NMAI059 Pravděpodobnost a statistika 1

Robert Šámal

27. června 2024

1 Pravděpodobnost – úvod

Jak si při čtení asi brzy všimnete, jedná se o pracovní verzi s mnohými překlepy a nedodělky, snad bude i tak užitečná ke studiu. Pokud objevíte nějaký překlep (a zejména nějakou závažnější chybu), budu vděčný za informaci na email samal@iuuk.mff.cuni.cz. Zatím chyby objevili Jan Adámek, Tomáš Boďa, Miroslav Cimerman, Šimon Dvořák, Dominik Farhan (ten i poradil zjednodušení důkazu Věty 83), Aurel Farkašovský, Adam Holda, Jakub Kubík, Matej Lieskovský, Jan Mitka, David Nguyen, Šimon Pajger, Pavol Rajczy, David Stanovský, Daniel Skýpala, Adam Šebesta, Milan Veselý, Pavel Valtr, Kryštof Višňák a Kateřina Vokálová. Děkuji!

Aplikace na rozehřátí Dříve než začneme s budováním pravděpodobnosti od začátku, ukažme si jednu rychlou aplikaci.

Příklad 1. *Dány dva polynomy f, g stupně d. Chceme zjistit, zda jsou stejné, a to co nejrychleji. Polynom f je dán svými koeficienty, tj. $f(x) = \sum_{i=0}^{d} a_i x^i$, polynom g lze psát jako $g_1 \cdot g_2$ a máme zadány koeficienty polynomů g_1 a g_2.*

Zjevné řešení je najít koeficienty polynomu g a pak srovnat, pro $i = 0, \ldots, d$, jestli koeficient v g u x^i je roven a_i. To je ale dost pomalé: $O(d^2)$ pokud budeme násobit přímočaře, nebo $O(d \log d)$ pokud použijeme pro násobení polynomů FFT. Ukážeme si, jak ověření provést v čase $O(d)$, pokud nám nevadí malá pravděpodobnost chyby.

Zvolíme parametr k, jako malé přirozené číslo. Vybereme náhodně a nezávisle (tj. opakovaně voláme funkci random.randint() nebo její ekvivalent) čísla $x_1, \ldots, x_k$ z množiny $S = \{1, 2, \ldots, 10d\}$. Pro každé z nich ověříme, zda $f(x_i) = g_1(x_i)g_2(x_i)$. (To lze jistě udělat v lineárním čase.) Pokud pro nějaké i rovnost neplatí, tak víme jistě, že $f \neq g$. Pokud rovnost platí pro všechna $i = 1, \ldots, k$, budeme mít za to, že $f = g$. Jaká je pravděpodobnost, že jsme se zmýlili? Všechna x_i jsou kořeny polynomu $f - g$. Ten má stupeň nejvýše d a není nulový (jinak jsme se nezmýlili). Proto má $f - g$ nejvýše d kořenů – a všechna čísla x_i jsou mezi nimi. To se pro jedno i stane s pravděpodobností $\leq \frac{d}{10d} = \frac{1}{10}$, pro všechna i s pravděpodobností $1/10^k$.

Nalezli jsme tzv. pravděpodobnostní algoritmus: v lineárním čase zjistíme, zda $f = g$, přičemž kladná odpověď může být špatně s pravděpodobností 10^{-k}.

Jedná se o nejjednodušší případ Schwartz-Zippelova algoritmu. Jeho obecná verze funguje pro polynomy ve více proměnných. Na tomto principu je založena i jedna metoda randomizovaného testování prvočíselnosti.

Pravděpodobnost – intuice, definice Některé jevy neumíme nebo nechceme popsat kauzálně: hod kostkou sice popisují fyzikální zákony, ale pokud by bylo možné změřit přesně způsob házení a předpovědět, které číslo padne, tak by to hry s kostkami pokazilo, obdobně při hodu šipkou na terč.

Počet emailů, které dostaneme za jeden den, typicky závisí na rozhodnutí mnoha jiných lidí. Doba běhu programu na reálném počítači zavisí zejména na tom, jak chytře je naprogramovaný, ale také na tom, co dělají ostatní programy, což ovlivňuje naplnění keší, atd.

Ponechme stranou fyzikálně-filosofickou otázku, zda něco může být opravdu náhodné, nebo zda je vesmír deterministický. (I v takovém případě by byla teorie pravděpodobnosti užitečná pro studium složitých systémů s chaotickým vývojem, viz příklady výše.) Místo toho přejděme k tomu, co pro pravděpodobnostní popis nějaké situace potřebujeme.

Elementární jevy V první řadě si musíme pořídit množinu Ω. Její prvky (budeme jim říkat elementární jevy) budou odpovídat jednotlivým výsledkům nějakého náhodného experimentu. Slovo experiment používáme hodně obecně, může zahrnovat jedno číslo, které padlo na hrací kostce, nebo celý průběh výpočtu v počítači, včetně všech stavů všech registrů a keší, nebo i (už bez nároku na konkrétní výpočet) celý stav vesmíru. Nicméně pro jeden hod kostkou budeme typicky používat $\Omega = [6] = \{1, 2, \ldots, 6\}$ – zajímá nás jen výsledek a ne třeba dráha letu kostky. Pro popis tři hodů kostkou bude $\Omega = [6]^3$, pro nekonečnou posloupnost hodů definujeme $\Omega = [6]^{\mathbb{N}}$ (množina všech zobrazení $\mathbb{N}$ do $[6]$, neboli všechny nekonečné poslouponosti čísel 1, 2, …, 6).

Pro popis počtu emailů můžeme brát $\Omega = \mathbb{N}_0$ (pokud nás opravdu zajímá jen jedno číslo v jednom dni), pro dobu běhu $\Omega = \mathbb{R}$. Při házení šipkou na terč bude Ω množina všech bodů terče, což můžeme popsat např. jako jednotkový kruh v rovině.

Prostor jevů Dále vybereme *prostor jevů (event space)* $\mathcal{F}$ jako podmnožinu potenční množiny $\mathscr{P}(\Omega)$. Jev je nějaká množina elementárních jevů, „něco, co nastalo“, u kterého budeme chtít měřit jejich pravděpodobnost. Často $\mathcal{F} = \mathscr{P}(\Omega)$, to je možné vždy, když Ω je spočetná a v takovém případě je to typická volba. Abychom mohli s měřenými jevy dobře pracovat, tak je potřeba, aby tvořily systém uzavřený na běžné operace. To popisuje následující definice.

Definice 2. *$\mathcal{F} \subseteq \mathscr{P}(\Omega)$ je prostor jevů (též σ-algebra), pokud*

1. *$\emptyset \in \mathcal{F}$ a $\Omega \in \mathcal{F}$,*
2. *$A \in \mathcal{F} \Rightarrow \Omega \setminus A \in \mathcal{F}$ (množinu $\Omega \setminus A$, tj. doplněk A, značíme zkráceně A^c)*

3. $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Z třetí podmínky speciálně plyne uzavřenost na sjednocení konečně mnoha množin: můžeme volit $A_{k+1} = A_{k+2} = \cdots = \emptyset$ a zjistíme, že pokud $A_1, \ldots, A_k \in \mathcal{F}$, tak i $A_1 \cup \cdots \cup A_k \in \mathcal{F}$. Kombinací s druhou podmínkou dostáváme také uzavřenost na průniky (konečné i nekonečné) a doplňky.

Množina $\mathscr{P}(\Omega)$ splňuje podmínky této definice. Nicméně pro $\Omega = \mathbb{R}$, nebo $\Omega = \{0,1\}^{\mathbb{N}}$ pro $\mathcal{F} = \mathscr{P}(\Omega)$ nedokážeme provést poslední krok popisu náhodných jevů, tj. definovat na $\mathcal{F}$ pravděpodobnost.

Axiomy pravděpodobnosti Stručně: co si představit pod pojmem pravděpodobnost? Pro jev $A \in \mathcal{F}$ číslo $P(A)$ určuje něco jako míru důvěry v to, že „nastane jev A“, neboli, že vybereme nějaký elementární jev, prvek $\omega \in \Omega$, takový, že $\omega \in A$. Pokud zkoumáme experiment, který lze opakovat, tak můžeme $P(A)$ chápat jako dlouhodobou úspěšnost, počet pokusů, kdy jsme vybrali prvek A. Pro jevy, které principiálně opakovat nelze, můžeme $P(A)$ chápat subjektivně, pomocí hraničního kurzu, se kterým jsme ještě ochotni si na A vsadit. Více se tím, co $P(A)$ vlastně znamená budeme zabývat v části o statistice. TODO: nebo už tady? TODO: příklad1 kostka TODO: příklad2 čaj Thomas Bayes

Definice 3. $P : \mathcal{F} \to [0,1]$ *se nazývá* pravděpodobnost (probability)*, pokud*

1. $P(\Omega) = 1$*, a dále*
2. $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$*, pro libovolnou (nekonečnou) posloupnost po dvou disjunktních jevů* $A_1, A_2, \ldots \in \mathcal{F}$.

Pozorování

- $P(\emptyset) = 0$: stačí v definici pravděpodobnosti vzít $A_i = \emptyset$ pro všechna i
- Druhá podmínka z Definice 3 platí i pro konečně mnoho množin. Jsou-li totiž $A_1, \ldots, A_n$ po dvou disjunktní, můžeme zvolit $A_{n+1} = A_{n+2} = \cdots = \emptyset$ a
$$P(\bigcup_{i=1}^{n} A_i) = P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^{n} P(A_i).$$

Definice 4. Pravděpodobnostní prostor (probability space) *je trojice* $(\Omega, \mathcal{F}, P)$ *taková, že*

- $\Omega \neq \emptyset$ *je libovolná množina elementárních jevů,*
- $\mathcal{F} \subseteq \mathscr{P}(\Omega)$ *je prostor jevů, a*
- P *je pravděpodobnost.*

Názvosloví

- „A je *jistý jev*“ znamená $P(A) = 1$. Také se říká, že A nastává *skoro jistě (almost surely)*, zkráceně s.j. (a.s.).
- „A je *nemožný jev*“ znamená $P(A) = 0$.
- *Šance (odds)* jevu A je $O(A) = \frac{P(A)}{P(A^c)}$. Např. šance na výhru je 1 ku 2 znamená, že pravděpodobnost výhry je 1/3; šance, že na kostce padne šestka je 1 ku 5.

K rozmyšlení 1. *Znamená $P(A) = 0$ totéž, jako $A = \emptyset$?*

Při všech výpočtech i teoretických úvahách budeme často používat následující vlastnosti pravděpodobnosti.

Věta 5. *V pravděpodobnostním prostoru $(\Omega, \mathcal{F}, P)$ platí pro $A, B \in \mathcal{F}$*

1. $P(A) + P(A^c) = 1 \qquad (A^c = \Omega \setminus A)$
2. $A \subseteq B \Rightarrow P(B \setminus A) = P(B) - P(A) \Rightarrow P(A) \leq P(B)$
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
4. $P(A_1 \cup A_2 \cup \dots) \leq \sum_i P(A_i)$ *(subaditivita, Booleova nerovnost)*

Důkaz.

1. Platí, že $\Omega = A \cup A^c$ a množiny A, A^c jsou disjunktní. Proto $P(\Omega) = P(A) + P(A^c)$. Dle definice je $P(\Omega) = 1$, proto jsme hotovi.
2. Obdobně jako v minulé části můžeme psát $B = A \cup (B \setminus A)$ a odsud plyne $P(B) = P(A) + P(B \setminus A) \geq P(A)$.
3. Označme $C_1 = A \setminus B$, $C_2 = A \cap B$ a $C_3 = B \setminus A$. Podle definice množinových operací platí $A \cup B = C_1 \cup C_2 \cup C_3$, $A = C_1 \cup C_2$ a $B = C_2 \cup C_3$. Navíc jsou množiny C_1, C_2, C_3 po dvou disjunktní. Z definice pravděpodobnosti tedy plyne, že $P(A \cup B) = P(C_1) + P(C_2) + P(C_3)$, zatímco $P(A) = P(C_1) + P(C_2)$ a $P(B) = P(C_2) + P(C_3)$. Odsud již snadno plyne, co potřebujeme.
4. Asi jste si už všimli, že všechny důkazy byly založeny na tom, že jsme hledali rozklad nějaké množiny na sjednocení po dvou disjunktních množin – to proto, abychom mohli použít druhou část definice pravděpodobnosti. I zde použijeme tento „trik zdisjunktnění“. Označme $B_1 = A_1$ a $B_i = A_i \setminus \bigcup_{j<i} A_j$ pro $i > 1$. Snadno si rozmyslíme, že
   - $B_i \subseteq A_i$ a tedy $P(B_i) \leq P(A_i)$ pro všechna i
   - $B_i \cap B_j = \emptyset$ pro $i \neq j$
   - $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i$

Odsud už snadno získáme, co potřebujeme:

$$P(\bigcup_{i=1}^{\infty} A_i) = P(\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} P(B_i) \leq \sum_{i=1}^{\infty} P(A_i).$$

□

Z diskrétní matematiky jistě znáte vzoreček podobný bodu 3 z předchozí věty: $|A \cup B| = |A| + |B| - |A \cap B|$. Za chvíli si ukážeme, že se jedná o speciální případ, kdy použijeme takzvaný klasický pravděpodobnostní prostor. Pravděpodobnostní variantu má i rozšíření na více množin, Princip inkluze a exkluze. Ten zde zmíníme jen bez důkazu (ten bude jako cvičení později, TODO odkaz).

Věta 6. *V pravděpodobnostním prostoru* $(\Omega, \mathcal{F}, P)$ *platí pro* $A_1, \ldots, A_n \in \mathcal{F}$ *vztah*

$$P(A_1 \cup \cdots \cup A_n) = \sum_{k=1}^{n} \Big((-1)^{k-1} \sum_{I \in \binom{[n]}{k}} P(\bigcap_{i \in I} A_i) \Big).$$

1.1 Příklady pravděpodobnostních prostorů

Klasický: neboli konečný s uniformní pravděpodobností: Ω je libovolná konečná množina, $\mathcal{F} = \mathscr{P}(\Omega)$, $P(A) = |A|/|\Omega|$. (To je ten klasický vzorec pro pravděpodobnost: počet dobrých možností děleno počtem všech.)

Představa: Ω je bedna s míčky, ty v množině A jsou červené. Ptáme se, jaká je pravděpodobnost, že vytáhneme červený míček.

Diskrétní: zobecnění předchozího případu. $\Omega = \{\omega_1, \omega_2, \ldots\}$ je libovolná nejvýše spočetná množina. Jsou dána $p_1, p_2, \ldots \in [0, 1]$ se součtem 1.
$P(A) = \sum_{i:\omega_i \in A} p_i$

Představa: Ω je opět bedna s míčky, ale každý míček má jinou šanci, že ho vytáhneme.

Varianta zápisu: funkci $p : \Omega \to \mathbb{R}$, pro kterou platí $p(\omega_i) = p_i$ nazveme pravděpodobnostní funkce pro příslušný pravděpodobnostní prostor. Pak můžeme psát přímočaře $P(A) = \sum_{a \in A} p(a)$.

Pokud Ω je konečná a $p(\omega) = 1/|\Omega|$ pro všechny $\omega \in \Omega$, dostáváme klasický pravděpodobnostní prostor.

Geometrický: „Hezká“ $\Omega \subseteq R^d$ pro $d \geq 1$. Nechť $V_d(A)$ označuje d-rozměrný objem množiny A a $\mathcal{F}$ obsahuje podmnožiny Ω, které mají tento objem definovaný. Definujeme $P(A) = V_d(A)/V_d(\Omega)$.

Představa pro $d = 2$: Ω je terč, do kterého střílíme hodně zdálky, takže všechny části zasahujeme stejně často. Pravděpodobnost, že nějakou množinu zasáhneme je přímo úměrná její velikosti: objemu či obsahu (podle dimenze).

Spojitý: Kombinace diskrétního a geometrického. Opět máme „hezkou“ $\Omega \subseteq R^d$ pro $d \geq 1$, navíc (taky hezkou) funkci $f : \Omega \to [0, \infty)$, která splňuje $\int_\Omega f = 1$. Definujeme $P(A) = \int_A f$.

Představa pro $d = 2$: Ω je terč, do kterého střílí kompetentní střelec, takže se častěji trefuje do středu. Funkce f popisuje, jak silný vliv to má.

Pokud $f(x) = 1/V_d(\Omega)$, dostáváme geometrický prostor jako výše.

Bernoulliho krychle – nekonečné posloupnosti Žádný z výše uvedených příkladů nezahrnuje prostor nekonečných posloupností: $\Omega = [6]^{\mathbb{N}}$ (nekonečně dlouhé házení kostkou), nebo pro informatiku přirozenější $\Omega = \{0, 1\}^{\mathbb{N}}$ (nekonečné házení mincí, tj. neomezený zdroj náhodných bitů. Ukážeme si obecně, jak pracovat s prostorem $\Omega = S^{\mathbb{N}}$, kde S je diskrétní postor s pravděpodobností Q.

Ani zde nemůžeme volit $\mathcal{F} = P(\Omega)$ a definovat pravděpodobnost splňující axiomy výše. Co nám typicky stačí je, aby pro množiny $A = A_1 \times \cdots \times A_k \times S \times S \times \cdots$ byla definována jejich pravděpodobnost a platilo

$$P(A) = Q(A_1) \cdots Q(A_k).$$

Intuice: jednotlivé souřadnice jsou nezávislé, můžeme klást požadavky na konečně mnoho z nich.

Dobrá zpráva je, že toto je možné: můžeme do množiny $\mathcal{F}$ dát všechny množiny tvaru $A_1 \times \cdots \times A_k \times S \times S \times \cdots$ (a jejich doplňky a disjunktní sjednocení), formulku pro P lze na všechny tyhle doplňky a disjunktní sjednocení rozšířit.

1.2 Nepříklady

Náhodné přirozené číslo můžeme vybrat mnoha způsoby. V přednášce poznáme geometrické a Poissonovo rozdělení. Nemůžeme ale požadovat, aby všechna přirozená čísla měla stejnou pravděpodobnost. (Proč?) „Náhodné přirozené číslo je sudé s pravděpodobností 1/2.“ To zní věrohodně, ale nemá jasný smysl. (A pro většinu voleb pravděpodobnosti na přirozených číslech to není pravda.)

Náhodné reálné číslo Opět není žádný preferovaný způsob, jak definovat pravděpodobnost pro $\Omega = \mathbb{R}$. Typicky bude každé reálné číslo mít pravděpodobnost 0! Navíc nejde definovat pravděpodobnost tak, aby nezáležela na posunu, tj. $P([0, 1]) = P([1, 2]) = \ldots$

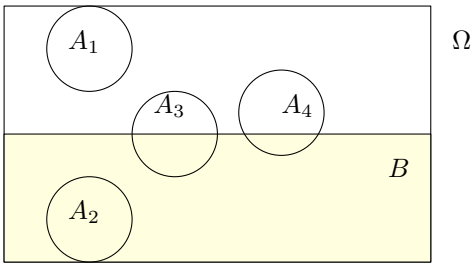
Náhodná tětiva kružnice – Bertrandův paradox Vybereme náhodnou tětivu zadané kružnice. Jaká je pravděpodobnost, že její délka je větší, než strana vepsaného rovnostranného trojúhelníku? To je známá úloha, jejíž paradoxnost je v tom, že není dobře zadaná: vybrat náhodnou tětivu je možno několika způsoby a pro každý vyjde pravděpodobnost jinak.

2 Podmíněná pravděpodobnost

Zatím jsme zavedli jakýsi jazyk pro mluvení o náhodných jevech, ale není jasné, jestli o nich můžeme říct něco pozoruhodného. Zajímavé to začne být až nyní, kdy začneme mluvit o podmiňování.

Definice 7. *Pokud $A, B \in \mathcal{F}$ a $P(B) > 0$, pak definujeme* podmíněnou pravděpodobnost *A při B* (probability of A given B) *jako*

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



K rozmyšlení 2. *Rozmyslete si, jaké jsou podmíněné pravděpodobnosti $P(A_i \mid B)$ a $P(B \mid A_i)$ pro $i = 1, 2, 3$ na obrázku.*

Věta 8. *V pravděpodobnostním prostoru $(\Omega, \mathcal{F}, P)$ pro pevné $B \in \mathcal{F}$ definujme $Q(A) := P(A \mid B)$ pro všechna $A \in \mathcal{F}$. Pak $(\Omega, \mathcal{F}, Q)$ je pravděpodobnostní prostor.*

Snadný důkaz vynecháváme. Důležité ale je si uvědomit, co věta znamená: podmiňování nějakým jevem znamená, že pracujeme v novém pravděpodobnostním prostoru. V něm chceme, aby pravděpodobnost B byla 1, musíme tedy všechny pravděpodobnosti „přeškálovat“ – vydělit $P(B)$.

TODO: co to udělá s klasickým a s diskrétním prostorem, co s geometrickým

Zřetězené podmiňování Přepíšeme si definici podmíněné pravděpodobnosti, $P(A \cap B) = P(B)P(A \mid B)$. Toto je tvar, který je často ten užitečný. Pro ilustraci, budiž A_1 jev „první karta v balíčku je srdcová“ a A_2 jev „druhá karta v balíčku je srdcová“ (obojí pro běžný balíček 32 karet). Jistě je $P(A_1) = 8/32$ (osm srdcových karet z celkem 32). Na $P(A_2 \mid A_1)$ se podíváme jako na výpočet v novém pravděpodobnostním prostoru, v tom, kde máme už jen 31 karet a mezi nimi jen 7 srdcových. Tudíž $P(A_2 \mid A_1) = 7/31$. Podle vzorce výše je $P(A_1 \cap A_2) = \frac{8}{32}\frac{7}{31}$.

** TODO: lépe

Zobecnění tohoto principu pro více množin udává následující věta.

Věta 9. *Pokud $A_1, \ldots, A_n \in \mathcal{F}$ a $P(A_1 \cap \cdots \cap A_n) > 0$, tak*

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = \\ P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2) \ldots P(A_n \mid \bigcap_{i=1}^{n-1} A_i)$$

Důkaz. Pokud $P(A_1 \cap \cdots \cap A_n) > 0$, tak i pro všechna k máme $P(A_1 \cap \cdots \cap A_k) > 0$ a tedy jsou všechny podmíněné pravděpodobnosti definované. Máme tedy na pravé straně výraz

$$P(A_1)\frac{P(A_1 \cap A_2)}{P(A_1)}\frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} \cdots \frac{P(A_1 \cap \cdots \cap A_n)}{P(A_1 \cap \cdots \cap A_{n-1})}$$

Snadno si všimneme, že v tomto výrazu se většina členů zkrátí, zbyde jen $P(A_1 \cap A_2 \cap \cdots \cap A_n)$, která nám zbýt má. (Formálně přesný důkaz pomocí matematické indukce si čtenář doplní v případě zájmu sám.) □

Všimněme si toho, že v této větě je na levé straně symetrický výraz – pokud seřadíme množiny $A_1, \ldots, A_n$ v jiném pořadí, tak se formule nezmění. Na pravé straně dostaneme tedy pro všech $n!$ pořadí stejný výsledek, ale pokaždé získaný jiným výpočtem. Máme tedy $n!$ vět za cenu jedné – můžeme si vybrat takové pořadí, které nám výpočet ulehčí.

K rozmyšlení 3. *O jevech A, B víme, že platí $A \Rightarrow B$, tedy pokud nastane A, tak nastane i B. Rozmyslete si, která z následujících tvrzení nutně platí také:*

1. $A \subseteq B$
2. $B \subseteq A$
3. $P(A \mid B) = 1$
4. $P(B \mid A) = 1$

Rozbor všech možností – věta o celkové pravděpodobnosti Připomeňme definici, kterou jste nejspíš potkali dříve, např. v diskrétní matematice.

Definice 10. *Spočetný systém množin $B_1, B_2, \ldots \in \mathcal{F}$ je* rozklad (partition) Ω, *pokud*

- $B_i \cap B_j = \emptyset$ *pro* $i \neq j$ *a*
- $\bigcup_i B_i = \Omega$.

Pokud rozklad vhodně zvolíme, tak pomocí něj rozdělíme množinu elementárních jevů do skupin, které budeme moci lépe zvládnout – uděláme jakýsi rozbor všech možností. V kombinatorice bychom tímto způsobem počítali velikost nějaké množiny A tak, že bychom spočítali velikosti průniků $A \cap B_i$. Následující věta dává analogii pro výpočet $P(A)$.

Věta 11 (Věta o celkové pravděpodobnosti)**.** *Pokud* $(\Omega, \mathcal{F}, P)$ *je pravděpodobnostní prostor,* $B_1, B_2, \ldots$ *je rozklad* Ω *a* $A \in \mathcal{F}$*, tak*

$$P(A) = \sum_i P(B_i)P(A \mid B_i)$$

(sčítance s $P(B_i) = 0$ *považujeme za 0).*

Důkaz. Protože množiny B_1, B_2, ... tvoří rozklad, můžeme A napsat jako disjunktní sjednocení

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \ldots$$

Podle definice pravděpodobnosti odsud máme

$$P(A) = \sum_i P(A \cap B_i)$$

Sčítance pro které je $P(B_i) > 0$ můžeme přepsat jako $P(B_i)P(A \mid B_i)$ (definice podmíněné pravděpodobnosti). Pokud pro nějaké i je $P(B_i) = 0$, tak je také $P(A \cap B_i) = 0$ a je tedy správně, že takový člen považujeme za nulu. □

Aplikace 1 – TODO

Aplikace – Gambler's ruin – zbankrotování hazardního hráče. Dáno $n \in \mathbb{N}$. Máme a korun, $0 \leq a \leq n$. Opakovaně hrajeme hru o 1 Kč se stejnou pravděpodobnosti výhry i prohry. Pokračujeme dokud nezbankrotujeme (nic nám nezbude), nebo nevyhrajeme (budeme mít n korun). Jaká je pravděpodobnost, že vyhrajeme?

Označme tuto pravděpodobnost jako $p(a)$. Jistě je $p(0) = 0$ a $p(n) = 1$. Co pro ostatní hodnoty a? Označíme V jev „vyhrajeme“ a PV je „poprvé vyhrajeme“. Podle pravidel je $P(PV) = 1/2$. Podle věty o celkové pravděpodobnosti je

$$P(V) = P(PV)P(V \mid PV) + P(PV^c)P(V \mid PV^c).$$

Všimneme si toho, že $P(V \mid PV)$ je vlastně $p(a+1)$ – jedná se o stejnou hru, ale začínáme z výhodnější pozice. Naopak $P(V \mid PV^c)$ je $p(a-1)$. Celkově tedy získáváme vztah

$$p(a) = \tfrac{1}{2}p(a+1) + \tfrac{1}{2}p(a-1).$$

Spolu s „okrajovými podmínkami“ pro $a \in \{0, n\}$ jsme našli $n+1$ rovnic pro $n+1$ neznámých. Pokud rovnici pro $a \in \{1, \ldots, n-1\}$ upravíme, dostaneme

$$p(a) - p(a-1) = p(a+1) - p(a).$$

Odsud již snadno získáme, že každý z těchto rozdílů je roven $1/n$ a tedy $p(a) = a/n$.

Bayesova věta

Věta 12. *Pokud $B_1, B_2, \ldots$ je rozklad Ω, $A \in \mathcal{F}$ a $P(A), P(B_j) > 0$, tak*

$$P(B_j \mid A) = \frac{P(B_j)P(A \mid B_j)}{\sum_i P(B_i)P(A \mid B_i)}$$

(sčítance s $P(B_i) = 0$ považujeme za 0).

Důkaz. Podle definice podmíněné pravděpodobnosti je $P(B_j \mid A) = \frac{P(B_j \cap A)}{P(A)} = \frac{P(B_j)P(A|B_j)}{P(A)}$. Ve jmenovateli nyní vyjádříme $P(A)$ podle věty o celkové pravděpodobnosti. □

U této věty je ale možná složitější než důkaz pochopení významu. Představme se, že modelujeme svět (nebo spíš nějakou jeho malou část) a množiny B_i popisují nějaké vzájemně se vylučující stavy, které ale nemůžeme přímo pozorovat. Víme ale, jaké mají pravděpodobnosti a jaký mají vliv na pravděpodobnost jevu A (neboli známe všechny podmíněné pravděpodobnosti $P(A \mid B_i)$). Jak se změní naše pravděpodobnosti různých stavů světa, pokud pozorujeme jev A? To je přesně to, co nám Bayesova věta říká. Ukažme si to na příkladu:

TODO

Nezávislost jevů

Definice 13. *Jevy $A, B \in \mathcal{F}$ jsou* nezávislé (independent)*, pokud $P(A \cap B) = P(A)P(B)$. Značíme $A \perp B$.*

Pokud $P(B) > 0$, můžeme podmínku vyjádřit ekvivalentně jako $P(A \mid B) = P(A)$ – pokud víme, že nastává B, tak to pravděpodobnost A neovlivní.

K rozmyšlení 4. *Hodíme dvakrát mincí. Označme A jev „poprvé padla panna“, B jev „podruhé padla panna“ a C jev „při každém ze dvou hodů padlo něco jiného.“ Rozhodněte, které ze tří dvojic uvažovaných jevů jsou nezávislé.*

Pojem nezávislosti nyní zobecníme na případ více jevů (případně i nekonečné mnoha).

Definice 14. *Buď I libovolná množina indexů. Jevy $\{A_i : i \in I\}$ jsou* (vzájemně) nezávislé*, pokud pro každou konečnou množinu $J \subseteq I$*

$$P(\bigcap_{j \in J} A_j) = \prod_{j \in J} P(A_j).$$

Pokud podmínka platí jen pro dvouprvkové množiny J, nazýváme jevy $\{A_i : i \in I\}$ po dvou nezávislé (pairwise independent).

3 Diskrétní náhodné veličiny

Často nás zajímá číslo dané výsledkem náhodného pokusu.

- Hodíme na terč a změříme vzdálenost od středu.
  Náhodný experiment je popsán místem dopadu (elementární jevy jsou body kruhu), ale nás zajímá jen ta vzdálenost.
- Házíme kostkou, dokud nepadne šestka, ale pak si všimneme jenom toho, kolik hodů to trvalo.
  Tady elementární jevy obsahují popis toho, co padlo, ale pak se ptáme jen na počet hodů.
- U quicksortu (algoritmus na třídění) měříme počet kroků (v závislosti na náhodných volbách pivotů).

První příklad necháme na později, ty další dva zobecníme následujícím způsobem.

Definice 15. *Mějme pravděpodobnostní prostor* $(\Omega, \mathcal{F}, P)$. *Funkci* $X : \Omega \to \mathbb{R}$ *nazveme* diskrétní náhodná veličina (discrete random variable)*, pokud* $Im(X)$ *(obor hodnot* X*) je spočetná množina a pokud pro všechna reálná* x *platí*

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

Množina na předchozí řádce lze zapsat i jinak, např. jako $X^{-1}(x)$. Nejčastěji ji ale budeme zapisovat stručně a přehledně jako $\{X = x\}$. Tak jako v této definici budeme i nadále náhodné veličiny značit velkými písmeny, zatímco jejich možné hodnoty obvykle odpovídajícími malými písmeny.

Představa náhodné veličiny: Ω je zase bedna s míčky, na každém z nich je teď napsané nějaké reálné číslo.

S diskrétní náhodnou veličinou máme přirozeně určený i nový pravděpodobnostní prostor: zapomeneme na to, jaký elementární jev nastal a budeme sledovat jen hodnotu funkce X. Neboli, místo $\omega \in \Omega$ budou teď elementární jevy hodnoty $X(\omega) \in Im(X) \subseteq \mathbb{R}$. Tento prostor, resp. na něm definovanou pravděpodobnost se nazývá *rozdělení náhodné veličiny* X, protože nám opravdu popisuje, jak jsou rozdělené hodnoty X – už ale bez vztahu k původnimu prostoru Ω. Pro popis tohoto prostoru budeme používat tzv. pravděpodobnostní funkci, která nám popisuje pravděpodobnosti jednotlivých bodů.

Definice 16. Pravděpodobnostní funkce (probability mass function, pmf) *diskrétní náhodné veličiny* X *je funkce* $p_X : \mathbb{R} \to [0, 1]$ *taková, že*

$$p_X(x) = P(\{X = x\})$$

Pravděpodobnost v předchozí definici budeme nadále stručně značit jen $P(X = x)$.

Pozorování 17. $\sum_{x \in Im(X)} p_X(x) = 1$

Pro důkaz stačí napsat množinu Ω jako disjunktní sjednocení množin $\{X = x\}$ pro všechna $x \in Im(X)$ (je jich jen spočetně mnoho!).

Pozorování 18. *Pro $S = \{s_i : i \in I\}$ spočetnou množinu reálných čísel a $c_i \in [0,1]$ splňující $\sum_{i\in I} c_i = 1$ existuje pravděpodobnostní prostor a diskrétní n.v. X na něm taková, že $p_X(s_i) = c_i$ pro $i \in I$.*

Můžeme například vybrat přímo diskrétní pravděpodobnostní prostor, kde $\Omega = S$ a pravděpodobnost bodu s_i je c_i. Pak můžeme vzít X jako identitu.

3.1 Příklady diskrétních rozdělení

Bernoulliho/alternativní rozdělení Typický příklad: X = počet šestek při jednom hodu kostkou. To, že X má Bernoulliho rozdělení s parametrem $p \in [0,1]$ značíme $X \sim Ber(p)$ (někdy $X \sim Alt(p)$). A jaká je příslušná pravděpodobnostní funkce?

- $p_X(1) = p$
- $p_X(0) = 1 - p$
- $p_X(x) = 0$ pro $x \neq 0, 1$

Můžeme ověřit Pozorování 17, neboli $p + (1 - p) = 1$.

Pro libovolný jev $A \in \mathcal{F}$ definujeme *indikátorovou n.v.* I_A jako pravdivostní hodnotu toho, že jev A nastal: $I_A(\omega) = 1$ pro $\omega \in A$ a $I_A(\omega) = 0$ jinak. Určitě platí $I_A \sim Bern(P(A))$. Pro mnoho úvah se nám budou takovéto náhodné veličiny hodit.

Geometrické rozdělení Tady bude typicky příklad X = kolikátým hodem kostkou padla první šestka (předpokládáme, že hody jsou nezávislé a házíme stále stejnou kostkou).

Značíme $X \sim Geom(p)$ (p bude značit „pravděpodobnost šestky“, neboli úspěchu, na který čekáme). Z nezávislosti jednotlivých pokusů odvodíme

$$p_X(k) = (1-p)^{k-1}p \qquad \text{pro } k = 1, 2, \ldots.$$

A jistě je $p_X(k) = 0$ jinak.

Ověříme identitu z Pozorování 17 (tím bychom mohli objevit chybu ve vzorci pro p_X). Podle vzorce pro součet geometrické řady platí

$$\sum_{k=1}^{\infty}(1-p)^{k-1}p = \frac{(1-p)^0 p}{1-(1-p)} = \frac{p}{p} = 1.$$

Varování Někdy se tomuto rozdělení říká posunuté geometrické, a za normální geometrické se považuje rozdělení $X - 1$, tj. počet neúspěšných hodů.

Binomické rozdělení Zde bude typický příklad X = počet šestek při n hodech kostkou. Obecněji, počet úspěchů při n nezávislých pokusech, z nichž každý má pravděpodobnost úspěchu p.

Značíme $X \sim Bin(n,p)$, kde $n \in \mathbb{N}$ a $p \in [0,1]$. Přímočará kombinatorika nám dává

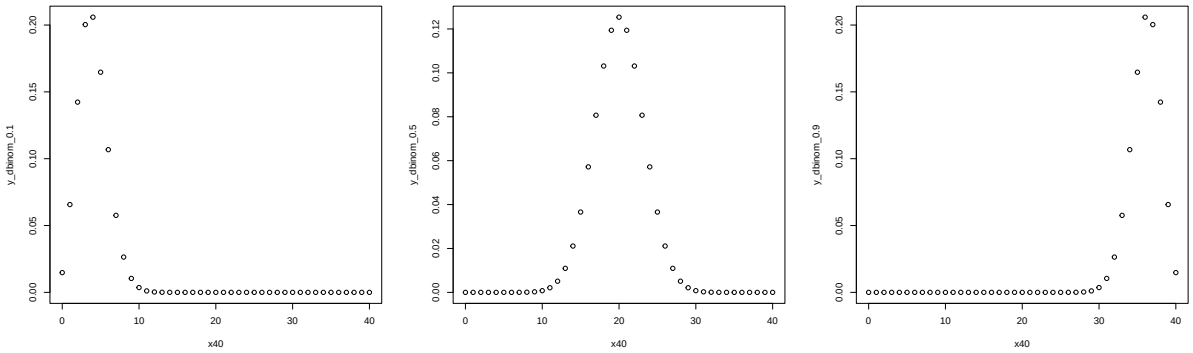
$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{pro } k \in \{0,1,\ldots,n\}$$

a jistě je $p_X(k) = 0$ jinak.

Pro kontrolu opět ověříme identitu z Pozorování 17: dle binomické věty platí

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = \big(p + (1-p)\big)^n$$

Pro ilustraci se podívejme na grafy této funkce. Ten pro $p = 1/2$ čtenář dobře zná ze studia binomických čísel, protože $p_{Bin(n,1/2)}(k) = \binom{n}{k} 2^{-n}$.



Vygenerováno následujícím kódem v R

```
x40 <- seq(0, 40, by = 1)
plot(x40,dbinom(x40, 40, 0.1))
plot(x40,dbinom(x40, 40, 0.5))
plot(x40,dbinom(x40, 40, 0.9))
```

Příklad 19. *V bedně je N míčků, z nich je K červených. Opakovaně vytáhneme míček (každý se stejnou pravděpodobnosti), podíváme se, zda je červený, a hodíme ho zpět. Označme X počet červených míčků po n opakováních. Jaká je distribuce náhodné veličiny X?*

Pravděpodobnost úspěchu při jednom tahu je K/N, je tedy $X \sim Bin(n, K/N)$.

Hypergeometrické rozdělení

Příklad 20. *V bedně je N míčků, z nich je K červených. Označme X počet červených míčků z n tažených míčků, kde* vytažené míčky nevracíme. *Jaká je distribuce náhodné veličiny X?*

Jedná se o takzvané hypergeometrické rozdělení, píšeme $X \sim Hyper(N, K, n)$. Jeho pravděpodobnostní funkce je

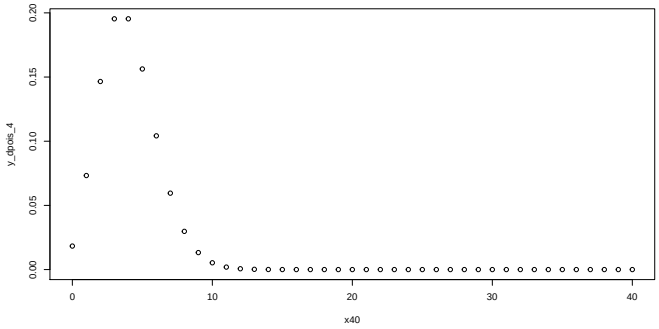
$$p_X(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, \qquad \text{pro } 0 \leq k \leq n.$$

Zdůvodnění je snadné: je potřeba vybrat množinu k červených míčků a $n-k$ nečervených. Pokud $n \ll N$, tak je snadné uvěřit (i ověřit), že vzorec pro p_X je přibližně stejný jako pro rozdělení binomické (je jedno, zda míčky vracíme, pokud jich celkem vytáhneme málo).

Poissonovo rozdělení Toto rozdělení popisuje například počet doručených emailů, dotazů na web server, atd. Zdůvodnění ale není tak přímočaré jako v předchozích případech. Popišme toto rozdělení napřed abstraktně. Značíme $X \sim Pois(\lambda)$, pro reálné $\lambda > 0$. Pro $k \in \mathbb{N}_0$ položíme

$$p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda},$$

pro jiná k je $p_X(k) = 0$. Abychom ověřili, že se jedná o pravděpodobnostní funkci, potřebujeme ověřit, že $\sum_{k=0}^{\infty} p_X(k) = 1$. To plyne přímo z Taylorova rozvoje exponenciální funkce, $e^{\lambda} = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}$.



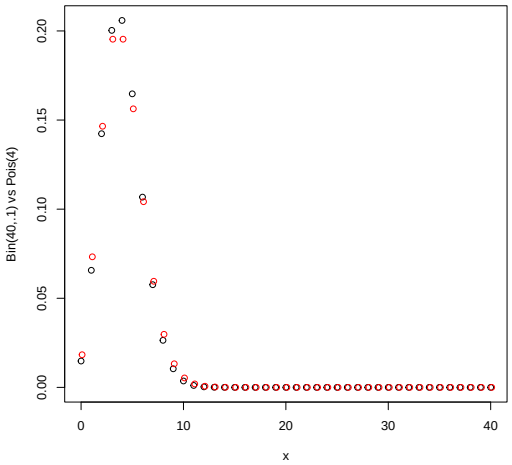
Vygenerováno následujícím kódem v R

```
x40 <- seq(0,40,by=1)
plot(x40,dpois(x40,4))
```

Pozorování 21. *$Pois(\lambda)$ je limitou $Bin(n, \lambda/n)$*

Důkaz. Co tím přesně myslíme: nechť $X_n \sim Bin(n, \lambda/n)$ a $X \sim Pois(\lambda)$. Zvolme pevné $k \in \mathbb{N}_0$. Pak

$$\lim_{n\to\infty} p_{X_n}(k) = p_X(k),$$



Srovnání binomického a Poissonova rozdělení: pravděpodobnostní funkce

neboli pro velká n můžeme binomické rozdělení aproximovat Poissonovým. Proč to platí? Podle definice,

$$\begin{aligned} p_{X_n}(k) &= \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n(n-1)\dots(n-k+1)}{k!}\frac{\lambda^k}{n^k}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k} \\ &= \frac{\lambda^k}{k!}\frac{n(n-1)\dots(n-k+1)}{n^k}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k}. \end{aligned}$$

Matematická analýza nám říká, že $(1-\lambda/n)^n \to e^{-\lambda}$, další dva členy zjevně konvergují k 1. □

Příklad 22. *Označme X počet emailů, které dostaneme za hodinu a λ typickou hodnotu X. Přepokládejme, že nám během té hodiny může email poslat libovolný z n našich přátel, nezávisle na sobě a každý se stejnou pravděpodobnosti. Navíc každý pošle nejvýše jeden email. Jaké je rozdělení X?*

Vygenerováno následujícím kódem v R

```
x = 0:40
bin = dbinom(x,40,0.1)
pois = dpois(x,4)
plot(x,bin, ylab="Bin(40,.1) vs Pois(4)")
points(x+.1,pois,col="red")
```

Jistě se jedná o binomické rozdělení $Bin(n,p)$. Zanedlouho (v části o střední hodnotě) zjistíme, že musí být $np=\lambda$, neboli $p=\lambda/n$. Pokud je n dost velké,

tak s ohledem na předchozí pozorování je rozdělení X přibližně o $Pois(\lambda/n)$. To může být pro počítání i pro porozumění této veličině praktičnější, než popis jako binomické rozdělení s velkým n a malou pravděpodobnosti. Tento příklad je poněkud uměly (předpoklad nezávislosti je hodně silný). Ale stejný závěr platí i za slabších předpokladů:

Poissonovo paradigma Nechť $A_1, \ldots, A_n$ jsou skoro-nezávislé[1] jevy, pro které platí $P(A_i) = p_i$ a $\sum_i p_i = \lambda$. Nechť n je velké, každé z p_i malé. Pak přibližně platí

$$\sum_{i=1}^{n} I_{A_i} \sim Pois(\lambda).$$

3.2 Střední hodnota

Představme si X jako výši výhry v jednom kole nějaké hazardní hry. Nechť $p_X(x_i) = p_i$ pro $i = 1, \ldots, k$. Teď tuto hru budeme hrát n-krát za sebou, a předpokádejme, že jednotlivá kola jsou nezávislá. [2] Kolik zhruba můžeme čekat, že získáme? Pokud n_i-krát vyhrajeme x_i, tak jsme celkem získali $\sum_{i=1}^{k} n_i x_i = n_1 x_1 + \cdots + n_k x_k$. Za jedno kolo to tedy průměrně je

$$\frac{1}{n} \sum_{i=1}^{k} n_i x_i = \sum_{i=1}^{k} \frac{n_i}{n} x_i.$$

Pro velká n je podíl n_i/n zhruba roven p_i (pravděpodobnost je dlouhodobá frekvence). Můžeme tedy očekávat, že průměrný zisk je $\sum_i p_i x_i$. Přesněji se na tuto situaci podíváme později (zákon velkých čísel), nyní nás to vede k následující obecné definici.

Definice 23. *Pokud X je diskrétní n.v., tak její* střední hodnota (expectation) *je označována $\mathbb{E}(X)$ a definována*

$$\mathbb{E}(X) = \sum_{x \in Im(X)} x \cdot P(X = x),$$

pokud řada konverguje absolutně, neboli pokud $\sum_{x \in Im(X)} |x| \cdot P(X = x) < \infty$. Lidsky: nechceme dostat výraz $\infty - \infty$. TODO: ale ∞ lze povolit!!!

Pokud by suma v definici $\mathbb{E}(X)$ obsahovala výraz typu $1 - 1 + 1 - 1 + \cdots$, tak $\mathbb{E}(X)$ nedefinujeme: Různým uzávorkováním tohoto výrazu můžeme dostat součet 0, 1, nebo (pokud povolíme i změnu pořadí) i jakékoli jiné celé číslo.

Všimněme si, že zde podstatně používáme toho, že X je diskrétní – jinak bychom nemohli sčítat pro všechna $x \in Im(X)$.

[1] přesnou definici zde dávat nebudeme

[2] Formálně bychom měli mluvit o posloupnosti n nezávislých náhodných veličin se stejným rozdělením. Ale k tomu nám ještě chybí názvosloví, proto zatím zůstaneme u intuitivního pohledu.

Za povšimnutí stojí i podobnost se vzorcem pro výpočet těžiště. Pokud máme na tyči k hmotných bodů, kde i-tý z nich má hmotnost m_i a souřadnici – polohu na tyči x_i, tak jejich těžiště má souřadnici $\sum_{i=1}^{k} \frac{m_i}{m} x_i$.

Pozorování 24. *Pokud je X diskrétní náhodná veličina na diskrétním pravděpodobnostním prostoru $(\Omega, \mathcal{F}, P)$ a $\mathbb{E}(X)$ je definována, tak platí*

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}).$$

Význam: můžeme počítat průměr přes všechny možné výsledky pokusu (pokud počet výsledků je spočetný). Tento vzorec je velmi přirozený, ale nelze použít, když je množina Ω moc velká. Nebo sdružíme dohromady všechny výsledky, které dají stejnou hodnotu X, tím dostaneme definici $\mathbb{E}(X)$ a obecnější vzorec. Toto lze zapsat i formálně a získat důkaz tohoto pozorovani:

Důkaz. Upravujme výraz na pravé straně jak bylo popsáno výše:

$$\begin{aligned}
\sum_{\omega \in \Omega} X(\omega) P(\{\omega\}) &= \sum_{x \in Im(X)} \sum_{\omega \in X^{-1}(x)} X(\omega) P(\{\omega\}) \\
&= \sum_{x \in Im(X)} \Big(x \sum_{\omega \in X^{-1}(x)} P(\{\omega\}) \Big) \\
&= \sum_{x \in Im(X)} \Big(x \cdot P(X^{-1}(x)) \Big)
\end{aligned}$$

A to je již definice $\mathbb{E}(X)$, kde $P(X^{-1}(x))$ je samozřejmě totéž jako $P(X = x)$. □

Pravidlo naivního statistika PNS Často budeme uvažovat nějakou funkci náhodné veličiny – pragmaticky vzato, výstup funkce random.randint() (nebo nějaké podobné) pošleme na vstup jiné funkci. Nebo hodíme „zobecněnou kostkou“ a výsledek upravíme nějakou funkcí. Formalizovat to budeme pomocí skládání funkcí:

Pozorování 25. *Pro reálnou funkci g a diskrétní n.v. X je $Y = g(X)$ také diskrétní n.v.*

Důkaz. Zápisem $g(X)$ myslíme náhodnou veličinu, která při výsledku náhodného experimentu $\omega \in \Omega$ řekne $g(X(\omega))$, neboli Y je složená funkce $g \circ X$. Počet hodnot, kterých Y nabývá je jistě nejvýše takový, jako pro X, tedy spočetný. Zbývá ověřit podmínku, že $Y^{-1}(y) = \{\omega \in \Omega : g(X(\omega)) = y\} \in \mathcal{F}$ pro všechna $y \in Im(Y)$. Tato množina je rovna

$$\bigcup_{x \in Im(X) : g(x) = y} X^{-1}(x). \tag{1}$$

(Omezení na $x \in Im(X)$ můžeme udělat, protože jinak by $X^{-1}(x)$ byla prázdná.) Zde musíme dát pozor: sice každá množina $X^{-1}(x)$ je v $\mathcal{F}$, ale máme povoleno sjednocovat jen spočetně mnoho množin. Naštěstí dle předpokladů je $Im(X)$ spočetná množina. □

Věta 26 (Pravidlo naivního statistika)**.** *Pokud X je diskrétní n.v. a g reálná funkce, tak*

$$\mathbb{E}(g(X)) = \sum_{x \in Im(X)} g(x)P(X = x)$$

pokud součet má smysl (přesněji, pokud řada konverguje absolutně).

(Smysl názvu věty (v angl. originále Law of the unconscious statistician) je, že na první pohled se mnoho lidí domnívá, že není co dokazovat, že se jedná o definici.)

Důkaz. Pišme opět $Y = g(X)$. Podle definice je $\mathbb{E}(Y) = \sum_{y \in Im(Y)} yP(Y = y)$. Pro množinu $\{Y = y\} = Y^{-1}(y)$ použijeme opět rovnici (1) a použijeme druhý axiom pravděpodobnosti:

$$P(Y^{-1}(y)) = \sum_{x \in Im(X), g(x)=y} P(X^{-1}(x)) \qquad \text{neboli}$$
$$P(Y = y) = \sum_{x \in Im(X), g(x)=y} P(X = x)$$

Zbytek je už snadné dosazení do vzorce. □

Věta 27 (Vlastnosti $\mathbb{E}$)**.** *Nechť X, Y jsou diskrétní n.v. a $a, b \in \mathbb{R}$.*

1. *Pokud $P(X \geq 0) = 1$, tak také $\mathbb{E}(X) \geq 0$. Pokud navíc $\mathbb{E}(X) = 0$, tak $P(X = 0) = 1$.*
2. *Pokud $\mathbb{E}(X) \geq 0$ tak $P(X \geq 0) > 0$.*
3. $\mathbb{E}(a \cdot X + b) = a \cdot \mathbb{E}(X) + b$.
4. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$.

Vlastnosti 3 a 4 se nazývají **linearita střední hodnoty.** Čtenář jistě snadno nahlédne, že z vlastnosti 4 plyne, že i střední hodnota součtu n proměnných je rovna součtu středních hodnot.

Důkaz. 1. Podle definice je

$$\mathbb{E}(X) = \sum_{x \in Im(X)} xP(X = x).$$

Podle předpokladů jsou v této sumě všechny členy nezáporné (pokud je $x < 0$, tak $P(X = x) = 0$), tedy je celá suma nezáporná. Pokud je součet nezáporných

členů nula, tak musí být všechny nulové. To ale znamená, že pro $x > 0$ je taky $P(X = x) = 0$.

2. Znovu vyjdeme přímo z definice. Kdyby neplatil závěr, tak pro všechna $x \geq 0$ je $P(X = x) = 0$, a tedy všechny nenulové členy v sumě jsou záporné.

3. Použijeme Větu 26 pro funkci $g(x) = ax + b$. Podle ní je

$$\begin{aligned}\mathbb{E}(aX + b) = \mathbb{E}(g(X)) &= \sum_{x \in Im(X)} (ax + b)P(X = x) \\ &= a \sum_{x \in Im(X)} xP(X = x) + b \sum_{x \in Im(X)} P(X = x) \\ &= a \cdot \mathbb{E}(X) + b \cdot 1\end{aligned}$$

4. Zatím jen pro případ diskrétního prostoru. Tam je vše průzračně jasné, když použijeme Pozorování 24:

$$\begin{aligned}\mathbb{E}(X + Y) &= \sum_{\omega \in \Omega} \big(X(\omega) + Y(\omega)\big)P(\{\omega\}) \\ &= \sum_{\omega \in \Omega} X(\omega)P(\{\omega\}) + \sum_{\omega \in \Omega} Y(\omega)P(\{\omega\}) \\ &= \mathbb{E}(X) + \mathbb{E}(Y)\end{aligned}$$

□

Příklad 28. *Hodíme kostkou a při šestce hodíme podruhé (potřetí už ne). Číslo, které padlo (nebo součet dvou čísel) budiž náhodná veličina X. Pak zkusíme totéž, ale házíme znovu při jedničce – příslušnou náhodnou veličinu označíme Y. Rozhodněte, zda je větší $\mathbb{E}(X)$ nebo $\mathbb{E}(Y)$, případně zda jsou stejné.*

Označme K_1 číslo, které padlo na první kostce. Číslo, které padlo na druhé kostce označíme K_2, ale jen, pokud se počítá (tj. pokud $K_1 = 6$), jinak bude $K_2 = 0$. Při takovém značení je $X = K_1 + K_2$ a podle Věty 27 je $\mathbb{E}(X) = \mathbb{E}(K_1) + \mathbb{E}(K_2)$. Podle definice je $\mathbb{E}(K_1) = (1 + \cdots + 6)/6 = 3.5$. Opět podle definice je $\mathbb{E}(K_2) = 0 \cdot P(K_2 = 0) + 1 \cdot P(K_2 = 1) + \cdots + 6 \cdot P(K_2 = 6)$. Přitom $P(K_2 = 0) = P(K_1 \neq 6) = 5/6$ a $P(K_2 = k) = 1/6^2$ pro $k = 1, \ldots, 6$ (musí být $K_1 = 6$ a na druhé kostce padnout správné číslo). Odsud $\mathbb{E}(K_2) = (1 + \cdots + 6)/6^2 = 7/12$ a tedy $\mathbb{E}(X) = 7/2 + 7/12 \doteq 4.08$.

Pro výpočet $\mathbb{E}(Y)$ musíme K_2 nahradit proměnnou K_2' – bude definovaná obdobně, jenom podmínku $K_1 = 6$ nahradíme $K_1 = 1$. Když se podíváme na předchozí odstavec, tak zjistíme, že rozdělení K_2' je stejné jako K_2 – jedná se o jiné náhodné veličiny, ale pro každé k je $P(K_2 = k) = P(K_2' = k)$. Mají tedy i stejnou střední hodnotu. Proto bez dalšího počítání je $\mathbb{E}(X) = \mathbb{E}(Y)$, pro 63 % účastníků přednášky nečekaný výsledek!

Podmíněná střední hodnota Je to vlastně střední hodnota náhodné veličiny zredukované na menší pravděpodobnostní prostor.

Definice 29. *Pokud X je diskrétní n.v. a $P(B) > 0$, tak* podmíněná střední hodnota X za předpokladu B (conditional expectation of X given B) *je*

$$\mathbb{E}(X \mid B) = \sum_{x \in Im(X)} x \cdot P(X = x \mid B),$$

pokud součet má smysl.

Věta 30 (Věta o celkové střední hodnotě)**.** *Pokud $B_1, B_2, \ldots$ je rozklad Ω a X diskrétní náhodná veličina. Pak platí*

$$\mathbb{E}(X) = \sum_i P(B_i) \cdot \mathbb{E}(X \mid B_i),$$

kdykoliv má součet smysl. (Sčítance s $P(B_i) = 0$ považujeme za 0.)

Důkaz. Vyjdeme z definice $\mathbb{E}(X)$, použijeme Větu 11 a nakonec prohodíme pořadí sčítání:

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{x \in Im(X)} x \cdot P(X = x) \\
&= \sum_{x \in Im(X)} x \cdot \sum_i P(B_i) \cdot P(X = x \mid B_i) \\
&= \sum_i P(B_i) \sum_{x \in Im(X)} x \cdot P(X = x \mid B_i) \\
&= \sum_i P(B_i) \mathbb{E}(X \mid B_i)
\end{aligned}$$

□

Příklad 31. *Nechť X je počet hodů, které musíme hodit, než nám padne šestka (včetně toho posledního). Určete $\mathbb{E}(X)$.*

Označme PS jev, že padla šestka hned prvním hodem. Použijeme Větu 30 pro $B_1 = PS$ a $B_2 = PS^c$. Určitě je $\mathbb{E}(X \mid PS) = 1$, protože $P(X = 1 \mid PS) = 1$. Dále $\mathbb{E}(X \mid PS^c) = 1 + \mathbb{E}(X)$ – po prvním neúspěchu jsme ve stejné situaci, jako na začátku, ale máme o jeden hod navíc. (Formálně bychom mohli psát $P(X = 1 + k \mid PS^c) = P(X = k)$.) Odsud pomocí věty

$$\begin{aligned}
\mathbb{E}(X) &= P(PS)\mathbb{E}(X \mid PS) + P(PS^c)\mathbb{E}(X \mid PS^c) \\
\mathbb{E}(X) &= \frac{1}{6} \cdot 1 + \frac{5}{6} \cdot (1 + \mathbb{E}(X)) \\
(1 - \frac{5}{6}) \cdot \mathbb{E}(X) &= 1
\end{aligned}$$

odsud již $\mathbb{E}(X) = 6$.

Alternativní postup: víme, že $X \sim Geom(1/6)$ a střední hodnotu geometrického rozdělení jsme už spočítali jinak, tj. nemuseli jsme počítat nic. Také jsme mohli naopak použít tento postup pro výpočet střední hodnoty obecného geometrického rozdělení. TODO: poznámka o existenci střední hodnoty. Jenom bychom měli pro úplnost doplnit ještě jednu věc – jakou?

Věta 32 (Alternativní formulka pro střední hodnotu)**.** *Nechť X je diskrétní n.v. nabývající jen hodnot z $\mathbb{N}_0 = \{0, 1, 2, \dots\}$. Pak platí*

$$\mathbb{E}(X) = \sum_{n=0}^{\infty} P(X > n).$$

Důkaz. Nejprve si rozmysleme, že platí následující vyjádření n.v. X pomocí indikátorových veličin:

$$X = \sum_{n=0}^{\infty} I_{X>n}.$$

Pokud totiž $X(\omega) = k$, tak je $I_{X>n}(\omega) = 1$ přesně pro $n = 0, 1, \dots, k-1$, tedy pro k hodnot. Nyní stačí použít větu o linearitě střední hodnoty. TODO: proč platí i pro nekonečné součty □

Důkaz. Vyjdeme z definice $\mathbb{E}(X)$, místo k napíšeme součet k jedniček a nakonec prohodíme pořadí sčítání:

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=0}^{\infty} k \cdot P(X = k) \\
&= \sum_{k=0}^{\infty} \sum_{0 \le n < k} P(X = k) \\
&= \sum_{n=0}^{\infty} \sum_{k > n} P(X = k) \\
&= \sum_{n=0}^{\infty} P(X > n)
\end{aligned}$$

□

Rozptyl

Definice 33. Rozptyl (variance) *n.v. X nazveme číslo* $\mathbb{E}((X - \mathbb{E}X)^2)$.

Značíme jej $\mathrm{var}(X)$. *Dále definujeme* směrodatnou odchylku (standard deviation) σ_X *jako* $\sqrt{\mathrm{var}(X)}$.

Někdy se také používá variační koeficient $CV_X = \sigma_X / \mathbb{E}(X)$ *(definujeme jen pro* $\mathbb{E}(X) > 0$*).*

Podle části 1 Věty 27, aplikované na $(X - \mathbb{E}(X))^2$, je $\text{var}(X) \geq 0$ a tedy σ_X je dobře definováno. Ze stejné části vyplývá i to, že $\text{var}(X) = 0$ jen pro funkce, které jsou s.j. konstantní $P(X = \mathbb{E}(X)) = 1$.

Jak rozptyl tak směrodatná odchylka měří to, jak moc X kolísá od své střední hodnoty. S rozptylem se lépe počítá, nicméně směrodatná odchylka má „stejné jednotky jako X“ (pokud je X v metrech, tak σ_X také), a to vede často k logičtějším vzorcům. Toto kolísání bychom mohli měřit i jinak: takzvaný k-tý centrální moment X je definován jako $\mathbb{E}(|X - \mathbb{E}(X)|^k)$. Nicméně rozptyl (případ $k = 2$) má nejhezčí vlastnosti, počínaje následující větou.

Věta 34. *Pro diskrétní náhodnou veličinu X platí*

$$\text{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mathbb{E}(X(X-1)) + \mathbb{E}(X) - \mathbb{E}(X)^2.$$

Důkaz. Pro přehlednost označíme $\mathbb{E}(X)$ písmenem μ.

$$\text{var}(X) = \mathbb{E}((X-\mu)^2) = \mathbb{E}(X^2 - 2\mu X + \mu^2) = \mathbb{E}(X^2) - 2\mu\mathbb{E}(X) + \mu^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

Tím jsme dokázali první část. Druhá odsud plyne snadno díky linearitě střední hodnoty. □

Věta 35. *Nechť a, b jsou reálná čísla a X diskrétní náhodná veličina. Pak platí*

$$\text{var}(aX + b) = a^2 \,\text{var}(X).$$

Důkaz. Použijeme přímo definici. Napřed upravíme její „vnitřek“. Pro $Y = aX + b$ je $Y - \mathbb{E}(Y) = (aX + b) - \mathbb{E}(aX + b) = (aX + b) - (a\mathbb{E}(X) + b) = a(X - \mathbb{E}(X))$. Proto je

$$\mathbb{E}((Y - \mathbb{E}(Y))^2) = \mathbb{E}(a^2(X - \mathbb{E}(X))^2) = a^2\mathbb{E}((X - \mathbb{E}(X))^2).$$

□

3.3 Vlastnosti konkrétních rozdělení

Bernoulliho Pro $X \sim Ber(p)$ je

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

Všimněme si, že $X^2 = X$, protože X nabývá jen hodnot 0, 1.

Geometrické Pro $X \sim Geom(p)$ je $\mathbb{E}(X) = 1/p$ a $\text{var}(X) = \frac{1-p}{p^2}$. Zdůvodníme si jen střední hodnotu, zato několika způsoby:

První zdůvodnění: použijeme Větu 32 (a součet geometrické řady):

$$\mathbb{E}(X) = \sum_{n=0}^{\infty} P(X > n) = \sum_{n=0}^{\infty} (1-p)^n = \frac{1}{1-(1-p)}.$$

Druhé zdůvodnění: použijeme definici $\mathbb{E}(X)$ (a pak si musíme poradit se vzniklou sumou).

$$\mathbb{E}(X) = \sum_{n=1}^{\infty} n \cdot P(X = n) = \sum_{n=1}^{\infty} n \cdot (1-p)^{n-1} p = \frac{p}{(1-(1-p))^2} = \frac{1}{p}.$$

Na místě pár teček může být postup analogický důkazu Věty 32, nebo některý z postupů, které znáte z kombinatoriky (třeba pomocí vytvořujících funkcí).

Třetí zdůvodnění: viz Příklad 31.

Čtvrté zdůvodnění: jen naznačíme, ale z jistého pohledu se jedná o to nejvýstižnější vysvětlení. Házíme kostkou tak dlouho, než dostaneme n šestek, označíme X celkový počet hodů. Co o X víme? Na jednu stranu je to součet n geometrických rozdělení, takže $\mathbb{E}(X)$ by měla být n-násobek střední hodnoty $Geom(1/6)$. Na druhou stranu, z X hodů by mělo padnout přibližně $X/6$ šestek, neboli $X/6 \approx n$. Symbol $\approx$ zde nebudeme upřesňovat – ale pokud budeme pro teď věřit, že X bude většinou blízko své střední hodnotě, tak by mělo platit $\mathbb{E}(X) \approx 6n$, a tedy $\mathbb{E}(Geom(1/6)) = 6$. Pro pořádnější zápis budeme potřebovat napřed pochopit nezávislost náhodných veličin a také tzv. zákony velkých čísel.

Binomické Pokud $X \sim Bin(n, p)$ označuje počet úspěchů při n hodech kostkou tak pišme $X = \sum_{i=1}^{n} X_i$, kde X_i je indikátorová náhodná veličina i-tého pokusu. Podle linearity je $\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{E}(X_i) = np$. Později si uvedeme větu pro rozptyl součtu, z něj bude plynou, že $\mathrm{var}(X) = \sum_{i=1}^{n} \mathrm{var}(X_i) = np(1-p)$.

Nabízí se otázka, zda můžeme každou binomickou veličinu napsat jako součet n veličin Bernoulliho. Odpověď je, že to nepotřebujeme! Jak $\mathbb{E}(X)$ tak $\mathrm{var}(X)$ závisí jen na rozdělení X, tj. na pravděpodobnostní funkci p_X. Můžeme si tedy vybrat takovou veličinu s tímto rozdělením, o které se nám lépe uvažuje!

Alternativní postup je použít přímo definici:

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=0}^{n} k \cdot P(X = k) \\
&= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} n \binom{n-1}{k-1} p \cdot p^{k-1} (1-p)^{(n-1)-(k-1)} \\
&= np(p + (1-p))^n = np
\end{aligned}$$

Obdobně zjistíme, že $\mathbb{E}(X(X-1)) = n(n-1)p^2$ a použitím Věty 34 získáme opět $\mathrm{var}(X) = np(1-p)$.

Hypergeometrické Pro $X \sim Hyper(N, K, n)$

- $\mathbb{E}(X) = n\frac{K}{N}$

- $\mathrm{var}(X) = n\frac{K}{N}(1-\frac{K}{N})\frac{N-n}{N-1}$

Ukážeme si dva způsoby výpočtu $\mathbb{E}(X)$.

První postup: $X = \sum_{i=1}^{n} X_i$, kde X_i je 1, pokud i-tý míček byl červený, a 0 jinak. Jistě je $X_1 \sim Ber(K/N)$ a tedy $\mathbb{E}(X_1) = K/N$. Ale při pozornějším pohledu zjistíme, že všechny veličiny $X_1, \ldots, X_n$ mají stejné rozdělení. (Pozor, to neznamená, že by byly nezávislé (což už brzy definujeme), ale jen, že $P(X_i = 1) = K/N$ platí pro všechna i.) Nejlepší způsob, jak to vidět, je když si představíme, že nám první a i-tý míček někdo prohodí. Tím se zamění hodnoty X_1 a X_i. Ale pokud toto prohození nastane vždy, tak se tím zjevně náhodnost procesu losování nepokazí.

Druhý postup: $X = \sum_{j=1}^{K} Y_j$, kde $Y_j = 1$, pokud jsme vytáhli j-tý míček (lhostejno, kolikátým tahem), a $Y_j = 0$ jinak. Tentokrát je $Y_j \sim Bin(n/N)$: počet všech možných množin tažených míčků je $\binom{N}{n}$, počet těch, které obsahují j-tý míček, je $\binom{N-1}{n-1}$. A elementární výpočet dává, že $\binom{N-1}{n-1}/\binom{N}{n} = n/N$. Proto

$$\mathbb{E}(Y_j) = P(Y_j = 1) = n/N$$

a linearita střední hodnoty opět dává $\mathbb{E}(X) = nK/N$.

Poissonovo Pro $X \sim Pois(\lambda)$ je

- $\mathbb{E}(X) = \lambda$
- $\mathrm{var}(X) = \lambda$

Přímo podle definice spočítáme

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=0}^{\infty} k \cdot P(X = k) \\
&= \sum_{k=0}^{\infty} k \cdot e^{-\lambda}\frac{\lambda^k}{k!} \\
&= \sum_{k=1}^{\infty} k \cdot e^{-\lambda}\frac{\lambda^k}{k!} \\
&= \lambda \sum_{k=1}^{\infty} e^{-\lambda}\frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} \\
&= \lambda \cdot 1 = \lambda.
\end{aligned}$$

Stejně jako u jiných rozdělení bychom obdobně mohli spočítat $\mathbb{E}(X(X-1))$ a použít Větu 32; detaily vynecháme.

4 Náhodné vektory

Nechť X a Y jsou (diskrétní) náhodné veličiny na stejném pravděpodobnostním prostoru $(\Omega, \mathcal{F}, P)$. Budeme chtít uvažovat (X, Y) jako jeden objekt – náhodný vektor. Jak to udělat? Příklad: házíme dvakrát čtyřstěnnou kostkou, X = první hod, Y = druhý hod.

Sdružené rozdělení

Definice 36. *Pro diskrétní n.v. X, Y na pravděpodobnostním prostoru $(\Omega, \mathcal{F}, P)$ definujeme jejich* sdruženou pravděpodobnostní funkci (joint pmf) $p_{X,Y} : \mathbb{R}^2 \to [0, 1]$ *předpisem*

$$p_{X,Y}(x, y) = P(\{\omega \in \Omega : X(\omega) = x \,\&\, Y(\omega) = y\}).$$

Tedy (X, Y) tvoří náhodný vektor, pokud jsou všechny tyto pravděpodobnosti definované, neboli pro každé $x, y \in \mathbb{R}$ platí $\{X = x \,\&\, Y = y\} \in \mathcal{F}$.

- Mohli bychom definovat i pro více než dvě n.v. $p_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$.

x \ y	1	2	3	4	p_X
1	1/16	1/16	1/16	1/16	1/4
2	1/16	1/16	1/16	1/16	1/4
3	1/16	1/16	1/16	1/16	1/4
4	1/16	1/16	1/16	1/16	1/4
p_Y	1/4	1/4	1/4	1/4	1

x \ y	1	2	3	4	p_X
1	1/4	0	0	0	1/4
2	0	1/4	0	0	1/4
3	0	0	1/4	0	1/4
4	0	0	0	1/4	1/4
p_Y	1/4	1/4	1/4	1/4	1

Marginální rozdělení Máme-li dáno $p_{X,Y}$, jak zjistit rozdělení jednotlivých složek, tj. p_X a p_Y. Jde to vůbec jednoznačně určit – a jde naopak určit $p_{X,Y}$ z p_X a p_Y?

Věta 37. *Nechť X, Y jsou diskrétní n.v. Pak*

$$p_X(x) = P(X = x) = \sum_{y \in Im(Y)} P(X = x \,\&\, Y = y) = \sum_{y \in Im(Y)} p_{X,Y}(x, y)$$

$$p_Y(y) = P(Y = y) = \sum_{x \in Im(X)} P(X = x \,\&\, Y = y) = \sum_{x \in Im(X)} p_{X,Y}(x, y)$$

Důkaz. Jev $\{X = x\}$ je disjunktní sjednocení jevů $\{X = x \,\&\, Y = y\}$, takže formule na prvním řádku plyne přímo z definice pravděpodobnosti. Druhý řádek analogicky. □

K rozmyšlení 5. *Jak by vypadala analogie pro vektory z více složkami? Např. jak zjistit p_X z $p_{X,Y,Z}$?*

Nezávislost náhodných veličin

Definice 38. *Diskrétní n.v. X, Y jsou* nezávislé (independent) *pokud pro každé $x, y \in \mathbb{R}$ jsou jevy $\{X = x\}$ a $\{Y = y\}$ nezávislé. To nastane, právě když*

$$P(X = x, Y = y) = P(X = x)P(Y = y).$$

K rozmyšlení 6. *Pro nezávislé n.v. platí opak Věty 37: z funkcí p_X a p_Y můžeme určit $p_{X,Y}$.*

Funkce náhodného vektoru Máme-li dáno $p_{X,Y}$, měli bychom být schopni určit pravděpodobnosti všech jevů, které závisí jen na hodnotách X a Y. Nechť g je nějaká funkce dvou proměnných a $Z = g(X, Y)$. Jak zjistit rozdělení Z, tj. funkci p_Z?

Věta 39. *Mějme náhodný vektor (X, Y) a funkci $g : \mathbb{R}^2 \to \mathbb{R}$. Pak $Z = g(X, Y)$ je n.v. na $(\Omega, \mathcal{F}, P)$ a*

$$p_Z(z) = P(Z = z) = \sum_{x \in Im(X), y \in Im(Y) : g(x,y) = z} P(X = x \,\&\, Y = y).$$

(Index při sčítání znamená, že sčítáme jen pro takové dvojice (x, y), pro které je $g(x, y) = z$.)

Důkaz. Stačí si uvědomit, že jev $\{Z = z\}$ je disjunktním sjednocením jevů $\{X = x \,\&\, Y = y\}$ pro dvojice, pro které $g(x, y) = z$. Jako v přechozích podobných důkazech, stačí se omezit na $x \in Im(X)$ a $y \in Im(Y)$, takže jde o nejvýše spočetné sjednocení a můžeme použít definici pravděpodobnosti. □

Speciální případ stojí za zvláštní povšimnutí:

Věta 40 (Konvoluční vzorec)**.** *Pokud X, Y jsou diskrétní náhodné veličiny (zkráceně d.n.v.), tak jejich součet $Z = X + Y$ má pravděpodobnostní funkci*

$$P(Z = z) = \sum_{x \in Im X} P(X = x \,\&\, Y = z - x).$$

Pokud jsou X a Y navíc nezávislé, tak dostáváme vyjádření pravděpodobnostní funkce p_Z jako konvoluce *funkcí p_X a p_Y:*

$$P(Z = z) = \sum_{x \in Im X} P(X = x)P(Y = z - x).$$

Důkaz. Stačí použít předchozí větu pro funkci danou předpisem $g(x,y) = x + y$. □

Příklad 41. *Nechť X a Y jsou dvě n.n.v. s uniformním rozdělením na $\{1,2,\ldots,6\}$ – neboli dva nezávislé hody hrací kostkou. Určete rozdělení $Z = X + Y$.*

$$P(X+Y=k) = \sum_{x=1}^{6} P(X=x)P(Y=k-x)$$

Každý člen v sumě je buď 1/36 nebo 0 (to druhé pokud není $1 \leq k-x \leq 6$). Snadno spočítáme, kolik takových x je a zjistíme, že p_Z je dáno následující tabulkou:

k	2	3	4	5	6	7	8	9	10	11	12
$p_Z(k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Příklad 42. *Nechť $X \sim Bin(m,p)$ a $Y \sim Bin(n,p)$, nechť navíc X a Y jsou nezávislé n.v. Určete rozdělení $Z = X + Y$.*

Z interpretace binomického rozdělení jako počtu úspěchů můžeme vidět, že by mělo být $Z \sim Bin(m+n,p)$: když X počítá úspěchy v m pokusech a Y počet v dalších n pokusech, tak jejich součet bude celkový počet úspěchů. Může nás ale nahlodat myšlenka, jestli každou binomickou n.v. můžeme chápat jako počet úspěchů, nebo jestli je to jenom speciální (byť důležitý) příklad. Nechme tuto otázkou zatím stranou a vyřešme úlohu pomocí Věty 40

$$\begin{aligned}
P(Z=z) &= \sum_{k=0}^{m} P(X=k)P(Y=z-k) \\
&= \sum_{k=0}^{m} \binom{m}{k} p^k (1-p)^{m-k} \binom{n}{z-k} p^{z-k} (1-p)^{n-(z-k)} \\
&= p^z (1-p)^{m+n-z} \sum_{k=0}^{m} \binom{m}{k} \binom{n}{z-k} \\
&= p^z (1-p)^{m+n-z} \binom{m+n}{z}
\end{aligned}$$

Pro snazší indexování zde chápeme $\binom{n}{\ell}$ jako 0, pokud není $0 \leq \ell \leq n$. Používáme zde standarní sumu kombinačních čísel ze Sekce 19.

Věta 43. *Při značení z Věty 39 platí*

$$\mathbb{E}(g(X,Y)) = \sum_{x \in ImX} \sum_{y \in ImY} g(x,y) P(X=x \,\&\, Y=y).$$

Důkaz. Položme $Z = g(X,Y)$. Podle definice je $\mathbb{E}(Z) = \sum_z z \cdot p_Z(z)$. Dosadíme

podle věty 39 a dostaneme

$$\begin{aligned}\mathbb{E}(Z) &= \sum_z z \sum_{x,y:g(x,y)=z} P(X = x \,\&\, Y = y) \\ &= \sum_z \sum_{x,y:g(x,y)=z} g(x,y) P(X = x \,\&\, Y = y) \\ &= \sum_{x,y:g(x,y)} g(x,y) P(X = x \,\&\, Y = y)\end{aligned}$$

Ve všech sumách sčítáme jen pro $x \in Im(X)$ a $y \in Im(Y)$. Poslední úprava plyne z toho, že každá dvojice (x, y) se vyskytne v právě jednom členu vnější sumy – jmenovitě v tom, kde je $z = g(x, y)$. □

Věta 44. *Pro X, Y n.v. a $a, b \in \mathbb{R}$ platí*

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$

Důkaz. Použijeme Větu 43 pro $g(x, y) = ax + by$. □

Věta 45. *Pro **nezávislé** diskrétní n.v. X, Y platí*

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

Důkaz. Využijeme Větu 43 pro funkci danou předpisem $g(x, y) = xy$. Platí tedy

$$\begin{aligned}\mathbb{E}(XY) &= \sum_{x \in Im(X)} \sum_{y \in Im(Y)} xy P(X = x \,\&\, Y = y) \\ &= \sum_{x \in Im(X)} \sum_{y \in Im(Y)} xy P(X = x) \cdot P(Y = y) \\ &= \sum_{x \in Im(X)} x P(X = x) \sum_{y \in Im(Y)} y \cdot P(Y = y) \\ &= \mathbb{E}(X) \cdot \mathbb{E}(Y).\end{aligned}$$

□

Kovariance

Definice 46. *Pro n.v. X, Y definujeme jejich* kovarianci (covariance) *předpisem*

$$\operatorname{cov}(X, Y) = \mathbb{E}\big((X - \mathbb{E}X)(Y - \mathbb{E}Y)\big).$$

Věta 47.

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

Důkaz. Snadno roznásobením a použitím linearity střední hodnoty. □

Jednoduché vlastnosti kovariance:

- $\operatorname{var}(X) = \operatorname{cov}(X, X)$ (přímo z definice)
- $\operatorname{cov}(X, aY+bZ+c) = a\operatorname{cov}(X, Y)+b\operatorname{cov}(X, Z)$ (linearita střední hodnoty)
- $\operatorname{cov}(X, Y) = 0$ pokud X,Y jsou nezávislé (Věta 47)
- ale nejen tehdy (viz níže)

Definice 48. Korelace (correlation) *náhodných veličin X, Y je definována předpisem*

$$\varrho(X, Y) = \frac{\operatorname{cov}(X, Y)}{\sqrt{\operatorname{var}(X)\operatorname{var}(Y)}}.$$

- je to přenormovaná kovariance
- $-1 \leq \varrho(X, Y) \leq 1$ (aplikace Cauchyho nerovnosti, necháme čtenáři k rozmyšlení)
- Korelace neznamená příčinnou souvislost! Např., korelace je symetrická, kauzalita nikoli! Podstatnější je, že veličiny spolu mohou korelovat (tj. mít vysokou korelaci) náhodou, a nejčastěji proto, že mají obě společnou příčinu (počet deštníků a výška kaluží na ulici spolu budou jistě korelovány, ale není mezi nimi příčinná souvislost).
- Naopak, nekorelace neznamená nezávislost. (Př: X libovolná, $Y = +X$ nebo $Y = -X$, obojí se stejnou pravděpodobností.)

Rozptyl součtu

Věta 49. *Nechť $X = \sum_{i=1}^{n} X_i$. Pak*

$$\operatorname{var}(X) = \sum_{i=1}^{n}\sum_{j=1}^{n}\operatorname{cov}(X_i, X_j) = \sum_{i=1}^{n}\operatorname{var}(X_i) + \sum_{i\neq j}\operatorname{cov}(X_i, X_j).$$

Speciálně, jsou-li $X_1, \ldots, X_n$ po dvou nezávislé, pak

$$\operatorname{var}(X) = \sum_{i=1}^{n}\operatorname{var}(X_i).$$

Důkaz. Spočítáme napřed $\mathbb{E}(X^2)$ a pak $\mathbb{E}(X)^2$, v obou případech použijeme linearitu střední hodnoty a roznásobení součinu součtů (pokaždé v jiném pořadí):

$$\mathbb{E}(X^2) = \mathbb{E}\Big(\big(\sum_{i=1}^{n} X_i\big)^2\Big) = \mathbb{E}\Big(\sum_{i=1}^{n}\sum_{j=1}^{n} X_i X_j\Big) = \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}(X_i X_j)$$

$$\mathbb{E}(X)^2 = \Big(\mathbb{E}\big(\sum_{i=1}^{n} X_i\big)\Big)^2 = \big(\sum_{i=1}^{n}\mathbb{E}(X_i)\big)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}(X_i)\mathbb{E}(X_j)$$

Odečtením obou rovnic a použitím Věty 47, dostáváme první rovnost ve větě. Další pak plynou použitím snadných vlastností uvedených výše. □

Podmíněné rozdělení Nechť X, Y jsou diskrétní náhodné veličiny na $(\Omega, \mathcal{F}, P)$, $A \in \mathcal{F}$ Budeme zkoumat rozdělení náhodné veličiny X, podmíněné jevem A, případně jevem definovaným pomocí n.v. Y. Protože podmíněnou pravděpodobnost už známe, nejde o nic nového, spíš o zavedení značení.

- $p_{X|A}(x) := P(X = x \mid A)$
  příklad: X je výsledek hodu kostkou, $A =$ padlo sudé číslo
- $p_{X|Y}(x|y) = P(X = x \mid Y = y)$ příklad: X, Z jsou výsledky dvou nezávislých hodů kostkou, $Y = X + Z$.

  $$p_{X|Y}(6|10) = \frac{p_{X,Y}(6, 10)}{p_Y(10)} = \frac{1/36}{3/36} = \frac{1}{3}.$$

  Nebo (možná lépe): pokud víme, že $Y = 10$, tak máme tři možnosti $4+6$, $5+5$, $6+4$, všechny stejně pravděpodobné, proto pravděpodobnost $1/3$. (Oproti výpočtu výše, tady jsme ušetřili výpočet, že každá možnost má pravděpodobnost $1/36$, což tady nebolí, ale ve složitějších případech by mohlo.)

Pozorování 50.

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x, y)}{p_Y(y)} = \frac{p_{X,Y}(x, y)}{\sum_{x'} p_{X,Y}(x', y)}$$

Důkaz. Definice podmíněné pravděpodobnosti a Věta 37. □

TODO: příklady

5 Spojité náhodné veličiny

V této kapitole budeme zkoumat obecné náhodné veličiny – tj. takové, které nemusí být diskrétní. Typický příklad je uniformně náhodné číslo z nějakého intervalu (programovací jazyky typicky simulují takovou náhodnou veličinu pro interval $(0, 1)$). Budeme ale zkoumat náhodné veličiny obecněji – začněme tedy definicí.

Definice 51. Náhodná veličina (random variable) *na* $(\Omega, \mathcal{F}, P)$ *je zobrazení* $X : \Omega \to \mathbb{R}$*, které pro každé* $x \in \mathbb{R}$ *splňuje*

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

Pozorování 52. *Diskrétní n.v. je n.v.*

Typicky se budeme zabývat spojitými náhodnými veličinami, neboli takovými, kde každé reálné číslo má nulovou pravděpodobnost, že ho dostaneme. Pravděpodobnostní funkce nám tedy nepomůže, a musíme použít nový způsob popisu náhodných veličin. Tím nejobecnějším je *distribuční funkce*.

Definice 53. Distribuční funkce (cumulative distribution function, CDF) *n.v. X je funkce $F_X : \mathbb{R} \to [0,1]$ definovaná předpisem*

$$F_X(x) := P(X \leq x) = P(\{\omega \in \Omega : X(\omega) \leq x\}).$$

Rozmysleme si napřed základní vlastnosti distribučních funkcí.

Věta 54. *Nechť X je náhodná veličina. Pak*

1. *F_X je neklesající funkce*
2. *$\lim_{x\to-\infty} F_X(x) = 0$*
3. *$\lim_{x\to+\infty} F_X(x) = 1$*
4. *F_X je zprava spojitá*

Důkaz. 1. Uvažme reálná čísla $x_1 < x_2$. Potřebujeme ukázat, že $P(X \leq x_1) \leq P(X \leq x_2)$. To ale plyne ze základních vlastností (Věta 5), protože pokud $X(\omega) \leq x_1$, tak také $X(\omega) \leq x_2$.

3. Označme $A_n = \{X \leq n\}$. Rozhodně platí $A_1 \subseteq A_2 \subseteq \cdots$ Podle Věty 103 (kterou jsme si nedokazovali, ale vypadá snad důvěryhodně) je

$$\lim_{n\to\infty} P(A_n) = P(\bigcup_{n\in\mathbb{N}} A_n) = P(\Omega) = 1.$$

2, 4: obdobně, detaily vynecháme. □

Spojité náhodné veličiny Nyní přejděme k nejběžnějším veličinám, které nejsou diskrétní: ke spojitým. Později si ukážeme, jaké další možnosti náhodných veličin jsou.

Definice 55. *N.v. X se nazývá spojitá (continuous), pokud existuje nezáporná reálná funkce f_X tak, že*

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt.$$

(Někdy se též používá pojem absolutně spojitá veličina.) Funkce f_X se nazývá hustota (probability density function, pdf) *náhodné veličiny X.*

Definici lépe porozumíme, když si rozmyslíme, jak ji můžeme použít ke generování náhodné veličiny s danou hustotou. Uvažme nezápornou reálnou funkci f, pro niž je $\int_{-\infty}^{\infty} f = 1$. Označme S množinu bodů pod grafem f, tj. $S = \{(x,y) \in \mathbb{R}^2 : 0 \leq y \leq f(x)\}$. Nyní uvažme náhodný bod b z množiny S, Vzpomeňme si na definici geometrického pravděpodobnostního prostoru, pro množinu $A \subseteq S$ je $P(b \in A) = V_2(A)/V_2(S) = V_2(A)$. (Zde V_2 označuje obsah (dvourozměrný objem) dané množiny. Podmínka na integrál f zařídí, že $V_2(S) = 1$.) Nyní označíme x-ovou souřadnici bodu b jako X. Tím jsme jistě popsali nějakou náhodnou veličinu. Prozkoumáme ji tím, že zjistíme, jakou má

distribuční funkci F_X. Označme $S_x = \{(t, y) \in \mathbb{R}^2 : 0 \leq y \leq f(t) \,\&\, t \leq x\}$. Jistě je

$$F_X(x) = P(X \leq x) = P(b \in S_x) = V_2(S_x) = \int_{-\infty}^{x} f(t)dt$$

Takže podle Definice 55 je funkce f hustota n.v. X.

TODO: fig

Věta 56. *Nechť spojitá n.v. X má hustotu f_X. Pak*

1. $P(X = x) = 0$ *pro každé* $x \in \mathbb{R}$.
2. $P(a \leq X \leq b) = \int_a^b f_X(t)dt$ *pro každé* $a < b \in \mathbb{R}$.

Důkaz. Přímo podle definice distribuční funkce a vlastností integrálu platí

$$\begin{aligned} P(a < X \leq b) &= P(X \leq b) - P(X \leq a) \\ &= F_X(b) - F_X(a) \\ &= \int_{-\infty}^{b} f_X - \int_{-\infty}^{a} f_X \\ &= \int_a^b f_X \end{aligned}$$

Odsud je možné dokázat bod 1 tak, že položíme $b = x$ a $a = b - 1/n$ a rozmyslíme si, že pro $n \to \infty$ se bude ten poslední integrál blížit nule. (Vidět je to snadno v případě, kdy f_X je omezená funkce, detaily vynecháme.) Protože

$$0 \leq P(X = x) \leq P(x - 1/n < X \leq x)$$

a pravá strana se blíží nule, je i $P(X = x) = 0$.

Důkaz 2 je už lehký:

$$P(a \leq X \leq b) = P(a < X \leq b) + P(X = a).$$

$P(X = a) = 0$ podle 1, a $P(a < X \leq b)$ je rovna kýženému integrálu. □

TODO napsat lépe! takže hustota náhodné veličiny je „limita histogramů“

$\frac{1}{2h} \int_{x-h}^{x+h} f_X(t)dt$ je průměrná hodnota funkce f_X na intervalu $(x - h, x + h)$. Pokud se tedy funkce moc neosciluje, je to pro malá h přibližně rovno $f_X(x)$. Na druhou stranu je to $P(x - h < X < x + h)/(2h)$ – čili opravdu hustota pravděpodobnosti, tedy pravděpodobnost intervalu vydělená jeho délkou.

Střední hodnota spojité n.v.

Definice 57. *Nechť spojitá n.v. X má hustotu f_X. Pak její* střední hodnota (expectation, expected value, mean) *je označována $\mathbb{E}(X)$ a definována*

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \; f_X(x)dx,$$

pokud integrál má smysl, tj. pokud se „nejedná o typ $\infty - \infty$“.

- Pro lidi s fyzikální intuicí: stejně se vypočte těžiště tyče ze znalosti hustoty v jednotlivých bodech.
- Opticky je vzorec pro $\mathbb{E}(X)$ podobný vzorci pro diskrétní n.v. Ukažme si, jak jde diskrétní n.v. použít pro aproximaci X. Zvolíme malé $\delta > 0$ a budeme měřit s přesností δ: místo X uvážíme $Y = \lfloor \frac{X}{\delta} \rfloor \delta$. Pokud je definice správně, tak čekáme že bude definovat $\mathbb{E}(X)$ blízkou hodnotě $\mathbb{E}(Y)$. Ale Y je diskrétní n.v., takže můžeme srovnat se vzorcem, který už známe.

$$\begin{aligned}\mathbb{E}(X) &= \int_{-\infty}^{\infty} x\ f_X(x)dx = \sum_{n\in\mathbb{Z}} \int_{n\delta}^{(n+1)\delta} x f_X(x)dx \\ &\doteq \sum_{n\in\mathbb{Z}} \int_{n\delta}^{(n+1)\delta} n\delta f_X(x)dx = \sum_{n\in\mathbb{Z}} n\delta \int_{n\delta}^{(n+1)\delta} f_X(x)dx \\ &= \sum_{n\in\mathbb{Z}} n\delta P(n\delta \le X < (n+1)\delta) = \sum_{n\in\mathbb{Z}} n\delta P(Y = n\delta)\end{aligned}$$

  A to je už podle definice střední hodnota veličiny Y. Při pečlivějším pohledu můžeme zjistit, že jsme se dopustili chyby nejvýše δ.

Věta 58 (Pravidlo naivního statistika)**.** *Pokud X je spojitá n.v. s hustotou f_X a g reálná funkce, tak*

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)dx,$$

pokud integrál má smysl.

(Důkaz vynecháme, dělal by se pomocí substituce v integrálu.)

Věta 59 (Linearita střední hodnoty)**.** *Pro $X_1, \ldots, X_n$ diskrétní nebo spojité n.v. platí*

$$\mathbb{E}(a_1X_1 + \cdots + a_nX_n) = a_1\mathbb{E}(X_1) + \cdots + a_n\mathbb{E}(X_n).$$

(Důkaz bude později.)

Rozptyl spojité n.v. Mějme n.v. X s $\mathbb{E}(X) = \mu$. Stejně jako pro diskrétní případ definujeme rozptyl jako $\mathbb{E}((X-\mu)^2)$ a proto podle Věty 58

$$\operatorname{var}(X) := \mathbb{E}((X-\mu)^2) = \int_{-\infty}^{\infty} (x-\mu)^2\ f_X(x)dx.$$

I pro spojité n.v. platí $\operatorname{var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$ (se stejným důkazem jako pro diskrétní n.v., protože i tady platí linearita střední hodnoty). Můžeme tedy rozptyl počítat snáze, když opět pomocí Věty 58 spočteme:

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2\ f_X(x)dx$$

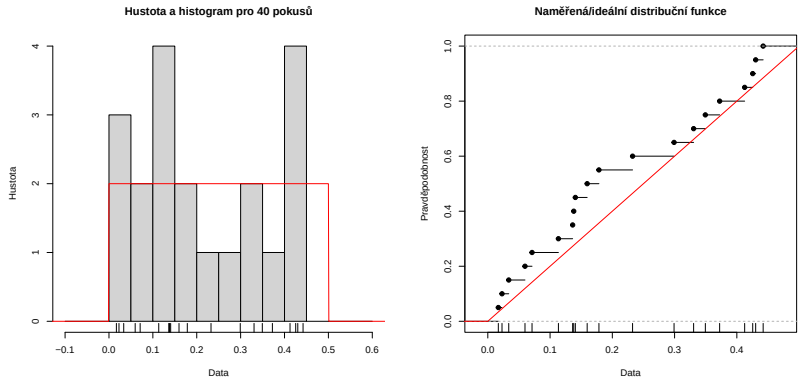
6 Konkrétní spojitá rozdělení a jejich parametry

6.1 Uniformní rozdělení

- N.v. X má uniformní rozdělení na intervalu $[a, b]$, píšeme $X \sim U(a, b)$, pokud $f_X(x) = 1/(b-a)$ pro $x \in [a, b]$ a $f_X(x) = 0$ jinak.
- Distribuční funkce má vzorec $F_X(x) = (x-a)/(b-a)$ pro $x \in [a, b]$, $F_X(x) = 0$ pro $x \leq a$ a $F_X(x) = 1$ pro $x \geq b$.

Pro ilustraci si ukažme spolu s grafem hustoty a distribuční funkce i naměřená data (40 hodnot vygenerovaných z uniformního rozdělení na intervalu $(0, 0.5)$. Pro vysvětlení: vlevo je kromě hustoty znázorněný i histogram – tj. pro každý interval délky 0.05 spočítáme kolik procent dat v intervalu leží a vydělíme délkou intervalu (odhadujeme hustotu pravděpodobnosti, nikoli pravděpodobnost). Pro kontrolu, naměřená data jsou vyznačená na vodorovné ose.

Na obrázku vpravo kromě distribuční funkce, tj. funkce která v bodě x má hodnotu $P(X \leq x)$, kreslíme její odhad pomocí nasamplovaných 40 hodnot – tj. kolik procent dat je $\leq x$. (Této funkci se říká ze zjevných důvodů empirická distribuční funkce. Ještě se jí budeme věnovat více ve statistické části.)



Střední hodnota $X \sim U(a, b)$ by jistě měla být $(a+b)/2$. Ověřme, že definice fungují, jak mají:

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x) = \int_a^b \frac{x}{b-a} = \left[\frac{x^2/2}{b-a}\right]_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}$$

Analogicky

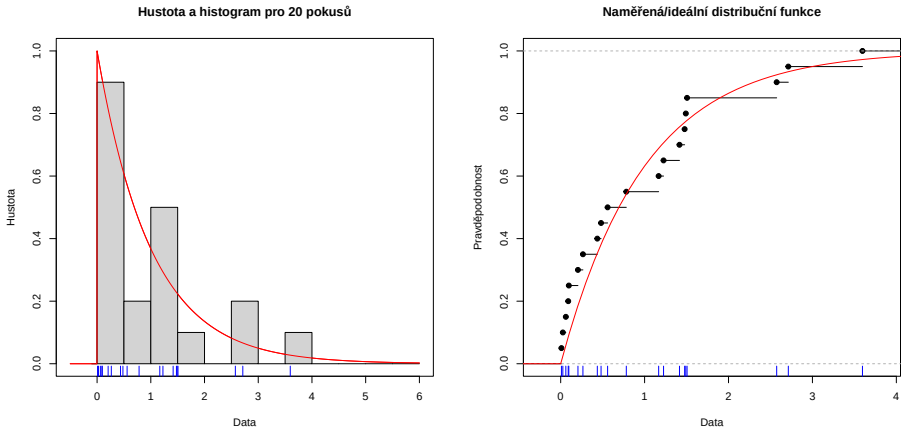
$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x) = \int_a^b \frac{x^2}{b-a} = \left[\frac{x^3/3}{b-a}\right]_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{a^2+ab+b^2}{3}$$

Odsud již snadným výpočtem získáme

$$\mathrm{var}(X) = \frac{(b-a)^2}{12}.$$

6.2 Exponenciální rozdělení

$$f_X(x) = \begin{cases} 0 & \text{pro } x \leq 0 \\ \lambda e^{-\lambda x} & \text{pro } x \geq 0 \end{cases} \qquad F_X(x) = \begin{cases} 0 & \text{pro } x \leq 0 \\ 1 - e^{-\lambda x} & \text{pro } x \geq 0 \end{cases}$$

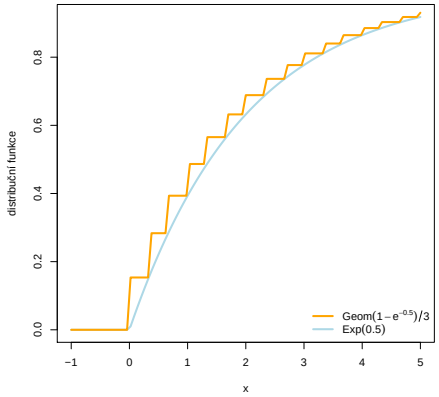


- X modeluje např. čas před příchodem dalšího telefonního hovoru do call-centra/dotazu na web-server/čas do dalšího blesku v bouřce/rozpadu atomu/...

Souvislost $X \sim Exp(\lambda)$ a $Y \sim Geom(p)$

- $P(X > n\delta) = e^{-\lambda n\delta}$ pro $\delta > 0$ a libovolné n
- $P(Y > n) = (1-p)^n$ pro $n \in \mathbb{N}$

Zkusíme pochopit a aproximovat X pomocí Y. Mějme dáno λ. Zvolíme vhodné malé $\delta > 0$ (délku časového intervalu), a vybereme p takové, aby $1-p = e^{-\lambda\delta}$, neboli $p \doteq \lambda\delta$. Pak pro každé $n \in \mathbb{N}$ je $P(X > n\delta) = P(Y > n)$ (zjevné porovnáním vzorců). Pokud se tedy spokojíme s hodnotou X „s přesností δ“ – tj. stačí nám zjistit $\lceil X/\delta \rceil$ – vystačíme s geometrickou náhodnou veličinou Y. Pro představu – atom uranu si každou milisekundu hodí korunou, jestli se má rozpadnout. Z tohoto pohledu je tedy exponenciální rozdělení limitou geometrického, kdy parametr δ zmenšujeme k nule.



Spočtěme nyní střední hodnotu $X \sim Exp(\lambda)$, obdobně jako výše pro uniformní rozdělení.

$$\begin{aligned}\mathbb{E}(X) &= \int_{-\infty}^{\infty} x f_X(x) = \int_0^{\infty} x\lambda e^{-\lambda x} \qquad \text{použijeme per-partes} \\ &= \left[-xe^{-\lambda x}\right]_0^{\infty} - \int_0^{\infty} -1 \cdot e^{-\lambda x} \\ &= 0 - 0 + \left[-e^{-\lambda x}/\lambda\right]_0^{\infty} = \frac{1}{\lambda}\end{aligned}$$

Analogicky spočteme druhý moment:

$$\begin{aligned}\mathbb{E}(X^2) &= \int_{-\infty}^{\infty} x^2 f_X(x) = \int_0^{\infty} x^2\lambda e^{-\lambda x} \qquad \text{použijeme opět per-partes} \\ &= \left[-x^2 e^{-\lambda x}\right]_0^{\infty} - \int_0^{\infty} -2x \cdot e^{-\lambda x} \\ &= \int_0^{\infty} 2xe^{-\lambda x} = \frac{2}{\lambda^2}\end{aligned}$$

Odsud již snadným výpočtem získáme

$$\operatorname{var}(X) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}.$$

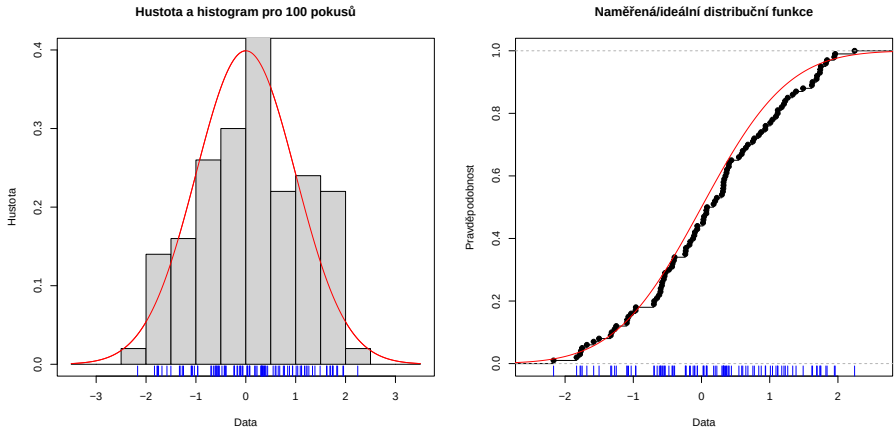
6.3 Normální rozdělení

Řekneme, že náhodná veličina X má *standardní normální rozdělení* (a píšeme $X \sim N(0,1)$), pokud $f_X = \varphi$, kde $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ (viz obrázek níže). Příslušnou distribuční funkci značíme Φ. Vzorec pro ni ale žádný není, víme jen že je to primitivní funkce k φ – a jde spočítat numericky. (Ta divná konstanta s $\sqrt{\pi}$ je tam proto, aby $\int_{-\infty}^{\infty} \varphi(t)dt = 1$.) Vzorec asi na první pohled

nevysvětluje, proč je to důležité rozdělení (i když spojení druhé mocniny a exponenciální funkce je asi přirozené). Ale jak uvidíme dále, je to asi nejdůležitější spojité rozdělení, možná hlavní důvod, proč spojitá rozdělení potřebujeme zkoumat.

Pro $X \sim N(0,1)$, platí $\mathbb{E}(X) = 0$, $\mathrm{var}(X) = 1$.

TODO: výpočet



Obecné normální rozdělení značíme $N(\mu, \sigma^2)$, jedná se o vhodně přeškálované standardní normální rozdělení. Řekneme, že $X \sim N(\mu, \sigma^2)$, pokud pro $Z = \frac{X-\mu}{\sigma}$ platí $Z \sim N(0,1)$. Nebo také obráceně: pro standardně normálně rozdělenou veličinu Z je $\mu + \sigma Z \sim N(\mu, \sigma^2)$. Není těžké odvodit, že normální rozdělení $N(\mu, \sigma^2)$ má hustotu $\frac{1}{\sigma}\varphi(\frac{x-\mu}{\sigma})$.

Z volby značení asi tušíte, že μ bude přislušná střední hodnota a σ^2 rozptyl. Že je to opravdu tak, plyne z Věty 27, a 35).

Odolnost vůči součtu Zajímavá a dost jedinečná vlastnost normálního rozdělení: Pokud $X_1, \ldots, X_k$ jsou n.n.v., kde $X_i \sim N(\mu_i, \sigma_i^2)$, pak jejich součet je také normální n.v., konkrétně

$$X_1 + \cdots + X_k \sim N(\mu, \sigma^2),$$

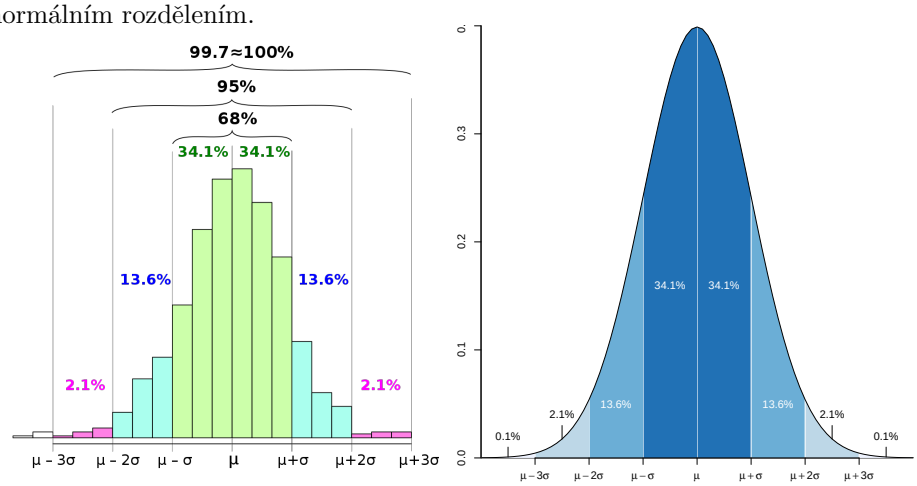
kde $\mu = \mu_1 + \cdots + \mu_k$ a $\sigma^2 = \sigma_1^2 + \cdots + \sigma_k^2$. Brzy si ukážeme dva způsoby, jak tuto vlastnost ověřit.

Význam normálního rozdělení Z centrální limitní věty (o které budeme mluvit později) je normální rozdělení dobrá aproximace pro součet mnoha nezávislých náhodných veličin. To trochu vysvětluje výše zmíněnou „odolnost vůči součtu“, ale zejména je to důvod, proč normální rozdělení je v praxi hodně používané. Mnoho v praxi pozorovaných veličin má totiž takový charakter: doba jízdy závisí na (víceméně nezávislých) dobách čekání na jednotlivých semaforech, výška člověka závisí na mnoha různých genech, atd.

Pro praktické použití je užitečné pravidlo 3σ (anglicky 68–95–99.7 rule). To říká, že

$$P(\mu - \sigma < X < \mu + \sigma) \doteq 0.68$$
$$P(\mu - 2\sigma < X < \mu + 2\sigma) \doteq 0.95$$
$$P(\mu - 3\sigma < X < \mu + 3\sigma) \doteq 0.997,$$

viz obrázek. Pokud narazíte na text, který říká, že tento index je větší než 2 pro 2.5% populace, budete hned vědět, že se jedná o veličinu se standardním normálním rozdělením.



(Obrázek vlevo z Wikipedie, autor Melikamp.)

6.4 Cauchyho rozdělení

Cauchyho rozdělení má hustotu $f(x) = \frac{1}{\pi(1+x^2)}$. Její distribuční funkce je tedy

$$F(x) = \int_{-\infty}^{x} f(t)dt = \frac{1}{\pi}\left[\operatorname{arctg} t\right]_{-\infty}^{x} = \frac{1}{\pi}\operatorname{arctg} x + \frac{1}{2}.$$

Uvádíme ji zde hlavně jako výstražný příklad. Přestože graf vypadá podobně jako normální rozdělení (jen trochu víc „rozplácnuté“), tak se zásadně liší: nemá střední hodnotu! Sice ze symetrie grafu bychom mohli usoudit, že by toto rozdělení mělo mít střední hodnotu 0, stejně jako $N(0,1)$, ale zde se jedná o nedefinovaný případ $\infty - \infty$. Je totiž

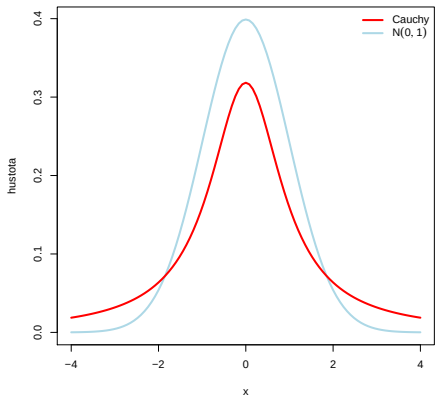
$$\int_0^{\infty} x \cdot f(x) = \left[\frac{1}{2\pi}\log 1 + x^2\right]_0^{\infty} = \infty.$$

Můžeme to i prakticky ověřit: když vygenerujeme mnoho náhodných hodnot z tohoto rozdělení a spočítáme jejich průměr, dostaneme hodnoty velice daleko od nuly – na rozdíl od $N(0,1)$.

Pro vyzkoušení v R: **mean**(**rcauchy**(10^6)) versus **mean**(**rnorm**(10^6)).

A totéž v Pythonu:

```
import numpy as np
print(np.mean(np.random.normal(0,1,10**6)))
print(np.mean(np.random.standard_cauchy(10**6)))
```



6.5 Paretovo rozdělení

Paretovo rozdělení má předpis distribuční funkce

$$F_X(x) = \begin{cases} 1 - \left(\frac{x_0}{x}\right)^\alpha & x \geq x_0 \\ 0 & x < x_0 \end{cases}.$$

Potřebujeme, aby $\alpha, x_0 > 0$, jinak by funkce F_X nemohla být distribuční funkce, viz Věta 54.

Za zmínku stojí jednodušší vzorec *funkce přežití (survival function)*, $S_X(x) = P(X > x) = (x_0/x)^\alpha$. Je vidět, že tato funkce klesá k nule mnohem pomaleji než pro exponenciální rozdělení, říkáme, že Paretovo rozdělení má *long tail* (velký chvost). Praktický příklad veličiny s tímto rozdělením je např. velikost souboru přenášeného po internetu nebo délka fungujícího úseku na disku.

Z distribuční funkce odvodíme hustotu $f_X(x) = F_X(x)' = \alpha x_0^\alpha / x^{\alpha+1}$ pro $x \geq x_0$, $f_X(x) = 0$ pro $x < x_0$.

Pro výpočet střední hodnoty použijeme definici

$$\int_{-\infty}^{\infty} x f_X(x) = \int_{x_0}^{\infty} \alpha \left(\frac{x_0}{x}\right)^\alpha = \left[\frac{\alpha}{1-\alpha} \frac{x_0^\alpha}{x^{\alpha-1}}\right]_{x_0}^{\infty} = x_0 \frac{\alpha}{\alpha - 1}.$$

Zde už potřebujeme, aby bylo $\alpha > 1$, pro $\alpha \leq 1$ vyjde střední hodnota ∞ (a pro $\alpha = 1$ i jiná primitivní funkce).

6.6 Přehled náhodných veličin

Diskrétní n.v.	**Spojitá n.v.**
$X : \Omega \to \mathbb{R}$	
obor hodnot $Im(X)$ konečný/spočetný	$Im(X)$ je typicky interval (i nekonečný)
$F_X(x) = P(X \leq x)$	
$p_X(x) = P(X = x)$	$f_X(x) = F'(x) \doteq P(x < X < x + h)/h$
$\sum_x p_X(x) = 1$	$\int_{-\infty}^{\infty} f_X(x)dx = 1$
$P(X \in A) = \sum_{x \in A} p_X(x)$	$P(X \in A) = \int_A f_X(x)dx$
$F_X(x) = \sum_{t \leq x} p_X(t)$	$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$
$\mathbb{E}(X) = \sum_x x \cdot p_X(x)$	$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \cdot f_X(x)dx$
$\mathbb{E}(g(X)) = \sum_x g(x) \cdot p_X(x)$	$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) \cdot f_X(x)dx$
Př: počet hodů než padne šestka	Př: doba běhu programu

Tabulka 1: Srovnání diskrétních a spojitých náhodných veličin. Ve sloupci vlevo jsou všechny sumační proměnné omezeny na $Im(X)$, tj. obor hodnot X.

6.7 Další rozdělení

Existuje mnoho dalších pojmenovaných rozdělení, na některá narazíme později. Zde jen stručný výčet:

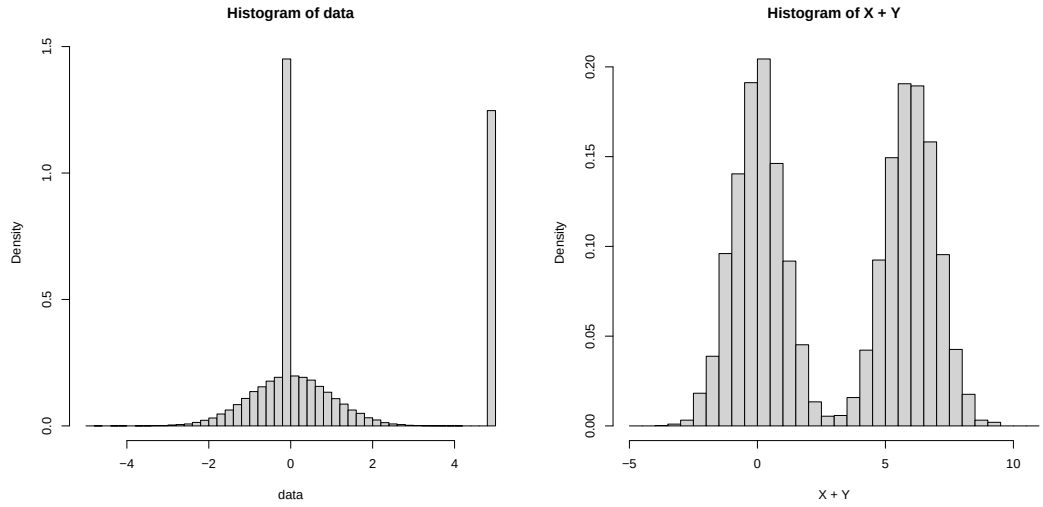
- Gamma rozdělení – rozdělení součtu několika nezávislých veličin s exponenciálním rozdělením. Modeluje jevy, kdy je potřeba „několikrát vystát frontu“.
- Beta rozdělení s parametry α, β: má hustotu $cx^{\alpha-1}(1-x)^{\beta-1}$ (pro $x \in (0,1)$).
- χ^2 rozdělení s k stupni volnosti = chí-kvadrát (χ^2_k) je speciální případ gamma rozdělení. Ale zejména je to rozdělení $Z_1^2 + \cdots + Z_k^2$, kde $Z_i \sim N(0,1)$ jsou n.n.v. – potkáme ve statistické části.
- Studentova t-distribuce – potkáme ve statistické části.

6.8 Kombinace spojitého a diskrétního rozdělení

Zkuste před dalším čtením tipnout, jaké veličiny jsou zachyceny na následujících obrázcích:

Diskrétní náh.vektory	**Spojité náh.vektory**
$X : \Omega \to \mathbb{R}^d$	
$F_{X,Y}(x,y) = P(X \leq x \,\&\, Y \leq y)$	
$p_{X,Y}(x,y) = P(X = x \,\&\, Y = y)$	$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}$
	$\doteq P(x < X < x+h \,\&\, y < Y < y+h)/h^2$
$\sum_y p_{X,Y}(y) = p_X(x)$	$\int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = f_X(x)$
$\sum_x p_{X,Y}(x) = p_Y(y)$	$\int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = f_Y(y)$
$P((X,Y) \in A) = \sum_{(x,y)\in A} p_{X,Y}(x,y)$	$P((X,Y) \in A) = \int_A f_{X,Y}(x,y)dxdy$
$p_{X+Y}(z) = \sum_x p_X(x)p_Y(z-x)$	$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx$
$\mathbb{E}(g(X,Y)) = \sum_{x,y} g(x,y) \cdot p_{X,Y}(x,y)$	$\mathbb{E}(g(X,Y)) = \int_{-\infty}^{\infty} g(x,y) \cdot f_{X,Y}(x,y)dxdy$

Tabulka 2: Srovnání diskrétních a spojitých náhodných vektorů.



7 Kvantilová funkce a Universalita uniformního rozdělení

Pro náhodnou veličinu X definujeme *kvantilovou funkci* $Q_X : [0,1] \to \mathbb{R}$ pomocí

$$Q_X(p) := \min\{x \in \mathbb{R} : p \leq F_X(x)\} \tag{2}$$

- Pokud F_X je spojitá, tak $Q_X = F_X^{-1}$ (inverzní funkce k F_X).

- Obecně platí:
$$Q_X(p) \leq x \Leftrightarrow p \leq F_X(x). \tag{3}$$

- Jako medián označujeme hodnotu $m = Q_X(1/2)$. Je to (nejmenší) takové číslo, že $P(X \leq m) \geq 1/2$ a $P(X > m) \leq 1/2$. Pokud je X diskrétní n.v., tak takových čísel je více, a v tom případě definice mediánu není ustálená. Dále budeme pro jednoduchost předpokládat spojité rozdělení.

- První kvartil je hodnota, $q = Q_X(1/4)$ taková, že jedna čtvrtina hodnot je $\leq q$ ($P(X \leq q) \leq 1/4$). Desátý percentil $Q_X(10/100)$ je hodnota větší než 10 % hodnot, atd.

Věta 60. *Nechť X je spojitá n.v., jejíž distribuční funkce $F = F_X$ je rostoucí. Pak $F(X) \sim U(0,1)$.*

Důkaz. Označme $Y = F(X)$ (tedy Y je náhodná veličina). Prozkoumáme, o jakou veličinu se jedná, pomocí její distribuční funkce. Pro $y \in (0,1)$ zvolme x takové, aby $y = F(x)$ (existuje díky spojitosti F).

$$P(Y \leq y) = P(F(X) \leq F(x)) = P(X \leq x) = F(x) = y.$$

Tudíž se F_Y shoduje s distribuční funkcí $U(0,1)$ na intervalu $(0,1)$, a tedy i všude jinde. Proto má Y uniformní rozdělení na $(0,1)$. □

Věta 61. *Nechť $U \sim U(0,1)$ a F je funkce „typu distribuční funkce“, tj. neklesající spojitá funkce s $\lim_{x\to-\infty} F(x) = 0$ a $\lim_{x\to+\infty} F(x) = 1$. Buď Q odpovídající kvantilová funkce (podle (2)). Pak $Q(U)$ je n.v. s distribuční funkcí F.*

Důkaz. Podobně jako v minulé větě, jenom jednodušeji. Označme $X = Q(U)$. Prozkoumáme distribuční funkci X:

$$P(X \leq x) = P(Q(U) \leq x) = P(U \leq F(x)) = F(x).$$

Tady první rovnost je definice X, druhá z vlastnosti (3), a poslední z vlastnosti uniformního rozdělení (a toho, že $F(x) \in [0,1]$). □

Větu 61 můžeme použít na generování náhodných veličin s daným rozdělením.

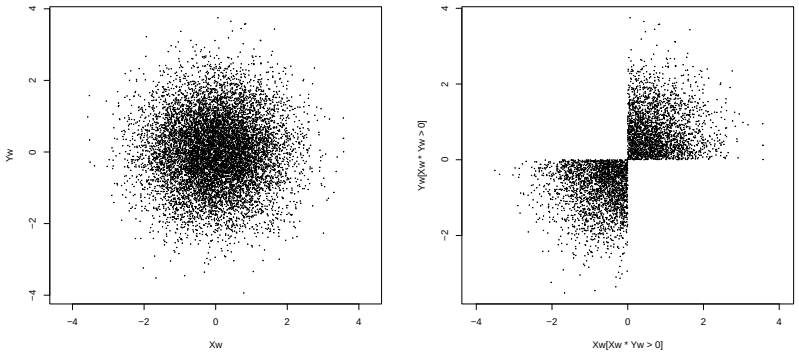
Příklad 62. *Vyrobte náhodnou veličinu s rozdělením $Exp(\lambda)$, pokud máte k dispozici $U \sim U(0,1)$.*

Víme, že distribuční funkce $Exp(\lambda)$ je $F(x) = 1 - e^{-\lambda x}$. Jedná se o spojitou funkci, příslušná kvantilová funkce je tedy její inverzní funkce. Snadno spočteme $Q(p) = \frac{\log(1-p)}{-\lambda}$. Podle Věty 61 má $Q(U) = \frac{\log(1-U)}{-\lambda}$ požadované rozdělení.

TODO: diskrétní n.v.

8 Náhodné vektory

Sdružená distribuční funkce (Joint cdf) Často chceme zkoumat několik náhodných veličin najednou. Budeme pak moci zkoumat jejich vzájemné závislosti, a např. rozlišit mezi dvěma následujícími obrázky (každá tečka znamená jednu vygenerovanou hodnotu náhodného vektoru).



Definice 63. *Pro n.v. X, Y na pravděpodobnostním prostoru $(\Omega, \mathcal{F}, P)$ definujeme jejich* sdruženou distribuční funkci (joint cdf) $F_{X,Y} : \mathbb{R}^2 \to [0,1]$ *předpisem*

$$F_{X,Y}(x,y) = P(\{\omega \in \Omega : X(\omega) \leq x \,\&\, Y(\omega) \leq y\}).$$

Mohli bychom definovat i pro více než dvě n.v. $F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = P(X_1 \leq x_1 \,\&\, \ldots X_n \leq x_n)$. Pro pohodlí značení budeme obvykle zkoumat jen dvojrozměrný případ.

Věta 64 (Pravděpodobnost obdélníku)**.** *Pokud $F = F_{X,Y}$, tak*

$$P(X \in (a,b] \,\&\, Y \in (c,d]) = F(b,d) - F(a,d) - F(b,c) + F(a,c).$$

Často můžeme sdruženou distribuční funkci psát jako integrál pomocí nezáporné funkce $f_{X,Y}$

$$F_{X,Y}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(x,y)dxdy.$$

Pak nazýváme n.v. X, Y sdruženě spojité. Funkce $f_{X,Y}$ je jejich *sdružená hustota.*

(Tak jako u jednorozměrného případu může být $f_{X,Y} > 1$, není to pravděpodobnost, ale její hustota – kolik připadá pravděpodobnosti na jednotku obsahu.)

Stejně jako u jednorozměrného případu můžeme pak pomocí hustoty vyjádřit i další pravděpodobnosti:

$$P((X,Y) \in S) = \int_S f_{X,Y}(x,y)dxdy.$$

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

Věta 65 (Pravidlo naivního statistika)**.** *Stejně jako v diskrétním případu platí pro střední hodnotu funkce dvou n.v.*

$$\mathbb{E}(g(X,Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)dxdy.$$

A stejně jako v diskrétním případu odsud odvodíme $\mathbb{E}(aX + bY + c) = a \cdot \mathbb{E}(X) + b \cdot \mathbb{E}(Y) + c$, a odsud indukcí Větu 59.

Nezávislost spojitých náhodných veličin

Definice 66. *Náhodné veličiny X, Y nazveme* nezávislé (independent)*, pokud jevy $\{X \leq x\}$ a $\{Y \leq y\}$ jsou nezávislé pro libovolná $x, y \in \mathbb{R}$. Ekvivalentně,*

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y), \qquad \textit{neboli}$$
$$F_{X,Y}(x,y) = F_X(x)F_Y(y)$$

Věta 67. *Nechť X, Y mají sdruženou hustotu $f_{X,Y}$ (a hustoty f_X, f_Y). Následující tvrzení jsou ekvivalentní:*

- *X, Y jsou nezávislé*
- *$f_{X,Y}(x,y) = f_X(x)f_Y(y)$*

Součet spojitých n.v.

Věta 68 (Konvoluce pro spojité n.v.)**.** *Nechť spojité X, Y jsou n.n.v. Pak $Z = X + Y$ je také spojitá n.v. a její hustotu dostaneme jako* konvoluci *funkcí f_X, f_Y, neboli*

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x)dx.$$

Příklad 69. *Nechť $X, Y \sim N(0,1)$ jsou n.n.v. Jaké je rozdělení jejich součtu $X + Y$?*

TODO: výpočet
Platí následující analogie Věty 37.

Věta 70 (Marginální hustota)**.**

$$f_X(x) = \int_{y\in\mathbb{R}} f_{X,Y}(x,y)dy$$
$$f_Y(y) = \int_{x\in\mathbb{R}} f_{X,Y}(x,y)dx$$

Podmíněná hustota

Definice 71. *Pro spojité n.v. X, Y definujeme* podmíněnou hustotu (conditional pdf) *předpisem*

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

pokud je $f_Y(y) > 0$, jinak ji nedefinujeme.

Vícerozměrné normální rozdělení Ukážeme si důležitý příklad spojitého náhodného vektoru. Připomeňme hustotu standardního normálního rozdělení, $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$. Položme

$$f(t_1, \ldots, t_n) = \varphi(t_1)\varphi(t_2)\cdots\varphi(t_n) = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{t_1^2+\cdots+t_n^2}{2}}$$

Všimněme si, že $f(t_1, \ldots, t_n) = (2\pi)^{-n/2} e^{-r^2/2}$, kde $r^2 = t_1^2 + \cdots + t_n^2$. Říkáme, že f je *radiálně symetrická funkce*, její hodnota záleží jen na vzdálenosti od počátku. (Na obrázku je pro ilustraci případ $n = 2$.)

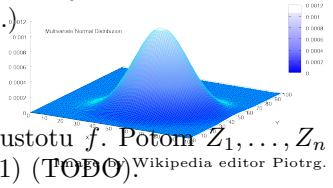


Image by Wikipedia editor Piotrg.

Nechť náhodný vektor $Z = (Z_1, \ldots, Z_n)$ má hustotu f. Potom $Z_1, \ldots, Z_n$ jsou n.n.v. (Věta 67) a pro všechna i je $Z_i \sim N(0,1)$ (TODO).

Z toho, že f je radiálně symetrická funkce plyne, že $Z/\|Z\|$ je uniformně náhodný bod na n-rozměrné sféře. (Což je nejlepší způsob, jak takový bod vygenerovat.)

Zároveň také skalární součin Z s libovolným jednotkovým vektorem má stejné rozdělení jako $\langle e_1, Z\rangle = Z_1$. Tudíž, pro každý jednotkový vektor u platí, že $\langle u, Z\rangle = \sum_{i=1}^n u_i Z_i$ má také rozdělení $N(0,1)$ To jsme si dříve uváděli jako „odolnost normálního rozdělení vůči sčítání“.

TODO: lépe vysvětlit?

Vícerozměrné normální rozdělení obecné Obecněji můžeme vzít náhodný vektor s hustotou $c \cdot e^{-Q(t)/2}$, kde $c > 0$ je vhodná konstanta a $Q(t) = (t-\mu)^T \Sigma^{-1} (t-\mu)$ je kvadratická funkce s positivně semidefinitní maticí Σ^{-1}. O takovém vektoru X říkáme, že má vícerozměrné normální rozdělení $N(\mu, \Sigma)$. Přitom $\mu \in \mathbb{R}^d$ je vektor středních hodnot, $\mu_i = \mathbb{E}(X_i)$ a Σ je matice kovariancí, $\Sigma_{i,j} = \mathrm{cov}(X_i, X_j)$. Standardní případ z předchozího odstavce odpovídá nulovému vektoru μ a jednotkové matici Σ. Pokud Σ není diagonální matice, tak souřadnice nejsou nezávislé – model tedy umožňuje popsat i složitější systémy náhodných veličin. Proto se často používá ve strojovém učení (tj. ve statistice).

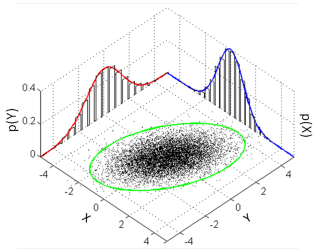


Image by Wikipedia editor Bscan.

Podmiňování Stejně jako v diskrétním případě zavedeme podmíněné pravděpodobnosti, tentokrát tedy v podobě distribuční funkce a hustoty.

Definice 72. *X je n.v. na $(\Omega, \mathcal{F}, P)$, $B \in \mathcal{F}$ a $P(B) > 0$.*

$$F_{X|B}(x) := P(X \leq x \mid B)$$

K tomu přísluší hustotní funkce $f_{X|B}$.

- pokud $B = \{X \in S\}$ (což, pro připomenutí, znamená přesněji $B = \{\omega \in \Omega : X(\omega) \in S\}$ pro nějakou množinu $S \subseteq \mathbb{R}$, tak

$$f_{X|B}(x) = \begin{cases} \frac{f_X(x)}{P(X \in S)} & \text{pokud } x \in S \\ 0 & \text{jinak} \end{cases}$$

Věta 73 (věta o rozkladu hustoty)**.** *Nechť X je spojitá n.v., nechť B_1, B_2, ... je rozklad Ω. Pak*

$$F_X(x) = \sum_i P(B_i) F_{X|B_i}(x) \quad a$$
$$f_X(x) = \sum_i P(B_i) f_{X|B_i}(x).$$

Důkaz. Věta o úplné pravděpodobnosti pro F_X, zderivujeme pro f_X. □

Podmíněná hustota a střední hodnota

- $\mathbb{E}(X \mid B) = \int_{-\infty}^{\infty} x \cdot f_{X|B}(x) dx$
- $\mathbb{E}(g(X)|B) = \int_{-\infty}^{\infty} g(x) f_{X|B}(x) dx$
- Pokud B_1, B_2, ... je rozklad, tak

$$\mathbb{E}(X) = \sum_i \mathbb{E}(X \mid B_i) P(B_i).$$

Podmíněná hustota a střední hodnota

- $f_{X|Y}(x|y) := \frac{f_{X,Y}(x,y)}{f_Y(y)}$ je hustota n.v. X, pokud $Y = y$
- $\mathbb{E}(X \mid Y = y) := \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y)dx$ je střední hodnota této veličiny
- $\mathbb{E}(g(X)|Y = y) = \int_{-\infty}^{\infty} g(x) \cdot f_{X|Y}(x|y)dx$
- $\mathbb{E}(X) = \int_{-\infty}^{\infty} \mathbb{E}(X \mid Y = y) f_Y(y)dy$

9 Nerovnosti

Věta 74 (Markovova nerovnost)**.** *Nechť náhodná veličina X splňuje $X \geq 0$, nechť $a > 0$ je reálné číslo. Pak*

$$P(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

Důkaz. Podle Věty 30

$$\begin{aligned}\mathbb{E}(X) &= P(X \geq a)\mathbb{E}(X \mid X \geq a) + P(X < a)\mathbb{E}(X \mid X < a) \\ &\geq P(X \geq a) \cdot a + 0\end{aligned}$$

Zde jsme použili toho, že $\mathbb{E}(X \mid X \geq a)$ je určitě alespoň a, dále $P(X < a) \geq 0$ a $\mathbb{E}(X \mid X < a) \geq 0$. Teď stačí jen vydělit a. □

Poznámky

- Pokud zvolíme $a = b \cdot \mathbb{E}(X)$, můžeme nerovnost psát ve tvaru

$$P(X \geq b \cdot \mathbb{E}(X)) \leq \frac{1}{b}.$$

- Extrémní případ je, pokud $X = a > 1$ s pravděpodobností $1/a$ a $X = 0$ jinak. Promyšlením tohoto příkladu vidíme, že k důkazu stačí obyčejné „kupecké počty“.
- Obrácený pohled: může být 51 % lidí dvakrát starší než průměr?

Příklad 75. *Sestavili jsme pravděpodobnostní algoritmus (tj., algoritmus, který si hází korunou) a zjistili jsme, že pro jeho dobu běhu X platí $\mathbb{E}(X) = n^2$ (n je velikost vstupu). Co můžeme odsud zjistit o tom, jak je pravděpodobné, že algoritmus poběží výrazně déle?*

I bez další analýzy algoritmu můžeme použít Markovovu nerovnost: je jistě $X \geq 0$, proto platí

$$P(X \geq cn^2) \leq \frac{1}{c}.$$

Věta 76 (Čebyševova nerovnost)**.** *Nechť X má konečnou střední hodnotu μ a rozptyl σ^2. Pak*

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

Důkaz. Položme $Y = (X-\mu)^2$. Jistě platí $Y \geq 0$ a $\mathbb{E}(Y) = \sigma^2$. Stačí tedy použít Markovovu nerovnost pro Y. □

Věta 77 (Chernoffova (Černovova) nerovnost)**.** *Nechť $X = \sum_{i=1}^{n} X_i$, kde X_i jsou n.n.v. nabývající hodnot ± 1 s pravděpodobností $1/2$. Pak pro $t > 0$ platí*

$$P(X \leq -t \cdot \sigma) = P(X \geq t \cdot \sigma) \leq e^{-t^2/2},$$

kde $\sigma = \sigma_X = \sqrt{n}$.

Větu nebudeme dokazovat, důkaz si uvedeme v dalším semestru (Pravděpodobnost a Statistika 2). Uvádíme ji zde jako ukázku toho, že pokud o veličině X máme speciální znalosti (jako zde, že je to součet nezávislých náhodných veličin), tak můžeme získat lepší odhad, než pro obecnou náhodnou veličinu, což dává Čebyševova nerovnost. Chernoffova nerovnost se často používá při analýze pravděpodobnostních algoritmů.

10 Limitní věty

10.1 Zákony velkých čísel

Věta 78 (Silný zákon velkých čísel (strong law of large numbers))**.** *Nechť $X_1, X_2, \ldots$ jsou stejně rozdělené n.n.v. se střední hodnotou μ a rozptylem σ^2. Označme $\bar{X}_n = (X_1 + \cdots + X_n)/n$ tzv.* výběrový průměr (sample mean)*. Pak platí*

$$\lim_{n\to\infty} \bar{X}_n = \mu \quad \textit{skoro jistě (tj. s pravděpodobností 1).}$$

Říkáme, že posloupnost $\bar{X}_n$ konverguje k μ skoro jistě (almost surely)*.*

Poznámky

- To je důvod, proč při dělání experimentů opakovat měření. Zároveň to ukazuje, že je třeba, aby měření byla nezávislá.
- Speciální případ: pro pravděpodobnostní prostor $(\Omega, \mathcal{F}, P)$ budeme zkoumat nějaký jev $A \in \mathcal{F}$. Experiment budeme nekonečněkrát opakovat, neboli uvážíme naše elementární jevy budou posloupnosti prvků Ω. $X_i(\omega) = 1$, pokud $\omega_i \in A$ (při i-tém opakování pokusu nastal jev A), a $X_i(\omega) = 0$ jinak.

  TODO

Monte Carlo integration Jak spočítat $\int_{x\in A} g(x)dx$?

Speciálně:

$$g(x) = \begin{cases} 1 & \text{pro } x \in S \\ 0 & \text{jinak} \end{cases}$$

... obsah kruhu

TODO: dopsat pořádně

Silný zákon velkých čísel dokazovat nebudeme, ale ukážeme si podobné tvrzení:

Věta 79 (Slabý zákon velkých čísel (weak law of large numbers))**.** *Nechť X_1, X_2, ... jsou stejně rozdělené n.n.v. se střední hodnotou μ a rozptylem σ^2. Označme $\bar{X}_n = (X_1 + \cdots + X_n)/n$. Pak pro každé $\varepsilon > 0$ platí*

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| > \varepsilon) = 0.$$

Říkáme, že posloupnost $\bar{X}_n$ konverguje k μ v pravděpodobnosti (in probability) *a píšeme $\bar{X}_n \xrightarrow{P} \mu$.*

Důkaz. Podle Věty 59 je $\mathbb{E}(\bar{X}_n) = (\mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n))/n = \mu$. A protože jsou jednotlivé sčítance nezávislé, můžeme použít i Větu 35 a dostaneme $\mathrm{var}(\bar{X}_n) = (\mathrm{var}(X_1) + \cdots + \mathrm{var}(X_n))/n^2 = \sigma^2/n$. Můžeme tedy použít Čebyševovu nerovnost pro $a = \sqrt{n}\varepsilon/\sigma$ a dostaneme, že

$$P(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

Tím jsme jednak dokázali větu, jednak i našli explicitní odhad na to, jak velká bude pravděpodobnost chyby pro konkrétní n, ε a σ. □

Jak naznačuje název předchozích vět, tak z Věty 78 plyne Věta 79. Uvědomte si ale, že jsme vlastně ukázali něco trochu víc – ukázali jsme, jak rychle se ta posloupnost pravděpodobností blíží k nule, neboli jsme získali nástroj, jak rozhodovat tvrzení následujícího typu:

Příklad 80. *Při měření hmotnosti uděláme chybu s normálním rozdělením a směrodatnou odchylkou 1 gram. Kolik měření musíme provést, aby průměr naměřených hodnot neměl větší chybu než* 0.1 *gramu?*

TODO dvě řešení – přes součet normálních i přes WLLN

10.2 Centrální Limitní věta

Nejdůležitější věta pravděpodobnosti a statistiky.

Věta 81. *Nechť X_1, X_2, ... jsou stejně rozdělené n.n.v. se střední hodnotou μ a rozptylem σ^2. Označme $Y_n = ((X_1 + \cdots + X_n) - n\mu)/(\sqrt{n} \cdot \sigma)$.*

Pak $Y_n \xrightarrow{d} N(0,1)$. *Neboli, pokud* F_n *je distribuční funkce* Y_n, *tak*

$$\lim_{n\to\infty} F_n(x) = \Phi(x) \quad \textit{pro každé } x \in \mathbb{R}.$$

Říkáme, že posloupnost Y_n *konverguje k* $N(0,1)$ v distribuci (in distribution).

Větu si zde dokazovat nebudeme, zájemce odkážeme na přednášku Pravděpodobnost a Statistika 2.

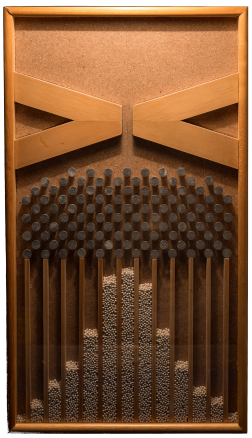
Rozmysleme si ale, proč má věta takové znění, jaké má.

- Stejným postupem jako v důkazu Věty 79 zjistíme, že $\mathbb{E}(Y_n) = 0$ a $\mathrm{var}(Y_n) = 1$. To je jednak klíčový rozdíl CLV a Zákona velkých čísel – zde se rozptyl neblíží nule, takže posloupnost Y_n se, oproti S_n neblíží jednobodovému rozdělení. To ještě nedokazuje, že Y_n se blíží $N(0,1)$, ale „má k tomu šanci".
- Pokud jsou všechna X_i normální s rozdělením $N(\mu, \sigma^2)$, tak (odolnost vůči součtu!) je i Y_n normální – s ohledem na předchozí bod bude přesně $Y_n \sim N(0,1)$. Neboli: jiné rozdělení než normální v CLV být nemůže!

TODO: obrázky
TODO: proč je to důležité TODO: co znamená konvergence v distribuci

Poznámky

- Speciální případ: pokud $X_1, \ldots, X_n \sim Ber(p)$, tak víme, že jejich součet se řídí rozdělením $Bin(n,p)$. V tomto případě je tedy Y_n přeškálované binomické rozdělení. To tedy ukazuje, že v distribuci se $(Bin(n,p) - np)/\sqrt{np(1-p)}$ blíží $N(0,1)$.
- Vizuální ukázku dává tzv. Galtonova deska: pravidelně rozmístěné hřebíky působí na padající písek jako posloupnost nezávislých náhodných veličin (posud o $+1$ nebo o -1 při každém spadnutí o jednu řadu dolů. Zaplněnost jednotlivých sloupečků vytváří histogram binomického rozdělení $Bin(n, 1/2)$ (kde n je počet řádek). Centrální limitní věta říká, že výsledné rozdělení je také přibližně normální.



- De Moivre–Laplaceho věta říká o něco silnější věc: v okolí pn je i pravděpodobnostní funkce $Bin(n,p)$ dobře aproximována hustotou $N(np, np(1-p))$, tedy

vhodně přeškálovanou Gaussovou funkcí φ. Přesněji:

$$\binom{n}{k} p^k (1-p)^{n-k} \simeq \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{(k-np)^2}{2np(1-p)}}$$

pro k blízké pn. Ještě přesněji: pokud pro nějaké c je $|k-pn| < c\sqrt{np(1-p)}$ a n se blíží nekonečnu, tak poměr dvou výrazů nahoře se blíží k jedné.

> Order in Apparent Chaos: I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the Law of Frequency of Error. The law would have been personified by the Greeks and deified, if they had known of it.
>
> Francis Galton: Natural Inheritance (1889)

11 Statistika – úvod

Doposud jsme se věnovali tomu, že pro zadaný model náhodné situace (např. dané náhodné veličiny, jejich distribuci danou (sdruženou) pravděpodobnostní funkcí/hustotou), jsme zjišťovali „co se stane“ – jaká je pravděpodobnost nějakého jevu, jaká je střední hodnota nějaké veličiny, atd. Nyní náš pohled otočíme: budeme chtít z pozorovaných dat odvodit, jaký model je asi vytvořil, případně jaké měl vlastnosti. Řešené úlohy si rozdělíme do několika typů:

- bodové odhady – chceme určit hodnotu nějakého parametru (často střední hodnoty neznámého rozdělení)
- intervalové odhady – chceme pro nějaký parametr najít interval, ve kterém bude ležet např. s pravděpodobností 99 %.
- testování hypotéz – chceme zjistit, zda naměřená data podporují naši hypotézu, nebo ji vylučují.
- (lineární) regrese – jaká je závislost mezi dvěma měřenými hodnotami?

Každé z těchto otázek se budeme věnovat podrobněji, ukažme si jen na příkladu, co to znamená. Řekněme, že chceme zjistit, kolik je v České republice leváků. V principu bychom mohli tuto otázku zjistit při příštím sčítání lidu, ale rádi bychom našli operativnější a levnější postup. Nabízí se použít náhodný vzorek: z celé populace ($N \doteq 10.7$ milionů) vybereme náhodný vzorek o rozsahu (velikosti) $n = 100$ (např.). U každého z nich zjistíme, zda je levák a odsud odvodíme něco o počtu leváků v celé ČR. Pokud máme za úkol vytvořit *bodový odhad*, tak najdeme jedno číslo, které bude nejlepší možný odhad. (Co to vlastně znamená si ukážeme v příští kapitole.) *Intervalový odhad* má skromnější úkol: najdeme interval, ve kterém bude neznámý počet s dost velkou pravděpodobností. Ukážeme si, jaký je optimální vztah mezi šířkou nalezeného intervalu a požadovanou spolehlivostí. Myslíte si, že je mezi programátory více

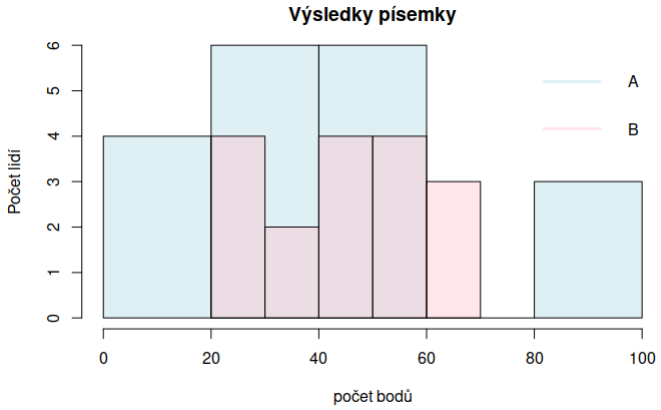
leváků než v celé populaci? Pokud to chcete ověřit, řešíte úlohu *testování hypotéz*. Pokud si myslíte, že je závislost mezi počtem programátorů v nějaké zemi a výší HDP na osobu, tak pro ověření potřebujete provést *regresi*.

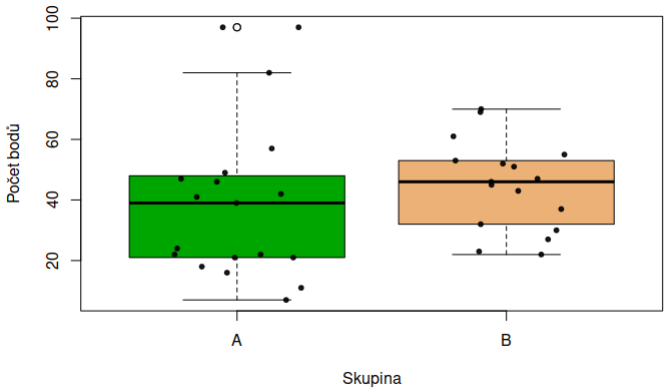
Než se ale pustíme do konkrétních metod a vět, podívejme se na několik tváří statistiky. Ta první (kterou nejčastěji uvidíme třeba v novinách) spočívá prostě v přehledném znázornění nějakých dat. Této části se říká *exploratční analýza dat*. Nebudeme si k ní říkat nic podnětného – kromě toho, že je potřeba pečlivost a poctivost. Možné neduhy:

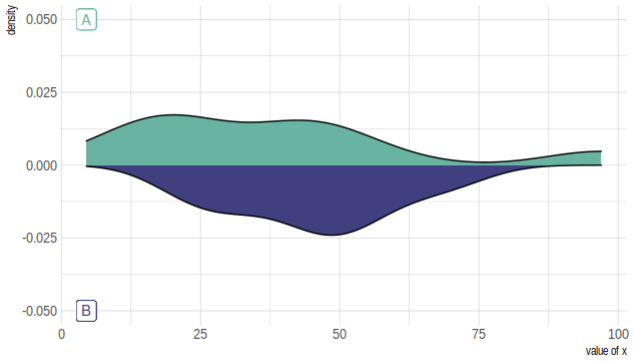
- co dělat s neúplnými daty (člověk co odpoví jenom na půlku otázek, student odejde v půlce písemky, firma zkrachuje v průběhu roku, pacient vypadne v půlce studi a vy nevíte, jestli se uzdravil nebo umřel, u některých programů jste nezměřili dobu běhu, protože se vám už nechtělo déle čekat).
- co dělat s daty, která jsou asi chybná
- pokud vybíráme jen část dat (náhodný vzorek) – vybíráme rovnoměrně náhodně? Nebo podle toho, co je snáze dostupné, nebo co se nám víc hodí?
- nejsou zvolené obrázky zavádějící?
- atd., atd.

Ukažme si zde jen tři možné způsoby znázornění stejných dat: dvě skupiny studentů píšící písemku.

TODO: vysvětlit boxplot TODO: u histogramu záleží na jemnosti!



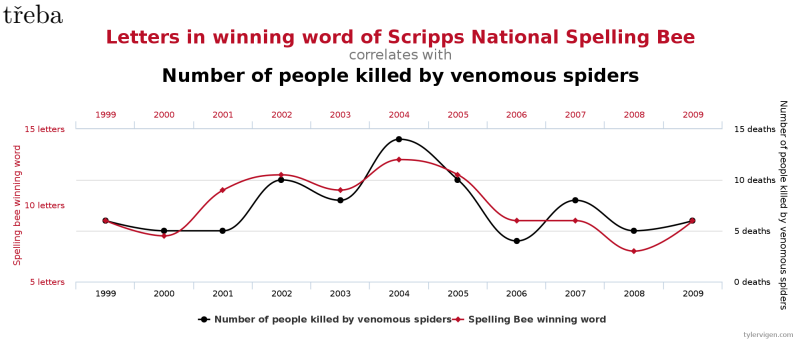




TODO: radar chart, stem chart, scatter plot, bubble chart

TODO: nominální proměnná (také kategorická), pořadové/ordinální, numerická

Explorační analýza může objevit zajímavé souvislosti. Problém ale je, jak zjistit, které jsou popis nějaké skryté skutečnosti a které jsou způsobeny náhodným kolísáním dat. Když budeme mít dat dostatečný počet, objevíme v nich cokoli, viz třeba



Tím se právě zabývá statistika ve smyslu aplikované matematické disciplíny;

budeme zkoumat tak zvanou *konfirmační analýzu*. Jejím základem je způsob, jak vybíráme z velké množiny (tzv. populace) objekty, které budeme měřit nebo jinak zkoumat. Logický postup je vybírat *uniformně náhodně bez vracení*. Jako jednodušší pro analýzu se ukazuje vybírat *uniformně náhodně s vracením*.

TODO: binom vs. hypergeom.

TODO: stratifikovaný výběr

Motivování předchozí diskuzí definujeme následující pojem:

Náhodný výběr Posloupnost n.n.v. $X_1, \ldots, X_n$ se stejným rozdělením nazýváme *náhodný výběr s rozsahem n*. Pokud mají tyto veličiny distribuční funkci F, píšeme $X_1, \ldots, X_n \sim F$. Někdy místo F píšeme krátce jen jméno rozdělení ($Ber(p)$, $Exp(\lambda)$, ...).

Uvědomme si, že ve skutečnosti tyto náhodné veličiny nevidíme, získáme jen jejich hodnoty pro naše konkrétní měření, neboli pozorování, neboli *realizaci*. Zpravidla značíme x_i naměřenou hodnotu X_i, neboli $x_i = X_i(\omega)$ pro elementární jev $\omega \in \Omega$, který skutečně nastal.

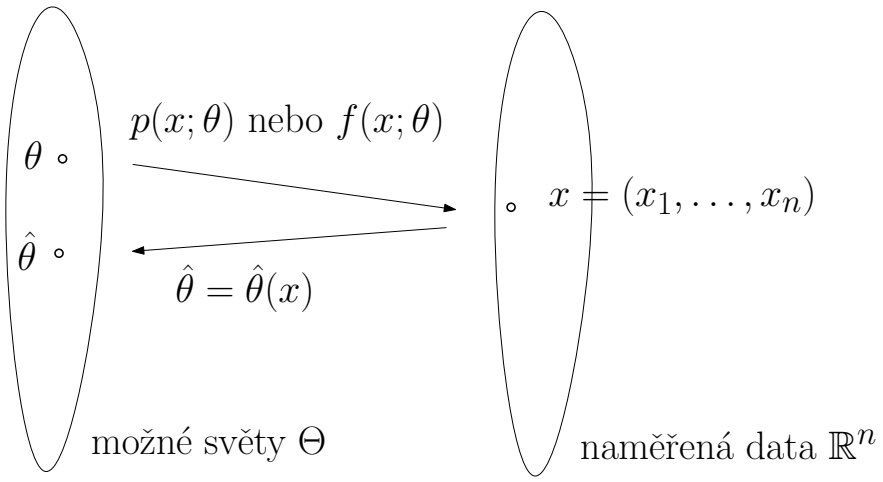
Prostor Ω nebudeme blíže zkoumat, zde si jej ale pro ujasnění ukážeme podrobně. Pokud stále počítáme podíl leváků, tak $\Omega = \{$všechny n-tice obyvatel ČR$\}$. Uvažovaná pravděpodobnost je uniformní a pro $\omega = (\omega_1, \ldots, \omega_n)$ zvolíme $X_i = I(\omega_i$ je levák$)$.

Parametrické vs. neparametrické modely V našem pátrání po modelu, který vysvětluje naměřená data, se můžeme rozhodnout, že budeme velice otevření – povolíme libovolnou distribuční funkci neznámé veličiny, nebo se třeba omezíme na diskrétní (nebo naopak spojité) veličiny. Nebo naopak můžeme použít tzv. *parametrické modely* a budeme zkoumat jen distribuční funkce F z množiny $\{F_\theta : \theta \in \Theta\}$ – θ je neznámý parametr, Θ množina možných hodnot toho parametru. Příklady:

- $Pois(\lambda)$ (parametr $\theta = \lambda$, $\Theta = \mathbb{R}^+$)
- $U(a, b)$ (parametr $\theta = (a, b)$, $\Theta = \mathbb{R}^2$)
- $N(\mu, \sigma^2)$ (parametr $\theta = (\mu, \sigma)$, $\Theta = \mathbb{R} \times \mathbb{R}^+$)

Nabízí se otázka, proč to děláme, odkud víme, že neznámá náhodná veličina má zrovna tenhle typ rozdělení. Typicky to děláme na základě předchozích zkušeností: buď zkoumání podobných problémů jinými lidmi, nebo naší vlastní explorační analýzy. Je k tomu ale třeba přidat komentář statistika George Boxe: „Všechny modely jsou špatné, ale některé jsou užitečné.“

Statistika Kromě toho, že statistika je název oboru, má i jistý technický význam. *Statistika* je libovolná funkce náhodného výběru, tj. např. aritmetický průměr, medián, maximum, atd. Takže $T = T(X_1, \ldots, X_n)$ je náhodná veličina – taková, která svoji náhodnost čerpá jen z veličin X_1, ..., X_n (při zpracování dat už neházíme kostkou). Dá se tedy říct, že úlohou statistiků je hledat vhodné statistiky, které řeší úlohy výše zmíněné (např. odhad neznámého parametru).



Obrázek 1: Ilustrace bodového odhadu. Naším úkolem je z naměřených dat (na obrázku jeden z puntíků v množině napravo) odvodit hodnotu parametru θ, který popisuje model, z něhož jsme data naměřili. (Např. skutečnou střední hodnotu, kterou bychom zjistili až po nekonečně mnoha opakováních.) V bodovém odhadu sestavíme jednu statistiku $\hat{\theta}$, která z naměřených dat x vyrobí náš odhad.

12 Bodové odhady

TODO: má θ popisovat jzn. model, nebo to má být něco, co odhadujeme?

Uvažujme nadále náhodný výběr $X_1, \ldots, X_n \sim F_\theta$, budeme chtít odhadnout nějakou funkci neznámého parametru θ. Můžeme zkoumat například veličiny s rozdělením $N(\mu, \sigma^2)$ a chtít odhadnout jejich rozptyl. Pak bude $\theta = (\mu, \sigma)$ a $g(\theta) = \sigma^2$. Ale i u rozdělení s jediným parametrem (např. $Geom(p)$) můžeme chtít místo parametru $\theta = p$ odhadovat střední hodnotu $1/\theta$.

Nyní se zaměříme na odhad pomocí jednoho čísla, neboli *bodový odhad*. Hledáme tedy statistiku $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$, která „dobře aproximuje“ hodnotu $g(\theta)$. Protože náš odhad je náhodná veličina, nemůžeme čekat, že by se kýžené hodnotě skutečně rovnal. Co tedy můžeme realisticky od bodových odhadů čekat? Jak porovnat, který odhad je lepší a který horší?

TODO $g(\theta)$???

Definice 82. *Pro náhodný výběr $X_1, \ldots, X_n \sim F_\theta$ a libovolnou funkci g nazveme bodový odhad $\hat{\theta}_n$*

- nevychýlený/nestranný (angl. unbiased) *pokud* $\mathbb{E}(\hat{\theta}_n) = g(\theta)$.
- asymptoticky nevychýlený/nestranný *pokud* $\lim_{n\to\infty} \mathbb{E}(\hat{\theta}_n) = g(\theta)$.
- konzistentní *pokud* $\hat{\theta}_n \xrightarrow{P} g(\theta)$. *(Definice konvergence v pravděpodobnosti je stejná jako ve Větě 79.)*

Dále definujeme

- vychýlení (bias) $\text{bias}(\hat{\theta}_n) = \mathbb{E}(\hat{\theta}_n) - \theta$.

- střední kvadratickou chybu (mean squared error) $\text{MSE}(\hat{\theta}_n) = \mathbb{E}((\hat{\theta}_n - \theta)^2)$

Uvědomme si, že všechny tyto pojmy závisí na neznámém parametru θ.

- Odhad může být nevychýlený a přitom k ničemu: pokud třeba v polovině případů dává hodnotu o milion vyšší a v polovině případů o milion nižší.
- To částečně řeší konzistence: ta říká, že pro rostoucí n pravděpodobnost velké chyby klesá k nule.
- Pokud nás zajímá nejen chování v limitě, ale i pro malá n, potřebujeme např. MSE. Pokud pro veličinu $Y = (\hat{\theta}_n - \theta)^2$ použijeme Markovovu nerovnost, zjistíme, že $P(Y > c\,\text{MSE}) < 1/c$. Takže pokud je MSE malé, máme i odhad na pravděpodobnost toho, že zrovna při našem měření vyjde odchylka od správné hodnoty velká.
- Další věta nám ukáže, jak docílit malé MSE: potřebujeme malé vychýlení i malý rozptyl našeho odhadu. Proto někdy může být vhodné uvažovat i (trochu) vychýlené odhady, pokud mají nižší rozptyl.

Obrázek 2 ilustruje zavedené pojmy. Pro náhodný výběr z $U(0, \theta)$ odhadujeme hodnotu neznámého parametru θ (na obrázku $\theta = 130$). Protože náš odhad $\hat{\theta}_n$ závisí na náhodných měřeních, jedná se o náhodnou veličinu. Na obrázku je histogram této náhodné veličiny (vždy 10^4 pokusů). Na horním obrázku je histogram symetrický kolem skutečné hodnoty θ, tj. zkoumáme nestranný odhad. Dokonce se zdá, že je rozdělení tohoto odhadu přibližně normální, což se bude hodit v další kapitole. Naopak na dolním obrázku je histogram silně nesymetrický, vždy dává nižší výsledek, než je skutečnost. Jedná se tedy o vychýlený odhad. V obou případech se histogram zúžil při zvýšení počtu měření z $n = 100$ na $n = 1000$. Pokud tento trend pokračuje i nadále pro $n \to \infty$, jedná se o konzistentní odhad (pokud se zároveň zužuje k té správné hodnotě). Odhad pro popis velikosti odchylky každého z bodových odhadů použijeme střední kvadratickou odchylku MSE. Z obrázku je patrné, že odhad na spodním obrázku má nižší MSE.

Zatím jsme neprozradili, jak tyto odhady z naměřených dat získáme, cílem bylo jen ilustrovat, jaké vlastnosti od odhadu můžeme chtít. Tento příklad jsme řešili/budete řešit na cvičeních, horní obrázek popisuje odhad získaný *metodou momentů*, spodní *metodou maximální věrohodnosti* – viz další kapitolky.

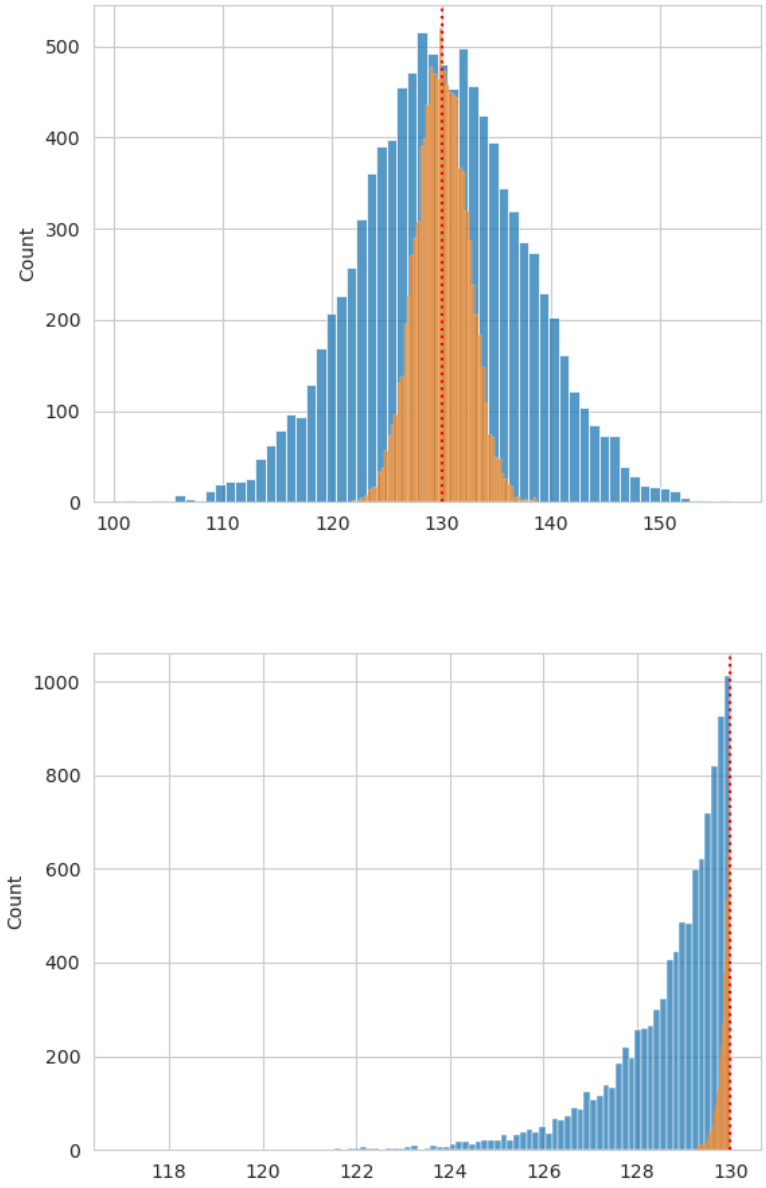
TODO: Porovnat dva odhady pro $\mathbb{E}(X)$: jedno měření a průměr.

Věta 83. $\text{MSE}(\hat{\theta}_n) = \text{bias}(\hat{\theta}_n)^2 + \text{var}(\hat{\theta}_n)$

Důkaz. Podle Věty 35 je $\text{var}(\hat{\theta}_n) = \text{var}(\hat{\theta}_n - \theta)$. To můžeme podle Věty 34 napsat jako

$$\mathbb{E}\big((\hat{\theta}_n - \theta)^2\big) - \big(\mathbb{E}(\hat{\theta}_n - \theta)\big)^2.$$

V předchozím vzorci je první člen $\text{MSE}(\hat{\theta}_n)$ a druhý $\text{bias}(\hat{\theta})^2$. □



Obrázek 2: Ilustrace bodového odhadu. Modrý histogram znázorňuje $\hat{\theta}_{100}$, neboli na výpočtu pomocí vzorku $n = 100$ měření. Oranžový histogram $\hat{\theta}_{1000}$, neboli $n = 1000$ měření. Červená čára znázorňuje skutečnou hodnotu θ. Na spodním obrázku totéž pro jinou metodu odhadu. (Který odhad je lepší?)

Výběrový průměr a rozptyl Ukažme si nyní několik konkrétních odhadů a podívejme se na jejich vlastnosti. Pokud x je proměnná typu numpy.array, můžeme použít k výpočtu příkazy vpravo.

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{výběrový průměr} \quad \text{x.mean()}$$

$$\bar{S}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \qquad \text{výběrový rozptyl} \quad \text{x.var(ddof=0)}$$

$$\widehat{S}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \qquad \text{výběrový rozptyl} \quad \text{x.var(ddof=1)}$$

Dvě varianty výběrového rozptylu jsou spojeny vztahem $\widehat{S}_n^2 = \frac{n}{n-1}\bar{S}_n^2$, zlomek $\frac{n}{n-1}$ se nazývá Besselova korekce.

Pro ilustraci malý příklad. Mějme X_1, X_2 náhodný výběr z uniformního rozdělení na množině $\{1, 2, 3\}$. Střední hodnota, v tomto kontextu též zvaná *populační průměr* je $\frac{1+2+3}{3} = 2$. V tabulce níže to odpovídá průměru prvního (a také druhého) sloupce.

Oproti tomu $\bar{X}_2$ je průměr naměřených hodnot. tj. $\frac{X_1+X_2}{2}$, tzv. *výběrový průměr* (průměr z náhodného výběru). Asi nepřekvapí, že průměrná hodnota ve třetím slupci, neboli $\mathbb{E}(\bar{X}_2)$ je také 2, věta níže nám to říká obecně – $\bar{X}_n$ je nestranný odhad.

O něco záludnější je odhad rozptylu. Známá definice rozptylu jako střední hodnoty $(X - \mathbb{E}(X))^2$ je v tabulce ukázána v posledním sloupci, jeho průměr je $\mathrm{var}(X_1) = \mathrm{var}(X_2)$, tzv. *populační rozptyl*. Oproti tomu výběrový rozptyl sčítá odchylky jednotlivých měření od výběrového průměru. (Protože $n - 1 = 1$, tak tady je $\bar{S}_2$ opravdu jen součet.)

V tabulce každý řádek popisuje jedno z možných měření: naměříme-li hodnoty $x_1 = 2, x_2 = 3$, odhadneme populační průměr výběrovým průměrem, tj. hodnotou $\bar{x} = 2.5$ a populační rozptyl výběrovým rozptylem $\bar{s}^2 = 0.5$.

X_1	X_2	$\bar{X}_2$	$(X_1 - \bar{X}_2)^2$	$(X_2 - \bar{X}_2)^2$	$\widehat{S}_2$	$(X_1 - E(X_1))^2$
1	1	1	0	0	0	1
1	2	$\frac{3}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	1
1	3	2	1	1	2	1
2	1	$\frac{3}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	0
2	2	2	0	0	0	0
2	3	$\frac{5}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	0
3	1	2	1	1	2	1
3	2	$\frac{5}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	1
3	3	3	0	0	0	1
2	2	2	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{2}{3}$

Věta 84. *Mějme náhodný výběr z rozdělení se střední hodnotou μ a rozptylem σ^2, přičemž μ i σ jsou reálná čísla (ne nekonečno). Pak*

1. *$\bar{X}_n$ je konzistentní nestranný odhad μ*
2. *$\bar{S}_n^2$ je konzistentní asymptoticky nestranný odhad σ^2*
3. *$\widehat{S}_n^2$ je konzistentní nestranný odhad σ^2*

Důkaz. 1. Zákon velkých čísel (Věta 79) nám dává konzistenci. Při jeho důkazu jsme spočítali střední hodnotu $\bar{X}_n$, takže jsme zjistili, že odhad je konzistentní.

2., 3. Důkaz konzistence vynecháme. Spočteme ale $\mathbb{E}(\bar{S}_n^2)$, abychom zjistili, jestli máme spíše používat $\bar{S}_n^2$ nebo $\widehat{S}_n^2$.

Označme $V = \sum_{i=1}^n (X_i - \bar{X}_n)^2$ a upravujme

$$\begin{aligned} V &= \sum_{i=1}^n \left(X_i^2 - 2X_i\bar{X}_n + \bar{X}_n^2\right) \\ &= \sum_{i=1}^n X_i^2 - 2\bar{X}_n \sum_{i=1}^n X_i + n\bar{X}_n^2 \end{aligned}$$

Zatím jsme jen použili vzorec pro druhou mocninu a přerovnali sumy. Nyní si uvědomíme, že $\sum_{i=1}^n X_i = n\bar{X}_n$ a získáme $V = \sum_{i=1}^n X_i^2 - n\bar{X}_n^2$.

Použijeme Větu 34, tj. $\mathbb{E}(Y^2) = \operatorname{var}(Y) + \mathbb{E}(Y)^2$. Výpočet z Věty 79 nám dá střední hodnotu a rozptyl $\bar{X}$. Takže

$$\mathbb{E}(X_i^2) = \sigma^2 + \mu^2 \qquad \text{a také}$$

$$\mathbb{E}(\bar{X}_n) = \frac{\sigma^2}{n} + \mu^2$$

Dohromady:

$$\mathbb{E}(V) = n(\sigma^2 + \mu^2) - n\Big(\frac{\sigma^2}{n} + \mu^2\Big) = (n-1)\sigma^2.$$

Z toho snadno plyne $\mathbb{E}(\widehat{S}_n^2) = \sigma^2$ (takže $\widehat{S}_n^2$ je nestranný odhad) a $\mathbb{E}(\bar{S}_n^2) = \frac{n}{n-1}\sigma^2$. Odhad $\bar{S}_n^2$ tedy není nestranný, ale jen asymptoticky nestranný (podíl $\frac{n}{n-1}$ má limitu 1). □

Z předchozí věty to vypadá, že $\widehat{S}_n^2$ je jasně lepší. Pro doplnění poznamenejme, že (např.) pro normální rozdělení je $\mathrm{MSE}(\bar{S}_n^2) < \mathrm{MSE}(\widehat{S}_n^2)$. To lze spočíst pomocí Věty 83 nebo nasimulovat na počítači:

```
v1 = np.array([stats.norm.rvs(size=5).var() for _ in range(10**3)])
v2 = v1*5/4
```

```
v1.mean(), v2.mean()
```

```
(0.8046533318454767, 1.0058166648068458)
```

```
((v1 - 1)**2).mean(), ((v2 - 1)**2).mean()
```

```
(0.3677967492918393, 0.515090753171922)
```

12.1 Metoda momentů

Podívejme se nyní na obecnou metodu, jak vytvořit bodové odhady pomocí takzvaných momentů. Připomeňme, že r-tý moment náhodné veličiny X je $\mathbb{E}(X^r)$ a zaveďme následující značení:

- $m_r(\theta) := \mathbb{E}(X^r)$ pro $X \sim F_\theta$ … r-tý (populační) moment
  Tj. děláme průměr z celé populace.
- $\widehat{m_r} := \frac{1}{n}\sum_{i=1}^n X_i^r$ … r-tý výběrový moment.
  Tj. děláme průměr z náhodného výběru $X_1, \ldots, X_n$ z F_θ.

Jinými slovy: $m_1(\theta)$ popisuje závislost $\mathbb{E}(X)$ na parametru θ, kterým rozdělení popisujeme. Pro každé θ je to konstanta. Naopak $\widehat{m_1}$ je náhodná veličina – výběrový průměr. Jeho vlastnosti také závisí na θ, ale to teď není důležité.

Věta 85. *$\widehat{m_r}$ je nestranný konzistentní odhad pro $m_r(\theta)$.*

Důkaz. Stačí použít Větu 84 pro veličiny $X_1^r, \ldots, X_n^r$. □

To motivuje následující postup:

Metoda momentů: Volíme takové θ, které řeší soustavu rovnic

$$m_r(\theta) = \widehat{m_r} \qquad r = 1, \ldots, k.$$

(Typicky k bude počet reálných čísel, která tvoří parametr θ.)

Mohli bychom zde uvádět věty, popisující kdy tento postup vytvoří konzistentní odhady (často) resp. nestranné (ne vždy). Ukažme si místo toho příklady užití:

Příklad 86. *Pokračujme v našem příkladu s počtem leváků: X_i bude 1 pokud je i-tý člověk levák, jinak 0. Takže $X_1, \dots, X_n$ je náhodný výběr z distribuce $Ber(\theta)$, kde θ je (neznámý) podíl leváků v populaci. Určete θ metodou momentů.*

Máme $m_1(\theta) = \theta$ a $\widehat{m_1} = \bar{X}_n$. Odvodili jsme tedy nepřekvapivý závěr, že je dobré zvolit jako náš bodový odhad statistiku $\hat{\theta}_n = \bar{X}_n$.

Příklad 87. *Nyní je $X_1, \dots, X_n$ je náhodný výběr z distribuce $N(\mu, \sigma^2)$ (například budeme měřit výšku náhodně vybraných lidí). Zde je tedy parametr $\theta = (\mu, \sigma)$. Určete θ metodou momentů.*

Tady je první moment $m_1(\theta) = \mu$. Dále druhý moment je

$$m_2(\theta) = \mathbb{E}(X^2) = \mathrm{var}(X) + \mathbb{E}(X)^2 = \sigma^2 + \mu^2.$$

Metoda momentů nám říká, že máme položit

$$\begin{aligned} \mu &= m_1(\theta) = \widehat{m_1} = \bar{X}_n \\ \sigma^2 + \mu^2 &= m_2(\theta) = \widehat{m_2} = \frac{X_1^2 + \dots + X_n^2}{n} \end{aligned}$$

Řešením je tedy $\hat{\Theta}_n = (\hat{\mu}, \hat{\sigma})$, kde

$$\begin{aligned} \hat{\mu} &= \bar{X}_n \\ \hat{\sigma} &= \sqrt{\frac{X_1^2 + \dots + X_n^2}{n} - \bar{X}_n^2}. \end{aligned}$$

Po krátké úpravě (resp. pohledem na důkaz Věty 84) zjistíme, že jsme dostali odhad $\bar{S}_n = \sqrt{\bar{S}_n^2}$, tj. „ten nesprávný odhad s $\frac{1}{n}$“.

12.2 Metoda maximální věrohodnosti (maximal likelihood, ML)

Nyní prozkoumejme další metodu tvorby odhadů. Někdy nám obě metody dají stejný odhad, někdy jiný – pak si vybereme, který nám více vyhovuje, například srovnáním velikosti MSE.

Uvažme opět náhodný výběr $X = (X_1, \dots, X_n)$ z modelu s parametrem θ. Buď $x = (x_1, \dots, x_n)$ některý možný výsledek, tzv. *realizace* náhodného výběru X. (To znamená, že pro nějaké $\omega \in \Omega$ je $x_i = X_i(\omega)$ pro $i = 1, \dots, n$.) Pokud se jedná o diskrétní náhodné veličiny, tak x má nějakou sdruženou pravděpodobnost

$$p_X(x; \theta) = \prod_{i=1}^{n} p_{X_i}(x_i; \theta).$$

Parametr θ píšeme jako speciální argument oddělený středníkem, abychom naznačili, že jde o principiálně jiný vstup do p_X než jsou naměřené hodnoty x. Vzorec pro p_X má tvar součinu, protože $X_1, \ldots, X_n$ jsou nezávislé náhodné veličiny.) Pokud zkoumáme spojité náhodné veličiny, tak x má přiřazenu sdruženou hustotu

$$f_X(x;\theta) = \prod_{i=1}^{n} f_{X_i}(x_i;\theta).$$

Abychom mohli oba případy diskutovat najednou, definujeme *věrohodnost (likelihood)* $L(x;\theta)$ jako $p_X(x;\theta)$ nebo $f_X(x;\theta)$, podle toho, jaký je typ náhodných veličin.

Uvědomme si rozdíl oproti obvyklé situaci v první částí přednášky: měli jsme zadané θ a pro různá x jsme určovali hodnotu $L(x;\theta)$. Nyní máme naopak pevné x – je to ten soubor hodnot, které jsme naměřili. Považujeme tedy $L(x;\theta)$ za funkci θ a hledáme vhodnou hodnotu θ.

Metoda MV (ML): volíme takové θ, pro které je $L(x;\theta)$ maximální.

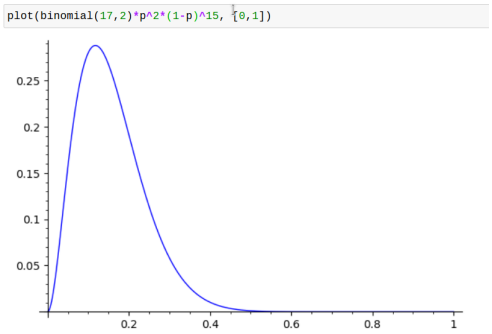
Výpočetně je často jednodušší místo hledání maxima $L(x;\theta)$ hledat maximum *logaritmické věrohodnosti* $\ell(x;\theta) = \log L(x;\theta)$ – a vzhledem k monotonii logaritmu najdeme stejný výsledek.

Příklad 88. *Opět budeme počítat leváky: X_i bude 1 pokud je i-tý člověk levák, jinak 0. Takže $X_1, \ldots, X_n$ je náhodný výběr z distribuce $Ber(\theta)$, kde θ je (neznámý) podíl leváků v populaci. Určete θ metodou maximální věrohodnosti.*

Předpokládejme, že ve vybraném vzorku bylo přesně k leváků, pro konkrétnost $x_1 = \cdots = x_k = 1$, $x_{k+1} = \cdots = x_n = 0$. Pořád předpokládáme, že každý člověk má nezávisle na ostatních pravděpodobnost θ, že je levák. Pravděpodobnost pozorovaného měření je tedy

$$\begin{aligned} L(x;\theta) &= \theta^k(1-\theta)^{n-k} && \text{neboli} \\ \ell(x;\theta) &= k\log\theta + (n-k)\log(1-\theta) && \text{a tedy} \\ \ell'(x;\theta) &= \frac{k}{\theta} - \frac{n-k}{1-\theta} \end{aligned}$$

Chceme najít θ, pro které bude $L(x;\theta)$ maximální, což je totéž, jako že bude $\ell(x;\theta)$ maximální. Podle základů analýzy je to buď na kraji intervalu (ale tam je L nulové a ℓ nedefinované) nebo v bodě, kde je derivace podle θ nulová. Platí tedy $\frac{k}{\theta} = \frac{n-k}{1-\theta}$ a odsud snadno $\theta = k/n$. Neboli, dospěli jsme opět k bodovému odhadu $\hat{\theta} = \bar{X}_n$.



Příklad 89. *Nyní je $X_1, \ldots, X_n$ je náhodný výběr z distribuce $N(\mu, \sigma^2)$ (například budeme měřit výšku náhodně vybraných lidí). Zde je tedy parametr $\theta = (\mu, \sigma)$. Určete θ metodou maximální věrohodnosti.*

Tady se jedná o spojité zobrazení, budeme tedy pracovat s hustotou $f = f_{X_i}$. Platí $f(x_i; \theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i-\mu)^2/(2\sigma^2)}$, takže

$$\ell(x_i; \theta) = -\frac{(x_i - \mu)^2}{2\sigma^2} - \log\sigma - \log\sqrt{2\pi} \qquad \text{a tedy}$$
$$\ell(x; \theta) = -\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2} - n\log\sigma - n\log\sqrt{2\pi}$$

V bodě kde funkce $\ell(x; \theta)$ nabývá maxima jsou derivace podle μ i podle σ rovny nule (definiční obor je otevřená množina, neboli nemá žádné „krajní hodnoty“). Je tedy

$$\frac{\partial}{\partial\mu}\ell(x; \theta) = \sum_{i=1}^{n} \frac{2(x_i - \mu)}{2\sigma^2} = 0 \qquad \text{a také}$$
$$\frac{\partial}{\partial\sigma}\ell(x; \theta) = \sum_{i=1}^{n} \frac{2(x_i - \mu)^2}{2\sigma^3} - \frac{n}{\sigma} = 0$$

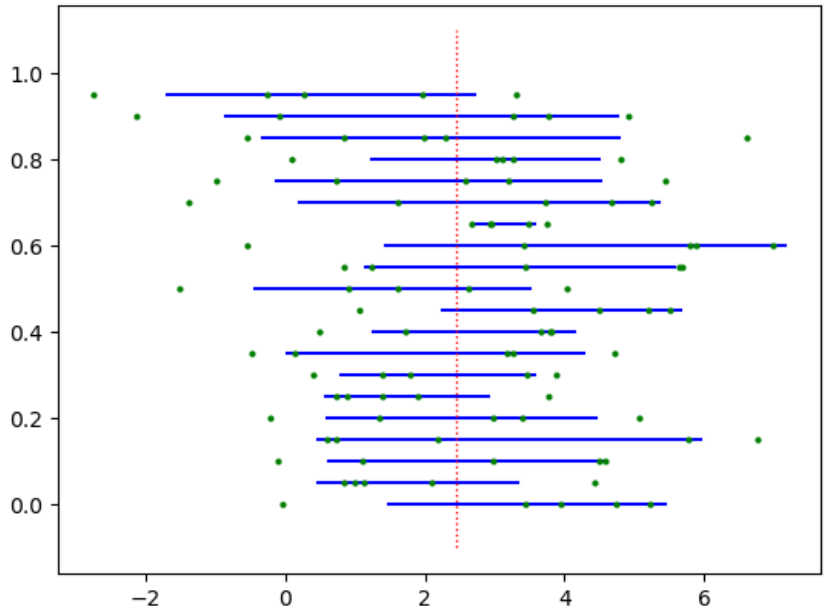
Z první rovnice dostaváme ihned $\hat{\mu} = \bar{X}_n$. Z druhé pak $\hat{\sigma}^2 = \bar{S}_n$.

13 Intervalové odhady

Nechť $D \leq H$ jsou dvě statistiky. Říkáme, že určují *intervalový odhad, též interval spolehlivosti, též konfidenční interval o spolehlivosti* $1-\alpha$ anglicky *$(1-\alpha)$-confidence interval*, zkráceně jen $(1-\alpha)$-CI, pokud

$$P(D \leq \theta \leq H) = 1 - \alpha.$$

Číslo α se nazývá koeficient spolehlivosti. Někdy také uvažujeme jednostranné intervalové odhady (případ $D = -\infty$, resp. $H = +\infty$). Můžeme také požadovat



Obrázek 3: Ilustrace intervalových odhadů. Každý řádek odpovídá jednomu statistickému šetření, ve kterém jsme provedli několik měření (zelené puntíky) a odhadli střední hodnotu pomocí modrého intervalu. Cílem je mít co nejpřesnější odhad (krátký interval), ale zároveň takový, aby většinou v předpovězeném intervalu ležela na správná hodnota (znázorněná červenou čárou).

silnější vlastnosti: $P(\theta > H) = P(\theta < D) = \alpha/2$. Než pokročíme dále, uvědomte si, co je v uvedených výrazech náhodné! Přepokládáme, že θ je parametr – jeho hodnota je nám sice neznámá, ale je to nějaká dobře definovaná vlastnost reality (třeba skutečný počet leváků v ČR). Náhodnost je v hraničních bodech intervalu: hodnoty D, H jsou náhodné veličiny. To, jaký interval řekneme jako naši aproximaci reality závisí na tom, jak dopadnou naše měření – neboli na realizaci náhodného výběru $X_1, \ldots, X_n$.

Hledaný interval budeme typicky uvažovat ve tvaru $[x-\delta, x+\delta]$ (pro vhodně určené x a δ). Tento interval pro zkrácení zapisujeme jako $x \pm \delta$.

Většina našich metod bude založena na následujícím přístupu:

- Máme nestranný bodový odhad $\hat{\theta}$ pro parametr θ
- $\hat{\theta}$ má normální rozdělení
- Pak $\hat{\theta} \pm \mathbf{z_{\alpha/2}} \cdot \mathbf{se}$ je $(1-\alpha)$-CI, kde

  $z_{\alpha/2} := \Phi^{-1}(1-\alpha/2)$,
  $\mathbf{se} := \sigma(\hat{\theta})$,

Tady $z_{\alpha/2}$ je zkratka za $\Phi^{-1}(1-\alpha/2)$, neboli je to takové číslo, aby platilo $P(Z \leq z_{\alpha/2}) = 1 - \alpha/2$.

Standardní chyba (standard error) je definováná jako $\mathbf{se} = \sigma(\hat{\theta})$, tj. je to směrodatná odchylka (standard deviation) náhodné veličiny $\hat{\theta}$.

Důkaz. Proveďme standardizaci $Z = \text{stand}(\hat{\theta}) = \frac{\hat{\theta}-\theta}{\sigma(\hat{\theta})}$. Z vlastností normálního rozdělení víme, že $Z \sim N(0,1)$, můžeme tedy pro jeho popis použít distribuční funkci Φ — nezávisle na hodnotě θ, **se**.

Z volby čísla $z_{\alpha/2}$ víme, že $P(Z \leq z_{\alpha/2}) = 1 - \alpha/2$, ze symetrie také platí $P(Z \leq -z_{\alpha/2}) = \alpha/2$, tudíž

$$P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha.$$

Když dosadíme definici Z a ekvivalentně upravíme nerovnici, zjistíme, že také

$$P(\hat{\theta} - z_{\alpha/2} \cdot \sigma(\hat{\theta}) \leq \theta \leq \hat{\theta} + z_{\alpha/2} \cdot \sigma(\hat{\theta})) = 1 - \alpha.$$

□

13.1 Intervalové odhady normální n.v. se známým rozptylem

Normální veličiny $N(\theta, \sigma^2)$ Máme náhodný výběr $X_1, \ldots, X_n \sim N(\theta, \sigma^2)$. Přitom θ neznáme, ale σ ano. (Realistický příklad je teploměr se známou chybou měření, kterým měříme neznámou teplotu.) Z Věty 84 víme, že $\bar{X}_n$ je nevychýlený odhad. Z kapitoly o normálním rozdělení víme, že $\bar{X}_n$ má normální rozdělení. Z důkazu Věty 79 víme, že $\mathbf{se} = \sigma(\bar{X}) = \sigma/\sqrt{n}$.

Závěr je, že $\bar{X} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ je $(1-\alpha)$-CI.

Jako konkrétní příklad: pokud naměříme hodnoty TODO $z_{0.025} = 1.96$.

Naše první statistická úloha bude vlastně cvičení na práci s normální veličinou. Větu formulujeme obecně, ale jako realistický příklad si můžeme představit fyzikální měření – měříme hmotnost opakovaným měřením na stejné váze, jejíž chyba měření má normální rozdělení se směrodatnou odchylkou σ. Známe σ (je

to vlastnost váhy), hledáme θ, neznámou hmotnost. Měření, která provádíme, tedy budou realizací náhodné veličiny s rozdělením $N(\theta, \sigma^2)$.

Věta 90 (Intervalové odhady normální náhodné veličiny)**.** *Nechť $X_1, \ldots, X_n$ je náhodný výběr z $N(\theta, \sigma^2)$. σ známe, θ chceme určit. Zvolíme $\alpha \in (0,1)$. Nechť $\Phi(z_{\alpha/2}) = 1 - \alpha/2$. Označme $S_n = (X_1 + \cdots + X_n)/n$ (tzv. výběrový průměr) a*

$$C_n := [S_n - z_{\alpha/2} \cdot \sigma/\sqrt{n}, S_n + z_{\alpha/2} \cdot \sigma/\sqrt{n}]$$

Pak $P(C_n \ni \theta) = 1 - \alpha$.

Důkaz. Označme $Y_n = \frac{S_n - \theta}{\sigma/\sqrt{n}}$. Uvědomme si, že $Y_n \sim N(0,1)$: spočteme $\mathbb{E}(Y_n) = 0$ a $\mathrm{var}(Y_n) = 1$ (stejně jako v důkazu Věty 79). Dále použijeme odolnost normálního rozdělení vůči součtu. Zbývá si uvědomit, že

$$P(\theta > S_n + z_{\alpha/2}\sigma/\sqrt{n}) = P(Y_n < -z_{\alpha/2}) = \Phi(-z_{\alpha/2}) = \alpha/2.$$

□

Libovolná veličina ze známým rozptylem σ^2 Nyní postoupíme k obecnějšímu případu: pokud máme dost sčítanců (abychom mohli použít centrální limitní větu), můžeme vytvořit intervalový odhad zcela stejně jako pro normální veličinu – nebude to ale odhad s přesně požadovanou pravděpodobností chyby, nýbrž tato pravděpodobnost bude mít jen požadovanou limitu.

Závěr je, že $\bar{X} \pm z_{\alpha/2} \cdot \sigma/\sqrt{n}$ je asymptoticky $(1-\alpha)$-CI, neboli

$$\lim_{n\to\infty} P(\bar{X}_n - z_{\alpha/2} \cdot \sigma/\sqrt{n} \le \theta \le \bar{X}_n + z_{\alpha/2} \cdot \sigma/\sqrt{n}) = 1 - \alpha.$$

Věta 91 (Intervalové odhady pomocí CLV)**.** *Nechť $X_1, \ldots, X_n$ je náhodný výběr z nějakého rozdělení se střední hodnotou θ a rozptylem σ^2. Opět σ známe, θ chceme určit a zvolíme $\alpha \in (0,1)$. Zvolíme $z_{\alpha/2}$, S_n a C_n jako v předchozí větě. Pak*

$$\lim_{n\to\infty} P(C_n \ni \theta) = 1 - \alpha.$$

Důkaz. Označíme Y_n stejně jako v přechozí větě a všimneme si, že CLV říká, že $Y_n \xrightarrow{d} N(0,1)$. Proto platí

$$P(\theta > S_n + z_{\alpha/2}\sigma/\sqrt{n}) = P(Y_n < -z_{\alpha/2}) \xrightarrow{n\to\infty} \Phi(-z_{\alpha/2}) = \alpha/2.$$

□

Dvouvýběrový test Zatím jsme se bavili o 1-výběrových testech (angl. 1-sample test), tj. měli jsme jednu sadu dat. Často ale potřebujeme porovnat dvě různé sady dat. Mějme nyní $X_1, \ldots, X_n$ náhodný výběr z rozdělení se střední hodnotou μ_X a rozptylem σ_X^2, a $Y_1, \ldots, Y_m$ náhodný výběr z rozdělení se střední hodnotou μ_Y a rozptylem σ_Y^2. TODO: realistický příklad

Chceme najít odhad pro $\theta = \mu_X - \mu_Y$. Je přirozené použít bodový odhad $\hat{\theta} = \bar{X}_n - \bar{Y}_m$, víme už, že se jedná o nevychýlený odhad. Pokud jsou všechny měřené hodnoty normální, tak je i $\hat{\theta}$ normální. I pokud nejsou, tak $\hat{\theta}$ bude přibližně normální díky centrální limitní větě (Věta 81). Zbývá nám spočíst standardní chybu. Protože $\bar{X}_n \perp \bar{Y}_m$ (přepokládáme nezávislost všech měření), tak

$$\mathrm{var}(\hat{\theta}) = \mathrm{var}(\bar{X}_n) + \mathrm{var}(\bar{Y}_m) = \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}$$

takže $\mathbf{se} = \sqrt{\mathrm{var}(\hat{\theta})} = \sqrt{\sigma_X^2/n + \sigma_Y^2/m}$.

Nyní už víme vše pro náš intervalový odhad o spolehlivosti $1 - \alpha$:

$$\bar{X}_n - \bar{Y}_m \pm z_{\alpha/2} \cdot \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}$$

13.2 Intervalové odhady normální n.v. s neznámým rozptylem

Odhad pravděpodobnosti Vraťme se k příkladu s odhadováním počtu leváků. Neboli: máme náhodný výběr $X_1, \ldots, X_n \sim Ber(p)$. V minulé kapitole jsme si několika způsoby nalezli odhad $\hat{p} = \bar{X}_n$ a ukázali, že je nestranný. Abychom nalezli intervalový odhad, potřebujeme zjistit standardní chybu, neboli směrodatnou odchylku $\bar{X}_n$. Z vlastností Bernoulliho rozdělení a vět o rozptylu víme, že

$$\mathbf{se} = \sigma(\bar{X}_n) = \sqrt{\frac{p(1-p)}{n}}.$$

V tomto vzorci ovšem neznáme skutečnou hodnotu p. Použijeme tzv. *plugin estimate*: odhadneme **se** pomocí vzorce nahoře, kde místo p dosadíme $\hat{p}$, neboli $\bar{X}_n$.

Jako přibližný $(1-\alpha)$-CI použijeme interval

$$\bar{X} \pm z_{\alpha/2} \cdot \widehat{\mathbf{se}},$$

kde

$$\widehat{\mathbf{se}} := \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}.$$

Normální veličiny s neznámým rozptylem – Studentův test Nechť $X_1, \ldots, X_n$ je náhodný výběr z $N(\mu, \sigma^2)$. Chceme odhadnout μ stejným principem jako dosud, tj. uvážit bodový odhad $\hat{\mu} = \bar{X}_n$ (nestranný a s normálním rozdělením). Jeho směrodatná odchylka je $\sigma(\hat{\mu}) = \sigma/\sqrt{n}$, kde ale σ tentokrát neznáme. Použijeme dříve odvozený odhad

$$\hat{\sigma} = \widehat{S}_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}.$$

A odsud

$$\widehat{\mathbf{se}} = \sigma(\hat{\mu}) = \frac{\widehat{S}_n}{\sqrt{n}}.$$

Nyní ale narazíme na problém: po úpravě jako dříve

$$T_n = \frac{\bar{X} - \mu}{\widehat{S}_n/\sqrt{n}}$$

nebude mít T normální rozdělení – nedělíme konstantou, ale náhodnou veličinou. Dobrá zpráva je, že to tolik navadí. Podstatné na dřívějším postupu bylo, že veličina Z měla stejné rozdělení nezávisle na neznámém μ a (známém) σ. Nová veličina T má opět rozdělení nezávislé na neznámém μ a neznámém σ. Toto rozdělení už není normální, říká se mu Studentovo rozdělení[3], podrobněji *Studentovo t-rozdělení s $n-1$ stupni volnosti*. Hustota Studentova rozdělení má přiměřeně hezký vzorec, který ale nebudeme potřebovat, jen si všimneme toho, že to je sudá funkce (pokud by bylo $\mu = 0$ a my změnili znaménko u každé X_i, dostali bychom bod se stejnou hustotou pravděpodobnosti díky symetrii normálního rozdělení). Distribuční funkce Studentova rozdělení je označovaná $\Psi_{n-1}(x)$. Má hodnoty v tabulkách a zejména v knihovních funkcích – pt(x,n−1) v R, scipy.stats.t.cdf(x,n−1) v pythonu).

Všimněme si, že vzorec pro Z a T se liší jen v tom, jestli že se místo skutečného σ napíše jeho konzistentní odhad $\widehat{S}_n$. Odsud lze nahlédnout (detaily vynecháváme), že $T_n \xrightarrow{d} Z$, neboli pro každé x platí $\lim_{n\to\infty} \Psi_{n-1}(x) = \Phi(x)$.

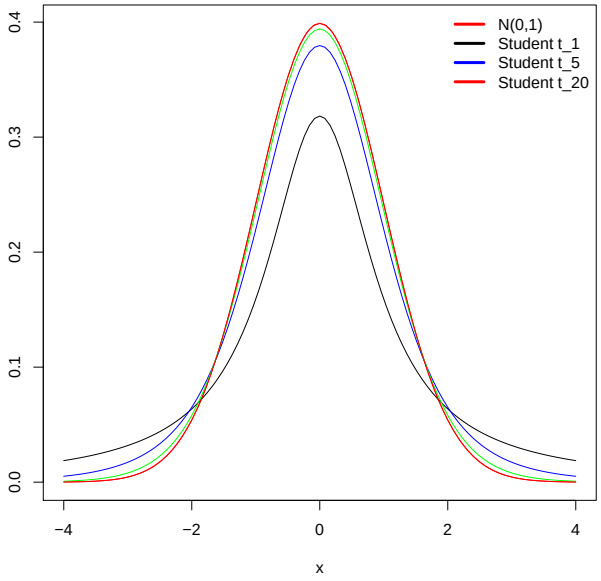
My nebudeme ale potřebovat ani tak Ψ_{n-1}, ale její inverzi (kvantilovou funkci), viz tabulka níže (všimněte si extrémních hodnot v první řádce: pokud máme dvě měření, tak už můžeme o rozptylu něco usuzovat, ale očividně dost nepřesně). Pomocí Φ^{-1} jsme definovali $z_{\alpha/2}$, analogicky zavedeme zkratku $t_{\alpha/2} := \Psi^{-1}_{n-1}(1-\alpha/2)$ (n známe z kontextu, ve značení se neuvádí).

$\hat{\theta} \pm \mathbf{t}_{\alpha/\mathbf{2}} \cdot \widehat{\mathbf{se}}$ je $(1-\alpha)$-CI, kde

$t_{\alpha/2} := \Psi^{-1}_{n-1}(1-\alpha/2)$,
$\widehat{\mathbf{se}} := \widehat{S}_n/\sqrt{n}$,

[3]pojmenované podle Williama Gosseta, statistika a hlavního experimentálního sládka v pivovaru Guinness, kterému jeho zaměstnavatel nechtěl dovolit publikovat vědecké práce pod pravým jménem

p	0.9	0.95	0.975	0.99	0.995
$\Psi_1^{-1}(p)$	3.08	6.31	12.71	31.82	63.66
$\Psi_2^{-1}(p)$	1.89	2.92	4.3	6.96	9.92
$\Psi_4^{-1}(p)$	1.53	2.13	2.78	3.75	4.6
$\Psi_8^{-1}(p)$	1.4	1.86	2.31	2.9	3.36
$\Psi_{30}^{-1}(p)$	1.31	1.7	2.04	2.46	2.75
$\Phi^{-1}(p)$	1.28	1.64	1.96	2.33	2.58



Věta 92. *Nechť $X_1, \ldots, X_n$ je náhodný výběr z $N(\mu, \sigma^2)$. Parametr $\theta = (\mu, \sigma)$ neznáme, chceme ale určit jen μ. Opět máme $\alpha \in (0,1)$. Nechť $\Psi_{n-1}(t_{\alpha/2}) = 1 - \alpha/2$. Pak*

$$\bar{X}_n \pm t_{\alpha/2} \frac{\widehat{S}_n}{\sqrt{n}} \quad \textit{je } (1-\alpha)\textit{-CI pro } \mu.$$

Důkaz. Podobně jako předchozí intervalové odhady: uvědomíme si, že pro $\delta = t_{\alpha/2} \frac{\widehat{S}_n}{\sqrt{n}}$.

$$\begin{aligned} \bar{X}_n - \delta < \mu < \bar{X}_n + \delta \qquad & \text{je totéž jako} \\ -\delta < \bar{X}_n - \mu < \delta \qquad & \text{neboli} \\ -t_{\alpha/2} < \frac{\bar{X}_n - \mu}{\widehat{S}_n/\sqrt{n}} < t_{\alpha/2} & \end{aligned}$$

Nyní zbývá použít definice $t_{\alpha/2}$ a symetrie, díky níž je $\Psi_{n-1}(-t_{\alpha/2}) = \alpha/2$. □

14 Testování hypotéz

V této kapitole si ukážeme základy statistického testování hypotéz, které tvoří základ velké části dnešní vědy. Jako úvodní ilustraci si uveďme tři příklady.

Příklad 93. *Rozhodněte, zda se (v dané zemi) rodí stejně chlapců a dívek.*

Napřed si ujasněme, co znamená otázka: pokud máme přístup ke všem informacím o narozeních, tak můžeme daný počet přesně spočítat. Ale je nutné předpokládat, že počty budou v průběhu času náhodně kolísat – naším úkolem je poznat, jestli naměřená data jsou vysvětlitelná náhodným kolísáním, nebo zda se detí nějakého pohlaví rodí víc.

Tuto úlohu řešil v roce 1710 skotský polyhistor John Arbuthnot. Jeho postup bychom dnes nazvali *znaménkový test*. Pokud se v jednom roce narodí více chapců, napíšeme si +, pokud více dívek, zapíšeme −. Pokud je počet stejný, zapíšeme 0 (a daný rok nebudeme počítat). Johnu Arbuthnotovi vyšlo 82 plusů z 82 dvou zkoumaných let, usoudil tedy, že se chlapců rodí více.[4] Takový výsledek bychom náhodou (tj. za předpokladu, že se chlapců a dívek rodí stejně) dostali s pravděpodobností cca $1/2^{82}$, tedy skoro nulovou. (K zamyšlení: kdyby poměr byl více vyvážený, třeba 50 plusů a 32 minusů, jak byste rozhodli? Jak určit správnou hraniici?)

Příklad 94. *TODO: pití čaje*

Příklad 95. *Z tisíce hodů korunou nám padl orel* 472*-krát. Máme minci (nebo toho, kdo s ní hází) podezírat z podjatosti?*

Jako rozumné vypadá stanovit nějakou hranici x a pokud bude O, počet orlů, menší než x, tak zahlásit, že je mince falešná. (Nebo možná budeme protestovat i pokud bude $O > 1000 - x$ – záleží na úhlu pohledu.) Výběr x nám umožňuje balancovat mezi dvěma možnými chybami: pokud zvolíme x příliš malé, tak poznáme jen hodně falešné mince. Pokud naopak bude x moc blízko k 500, tak budeme varovat před falešnou mincí bezdůvodně, při každé „náhodné oscilaci“. Obvyklý přístup k tomuto dilamatu je ten, že napřed stanovíme tzv. hladinu spolehlivosti/významnosti α a pak x vybereme tak, aby chybné zamítnutí (tj. chybné varování před falešnou mincí) mělo pravděpodobnost rovnou (nebo blízkou) α. Volba α závisí na oboru TODO.

Další možné otázky, které můžeme pomocí statistického testovní hypotéz zkoumat: Má vylepšený program kratší dobu běhu než původní? Je léčba nemoci metodou X dobrá? (Lepší než placebo, lepší než metoda Y, ...) Jsou leváci lepší boxeři než praváci?

Jak k takovým otázkám obecně přistupovat? Popíšeme si metodu prosazovanou statistikem R.A. Fisherem.

[4] Poměr je typicky kolem 1.05 chlapců na jednu dívku, ale závisí na mnoha okolnostech.

- Uvážíme dvě hypotézy: H_0, H_1
- H_0, tzv. *nulová hypotéza (null hypothesis)*, značí defaultní, konzervativní model (lék nefunguje, mince je spravedlivá)
- H_1, tzv. *alternativní hypotéza (alternative hypothesis)* – značí alternativní model „zajímavost“
- výstup našeho testování je, že buď nulovou hypotézu H_0 zamítneme (data nás „donutí“ myslet si, že náš popis světa není správný), nebo nezamítneme (tj. pozorovaná data jsou s nulovou hypotézou v souladu).

Protože všechna naše pozorování jsou výsledek náhodného procesu, nemůžeme nic tvrdit s jistotou. Můžeme se tedy dopustit dvou typů chyb:

- Chyba 1. druhu: chybné zamítnutí. Zamítneme H_0, i když platí. „Trapas.“
- Chyba 2. druhu: chybné přijetí. Nezamítneme H_0, ale ona neplatí. „Promarněná příležitost.“

Postupujeme tedy takto:

- Vybereme vhodný statistický model.
- Volíme *hladinu významnosti (significance level)* α. Test budeme konstruovat tak, aby pravděpodobnost chyby 1. druhu, tj. chybného zamítnutí H_0, byla α. Typicky $\alpha = 0.05$.
- Určíme *testovou statistiku* $T = h(X_1, \ldots, X_n)$, kterou budeme určovat z naměřených dat.
- Určíme *kritický obor (rejection region)* – množinu W.
- Naměříme hodnoty $x_1, \ldots, x_n$, realizace náh. veličin $X_1, \ldots, X_n$ – až teď, po určení všech detailů testu. Jinak bychom byli v pokušení vybrat z několika možných testů ten, který nám dává správné výsledky.
- Rozhodovací pravidlo: zamítneme H_0 pokud $t = h(x_1, \ldots, x_n) \in W$.
- Je tedy $\alpha = P(h(X) \in W; H_0)$.
- Dále označíme $\beta = P(h(X) \notin W; H_1)$ pravděpodobnost chyby 2. druhu. Hodnotu $1 - \beta$ označujeme jako *sílu testu*.

Výše uvedený postup dává jen dva výstupy: ano/ne. Můžeme ale chtít z naměřených dat zjistit něco více. Často používaný postup je výpočet tzv. p-hodnoty: minimální α, pro které bychom H_0 zamítli. Jinak řečeno, je to pravděpodobnost, že naměříme data, která jsou alespoň „tak špatná“ jako ta, co vidíme.

Všimněte si, že nepíšeme $P(h(X) \in W|H_0)$. Středník v zápise má naznačit, že počítáme pravděpodobnost ve světě, kde platí hypotéza H_0. Protože ale to, v jakém žijeme světě (a jaké hypotézy ho popisují) nepovažujeme za náhodný jev, nemůžeme zde použít značení podmíněné pravděpodobnosti. Toto se změní v tzv. Bayesovské statistice, ke které se ale tento semestr nedostaneme.

Příklad 96. *Nalili nám správnou míru? Nebo přesněji: je střední hodnota taková, jaká je v nápojovém lístku?*

- $X_1, \ldots, X_n$ *náhodný výběr z* $N(\theta, \sigma^2)$
- *Předpokládejme, že* σ^2 *známe ze zkušenosti.*
- $H_0 : \theta = \mu$ *(např.* $\mu = 0.5\ \ell$*)* $\qquad H_1 : \theta \neq \mu$

Postup: volíme statistiku $Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$. Víme, že $Z \sim N(0,1)$, *pokud platí* H_0. Proto pokud zvolíme $W = \{x : |x| > z_{\alpha/2}\}$, kde opět $z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2)$, tak bude pravděpodobnost chyby 1. druhu přesně α. (Pravděpodobnost chyby 2. druhu, a tedy síla testu, závisí na tom, jak je θ blízké k μ.)

Toto je příklad tzv. *jednovýběrového testu (one-sample test)* – vybíráme z jedné populace, testujeme její parametr. Koncepčně jiný je tzv. *dvouvýběrový test (two-sample test)*, kde vybíráme ze dvou populací a testujeme, zda se jejich statistické vlastnosti liší.

Příklad 97. *Nalili nám stejně jako kamarádovi? Nebo přesněji: je střední hodnota stejná pro nás oba?*

- $X_1, \ldots, X_n$ *náhodný výběr z* $N(\theta_X, \sigma^2)$
- $Y_1, \ldots, Y_m$ *náhodný výběr z* $N(\theta_Y, \sigma^2)$
- *Předpokládejme stále, že* σ^2 *známe.*
- $H_0 : \theta_X = \theta_Y \qquad H_1 : \theta_X \neq \theta_Y$

Označme $S = \bar{X}_n - \bar{Y}_m$. Opět jistě platí $\mathbb{E}(S) = 0$, *pokud platí* H_0. Spočteme nyní rozptyl:

$$\operatorname{var}(S) = \operatorname{var}(\bar{X}_n) + \operatorname{var}(\bar{Y}_m) = \sigma^2/n + \sigma^2/m.$$

Označíme tedy

$$Z = \frac{\bar{X}_n - \bar{Y}_m}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}}.$$

Podle předchozího výpočtu je $Z \sim N(0,1)$. Můžeme tedy volit W stejně jako v předchozím případu.

Test, který jsme vytvořili, se nazývá Z-test.

Co si počít s nešťastným předpokladem znalosti rozptylu? Použijeme výběrový rozptyl, neboli bodový odhad rozptylu z naměřených data. V případě jednovýběrového testu tedy označíme $\widehat{S}_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ a

$$T = \frac{\bar{X}_n - \mu}{\widehat{S}_n/\sqrt{n}}.$$

Pokud je totiž $\widehat{S}_n^2$ odhad rozptylu jednotivých veličin, tak jejich průměr má rozptyl n-krát menší, proto ve jmenovateli vystupuje $\widehat{S}_n/\sqrt{n}$. Jak už jsme si říkali

u intervalových odhadů, rozdělení veličiny T je známé, je to tzv. Studentovo t-rozdělení. Zejména jeho parametry nezávisí na μ ani na σ. Upravíme tedy kritický obor na

$$W = \{x : |x| > t_{\alpha/2}\}, \qquad \text{kde } t_{\alpha/2} = \Psi_{n-1}^{-1}(1 - \alpha/2).$$

TODO: Co v případě dvouvýběrového testu?

Upravený test, se nazývá T-test (a používá se častěji než Z-test, protože předpoklad znalosti σ^2 je trochu umělý). Historicky se tyto testy používaly s tabukami funkcí Φ a Ψ_{n-1}. Dnes místo tabulek obvykle použijeme knihovní funkce ve vhodném softwaru, nebo spíš rovnou funkci, které předhodíme vstupní data a ona spočítá, zda hypotézu zamítnout, resp. jaká je p-hodnota. V Pythonu tak použijeme scipy.stats.ttest_1samp() nebo scipy.stats.ttest_ind(), v Rku t.test().

TODO párové testy!

TODO: Ilustrace s čísly na computeru

- $X_1, \ldots, X_{n_1}$ náhodný výběr z $Ber(\theta_X)$
- $Y_1, \ldots, Y_{n_2}$ náhodný výběr z $Ber(\theta_Y)$
- $H_0 : \theta_X = \theta_Y \qquad H_1 : \theta_X \neq \theta_Y$

TODO

p-hacking Lákavý postup je naměřit nějaká data, hledat v nich zajímavosti, a ty pak považovat za prokázanou pravdu. Je ale třeba si uvědomit, že když máme dost dat, tak tam nějaké zajímavosti budou „shodou okolností“ (něco jako Ramseyova věta v diskrétní matematice) – viz obrázek o pavoucích v úvodu statistické kapitoly.

Ještě hůře: můžeme být v pokušení test opakovat, dokud neobjevíme žádaný výsledek. Je ale třeba si uvědomit, že to je ekvivalent toho, když hrajeme „Člověče, nezlob se“ a když nám nepadne šestka, tak se dožadujeme opakovaného hodu.

Správný přístup je usilovat o *reprodukovatelnost* – po explorační analýze dat uděláme nezávislý sběr dat a ten analyzujeme konfirmačně. Případně dopředu náhodně rozdělíme data na část pro tvorbu hypotéz a část pro jejich potvrzení ... jednoduchý případ křížové validace (cross validation). Z pohledu celé vědecké komunity jsou kontrolou reprodukovatelnosti tzv. metaanalýzy — přehled různých pozorování téhož experimentu a souhrn toho, co kolikrát vyšlo.

14.1 Testy dobré shody

Zatím jsme se převážně bavili o měření *numerických dat* – takových, kde dává smysl z naměřených hodnot počítat průměr. Nyní se podíváme na *kategoriální*, taková, kde případné přirazení čísel jednotlivým variantám má jen smysl pomocného kódování: barva očí, politická preference, místo narození. Pak jediné, co můžeme zkoumat, je počet opakování jednotlivých hodnot. Připomeňme si definici multinomického rozdělení.

Definice 98. *Dána $p_1, \ldots, p_k \geq 0$ tak, že $p_1 + p_2 + \cdots + p_k = 1$. n-krát zopakuji pokus, kde může nastat jedna z k možností, i-tá má pravděpodobnost p_i. $X_i :=$ kolikát nastala i-tá možnost $(X_1, \ldots, X_k)$ má multinomické rozdělení s parametry $n, (p_1, \ldots, p_k)$.*

- triviální případ: $X_i =$ počet hodů kostkou, kdy padlo i
- důležitý případ: $X_i =$ počet výskytů i-tého písmene, i-tého slovního druhu, ...
- $P(X_1 = x_1, \ldots, X_k = x_k) = \binom{n}{x_1, \ldots, x_k} p_1^{x_1} \ldots p_k^{x_k} \qquad (*)$

Budeme teď uvažovat situaci, kdy jsou jednotlivé pravděpodobnosti neznámý parametr, budeme je tedy označovat $\theta = (\theta_1, \ldots, \theta_k) \in \Theta$ (množina Θ tedy obsahuje všechny k-tice nezáporných čísel se součtem 1). Vybereme si oblíbenou hodnotu $\theta^* \in \Theta$ a budeme testovat nulovou hypotézu $H_0 : \theta = \theta^*$. Tak jako v předchozí kapitole bude naším cílem stanovit pravidlo, která naměřená data jsou „příliš extrémní“ na to, abychom věřili, že pocházejí z multinomického rozdělení s parametry n, θ^*. I tady to uděláme tak, že vybereme vhodnou statistiku $T = T(X_1, \ldots, X_n)$ a zamítneme pokud bude T příliš velká, tedy bude $W = (\gamma, \infty)$ pro vhodnou hodnotu γ.

Oproti hodu mincí v minulé kapitole, tady není vůbec jasné, jakou T zvolit, neboli jak měřit, že hody kostkou neodpovídají modelu: asi bychom byli nejraději, kdyby $X_i = \theta_i^* n$ – ale co když to neplatí, jak máme odchylky jednotlivých měření kombinovat?

Obecný postup (který se hodí nejen u testů dobré shody) je tzv. *test podílem věrohodností* (likelihood-ratio test). Vycházíme z pojmu věrohodnosti, který jsme už potkali v kapitole o bodových odhadech. Až budeme mít naměřená data $x = (x_1, \ldots, x_k)$, tak každému parametru $\theta \in \Theta$ přiradíme věrohodnost $L(x; \theta)$. Připomeňme, že takhle jsme metodou maximální věrohodnosti hledali takové $\hat{\theta}$, pro které je $L(x; \hat{\theta})$ maximální. Pokud $\hat{\theta} \neq \theta^*$, tak je věrohodnost θ^* menší — pokud bude menší o hodně, tak bychom asi rádi zamítli nulovou hypotézu. Budeme tedy posuzovat podíl $L(x; \hat{\theta})/L(x; \theta^*)$. Z technických důvodů formulku ještě upravíme:

$$G = 2 \log \frac{L(x; \hat{\theta})}{L(x; \theta^*)}.$$

Podívejme se nyní, kam nás tato metoda zavede pro multinomické rozdělení. Podle vzorce $(*)$ je $L(x; \theta) = \prod_{i=1}^{k} \theta_i^{x_i}$. Pro nalezení nejvěrohodnějšího parametru $\hat{\theta}$ můžeme použít postup z matematické analýzy – metodu Lagrangeových multiplikátorů pro nalezení vázaných extrémů. TODO Touto metodou bychom došli k očekávanému výsledku $\hat{\theta} = x/n$. Explicitně, pro každé i je nejvíc věrohodná pravděpodobnost i-tého výsledku $\hat{\theta}_i = x_i/n$, neboli se jedná o „nasamplovanou pravděpodobnost“ získanou naším měřením. (Vsimněte si, že to souhlasí s výpočtem pro $k = 2$ (házení mincí) v kapitole TODO.)

Po dosazení získáváme

$$G = 2\log\prod_{i=1}^{k}\frac{(x_i/n)^{x_i}}{(\theta_i^*)^{x_i}}$$
$$= 2\sum_{i=1}^{k} x_i \log\frac{x_i}{n\theta_i^*}$$

To ještě pro přehlednost přeznačíme – označíme E_i očekávanou (expected) hodnotu X_i a O_i pozorovanou (observed) hodnotu. Je tedy $E_i = n\theta_i^*$ a $O_i = x_i$. V tomto značení máme

$$G = 2\sum_{i=1}^{k} O_i \log\frac{O_i}{E_i}.$$

Tento výraz se dá aproximovat použitím Taylorova polynomu a dostaneme

$$G \doteq \chi^2 = \sum_{i=1}^{k}\frac{(E_i - O_i)^2}{E_i}.$$

Teď máme hned dvě statistiky, které nějakým způsobem měří odklon naměřených dat od ideálního stavu $E_i = O_i$ (v něm bude $G = \chi^2 = 0$) až po extrémní případ, kdy všechny pokusy dopadly stejným výsledkem.

Zbývá poslední krok: stanovit kritický obor, ve kterém budeme nulovou hypotézu zamítat. Tak jako v jiných případech testování hypotéz musíme vyvážit pravděpodobnost chyby I. a II. druhu. Zvolíme tedy α a hledáme takové γ, aby $P(T > \gamma; H_0) = \alpha$ (pro $T = G$ nebo $T = \chi^2$). Jak to udělat? Máme tři možnosti:

- pro malá n můžeme projít postupně všech k^n možných výstupů n pokusů (nebo o něco efektivněji, všech $\binom{n+k-1}{n}$ možných výsledků multinomického rozdělení). Pro každý z nich vyhodnotíme statistiku T a pak vybereme takovou hodnotu γ, aby $T > \gamma$ jen ve 100α procentech případů. Zde budeme znát pravděpodobnost chyby prvního druhu přesně, proto se tomuto postupu říká přesný, nebo též *exaktní* test. Je ale poněkud nepraktický.

- pokud je n větší, můžeme postup podle předchozího bodu simulovat pomocí dostatečně velkého počtu náhodných vzorků. Pro případ s kostkou: hodíme tisíckrát n ideálními kostkami, pokaždé změříme hodnotu T. Jako γ pak zvolíme 51. nejhorší naměřenou hodnotu.

- konečně pro velké n a statistiku χ^2 můžeme použít aproximaci pomocí tzv. χ^2-rozdělení s $k-1$ stupni volnosti. Jde o rozdělení náhodné veličiny $Q = Z_1^2 + \cdots + Z_{k-1}^2$, kde $Z_1, \ldots, Z_{k-1}$ jsou n.n.v. s rozdělením $N(0,1)$. Dá se ukázat, že pro velká n je rozdělení (diskrétní n.v.) statistiky χ^2 (za platnosti hypotézy H_0) blízké rozdělení (spojité n.v.) Q. Budeme tedy volit $\gamma = F_Q^{-1}(1-\alpha)$ a pak bude platit $P(Q > \gamma) \doteq \alpha$.

Příklad 99. *Házíme opakovaně kostkou. Jednotlivá čísla padla s četností 92, 120, 88, 98, 95, 107. Je kostka spravedlivá?*

Máme tedy $n = 600$, $x = (92, 120, 88, 98, 95, 107)$. Testujeme nulovou hypotézu $H_0 : \theta = \theta^* = (1/6, \dots, 1/6)$. Nejvěrohodnější parametr je

$$\hat{\theta} = (92/600, 120/600, 88/600, \dots).$$

Nicméně možná je θ^* také dost věrohodné? Dosadíme do vzorce

$$\begin{aligned}\chi^2 =& \frac{(92-100)^2}{100} + \frac{(120-100)^2}{100} + \frac{(88-100)^2}{100} + \frac{(98-100)^2}{100} \\ &+ \frac{(95-100)^2}{100} + \frac{(107-100)^2}{100} = 6.86\end{aligned}$$

Ke zjištění, zda 6.86 je příliš velká použijeme aproximaci pomocí χ^2 rozdělení s pěti stupni volnosti. Pomocí knihovní funkce **qchisq**(0.95,5) získáme $\gamma \doteq 11.1$. Takže výsledek našeho házení kostkou rozhodně neumožňuje nulovou hypotézu zamítnout. Můžeme i spočítat p-hodnotu pomocí 1−**pchisq**(6.86,5) získáme přibližně 0.23 – takže cca 23 procent hodů kostkou je ještě víc nevyvážených, než ten, který analyzujeme.

Další rozšíření

- Pro zkoumání rozdělení libovolné n.v. Y můžeme vybrat „přihrádky“ $B_1, \dots, B_k$ (rozklad $\mathbb{R}$) a zkoumat, kolikrát je $Y \in B_i$
- Obdobný test pro nezávislost (diskrétních) náhodných veličin

15 Lineární regrese

TODO

TODO regrese jako testování hypotéz

16 Neparametrická statistika

16.1 Empirická distribuční funkce

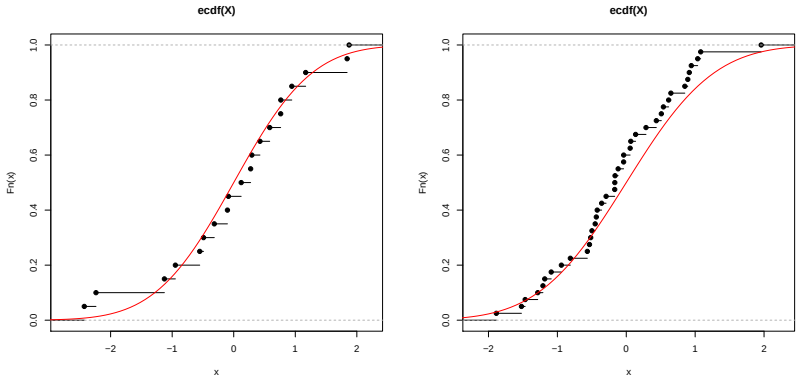
Zatím jsme převážně mluvili o tzv. *parametrické statistice* – uvažovali jsme modely popsané pomocí parametrů, typicky jich nebylo mnoho (například střední hodnota a rozptyl). To mnoho věcí zjednodušuje (mj. nám pak stačí méně dat) – ale nepomůže nám to v situaci, kdy zkoumaná situace takový jednoduchy model nemá.

Pokud tedy nemůžeme hledat parametry, co nám zbývá? Každá náhodná veličina, resp. její distribuce, je přesně popsaná pomocí distribuční funkce. Můžeme se tedy pokusit tuto distribuční funkci aproximovat.

Definice 100. *Nechť $X_1, \dots, X_n \sim F$ je náhodný výběr.* Empirická distribuční funkce (empirical CDF) *je definována*

$$\widehat{F}_n(x) = \frac{\sum_{i=1}^n I(X_i \leq x)}{n},$$

kde $I(X_i \leq x) = 1$ pokud $X_i \leq x$ a 0 jinak.



Věta 101 (Vlastnosti ecdf)**.** *Pro pevné x platí*

- $\mathbb{E}(\widehat{F}_n(x)) = F(x)$
- $\mathrm{var}(\widehat{F}_n(x)) = \frac{F(x)(1-F(x))}{n}$
- *$\widehat{F}_n(x)$ konverguje k $F(x)$ v pravděpodobnosti, píšeme $\widehat{F}_n(x) \xrightarrow{P} F(x)$.*

Důkaz. Slabý zákon velkých čísel. □

Poznámky:

- platí i silnější Glivenko–Cantelliho věta, která říká, že konvergence k nule platí „pro všechna x najednou“:

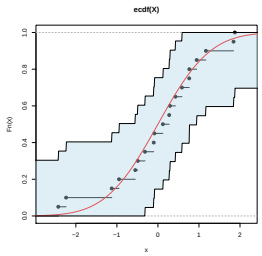
$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F(x) - \widehat{F}_n(x)| = 0 \qquad \text{skoro jistě.}$$

- ještě silnější je následující věta, která dává i efektivní odhad pro rychlost konvergence.

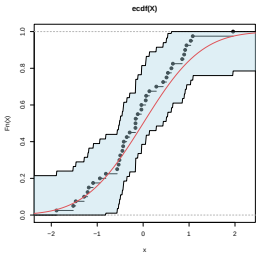
Věta 102 (Dvoretzky-Kiefer-Wolfowitz)**.** *Nechť $X_1, \dots, X_n \sim F$ jsou n.n.v., $\widehat{F}_n$ jejich empirická distribuční funkce. Nechť $\mathbb{E}(X_i)$ je konečná pro každé $i = 1, \dots, n$. Zvolme $\alpha \in (0,1)$ a označme $\varepsilon = \sqrt{\frac{1}{2n} \log \frac{2}{\alpha}}$. Pak platí*

$$P(\widehat{F}_n(x) - \varepsilon \leq F(x) \leq \widehat{F}_n(x) + \varepsilon) \geq 1 - \alpha.$$

TODO: důsledek: KS test (Kolmogorov Smirnov)



(Obrázek vytvořil wiki-editor Sivaji12331, $\alpha = 0.05$.)



16.2 Permutační test

- Máme k dispozici dvě sady nezávislých náhodných veličin (náhodné výběry):
- $X_1, \ldots, X_n \sim F_X$ a $Y_1, \ldots, Y_m \sim F_Y$
- Chceme rozhodnout, zda platí $H_0 : F_X = F_Y$ nebo $H_1 : F_X \neq F_Y$
- Příklady: doba běhu programu před/po vylepšení, hladina cholesterolu u lidí co jedí/nejedí Zázračnou Superpotravu$^{\mathrm{TM}}$, frekvenci krátkých slov v textu autora X a Y.
- Nevíme nic o vlastnostech F_X, F_Y (zejména nečekáme, že je normální)

Postup:

- Zvolíme vhodnou statistiku, např.

$$T(X_1, \ldots, X_n, Y_1, \ldots, Y_m) = |\bar{X}_n - \bar{Y}_m|$$

- $t_{\mathrm{obs}} := T(X_1, \ldots, X_n, Y_1, \ldots, Y_m)$
- Za předpokladu H_0 jsou „všechny permutace stejné“: X_i i Y_j se generovaly ze stejného rozdělení.
- Náhodně zpermutujeme zadaných $m + n$ čísel a pro každou permutaci výčíslíme T – dostaneme $T_1, T_2, \ldots, T_{(m+n)!}$.

- Jako p-hodnotu vezmeme pravděpodobnost, že $T > t_{\text{obs}}$, neboli

$$p = \frac{1}{(m+n)!} \sum_j I(T_j > t_{\text{obs}}).$$

- To je pravděpodobnost chyby 1. druhu, neboli H_0 zamítneme, pokud je $p < \alpha$ (pro naši zvolenou hodnotu α, např. $\alpha = 0.05$).

Vylepšení

- Zkoušet všechny permutace může trvat moc dlouho. Vezmeme tedy jen vhodný počet B nezávisle náhodně vygenerovaných permutací a spočítáme jenom B hodnot $T_1, \ldots, T_B$.
- Jako p-hodnotu vezmeme odhad pravděpodobnost, že $T > t_{\text{obs}}$, neboli

$$\frac{1}{B} \sum_{j=1}^{B} I(T_j > t_{\text{obs}}).$$

- Pro dostatečně velké m, n dává podobné výsledky jako testy založené na CLV, vhodné je tedy zejména pro středně velké počty.

17 Generování n.v.

Základní metodou je tzv. *inverzní samplování (inverse sampling)*, neboli použití Věty 61. (Jméno je odvozeno z toho, že pro spojité n.v. je kvantilová funkce inverzí funkce distribuční.)

Není těžké si rozmyslet, že metoda funguje i pro diskrétní n.v. a dělá přesně to, co by člověk čekal: pokud požadovaná n.v. nabývá hodnoty $x_1, x_2, \ldots$ s pravděpodobnostmi $p_1, p_2, \ldots$, tak rozdělíme interval $[0, 1]$ na intervaly délek $p_1, p_2, \ldots$ a pokud U, tj. uniformní n.v., padne do i-tého intervalu, tak vygenerujeme hodnotu x_i.

Často se ale stane, že neumíme dobře spočítat kvantilovou funkci náhodné veličiny, ale hustotu umíme. Pro tyto případy je vhodná metoda zvaná *zamítací samplování (rejection sampling)*. Je založená na tom, že chceme generovat n.v. X za pomoci n.v. Y, kterou generovat umíme. Přitom X a Y mají podobné rozdělení, v tom smyslu, že pro nějakou konstantu $c > 0$ platí pro všechna reálná t nerovnost $f_X(t) \leq c f_Y(t)$.

K pochopení metody se hodí si připomenout úvodní pohled na hustotu n.v.: pokud $B = (B_1, B_2)$ je náhodný bod pod grafem funkce f_X, tak B_1, první souřadnice bodu B, má rozdělení X. Pokud zafixujeme hodnotu první souřadnice $B_1 = x$, tak platí $0 \leq B_2 \leq f_X(x)$ (protože B je pod křivkou f_X), navíc je B_2 v tomto rozsahu uniformně rozdělený.

1. Vygenerujeme y, u coby realizace n.v. Y s hustotou f_Y, a $U \sim U(0, 1)$.

2. Pokud $u \leq \frac{f_X(y)}{cf_Y(y)}$, tak $X := y$.

3. Jinak hodnotu Y, U zamítneme a opakujeme od bodu 1.

Vysvětlení: pro dané y je $cuf_Y(y)$ rovnoměrně rozděleno na $[0, cf_Y(y)]$. Pokud je tato hodnota menší než $f_X(y)$, tak dvojice $(y, cuf_Y(y))$ je náhodný bod pod grafem f_X, a tedy ji můžeme použít k vygenerování n.v. X. Pokud je $cuf_Y(y) > f_X(y)$, musíme y i u zahodit a začít znovu.

Někdy se též používají speciální úpravy pro jednotlivé druhy náhodných veličin: tzv. rozdělení gamma je součtem několika nezávislých exponenciálních rozdělení. Můžeme ho tedy tak i generovat. Normální rozdělení je vhodné generovat po dvou v polárních souřadnicích, atd. To však uvádíme jen pro zajímavost, k podrobnějšímu zkoumání to necháme specialistům.

18 Seznam značení

- $\{X = x\} = \{\omega \in \Omega : X(\omega) = x\}$
- $p_X(x) = P(X = x) = P(\{X = x\})$ – pravděpodobnostní funkce X
- $p_{X,Y}(x, y) = P(X = x \,\&\, Y = y)$ – sdružená pravděpodobnostní funkce X, Y
- $p_{X|Y}(x|y) = P(X = x \mid Y = y)$ – podmíněná pravděpodobnostní funkce X
- $F_X(x) = P(X \leq x) = P(\{X \leq x\})$ – distribuční funkce X
- $F_{X,Y}(x, y) = P(X \leq x \,\&\, Y \leq y)$ – sdružená distribuční funkce X, Y
- $f_X(x)$ – hustota X
- $f_{X,Y}(x, y)$ – sdružená hustota X, Y
- $f_{X|Y}(x|y) = f_{X,Y}(x, y)/f_Y(y)$ – podmíněná hustota X, Y
- $(a, b), [a, b]$ – otevřený, uzavřený interval
- $\langle a, b \rangle$ – skalární součin

19 Prerekvizity

Aneb co používáme z jiných částí matematiky (zatím jen heslovitě).

- $\sum_{k=0}^{n} q^k = \frac{1-q^{n+1}}{1-q}$
- $\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}$
- $\sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} = (x + y)^n$

- $\sum_{k=0}^{t} \binom{m}{k}\binom{n}{t-k} = \binom{m+n}{t}$
- Definice $\int_a^b f$
- Pokud $f : [a, b] \to [0, \infty)$ je funkce, tak $\int_a^b f(x)dx$ je obsah plochy omezené grafem funkce f na intervalu $[a, b]$. (Pokud tedy ten integrál existuje.)

  TODO: obrázek
- vztah integrálu a derivace

20 Bonusy

Nepřednášelo se, nemusíte umět. Ale třeba to někoho zaujme.

Spojitost pravděpodobnosti

Věta 103. *Nechť pro množiny z prostoru jevů platí*

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$$

a $A = \cup_{i=1}^{\infty} A_i$*. Pak platí*

$$P(A) = \lim_{i \to \infty} P(A_i).$$

- $A_n \subset \{P, O\}^{\mathbb{N}}$, $A_n =$ mezi prvními n hody padl aspoň jednou orel.

Borel–Cantelliho lemma

Věta 104. *Nechť jevy* $A_1, A_2, \ldots$ *splňují* $P(A_i) = p_i > 0$ *pro každé* i*. Označme Nic jev „nenastal žádný z jevů* $\{A_i\}$*“ a Inf jev „nastalo nekonečně mnoho z jevů* $\{A_i\}$*“.*

1. *Pokud* $\sum_i p_i < \infty$*, tak* $P(Inf) = 0$*.*
2. *Pokud jsou jevy* $A_1, A_2, \ldots$ *nezávislé a* $\sum_i p_i = \infty$*, tak* $P(Nic) = 0$*,* $P(Inf) = 1$*.*