $\{Y \leq 11.9\} = \Omega \in \mathcal{F}^*$.

Takže aj v tomto prípade je menšia $\mathcal{F}^*$ stále dostatočne bohatá na to, aby na nej bola táto funkcia Y náhodnou premennou. Je však funkcia X z predošlého príkladu náhodnou premennou aj na $(\Omega, \mathcal{F}^*)$?

Nie je. Množina $\{X \leq 2.3\} = \{1, 2\} \notin \mathcal{F}^*$ a preto nespĺňa definíciu náhodnej premennej. V definícii sme zvolili $x = 2.3$.

[16] Pri diskrétnej množine Ω nám stačilo uvažovať $\mathcal{F} = 2^\Omega$. Nemali sme problém. Pri spojitej množine Ω toto nejde. Vo všeobecnosti platí, že čím väčšia je $\mathcal{F}$, tým podrobnejšia môže byť náhodná premenná X. Viacej sa týmto budeme zaoberať na pokročilejšom kurze.

Poučenie: kým množina udalostí $\mathcal{F}$ bola dosť veľká, pri $\mathcal{F}^*$ už nevieme priraďovať pravdepodobnosti hocijakým otázkam. Napriklad na to, aby sme vedeli priradiť pravdepodobnosť udalosti "Padne na kocke číslo menšie ako 2.3?" potrebujem, aby množina $\{1,2\}$ bola tiež udalosťou. Ale to pri $\mathcal{F}^*$ nie je. Množina udalostí $\mathcal{F}^*$ totiž vie jedine rozlišovať, či pádne párne $\{2,4,6\} \in \mathcal{F}^*$ alebo nepárne číslo $\{1,3,5\} \in \mathcal{F}^*$. Náhodnej premennej Y to ale nevadí, lebo ona "zlepí" hodnoty 1, 3 aj 5 do hodnoty 0, lebo všetky tieto hodnoty sú pre ňu rovnako nepárne.

Príklad 4.3. Hádžeme dvoma férovými kockami. Označme $\Omega = \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(2,2),(2,3),\cdots,(6$ kde napr. $(3,2)$ značí, že na prvej kocke padlo číslo 3 a na druhej kocke padlo číslo 2. Nech $\mathcal{F} = 2^\Omega$ a nech $P(A) = \frac{|A|}{36}$. Označme písmenom X náhodnú premennú, ktorá bude označovať súčet dvoch čísiel, ktoré padnú na kockách. Náhodná premenná X priradí každému elementu (k_1,k_2) z Ω reálne číslo nasledovne: $X((k_1,k_2)) = k_1 + k_2$.

4.2 Kumulatívna distribučná funkcia

Na charakterizáciu náhodnej premennej budeme používať **kumulatívnu distribučnú funkciu**. Distribučná funkcia náhodnej premennej X je funkcia $F : \mathbb{R} \to \mathbb{R}$, ktorá je definovaná nasledovne:

$$F_X(x) = P(X \leq x).$$

Už z tejto definície máme niektoré jej vlastnosti

- F_X je neklesajúca funkcia,
- F_X je sprava spojitá,
- $F_X(x) \in [0,1]$,
- $\lim_{x\to-\infty} F_X(x) = 0$ a $\lim_{x\to\infty} F_X(x) = 1$.

Ďalšie vlastnosti kumulatívnej distribučnej funkcie sú nasledovné:[17]

- $P(X \in (a,b]) = F_X(b) - F_X(a)$,
- $P(X \in (a,b)) = \lim_{x\to b^-} F_X(x) - F_X(a)$,
- $P(X \in [a,b)) = \lim_{x\to b^-} F_X(x) - \lim_{x\to a} F_X(x)$,
- $P(X \in [a,b]) = F_X(b) - \lim_{x\to a} F_X(x)$,
- $P(X < a) = \lim_{x\to a^-} F_X(x)$,
- $P(X > a) = 1 - F_X(a)$,
- $P(X = a) = F_X(a) - \lim_{x\to a^-} F_X(x)$.

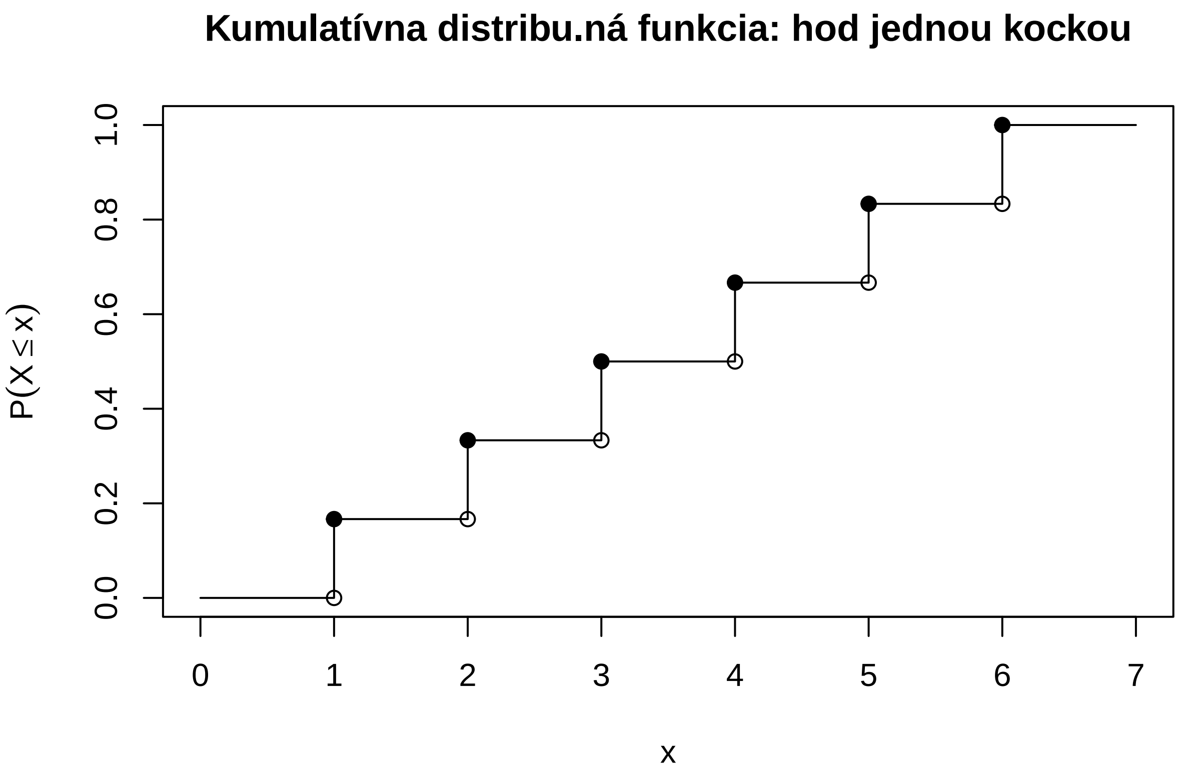
Tieto vlastnosti nám ukazujú, že pomocou kumulatívnej distribučnej funkcie vieme vyjadriť pravdepodobnosť toho, že náhodná premenná nabobudne hodnotu v akomkoľvek intervale. Pripomeňme, že napríklad $P(X \in [a,b))$ je len skrátený zápis pre pravdepodobnosť nasledovnej množiny $P\left(\{\omega \in \Omega : a \leq X(\omega) < b\}\right)$.

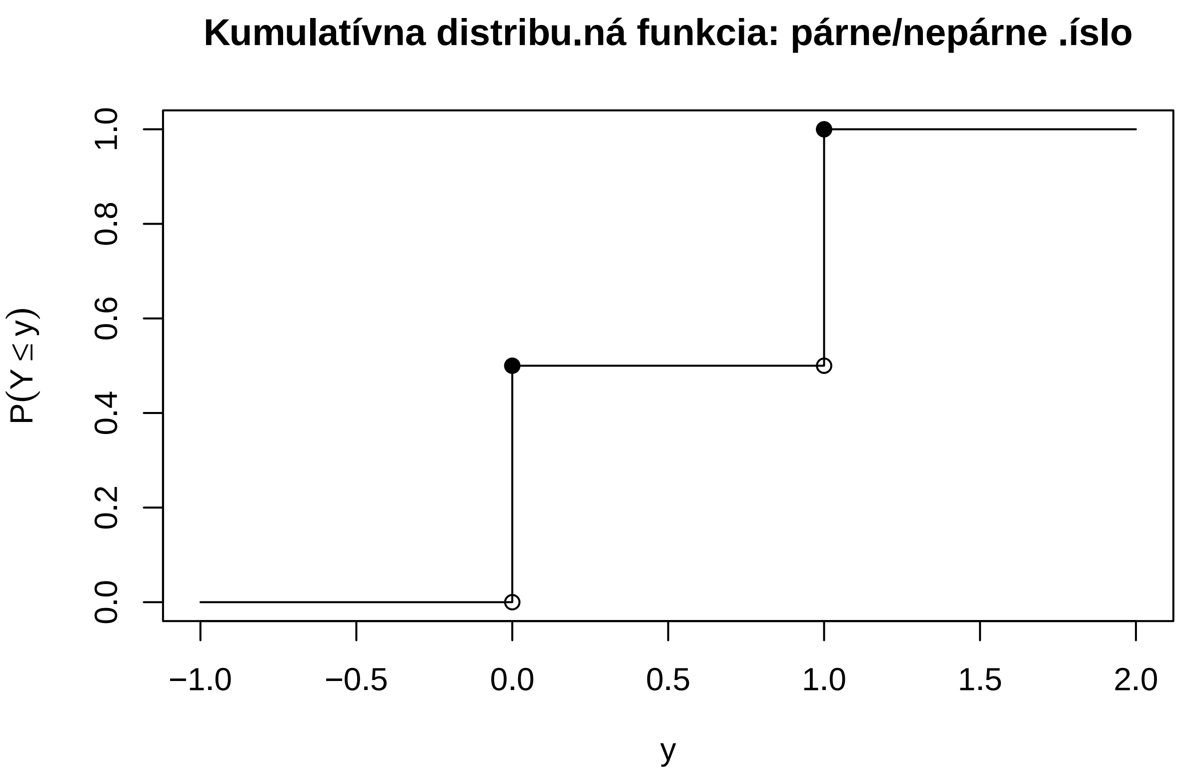
Kumulatívna distribučná funkcia poskytuje kompletnú informáciu o pravdepodobnostnom správaní náhodnej premenej. Inými slovami, úplne popisuje typ náhodnosti.

Príklad 4.4. Hádžeme férovou kockou a máme náhodnu premennú číslo, ktoré padlo na kocke X a identifikátorovú premennú Y čí padlo párne číslo ako v príkladoch 4.1 a 4.2.

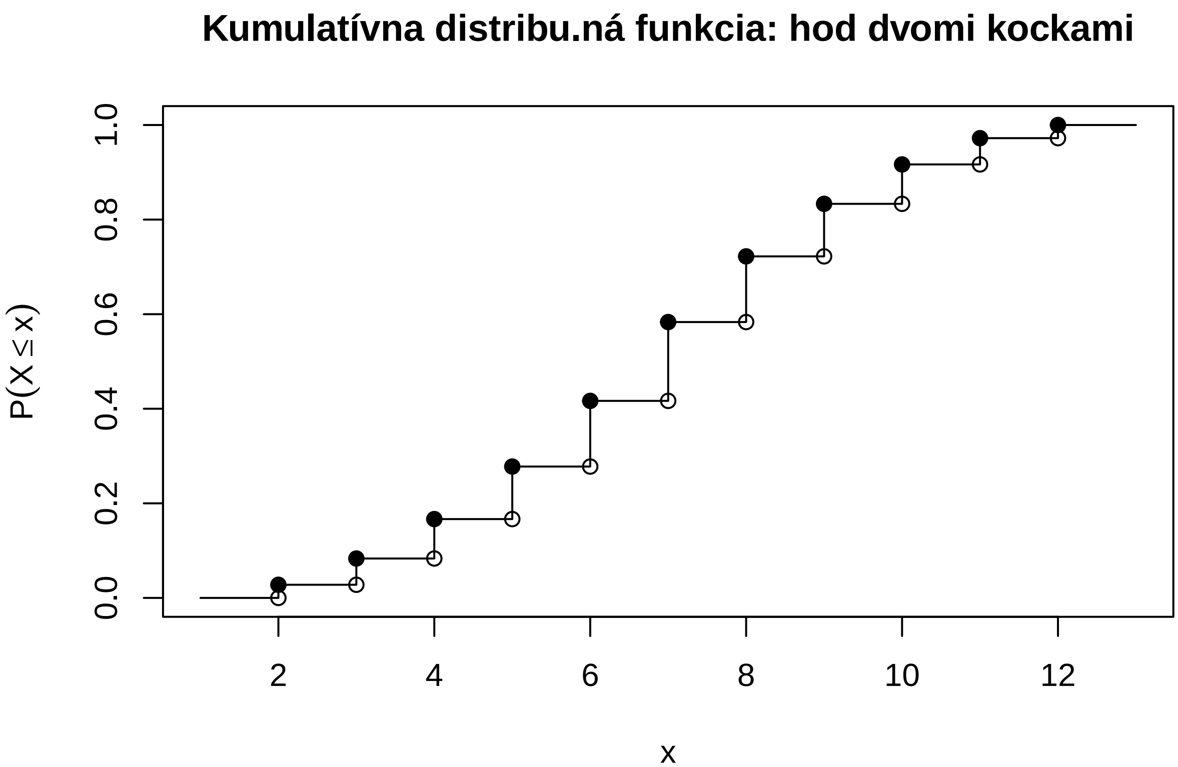
[17] Na dôkaz je potrebná veta o spojitosti pravdepodobnosti, ktorá hovorí, že pre postupnosť udalosti, pre ktorú platí $A_n \subset A_{n+1}$ máme, že

$$\lim_{n\to\infty} P(A_n) = P\left(\cup_{n=1}^{\infty} A_n\right).$$

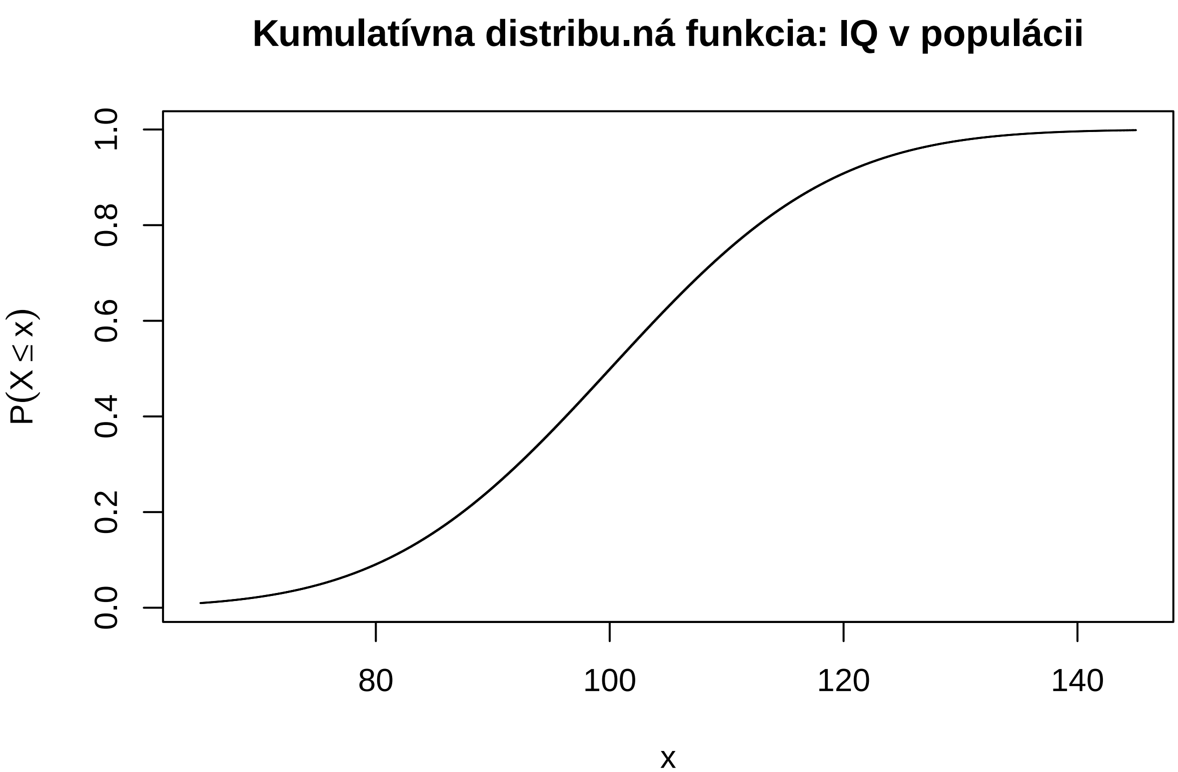




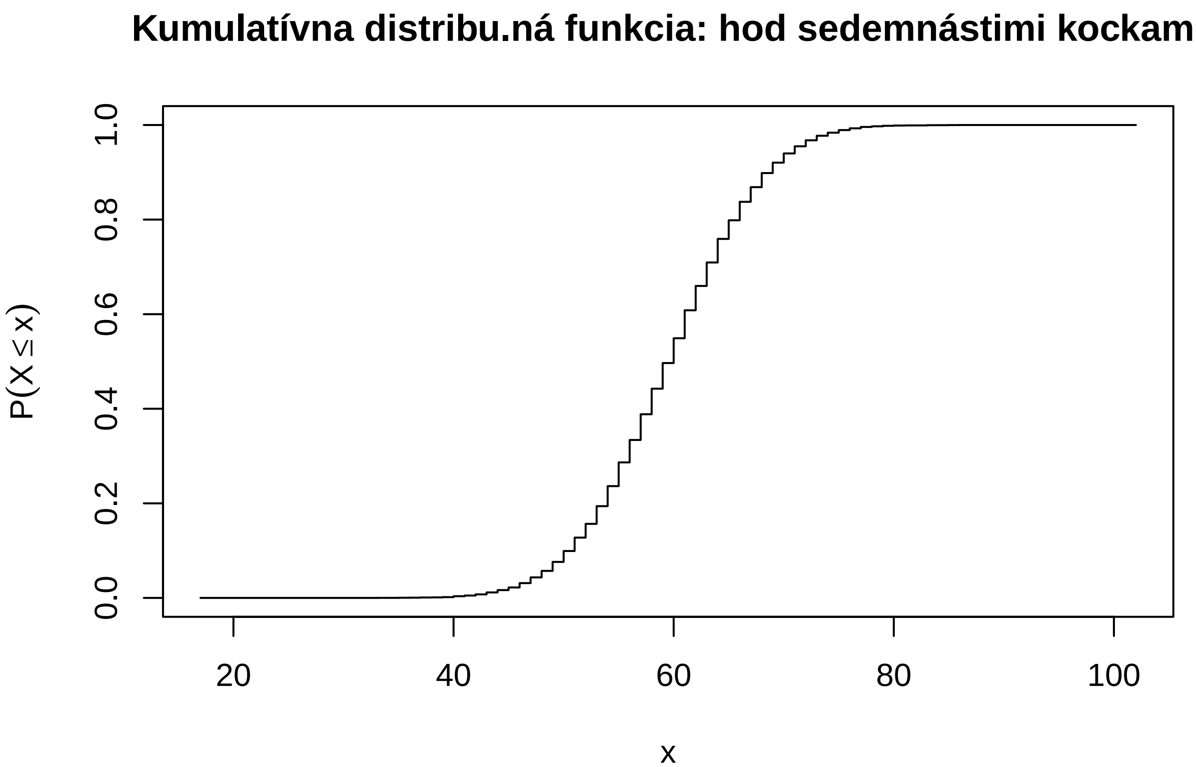
Príklad 4.5. Hádžeme dvoma férovými kockami a máme náhodnu premennú súčet dvoch kociek X ako v príklade 4.3.



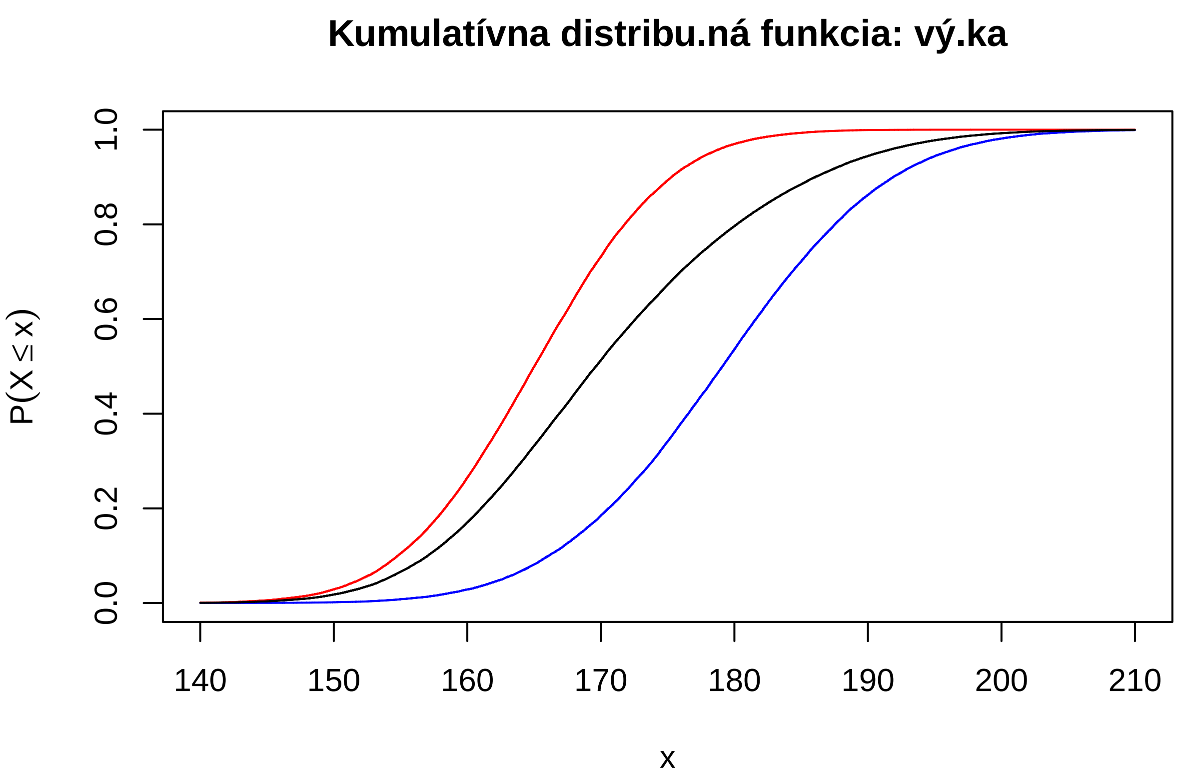
Príklad 4.6. Predpokladajme, že náhodná premenná X reprezentuje IQ v populácii, jej kumulatívna distribučná funkcia môže vyzerať napríklad takto:



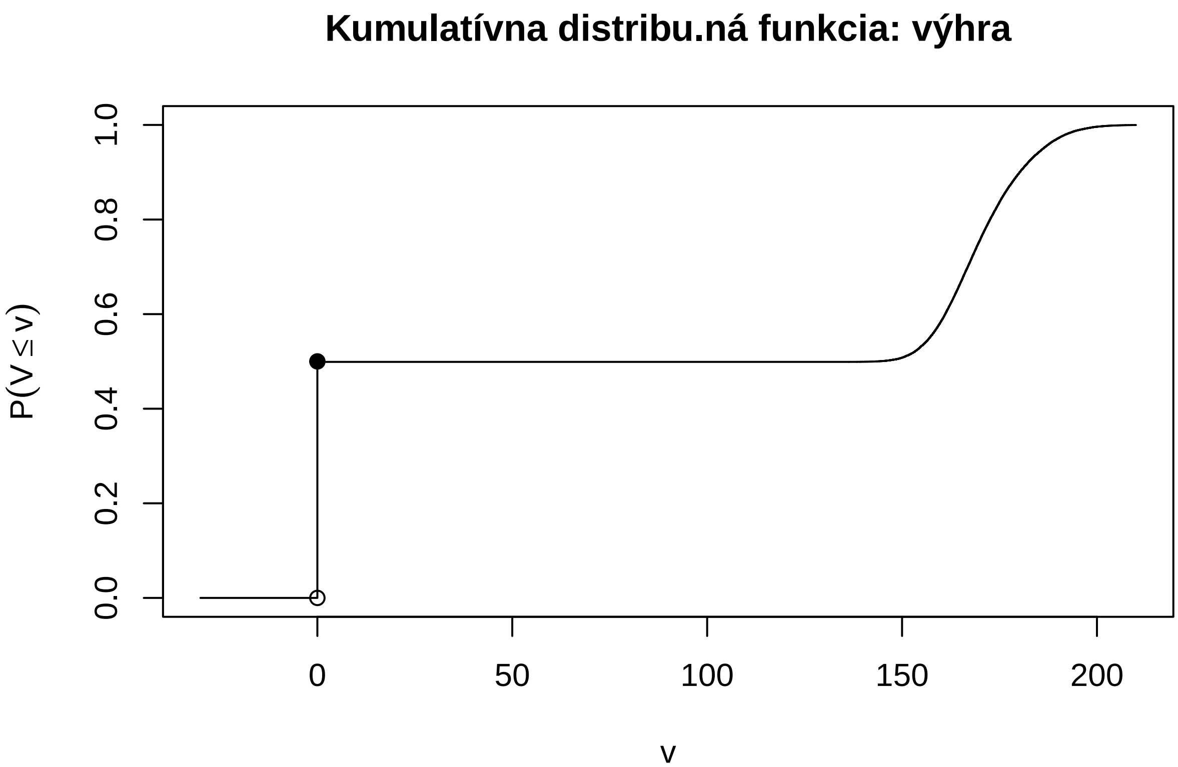
Príklad 4.7. Hádžeme sedemnástimi férovými kockami a máme náhodnú premennú súčet sedemnástich kociek X ako v príklade 4.3, kde boli len dve.



Príklad 4.8. Vyška mužov (modrou) a žien (červenou), v populácii kde je 60% žien. Celková kumulatívna distribučná funkcia pre mužov a ženy spolu je čiernou.



Príklad 4.9. Uvažujme výšky z predchádzajúceho príkladu. Majte nasledovnú súťaž. Človek hodí mincou: ak padne hlava ($M = 1$), tak vyhrá cenu vo výške (v eurách) svojej výšky v cm H. Ak padne znak ($M = 0$), tak nevyhrá nič. Výhra je teda $V = H \cdot M$.



4.3 Zhrnutie

Náhodná premenná je nástroj, ktorý nám umožnuje dostať abstraktný pravdepodobnostný priestor na reálnu os a počítať s náhodnosťou. Jedným zo spôsobov ako ju úplne popísať je **kumulatívna distribučná funkcia**.

4.4 Cvičenia

Cvičenie 4.1. Podrobne vysvetlite, čo je chybné na nasledovných označeniach/tvrdeniach:

- $F_X(0.5) = 1.3$,
- $F_X(0.5) = -\pi/3$,
- $F_X(8) = \infty$,
- $F_X(8) \cdot F_Y(2) = 16$,
- $\lim_{x\to 2} F_X(x) = \infty$,
- $F_Z(\{Z \in (-\infty, 3]\}) = 0.5$,
- $F_Y(A) = P(\{Y \in A\})$,
- $\Omega = 5$,
- $\mathcal{F} = 2^\Omega = 32$.

Cvičenie 4.2. Majme náhodnú premennú M, ktorá nadobúda hodnoty $1, 2, 3, \ldots$ s pravdepodobnosťami $P(M = m) = \frac{1}{2^m}$.

- Nájdite kumulatívnu distribučnú funkciu M,
- Vypočítajte $P(3 < M \leq 7)$,
- Vypočítajte $P(M > 3)$.

Cvičenie 4.3. Overte, či nasledujúce funkcie môžu byť kumulatívnou distribučnou funkciou.

- $F(x) = \begin{cases} 1 - 4^{-x} & \text{ak } x \geq 0, \\ 0 & \text{ak } x < 0, \end{cases}$

- $F(x) = \begin{cases} 0 & \text{pre } x < a, \\ \frac{x-a}{b-a} & \text{pre } x \in [a, b), \\ 1 & \text{pre } x \geq b, \end{cases}$
- $F(x) = \frac{e^x}{1+e^x}$,
- $F(k) = \sum_{i=1}^{k} \frac{1}{2^k}$,
- $F(k) = \sum_{i=1}^{k} \exp\left(\frac{(-1)^{k+1}}{k}\right)$,
- $F(k) = \frac{3}{\pi} \sum_{i=1}^{k} \frac{1}{k^2}$.

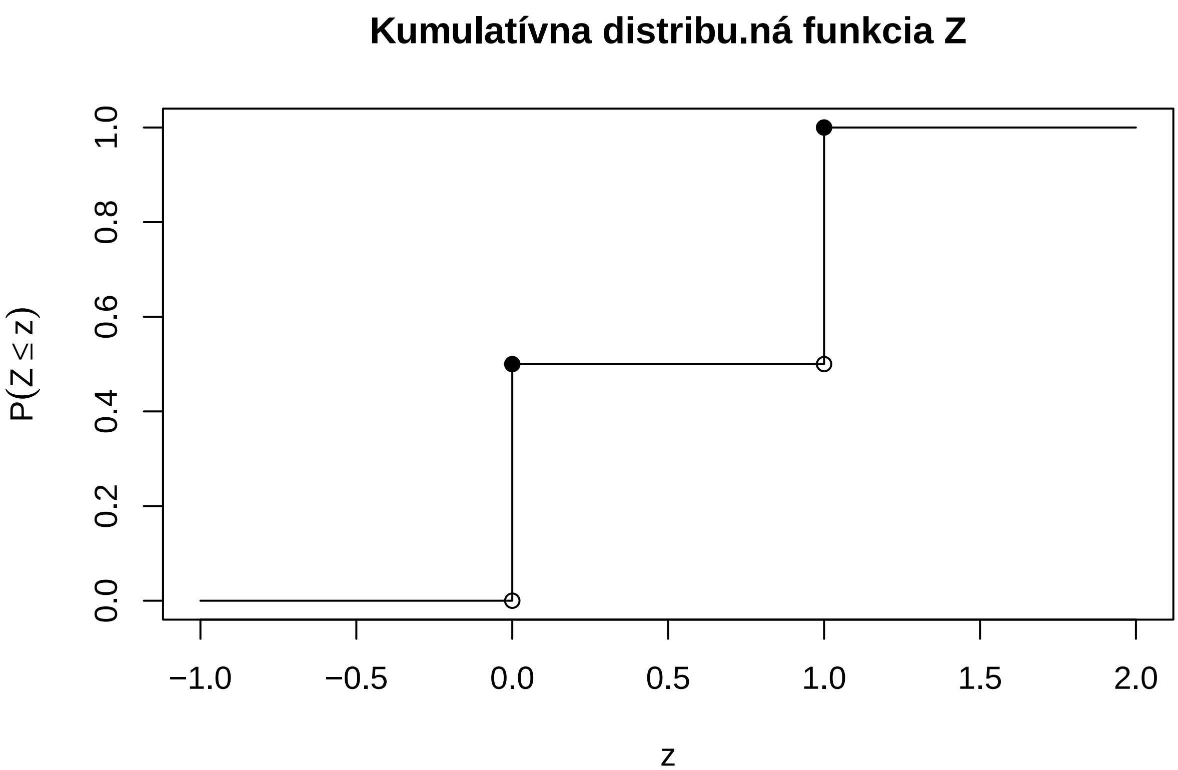
Cvičenie 4.4. Uvažujme hádzanie jednou kockou. Nech $\Omega = \{1, 2, 3, 4, 5, 6\}$ a nech $\mathcal{F} = \{\Omega, \emptyset, \{1\}, \{2, 3, 4, 5, 6\}\}$. Zostrojte dve funkcie $X_1 : \Omega \to \mathbb{R}$ a $X_2 : \Omega \to \mathbb{R}$ také, že

- X_1 je náhodná premenná,
- X_2 nie je náhodná premenná.

Cvičenie 4.5. Uvažujme náhodnú premennú Y ktorá nadobúda hodnoty $0, 1, 2, 3$ s pravdepodobnosťami $0.5, 0.25, 0.125, 0.125$. Zobrazte jej kumulatívnu distribučnú funkciu.

Cvičenie 4.6. Načrtnite dve rôzne kumulatívne distribučné funkcie F_X, F_Y (pre náhodné premenné X a Y) pre ktoré platí $F_X(120) = F_Y(120) = 0.5$, $F_X(140) > F_Y(140)$ a $F_X(90) < F_Y(90)$.

Načrtnite, ako by vyzerala kumulatívna distribučná $S = ZX + (1 - Z)Y$, kde náhodná premenná Z má nasledovnú kumulatívnu distribučnú funkciu:



Cvičenie 4.7. Na Univerzite je 60% žien a 40% mužov. Náhodne vybratý študent/ka odpovedá na dotazník. Dotazník pre mužov má 20 otázok, dotazník pre ženy má 25 otázok. Nech X označuje počet opýtaných otázok pre náhodne zvoleného človeka. Určite $P(X \leq x)$ pre všetky $x \in \mathbb{R}$.

Cvičenie 4.8. Nech X je číslo, ktoré padne na vrchu férovej kocky. Nech $Y = X^3 + 1$ a $Z = \sqrt{X}$. Určite

- $P(Y = y)$ pre všetky reálne čísla $y \in \mathbb{R}$,
- $P(Z = z)$ pre všetky reálne čísla $z \in \mathbb{R}$,
- $P(YZ = x)$ pre všetky reálne čísla $x \in \mathbb{R}$,

- $P(Y + Z = v)$ pre všetky reálne čísla $v \in \mathbb{R}$.

5 Diskrétna náhodná premenná

Zatiaľ sme si hovorili, že čo je to náhodná premenná. Je to spôsob ako usporiadať/očíslovať náhodnosť tak, aby sa nám s ňou potom vhodne pracovalo. Existujú rôzne druhy/skupiny náhodných premenných:

- diskrétne náhodné premenné,
- spojité náhodné premenné,
- zmiešané náhodné premenné.

Toto rozdelenie je na základe toho, aké hodnoty môžu tieto náhodné premenné nadobúdať. Delenie dáva zmysel aj preto, že s týmito skupinami náhodných premenných sa pracuje o dosť inak. Prvým dvom skupinám sa budeme venovať osobitne. V rámci tejto kapitoly budeme skúmať *diskrétne náhodné premenné* s nasledovnými pravdepodobnostnými rozdeleniami:

- rovnomerné rozdelenie,
- Bernoulliho rozdelenie,
- binomické rozdelenie,
- Poissonovo rozdelenie,
- geometrické rozdelenie,
- hypergeometrické rozdelenie,
- negatívne binomické rozdelenie.

Budú to také náhodné premenné, ktorých obor hodnôt je konečná alebo nanajvýš spočitateľná množina. Každé takéto rozdelenie je vhodné na modelovanie akýchsi špecifických situácií.

5.1 Pravdepodobnostná funkcia diskrétnej náhodnej premennej

Vieme, že kumulatívna distribučná funkcia *plne* charakterizuje pravdepodobnostné správanie náhodnej premennej. Je to funkcia a vieme ju zobraziť. V prípade diskrétnej náhodnej premennej však vieme popísať pravdepodobnostné správanie náhodnej premennej aj alternatívne. Jednoducho tak, že vyčíslime pravdepodobnosti $P(X = x)$ pre všetky možné $x \in \mathcal{S}_X$, kde $\mathcal{S}_X = \{X(\omega) : \omega \in \Omega\}$. Množina $\mathcal{S}_X$ je teda obor hodnôt funkcie X, zahŕňa všetky možné hodnoty, aké môže funkcia X nadobúdať. Pre diskrétnu náhodnú premennú je $\mathcal{S}_X$ konečná alebo nanajvýš spočitateľná množina.

Funkciu $p_X : \mathcal{S}_X \to [0, 1]$ definovanú nasledovne

$$p_X(x) = P(X = x)$$

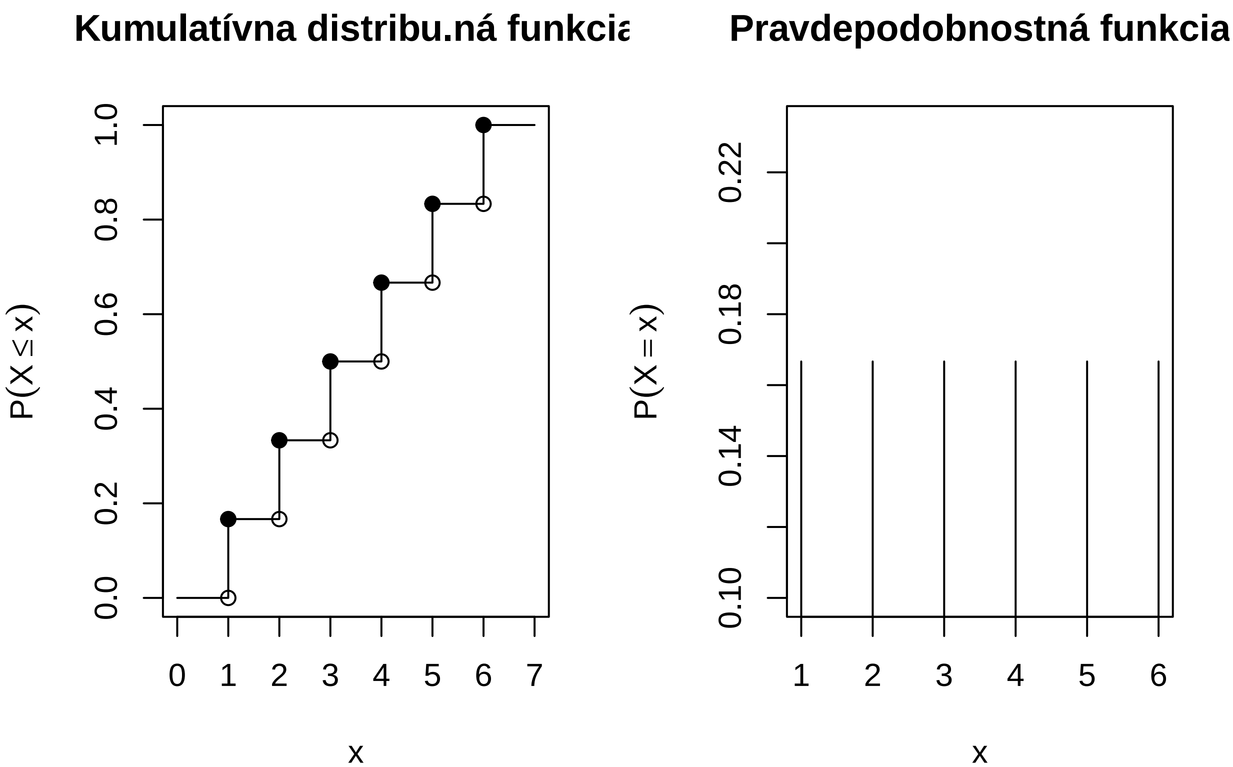
nazývame **pravdepodobnostná funkcia diskrétnej náhodnej premennej** X.

Platí

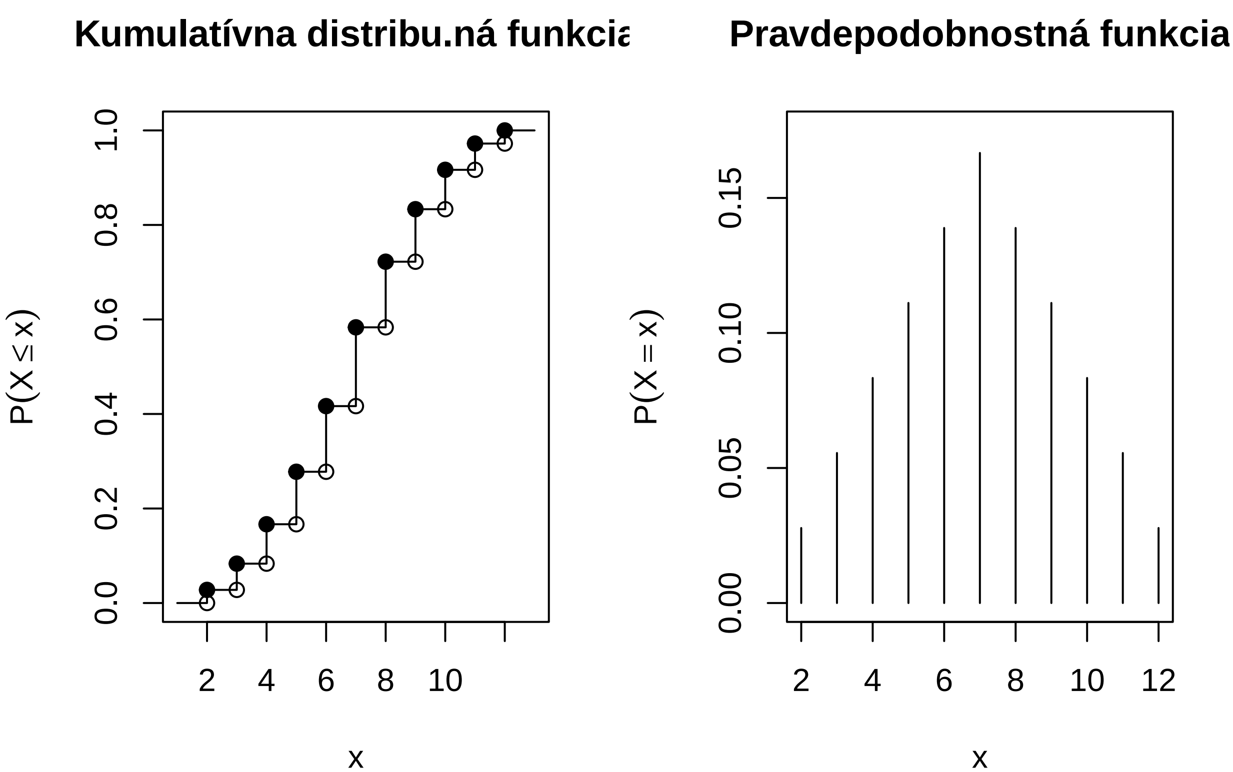
- $0 \leq p_X(x) \leq 1$,
- $\sum_{x \in \mathcal{S}_X} p_X(x) = 1$.

Funkcia $p_X(x)$ tiež kvantifikuje "výšku schodu" v bodoch nespojitosti kumulatívnej distribučnej funkcie diskrétnej náhodnej premennej.

Príklad 5.1. Hádžeme férovou kockou. Označme $\Omega = \{1, 2, 3, 4, 5, 6\}$, kde napr. 4 značí, že na kocke padlo číslo 4. Nech $\mathcal{F} = 2^\Omega$ a nech $P(A) = \frac{|A|}{6}$. Označme písmenom X náhodnú premennú, ktorá bude označovať číslo, ktoré padne na kocke. Náhodná premenná X priradí každému elementu $k \in \Omega$ reálne číslo nasledovne: $X(k) = k$.



Príklad 5.2. Hádžeme dvoma férovými kockami. Označme $\Omega = \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(2,2),(2,3),\cdots,(6$
kde napr. $(3,2)$ značí, že na prvej kocke padlo číslo 3 a na druhej kocke padlo číslo 2. Nech $\mathcal{F} = 2^{\Omega}$ a nech $P(A) = \frac{|A|}{36}$. Označme písmenom X náhodnú premennú, ktorá bude označovať súčet dvoch čísiel, ktoré padnú na kockách. Náhodná premenná X priradí každému elementu (k_1, k_2) z Ω reálne číslo nasledovne: $X((k_1, k_2)) = k_1 + k_2$.



5.2 Charakteristiky diskrétnych náhodných premenných

Niekedy je užitočné reprezentovať nejaký aspekt tohoto správania jedným číslom. Okolo akej hodnoty je toto rozdelenie centrované (stredná hodnota) a ako veľmi je koncentrované/roztýlené (variancia/rozptyl).

5.2.1 Stredná hodnota

Stredná hodnota diskrétnej náhodnej premennej je zadefinovaná ako

$$\mathrm{E}[X] = \sum_{x \in \mathcal{S}_X} x \cdot p_X(x),$$

pokiaľ platí $\sum_{x \in \mathcal{S}_X} |x| \cdot p_X(x) < \infty$.[18]

Ak pretransformujeme náhodnú premennú $Y = f(X)$, kde $f : \mathbb{R} \to \mathbb{R}$, tak jej stredná hodnota je

$$\mathrm{E}[Y] = E\left[f(X)\right] = \sum_{x \in \mathcal{S}_X} f(x) \cdot p_X(x).$$

Príklad 5.3. Hádžeme férovou kockou. Nech X označuje číslo, ktoré padlo na kocke.

$$\mathrm{E}[X] = \sum_{x \in \mathcal{S}_X} x \cdot p_X(x) = \sum_{x \in \{1,2,3,4,5,6\}} x \cdot p_X(x) = \sum_{x=1}^{6} x \cdot \frac{1}{6} = 3.5.$$

Príklad 5.4. Hádžeme dvomi férovými kockami. Nech X označuje súčet čísel na dvoch kockách.

$$\begin{aligned}
\mathrm{E}[X] &= \sum_{x \in \mathcal{S}_X} x \cdot p_X(x) = \sum_{x \in \{2,3,\cdots,12\}} x \cdot p_X(x) \\
&= 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + 5 \cdot \frac{4}{36} + 6 \cdot \frac{5}{36} + 7 \cdot \frac{6}{36} \\
&\quad + 8 \cdot \frac{5}{36} + 9 \cdot \frac{4}{36} + 10 \cdot \frac{3}{36} + 11 \cdot \frac{2}{36} + 12 \cdot \frac{1}{36} = 7.
\end{aligned}$$

Stredná hodnota charakterizuje náhodnú premennú jedným číslom. Je to vážený priemer, kde hodnoty sú váhované prislúchajúcimi pravdepodobnosťami. Fyzikálnou interpretáciou je *ťažisko*.

Príklad 5.5. Hádžeme neférovou mincou. $\Omega = \{H, Z\}, \mathcal{F} = 2^{\Omega}$ a $P(\{H\}) = 0.8 = 1 - P(\{Z\})$. Nech $X(H) = 1$ a $X(Z) = 0$.

$$\mathrm{E}[X] = 1 \cdot 0.8 + 0 \cdot 0.2 = 0.8.$$

Pre strednú hodnotu platí, že je *lineárna*. Uvažujme dve náhodné premenné X a Y a tretiu náhodnú premennú Z, pre ktorú platí $Z = aX + bY$. Náhodná premenná Z nadobúda hodnoty v $\mathcal{S}_Z = \{ax + by : x \in \mathcal{S}_X, y \in \mathcal{S}_Y\}$.

[18] Túto podmienku potrebujeme na to, aby bola táto suma konvergentná. V prípade, že $\mathcal{S}_X$ je konečná množina je táto podmienka splnená vždy. Pri nekonečnej spočítateľnej množine $\mathcal{S}_X$ však čiastočné súčty radu môžu divergovať.

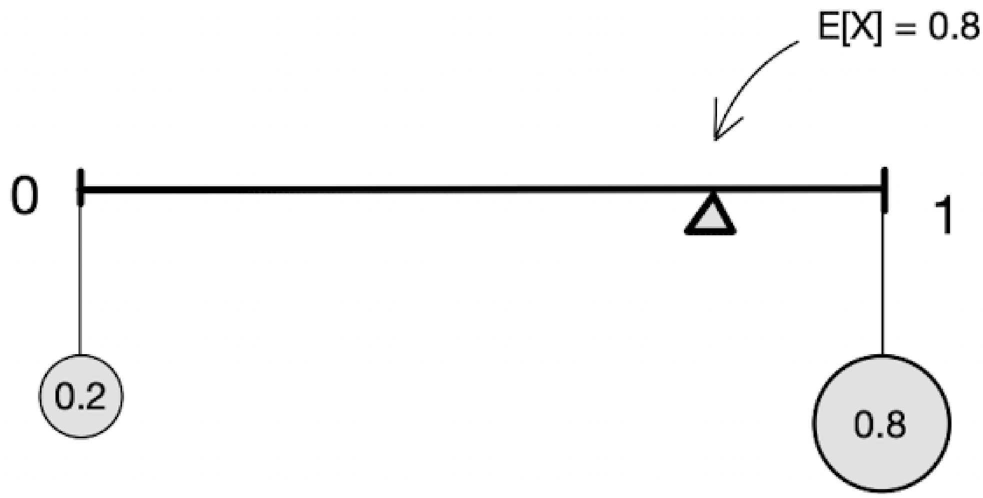


Figure 11: Ilustrácia strednej hodnoty pre neférovú mincu. Stredná hodnota je číslo, kde treba podložiť hojdačku tak, aby bola v rovnováhe.

$$
\begin{aligned}
E[Z] &= \sum_{z\in\mathcal{S}_Z} z\cdot p_Z(z) = \sum_{x\in\mathcal{S}_X, y\in\mathcal{S}_Y} (ax+by)\cdot P(aX=ax\cap bY=by)\\
&= \sum_{x\in\mathcal{S}_X, y\in\mathcal{S}_Y} ax\cdot P(X=x\cap Y=y) + \sum_{x\in\mathcal{S}_X, y\in\mathcal{S}_Y} by\cdot P(X=x\cap Y=y)\\
&= \sum_{x\in\mathcal{S}_X} ax\cdot P(X=x) + \sum_{y\in\mathcal{S}_Y} by\cdot P(Y=y)\\
&= a\sum_{x\in\mathcal{S}_X} x\cdot p_X(x) + b\sum_{y\in\mathcal{S}_Y} y\cdot p_Y(y)\\
&= aE(X)+bE(Y).
\end{aligned}
$$

Využili sme, že platí

$$\sum_{x\in\mathcal{S}_X, y\in\mathcal{S}_Y} P(X=x\cap Y=y) = \sum_{x\in\mathcal{S}_X} P(X=x) = \sum_{y\in\mathcal{S}_Y} P(Y=y),$$

lebo množiny $\{X=x\}$ pre $x\in\mathcal{S}_X$ a $\{Y=y\}$ pre $y\in\mathcal{S}_Y$ tvoria rozklad množiny Ω.

$$\Omega = \cup_{x\in\mathcal{S}_X}\{X=x\} = \cup_{y\in\mathcal{S}_Y}\{Y=y\}.$$

Pripomeňme, že $\{X=x\}$ je len skrátený zápis pre množinu $\{\omega\in\Omega : X(\omega)=x\}$.

Z tohoto priamo vyplýva aj

$$E[aX+b] = a\mathrm{E}[X]+b,$$

nakoľko môžeme uvažovať náhodnú premennú $Y=1$ ako konštantnú jednotku.

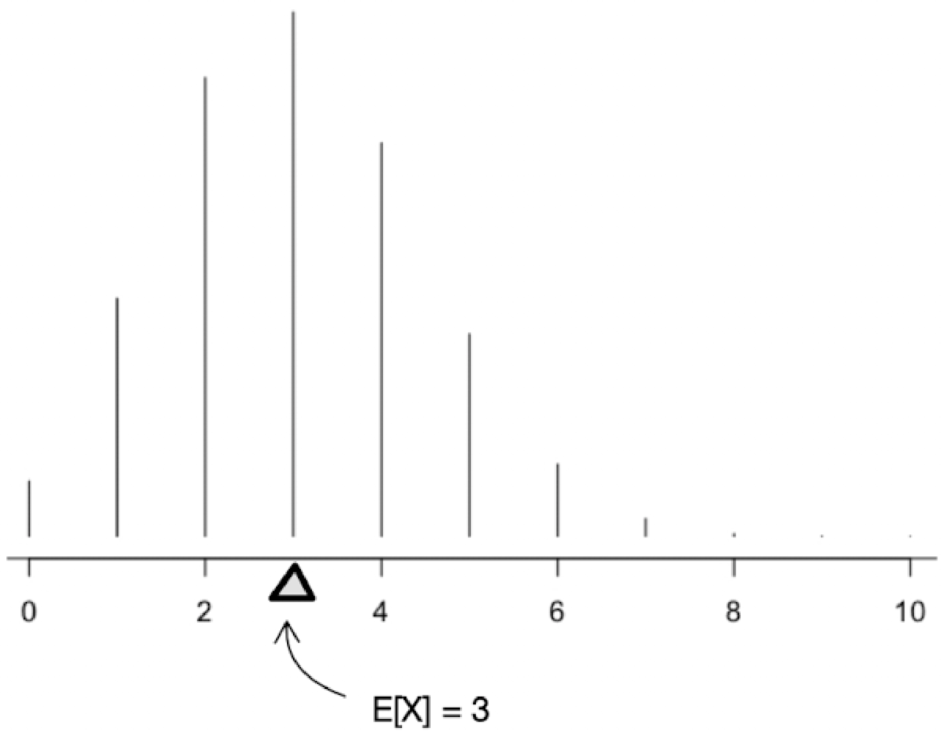


Figure 12: Ilustrácia strednej hodnoty pre počet úspechov z celkového počtu 10 nezávislých pokusov, kde každý úspech má pravdepodobnosť 0.3.

5.2.2 Variancia (rozptyl)

Variancia diskrétnej náhodnej premennej je zadefinovaná ako

$$\mathrm{Var}[X] = E\left[(X - \mathrm{E}[X])^2\right].$$

Takže variancia nie je nič iné ako stredná hodnota kvadratických odchýlok od strednej hodnoty, teda $\mathrm{E}[Y]$, kde $Y = (X - \mathrm{E}[X])^2$.

Z definície variancie máme priamo:

$$\mathrm{Var}[X] = \sum_{x \in \mathcal{S}_X} (x - \mathrm{E}[X])^2 \cdot p_X(x).$$

Naviac platí

$$\begin{aligned}
\mathrm{Var}[X] &= E\left[X^2 - 2X \cdot \mathrm{E}[X] + (\mathrm{E}[X])^2\right] \\
&= E[X^2] - 2\mathrm{E}[X] \cdot \mathrm{E}[X] + (\mathrm{E}[X])^2 \\
&= E[X^2] - (\mathrm{E}[X])^2,
\end{aligned}$$

kde sme využili linearitu strednej hodnoty náhodnej premennej.

Smerodajná odchýlka náhodnej premennej je odmocnina z jej variancie

$$\mathrm{sd}[X] = \sqrt{\mathrm{Var}[X]}.$$

Kým interpretácia strednej hodnoty bola, že ide o ťažisko, pri variancii alebo smerodajnej odchýlke je to o čosi zložitejšie.

Príklad 5.6. Variancia a smerodajná odchýlka výsledku hodu neférovej mince.

$$\begin{aligned}
\mathrm{E}[X] &= 0 \cdot 0.2 + 1 \cdot 0.8 = 0.8, \\
E[X^2] &= 0^2 \cdot 0.2 + 1^2 \cdot 0.8 = 0.8, \\
\mathrm{Var}[X] &= E[X^2] - (\mathrm{E}[X])^2 = 0.8 - 0.8^2 = 0.16, \\
\mathrm{sd}[X] &= \sqrt{\mathrm{Var}[X]} = \sqrt{0.16} = 0.4.
\end{aligned}$$

Príklad 5.7. Uvažujme náhodnú premennú Y o ktorej vieme, že

$$\begin{aligned}
p_Y(1) &= 0.3, \\
p_Y(2) &= 0.5, \\
p_Y(3) &= 0.2.
\end{aligned}$$

Potom platí

$$\begin{aligned}
\mathrm{E}[Y] &= 1 \cdot 0.3 + 2 \cdot 0.5 + 3 \cdot 0.2 = 1.9, \\
E[Y^2] &= 1^2 \cdot 0.3 + 2^2 \cdot 0.5 + 3^2 \cdot 0.2 = 4.1, \\
\mathrm{Var}[Y] &= E[Y^2] - (\mathrm{E}[Y])^2 = 4.1 - 3.61 = 0.49.
\end{aligned}$$

môžeme skontrolovať, že vskutku platí

$$\begin{aligned}
\mathrm{Var}[Y] &= E[(Y-\mathrm{E}[Y])^2] \\
&= (1-1.9)^2\cdot 0.3 + (2-1.9)^2\cdot 0.5 + (3-1.9)^2\cdot 0.2 \\
&= 0.243 + 0.005 + 0.242 \\
&= 0.49.
\end{aligned}$$

Smerodajná odchýlka je preto $\mathrm{sd}[Y] = \sqrt{\mathrm{Var}[Y]} = 0.7$.

Na tomto príklade si možno všimnúť aj to, že na jeho výpočet nepotrebujeme priamo pracovať s pravdepodobnostným priestorom.

5.3 Rovnomerné rozdelenie

Hovoríme, že diskrétna náhodná premenná X má **rovnomerné rozdelenie** na hodnotách $1, 2, 3, \cdots, m$ ak platí[19]

$$p_X(k) = \begin{cases} \frac{1}{m}, & \text{ak } k \in \{1, 2, \cdots, m\} \\ 0, & \text{inak} \end{cases}$$

Takúto náhodnú premennú označujeme ako $X \sim \mathrm{Unif}(\{1, 2, \cdots, m\})$,

Toto rozdelenie modeluje situácie, keď sú výsledky experimentu prislúchajúce $X = 1, X = 2$ alebo $X = k$ rovnako pravdepodobné. Inými slovami, keď veci nastávajú "náhodne" - v zmysle nesystematicky. Argument symetrie je niekedy použitý, a ak nie je povedané inak, tak použitím slova "náhodne" sa potichu predpokladá, že udalosti sú rovnako pravdepodobné.

Pre X s rovnomerným rozdelením musíme mať množinu $\mathcal{S}_X$ konečnú, nakoľko potrebujeme aby platilo $\sum_{x\in\mathcal{S}_X} p_X(x)$.

Príkladov na toto rozdelenie sme videli už viacero: hod férovou kockou alebo hod férovou mincou.

Stredná hodnota a variancia pre takúto náhodnú premennú sú

$$\begin{aligned}
\mathrm{E}[X] &= \sum_{k=1}^{m} k\frac{1}{m} = \frac{1}{m}\frac{m(m+1)}{2} = \frac{m+1}{2}, \\
\mathrm{Var}[X] &= \cdots = \frac{m^2-1}{12}.
\end{aligned}$$

5.4 Bernoulliho rozdelenie

Hovoríme, že diskrétna náhodná premenná X má **Bernoulliho rozdelenie** s parametrom p ak platí

$$p_X(k) = \begin{cases} p, & \text{ak } k = 1, \\ (1-p), & \text{ak } k = 0, \\ 0, & \text{inak.} \end{cases}$$

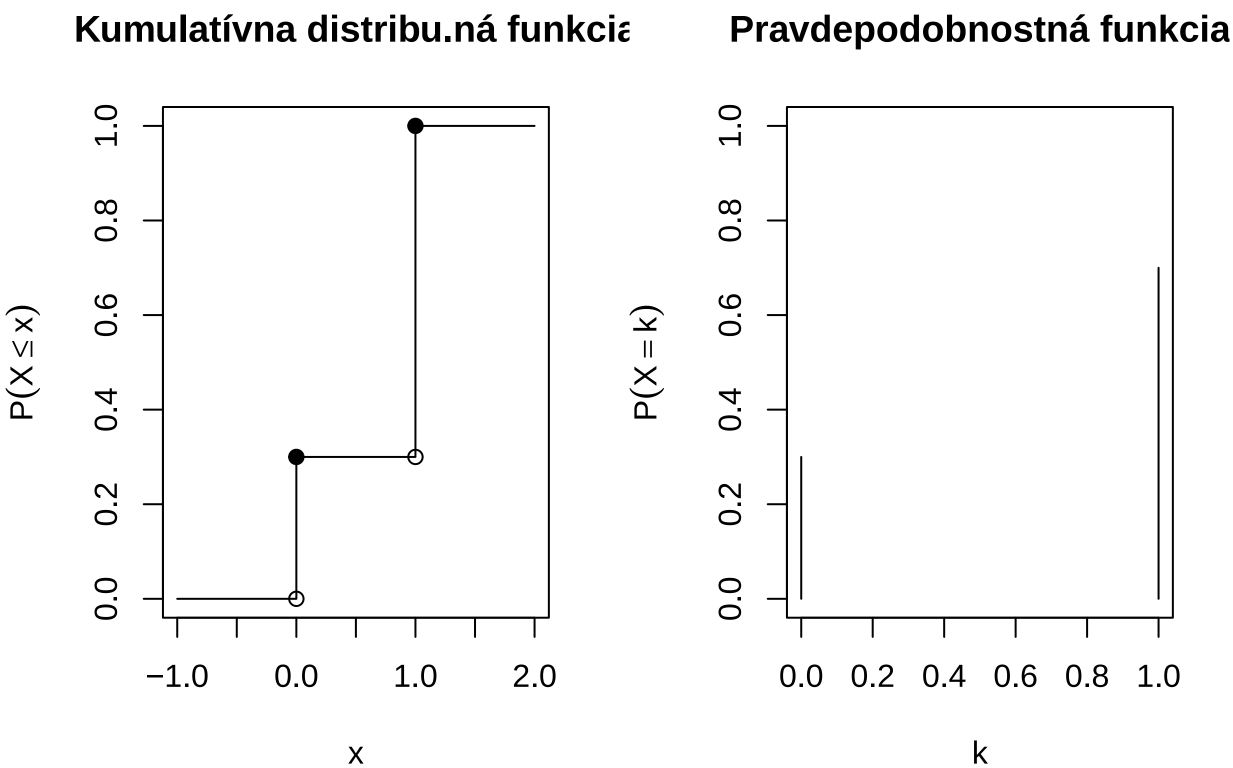
To, že náhodná premenná má takéto rozdelenie označujeme $X \sim Bern(p)$.

Stredná hodnota a variancia pre takúto náhodnú premennú sú

[19] Všimnime si, že teraz budeme používať k namiesto x, teda $p_X(k)$ namiesto $p_X(x)$. Zvykom je, že písmená ako k, l, m, n sa používajú na označenie prirodzených čísel, zatiaľčo x, y, z sú používané skôr na označenie reálnych čísel.

$$
\begin{aligned}
\mathrm{E}[X] &= p \cdot 1 + (1-p) \cdot 0 = p, \\
\mathrm{Var}[X] &= p(1-p).
\end{aligned}
$$

Bernoulliho rozdelenie modeluje výsledok hodu (potenciálne) neférovej mince. Ale nielen mince, čohokoľvek, čoho pravdepodobnosť je nejaké fixné číslo.



5.5 Binomické rozdelenie

Hovoríme, že diskrétna náhodná premenná X má **binomické rozdelenie** s parametrami n a p ak platí

$$
p_X(k) = P(X = k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & \text{ak } k \in \{0, 1, 2, \cdots, n\} \\ 0, & \text{inak.} \end{cases}
$$

Budeme to označovať $X \sim \mathrm{Bin}(n, p)$.

Stredná hodnota a variancia pre takúto náhodnú premennú sú

$$\begin{aligned}
\mathrm{E}[X] &= \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} \\
&= \sum_{k=1}^{n} k\binom{n}{k}p^k(1-p)^{n-k} \\
&= \sum_{k=1}^{n} n\binom{n-1}{k-1}p^k(1-p)^{n-k} \\
&= np\sum_{k=1}^{n} \binom{n-1}{k-1}p^{k-1}(1-p)^{(n-1)-(k-1)} \\
&= np\sum_{j=0}^{m} \binom{m}{j}p^{j}(1-p)^{m-j} \\
&= np, \\
\mathrm{Var}[X] &= \cdots = np(1-p).
\end{aligned}$$

Príklad 5.8. Majme náhodnú premennú X, ktorá je rozdelená binomicky s parametrami $n = 5$ a $p = 0.5$. Vypočítajte $P(X < 2)$.

$$P(X<2) = P(X=0) + P(X=1) = \binom{5}{0}0.5^0 0.5^5 + \binom{5}{1}0.5^1 0.5^4 = 0.03125 + 0.15625 = 18.75\%$$

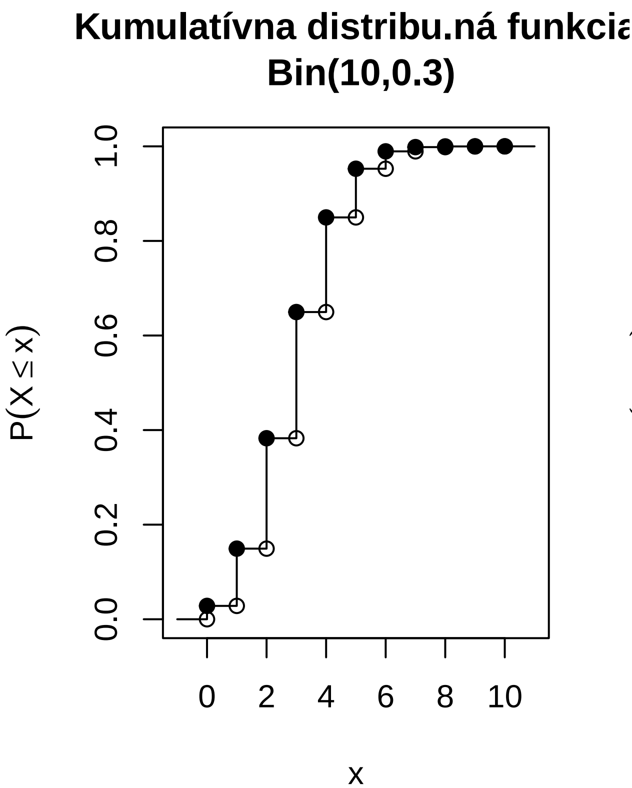
Aké situácie sú typicky modelované binomickými rozdelením? Musíme mať

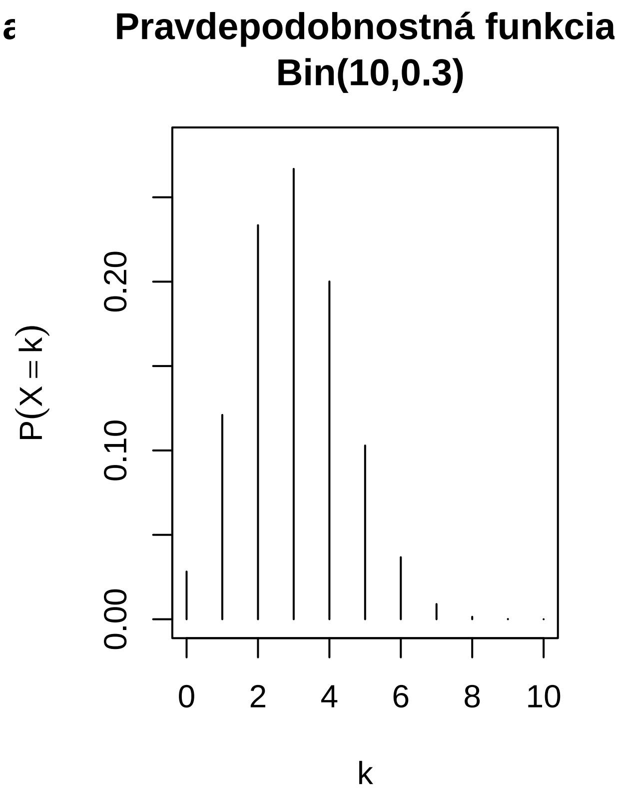
- istý *fixný* počet pokusov
- ktoré majú rovnaký pravdepodobnosť "uspechu" alebo "neúspechu",
- ktoré sú nezávislé.

Je dôležité poznamenať, že ide o *zjednodušenie*, ktorého ambíciou je byť užitočné ale nie byť pravdou. Skoro žiadne pravdepodobnosti v reálnom svete nebudú úplne presne rovnaké a udalosti úplne nezávislé. Binomické rozdelenie je užitočná zjednodušenina, ktorá nám pomáha kvantifikovať náhodnosť spojenú s takýmito experimentami v myšlienkovom experimente, kde tieto predpoklady sú splnené presne.

Tu je nejaká skupina situácií, ktoré sú môžu modelované binomickým rozdelením:

- Aká je rozdelenie počtu pokazených súčiastok z celkového počtu 50 kusov? Chybovosť je rovnaká pre všetky súčiastky a tie sa navzájom neovplyvňujú.
- Aká je rozdelenie počtu chorých ľudí zo vzorky 100 ľudí? Pravdepodobnosť ochorenie je pre každého rovnaká a ľudia navzájom neinteragujú.
- Aká je rozdelenie počtu hláv z celkového počtu 10 nezávislých hodov mincou? Minca je stále tá istá.
- Aké je rozdelenie počtu uhádnutých otázok na teste? Pravdepodobnosť uhádnutia každej otázky je rovnaká.
- Aká je rozdelenie počtu vyliečených pacientov zo vzorky veľkosti 33? Pravdepodobnosť vyliečenia je rovnaká a nezávislá od vyliečenia iných pacientov.
- Aké je rozdelenie počtu voličov pre konkrétneho/nu kandidáta/tku v danom volebnom okrsku? Vzorka voličov je homogénna.
- Aké je rozdelenie počtu žien medzi 30 náhodne vybranými učiteľmi/kami na strednej škole?
- V danej vekovej skupine má 10% pacientov nežiadúce účinky po podaní liečby. Aké je rozdelenie počtu hlásení pri vzorke 80 pacientov?
- Aké je rozdelenie počtu SPAMových emailov z celkového počtu 100 emailov? Každý email počas dňa má rovnakú pravdepodobnosť, že bude SPAMom.
- Aké je rozdelenie počtu vrátených kusov tovaru z celkového počtu 234 kusov pre internetový obchod ak je odpozorované, že zákazníci vracajú tovar s pravdepodobnosťou 3%. Uvažujeme, že pravdepodobnosť vrátenia nesúvisí s tým o aký tovar sa jedná a vrátenie jedného kusu tovaru neovplyvní vrátenie iného.





5.6 Poissonovo rozdelenie

Hovoríme, že diskrétna náhodná premenná X má **Poissonovo rozdelenie** s parametrom λ ak platí

$$p_X(k) = P(X = k) = \begin{cases} \frac{\lambda^k e^{-\lambda}}{k!}, & \text{ak } k \in \{0, 1, 2, \cdots\} \\ 0, & \text{inak.} \end{cases}$$

Zapisujeme to ako $X \sim Pois(\lambda)$.

Stredná hodnota pre Poissoovsky rozdelenú náhodnú premennú je

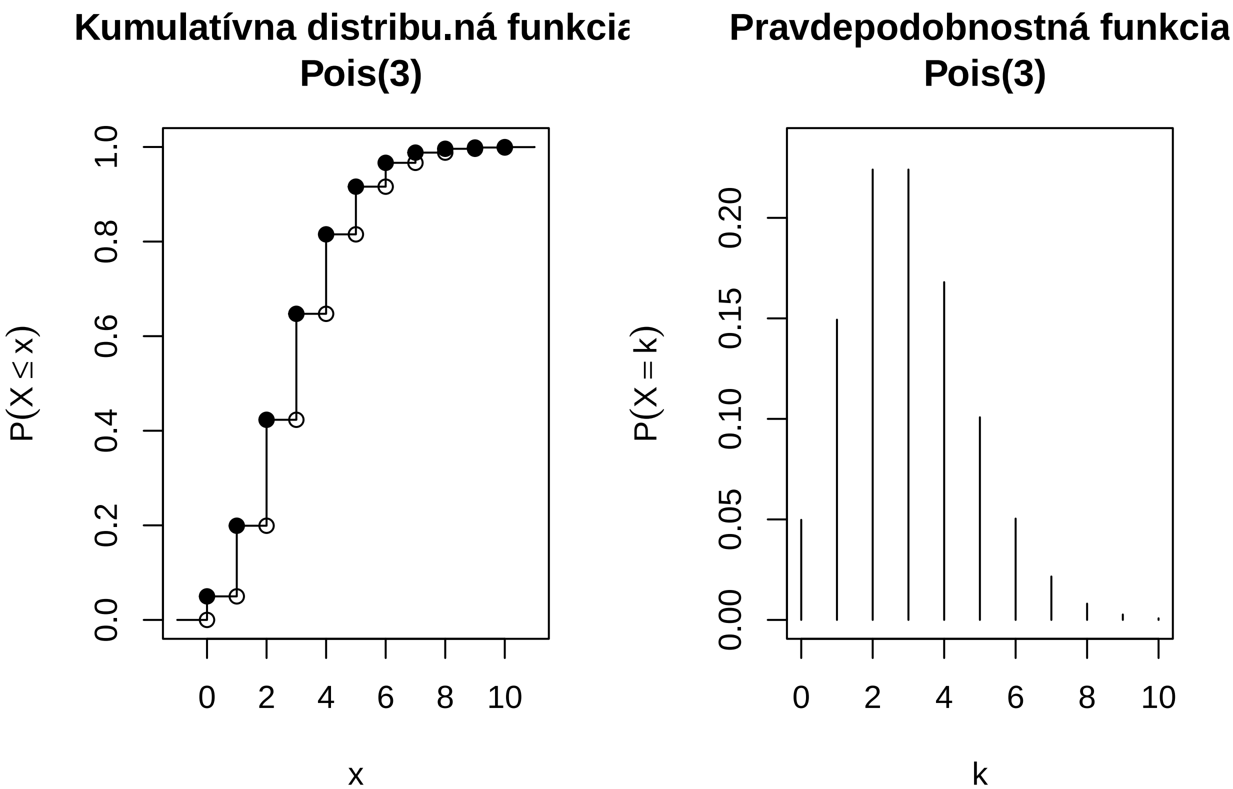
$$\begin{aligned} \mathrm{E}[X] &= \sum_{k=0}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!} \\ &= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\ &= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \\ &= \lambda e^{-\lambda} e^{\lambda} = \lambda, \\ \mathrm{Var}[X] &= \cdots = \lambda. \end{aligned}$$

Príklad 5.9. Majme náhodnú premennú X, ktorá je rozdelená Poissonovsky s parametrom $\lambda = 4$. Vypočítajte $P(X = 1)$.

$$P(X = 1) = \frac{4^1 e^{-4}}{1!} \approx 7.326\%.$$

Na to aby bola nejaká situácia, kde počítame koľko udalostí nastalo, dobre modelovaná Poissonovým rozdelením, musia platiť tieto skutočnosti:

- Hovoríme o počte nejakých udalostí, vecí v danom čase, priestore, ploche.
- Skutočnosť, že nastane jedna udalosť neovplyvní nastatie tej ďalšej.
- Udalosti nastávajú s rovnakou frekvenciou.
- Dve udalosti nemôžu nastať v ten istý čas, tom istom mieste, ploche.[20]

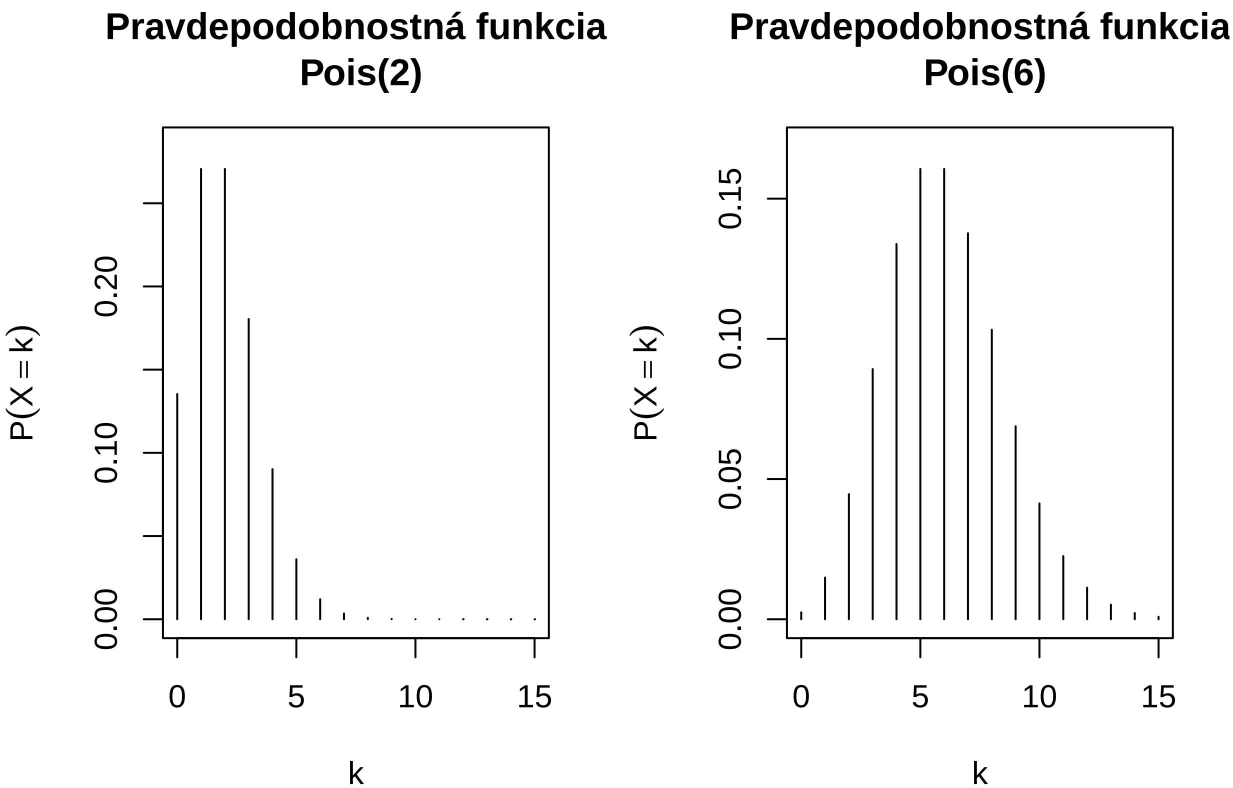


Nasledovné situácie môžu byť modelované Poissonovým rozdelením:

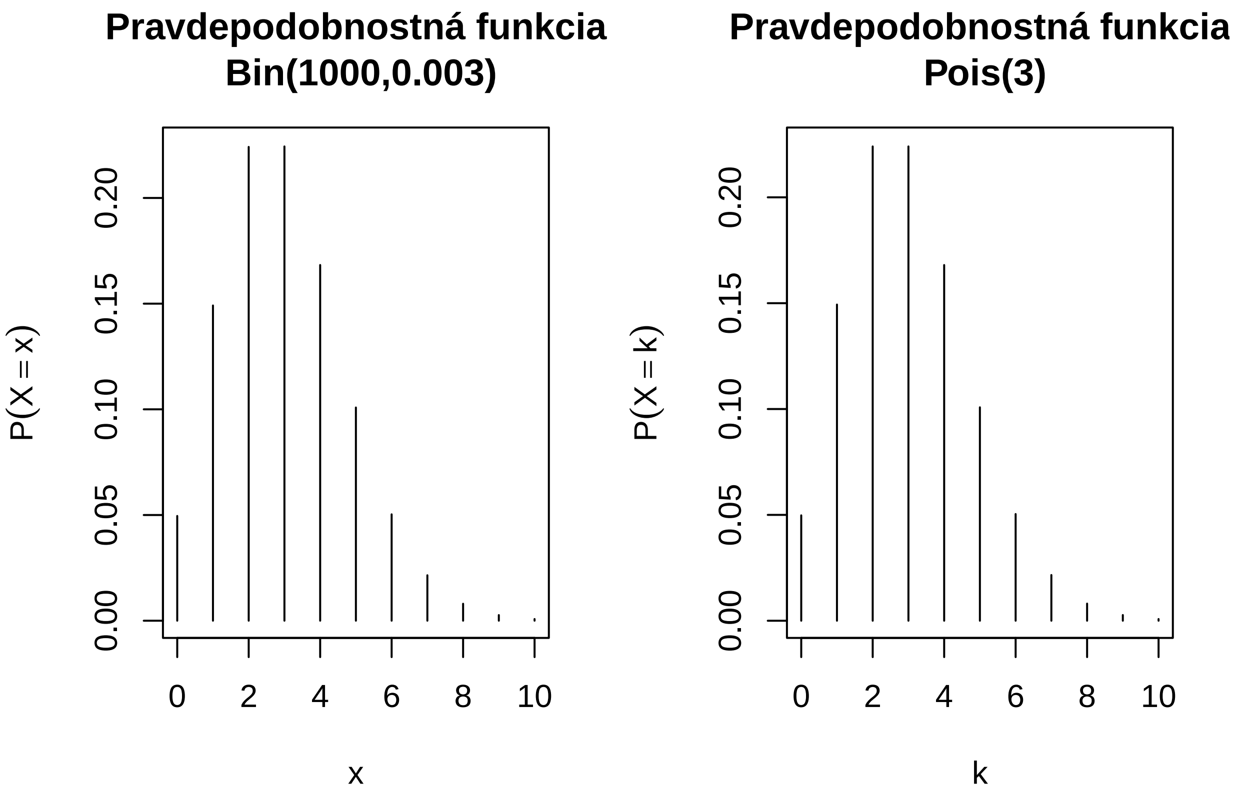
- Počet dopravných nehôd.
- Počet vzácnych druhov rakoviny.
- Počet návštevníkov obchodu.
- Počet zemetrasení.
- Počet rastlinných druhov na nejakom úseku zeme.
- Počet vyžiarených častíc z rádioaktívneho zdroja.
- Počet zranení spôsobených kopnutím koňa počas roka.
- Počet narodených detí v daný deň.
- Počet medveďov, ktoré stretneme v daný rok.

Nasledujúci obrázok porovnáva Poissonovsky rozdelené náhodné premenné pre rôzne parametre.

[20] Náhodný proces, ktorý modeluje takéto udalosti sa nazýva *Poissonov proces* a je dôležitým matematickým nástrojom v poisťovníctve. Budeme sa o ňom dopodrobna učiť neskôr.



Poissonovské rozdelenie $Pois(\lambda)$ veľmi dobre aproximuje binomické $\mathrm{Bin}(n,p)$ kde priemerná hodnota $\lambda = np$ a $n \to \infty$ a súčasne $p \to 0$. Nasledujúci obrázok porovnáva rozdelenia $\mathrm{Bin}(1000, 0.003)$ a $Pois(3)$. Sú na nerozoznanie.



Toto nám zároveň aj pomáha odpovedať na otázku, kedy je Poissonovské rozdelenie adekvátne. Napríklad vtedy, keď je binomický model adekvátny ale počet pokusov je veľký a pravdepodobnosť malá. Aká je pravdepodobnosť, že najbližsiu sekundu príde niekto na môj webovský server? Malá (p je malé číslo). Ale na druhej strane tých sekúnd je veľa (n je veľké čislo).[21]

5.7 Geometrické rozdelenie

Hovoríme, že diskrétna náhodná premenná X má **geometrické rozdelenie** s parametrom p ak platí

$$p_X(k) = \begin{cases} (1-p)^k p, & \text{ak } k \in \{0, 1, 2, \cdots\} \\ 0, & \text{inak.} \end{cases}$$

V tomto prípade používame nasledovné značenie $X \sim \text{Geom}(p)$.

Príklad 5.10. Majme náhodnú premennú X, ktorá je rozdelená geometricky s parametrom $p = 0.1$. Vypočítajte $P(X = 10)$. (Teda, že "úspech" nastane v 11. pokuse)

$$P(X = 10) = (1-p)^{10} p = 0.9^1 00.1 \approx 3.48\%.$$

Príklad 5.11. Majme náhodnú premennú X, ktorá je rozdelená geometricky s parametrom $p = 0.3$. Vypočítajte $P(X < 4)$.

$$\begin{aligned} P(X < 4) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= 0.7^0 0.3 + 0.7^1 0.3 + 0.7^2 0.3 + 0.7^3 0.3 \\ &= 0.3 + 0.21 + 0.147 + 0.1029 \approx 75.99\%. \end{aligned}$$

Geometrické rozdelenie je vhodné na modelovanie situácií, kedy čakáme na prvý úspech, ktorý prichádza v každom kroku s pravdepodobnosťou p. Náhodná premenná hovorí o počte neúspechov, kým nenastane prvý úspech.

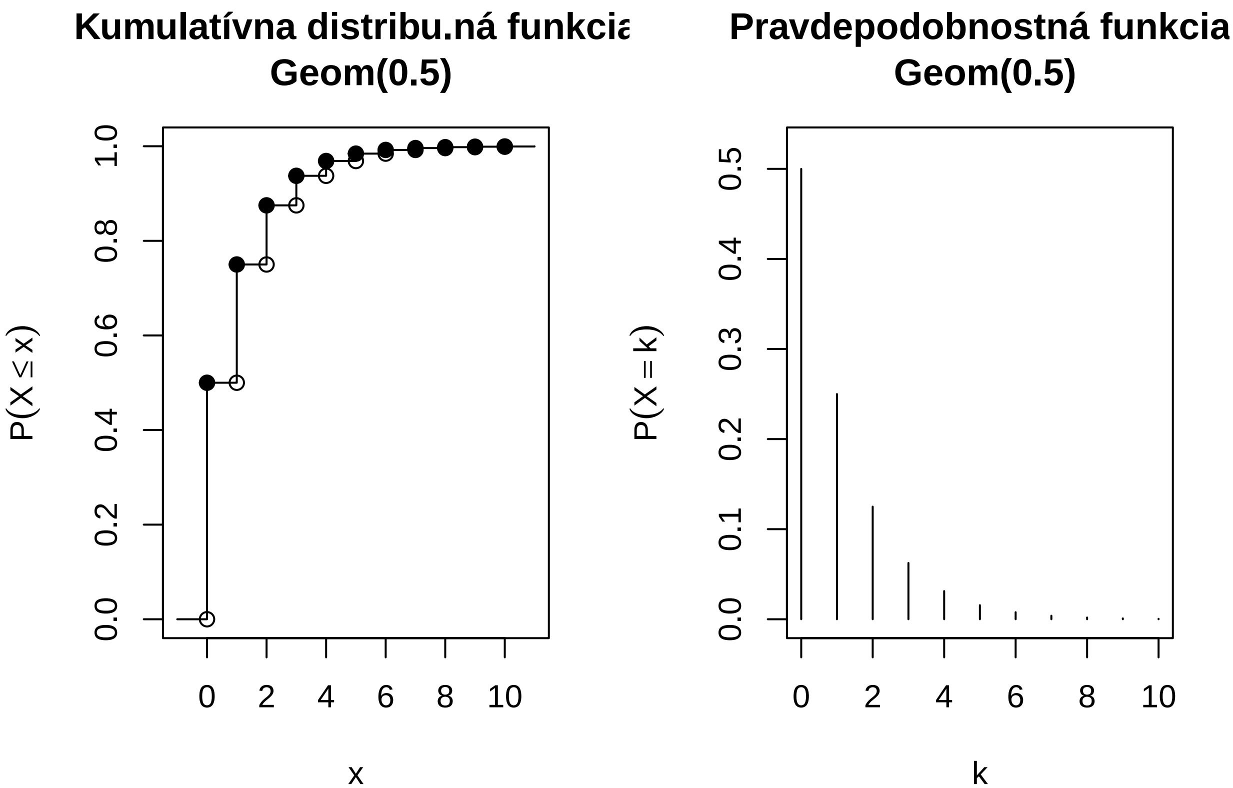
Stredná hodnota a variancia pre geometricky rozdelenú náhodnú premennú s parametrom p sú

$$\begin{aligned} \mathrm{E}[X] &= \sum_{k=0}^{\infty} k(1-p)^k p \\ &= p \sum_{k=0}^{\infty} k(1-p)^k \\ &= \frac{1-p}{p}, \\ \mathrm{Var}[X] &= \cdots = \frac{1-p}{p^2}. \end{aligned}$$

Čo musí byť splnené aby počet neúspechov, kým nastane prvý úspech bol adekvátne popísaný geometrickým rozdelením?

- jednotlivé pokusy musia byť nezávislé,
- jednotlivé pokusy musia mať rovnakú a fixnú pravdepodobnosť úspechu.

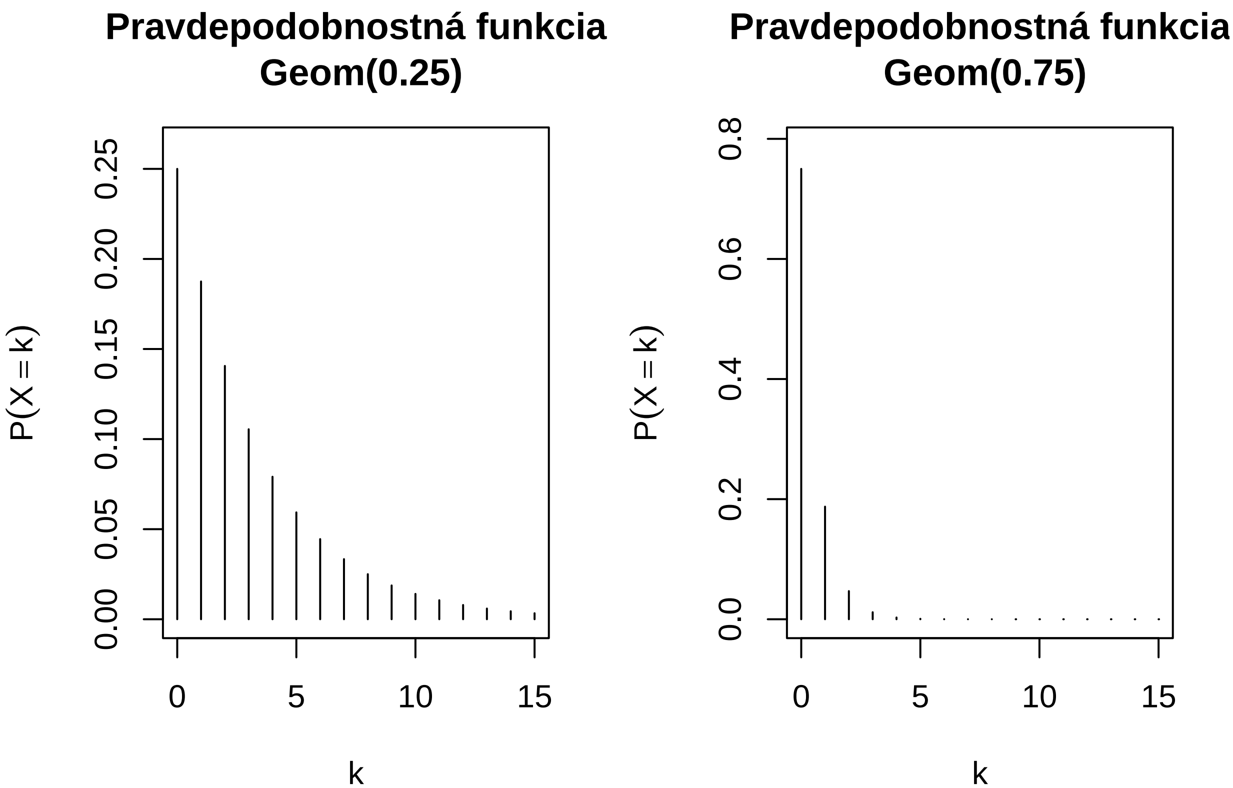
[21] Binomické rozdelenie pri veľkom počte pokusov má aj tú nevýhodu, že sa ťažšie vyčísluje numericky. Pre $X \sim \text{Bin}(1000, 0.001)$ vypočítať napríklad $p_X(20) = \binom{1000}{20} 0.001^{20} 0.999^{980}$ nie je jednoduché, lebo spolu násobíme obrovské aj maličké čísla.



Nasledovné situácie môžu byť modelované (ich adekvátnosť záleží od konkrétneho príkladu) geometrickým rozdelením:

- Počet ľudí, ktorých sa musím opýtať otázku, kým nenarazím na niekoho, kto bude vedieť odpovedať.
- Počet striel na bránu kým nepadne gól.
- Počet zákazníkov, kým nepríde nejaký, ktorý sa bude sťažovať.
- Počet vyrobených súčiastok, kým nenastane chyba.
- Počet opravených testov, kým niektorý študent dostane A.
- Počet rizikových startupov do ktorých treba investovať, kým nejaký z nich bude úspešný.
- Počet talentov, ktoré musí hľadač talentov vyskúšať, kým nenájde superhviezdu.

Tento obrázok porovnáva geometricky rozdelené náhodné premenné pre rôzne parametre.



5.8 Hypergeometrické rozdelenie

Hovoríme, že diskrétna náhodná premenná X má **hypergeometrické rozdelenie** s parametrami N, K a n ak platí

$$p_X(k) = \begin{cases} \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, & \text{ak } k \in \{\max\{0, n+K-N\}, \cdots, \min\{n, K\}\} \\ 0, & \text{inak.} \end{cases}$$

Označovať ju budeme ako $X \sim \text{HyperGeom}(N, K, n)$

Stredná hodnota pre hypergeometricky rozdelenú náhodnú premennú (dôvodenie vynechávame) je

$$\begin{aligned} \text{E}[X] &= \sum_{k=\max\{0,n+K-N\}}^{\min\{n,K\}} k \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} = \cdots = n\frac{K}{N}. \\ \text{Var}[X] &= \cdots = n\frac{K}{N}\frac{N-K}{N}\frac{N-n}{N-1}. \end{aligned}$$

Vo veľkom vreci máme N guľôčok, z toho K červených. Načrieme rukou a vyberieme hrsť o veľkosti n guličiek. Aká je šanca, že spomedzi týchto n guličiek je práve k červených?

Dôvodenie pre tvar $p_X(k)$ je nasledovné. Existuje práve $\binom{N}{n}$ rôznych spôsobov ako vybrať hrsť veľkosti n z celkovej množiny N guličiek, takže to je celkový počet hrstí. Koľko je "úspešných" hrstí? No existuje $\binom{K}{k}$ možností ako vybrať červené guličky a zároveň (preto je to súčin) $\binom{N-K}{n-k}$ možností ako vybrať modré guličky.

Príklad 5.12. Majme náhodnú premennú X, ktorá je rozdelená hypergeometricky s parametrami $N = 30, K = 8, n = 5$. Vypočítajte $P(X = 3)$.

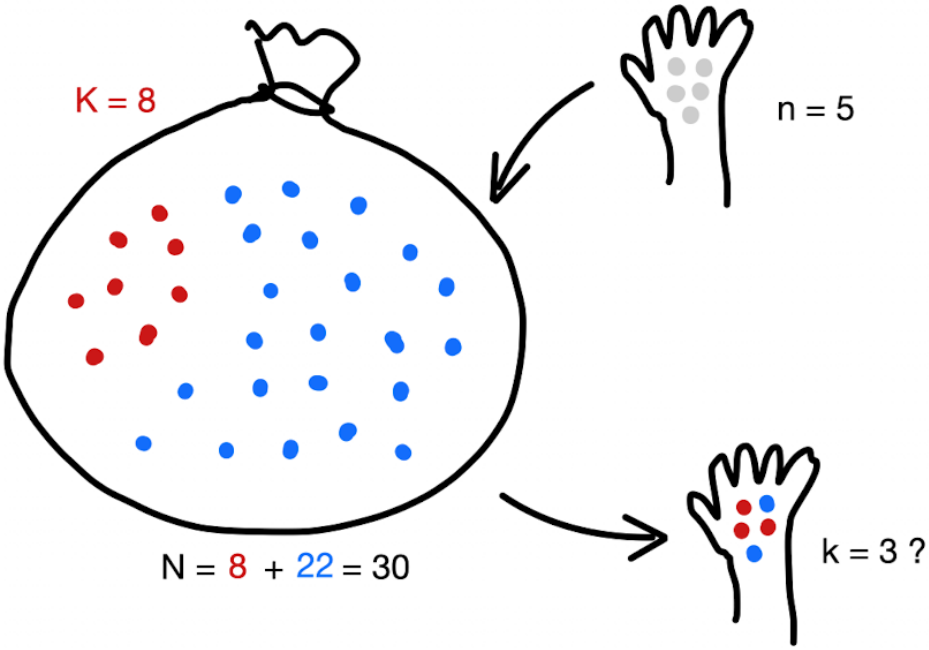
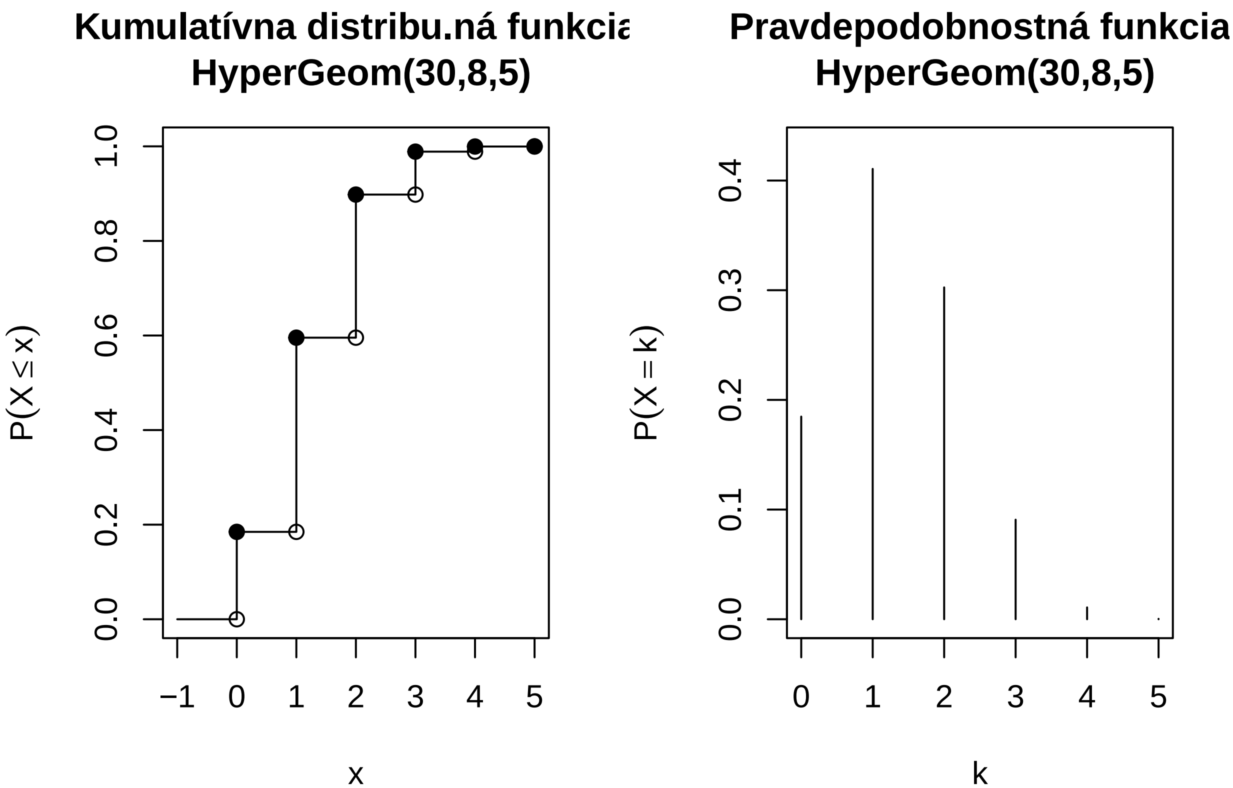


Figure 13: Ilustrácia hypergeometrickej distribúcie.

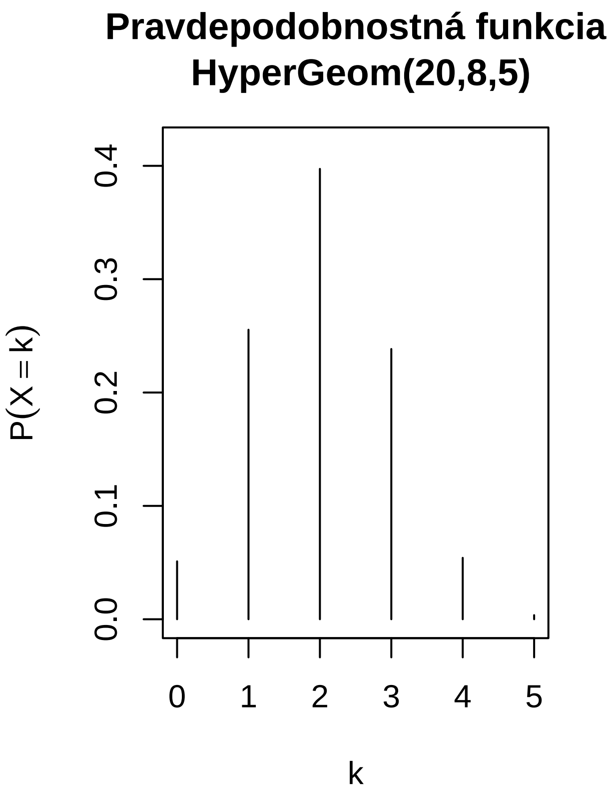
$$
\begin{aligned}
P(X=3) &= \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \\
&= \frac{\binom{8}{3}\binom{22}{2}}{\binom{30}{5}} \approx 9.08\%.
\end{aligned}
$$

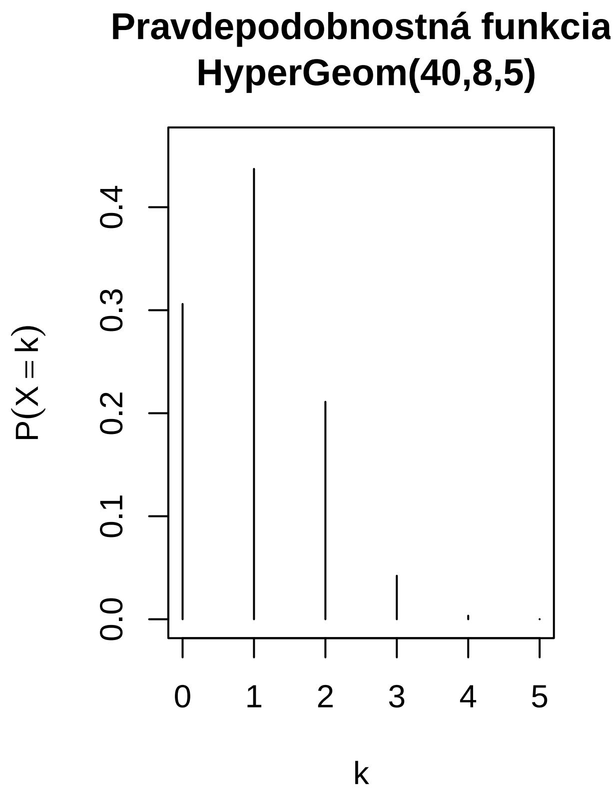


Hypergeometrické rozdelenie môže vhodne modelovať napríklad nasledovné situácie:

- Koľko je žien medzi náhodne vybranými 10 ľuďmi ako je v danom okrsku 100 mužov a 200 žien?
- Vieme, že medzi 200 súčiastkami je 10 chybných. Koľko chybných bude z náhodnej vybranej vzorky 20 súčiastok?
- Z 25 uchádzačov (10 so skúsenosťami, 15 bez skúseností) o prácu vyberú 4. Aká je šanca, že medzi náhodne vybranými 4 kandidátmi budú práve dvaja so skúsenosťami?

Nižšie sú pravdepodobnostné funkcie pre rôzne hodnoty N.





5.9 Negatívne binomické rozdelenie

Hovoríme, že diskrétna náhodná premenná X má **negatívne binomické rozdelenie** s parametrami r a p ak platí

$$p_X(k) = \begin{cases} \binom{k+r-1}{r-1}(1-p)^k p^r, & \text{ak } k \in \{0,1,2,3,\cdots\} \\ 0, & \text{inak.} \end{cases}$$

Označovať ju budeme ako $X \sim \mathrm{NegBin}(r,p)$

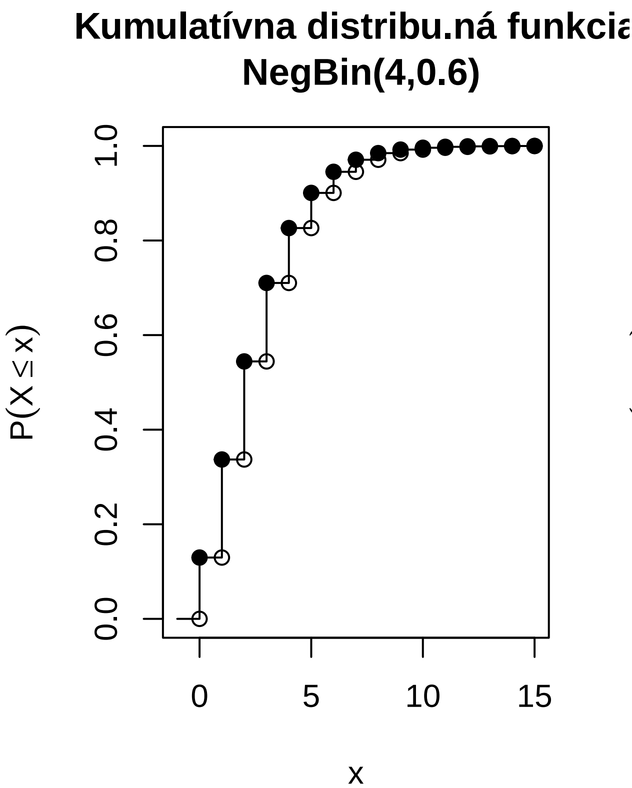
Stredná hodnota a variancia pre negatívne binomicky rozdelenú náhodnú premennú (dôvodenie vynechávame) sú

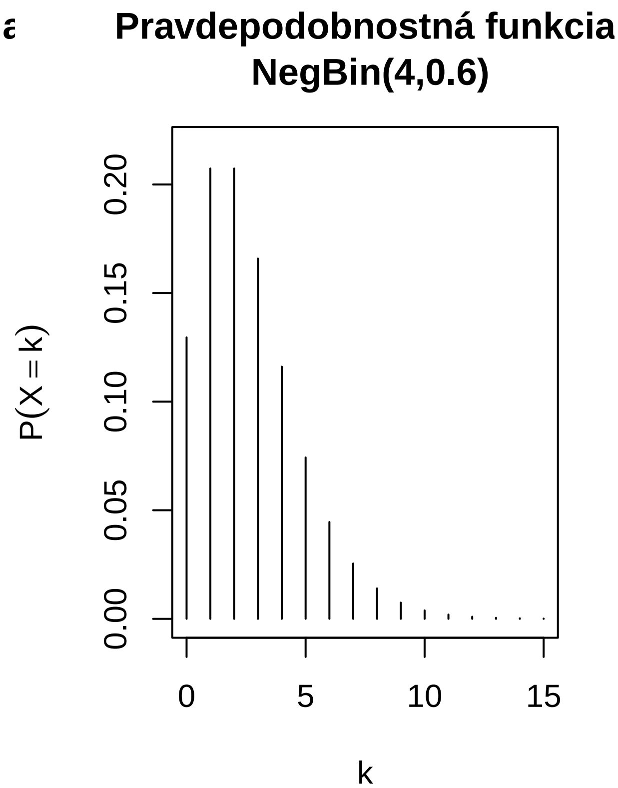
$$\begin{aligned} \mathrm{E}[X] &= \sum_{k=0}^{\infty} k\binom{k+r-1}{r-1}(1-p)^k p^r = \cdots = \frac{(1-p)r}{p}, \\ \mathrm{Var}[X] &= \cdots = \frac{(1-p)r}{p^2}. \end{aligned}$$

Takáto náhodná premenná popisuje počet neúspechov, kým nenastane r-tý úspech.

Príklad 5.13. Majme náhodnú premennú X, ktorá je rozdelená negatívne binomicky s parametrami $r = 4$ a $p = 0.6$. Vypočítajte $P(X = 3)$. (Toto zodpovedá 3 neúspechom, kým nenastanú 4 úspechy, takže 4. úspech nastane pri 7. pokuse.)

$$
\begin{aligned}
P(X=3) &= \binom{k+r-1}{r-1}(1-p)^k p^r \\
&= \binom{3+4-1}{4-1}(1-0.6)^3 0.6^4 \approx 16.59\%.
\end{aligned}
$$

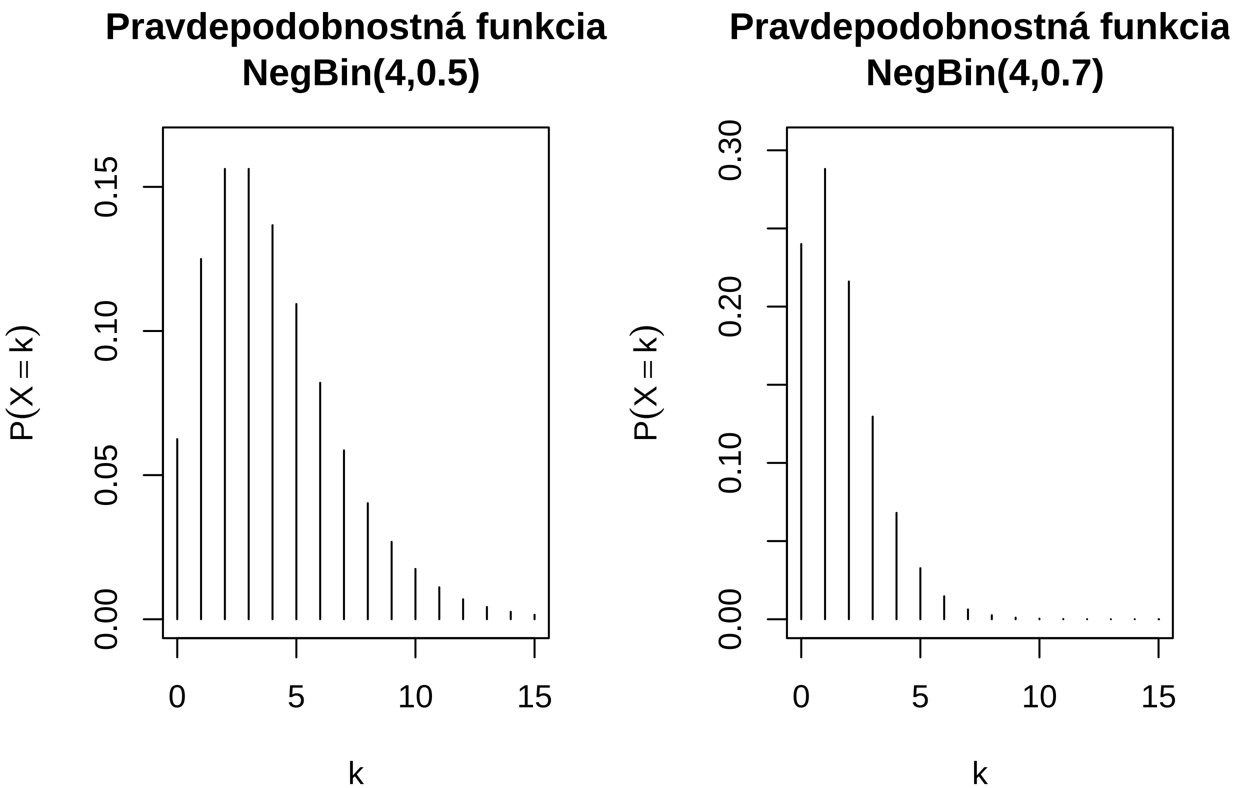




Tieto situácie môžu byť popísané negatívne binomickým rozdelením:

- Stroj zvládne 10 zlyhaní. Ako dlho bude bežať ak je pravdepodobnosť zlyhania na jednu časovú jednotku p?
- Ropná spoločnosť má prostriedky na 5 výskumných vrtov. Úspešnosť každého vrtu je 15%. Aká je šanca, že prvýkrát narazí na ropu pri štvrtom vrte?

Nižšie sú pravdepodobnostné funkcie pre rôzne parametre p:



5.10 Zhrnutie

Existujú rôzne diskrétne náhodné premenné. Niektoré skupiny náhodých premenných sú natoľko zaujímavé, že majú aj svoje špeciálne mená.

5.11 Cvičenia

Cvičenie 5.1. Ukážte, že nemôže existovať taká konštanta $c \in \mathbb{R}$, že by nasledovná funkcia popisovala pravdepodobnostnú funkciu náhodnej premennej X:

$$p_X(k) = \begin{cases} \frac{c}{k}, & \text{ak } k \in \{1, 2, \cdots\} \\ 0, & \text{inak.} \end{cases}$$

Cvičenie 5.2. Vypočítajte varianciu diskrétnej rovnomerne rozdelenej náhodnej premennej.

Cvičenie 5.3. Majme náhodnú premennú X, ktorá je rozdelená binomicky s parametrami $n = 10$ a $p = 0.25$. Vypočítajte $P(X < 4)$.

Cvičenie 5.4. Majme náhodnú premennú X, ktorá je rozdelená binomicky s parametrami $n = 8$ a $p = 0.8$. Vypočítajte $P(X > 6)$.

Cvičenie 5.5. Raketový modul má 6 tesnení. Každé z nich môže zlyhať s pravdepodobnosťou 0.01. Aká je pravdepodobnosť, že zlyhá nie viacej ako jedno tesnenie ?

Cvičenie 5.6. Na prejdenie desaťotázkového testu s možnosťami A, B, C, D, kde práve jedna odpoveď je správna, treba správne odpovedať na aspoň 5 otázok. Aká je šanca, že to študent zvládne bez učenia?

Cvičenie 5.7. Na prejdenie desaťotázkového testu s možnosťami A, B, C, D, kde práve jedna odpoveď je správna, treba správne odpovedať na aspoň 5 otázok. Na prvých 5 otázok vie študent odpovedať správne s pravdepodobnosťou 80%, druhých päť sa týka látky, ktorú nikdy nevidel. Aká je šanca, že tento študent zvládne tento test?

Cvičenie 5.8. Majme náhodnú premennú X, ktorá je rozdelená Poissonovsky s parametrom $\lambda = 8$. Vypočítajte $P(X = 2)$.

Cvičenie 5.9. Počet rastlinných druhov na exotickom ostrove je modelovaný $Pois(10)$ na každých $1m^2$. Aká je pravdepodobnosť, že na náhodne zvolenom metri štvorcovom nájdeme viacej ako 12 ale menej ako 15 rôznych rastlinných druhov?

Cvičenie 5.10. Majme náhodnú premennú X, ktorá je rozdelená geometricky s parametrom $p = 0.8$. Vypočítajte $P(X \geq 5)$.

Cvičenie 5.11. Majme náhodnú premennú X, ktorá je rozdelená geometricky s parametrom $p = 0.2$. Vypočítajte $P(X < 3)$.

Cvičenie 5.12. Majme náhodnú premennú X, ktorá je rozdelená hypergeometricky s parametrami $N = 10, K = 6, n = 5$. Vypočítajte $P(X = 3)$.

Cvičenie 5.13. Majme náhodnú premennú X, ktorá je rozdelená hypergeometricky s parametrami $N = 12, K = 5, n = 5$. Vypočítajte $P(X = 3)$.

Cvičenie 5.14 (*)**.** Zdôvodnite, prečo sa $\mathrm{Bin}(n,p)$ rozdelená náhodná premenná správa podobne ako $Pois(\lambda)$, pre $n \to \infty$ a $p \to 0$ a súčasne $\lambda = np$.

Cvičenie 5.15. Ak je šanca katastrofy každých sto rokov rovná $1/6$, aká je pravdepodobnosť, že nejaká katastrofa nastane počas nasledujúcich 500 rokov?

Cvičenie 5.16. Vypočítajte hodnotu

$$\sum_{j=2}^{n} j(j-1)\binom{n}{j}p^j(1-p)^{n-j}.$$

Cvičenie 5.17. Aké je pravdepodobnostné rozdelenie počtu uhádnutých otázok na ABCD teste z celkového počtu 10 otázok, ak vieme, že z prvých dvoch otázok otázok bola správne zodpovedaná práve jedna otázka?

Cvičenie 5.18. V klinickom skúšaní máme dve skupiny po 10 pacientov. V prvej skupine je pravdepodobnosť úspešnosti liečby 0.5, v druhej skupine je to 0.7. Uvažujme, že výsledky všetkých pacientov sú od seba nezávislé. Aká je pravdepodobnosť, že v prvej skupine bude aspoň toľko úspešne vyliečených ako a v druhej skupine?

Cvičenie 5.19. Presné testy sú drahé. Uvažujme nasledujúcu situáciu. Máme 1000 ľudí, ktorých potrebujeme otestovať. Pravdepodobnosť pozititívneho testu nech je 0.01. Namiesto toho, aby sme otestovali každého človeka osobitne. Tak spojíme odberové vzorky do 10 skupín po 100. Ak bude v nejakej skupine detekovaná pozitivita, tak pretestujeme všetkých 100 ľudí v tejto skupine.

- Koľko testov v priemere urobíme?
- Aká je pravdepodobnosť, že týmto spôsobom urobíme viacej testov ako 1000?

Cvičenie 5.20. Nech je pravdepodobnosť narodenia trojičiek $1/10000$. Aká je pravdepodobnosť, že z 8000 pôrodov sa narodia trojičky práve jedenkrát?

Cvičenie 5.21. Letecká prepravná spoločnosť predáva 200 lístkov napriek tomu, že v lietadle je len 198 miest nakoľko v priemere 3% ľudí neprídu. Toto je bežnou praxou a nazýva sa to *overbooking*. Aká je pravdepodobnosť, že si všetci pasažieri budú mať kde v lietadle sadnúť?

Cvičenie 5.22. Nech je v populácii 0.1% ľudí farboslepých. Aká je šanca, že v náhodnej vzorke 800 ľudí nebude viacej ako jeden človek farboslepý?

Cvičenie 5.23. V balíku je 37 gumených žížal. Z nich 13 sú kyslé gumené žížaly a 24 sú nekyslé gumené žížaly. Aká je šanca, že náhodne vybratá hrsť veľkosti 8 žížal bude mať práve 3 kyslé gumené žížaly.

Cvičenie 5.24. Uvažujme boxera, ktorý potrebuje 4 víťazné údery na to, aby knockoutoval súpera. Sám však znesie úderov 6, siedmy už nie. Je o čosi lepší ako jeho súper a šanca, že daná výmena skončí v jeho prospech je 0.53. Aká je šanca, že zápas skončí v jeho prospech?

Cvičenie 5.25. Chyba pri každom produkčnom cykle stroja je 0.015, denne zvládne stroj 8 cyklov. Stroj sa po 3 chybách zasekne a treba ho servisovať. Aká je šanca, že stroj bude fungovať bez zastavenia celé dva pracovné týždne?

Cvičenie 5.26. Výskumníčka musí získať 20 dotazníkových odpovedí. Každý človek odpovie na žiadosť odoslať dotazník s pravdepodobnosťou 40%. Aká je pravdepodobnosť, že musí výskumníčka osloviť viacej ako 50 ľudí?

Cvičenie 5.27. Ukážte, že limita pravdepodobnostnej funkcie hypergeometricky rozdelenej náhodnej premennej sa blíži k pravdepodobnostnej funkcii binomicky rozdelenej náhodnej premennej pre $M/N \to p$.

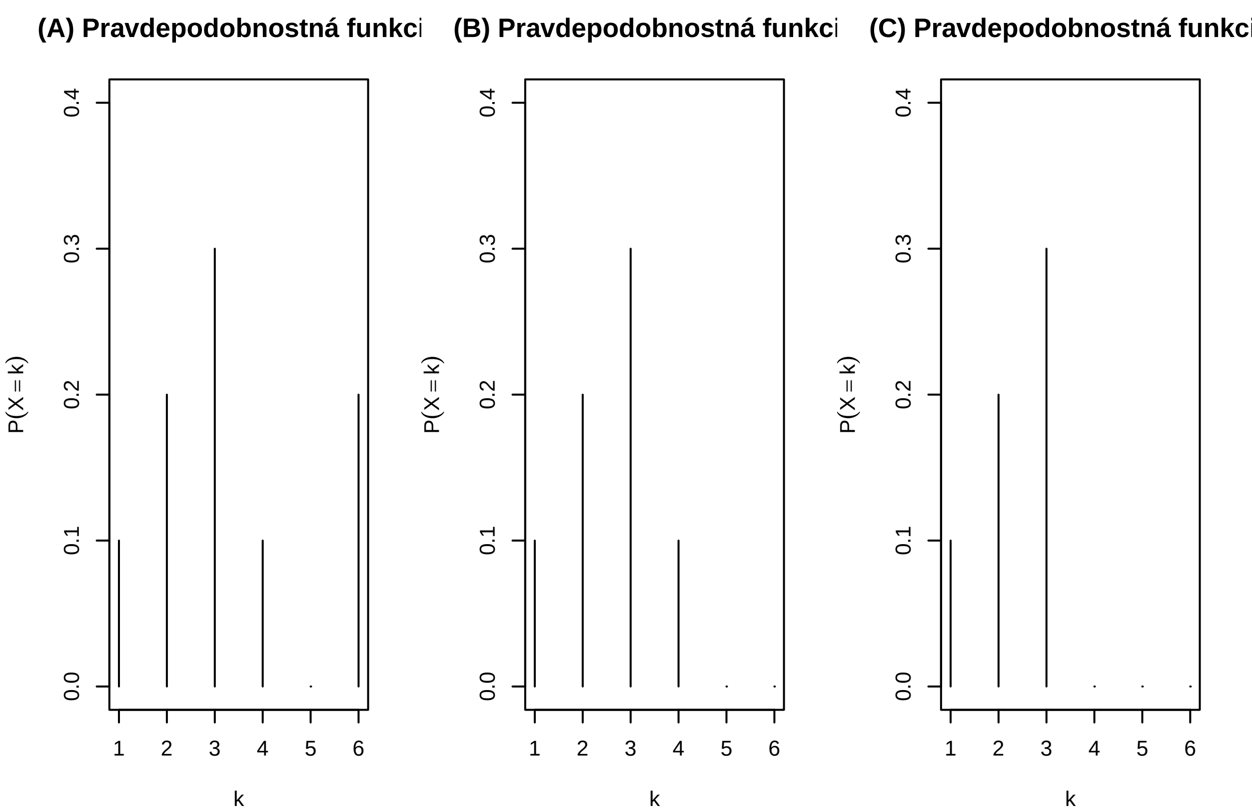
Cvičenie 5.28. Odvoďte varianciu pre náhodnú premennú $X \sim \text{Bin}(n, p)$.

Cvičenie 5.29.

(A) Doplňte veľkosť chýbajúcej hodnoty $p_X(5)$.

(B) Doplňte veľkosť chýbajúcej hodnoty $p_X(5)$ a $p_X(6)$ tak, aby $\text{E}[X] = 3.6$.

(C) Doplňte veľkosť chýbajúcej hodnoty $p_X(4), p_X(5)$ a $p_X(6)$ tak, aby $\text{E}[X] = 3.2$ a $\text{Var}[X] = 1.56$.



Cvičenie 5.30. Máme náhodnú premennú, ktorej realizácie sú 7,8,8,9,8,7,8,8,7,6,7,8,8,8,7,6,7,6,7,7,8,9,8,7,9,6,9,9,6,6,9,...

Čo by ste povedali, aká je približne stredná hodnota?

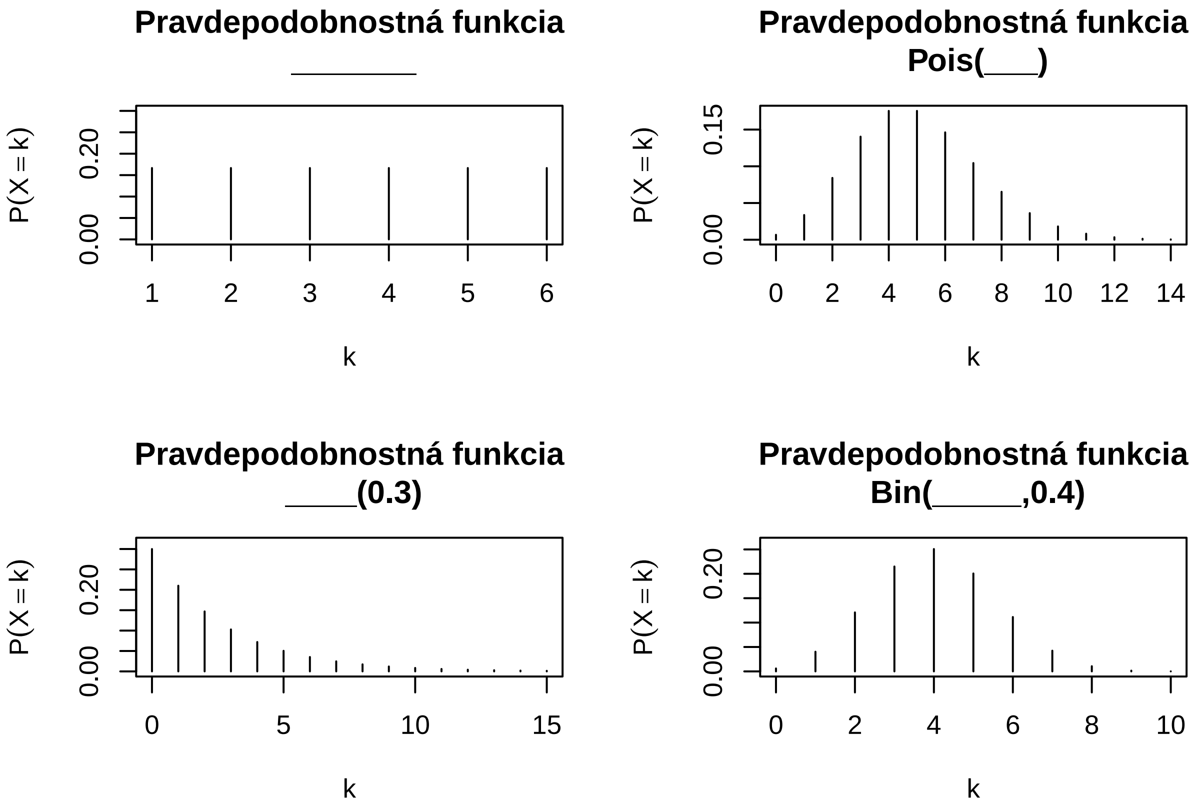
- 6
- 6.5
- 7
- 7.5
- 8

- 8.5
- 9

Čo by ste povedali, aká je približne variancia tejto náhodnej premennej?

- -1
- 0
- 1
- 4
- 8
- 16

Cvičenie 5.31. Doplňte



Cvičenie 5.32. Prepojte

(A) Rovnomerné rozdelenie.

(B) Bernoulliho rozdelenie.

(C) Binomické rozdelenie.

(D) Poissonovo rozdelenie.

(E) Geometrické rozdelenie.

(F) Hypergeometrické rozdelenie.

(G) Negatívne binomické rozdelenie.

s nasledujúcimi situáciami/úlohami

(1) Počet správne uhádnutých odpovedí na externej maturitnej skúške.

(2) Šanca, že zo siedmych gumených medvedíkov, ktoré si vyberieme z balíčka je práve 5 červenej farby. V balíčku je 80 macíkov (rodinné balenie) a z nich 33 je červených.

(3) Šanca, že sa nám podarí urobiť skúšku bez učenia, keď máme právo na dva opravné termíny.

(4) Počet rastlinných druhov na náhodne vybranom kúsku pôdy.

(5) Šanca, že náhodne vybraná fixka je modrá.

(6) Výhra v stávkovej spoločnosti, kde kurz implikuje pravdepodobnosť výhry 36%.

(7) Rok, kedy konečne prvýkrát stihnem urobiť daňové priznanie bez odkladu.

6 Spojitá náhodná premenná

Doteraz sme hovorili o náhodných premenných, pre ktoré bola množina hodnôt, ktoré nadobúdali, konečná alebo nanajvýš spočitateľná. Existuje veľká skupina situácií, kedy je lepšie uvažovať o náhodnej premennej, ktorá nadobúda nespočitateľne veľa hodnôt, napríklad akékoľvek reálne čislo.

6.1 Funkcia hustoty spojite rozdelenej náhodnej premennej

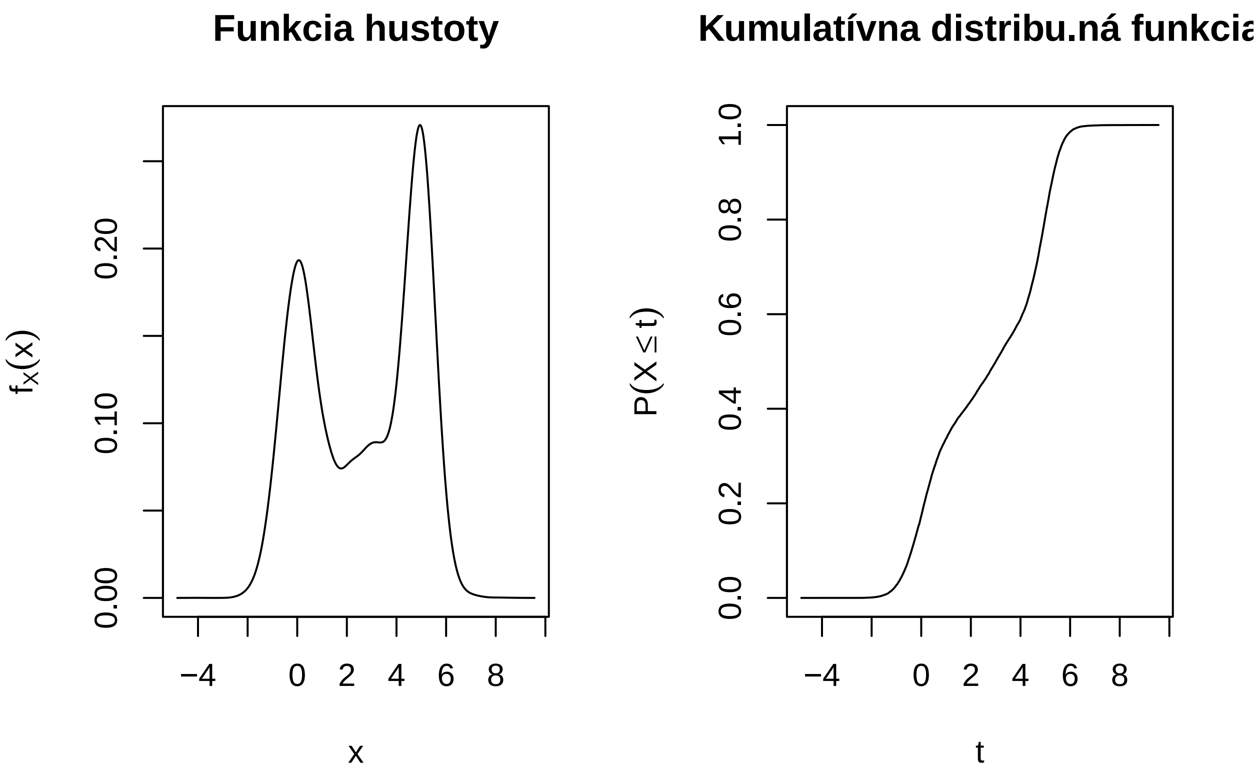
Popisovať náhodnosť pre tieto premenné si vyžaduje iný prístup. Pre diskrétne náhodné premenné stačilo pravdepodobnosti *vymenovať*. Teraz to nie je možné, lebo ich je príliš veľa (konkrétne nespočitateľne veľa) a pravdepodobnosť toho, že náhodná premenná bude nadobúdať práve nejakú konkrétnu hodnotu bude 0.

Na popis náhodnosti môžeme stále používať kumulatívnu distribučnú funkciu, ale namiesto pravdepodobnostnej funkcie budeme používať **funkciu hustoty** (alternatívne **funkciu hustoty spojite rozdelenej náhodnej premennej**.)

Hovoríme, že spojite rozdelená náhodná premenná X má funkciu hustoty $f_X(x)$, kde $f_X : \mathbb{R} \to \mathbb{R}$, ak platí:

$$P(a \leq X \leq b) = P(X \in [a, b]) = \int_a^b f_X(x)dx$$

Nižšie je príklad, ako môže vyzerať funkcia hustoty spojite rozdelenej náhodnej premennej a jej príslušná kumulatívna distribučná funkcia.



Funkcia hustoty f_X má rôzne vlastnosti:

- $$P(X = a) = P(a \leq X \leq a) = \int_a^a f_X(x)dx = 0,$$

  teda pravdepodobnosť udalosti, že náhodná premenná nadobúda konkrétnu hodnotu a je nula, a to dokonca pre úplne hocijakú hodnotu $a \in \mathbb{R}$.
- $f_X(x) \geq 0$ pre všetky $x \in \mathbb{R}$ - hustota pravdepodobnosti musí byť všade nezáporná.
- $F_X(t) = \int_{-\infty}^t f_X(x)dx$ - hovorí o vzťahu medzi kumulatívnou distribučnou funkciou a funkciou hustoty. Pri spojite rozdelenej náhodnej premennej nemáme problém s derivovaním funkcie F_X. Pripomeňme, že pri diskrétne rozdelenej náhodnej premennej má funkcia F_X "skoky" v bodoch x, kde $p_X(x) > 0$, a teda v týchto bodoch nie je funkcia F_X spojitá, takže ani diferencovateľná.
- $P(X \in \mathbb{R}) = \int_{-\infty}^{\infty} f_X(x)dx = 1$ - tu ide o analógiu k $\sum_{x \in \mathcal{S}_X} p_X(x) = 1$ v diskrétnom prípade. Pravdepodobnosť, že náhodná premenná nadobúda *nejakú* hodnotu, je 1.
- $f_X(x) = \frac{d}{dx}F_X(x)$ - toto vyplýva zo vzťahu medzi deriváciou a integrálom, nazýva sa *základná (fundalmentálna) veta integrálneho počtu.* Kumulatívna distribučná funkcia F_X je primitívnou funkciou funkcie hustoty f_X.

O funkcii hustoty môžeme uvažovať tak, že

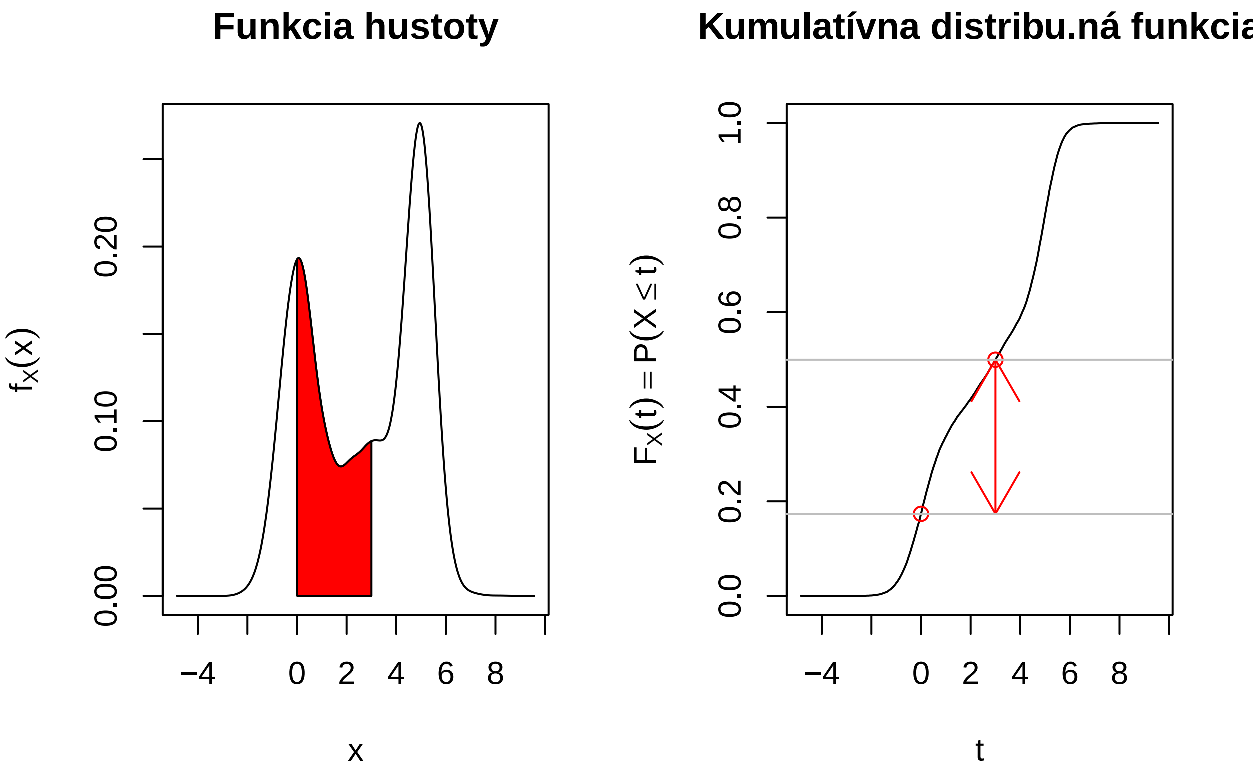
$f(x)\Delta \approx \int_x^{x+\Delta} f(s)ds = P(x \leq X \leq x + \Delta) = F_X(x + \Delta) - F_X(x)$, pre nejaké malé číslo Δ.

Vo všeobecnosti máme

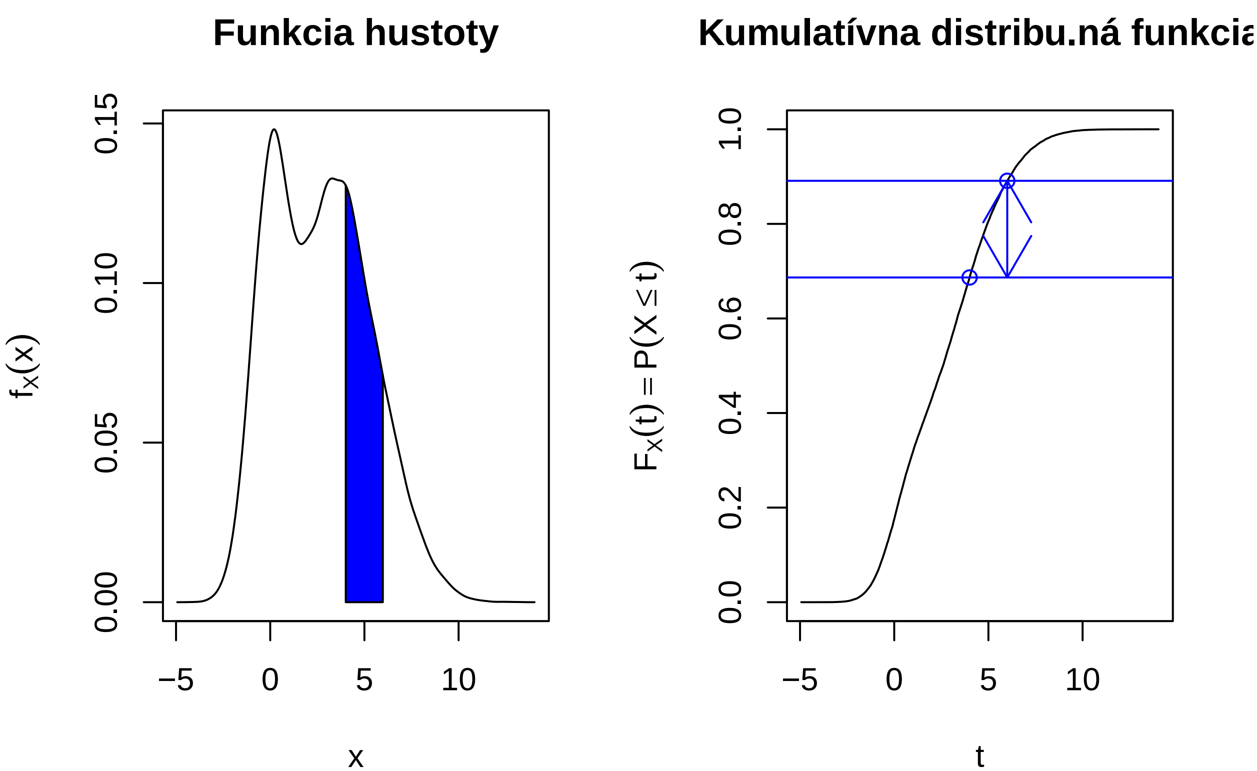
$$P(X \in A) = \int_A f_X(x)dx,$$

kde symbol $\int_A$ značí integrovanie cez oblasť A.

Nasledujúci obrázok ilustruje ako sa vypočíta pravdepodobnosť udalosti, že náhodná premenná padne do intervalu $(0,3)$, teda $P(X \in [0,3]) = P(0 \leq X \leq 3) = \int_0^3 f_X(x)dx = F_X(3) - F_X(0)$

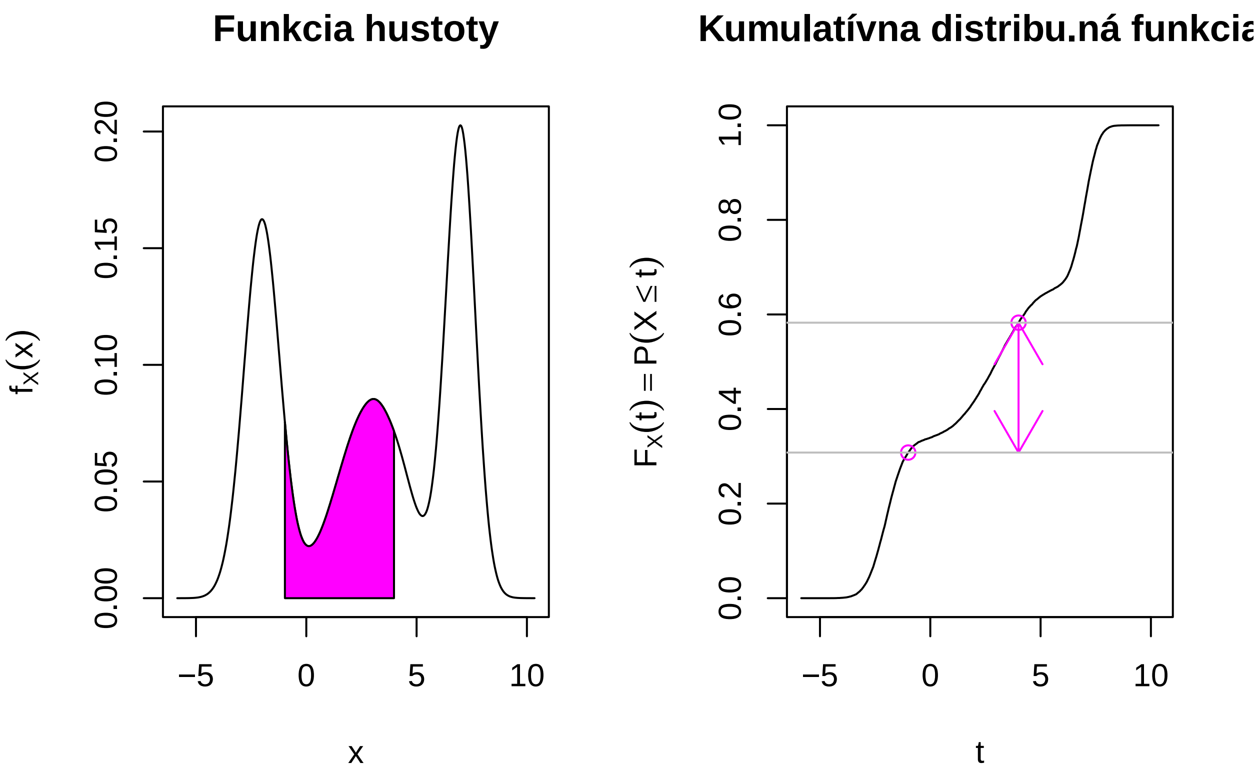


Ďalší obrázok ukazuje $P(X \in [4,6]) = P(4 \leq X \leq 6) = \int_4^6 f_X(x)dx = F_X(6) - F_X(4)$ pre inú funkciu hustoty:

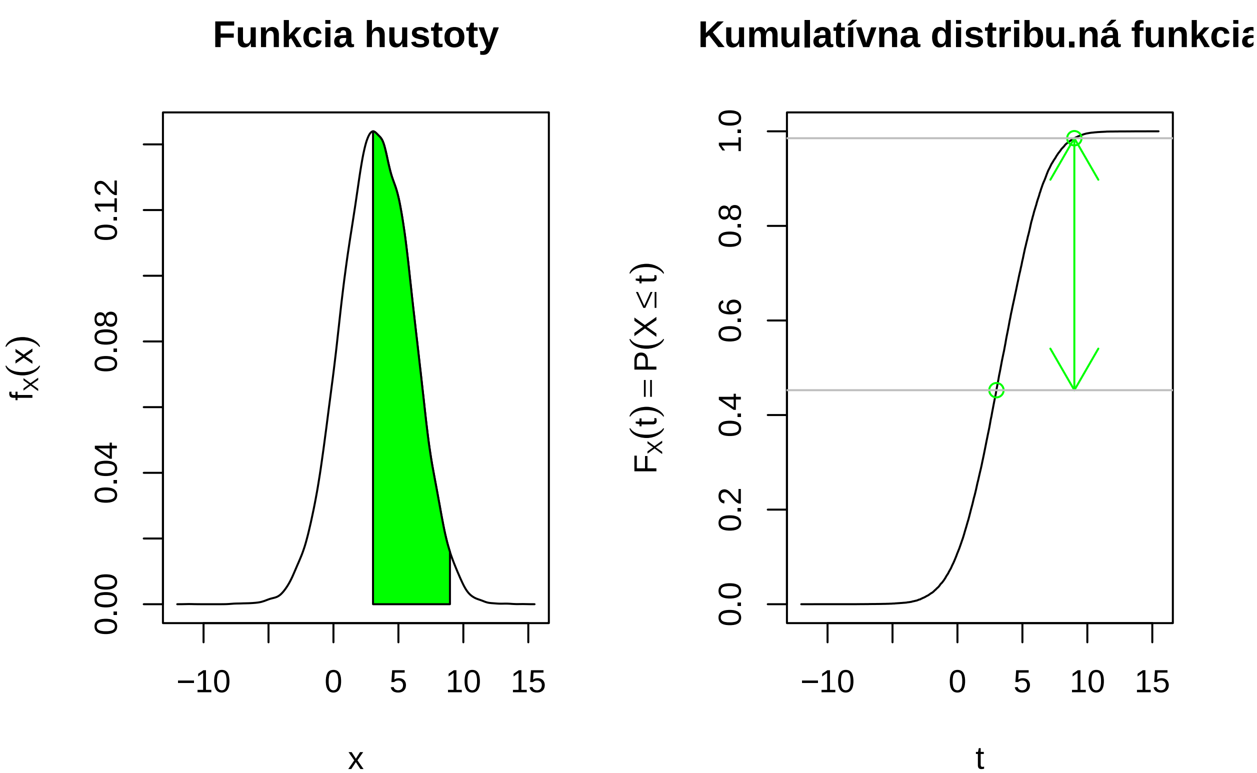


Plocha pod grafom funkcie hustoty, teda integrál funkcie f_X cez nejaký interval, určuje pravdepodobnosť, že náhodná premenná nadobudne hodnotu v tomto intervale.

Tu je ilustrácia pre nejakú inú funkciu hustoty a $P(X \in [-1,4]) = P(-1 \leq X \leq 4) = \int_{-1}^{4} f_X(x)dx = F_X(4) - F_X(-1)$

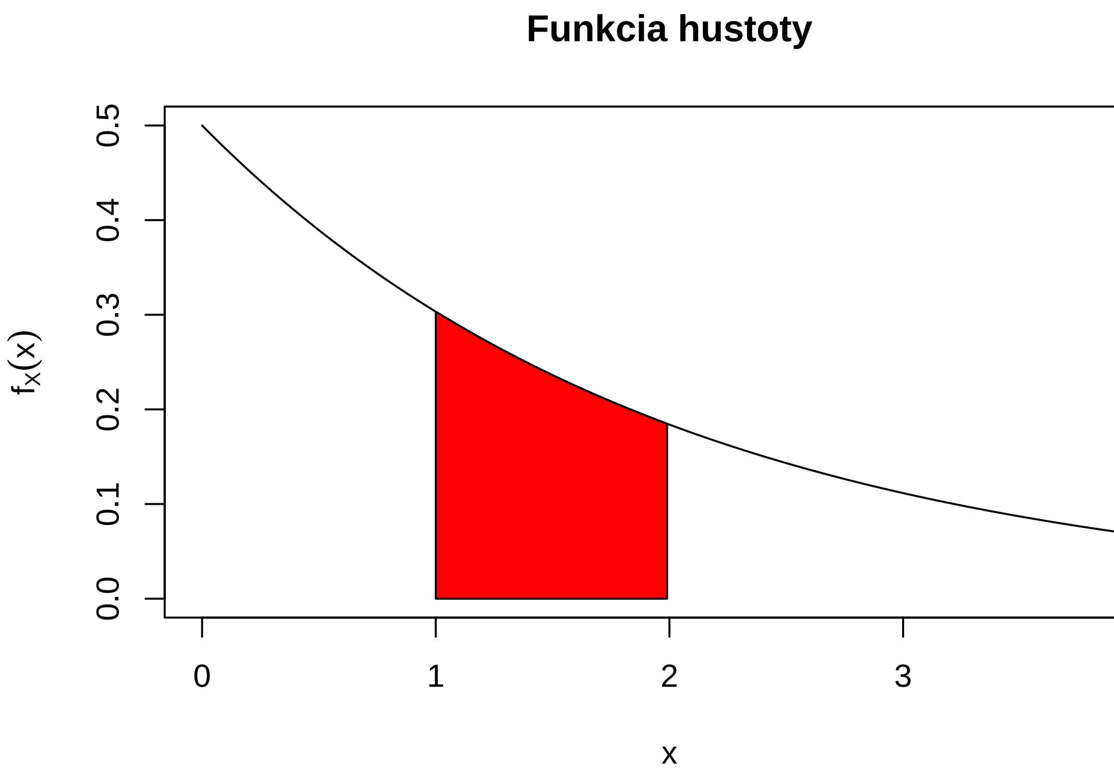


A ďalší obrázok ukazuje inú funkciu hustoty a $P(X \in [3,9]) = P(3 \leq X \leq 9) = \int_{3}^{9} f_X(x)dx = F_X(9) - F_X(3)$



Príklad 6.1. Majme náhodnú premennú X, ktorá má nasledovnú funkciu hustoty: $f_X(x) = \frac{1}{2}e^{-x/2}$ pre $x > 0$, inak $f_X(x) = 0$. Vypočítajte $P(1 \leq X \leq 2)$.

$$\begin{aligned} P(1 \leq X \leq 2) &= \int_1^2 f_X(x)dx = \int_1^2 \frac{1}{2}e^{-x/2}dx \\ &= [-e^{-x/2}]_1^2 = [-e^{-2/2}] - [-e^{-1/2}] \\ &= e^{-1/2} - e^{-1} \approx 23.87\%. \end{aligned}$$

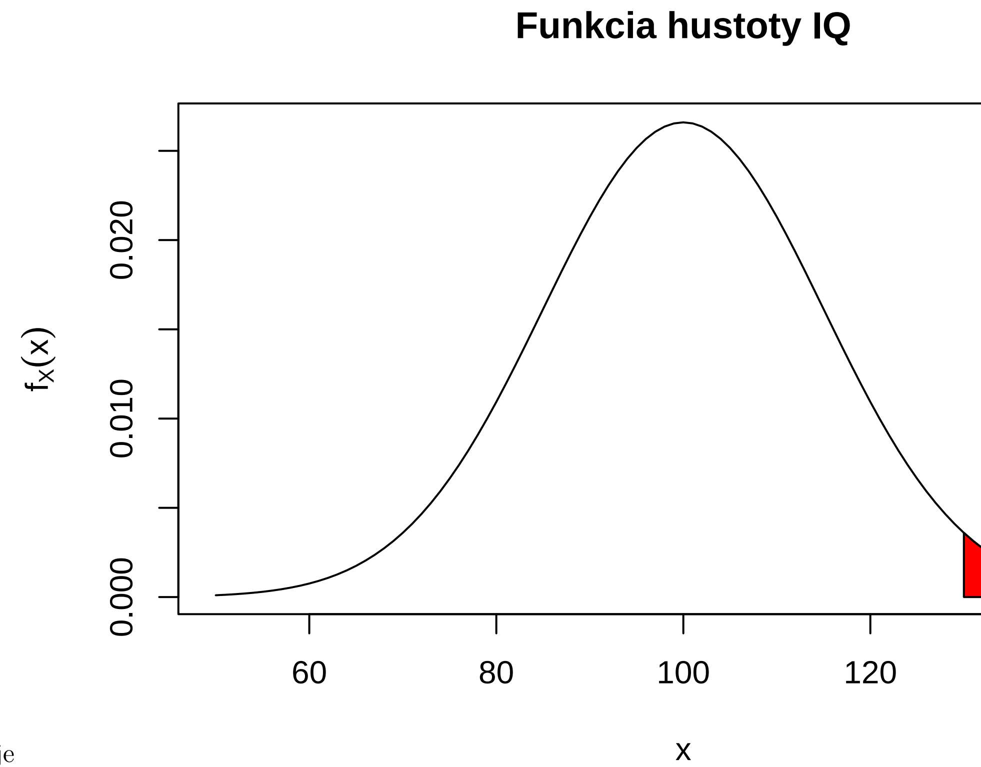


Situácia je ilustrovaná nižšie:

Príklad 6.2. Uvažujme náhodnú premennú X, ktorá modeluje IQ náhodne vybraného človeka z danej populácie. Táto náhodná premenná má nasledovnú funkciu hustoty $f_X(x) = \frac{1}{15\sqrt{2\pi}}e^{-\frac{1}{2}\frac{(x-100)^2}{15^2}}$. Vypočítajte pravdepodobnosť, že náhodne vybraný človek z tejto populácie bude mať IQ väčšie ako 130.

$$\begin{aligned} P(X > 130) &= 1 - P(X \leq 130) = 1 - \int_{-\infty}^{130} f_X(x)dx = 1 - \int_{-\infty}^{130} \frac{1}{15\sqrt{2\pi}}e^{-\frac{1}{2}\frac{(x-100)^2}{15^2}}dx \\ &= \cdots \approx 1 - 0.9772 = 2.28\%. \end{aligned}$$

Tento integrál nemá analytické ("pekné") riešenie a treba ho zrátať numericky pomocou počítačového programu.



Nasledujúci obrázok túto situáciu ilustruje

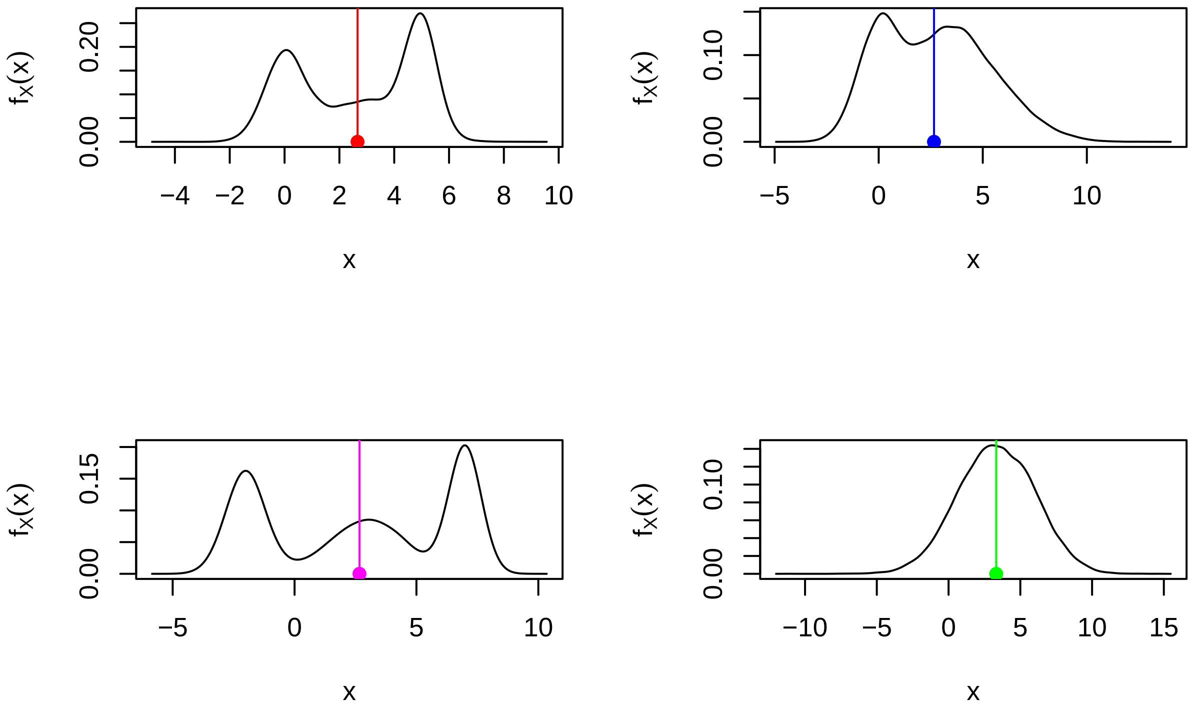
6.2 Charakteristiky spojitých náhodných premenných

Stredná hodnota spojite rozdelenej náhodnej premennej je

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx,$$

ak $\int_{-\infty}^{\infty} |x| f_X(x) dx < \infty$. Teda ide o hodnoty x váhované hustotou pravdepodobnosti $f_X(x)$.

Nasledujúci obrázok ukazuje stredné hodnoty pre náhodné premenné s rôznymi funkciami hustoty:



Stredná hodnota má aj pre spojite rozdelenú náhodnú premennú interpretáciu ťažiska. Ak by sme si predstavili funkciu hustoty ako tenký kovový plát, tak stredná hodnota je miesto, v ktorom by sme museli tento plát podoprieť tak, aby bol v rovnováhe.

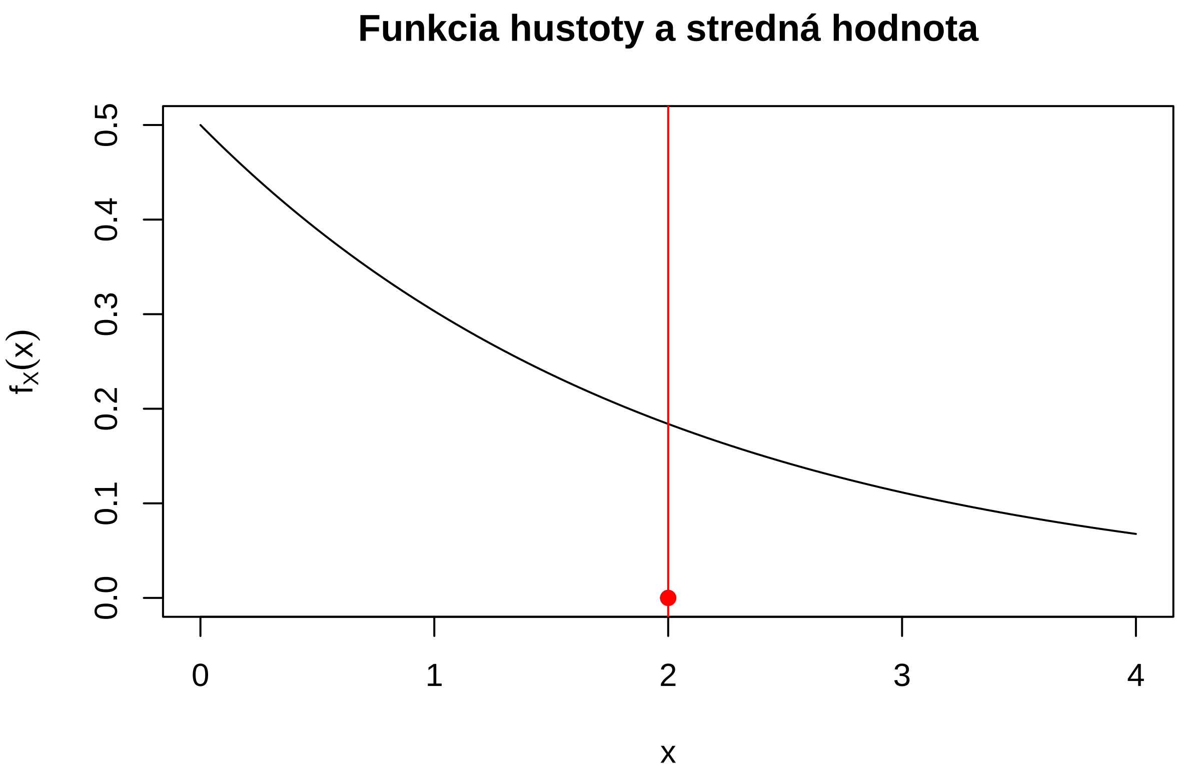
Pri počítaní stredných hodnôt ale aj pri počítaní pravdepodobností si budeme musieť spomenúť ako sa *integruje*. Existuje konečný počet trikov, ktoré sú postačujúce na vypočítanie veľkej väčšiny intergrálov, s ktorými sa stretnete v rámci pravdepodobnosti a štatistiky:

- nájdenie primitívnej funkcie
- integrovanie per partes
- substitúcia

Iné techniky, ako napríklad využitie trigonimetrických identít alebo rozklad na čiastočné zlomky, sa pre naše potreby budú využívať len zriedka, až vôbec.

Príklad 6.3. Majme náhodnú premennú X, ktorá má nasledovnú funkciu hustoty $f_X(x) = \frac{1}{2}e^{-x/2}$ pre $x > 0$ inak $f_X(x) = 0$. Vypočítajte $\mathrm{E}[X]$.

$$
\begin{aligned}
\mathrm{E}[X] &= \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^{\infty} x \frac{1}{2} e^{-x/2} dx \\
&= \left[x.\left(-e^{-\frac{x}{2}}\right)\right]_0^{\infty} - \int_0^{\infty} 1.\left(-e^{-\frac{x}{2}}\right) dx \\
&= 0 - 0 - \left[2e^{-\frac{x}{2}}\right]_0^{\infty} = 0 - (-2) = 2.
\end{aligned}
$$



Stredná hodnota stransformovanej náhodnej premennej $g(X)$, kde $g : \mathbb{R} \to \mathbb{R}$, je

$$\mathrm{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx.$$

Poznamenajme, že aj pre spojitú náhodnú premennú platí $\mathrm{E}[aX + bY] = a\mathrm{E}[X] + b\mathrm{E}[Y]$, teda stredná hodnota je lineárny operátor.

Variancia hodnota spojite rozdelenej náhodnej premennej je

$$\mathrm{Var}[X] = \mathrm{E}[(X - \mathrm{E}[X])^2] = \int_{-\infty}^{\infty} (x - \mathrm{E}(X))^2 f_X(x) dx,$$

ak $\int_{-\infty}^{\infty} |x^2| f_X(x) dx < \infty$. Aj pre spojite rozdelenú náhodnú premennú platí $\mathrm{Var}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$.

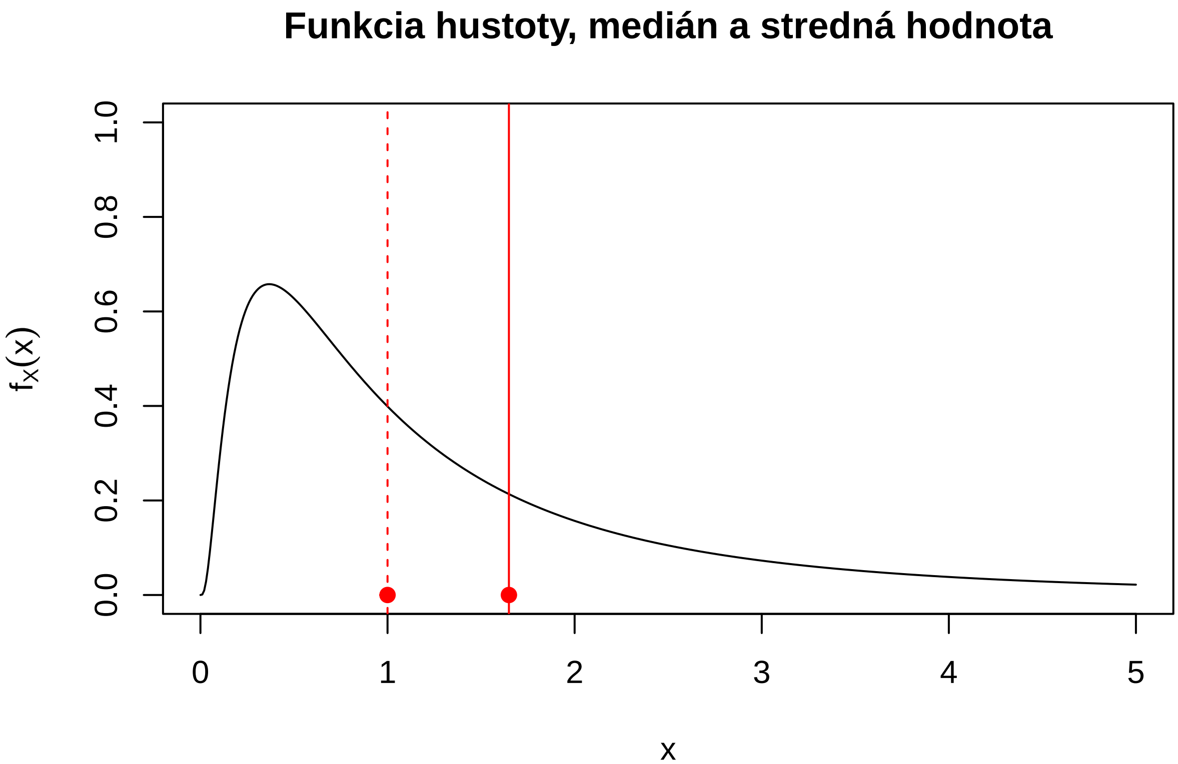
Príklad 6.4. Majme náhodnú premennú X, ktorá má nasledovnú funkciu hustoty $f_X(x) = \frac{1}{2} e^{-x/2}$ pre $x > 0$, inak $f_X(x) = 0$. Vypočítajte $\mathrm{Var}[X]$.

$$\begin{aligned} \mathrm{E}[X] &= \ldots [\text{ukázali sme vyššie}] \cdots = 2 \\ \mathrm{E}[X^2] &= \ldots [\text{dva krát per partes}] \cdots = 8 \\ \mathrm{Var}[X] &= \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = 8 - 2^2 = 4 \end{aligned}$$

Medián spojite rozdelenej náhodnej premennej X je taká hodnota M, pre ktorú platí

$$P(X \le M) = P(X \ge M) = 0.5.$$

Medián je taká hodnota, že napravo aj naľavo od nej je 50% pravdepodobnostnej masy. Ľudia sa často čudujú, keď v médiách zaznie informácia o priemernej mzde v hospodárstve. Zdá sa im priveľká. Mnoho ľudí totiž intuitívne stotožňuje priemer s mediánom, tieto hodnoty však môžu byť výrazne iné, najmä pri rozdeleniach, ktoré sú typické pre mzdy. Na nasledujúcom obrázku je taká naklonená funkcia hustoty (nech sú to mesačné mzdy v tisísoch eur). Kým medián (prerušovaná čiara) je 1, tak stredná hodnota (plná čiara) je približne 1.65. Takže priemerná mzda v tomto hospodárstve je 1650eur ale len polovica ľudí má plat aspoň 1000eur. Preto informácia v televíznych správach o priemernej mzde sa môže zdať mnohým ľuďom privysoká.



Príklad 6.5. Majme náhodnú premennú X, ktorá má nasledovnú funkciu hustoty $f_X(x) = \frac{1}{2}e^{-x/2}$ pre $x > 0$, inak $f_X(x) = 0$. Vypočítajte jej medián.

Vieme, že kumulatívna distribučná funkcia vyzerá nasledovne

$$F_X(t) = \begin{cases} 0, & \text{ak } t < 0, \\ 1 - e^{-t/2}, & \text{ak } t \geq 0. \end{cases}$$

Preto medián je hodnota M, ktoré rieši nasledovnú rovnicu

$$F_X(M) = 1 - e^{-M/2} = 0.5,$$

preto $M = 2\log(2)$.

Existujú rôzne špecifické typy spojite rozdelených náhodných premenných

- rovnomerné rozdelenie na intervale (a, b),
- normálne rozdelenie,

- exponenciálne rozdelenie,
- t-rozdelenie,
- chí-kvadrát rozdelenie,

a mnoho mnoho iných.

6.3 Rovnomerné rozdelenie

Hovoríme, že spojitá náhodná premenná X má **rovnomerné rozdelenie** na intervale $[a,b]$, ak má nasledovnú funkciu hustoty:

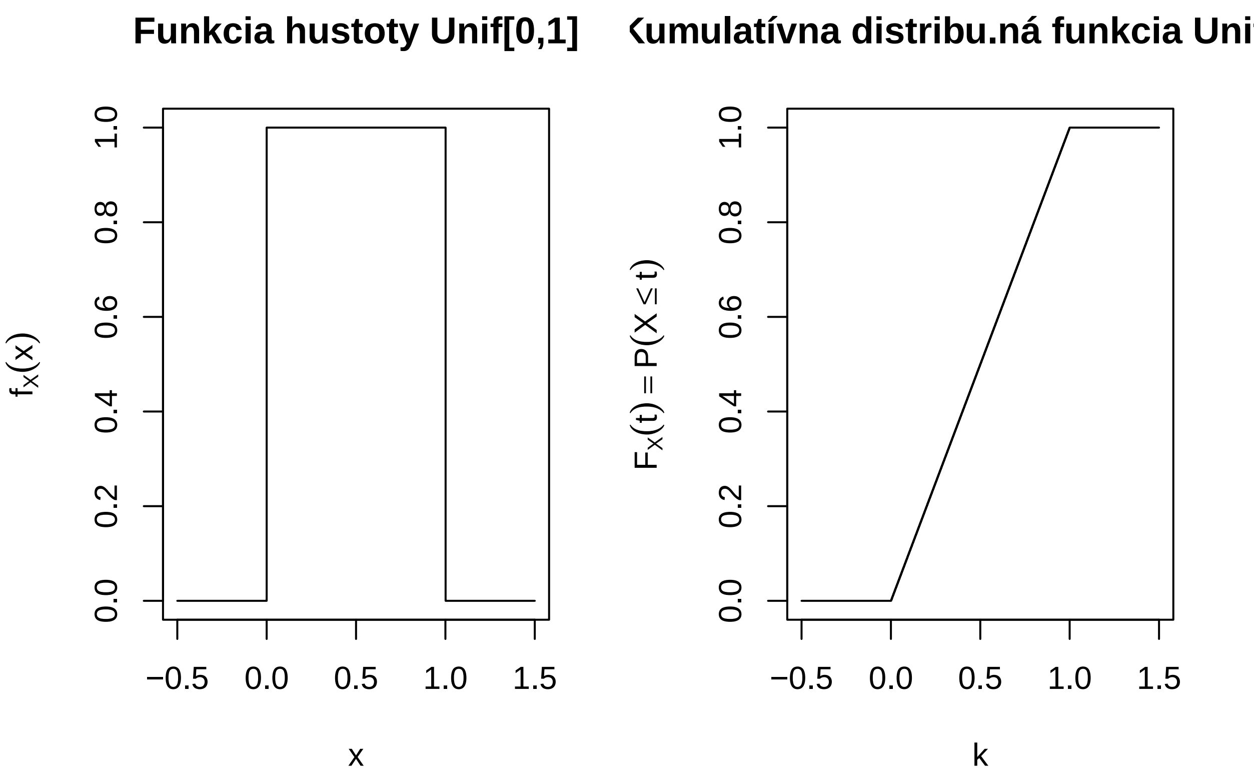
$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{ak } x \in [a,b], \\ 0. & \text{inak} \end{cases}$$

Tomu zodpovedá kumulatívna distribučná funkcia:

$$F_X(t) = \begin{cases} 0, & \text{ak } t < a, \\ \frac{t-a}{b-a}, & \text{ak } t \in [a,b], \\ 1, & \text{ak } t > b. \end{cases}$$

Takúto náhodnú premennú označujeme ako $X \sim \text{Unif}[a,b]$,

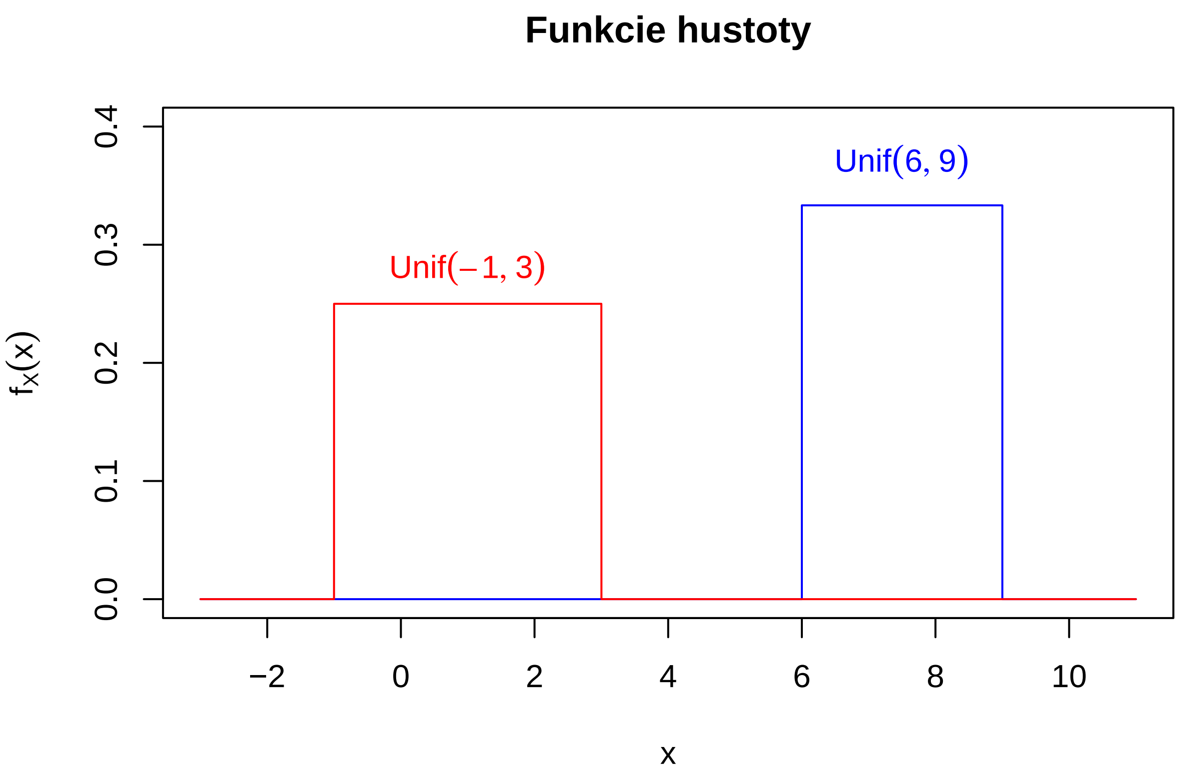
Toto rozdelenie je vhodné, keď vyberáme rovnomerne náhodne číslo z intervalu. *Rovnomerne* znamená, že pravdepodobnosť, že náhodná premenná nadobudne hodnotu v dvoch intervaloch rovnakej dĺžky, je rovnaká. Teda nijakým spôsobom neuprednostňujeme žiadne hodnoty pred inými.



Stredná hodnota a variancia pre takúto náhodnú premennú sú

$$\begin{aligned} \mathrm{E}[X] &= \int_{-\infty}^{\infty} x f_X(x) dx = \int_a^b x \frac{1}{b-a} dx \\ &= \left[\frac{x^2}{2(b-a)}\right]_a^b = \frac{b+a}{2}, \\ \mathrm{Var}[X] &= \dots = \frac{1}{12}(b-a)^2. \end{aligned}$$

Nasledujúci obrázok porovnáva rovnomerne rozdelené náhodné premenné pre rôzne intervaly.



6.4 Normálne rozdelenie

Hovoríme, že spojitá náhodná premenná X má **normálne rozdelenie** s parametrami μ a σ^2, ak má nasledovnú funkciu hustoty:

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

Tomu zodpovedá kumulatívna distribučná funkcia:

$$F_X(t) = \int_{-\infty}^{t} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx.$$

Takúto náhodnú premennú označujeme ako $X \sim \mathrm{N}(\mu, \sigma^2)$.

Kumulatívna distribučná funkcia normálneho rozdelnia nemá pekné analytické riešenie. Pre špeciálny prípad, keď $\mu = 0$ a $\sigma = 1$, teda pre $\mathrm{N}(0, 1)$, rezervujeme pre funkciu hustoty aj pre kumulatívnu distribučnú funkciu špeciálne symboly:

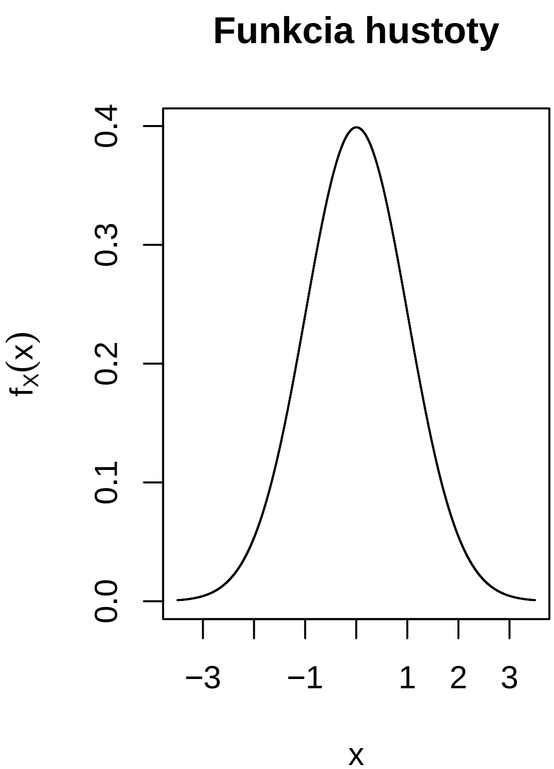
$$\phi(x) \equiv \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

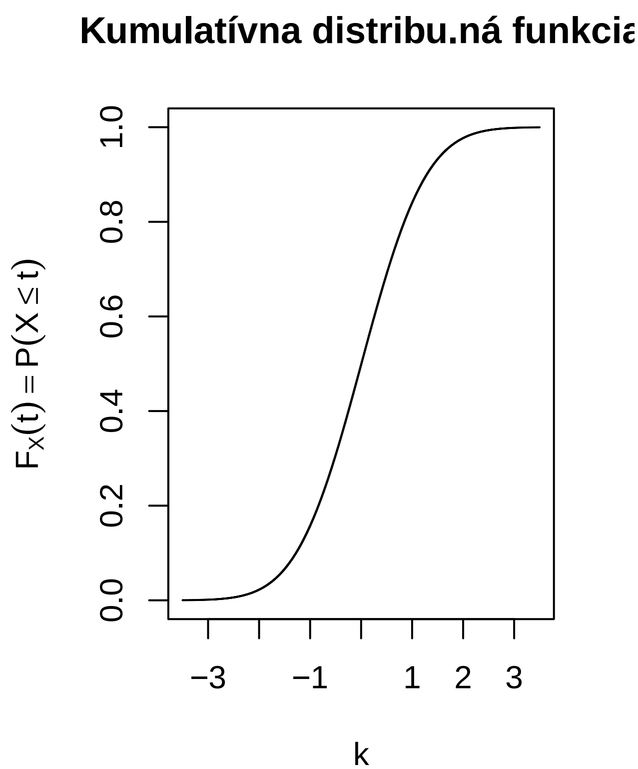
$$\Phi(t) \equiv \int_{-\infty}^{t} \phi(x) dx = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \int_{-\infty}^{t} \frac{1}{1\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-0)^2}{1^2}} dx.$$

Rozdelenie $\mathrm{N}(0,1)$ nazývame **normované normálne rozdelenie** alebo aj **štandardizované normálne rozdelenie**.

Normálne (Gaussovské) rozdelenie je jedno z najdôležitejších pravdepodobnostných rozdelení v štatistike. Tvar jeho hustoty pravdepodobnosti je známy Gaussovský zvon. Jeho dôležitosť plynie najmä z toho, že je to limitné rozdelenie pre aritmetický priemer pre nezávislé a rovnako rozdelené pozorovania zo (skoro) *hocijakej* distribúcie. O tomto výsledku sa budeme rozprávať ešte veľa, nazýva sa Centrálna limitná veta, ale o tomto až neskôr. Prekvapivo rôzne veci v prírode alebo v spoločnosti sa správajú ako normálne rozdelenie, alebo aspoň podobne ako normálne rozdelenie.

- IQ v populácii,
- Krvný tlak v populácii,
- Rôzne biologické merania - dĺžky, povrchy, objemy,
- Chyby merania pri fyzikálnych experimentoch,
- Výsledky maturitných testov,
- Nárast hodnoty akcií na burze cenných papierov,
- Ak zahrievate tyč v jednom bode, tak teplota v nejakom čase bude vyzerať ako toto rozdelenie. Toto je riešenie diferenciálnej rovnice, ktorá sa nazýva *rovnica vedenia tepla*,
- čokoľvek sa správa ako *difúzia*, napríklad pohyb plynu v priestore - ako ďaleko môže antilopa zacítiť šelmu (viacrozmerné normálne rozdelenie).

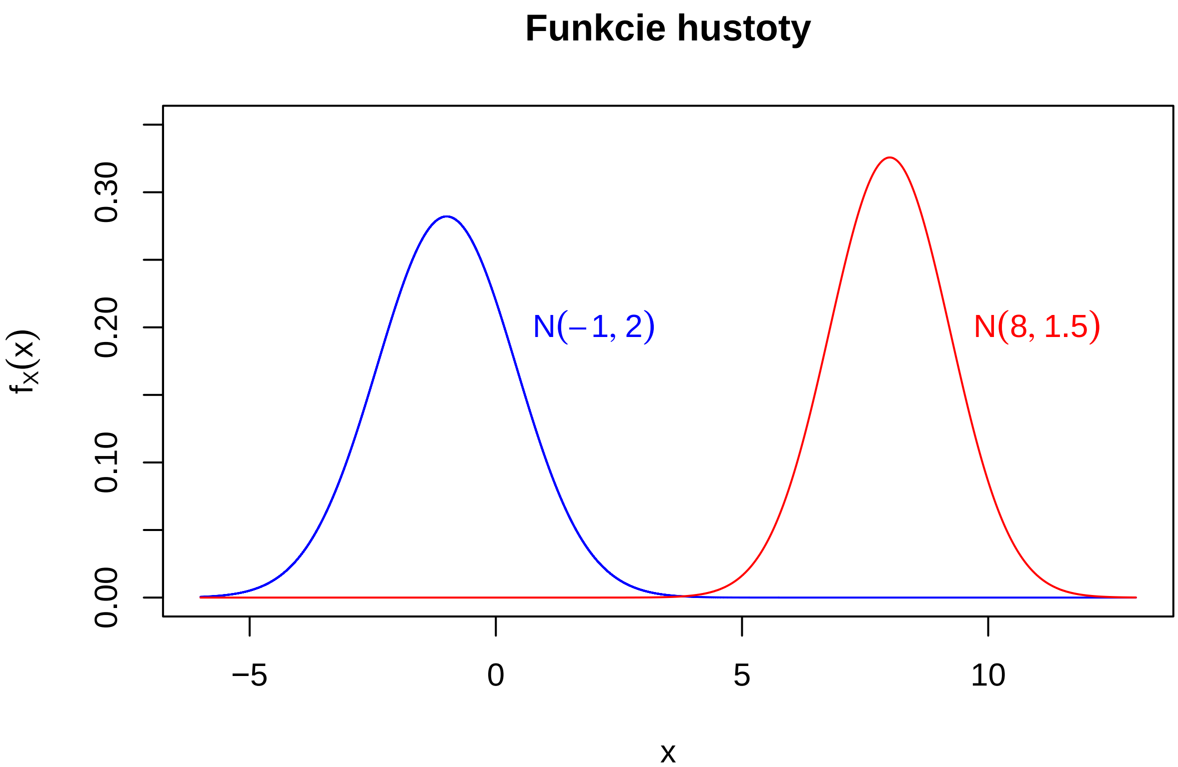




Stredná hodnota a variancia pre takúto náhodnú premennú sú

$$\begin{aligned} \mathrm{E}[X] &= \cdots = \mu, \\ \mathrm{Var}[X] &= \cdots = \sigma^2. \end{aligned}$$

Nasledujúci obrázok porovnáva Normálne rozdelené náhodné premenné pre rôzne parametre.



Normálne rozdelenie má rôzne vlastnosti, vďaka ktorým sa s ním dobre pracuje, napríklad:

- Ak je $X \sim \mathrm{N}(\mu, \sigma^2)$, potom platí, že $aX + b \sim N(a\mu + b, a^2\sigma^2)$. Teda lineárna transformácia normálne rozdelenej náhodnej premennej je stále normálne rozdelená,
- Ak $X_1 \sim \mathrm{N}(\mu_1, \sigma_1^2)$ a $X_2 \sim \mathrm{N}(\mu_2, \sigma_2^2)$ a sú nezávislé, potom je $aX_1 + bX_2 \sim \mathrm{N}(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$.

6.5 Exponenciálne rozdelenie

Hovoríme, že spojitá náhodná premenná X má **exponenciálne rozdelenie** s parametrom λ, ak má nasledovnú funkciu hustoty:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{ak } x > 0, \\ 0. & \text{inak} \end{cases}$$

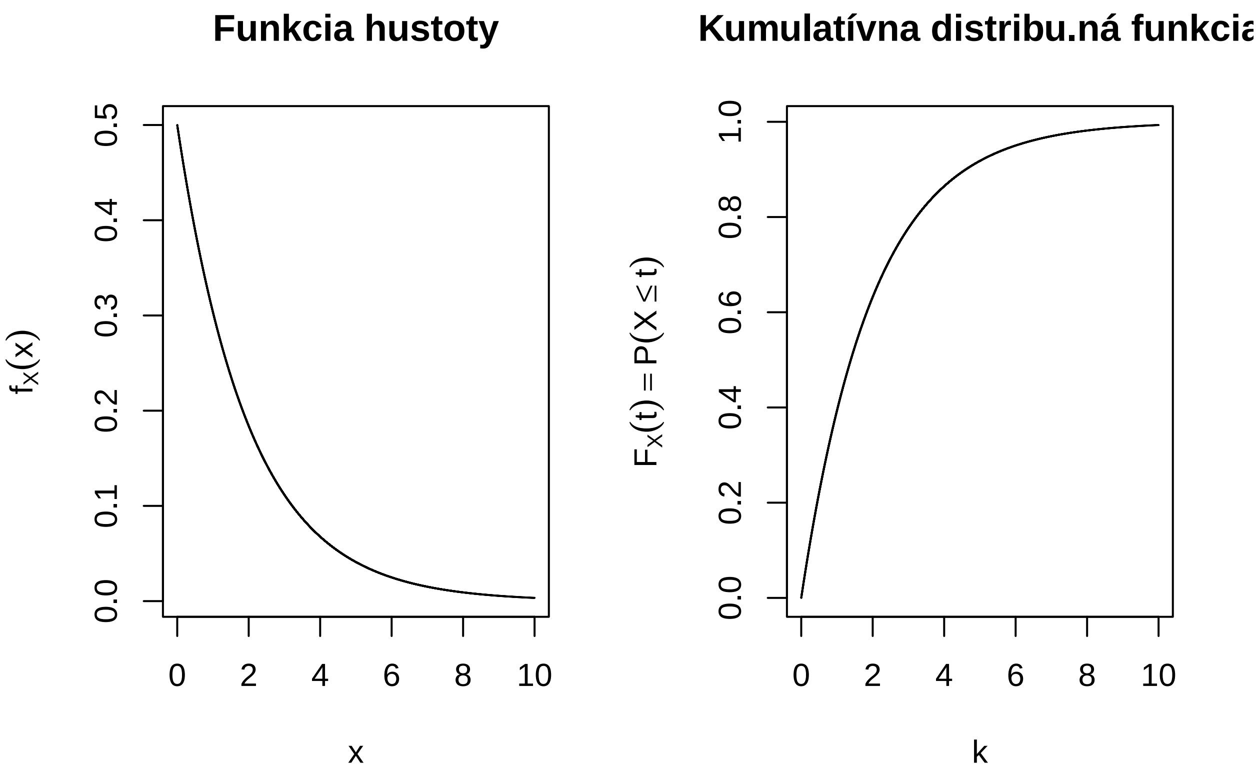
Tomu zodpovedá nasledovná kumulatívna distribučná funkcia:

$$F_X(t) = \begin{cases} 0, & \text{ak } t < 0, \\ 1 - e^{-\lambda t}, & \text{ak } t \geq 0, \end{cases}$$

Takúto náhodnú premennú označujeme ako $X \sim \mathrm{Exp}(\lambda)$, a má jeden parameter, λ.

Toto rozdelenie modeluje napríklad

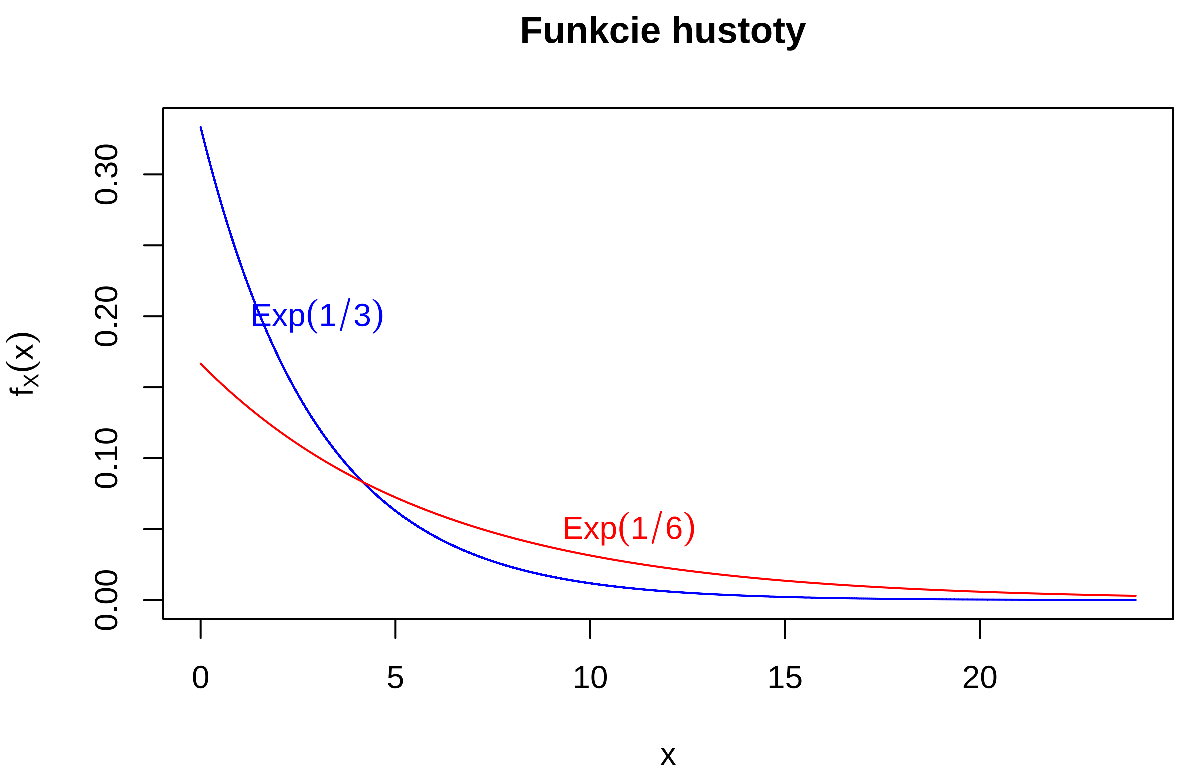
- čas dokým nenastane nejaká situácia/udalosť (príde zákazník, zaplava, zemetrasenie),
- čas rozpadu rádioaktívnych častíc,
- doba obslúženia zákazníka,
- čas medzi udalosťami, ktorých počet je modelovaný Poissonovým rozdelením.



Stredná hodnota a variancia pre takúto náhodnú premennú sú

$$\begin{aligned}
\mathrm{E}[X] &= \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^{\infty} x\lambda e^{-\lambda x} dx \\
&= \left[x.\left(-e^{-\lambda x}\right)\right]_0^{\infty} - \int_0^{\infty} 1.\left(-e^{-\lambda x}\right) dx \\
&= 0 - 0 - \left[\frac{1}{\lambda} e^{-\lambda x}\right]_0^{\infty} = 0 - \left(-\frac{1}{\lambda}\right) = \frac{1}{\lambda}, \\
\mathrm{Var}[X] &= \cdots = \frac{1}{\lambda^2}.
\end{aligned}$$

Tu je porovnanie exponenciálne rozdelených náhodných premenných pre rôzne parametre $\lambda = 3$ (modrou farbou) a $\lambda = 6$ (červenou farbou).



Exponenciálne rozdelenie má jednu dôležitú vlastnosť a to, že si *nepamätá*. Majme žiarovku, ktorej životnosť (označená ako W) je modelovaná exponenciálnym rozdelením s parametrom λ. Napríklad ak priemerná životnosť je 1000 hodín, tak $\lambda = 1/1000$. Vieme, že žiarovka už funguje 500 hodín (nech je toto a). Aké je pravdepodobnostné rozdelenie jej ďalšej životnosti?

$$\begin{aligned} P(W > t + a | W > a) &= \frac{P(W > t + a \cap W > a)}{P(W > a)} = \frac{P(W > t + a)}{P(W > a)} \\ &= \frac{e^{-\lambda(t+a)}}{e^{-\lambda a}} \\ &= e^{-\lambda t} = P(W > t). \end{aligned}$$

Používaná žiarovka je rovnako dobrá ako nová (!). A je úplne jedno, ako dlho už funguje. Exponenciálne rozdelenie preto dobre modeluje čas medzi udalosťami, ktoré nastanú *zrazu* , alebo životnosť objektov, ktoré *nestarnú*. Dá sa dokonca ukázať, že toto je jediné rozdelenie, ktoré má takúto vlastnosť. Pre modelovanie životnosti vecí, ktorých šanca zlyhania v čase rastie alebo klesá, sa používa nejaké bohatšie rozdelenie, napríklad Weibullovo.

6.6 Chí-kvadrát rozdelenie

Hovoríme, že spojitá náhodná premenná X má **chí-kvadrát rozdelenie** s k stupňami voľnosti, ak má nasledovnú funkciu hustoty:

$$f_X(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}.$$

Distribučná funkcia, podobne ako pre normálne rozdelenie, nemá peknú analytickú formulu. Pri práci s ňou sa spoliehame na softvér.

Takúto náhodnú premennú označujeme ako $X \sim \chi^2_n$ a toto rozdelenie popisuje pravdepodobnostné správanie

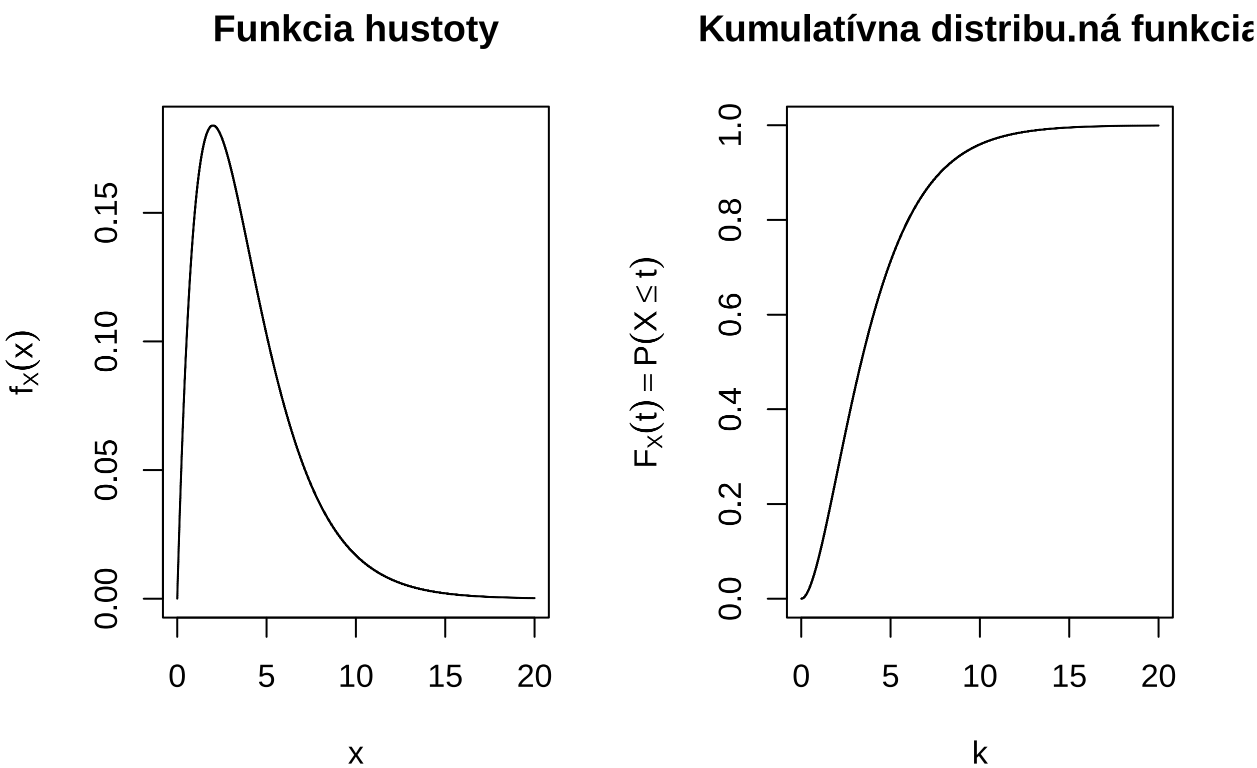
$$X_1^2 + X_2^2 + \cdots + X_k^2$$

kde $X_i \sim \mathrm{N}(0,1)$ a zároveň sú $\{X_i\}_{i=1}^k$ nezávislé náhodné premenné.

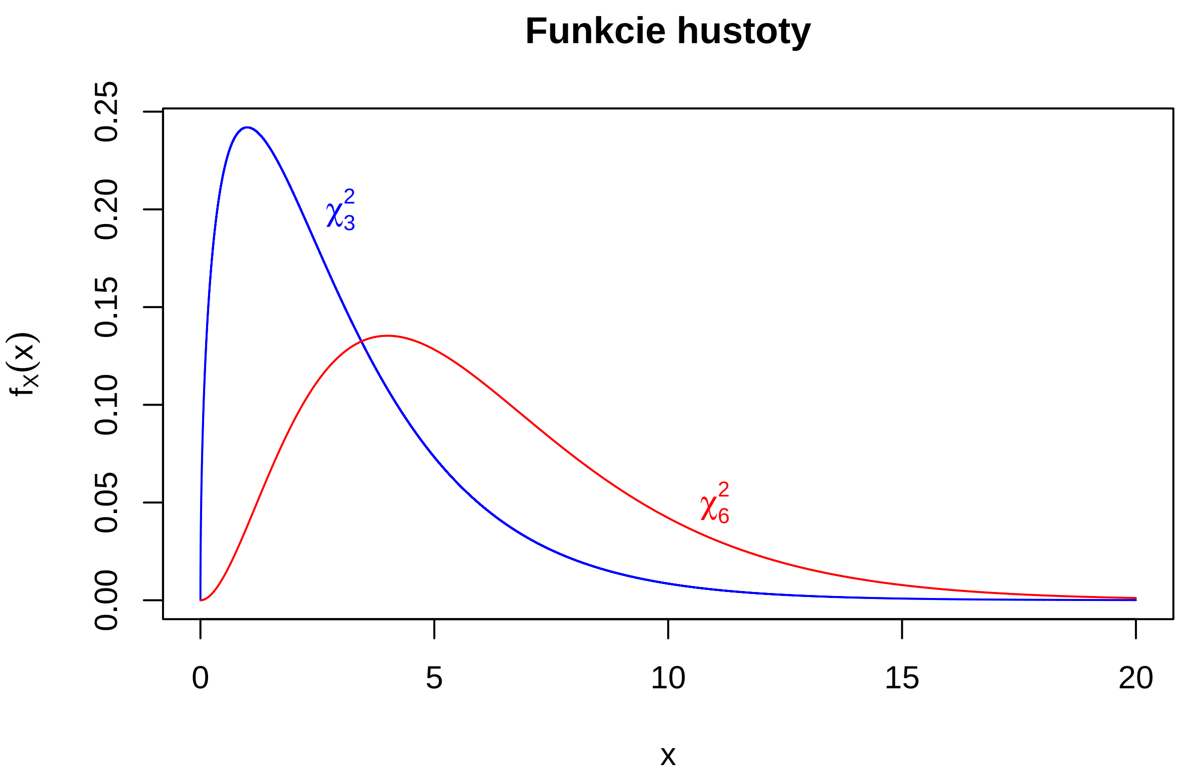
Rozdelenie χ_k^2 je užitočné pri štatistickom testovaní hypotéz. Konkrétne odhad variancie bude viesť na štatistiku, ktorá bude po vhodnej transformácii rozdelená ako χ_k^2.

Stredná hodnota a variancia pre takúto náhodnú premennú sú

$$\begin{aligned}
\mathrm{E}[X] &= \mathrm{E}[X_1^2] + \mathrm{E}[X_2^2] + \cdots + \mathrm{E}[X_k^2] = 1 + 1 + \cdots + 1 = k, \\
\mathrm{Var}[X] &= \cdots = 2k.
\end{aligned}$$



Tu je porovnanie χ^2 rozdelení náhodných premenných pre rôzne parametre $k = 3, 6$:



6.7 Studentovo rozdelenie (t-rozdelenie)

Hovoríme, že spojitá náhodná premenná X má **Studentovo rozdelenie** s k stupňami voľnosti, ak má nasledovnú funkciu hustoty:

$$f_X(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\,\Gamma\left(\frac{k}{2}\right)} \left(1 + \frac{x^2}{k}\right)^{\frac{k+1}{2}}.$$

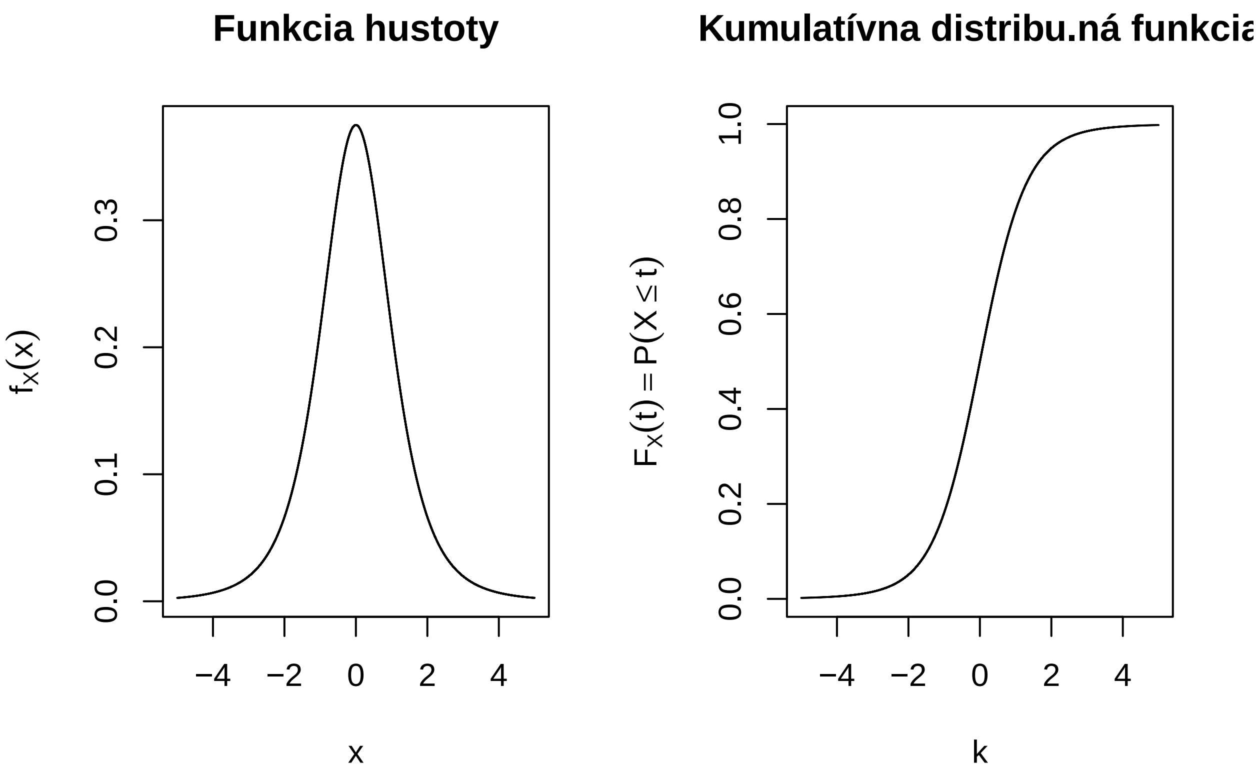
Predpis kumulatívnej distribučnej funkcie je tiež komplikovaný a nie je užitočné si ho pamätať.

Takúto náhodnú premennú označujeme ako $X \sim T_k$.

Studentove t-rozdelenie bude užitočné pri testovaní štatistických hypotéz, napríklad keď chceme porovnať, či majú dva súbory pozorovaní z nejakých náhodných premenných rovnakú strednú hodnotu. Uvažujme $Z \sim N(0,1)$ a $V \sim \chi^2_k$, ktoré sú nezávislé. Potom T

$$T = \frac{Z}{\sqrt{\frac{V}{k}}}$$

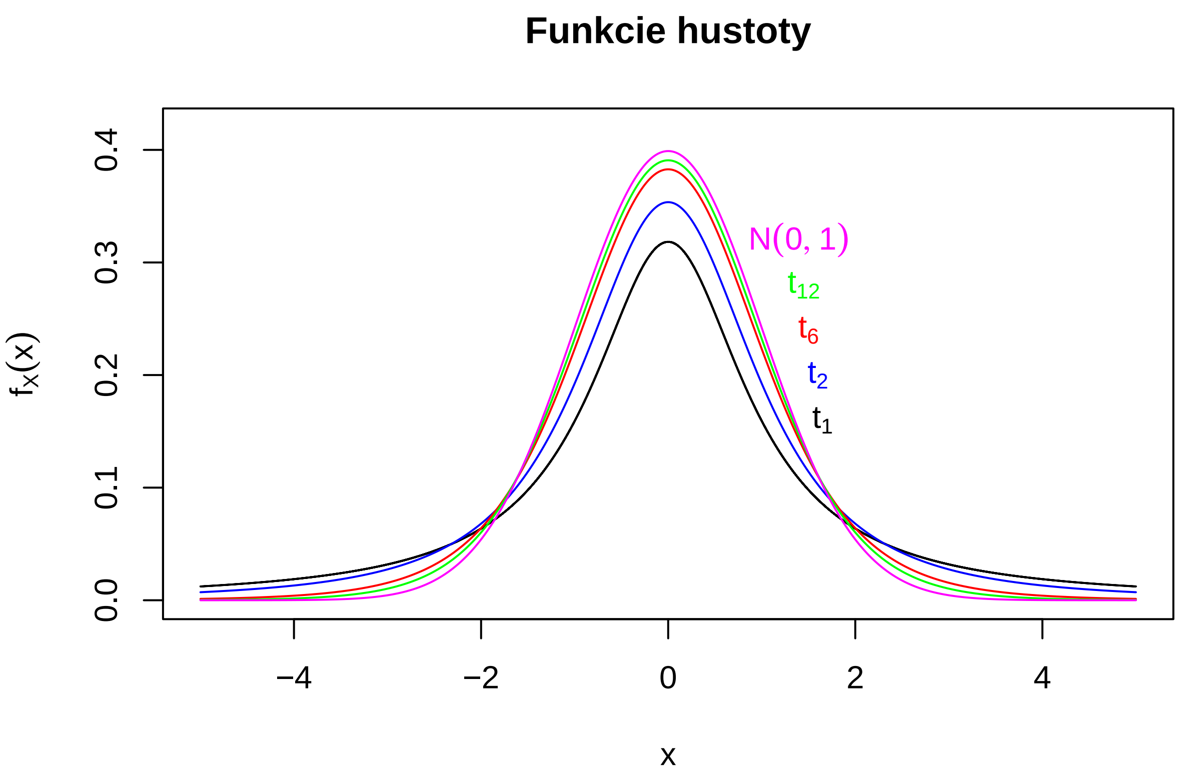
má Studentovo t-rozdelenie s k stupňami voľnosti.



Stredná hodnota (pre $k > 1$) a variancia (pre $k > 2$) pre takúto náhodnú premennú sú

$$\begin{aligned} \mathrm{E}[X] &= 0, \\ \mathrm{Var}[X] &= \cdots = \frac{k}{k-2}. \end{aligned}$$

Nasledujúci obrázok porovnáva t-rozdelené náhodné premenné pre rôzne parametre $k = 1, 2, 6, 12$ a limitné rozdelenie pre $k \to \infty$, teda $\mathrm{N}(0, 1)$.



Rozdelenie t_1 má špeciálne meno. Nazýva sa **Cauchyho rozdelenie** a je zaujímavé tým, že $\mathrm{E}[X]$ neexistuje, lebo nie je splnená podmienka $\int_{-\infty}^{\infty} |x| f_X(x) dx < \infty$. O takomto rozdelení hovoríme, že má **ťažké chvosty**, lebo pravdepodobnosť veľkej alebo malej hodnoty x síce ide k nule, ale príliš pomaly na to, aby bol integrál hodnôt váhovaných pravdepodobnosťami konečný.

6.8 Cvičenia

Cvičenie 6.1. Majme náhodnú premennú s nasledovnou kumulatívnou distribučnou funkciou:

$$F_X(x) = \begin{cases} 0, & \text{ak } x \leq 0, \\ \frac{x^2}{9}, & \text{ak } x \in (0, 3], \\ 1, & \text{ak } x > 3. \end{cases}$$

Nájdite a načrtnite jej funkciu hustoty pravdepodobnosti.

Vypočítajte jej strednú hodnotu a medián.

Cvičenie 6.2. Majme náhodnú premennú s nasledovnou kumulatívnou distribučnou funkciou:

$$F_X(x) = \begin{cases} 0, & \text{ak } x \leq 0, \\ \sin(2x), & \text{ak } x \in \left(0, \frac{\pi}{4}\right), \\ 1, & \text{ak } x > \frac{\pi}{4}. \end{cases}$$

Nájdite a načrtnite jej funkciu hustoty pravdepodobnosti.

Vypočítajte jej strednú hodnotu a medián.

Cvičenie 6.3. Majme náhodnú premennú s nasledovnou funkciu hustoty pravdepodobnosti:

$$f_X(x) = \begin{cases} 2x, & \text{ak } x \in [0,1], \\ 0, & \text{inak.} \end{cases}$$

Skontrolujte, že táto f_X je korektná funkcia hustoty.

Nájdite a načrtnite jej kumulatívnu distribučnú funkciu.

Cvičenie 6.4. Majme náhodnú premennú s nasledovnou funkciu hustoty pravdepodobnosti:

$$f_X(x) = \begin{cases} \frac{1}{3}, & \text{ak } x \in [0,1], \\ \frac{2}{3}, & \text{ak } x \in (1,2], \\ 0, & \text{inak.} \end{cases}$$

Skontrolujte, že táto f_X je korektná funkcia hustoty.

Nájdite a načrtnite jej kumulatívnu distribučnú funkciu.

Vypočítajte jej strednú hodnotu, varianciu a medián.

Cvičenie 6.5. Majme náhodnú premennú s nasledovnou funkciu hustoty pravdepodobnosti:

$$f_X(x) = \begin{cases} 0, & \text{ak } x \leq 0, \\ a \sin x, & \text{ak } x \in (0,\pi], \\ 0, & \text{ak } x > \pi. \end{cases}$$

Vypočítajte a tak, aby f_X bola korektná funkcia hustoty.

Nájdite a načrtnite jej kumulatívnu distribučnú funkciu.

Vypočítajte jej strednú hodnotu, varianciu a $P(X \geq \pi/4)$

Cvičenie 6.6. Majme náhodnú premennú s nasledovnou funkciu hustoty pravdepodobnosti:

$$f_X(x) = \frac{4c}{e^x + e^{-x}}$$

Vypočítajte c tak, aby f_X bola korektná funkcia hustoty.

Nájdite a načrtnite jej kumulatívnu distribučnú funkciu.

Vypočítajte jej strednú hodnotu, varianciu, medián a $P(X \geq 1)$.

Cvičenie 6.7. Majme náhodnú premennú s nasledovnou funkciu hustoty pravdepodobnosti:

$$f_X(x) = \begin{cases} x - \frac{x^3}{4}, & \text{ak } x \in (0,2), \\ 0, & \text{inak.} \end{cases}$$

Nájdite a načrtnite jej kumulatívnu distribučnú funkciu.

Vypočítajte jej strednú hodnotu, varianciu, medián a $P(|X-1| \leq 0.5)$

Cvičenie 6.8. Nech $X \sim \text{Unif}[4,6]$

Nájdite kumulatívnu distribučnú funkciu náhodnej premennej X.

Nájdite kumulatívnu distribučnú funkciu náhodnej premennej X^2.

Cvičenie 6.9. Využitím linearity strednej hodnoty ukážte, že platí

$$\mathrm{Var}[aX+b] = a^2\mathrm{Var}[X].$$

Cvičenie 6.10. Majme náhodnú premennú s nasledovnou funkciu hustoty pravdepodobnosti:

$$f_X(x) = \begin{cases} c.e^{-x}, & \text{ak } x > 0, \\ 0, & \text{inak.} \end{cases}$$

Určite konštantu c.

Cvičenie 6.11. Ukážte, že neexistuje žiadna náhodná premenná, pre ktorú platí $\mathrm{E}[X] = 4$ a súčasne $\mathrm{E}[X^2] = 15$.

Cvičenie 6.12. Nech $X \sim \mathrm{N}(0,1)$ a $Y \sim \mathrm{N}(1,1)$

Ukážte, že $P(X < 3) > P(Y < 3)$.

Cvičenie 6.13. Ukážte, že nasledujúce funkcie hustoty sú korektné

- (Weibullovo rozdelenie)

$$f_X(x) = \begin{cases} \alpha x^{\alpha-1}e^{-x^\alpha}, & \text{ak } x > 0, \\ 0, & \text{inak} \end{cases}$$

pre $\alpha > 0$,

- (Pareto rozdelenie)

$$f_X(x) = \begin{cases} \alpha(1+x)^{-\alpha-1}, & \text{ak } x > 0, \\ 0, & \text{inak} \end{cases}$$

pre $\alpha > 0$,

- (Cauchyho rozdelenie)

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

(pomôcka, pozrite sa do tabuliek derivácie pre arctan.)

- (Laplaceho rozdelenie)

$$f_X(x) = e^{-|x|}/2,$$

- (Extreme value rozdelenie)

$$f_X(x) = e^{-x}\exp\{-e^{-x}\}.$$

Cvičenie 6.14. Majme náhodnú premennú s takouto funkciou hustoty

$$f_X(x) = \begin{cases} cx^2, & \text{ak } |x| < 1, \\ 0, & \text{inak.} \end{cases}$$

- Nájdite konštantu c.
- Vypočítajte $\mathrm{E}[X]$ a $\mathrm{Var}[X]$.

- Vypočítajte $P\left(X \geq \frac{1}{2}\right)$

Cvičenie 6.15. Majme náhodnú premennú s takouto funkciou hustoty

$$f_X(x) = \frac{1}{2}e^{-|x|}.$$

Vypočítajte kumulatívnu distribučnú funkciu pre $Y = X^2$.

Cvičenie 6.16. Majme náhodnú premennú s takouto funkciou hustoty

$$f_X(x) = \begin{cases} 4x^3, & \text{ak } 0 < x < 1, \\ 0, & \text{inak.} \end{cases}$$

- Vypočítajte $P\left(X \geq \frac{2}{3} | X > \frac{1}{3}\right)$.

Cvičenie 6.17. Majme náhodnú premennú s takouto funkciou hustoty

$$f_X(x) = \begin{cases} x^2\left(2x + \frac{3}{2}\right), & \text{ak } 0 < x \leq 1, \\ 0, & \text{inak.} \end{cases}$$

- Vypočítajte $\mathrm{Var}[Y]$ pre $Y = \frac{2}{X} + 3$.

Cvičenie 6.18. Majme kladnú spojitú náhodnú premennú. Ukážte, že platí

$$\mathrm{E}[X] = \int_0^\infty P(X \geq x)dx.$$

Cvičenie 6.19. Nech f_{X_1} a f_{X_2} sú hustoty pravdepodobnosti náhodných premenných X_1 a X_2. Rozhodnite, či môže platiť $f_{X_1}(x) > f_{X_2}(x)$ pre všetky $x \in \mathbb{R}$.

Cvičenie 6.20. Životnosť súčiastky je exponenciálne rozdelená so strednou hodnotou 1 rok. Aká je pravdepodobnosť, že ak funguje už 2 roky, že bude fungovať ďalšie 3 roky?

Cvičenie 6.21. Na náš web príde človek približne každých 10min. Aká je šanca, ze počas polhodinového výpadku nikto nechcel navštíviť náš web? (Použite exponenciálne rozdelenie.)

7 Súvis medzi náhodnými premennými

Potrebujeme matematický aparát na to, aby sme vedeli pracovať s viacerými náhodnými premennými naraz. Potrebujeme vedieť, ako rôzne náhodné premenné spolu súvisia. Začneme s dvomi náhodnými premennými X a Y. Dve náhodné premenné dokopy tvoria dvojrozmerný **náhodný vektor** $(X, Y)^T$. Skúmať súvis medzi náhodnými premennými sa preto nedá bez toho, že by sme vedeli ako udalosti nastávajú *spolu*.

Združená kumulatívna distribučná funkcia $F_{XY} : \mathbb{R}^2 \to [0, 1]$ náhodného vektora $(X, Y)^T$ je definovaná nasledovne:

$$F_{XY}(s, t) = P(X \leq s \cap Y \leq t) = P(X \leq s, Y \leq t).$$

Ide len o dvojrozmerný ekvivalent toho, čo sme už videli predtým.

Diskrétne náhodné premenné

Ak sú elementy náhodného vektora diskrétne rozdelené náhodné premenné, potom alternatívne môžeme na popis náhodnosti použiť **združenú pravdepodobnostnú funkciu**

$$p_{XY}(x, y) = P(X = x \cap Y = y) = P(X = x, Y = y).$$

Aby táto bola korektná, tak musí platiť

- $\forall x \in \mathcal{S}_X, y \in \mathcal{S}_Y : p_{XY}(x,y) \geq 0$,
- $\sum_{x \in \mathcal{S}_X} \sum_{y \in \mathcal{S}_Y} p_{XY}(x,y) = 1$,
- $\forall x \in \mathcal{S}_X : \sum_{y \in \mathcal{S}_Y} p_{XY}(x,y) = p_X(x)$,
- $\forall y \in \mathcal{S}_Y : \sum_{x \in \mathcal{S}_X} p_{XY}(x,y) = p_Y(y)$.

Spojité náhodné premenné

Ak sú elementy náhodného vektora spojite rozdelené, tak potrebujeme **združenú funkciu hustoty** $f_{XY}(x,y) : \mathbb{R}^2 \to \mathbb{R}^+$, ktorá spĺňa nasledovnú vlastnosť

$$P\left(X \in [a,b] \cap Y \in [c,d]\right) = P\left(X \in [a,b], Y \in [c,d]\right) = \int_a^b \int_c^d f_{XY}(x,y) dy dx.$$

Funkcia hustoty musí spĺňať nasledovné vlastnosti

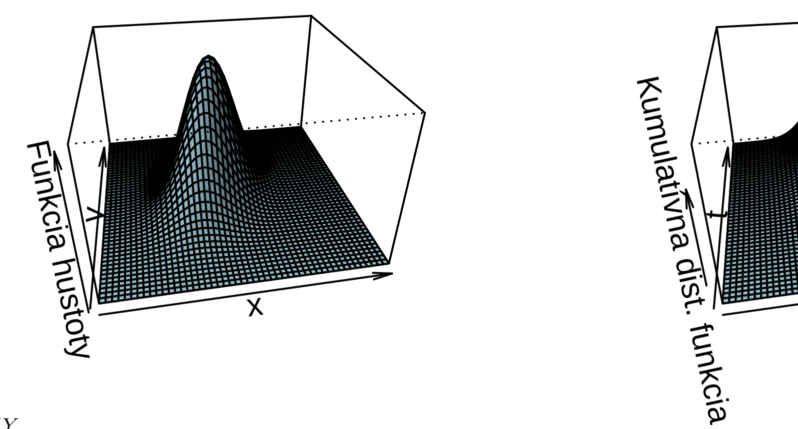
- $\forall x \in \mathbb{R}, y \in \mathbb{R} : f_{XY}(x,y) \geq 0$,
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x,y) dx dy = 1$,
- $\forall x \in \mathbb{R} : \int_{-\infty}^{\infty} f_{XY}(x,y) dy = f_X(x)$,
- $\forall y \in \mathbb{R} : \int_{-\infty}^{\infty} f_{XY}(x,y) dx = f_Y(y)$.

Vzťah medzi f_{XY} a F_{XY} je nasledovný:

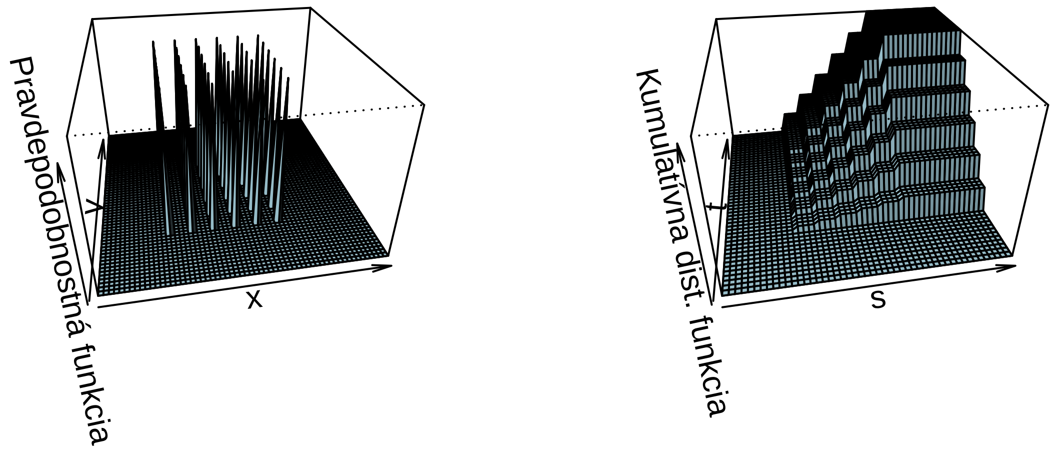
$$F_{XY}(s,t) = \int_{-\infty}^{s} \int_{-\infty}^{t} f_{XY}(x,y) dy dx$$

a

$$f_{XY}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x,y).$$



Tu je vizualizovanú jeden konkrétny príklad pre f_{XY} a F_{XY}
Tu je vizualizovaná p_{XY} a F_{XY} pre diskrétne rozdelený náhodný vektor (X,Y), kde X a Y sú čísla, ktoré padnú



na dvoch nezávislých kockách.

7.1 Nezávislé náhodné premenné

Hovoríme, že dve náhodné premenné X a Y sú nezávislé ak

$$\forall x \in \mathcal{S}_X, y \in \mathcal{S}_Y : p_{XY}(x,y) = p_X(x) \cdot p_Y(y),$$

ak sú diskrétne rozdelené a

$$\forall x \in \mathbb{R}, y \in \mathbb{R} : f_{XY}(x,y) = f_X(x) \cdot f_Y(y),$$

ak sú spojite rozdelené. Informácia o pravdepodobnostných správaniach X a Y je preto dostatočná na to, aby sme vedeli, ako sa budú správať spolu.

Alternatívna definícia nezávislosti je

$$\forall B_1, B_2 \subseteq \mathbb{R} : P(X \in B_1, Y \in B_2) = P(X \in B_1) \cdot P(Y \in B_2).$$

[22]

alebo, taktiež ekvivalentne

$$\forall s, t \in \mathbb{R} : F_{XY}(s,t) = F_X(s) \cdot F_Y(t).$$

Poľahky si všimneme, že voľbou $B_1 = (-\infty, s]$ a $B_2 = (-\infty, t]$ dostávame vzťah s predošlou definíciou.

Pre nezávislé náhodné premenné platí, že ak X a Y sú nezávislé potom sú aj $h_1(X)$ a $h_2(Y)$ sú nezávislé náhodné premenné, kde h_1 a h_2 sú nejaké transformácie.[23] Napríklad, X^2 a Y, kde $h_1(x) = x^2$ a $h_2(y) = y$.

Príklad 7.1. Majme nasledovnú združenú hustotu pravdepodobnosti pre náhodný vektor (X,Y).

$$f_{XY}(x,y) = \begin{cases} \frac{1}{4}, & \text{ak } x, y \in [0,2], \\ 0, & \text{inak} \end{cases}$$

Sú X a Y nezávislé náhodné premenné?

[22] Tieto množiny B_1 a B_2 nemôžu byť hocijaké ale musia byť borelovské. O podrobnostiach sa učíme v kurze Teórie pravdepodobnosti.

[23] Tieto transformácie nemôžu byť hocijaké ale musia byť borelovsky merateľné. O podrobnostiach sa učíme, neprekvapivo, v kurze Teórie pravdepodobnosti.

7.2 Miera závislosti

Zatiaľčo p_{XY}, f_{XY}, F_{XY} popisujú súvis náhodných premenných *úplne*, niekedy máme potrebu charakterizovať súvis jediným číslom. Podobne ako sme pomocou strednej hodnoty vyjadrovali centrum distribúcie a pomocou variancie to, ako veľmi sa náhodná premenná menila.

Kovarianciou dvoch náhodných premenných nazývame

$$\mathrm{Cov}[X,Y] \equiv \mathrm{E}[(X-\mathrm{E}[X])(Y-\mathrm{E}[Y])].$$

Pre kovarianciu platí $\mathrm{Cov}[X,Y] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$ a $\mathrm{Cov}[X,X] = \mathrm{Var}[X]$.

Pre nezávislé náhodné premenné platí $\mathrm{E}[XY] = \mathrm{E}[X]\mathrm{E}[Y]$ a preto $\mathrm{Cov}[X,Y] = 0$.

Pre varianciu súčtu dvoch náhodných premenných platí:

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}[X,Y].$$

Vysvetlenie je tu a je založené na linearite strednej hodnoty:

$$\begin{aligned}
\mathrm{Var}[X+Y] &= \mathrm{E}\left[((X+Y)-\mathrm{E}[X+Y])^2\right] \\
&= \mathrm{E}\left[([X-\mathrm{E}[X]) + (Y-\mathrm{E}[Y]])^2\right] \\
&= \mathrm{E}\left[(X-\mathrm{E}[X])^2 + (Y-\mathrm{E}[Y])^2 + 2\,(X-\mathrm{E}[X])(Y-\mathrm{E}[Y])\right] \\
&= \mathrm{E}\left[(X-\mathrm{E}[X])^2\right] + \mathrm{E}\left[(Y-\mathrm{E}[Y])^2\right] + \mathrm{E}\left[2\,(X-\mathrm{E}[X])(Y-\mathrm{E}[Y])\right] \\
&= \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\,\mathrm{Cov}[X,Y]\;.
\end{aligned}$$

Takže pre nezávislé náhodné premenné platí

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\underbrace{\mathrm{Cov}[X,Y]}_{=0} = \mathrm{Var}[X] + \mathrm{Var}[Y].$$

Koreláciou dvoch náhodných premenných nazývame

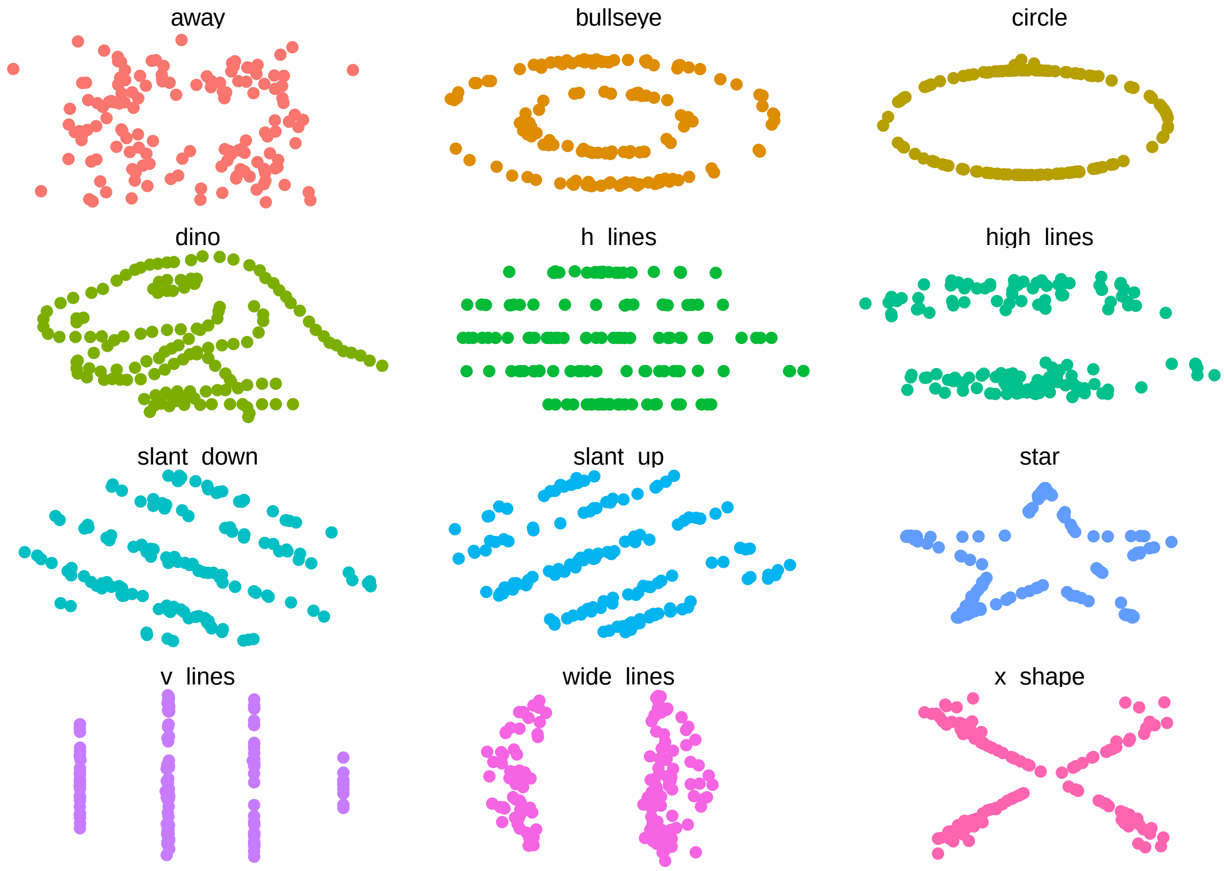
$$\mathrm{Corr}[X,Y] \equiv \frac{\mathrm{Cov}[X,Y]}{\mathrm{sd}[X]\cdot\mathrm{sd}[Y]} = \frac{\mathrm{E}[(X-\mathrm{E}[X])(Y-\mathrm{E}[Y])]}{\sqrt{\mathrm{E}[(X-\mathrm{E}[X])^2]}\cdot\sqrt{\mathrm{E}[(Y-\mathrm{E}[Y])^2]}}.$$

Pre koreláciu platí:

- $-1 \le \mathrm{Corr}[X,Y] \le 1$, je bezrozmerná, t.j. nemá žiadne jednotky,
- $\mathrm{Corr}[X,Y] = \mathrm{Corr}[Y,X]$ takže korelácia je symetrická,
- $\mathrm{Corr}[X,Y] = \pm 1 \implies \exists a,b \in \mathbf{R} : Y = aX + b$, nadobúda hodnoty ± 1 práve vtedy, keď je jedna náhodná premenná lineárnou funkciou druhej,

- X a Y sú nezávislé $\implies \mathrm{E}[XY] = \mathrm{E}[X]\mathrm{E}[Y] \implies \mathrm{Cov}[X,Y] = 0 \implies \mathrm{Corr}[X,Y] = 0$

Skutočnosť, že súvis dvoch náhodných premenných vyjadríme jediným číslom, so sebou nesie aj náklady. Kompaktnejší popis musí nutne nejakú informáciu vynechať, čo môže, ale nemusí byť problematické. Nasledujúci obrázok demonštruje realizácie 12 rôznych dvojích náhodných premenných (X,Y), ktoré majú rovnaké $\mathrm{E}[X], \mathrm{E}[Y], \mathrm{Var}[X], \mathrm{Var}[Y], \mathrm{Corr}[X,Y]$. (Zdroj: https://cran.r-project.org/web/packages/datasauRus/vignettes/Datasaurus.html). V týchto prípadoch sú závislosti medzi týmito premennými veľmi veľmi rôzne. Pozerať sa len na sumárne charakteristiky je preto zavádzajúce.



Korelácia je miera *lineárnej* závislosti. Z toho, že $\text{Corr}[X, Y] = 0$, **nevyplýva**, že X a Y sú nezávislé. A to jednoducho preto, že medzi nimi môže byť aj iná ako lineárna závislosť, ako demonštruje nasledovný príklad.

Príklad 7.2. Majme $X \sim \text{Unif}[-1, 1]$ a $Y = X^2$. Tomuto zodpovedajú nasledovné funkcie hustoty

$$f_X(x) = \begin{cases} \frac{1}{2}, & \text{ak } x \in [-1, 1], \\ 0, & \text{inak} \end{cases}$$

a

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}}, & \text{ak } y \in [0, 1], \\ 0, & \text{inak.} \end{cases}$$

Združená hustota f_{XY} je 0 všade tam, kde $Y > X$, ale $f_X \times f_Y$ tam nie je nutne 0, preto X a Y nie sú nezávislé. Zároveň však platí (ukážte prečo) $\text{E}[X] = 0, \text{E}[Y] = \frac{1}{3}$ a $\text{Cov[X,Y]} = \text{E}[(X - \text{E}[X])(Y - \text{E}[Y])] = \text{E}\left[(X - 0)\left(X^2 - \frac{1}{3}\right)\right] = 0$.

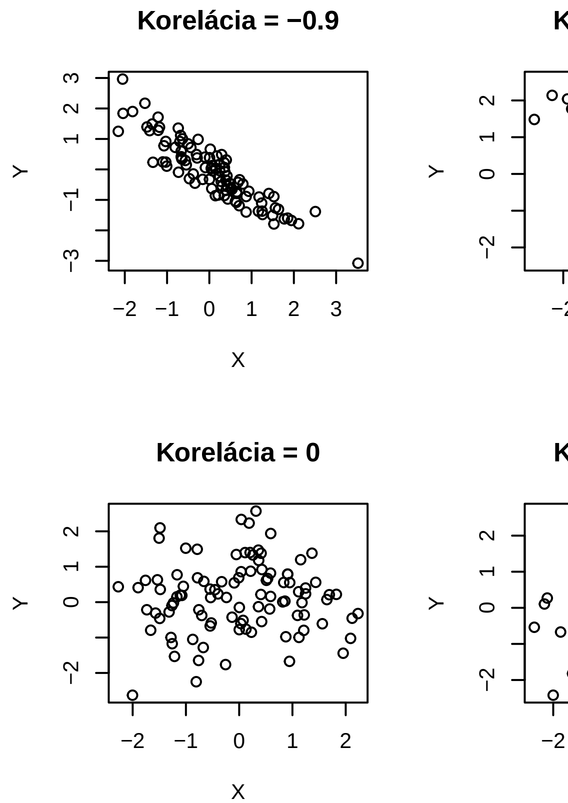
Teda X a Y sú závislé (však Y je priamo funkciou X!), ale nekorelované (teda *lineárne* nezávislé).

Príklad 7.3. Nech $X \sim N(0, 1)$ a nech

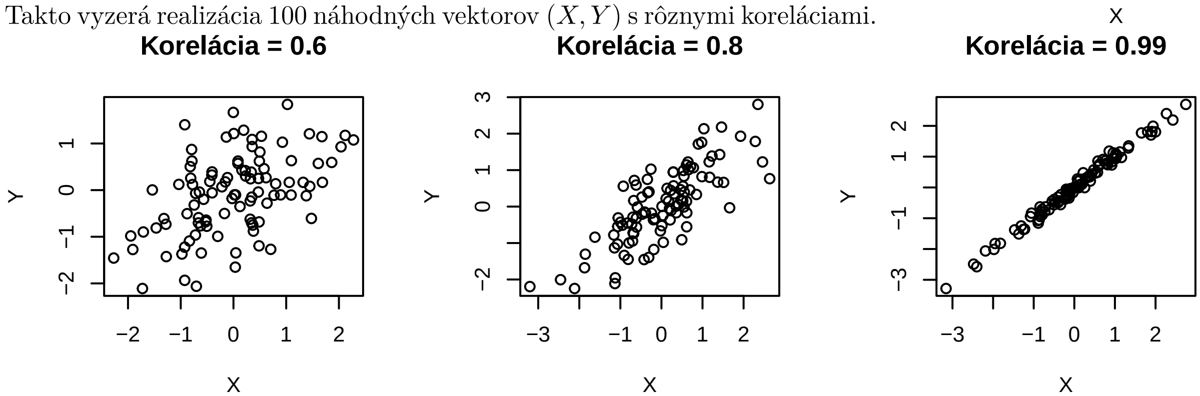
$$Y = \begin{cases} X, & \text{ak } |X| \leq c, \\ -X, & \text{inak,} \end{cases}$$

Pre hodnotu c veľmi malú je $\text{Corr}[X, Y] \approx -1$, naopak, pre c veľmi veľké je to $\text{Corr}[X, Y] \approx 1$. Nakoľko sa táto korelácia spojite mení s c, podľa Vety o strednej hodnote musí existovať hodnota c taká, že $\text{Corr}[X, Y] = 0$.

Na druhej strane, X a Y nemôžu byť nezávislé, nakoľko Y je deterministickou funkciou X.



Takto vyzerá realizácia 100 náhodných vektorov (X, Y) s rôznymi koreláciami.



Korelácia hovorí o asociácii, ale pozor, nie o kauzalite. Skutočnosť, že hodnoty X a Y nejakým spôsobom nastávajú naraz neznamená, že X spôsobuje Y alebo naopak. Napríklad predaje zmrzliny (X) sú korelované s napadnutiami žralakom (Y). Neznamená to ale, že tieto premenné spolu kauzálne súvisia. Ľudia skrátka jedia zmrzlinu ako aj surfujú viacej vtedy, keď je teplo.

Teraz nejaký príklad na počítanie:

Príklad 7.4.

Majme X, Y pre ktoré je p_{XY} vyjadrená nasledovnou tabuľkou.

Vypočítajte $\mathrm{Corr}[X, Y]$.

Pravdepodobnostná funkcia				
	Y=1	Y=2	Y=3	Y=4
X=1	0.1	0	0.1	0
X=2	0.3	0	0.1	0.2
X=3	0	0.2	0	0

$$\begin{aligned}
\mathrm{E}[X] &= 1\cdot(0.1+0+0.1+0)+2\cdot(0.3+0+0.1+0.2)+\\
&\quad 3\cdot(0+0.2+0+0)=2,\\
\mathrm{E}[X^2] &= 1^2\cdot(0.1+0+0.1+0)+2^2\cdot(0.3+0+0.1+0.2)+\\
&\quad 3^2\cdot(0+0.2+0+0)=4.4,\\
\mathrm{Var}[X] &= \mathrm{E}[X^2]-(\mathrm{E}[X])^2=4.4-4=0.4,\\
\mathrm{sd}[X] &= \sqrt{\mathrm{Var}[X]}\approx 0.632,\\
\mathrm{E}[Y] &= 1\cdot(0.1+0.3+0)+2\cdot(0+0+0.2)+\\
&\quad 3\cdot(0.1+0.1+0)+4\cdot(0+0.2+0)=2.2,\\
\mathrm{E}[Y^2] &= 1^2\cdot(0.1+0.3+0)+2^2\cdot(0+0+0.2)+\\
&\quad 3^2\cdot(0.1+0.1+0)+4^2\cdot(0+0.2+0)=6.2,\\
\mathrm{Var}[Y] &= \mathrm{E}[Y^2]-(\mathrm{E}[Y])^2=6.2-4.84=1.36,\\
\mathrm{sd}[Y] &= \sqrt{\mathrm{Var}[Y]}\approx 1.166,\\
\mathrm{E}[XY] &= 1\cdot1\cdot0.1+1\cdot2\cdot0+1\cdot3\cdot0.1+1\cdot4\cdot0+\\
&\quad 2\cdot1\cdot0.3+2\cdot2\cdot0+2\cdot3\cdot0.1+2\cdot4\cdot0.2+\\
&\quad 3\cdot1\cdot0+3\cdot2\cdot0.2+3\cdot3\cdot0+3\cdot4\cdot0=4.4,\\
\mathrm{Cov}[X,Y] &= \mathrm{E}[XY]-\mathrm{E}[X]\mathrm{E}[Y]=4.4-2\cdot2.2=0,\\
\mathrm{Corr}[X,Y] &= 0.
\end{aligned}$$

7.3 Cvičenia

Cvičenie 7.1. Zostrojte združenú pravdepodobnostnú funkciu pre X, Y tak, aby súčasne platilo:

- $\mathrm{E}[X] = 2$,
- $\mathrm{E}[Y] = 2$,
- $\mathrm{Cov}[X, Y] = 0$,
- $P(X \geq Y) = 0.5$.

Ak sa to nedá dokážte prečo.

Cvičenie 7.2. Majme náhodné premenné X, Y s nasledovnou združenou funkciou hustoty

$$f_{XY}(x,y) = \begin{cases} cy^2, & \text{ak } x \in [0,2],\ y \in [0,1], \\ 0, & \text{inak.} \end{cases}$$

Vypočítajte

(a) hodnotu konštanty c,
(b) $P(X \leq 1)$,
(c) $P(X + Y > 2)$,
(d) $P(X > Y)$,
(e) $P(X = 3Y)$,

(f) $\mathrm{E}[Y]$,
(g) $\mathrm{Cov}[X, Y]$.

Cvičenie 7.3. Majme náhodné premenné X, Y s nasledovnou združenou funkciou hustoty

$$f_{XY}(x, y) = \begin{cases} c(x^2 + y), & \text{ak } 0 \leq y \leq 1 - x^2, \\ 0, & \text{inak.} \end{cases}$$

Vypočítajte

(a) hodnotu konštanty c,
(b) $P(0 \leq X \leq \frac{1}{2})$,
(c) $P(Y \leq X + 1)$.

Cvičenie 7.4. Majme náhodné premenné X, Y s nasledovnou združenou funkciou hustoty

$$f_{XY}(x, y) = \begin{cases} \frac{15}{4} x^2, & \text{ak } 0 \leq y \leq 1 - x^2, \\ 0, & \text{inak.} \end{cases}$$

(a) Určte hustotu náhodnej premennej X.
(b) Určte hustotu náhodnej premennej Y.
(c) Zistite, či X a Y sú nezávislé.

Cvičenie 7.5. Nech $X \sim N(0, 1)$ a nech W má nasledovnú pravdepodobnostnú funkciu

$$p_W(w) = P(W = w) = \begin{cases} \frac{1}{2}, & \text{ak } w = 1, \\ \frac{1}{2}, & \text{ak } w = -1, \\ 0, & \text{inak.} \end{cases}$$

Nech naviac X a W sú nezávislé náhodné premenné. Zadefinujme $Y = XW$. Ukážte, že $\mathrm{Cov}[X, Y] = 0$ a že X a Y nie sú nezávislé.

8 Zákon veľkých čísel

Zákon veľkých čisel je jedným z najdôležitejších výsledkov v pravdepodobnosti a štatistike. Hovorí o tom, že aritmetický priemer vypočítaný zo stále väčšieho a väčšieho počtu nezávislých a rovnako rozdelených náhodných premenných sa blíži k skutočnej strednej hodnote.[24] Zároveň je to aj výsledok, ktorý sme potichu používali len sme o tom nehovorili explicitne. Vždy, keď používame počítačovú simuláciu na ukázanie akýchsi vlastností náhodnej premennej, využívame fakt, že pri dostatočne veľkom množstve simulácií sú tieto odsimulované vlastnosti výpovedné o skutočných vlastnostiach.

Príklad 8.1. Majme mincu o ktorej nevieme, či je férová alebo nie. Chceli by sme vedieť aká je pravdepodobnosť toho, že padne hlava ($X = 1$) alebo znak ($X = 0$). Teda uvažujeme $X \sim \mathrm{Bern}(p)$. Hádžeme mincou, napríklad 8 krát a dostávame realizáciu siedmych náhodných premenných $X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8$. Hádžeme tak, že výsledok každého hodu nesúvisí s tým predošlým, takže $\{X_i\}_{i=1}^{8}$ sú nezávislé. Každé X_i je hod mincou, takže $X_i \sim \mathrm{Bern}(p)$ Nevieme aké je p ale intuícia nám hovorí, že to bude blízke $\sum_{i=1}^{8} X_i/8$. Vskutku, $Y = \sum_{i=1}^{8} X_i \sim \mathrm{Bin}(8, p)$ a $\mathrm{E}[Y] = 8p$ preto $\mathrm{E}[Y/8] = p$. Osem realizácií je dosť málo, čím viacej by sme ich mali, tým bližšie by bol aritmetický priemer skutočnému priemeru.

Zákon veľkých čisel upresňuje túto intuíciu.

[24] Ak táto existuje.

V prvom rade treba zadefinovať, čo znamenená *blízke* alebo *blížiť sa.* Cudzím slovom *konvergovať.* Ešte predtým, čo znamená nezávislosť náhodných premenných pre viac ako len dve náhodné premenné.

Hovoríme, že náhodné premenné $X_1, X_2, \ldots, X_n$ sú **nezávislé** ak platí,

$$P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n) = P(X_1 \leq x_1) \cdot P(X_2 \leq x_2) \cdots P(X_n \leq x_n)$$

pre všetky $x_1, x_2, \ldots, x_n \in \mathbb{R}$.

Alternatívne by sme mohli zapísať

$$p_{X_1 X_2 \cdots X_n}(x_1, x_2, \ldots, x_n) \equiv P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n) = P(X_1 = x_1) \cdot P(X_2 = x_2) \cdots P(X_n = x_n),$$

pre diskrétne rozdelené náhodné premenné a všetky $x_i \in \mathcal{S}_{X_i}$ a

$$f_{X_1 X_2 \cdots X_n}(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdots f_{X_n}(x_n),$$

pre spojité náhodné premenné a všetky $x_1, x_2, \ldots, x_n \in \mathbb{R}$, kde $f_{X_1 X_2 \cdots X_n}(x_1, x_2, \ldots, x_n)$ označuje združenú mnohorozmernú funkciu hustoty. Platí pre ňu:

$$P\left(X_1 \in [a_1, b_1], \cdots, X_n \in [a_n, b_n]\right) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f_{X_1 X_2 \cdots X_n}(x_1, x_2, \ldots, x_n) dx_n \ldots dx_1.$$

Ide len o obyčajné zovšeobecnenie vzťahu z predošlej kapitoly pre viac ako dve premenné.

8.1 Konvergencia podľa pravdepodobnosti

Existujú rôzne spôsoby *blíženia sa* keď hovoríme o náhodných premenných. Jeden takýto koncept si teraz zadefinujeme. Majme postupnosť náhodných premenných $\{X_i\}_{i=1}^{\infty} = X_1, X_2, X_3 \cdots$.

Hovoríme, že postupnosť náhodných premenných $\{X_i\}_{i=1}^{\infty}$ **konverguje podľa pravdepodobnosti** k náhodnej premennej X ak platí:

$$\forall \epsilon > 0 : \lim_{n \to \infty} P(|X_n - X| < \epsilon) = 1.$$

Toto označujeme $X_n \to_P X$.

Pripomeňme si, čo je vlastne tento objekt $P(|X_n - X| < \epsilon)$:

$$P(|X_n - X| < \epsilon) = P\left(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| < \epsilon\}\right).$$

To znamená, že pre akékoľvek malé ϵ existuje nejaké dostatočne veľké n_0 také, že pre všetky $n \geq n_0$ bude pravdepodobnosť toho, že X_n bude ϵ-blízko X ľubovoľne blízka nule.[25] Pri fixnom ϵ sa pozeráme na čísla $a_n \equiv P(|X_n - X| < \epsilon)$ len ako na nejakú postupnosť a_n, ktorá sa mení s n.

Špeciálnym prípadom je, ak je limitná náhodná premenná X rovná nejakej konštante, teda ak $X_n \to_P c$.

[25] Tých idenfikátorov je naozaj veľa. Plný zápis je nasledovný:

$$\forall \epsilon > 0 : \forall \epsilon_1 > 0 : \exists n_0 \in \mathbb{N} : \forall n \geq n_0 : |P\left(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| < \epsilon\}\right) - 1| \leq \epsilon_1.$$

Všimnite si, že rola týchto dvoch malých ϵ a ϵ_1 je rôzna. Kým ϵ kontroluje ako blízko je X_n of X, ϵ_1 kontroluje ako blízko je hodnota $P(|X_n - X| < \epsilon)$ od nuly.

8.2 Markovova nerovnosť

Veta 8.1. *Majme nezápornú náhodnú premennú X a číslo $c \in \mathbb{R}$. Potom platí*

$$P(X \geq c) \leq \frac{E[X]}{c}.$$

Proof. Skonštruujme náhodnú premennú Y nasledovne:

$$Y = \begin{cases} 0, & \text{ak } X < c, \\ c, & \text{ak } X \geq c. \end{cases}$$

Z konštrukcie vyplýva, že $Y \leq X$ preto platí $\mathrm{E}[Y] \leq \mathrm{E}[X]$. Preto platí

$$\mathrm{E}[Y] = c \cdot P(X \geq c) + 0 \cdot P(X < c),$$

a preusporiadaním dostaneme želanú nerovnosť. □

Markovova nerovnosť je **tesná**, to znamená, že už ju nemôžeme vylepšiť. Ona totiž platí pre všetky nezáporné náhodné premenné, takže aj pre Y z dôkazu pre ktoré nastáva priamo rovnosť. Ak by sme chceli vylepšiť MN a nájsť menšiu hornú medzu ako $\frac{\mathrm{E}[X]}{c}$ pre všetky možné nezáporné náhodné premenné, došli by sme k sporu, lebo pre Y z dôkazu by neplatila.

8.3 Čebyševova nerovnosť

Veta 8.2. *Majme náhodnú premennú so strednou hodnotou μ a konečnou varianciou σ^2. Potom pre akékoľvek číslo $k > 0$ platí:*

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Alternatívnou formuláciou, ak zvolíme $k = \frac{c}{\sigma}$ je

$$P(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2}.$$

Proof. Označme $Y = (X - \mu)^2$, ktorá je nezáporná náhodná premenná a zároveň platí $\mathrm{E}[Y] = \sigma^2$. Naviac platí

$$P(|X - \mu| \geq k\sigma) = P((X - \mu)^2 \geq k^2\sigma^2) \leq \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2},$$

kde nerovnosť vyplýva z Markovovej nerovnosti. □

Nezávisle od typu rozdelenia:

- pravdepodobnosť, že sme viac ako 2 smerodajné odchýlky vzdialený od priemeru je menšia ako 1/4,
- pravdepodobnosť, že sme viac ako 3 smerodajné odchýlky vzdialený od priemeru je menšia ako 1/9,
- pravdepodobnosť, že sme viac ako 4 smerodajné odchýlky vzdialený od priemeru je menšia ako 1/16,
- pravdepodobnosť, že sme viac ako 5 smerodajné odchýlky vzdialený od priemeru je menšia ako 1/25.

Toto je ale len horná medza. V skutočnosti môže byť táto pravdepodobnosť oveľa menšia. Prečo je tomu tak? No táto nerovnosť je len tak dobrá ako je dobrá Markovovská nerovnosť v dôkaze Čebyševovej nerovnosti.

8.4 Slabý Zákon Veľkých čísel

Pripomeňme, že pre nezávislé náhodné premenné X, Y platí $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$.

Pre n nezávislých náhodných premenných $X_1, ..., X_n$ analogicky platí $\mathrm{Var}[X_1 + ... + X_n] = \mathrm{Var}[X_1] + ... + \mathrm{Var}[X_n]$.

Pripomeňme tiež, že $\mathrm{Var}[aX+b] = a^2\mathrm{Var}[X]$.[26] Ide o použitie tohoto výsledku, ak Y je konštantná 1 a teda $\mathrm{Var}(Y) = 0$ a $\mathrm{Cov}(X,Y) = 0$.

$$\begin{aligned}\mathrm{Var}[aX+bY] &= \mathrm{E}\left[[(aX+bY] - \mathrm{E}[aX+bY)]^2\right] \\ &= \mathrm{E}\left[(a(X-\mathrm{E}[X]) + b(Y-\mathrm{E}[Y])^2\right] \\ &= \mathrm{E}\left[a^2\,(X-\mathrm{E}[X])^2 + b^2\,(Y-\mathrm{E}[Y])^2 + 2ab\,(X-\mathrm{E}[X])(Y-\mathrm{E}[Y])\right] \\ &= \mathrm{E}\left[a^2\,(X-\mathrm{E}[X])^2\right] + \mathrm{E}\left[b^2\,(Y-\mathrm{E}[Y])^2\right] + \mathrm{E}\left[2ab\,(X-\mathrm{E}[X])(Y-\mathrm{E}[Y])\right] \\ &= a^2\,\mathrm{Var}[X] + b^2\,\mathrm{Var}[Y] + 2ab\,\mathrm{Cov}[X,Y]\ .\end{aligned}$$

Teraz príde jeden z najdôležitejších výsledkov prezentovaných v rámci tohoto kurzu.

[Dramatická pauza.]

Veta 8.3. *Majme postupnosť nezávislých náhodných premenných s rovnakou strednou hodnotou μ a konečnou varianciou σ^2. Potom platí*

$$\bar{X}_n \equiv \frac{\sum_{i=1}^n X_i}{n} \to_P \mu.$$

Proof. Nakoľko $X_1, X_2, X_3, \cdots$ sú nezávislé platí

$$\mathrm{Var}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathrm{Var}[X_i] = n\sigma^2,$$

preto

$$\mathrm{Var}\left[\bar{X}_n\right] = \mathrm{Var}\left[\sum_{i=1}^n X_i/n\right] = \frac{1}{n^2}\mathrm{Var}\left[\sum_{i=1}^n X_i\right] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Naviac

$$\mathrm{E}\left[\bar{X}_n\right] = \mathrm{E}\left[\sum_{i=1}^n X_i/n\right] = \sum_{i=1}^n \mathrm{E}[X_i]/n = \frac{n\mu}{n} = \mu.$$

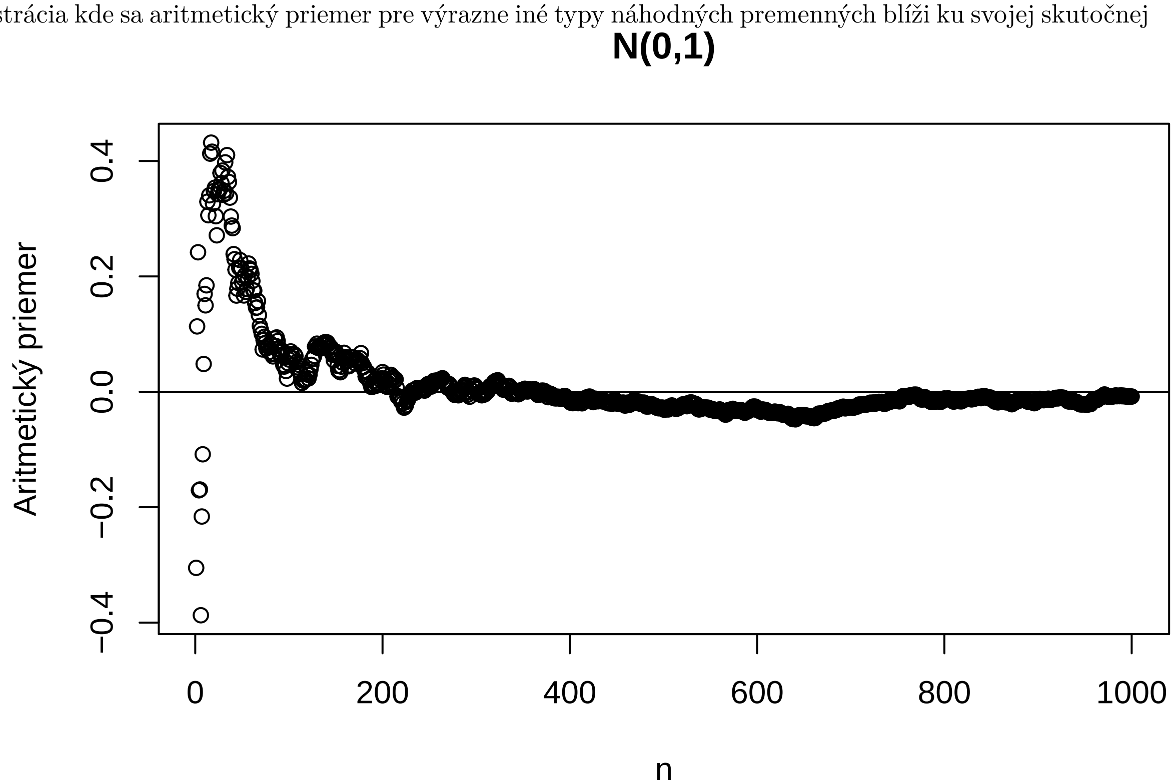
Použitím Čebyševovej nerovnosti dostávame

$$P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2 n},$$

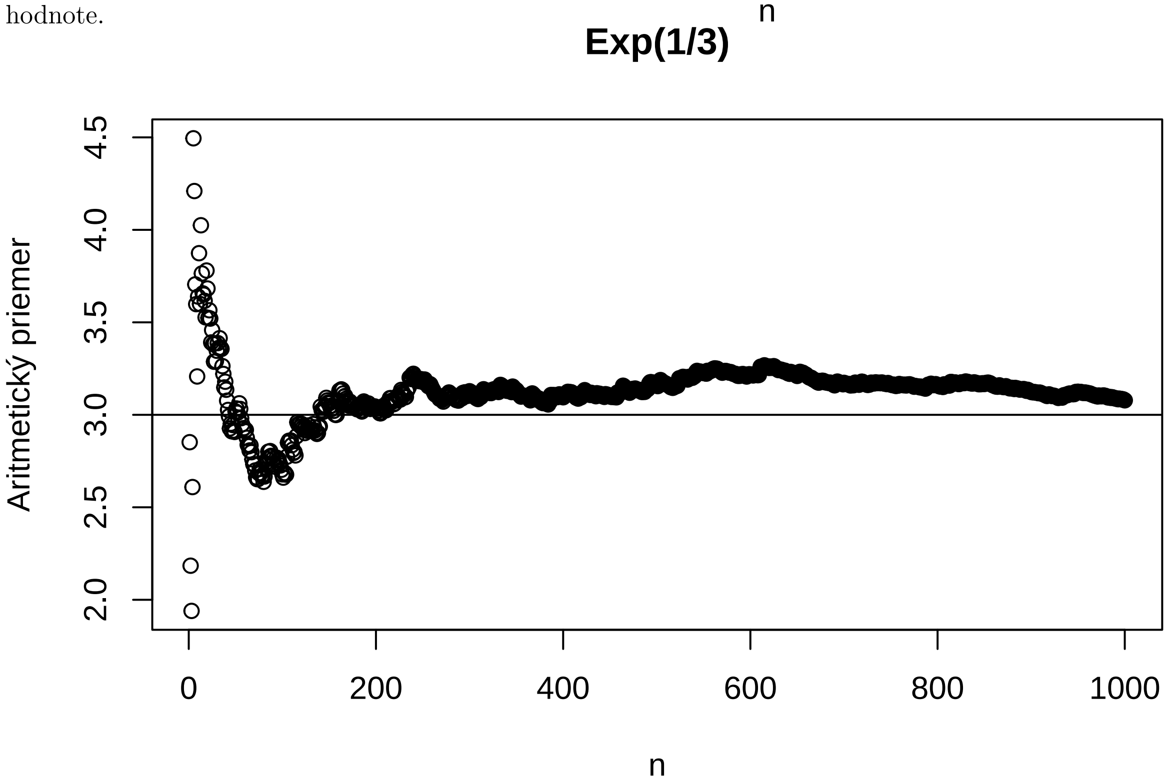
a preto $\lim_{n\to\infty} P(|\bar{X}_n - \mu| \geq \epsilon) = 0$, teda $\lim_{n\to\infty} P(|\bar{X}_n - \mu| < \epsilon) = 1$, pre akékoľvek $\epsilon > 0$, čo sme chceli ukázať. □

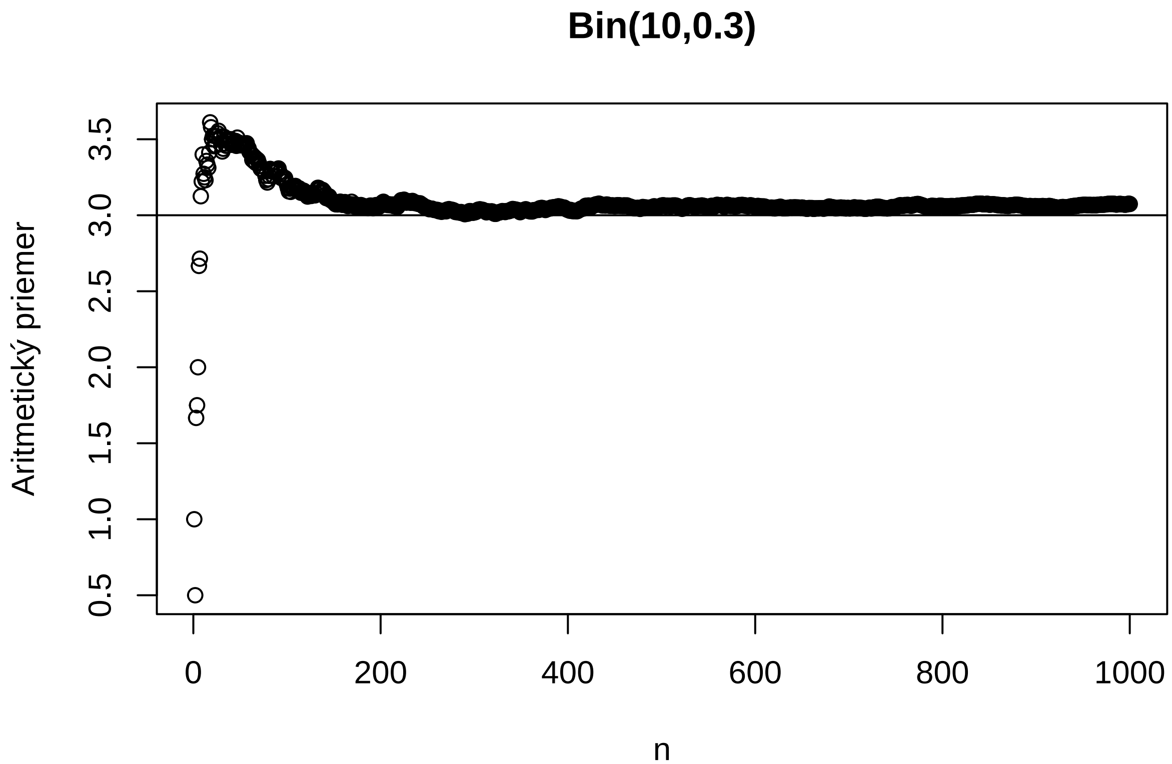
[26] Ak $\mathrm{Var}[X] < \infty$.

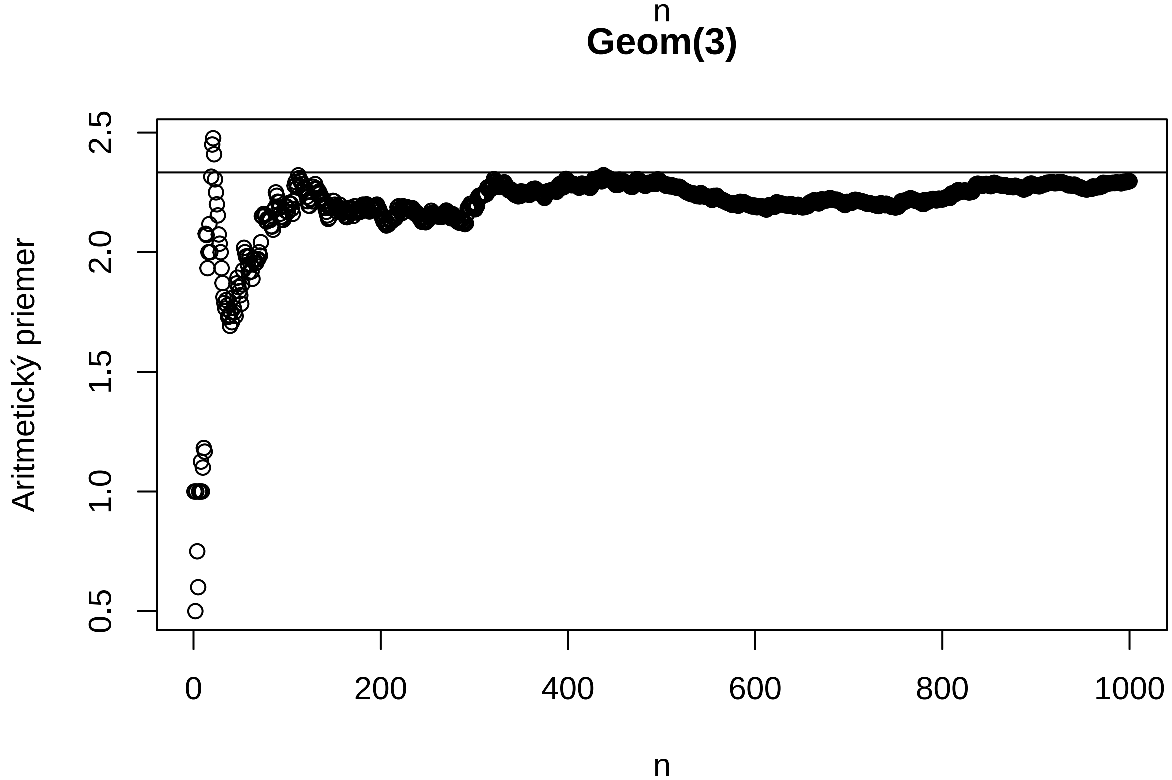
Tu je ilustrácia kde sa aritmetický priemer pre výrazne iné typy náhodných premenných blíži ku svojej skutočnej



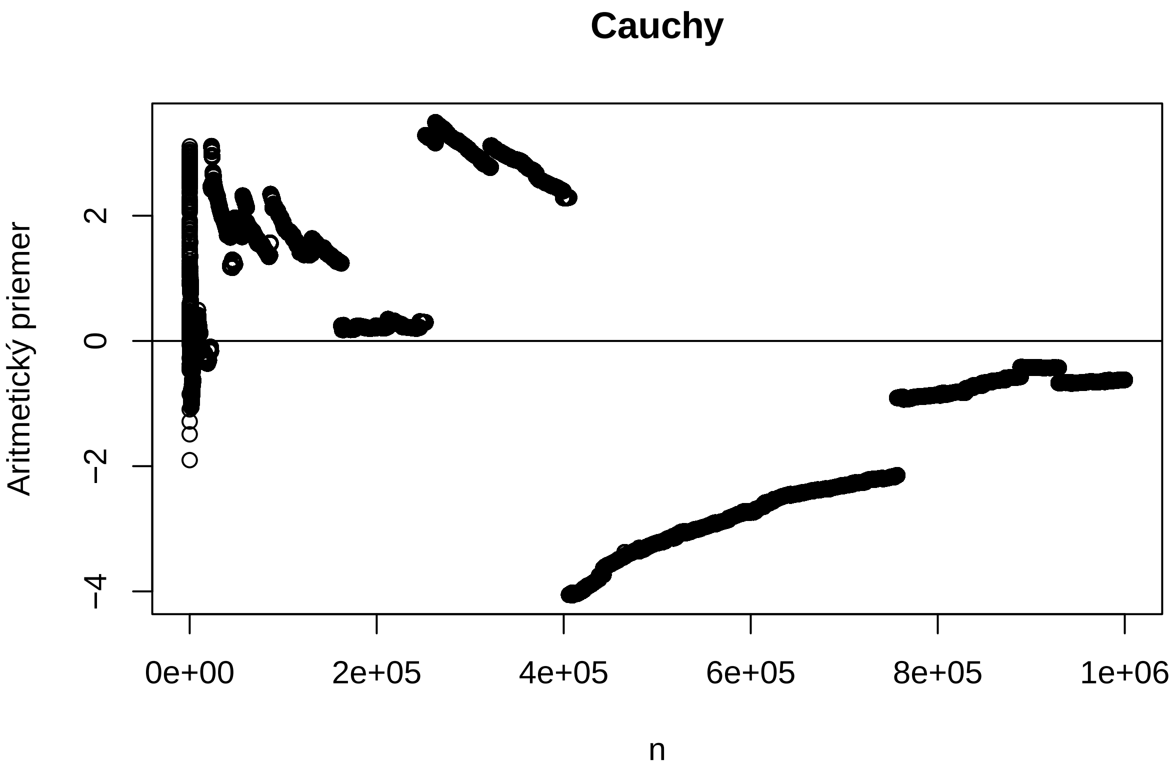
hodnote.







Ak však vyskúšame Cauchyho rozdelenie, ktoré nemá strednú hodnotu, dostaneme nasledujúci obrázok.



Zákon veľkých čísel o tomto prípade nehovorí nič.

Príklad 8.2. Nech A je nejaká udalosť a nech

$$1_A(\omega) = \begin{cases} 1, & \text{ak } \omega \in A, \\ 0, & \text{inak,} \end{cases}$$

je identifikátorová funkcia udalosti A.

Potom platí $\mathrm{E}[1_A] = 1 \cdot P(A) + 0 \cdot P(A^C) = P(A)$.

Preto ak odsimulujeme na počítači 10000 simulácií a v nich A nastala 13 krát, odhadujeme, že $P(A)$ je rovné 0.0013. Tento odhad sa bude blížiť ku skutočnej hodnote pre stále väčšie množstvo simulácií. Je to kvôli ZVČ, lebo $\bar{X}_n = \frac{0+0+1+\cdots+0}{10000} = 0.0013$, kde X_i majú rovnaké pravdepodobnostné rozdelenie ako 1_A a sú nezávislé.

Príklad 8.3. Chceli by sme simulačne vypočítať číslo π. Obsah štvrťkruhu so stredom v $[0,0]$ s polomerom 1 vo štvorci $[0,1] \times [0,1]$ je rovný $\pi/4$. Vygenerujeme mnoho čísel z $[0,1] \times [0,1]$ a pozrieme sa na proporciu tých, ktoré skončia vo štvrťkruhu. Nech $X_i = 1$ vtedy ak dané číslo patrí štvrťkruhu. Potom

$$P\left(\left|\frac{X_1 + X_2 + \cdots X_n}{n} - \frac{\pi}{4}\right| \geq \epsilon\right) \leq \frac{\mathrm{Var}[X_i]}{n\epsilon^2} = \frac{\frac{\pi}{4}(1-\frac{\pi}{4})}{n\epsilon^2}.$$

Nevieme však koľko je $\frac{\pi}{4}(1-\frac{\pi}{4})$ (toto číslo chceme aproximovať) ale funkcia $p(1-p)$ nadobúda svoje maximum v 1/4. Preto

$$P\left(\left|\frac{X_1 + X_2 + \cdots X_n}{n} - \frac{\pi}{4}\right| \geq \epsilon\right) \leq \frac{1}{4n\epsilon^2}.$$

Ak napríklad chceme dostať simulačný odhad na 3 desatinné miesta, teda s presnosťou 1/1000, musíme zvoliť $\epsilon = 1/4000$.[27]: Ak si chceme byť istý, že náš odhad bude v tomto rozmedzí s pravdepodobnosťou 1%,

[27]Je tam $\epsilon = 1/4000$ a nie $\epsilon = 1/1000$ kvôli tomu, že chceme mať hodnotu π dostatočne presnú. Ale v rovnici je $\frac{\pi}{4}$.

potrebujeme aby

$$\frac{4000^2}{4n} = \frac{4,000,000}{n} \leq 0.01 \implies n \geq 400,000,000.$$

Takže potrebujeme aspoň 400 miliónov simulácií.[28]

8.5 Čo zákon o veľkých číslach nehovorí

Hádžeme férovou mincou. Za posledných 20 hodov nám nepadla ani raz hlava. Ale vieme, že platí ZVČ, preto by teraz malo padať viacej hláv, nie?

Veru nie. ZVČ hovorí o *limitnom prípade* (a nekonečno je o dosť viacej ako 20). Skutočnosť, že sa nám teraz nedarilo hádzať hlavy (alebo hádzať 6ku v doskovej hre) nijakovsky nesúví s tým, čo sa bude diať neskôr. Tie hody sú stále nezávislé a šanca uvidieť hlavu je rovnako veľká ako na začiatku hádzania.

8.6 Cvičenia

Cvičenie 8.1. Porovnajte Markovovskú nerovnosť pre $P(X \geq 4)$ so skutočnou hodnotou, ak $X \sim \text{Exp}(1/2)$. Kedy je Markovovská nerovnosť neinformatívna?

Cvičenie 8.2. Majme náhodnú premennú, pre ktorú platí $P(X \geq 0) = 1$ a $P(X \geq 10) = 1/5$. Ukážte, že $\mathrm{E}[X] \geq 2$.

Cvičenie 8.3. Majme náhodnú premennú, pre ktorú platí $\mathrm{E}[X] = 10, P(X \leq 7) = 0.2, P(X \geq 13) = 0.3$. Dokážte, že $\mathrm{Var}[X] \geq 9/2$.

Cvičenie 8.4. Akú veľkú musíme zvoliť vzorku nezávislých náhodných premenných (s konečnou strednou hodnotou a varianciou), aby bola pravdepodobnosť, že sa bude aritmetický priemer nachádzať bližšie ako dve smerodajné odchýlky od strednej hodnoty, aspoň 99%?

Cvičenie 8.5. Majme postupnosť $X_1, X_2, \ldots$ nezávislých náhodných premenných so strednou hodnotou 3.5 a varianciou 35/12. Akú veľkú musíme nastaviť hodnotu n, aby sme si boli istí, že

$$P(3 < \bar{X}_n < 4) \geq 0.8.$$

Cvičenie 8.6. Majme postupnosť $X_1, X_2, \ldots$ náhodných premenných nezávislých hodov férovou kockou. Koľkokrát musíme hodiť hockou, aby sme si boli istí, že

$$P(3 < \bar{X}_n < 4) \geq 0.8.$$

Porovnajte s predošlým výsledkom.

Cvičenie 8.7. Majme postupnosť $X_1, X_2, \ldots$ náhodných premenných, pre ktoré platí $P(X_n = n^2) = 1/n$ a $P(X_n = 0) = 1 - 1/n$. Ukážte, že $X_n \to_P 0$ a zároveň $\lim_{n\to\infty} \mathrm{E}[X_n] = \infty$.

Cvičenie 8.8. Férovou mincou hádžete nezávisle 100 krát. Odhadnite pravdepodobnosť pomocou Čebyševovej nerovnosti, že hlavou hodíte viac ako 30, ale menej ako 70 krát. Porovnajte so skutočnou pravdepodobnosťou.

9 Centrálna limitná veta

Normálne rozdelenie hrá prekvapivo dôležitú úlohu v teórii pravdepodobnosti a v štatistike. Ukazuje sa, že aritmetický priemer z nezávislých, ale rovnako rozdelených náhodných premenných, pokiaľ je počítaný z dostatočne veľkého množstva premenných, sa správa ako normálne rozdelená náhodná premenná.

[28] Čo je mimochodom dosť veľa.

9.1 Konvergencia podľa distribúcie

V minulej kapitole sme si predstavili jeden zo spôsobov, ako sa môže postupnosť náhodných premenných blížiť k nejakej náhodnej premennej, konkrétne konvergenciu podľa pravdepodobnosti. Existuje aj iný spôsob. Niekedy chceme uvažovať situáciu, že pravdepodobnostné správnanie prvkov postupnosti náhodných premenných sa stále viacej a viacej podobá na pravdepodobnostné správanie akejsi limitnej náhodnej premennej.

Hovoríme, že postupnosť náhodných premenných $\{X_i\}_{i=1}^{\infty}$ **konverguje podľa distribúcie** k náhodnej premennej X, ak platí pre všetky body x spojitosti funkcie F_X:

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x),$$

Označujeme $X_n \to_D X$.

Veta 9.1. *Platí*

$$X_n \to_P X \implies X_n \to_D X.$$

Proof. Nakoľko platí $\{X_n \leq x\} \subset \{X \leq x+\epsilon\} \cup \{|X - X_n| > \epsilon\}$ (lebo ak platí $X_n \leq x$ a $X > x+\epsilon$, potom nutne aj $|X - X_n| > \epsilon$), dostávame:

$$P(X_n \leq x) \quad \leq \quad P(X \leq x+\epsilon) + P(|X_n - X| > \epsilon).$$

Podobnou úvahou dostaneme

$$P(X \leq x-\epsilon) \quad \leq \quad P(X_n \leq x) + P(|X_n - X| > \epsilon).$$

Preto platí

$$P(X \leq x-\epsilon) - P(|X_n - X| > \epsilon) \leq P(X_n \leq x) \leq P(X \leq x+\epsilon) + P(|X_n - X| > \epsilon).$$

Ak aplikujeme operátor limity na tieto nerovnosti, spolu s využitím definície konvergencie podľa pravdepodobnosti dostávame

$$F_X(x-\epsilon) \leq \lim_{n\to\infty} P(X_n \leq x) \leq F_X(x+\epsilon),$$

a tieto nerovnosti platia pre akékoľvek ϵ.

V bode spojitosti F_X platí $\lim_{\epsilon\to 0} F_X(x-\epsilon) = \lim_{\epsilon\to 0} F_X(x+\epsilon) = F_X(x)$, a preto $\lim_{n\to\infty} P(X_n \leq x) = F_X(x)$, čo sme chceli ukázať. □

9.2 Centrálna limitná veta

Veta 9.2. *Majme postupnosť nezávislých a rovnako rozdelených náhodných premenných $X_1, X_2, X_3, \ldots$ s konečnou strednou hodnotou μ a konečnou varianciou σ^2. Potom platí*

$$\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \to_D Z,$$

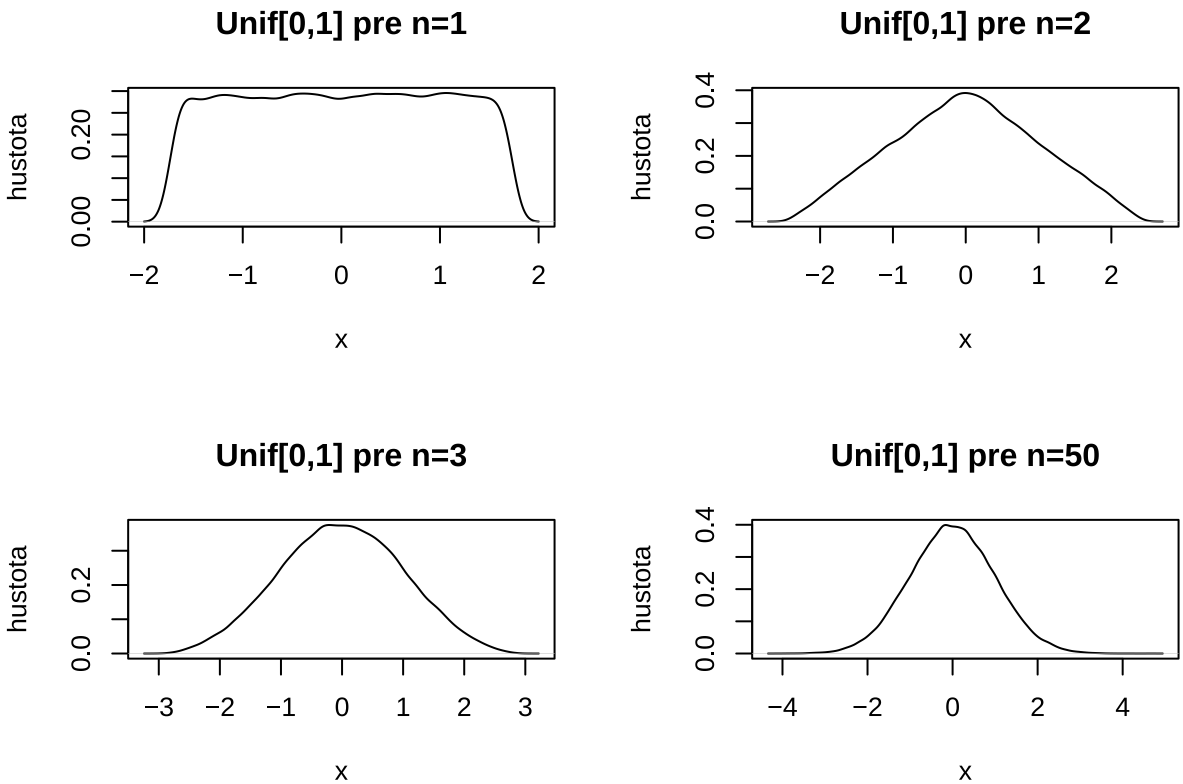
kde $Z \sim N(0,1)$.

Alternatívnym zápisom je, že ak označíme $Y_n = \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma}$, potom

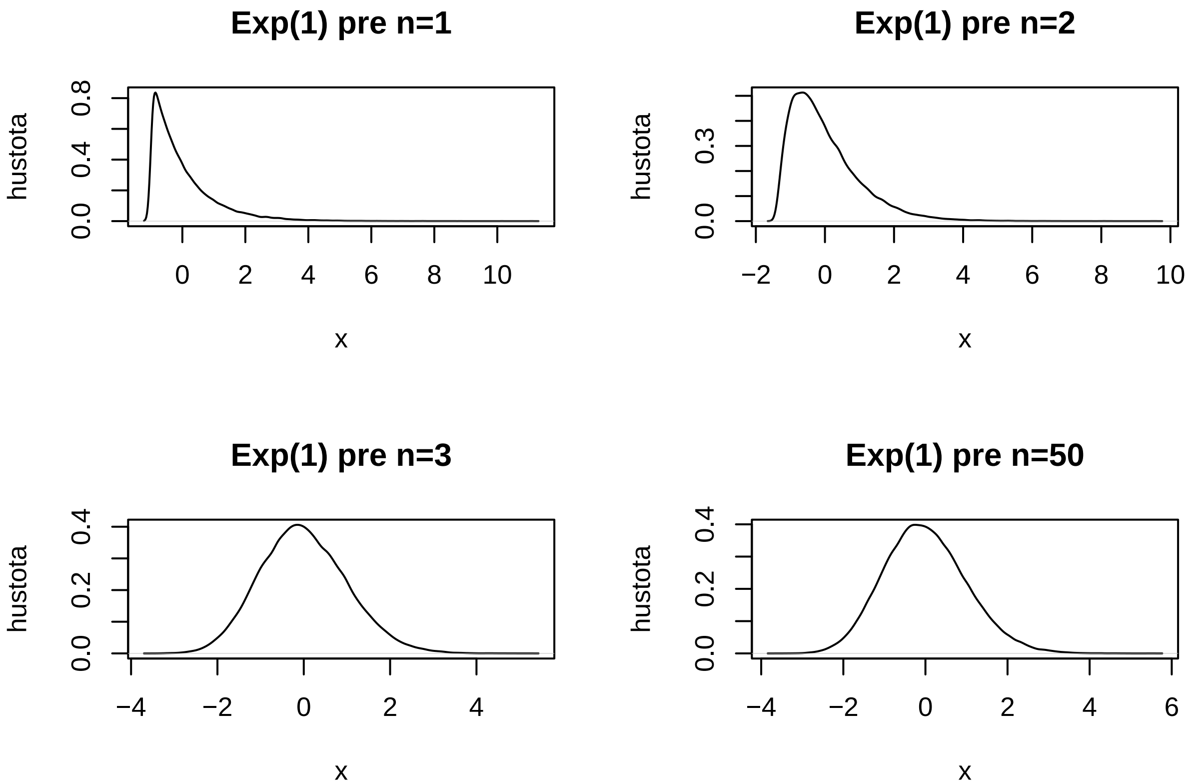
$$\forall x \in \mathbb{R} : \lim_{n\to\infty} F_{Y_n}(x) = \Phi(x).$$

Dôkaz je nad rámec tohoto kurzu, a preto ho vynechávame.

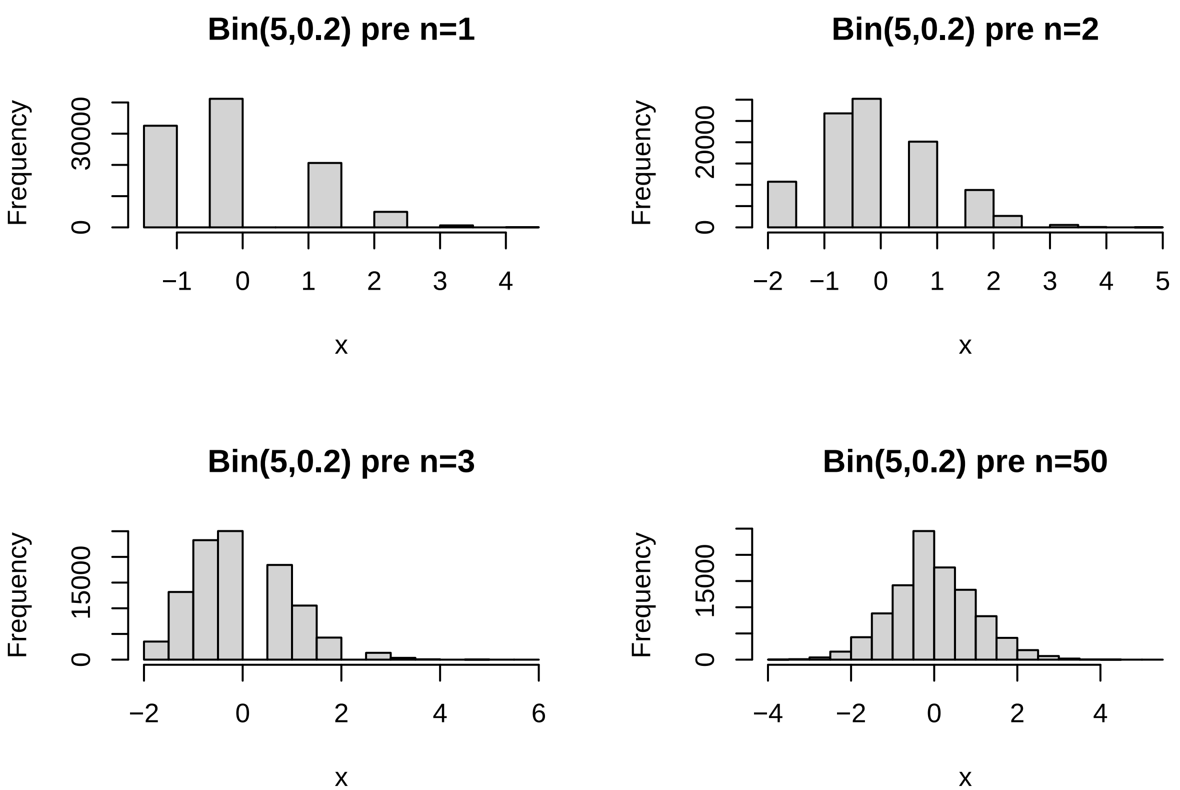
Tu je ilustrácie pre rovnomerné rozdelenie. Ide o obrázky odhadov funkcie hustoty. Napriek veľkému množstvu simulácií nie sú úplne hladké, viacej o takýchto odhadoch hustôt sa naučíme na druhom kurze regresie.



Tu pre exponenciálne rozdelenie:



A tu pre binomické rozdelenie:



Toto má dôležité praktické dôsledky. My vôbec nemusíme vedieť, aké má nejaká náhodná pravdepodobnostné rozdelenie. Ale vieme, že priemer nezávislých náhodných premenných sa už bude správať systematicky(!). Toto je veľmi všeobecný výsledok. Na tomto poznatku je založených mnoho štatistických testov.

Príklad 9.1. Hádžeme férovou kockou 900 krát. Ideme aproximovať pravdepodobnosť, že uvidíme viacej ako 495 hláv. Každý hod mincou $X_i \sim \text{Bern}(0.5)$ a $\text{E}[X_i] = 0.5, \text{Var}[X_i] = 0.25$.

Preto

$$\begin{aligned}
P\left(\sum_{i=1}^{900} X_i > 495\right) &= P\left(\frac{\sum_{i=1}^{900} X_i}{900} > \frac{495}{900}\right) \\
&= P\left(\bar{X}_n - 0.5 > \frac{495}{900} - 0.5\right) \\
&= P\left(\frac{\bar{X}_n - 0.5}{0.5} > \frac{\frac{495}{900} - 0.5}{0.5}\right) \\
&= P\left(\sqrt{900}\frac{\bar{X}_n - 0.5}{0.5} > \sqrt{900}\frac{\frac{495}{900} - 0.5}{0.5}\right) \\
&= P\left(\sqrt{900}\frac{\bar{X}_n - 0.5}{0.5} > 30\frac{\frac{495}{900} - 0.5}{0.5}\right) \\
&= P\left(\sqrt{900}\frac{\bar{X}_n - 0.5}{0.5} > 3\right) \\
&\approx 1 - \Phi(3) = 0.0013
\end{aligned}$$

Princíp výpočtu takýchto aproximácií je vždy rovnaký. Začneme s tým, čo chceme vypočítať a ekvivalentnými úpravami to prevedieme na formuláciu CLV.

Príklad 9.2. Majme nezávislé $X_1, X_2, \cdots X_{12}$ kde $X_i \sim \text{Unif}[0,1]$. Pomocou CLV aproximujte $P(|\bar{X}_n - 0.5| \leq 0.1)$.

$$\begin{aligned}
P(|\bar{X}_n - 0.5| \leq 0.1) &= P(\sqrt{12}(\bar{X}_n - 0.5)| \leq \sqrt{12} \cdot 0.1) \\
&= P\left(\left|\sqrt{12}\frac{\bar{X}_n - 0.5}{\sqrt{\frac{1}{12}}}\right| \leq \sqrt{12}\frac{0.1}{\sqrt{\frac{1}{12}}}\right) \\
&= P\left(\left|\sqrt{12}\frac{\bar{X}_n - 0.5}{\sqrt{\frac{1}{12}}}\right| \leq 1.2\right) \\
&\approx \Phi(1.2) - \Phi(-1.2) = 0.7698.
\end{aligned}$$

Využili sme skutočnosť, že $\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma}$ je približne normovane normálne rozdelená náhodná premenná ($\text{N}(0,1)$).

9.3 Cvičenia

Cvičenie 9.1. Majme postupnosť nezávislých a rovnako rozdelených náhodných premenných $X_1, X_2, X_3, \cdots$, kde $X_1 \sim \text{Exp}(1)$.

Aké veľké musí byť n aby platilo $P(0.9 \leq \bar{X}_n \leq 1.1) \geq 0.9$?

Cvičenie 9.2. Výťah unesie 4tony nákladu. Majme 100 krabíc, každá s priemernou váhou 39kg a so smerodajnou odchýlkou 2kg. Aproximujte pravdepodobnosť, že výťah bude preťažený.

Cvičenie 9.3. Majme postupnosť nezávislých a rovnako rozdelených náhodných premenných $X_1, X_2, X_3, \cdots$, kde $\text{E}[X_1] = 200, \text{Var}[X_1] = 40, n = 100$.

Aproximujte pomocou CLV hodnotu $P(190 \leq \bar{X}_n \leq 210)$.

10 Opakovanie

Prebrali sme tieto *témy*:

- Interpretácia pravdepodobnosti,
- Pravdepodobnostný priestor a jeho vlastnosti,
- Podmienená pravdepodobnosť a Bayesova veta,
- Náhodná premenná,
- Diskrétne náhodné premenné,
- Spojité náhodné premenné,
- Súvis medzi náhodnými premennými,
- Zákon veľkých čísel,
- Centrálna limitná veta.

Zaviedli sme si **nasledovné pojmy**, ktorých definíciu musíme poznať. Nové pojmy boli vždy **označované hrubým písmom**:

- Pravdepodobnostný priestor - množina potenciálnych dopadnutí experimentu Ω, množina udalostí $\mathcal{F}$, pravdepodobnosť P,
- Rozklad množiny,
- Nezávislosť udalostí,
- Podmienená pravdepodobnosť,
- Náhodná premenná,
- Kumulatívna distribučná funkcia,
- Pravdepodobnostná funkcia,
- Funkcia hustoty pravdepodobnosti,
- Stredná hodnota,
- Variancia,
- Smerodajná odchýlka,
- Medián,
- Združená kumulatívna distribučná funkcia,
- Združená pravdepodobnostná funkcia,
- Združená funkcia hustoty pravdepodobnosti,
- Nezávislosť náhodných premenných,
- Kovariancia,
- Korelácia,
- Konvergencia podľa pravdepodobnosti,
- Konvergencia podľa distribúcie.

Taktiež musíme poznať vlastnosti týchto objektov a aké sú medzi nimi vzťahy. Je dôležité poznať tieto vzťahy:

- Náhodná premenná a pravdepodobnostný priestor,
- Kumulatívna distribučná funkcia a pravdepododobnosť,
- Kumulatívna distribučná funkcia a pravdepodobnostná funkcia,
- Kumulatívna distribučná funkcia a funkcia hustoty,
- Stredná hodnota, variancia a pravdepodobnosntná funkcia,
- Stredná hodnota, variancia a funkcia hustoty,
- Stredná hodnota a kovariancia.
- Stredná hodnota, smerodajná odchýlka a korelácia.

Hovorili sme tiež o týchto vetách/tvrdeniach:

- Bayesova veta,
- Linearita strednej hodnoty: $\mathrm{E}[aX + bY] = a\mathrm{E}[X] + b\mathrm{E}[Y]$,
- Vzťah pre varianciu: $\mathrm{Var}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$,
- Variancia lineárnej transformácie: $\mathrm{Var}[aX + b] = a^2\mathrm{Var}[X]$,
- Z nezávislosti náhodných premenných vyplýva, že sú nekorelované (naopak to neplatí),
- Pre varianciu platí $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}[X, Y]$,

- Markovova nerovnosť,
- Čebyševova nerovnosť,
- Zákon veľkých čísel,
- Z konvergencie podľa pravdepodobnosti vyplýva konvergencia podľa distribúcie (naopak to neplatí),
- Centrálna limitná veta.

Dosť času sme strávili aj spoznávaním rôznych pravdepodobnostných distribúcií, či už diskrétne alebo spojite rozdelených:

- Diskrétne: Rovnomerné, Bernoulliho, Binomické, Poissonovo, Geometrické, Hypergeometrické a Negatívne binomické,
- Spojité: Rovnomerné, Normálne, Exponenciálne, Chí-kvadrát a Studentovo.

Ambíciou tohoto kurzu bolo zjednodušiť prechody medzi rôznymi spôsobmi popisu typu náhodnosti. Ak máme kompletnú informáciu o pravdepodobnostnom správaní, napríklad formou kumulatívnej distribučnej funkcie, musíme byť schopný odvodiť všetko ostatné. Napríklad funkciu hustoty alebo jej charakteristiky ako napríklad strednú hodnotu, varianciu, smerodajnú odchýlku alebo medián. Familiarita a plynulosť prechodu je to, o čo sme sa snažili. Znalosť tohoto jazyka je nutnou podmienkou úspešného pokračovania v štúdiu pravdepodobnosti a štatistiky, ako aj praktickej dátovej analýzy.

Cvičné otázky

- Zadefinujte pravdepodobnostný priestor a náhodnú premennú, ktoré budú zodpovedať súčtu počtu bodiek na dvoch nezávisle hodených kociek.
- Dokážte, že so spojitej aditivity pravdepodobnosti vyplýva aj konečná aditivita.
- Aká je stredná hodnota náhodnej premennej ktorá nadobúda hodnoty $1, 2, 3, 4, 5$ s pravdepodobnosťami $\frac{c}{1}, \frac{c}{2}, \frac{c}{3}, \frac{c}{4}, \frac{c}{5}$, kde c je konštanta.
- Načrtnite kumulatívnu distribučnú funkciu náhodnej premennej, ktorá označuje počet bodiek, ktorý padne na férovej kocke.
- Vymyslite pravdepodobnostný priestor a dve funkcie $\Omega \to \mathbb{R}$, také, že jedna z nich bude náhodná premenná ale druhá nebude.
- Majme diagnostický test so špecificitou 90% a senzitivitou 99%. Vypočítajte pravdepodobnosť choroby v prípade, že test je pozitívny.
- Majme dve náhodné premenné X a Y také, že $P(X = 0, Y = 0) = 0.06$, $P(X = 1, Y = 0) = 0.24$, $P(X = 0, Y = 1) = 0.14$, $P(X = 1, Y = 1) = 0.56$. Vypočítajte $\mathrm{Cov}[X, Y]$. Sú náhodné premenné X a Y nezávislé?
- Uvažujme náhodnú premennú Y o ktorej vieme, že

$$\begin{aligned} p_Y(1) &= 0.2, \\ p_Y(2) &= 0.3, \\ p_Y(3) &= 0.5. \end{aligned}$$

  Vypočítajte $\mathrm{E}[Y], \mathrm{Var}[Y], \mathrm{sd}[Y]$.
- Majme náhodnú premennú s nasledovnou funkciu hustoty pravdepodobnosti:

$$f_X(x) = \begin{cases} c(x - \frac{x^3}{4}), & \text{ak } x \in (0, 2), \\ 0, & \text{inak.} \end{cases}$$

  Vypočítajte hodnotu konštanty c. Nájdite a načrtnite jej kumulatívnu distribučnú funkciu. Vypočítajte jej strednú hodnotu, varianciu, medián a $P(|X - 1| \leq 0.5)$
- Majme náhodné premenné X, Y s nasledovnou združenou funkciou hustoty

$$f_{XY}(x, y) = \begin{cases} cy^3, & \text{ak } x \in [0, 3],\ y \in [0, 1], \\ 0, & \text{inak.} \end{cases}$$

  Vypočítajte hodnotu konštanty c, $P(X + Y > 2)$, $P(X > Y)$, $P(X = 3Y)$, $\mathrm{E}[Y]$, $\mathrm{Cov}[X, Y]$.

- Aké je pravdepodobnostné rozdelenie počtu uhádnutých otázok na ABC teste z celkového počtu 10 otázok, ak vieme, že z prvých troch otázok otázok bola správne zodpovedaná práve jedna otázka?
- Nech je pravdepodobnosť nárazu asteroidu na zem 1/10000 za rok. Aká je pravdepodobnosť, že za 200 rokov narazí asteroid práve jedenkrát?
- Ktorým pravdepodobnostným rozdelením by ste modelovali: počet dopravných nehôd? Dobu kým nastavne ďalšia dopravná nehoda? Počet neúspešných žiadostí o grant, kým sa Vám to nepodarí? Výšku afrických slonov? Chybu merania? Priemer z veľkého množstva nezávislých náhodných premenných?
- Napíšte názov pravdepodobnostného rozdelenie, pre ktoré nemôžeme použiť Zákon veľkých čísel.
- Skonštruujte dve nekorelované náhodné premenné, ktoré nie sú nezávislé.
- Majme postupnosť $X_1, X_2, \ldots$ náhodných premenných so strednou hodnotou 3 a varianciou 1. Aspoň akú veľkú musíme nastaviť hodnotu n tak, aby platilo, že

  $$P(3 < \bar{X}_n < 4) \geq 0.8.$$

  Porovnajte výsledky založené na základe Čebyševovej nerovnosti a Centrálnej limitnej vety.
- Zo skúseností vieme, že na matematický ples sa lístky veľmi rýchlo vypredajú: každý človek v rade si kúpi v priemere 2.3 lístkov zo smerodajnou odchýlkou 2. Máme 250 voľných miest a v rade čaká 100 ľudí. Aproximujte pravdepodobosť, že sa každému ujde toľko lístkov, koľko chce. Explicitne pomenujte zjednodušujúce predpoklady, ktoré urobíte.
- Majme 100 mužov na palube lietadla, hmotnosť každého z nich má strednú hodnotu 80 a smerodajnú odchýlku 10. Pomocou CLV aproximujte pravdepodobnosť, že ich celková hmotnosť nepresiahne 9000kg. Uvažujte, že ich váhy sú nezávislé.

[Budem veľmi rád za akúkoľvek spätnú väzbu. Preklepy, logické chyby, nejasnosti. Ďakujem!]