

NMAI059 Pravděpodobnost a statistika
Příručka k přednášce.
11. října 2018

Jak používat tuto příručku. Jde o postupně vznikající text, který má obsahovat všechny definice a věty z přednášky. Až na výjimky bude obsahovat jen stručná vysvětlení, doplňující příklady, důkazy nemusí být všechny a úplné. Proto **tato příručka nemůže nahradit účast na přednášce** a není tak zamýšlena.

Definice a věty se snažím pojmenovat a strukturovat co nejvíce. Na přednášce se pokusím tohoto držet, ale zároveň se budu snažit vysvětlovat význam jednotlivých pojmů a jejich používání. Zde by vše mlo být korektně (přestože překlepy se vyskytnou s pravděpodobností 1), zatímco na přednášce občas dojde k chybě (nejsem robot) a případně použiji intuitivní výklad k lepšímu pochopení. Text má sloužit hlavně k tomu, abyste měli k dispozici základní poznámky a ty si poté porovnali s poznámkami z přednášek.

Některé části textu jsou psány *malým písmem*. To jsou části, které pro první čtení představují doplňkovou informaci. Je to proto, že plně matematicky rigorózní výklad by vyžadoval větší rozsah přednášky a více pojmů, než je pro základní pochopení v danou chvíli nutné. Zájemci o matematickou přesnost a důkladnost tak nejsou ochuzeni, tyto poznámky si přečtou a budou o nich přemýšlet, ostatní se k nim vrátí při druhém čtení příručky.

1. AXIOMY; PRAVDĚPODOBNOTNOSTNÍ PROSTOR; NÁHODNÉ JEVY

1.1. Axiomatika. Intuitivní výklad pravděpodobnosti: Nemůžeme předem říci, jak nějaký pokus dopadne, ze zkušenosti víme, že výsledek závisí na mnoha faktorech a nemáme možnost jej přesně vypočítat. Výsledek tedy chápeme jako *náhodný*. Pravděpodobnost je míra „četnosti“: *budeme-li opakovat stejný pokus za stejných podmínek a výsledky jednotlivých pokusů se navzájem neovlivňují, pak podíl jednotlivých výsledků na celkovém počtu pokusů se blíží teoretické pravděpodobnosti.*

Takto o pravděpodobnosti uvažujeme intuitivně. Rigorózní matematický model v roce 1933 zavedl Andrej N. Kolmogorov.

Definice 1 (Pravděpodobnostní prostor). Mějme neprázdnou množinu Ω a na ní systém jevů (podmnožin) \mathcal{F} . Na \mathcal{F} uvažujme pravděpodobnostní míru P , tedy zobrazení $P : \mathcal{F} \rightarrow [0, 1]$ splňující

- (1) $P(\Omega) = 1$
- (2) pro libovolné po dvou disjunktní $A_1, A_2, \dots, A_n \in \mathcal{F}$ platí

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

(konečná aditivita)

- (3) pro libovolné po dvou disjunktní $A_1, A_2, \dots \in \mathcal{F}$ platí

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

(spočetná, neboli σ -aditivita)

Pro tuto chvíli jsme ponechali otázku \mathcal{F} otevřenou. Z definice však vidíme, že je třeba, aby sjednocení po dvou disjunktních množin z \mathcal{F} byla opět množina v \mathcal{F} . Jsou i další požadavky na \mathcal{F} , které splňuje σ -algebra množin.

Definice (σ -algebra). Nechť Ω je neprázdna množina. Třídu podmnožin Ω označená \mathcal{F} nazveme σ -algebrou pokud

- (1) $\emptyset \in \mathcal{F}, \Omega \in \mathcal{F}$.
- (2) $A \in \mathcal{F} \Rightarrow \Omega \setminus A \in \mathcal{F}$.
- (3) $A_1, A_2, \dots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Vidíte, že kromě spočetné aditivity chceme ještě přítomnost doplňku a celé množiny.

Definice 2 (Klasický pravděpodobnostní prostor). Buď Ω neprázdná konečná množina, $\mathcal{F} = 2^\Omega$. Definujeme $P(A) = \frac{|A|}{|\Omega|} \forall A \subset \Omega$. Potom (Ω, \mathcal{F}, P) nazýváme *klasický pravděpodobnostní prostor*.

Definice 3 (Diskrétní pravděpodobnostní prostor). Buď Ω neprázdná konečná nebo spočetná množina, $\mathcal{F} = 2^\Omega$. Buď $p : \omega \rightarrow \mathbb{R}$ taková funkce, že $p(\omega) \in [0, 1] \forall \omega \in \Omega$ a $\sum_{\omega \in \Omega} p(\omega) = 1$. Dále definujeme $P(A) = \sum_{\omega \in A} p(\omega)$. Potom (Ω, \mathcal{F}, P) nazýváme *diskrétní pravděpodobnostní prostor*.

Definice 4 (Reálný spojitý pravděpodobnostní prostor). Buď Ω omezený či neomezený reálný interval, \mathcal{F} nechť obsahuje všechny otevřené i uzavřené podmnožiny Ω a jejich spočetná sjednocení a doplňky. Buď $f : \omega \rightarrow \mathbb{R}$ taková funkce, že $f(\omega) \geq 0$ a $\int_{\Omega} f(\omega) d\omega = 1$. Dále pro $A \in \mathcal{F}$ definujeme $P(A) = \int_A f(\omega) d\omega$. Potom (Ω, \mathcal{F}, P) nazýváme *reálný spojitý pravděpodobnostní prostor*.

V posledním případě může být Ω i podmnožinou \mathbb{R}^k pro $k > 1$. V tom případě je nutné uvažovat násobné integrály. Pro účely tohoto seznámení s pravděpodobností nám budou stačit jen dvounásobné integrály, ale rozšíření není tak složité, jak by se na první pohled mohlo zdát.

1.2. Základní názvosloví a početní věty pro pravděpodobnost.

- $\omega \in \Omega$ se nazývá *elementární jev*. $A \in \mathcal{F}$ se nazývá *náhodný jev*.
- Jestliže $P(A) = 1$, potom A je jev *jistý*.
- Jestliže $P(A) = 0$, potom A je jev *nemožný*.
- Je-li A náhodný jev, pak $A^C = \Omega \setminus A$ je *doplňkový jev* jevu A .

Věta 1 (Základní výpočetní vzorce). Buď P pravděpodobnostní míra na \mathcal{F} , tedy P každému náhodnému jevu A přiřadí jeho pravděpodobnost.

- (1) $P(A^C) = 1 - P(A) \forall A \in \mathcal{F}$
- (2) $A, B \in \mathcal{F} : A \subset B$, pak $P(A) \leq P(B)$ a $P(B \setminus A) = P(B) - P(A)$

Důkaz. Jednotlivé body:

- (1) $P(A \cup A^C) = P(A) + P(A^C)$. Jelikož $A \cup A^C = \Omega$ a podle vlastností pravděpodobnostní míry $P(\Omega) = 1$, $P(A) + P(A^C) = 1$.
- (2) $A \subset B \Rightarrow B = A \cup (B \setminus A)$. To jsou dvě disjunktní množiny. Potom $P(B) = P(A) + P(B \setminus A) \geq P(A)$.

□

Důležitou vlastností pravděpodobnostní míry je její spojitost. My ji využijeme jenom v důkazu spojitosti distribuční funkce zprava, proto je zde uvedena jako poznámka.

Vezmeme si systém $\{A_i\}_{i=1}^\infty \subset \mathcal{F}$ takový, že $A_i \subset A_{i+1}$. Potom pro $A = \bigcup_{i=1}^\infty A_i$ píšeme $A_i \nearrow A$ a říkáme, že systém $\{A_i\}_{i=1}^\infty$ konverguje monotónně (vzhůru) k A .

Analogicky systém, pro který platí $A_i \supset A_{i+1}$ konverguje monotónně (dolů) k $A = \bigcap_{i=1}^\infty A_i$, což zapisujeme $A_i \searrow A$. Uvědomme si, že v obou případech je $A \in \mathcal{F}$, protože \mathcal{F} je σ -algebra.

Věta (Spojitost pravděpodobnostní míry). Nechť máme $\{A_i\}_{i=1}^\infty \subset \mathcal{F}$ takovou, že $A_i \searrow \emptyset$. Pak

$$\lim_{i \rightarrow \infty} P(A_i) = 0.$$

Důkaz. Víme, že $P(\bigcap_{i=1}^\infty A_i) = P(\emptyset) = 0$. To je ekvivalentní s $1 - P((\bigcap_{i=1}^\infty A_i)^C) = 1 - P(\bigcup_{i=1}^\infty A_i^C) = 1$. Dále víme, že $A_i \supseteq A_{i+1}$ a díky tomu $A_i^C \subseteq A_{i+1}^C$, tedy $A_i^C \nearrow \Omega$. Nadefinujeme si posloupnost $B_1 = A_1^C, B_{i+1} = A_{i+1}^C \setminus A_i^C$. Zřejmě B_i jsou disjunktní, $\bigcup_{i=1}^\infty B_i = \Omega$ a $P(\bigcup_{i=1}^\infty B_i) = P(\omega) = 1$ a proto $\sum_{i=1}^\infty P(B_i) = 1$. Poslední součet rozložíme na dva a pro $n \rightarrow \infty$ dostaneme

$$1 = \underbrace{\sum_{i=1}^n P(B_i)}_{\rightarrow 1} + \underbrace{\sum_{i=n+1}^\infty P(B_i)}_{\rightarrow 0}$$

což je důsledkem konečnosti a σ -aditivity pravděpodobnosti. Z definice B_i a z vlastností A_i plyne $P(\bigcup_{i=1}^n B_i) = P(\bigcup_{i=1}^n A_i^C) = 1 - P(\bigcap_{i=1}^n A_i) = 1 - P(A_n)$. Víme, že levá strana konverguje k 1, tedy proto $P(A_n) \rightarrow 0$. □

Věta 2 (Inkluze a exkluze). *Mějme A_1, \dots, A_n náhodné jevy. Pak*

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) + \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) - \dots + (-1)^{n-1} P\left(\bigcap_{i=1}^n A_i\right)$$

Důkaz. Matematickou indukcí.

První krok pro $n = 2$: zřejmě $A = (A \setminus B) \cup (A \cap B)$, $B = (B \setminus A) \cup (A \cap B)$ a proto

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Indukční krok pro $n - 1 \rightarrow n$:

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= P\left(\left(\bigcup_{i=1}^{n-1} A_i\right) \cup A_n\right) = P\left(\bigcup_{i=1}^{n-1} A_i\right) + P(A_n) - P\left(\bigcup_{i=1}^{n-1} (A_i \cap A_n)\right) \\ &= \sum_{i=1}^{n-1} P(A_i) - \sum_{1 \leq i < j \leq n-1} P(A_i \cap A_j) + \dots + (-1)^{n-2} P\left(\bigcap_{i=1}^{n-1} A_i\right) + P(A_n) \\ &\quad - \sum_{1 \leq i \leq j \leq n-1} P(A_i \cap A_n) + \dots + (-1)^{n-2} P\left(\bigcap_{i=1}^n A_i\right) \end{aligned}$$

odkud po vhodném přeuspořádání a sečtení dostaneme

$$\sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) + \dots + (-1)^{n-1} P\left(\bigcap_{i=1}^n A_i\right)$$

□

Než přikročíme k dalším třem významným výpočetním větám, musíme zavést podmíněnou pravděpodobnost.

Řekněme, že jev A nastane s pravděpodobností $P(A)$ nebo nenastane s $P(A^c) = 1 - P(A)$. Nyní dostaneme informaci o tom, že nastal jev B . Pomůže nám tato informace zpřesnit představu o pravděpodobnosti jevu A ? Triviálně ano v těchto případech:

- Jestliže $B \subset A$.
- Jestliže $B \cap A = \emptyset$.

V prvním případě jev A nastat musí, protože víme, že se stal jev B . Ve druhém případě jev A nemůže nastat, jev B jej vylučuje. V ostatních případech to však může být všelijaké.

Definice 5 (Podmíněná pravděpodobnost). Mějme jevy $A, B \in \mathcal{F}$, $P(B) > 0$. Definujeme podmíněnou pravděpodobnost jevu A při B jako $P(A|B) = \frac{P(A \cap B)}{P(B)}$.

Rozmyslete. Ukažte, že podmíněná pravděpodobnost splňuje vlastnosti předepsané pro pravděpodobnosti míry.

Neplatí ale $P(A|B \cup C) \neq P(A|B) + P(A|C)$. Najděte protipříklad.

Podmíněná pravděpodobnost za podmínky B udává pravděpodobnost při dodatečné informaci, že pozorovaný (i když neznámý) elementární jev ω splňuje $\omega \in B$. Všimněte si, že platí $P(A|\Omega) = P(A)$. Intuitivně, znalost faktu $\omega \in \Omega$ je totiž triviální a nedává žádnou další informaci.

Věta 3 (O postupném podmiňování, o násobení pravděpodobností). *Mějme A_1, \dots, A_n náhodné jevy (prvky \mathcal{F}) takové, že $P(\bigcap_{i=1}^n A_i) > 0$. Potom*

$$P\left(\bigcap_{i=1}^n A_i\right) = P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_1 \cap A_2) \cdots P(A_{n-1} | A_1 \cap \dots \cap A_{n-2}) \cdot P(A_n | A_1 \cap \dots \cap A_{n-1}).$$

Důkaz. Matematickou indukcí.

□

Definice 6 (Disjunktní rozklad). Spočetný systém náhodných jevů $\{B_i\}_{i=1}^\infty \subset \mathcal{F}$ nazveme *disjunktním rozkladem* Ω , pokud:

- (1) $B_i \cap B_j = \emptyset \forall i \neq j$

(2) $\bigcup_i B_i = \Omega$ (stačí $P(\bigcup_i B_i) = 1$).

(3) $P(B_i) > 0 \forall i$

Věta 4 (O úplné pravděpodobnosti). *Mějme A náhodný jev a $\{B_i\}$ disjunktní rozklad. Pak*

$$P(A) = \sum_i P(A|B_i) \cdot P(B_i)$$

Důkaz. Víme, že $A \cap B_i, i = 1, 2, \dots$ jsou po dvou disjunktní množiny, proto $\bigcup_i A \cap B_i = A \cap \bigcup_i B_i = A \cap \Omega = A$. Odtud

$$P(A) = P\left(\bigcup_i A \cap B_i\right) = \sum_i P(A \cap B_i) = \sum_i P(A|B_i) \cdot P(B_i)$$

□

Věta 5 (Bayesova). *Mějme A náhodný jev a $\{B_i\}_i$ disjunktní rozklad Ω , nechť také $P(A) > 0$. Potom*

$$P(B_i|A) = \frac{P(A|B_i) \cdot P(B_i)}{\sum_j P(A|B_j) \cdot P(B_j)}.$$

Důkaz. Využijeme definici podmíněné pravděpodobnosti a větu o úplné pravděpodobnosti. □

5.10.2018

Věta 3 nám dává přesný výpočet pravděpodobnosti současného výskytu n náhodných jevů. V praxi však může být výpočet podmíněných pravděpodobností ve větě 3 velmi náročný. V teorii pravděpodobnosti se proto používají i různé odhady pravděpodobností složitých jevů. Uvedeme zde jen jednu z nich.

Věta 6 (Bonferroniho nerovnost). *Mějme A_1, \dots, A_n náhodné jevy. Pak*

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

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

1.3. Nezávislost. Nezávislost (stochastická nezávislost) náhodných jevů znamená, že výskyt jednoho náhodného jevu neovlivní pravděpodobnost výskytu druhého náhodného jevu.

Definice 7 (Nezávislost). *Jevy $A, B \in \mathcal{F}$ jsou nezávislé, pokud $P(A \cap B) = P(A) \cdot P(B)$.*

Tato definice je v souladu s intuitivním popisem nezávislosti. Když jevy A, B jsou nezávislé, potom

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

a naopak, tato rovnost implikuje nezávislost. O nezávislosti však můžeme hovořit i u jevů s nulovou pravděpodobností (které jsme zatím z podmiňování vyloučili).

Nezávislost ale potřebujeme definovat i pro více než dva jevy.

Definice 8 (Vzájemná nezávislost). *Bud'ťe A_1, \dots, A_n náhodné jevy, pak A_i jsou vzájemně nezávislé, pokud $\forall i_1, \dots, i_k \subset \{1, \dots, n\}$ platí*

$$P\left(\bigcap_{j=1}^k A_{i_j}\right) = \prod_{j=1}^k P(A_{i_j}).$$

V této definici si všimněte, že rovnost *musí platit pro všechny podmnožiny indexů*.

Rozmyslete. Najděte trojici náhodných jevů A, B a C takových, že každé dva jevy jsou nezávislé, ale přitom $P(A \cap B \cap C) \neq P(A)P(B)P(C)$.

K tomu stačí uvažovat jen *klasický* pravděpodobnostní prostor s několika elementárními jevy $\omega_1, \dots, \omega_n$ a vhodně volit A, B a C . Jakou nejmenší mohutnost Ω jste ke svému příkladu potřebovali?

Někdy potřebujeme rozšířit pojem nezávislosti na *libovolný* počet náhodných jevů. Pak samozřejmě nemůžeme požadovat platnost definující rovnosti z věty 8 pro nekonečné (zvláště pak i nespočetné) průniky. Nezávislost je ale dobře definována pomocí *všech konečných* indexových podmnožin. Ještě se k tomuto vrátíme u nezávislosti náhodných veličin.

Uvědomte si, jaké (zjednodušující) důsledky má nezávislost na výpočty pravděpodobností—viz výše uvedené výpočetní věty.

2. NÁHODNÉ VELIČINY A VEKTORY

2.1. Náhodná veličina a její rozdělení. Spočítat pravděpodobnost nějakého jevu je jedna věc. Jiný problém je ale pracovat s obecně neznámým, ale náhodným výsledkem. K tomu slouží náhodná veličina (a je třeba si dobře rozmyslet, co znamená).

Definice 9. Mějme (Ω, \mathcal{F}, P) pravděpodobnostní prostor. Zobrazení $X : \Omega \rightarrow \mathbb{R}$ takové, že $X^{-1}(-\infty, a] = \{\omega, X(\omega) \leq a\} \in \mathcal{F} \forall a \in \mathbb{R}$ se nazývá *náhodná veličina*.

Díky definici náhodné veličiny a díky vlastnostem pravděpodobnostní míry P tak dovedeme pro *jakýkoliv* interval říci, s jakou pravděpodobností je hodnota náhodné veličiny v *tomto* intervalu.

Matematicky jde o požadavek *měřitelnosti* zobrazení X . Tím dostáváme do rukou silný nástroj, který nám umožní odvozování nejrůznějších vlastností náhodné veličiny, korektní definice charakteristik a další. Pro základní pochopení to však není nezbytné. Dokud se budeme věnovat diskrétní náhodné veličině, můžeme se bez pojmu měřitelnost obejít (téměř) úplně.

Definice 10 (Rozdělení náhodné veličiny). Buď $X : \Omega \rightarrow \mathbb{R}$ náhodná veličina. Pravděpodobnostní míra P_X definovaná na \mathbb{R} předpisem $P_X(-\infty, a] = P[X \leq a]$ se nazývá *rozdělení náhodné veličiny* X .

Pravděpodobnostní prostor je tak světem, ve kterém se uskutečňuje náhoda a náhodná veličina je modelem, který zviditelňuje náhodu výsledky v reálných číslech. **Na jednom pravděpodobnostním prostoru můžeme definovat více náhodných veličin.** K tomu se dostaneme zanedlouho. Díky náhodné veličině nám vzniká *kanonický* pravděpodobnostní prostor $(\mathbb{R}, \mathcal{B}, P_X)$, kde \mathcal{B} obsahuje všechny otevřené i uzavřené reálné množiny, jejich spočetná sjednocení a doplňky.

Příklad. Mějme

$$\Omega = \{1, 2, 3, 4, 5, 6\}^2, \mathcal{F} = 2^\Omega, P(\{\omega\}) = \frac{1}{36}.$$

Definujme

$$X_1(\omega_1, \omega_2) = \omega_1, X_2(\omega_1, \omega_2) = \omega_2, Y(\omega_1, \omega_2) = \omega_1 + \omega_2.$$

Potom X_1 a X_2 mají stejné rozdělení, ale jsou různé, a X_i a Y mají různá rozdělení. Na jednom pravděpodobnostním prostoru tedy můžeme definovat více různých modelů, případně různých náhodných veličin se stejným rozdělením.

Určete rozdělení X_i a Y .

Pro jaké množiny $A \subset \mathbb{R}$ umíme najít P_X ?

- (1) $(-\infty, a]$
- (2) $(a, b], a < b = (-\infty, b] \setminus (-\infty, a]$
- (3) $(a, b) = \bigcup_{n=1}^{\infty} (a, b - \frac{1}{n}]$
- (4) Všechny otevřené množiny
- (5) $A \in \mathcal{B} \Rightarrow X^{-1} \in \mathcal{F}$ pro všechny $A \in \mathcal{B}$, X je Borelovsky měřitelná

Definice (Náhodné jevy generované X). Mějme X náhodnou veličinu, označme množinu \mathcal{F}_X takovou, že $\mathcal{F}_X = \{B : B = X^{-1}(A) \text{ pro nějakou } A \in \mathcal{B}\}$.

Věta. \mathcal{F}_X je také σ -algebra náhodných jevů generovaných v X .

\mathcal{F}_X je σ -algebra a $\mathcal{F}_X \subset \mathcal{F}$. Potom

$$P_X(A) = P[X \in A] = \underbrace{P(X^{-1}(A))}_{\in \mathcal{F}_X \subset \mathcal{F}}.$$

Na levé straně je míra na $(\mathbb{R}, \mathcal{B})$, na pravé straně je míra na (Ω, \mathcal{F}) .

Z definice 10 je vidět, že k jednoznačnému určení (pravděpodobnostního) rozdělení P_X náhodné veličiny X stačí znát pravděpodobnosti množin typu $(-\infty, a]$ pro všechna $a \in \mathbb{R}$. To nás vede k definici distribuční funkce.

Definice 11. Buď X náhodná veličina a P_X její rozdělení. Funkce $F_X : \mathbb{R} \rightarrow [0, 1]$ definovaná jako $F_X(x) = P_X[-\infty, x] = P[X \leq x]$ se nazývá *distribuční funkce* náhodné veličiny X .

Distribuční funkce F plně popisuje rozdělení P v tom smyslu, že mají-li dvě náhodné veličiny stejnou distribuční funkci, jsou stejně rozdělené.

Věta 7 (Vlastnosti distribuční funkce). *Mějme náhodnou veličinu X a její distribuční funkci F_X . Pak*

- (1) $\lim_{x \rightarrow -\infty} F_X(x) = 0$
- (2) $\lim_{x \rightarrow \infty} F_X(x) = 1$
- (3) F_X je neklesající a zprava spojitá

Důkaz. Plyne z vlastností pravděpodobnostní míry.

K důkazu vlastností distribuční funkce ovšem potřebujeme spojitost pravděpodobnostní míry v prázdné množině a proto se zde omezíme na konstatování, že tomu tak je. Zájemci si najdou poznámku o spojitosti pravděpodobnostní míry výše. \square

Je tomu ale i naopak. Každá funkce, která má vlastnosti distribuční funkce je distribuční funkcí.

Věta 8. *Nechť F splňuje vlastnosti z věty 7. Pak existuje pravděpodobnostní prostor (Ω, \mathcal{F}, P) a náhodná veličina X takové, že F je distribuční funkcí náhodné veličiny X .*

Důkaz. Volme $\Omega = \mathbb{R}$, \mathcal{F} systém obsahující všechny otevřené množiny, jejich spočetná sjednocení i průniky a všechny doplňky, P volme jako míru na \mathcal{F} definovanou předpisem $P(-\infty, x] = F(x)$.

Nyní definujme $X : \Omega \rightarrow \mathbb{R}$ takové, že $X(\omega) = \omega$. Potom $F_X(a) = P[X \leq a] = P(-\infty, a] = F(a)$.

Jde o takzvanou kanonickou konstrukci, zde jsme ukázali hlavní krok. \square

2.2. Náhodný vektor a jeho rozdělení. V následujícím budeme užívat pro jakékoliv dva vektory $a, b \in \mathbb{R}^d$ úmluvu

$$\begin{aligned} a \leq b &\Leftrightarrow a_i \leq b_i \quad \forall i = 1, 2, \dots, d \\ a < b &\Leftrightarrow a_i \leq b_i \quad \forall i = 1, 2, \dots, d, \text{ a existuje } j : a_j < b_j. \end{aligned}$$

Definice 12 (Náhodný vektor). Zobrazení $X : \Omega \rightarrow \mathbb{R}^d$, kde (Ω, \mathcal{F}, P) je pravděpodobnostní prostor, $d \in \mathbb{N}$, $d \geq 2$ takové, že $\{\omega : X(\omega) \leq a\} \in \mathcal{F} \quad \forall a \in \mathbb{R}^d$ nazveme *náhodný vektor*.

Opět jde o požadavek měřitelnosti zobrazení X .

Definice 13 (Rozdělení a distribuční funkce náhodného vektoru). Buď X d -rozměrný náhodný vektor. Pravděpodobnostní míra P_X definovaná na \mathbb{R}^d předpisem

$$P_X\left(\prod_{i=1}^d (-\infty, a_i]\right) = P[X \leq a] = P\left(\bigcap_{i=1}^d [X_i \leq a_i]\right)$$

nazveme rozdělení náhodného vektoru.

Funkce $F_X : \mathbb{R}^d \rightarrow [0, 1]$ definovaná $F_X(a) = P[X \leq a]$ se nazývá *distribuční funkce* náhodného vektoru X .

Pravděpodobnostní míra P_X samozřejmě musí být definovaná i na všech otevřených podmnožinách \mathbb{R}^d . Toto rozšíření z uvedených intervalů typu $(-\infty, a]$ je však jednoznačné a proto je postačující k popisu rozdělení právě distribuční funkce. Pro náhodné vektory však bývá i distribuční funkce někdy velmi složitá.

Zaveďme si značení: Pro $a < b$ buď $\Delta_k(a, b)$ množina těch c , pro které existuje právě k indexů i_1, i_2, \dots, i_k takových, že $c_{i_j} = a_{i_j}$ a pro zbytek indexů je $c_l = b_l$.

Tedy například pro $a = (a_1, a_2, a_3)$, $b = (b_1, b_2, b_3)$ je $\Delta_1(a, b) = \{(a_1, b_2, b_3), (b_1, a_2, b_3), (b_1, b_2, a_3)\}$.

Věta 9 (Vlastnosti distribuční funkce). *Buď F_X distribuční funkce náhodného vektoru X . Pak:*

- (1) $\lim_{a_i \rightarrow -\infty} F_X(\mathbf{a}) = 0$ pro libovolné i .
- (2) $\lim_{a_i \rightarrow \infty} \forall i F_X(\mathbf{a}) = 1$.
- (3) V každé složce argumentu je F_X zprava spojitá a neklesající
- (4) $\forall \mathbf{a} < \mathbf{b}$ platí $\sum_{k=0}^d (-1)^k \sum_{c \in \Delta k(\mathbf{a}, \mathbf{b})} F_X(c) \geq 0$

Podívejme se na čtvrtou podmínku pro $\mathbf{a} = (a_1, a_2)$, $\mathbf{b} = (b_1, b_2)$. Pak tato podmínka říká, že pro $\mathbf{a} < \mathbf{b}$ platí, že $F_X(b_1, b_2) - F_X(a_1, b_2) - F_X(b_1, a_2) + F_X(a_1, a_2) \geq 0$. Jedná se tedy o pravděpodobnost obdélníku $(a_1, b_1] \times (a_2, b_2]$ a ta musí být pochopitelně nezáporná.

I zde platí, že funkce splňující vlastnosti (1)–(4) věty 9 je distribuční funkcí nějaké náhodné veličiny. Rozmyslete si, že vlastnost (4) neplyne z vlastností (3) a je pro distribuční funkci podstatná—tedy funkce splňující jen (1)–(3) nemusí odpovídat pravděpodobnostní míře, neboť může dávat nějaké množině zápornou míru.

Definice 14 (Marginální rozdělení). Buď X náhodný vektor a P_X jeho rozdělení takové, že $P_{X_i}(-\infty, a] = \lim_{a_j \rightarrow \infty, j \neq i} P_X(X_{j=1}^d(-\infty, a_j])$ se nazývá *marginální rozdělení* X_i a $F_{X_i} = \lim_{a_j \rightarrow \infty, j \neq i} F_X(\mathbf{a})$ se nazývá jeho marginální distribuční funkce X_i .

Terminologie: Náhodný vektor (veličina) X , je *diskrétní*, pokud nabývá nejvýše spočetně mnoha hodnot. V tom případě existuje nejvýše spočetná množina \mathbb{S} a nezáporné hodnoty $p_s, s \in \mathbb{S}$ takové, že $P[X = s] = p_s$ a $P[\in A] = \sum_{s \in A} p_s$. Jinými slovy $X(\omega) \in \mathbb{S} \forall \omega$.

Náhodný vektor (veličina) X je *spojitý*, pokud pro libovolnou hodnotu platí $P[X = a] = 0$ a existuje-li nezáporná funkce f taková, že $P[X \in A] = \int_A f(x) dx$.

Rozmyslete. Jistě musí platit

$$\sum_{x \in \mathbb{S}} p_x = 1, \quad \int_{\mathbb{R}^d} f(x) dx = 1.$$

Uvědomte si, že uvedený integrál je násobný, což je podobné, jako násobné sumy. Nejprve se musí vyčíslit vnitřní integrál a postupně se jde až ke vnějšímu.

Všimněte si také, že pro spojitý vektor nezávisí na tom, změníme-li hustotu f v konečně mnoha bodech.

Dohodněme se, že pro naše účely postačí, pokud diskretní náhodný vektor nabývá hodnoty vždy v podmnožině \mathbb{N}_0^d , neřekneme-li jinak.

Rozdělení diskretního náhodného vektoru X je zcela charakterizováno souborem hodnot $\{p_s\}_{s \in \mathbb{S}}$, rozdělení spojitého náhodného vektoru je plně charakterizováno funkcí f . Hodnoty p_s v případě diskretního náhodného vektoru a funkci f v případě spojitého náhodného vektoru nazveme *hustotou* náhodného vektoru X .

V angličtině se distribuční funkce nazývá *cummulative distribution function* a zkracuje obvykle *cdf*. U hustoty obvykle dodáváme, vůči jaké referenční míře tato hustota je definována. Zde jde o aritmetickou = čítací míru u diskretních a o Lebesgueovu míru v případě spojitých náhodných vektorů. V angličtině se hustota nazývá *probability density function* a zkracuje *pdf*.

Svět však není tak přehledný. Náhodné vektory mohou mít jak kombinovaně spojitě i diskretní složky, dokonce i náhodná veličina může mít diskretní a spojitou část—například denní úhrn srážek. Ještě jinou možností je náhodná veličina, která není diskretní, ani spojitá: nabývá nespočetně mnoha hodnot, ale neexistuje žádná f taková, že by $P[X \in A] = \int_A f(x) dx$.

Příklad. Základními modely rozdělení diskretní náhodné veličiny jsou:

- (1) Alternativní (Bernoulliho). X nabývá hodnot z množiny $\{0, 1\}$, $P[X = 1] = P_x(\{1\}) = p \in (0, 1)$ a $P[X = 0] = 1 - p$. Parametr p je interpretován jako pravděpodobnost úspěchu. Značíme $\text{Alt}(p)$.
- (2) Binomické. Rozdělení počtu úspěchů do n nezávislých pokusů. $P[X = k] = \binom{n}{k} p^k (1 - p)^{n-k}$, značíme $\text{Bi}(n, p)$.
- (3) Geometrické. Počet neúspěchů před prvním úspěchem v nezávislých pokusech. $P[X = k] = p(1 - p)^k$, značíme $\text{Geom}(p)$.
- (4) Poissonovo. Počet událostí v jednotkovém časovém intervalu. $P[X = k] = \exp(-\lambda) \lambda^k / k!$, kde $\lambda > 0, k = 0, 1, \dots$. Značíme $\text{Po}(\lambda)$

Příklad. Základními modely rozdělení spojitě náhodné veličiny jsou:

- (1) Rovnoměrné rozdělení na intervalu (a, b) , $-\infty < a < b < \infty$. Jeho hustota je

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a, b) \\ 0 & x \notin (a, b). \end{cases}$$

Toto rozdělení modeluje nulovou informaci o výskytu hodnoty v daném intervalu a značíme jej $U(a, b)$.

- (2) Exponenciální rozdělení je dáno hustotou

$$f(x) = \begin{cases} \lambda \exp(-\lambda x) & x > 0 \\ 0 & x < 0, \end{cases}$$

kde $\lambda > 0$ je parametrem tohoto rozdělení. Toto rozdělení je základním modelem pro dobu do události. Jeho označení je $\text{Exp}(\lambda)$.

- (3) Normální rozdělení je dáno hustotou

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), x \in \mathbb{R}.$$

Zde $\mu \in \mathbb{R}$ a $\sigma^2 > 0$ jsou parametry. Normální rozdělení se též nazývá Gaussovo a je základním modelem pro spojitě náhodné veličiny (někdy až nadužívaným). Značíme jej $N(\mu, \sigma^2)$.

V případě, že $\mu = 0$ a $\sigma^2 = 1$, mluvíme o *standardním* či *normovaném* normálním rozdělení, značíme $N(0, 1)$ a jeho distribuční funkce a hustota se označují Φ , případně ϕ . Tím se dává najevo důležitost postavení tohoto rozdělení v pravděpodobnosti a statistice.

Příklad. Pro náhodné vektory uveďme jedno diskrétní a jedno spojitě rozdělení.

- (1) Multinomické rozdělení je mnohorozměrnou obdobou binomického. Buď n počet pokusů v nichž může dojít k jednomu z k výsledků. Výsledek i má v každém pokusu pravděpodobnost p_i , $\sum_{i=1}^k p_i = 1$, výsledky pokusů se nijak neovlivňují. Náhodný vektor $\mathbf{X} = (X_1, X_2, \dots, X_k)$ zachycující počty výsledků v n pokusech má rozdělení s hustotou

$$P[\mathbf{X} = (n_1, n_2, \dots, n_k)] = \begin{cases} \frac{n!}{n_1! n_2! \dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k} & n_i \in \mathbb{N}_0, \sum_{i=1}^k n_i = n \\ 0 & \text{jinak.} \end{cases}$$

Toto rozdělení označujeme zkratkou $\text{Mult}(n, p_1, \dots, p_k)$.

- (2) Mnohorozměrné (zde d -rozměrné) normální rozdělení je dáno hustotou

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} \sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right), \quad \mathbf{x} \in \mathbb{R}^d,$$

a parametry rozdělení jsou $\boldsymbol{\mu} \in \mathbb{R}^d$ a Σ symetrická pozitivně definitní matice typu $d \times d$. Označení tohoto rozdělení je $N_d(\boldsymbol{\mu}, \Sigma)$.

Pro další účely postačí, pokud se budeme věnovat dvourozměrným spojitým náhodným vektorům. Vše podstatné na nich lze pochopit a omezíme se tak „jen“ na dvounásobné integrování (kterého se nebojte!!). V tom případě je možné omezit se na dvourozměrné normální rozdělení a zkoumat jeho vlastnosti.

Příklad (Dvourozměrné normální rozdělení). Uveďme si základní tři varianty tohoto rozdělení.

- (1) *Standardní normované* dvourozměrné normální rozdělení je charakterizováno $\boldsymbol{\mu} = \mathbf{0}$ a $\Sigma = I_2$, jednotková matice. Jeho hustota má tedy jednoduchý tvar

$$f(\mathbf{x}) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x_1^2 + x_2^2)\right).$$

(2) V případě, kdy $\mu = 0$ a matice Σ má speciální tvar

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \text{ kde } |\rho| < 1$$

pak mluvíme často také o normovaném normálním rozdělení a jeho hustota má tvar (odvoďte si z obecného normálního rozdělení)

$$f(x) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2(1-\rho^2)}\right).$$

(3) V obecném případě, je hustota dána vzorcem

$$f(x) = \frac{1}{2\pi\sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(x_1 - \mu_1, x_2 - \mu_2)\Sigma^{-1} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}\right).$$

2.3. Nezávislost, náhodný výběr a empirické rozdělení. Začneme jednoduchým pozorováním.

Věta 10 (O marginálním rozdělení). *Máme-li náhodný vektor X a jeho rozdělení P_X , pak marginální rozdělení složek X_1, \dots, X_d jsou jednoznačně určeny rozdělením P_X .*

Důkaz. Očividné. □

Rozmyslete. Naopak ale tvrzení **neplatí**.

Hodíme dvěma kostkami. Jejich výsledky jsou A, B . Dále mějme $C = A - 1$ pro B sudé a $C = A + 1$ pro B liché. Zapišme si je do tabulky:

$A \backslash B$	1	2	3	4	5	6	A^C	1	2	3	4	5	6	
1	1/36	1/36	1/36	1/36	1/36	1/36	1/6	1	0	1/12	0	0	1/12	1/6
2	1/36	1/36	1/36	1/36	1/36	1/36	1/6	2	1/12	0	1/12	0	0	1/6
3	1/36	1/36	1/36	1/36	1/36	1/36	1/6	3	0	1/12	0	1/12	0	1/6
4	1/36	1/36	1/36	1/36	1/36	1/36	1/6	4	0	0	1/12	0	1/12	1/6
5	1/36	1/36	1/36	1/36	1/36	1/36	1/6	5	0	0	0	1/12	0	1/6
6	1/36	1/36	1/36	1/36	1/36	1/36	1/6	6	1/12	0	0	0	1/12	1/6
	1/6	1/6	1/6	1/6	1/6	1/6			1/6	1/6	1/6	1/6	1/6	

Vidíme, že pro náhodné vektory (A, B) a (A, C) jsou jejich marginální rozdělení stejná, ale sdružená rozdělení jsou velmi odlišná.

Je krásným matematickým výsledkem, že jakékoliv mnohorozměrné rozdělení je složeno ze svých marginálních pomocí (v jistém smyslu jednoznačně dané) funkce zvané *kopula* a to následovně. Buď F_X sdružená distribuční funkce a buď $F_{X_i}, i = 1, \dots, d$ její marginální distribuční funkce. Pak existuje (v jistém smyslu jediná, pro spojitě F_X opravdu jediná) funkce $C : [0, 1]^d \rightarrow [0, 1]$ taková, že

$$F_X(x) = C(F_1(x_1), F_2(x_2), \dots, F_d(x_d)).$$

Funkce C je sama distribuční funkcí, jejíž marginální rozdělení jsou rovnoměrná na intervalu $(0, 1)$. Modelování a odhadování funkce C je předmětem mnoha studií.

Sdružené rozdělení náhodného vektoru X určuje, jaký *stochastický* vztah mezi sebou mají jednotlivé náhodné veličiny v X . Při jednom speciálním vztahu ale platí, že sdružené rozdělení z marginálního určíme vždy.

Definice 15 (Nezávislost náhodných veličin). Mějme X_1, X_2, \dots, X_k náhodné veličiny definované na (Ω, \mathcal{F}, P) . Tyto veličiny jsou (stochasticky) *vzájemně nezávislé*, pokud platí

$$F_X(x) = \prod_{i=1}^k F_{X_i}(x_i) \quad \forall x = (x_1, \dots, x_k) \in \mathbb{R}^k,$$

kde $X = (X_1, \dots, X_k)$.

Věta 11 (Rozdělení nezávislých náhodných veličin). *Náhodné veličiny X_1, \dots, X_k jsou nezávislé, když $\forall A_1, \dots, A_k$ platí*

$$P\left(\bigcap_{i=1}^k [X_i \in A_i]\right) = \prod_{i=1}^k P[X_i \in A_i].$$

Důkaz. Plyne z charakterizace rozdělení náhodného vektoru pomocí distribuční funkce. \square

Správně bychom měli uvažovat jen $A_i \in \mathcal{B}$, protože pro neměřitelné množiny nejsou hodnoty pravděpodobnosti definovány a dostali bychom se do velkých potíží. Ale s neměřitelnými množinami se zde nebudeme potýkat.

Věta 12 (Ekvivalentní podmínky nezávislosti). *Diskrétní náhodné veličiny X_1, \dots, X_k jsou nezávislé právě tehdy, když $p_{\mathbf{X}}(a_1, \dots, a_k) = \prod_{i=1}^k p_{X_i}(a_i)$.*

Spojitě náhodné veličiny X_1, \dots, X_k jsou nezávislé právě tehdy, když $f_{\mathbf{X}}(x_1, \dots, x_k) = \prod_{i=1}^k f_{X_i}(x_i)$.

Důkaz. Plyne ze vztahu mezi hustotou a distribuční funkcí. \square

Problém: Modely pro náhodné veličiny a vektory jsou jen teoretické konstrukty. Někdy mohou být velmi blízké (binomické rozdělení, pokud máme „zaručenu“—ale jak?—nezávislost a homogenitu pokusů), někdy prostě „jen rozumné“.

Obvykle se snažíme model najít, navrhnout, posoudit, odhadnout, otestovat podle *empirických* zkušeností. K tomu ale potřebujeme zkušenost získat a to lze opakovaným pozorováním náhodného jevu či vektoru o kterém chceme něco zjistit.

Základním modelem je *náhodný výběr*. Jeho podstatou je, že stejnou „náhodnost“ pozorujeme v nezávislých opakováních a z těchto pak můžeme zevšeobecňovat.

Definice 16 (Náhodný výběr). Posloupnost X_1, X_2, \dots, X_n náhodných veličin či vektorů takových, že jsou nezávislé a mají stejné rozdělení P , nazvěme náhodný výběr velikosti n z rozdělení P . Ve zkratce říkáme, že X_1, \dots, X_n jsou iid (independent and identically distributed).

Vraťme se k problematice nezávislosti a porovnejme nezávislost náhodných veličin a jevů. Pro náhodné jevy jsme potřebovali rovnost mezi pravděpodobnostmi průniku a součinem pravděpodobností *pro všechny podmnožiny indexů*. Pro náhodné veličiny je toto nahrazeno požadavkem na distribuční *funkci*.

I zde se můžeme ptát na nezávislost *libovolné množiny* náhodných veličin. A i zde ji lze definovat pomocí nezávislosti *všech konečných podmnožin*. Pro nás ale bude stačit jediné: nezávislost náhodných veličin X_1, X_2, \dots , což budeme potřebovat při limitních větách a jejich použití. K tomu ale stačí, aby pro jakékoliv konečné n byly nezávislé náhodné veličiny X_1, \dots, X_n .

Na základě náhodného výběru můžeme hledat vhodné modely a odhady pro neznámé rozdělení náhodné veličiny. Například jednoduše lze definovat odhad distribuční funkce.

Definice 17 (Empirická distribuční funkce). Buď X_1, \dots, X_n náhodný výběr z rozdělení P s distribuční funkcí F . Pak $\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \chi(X_i \leq x)$, kde χ je indikátorová funkce, se nazývá *empirická distribuční funkce*.

Empirická distribuční funkce je docela dobrým odhadem skutečné distribuční funkce náhodné veličiny. Než však budeme moci říci, co to znamená *býti dobrým odhadem* a proč tomu tak je, musíme si pro náhodné veličiny zavést další charakteristiky.

Poznámka. Zde vidíme význam pojmu náhodná veličina. Empirická distribuční funkce je funkcí náhodných veličin a proto je sama náhodnou veličinou (to si rozmyslete!). Máme tedy vzorec, který pracuje s předem neznámými hodnotami (náhodnou veličinou) a při opakovaných experimentech (náhodných výběrech) dává různé výsledky, jejichž rozdělení se také dá odvodit. Což částečně hned uděláme.

Věta 13 (Rozdělení empirické distribuční funkce I). *Buď X_1, \dots, X_n náhodný výběr z rozdělení P s distribuční funkcí F a zvolme pevné x . Pak $\widehat{F}_n(x)$ má rozdělení dané předpisem*

$$P \left[\widehat{F}_n(x) = \frac{k}{n} \right] = \binom{n}{k} (F(x))^k (1 - F(x))^{n-k}.$$

Důkaz. Jistě platí, že $n\widehat{F}_n(x) = k$, když právě k z n hodnot v náhodném výběru je nejvýše x . Pro všechny náhodné veličiny X_i z náhodného výběru platí

$$P[\chi(X_i \leq x) = 1] = P[X_i \leq x] = F(x) = 1 - P[\chi(X_i \leq x) = 0],$$

proto součet $\sum_{i=1}^n \chi(X_i \leq x)$ má binomické rozdělení a z toho již tvrzení věty snadno plyne. \square

3. STŘEDNÍ HODNOTA A DALŠÍ MOMENTY

Definujeme číselné charakteristiky, které dostatečně vypovídají o chování náhodné veličiny. K čemu je potřebujeme, když chování náhodné veličiny *plně popisuje* její rozdělení? Některá použití uvidíme později, ale jeden z důvodů je i ten, že rozdělení (pravděpodobnostní míra či distribuční funkce) jsou poměrně nepřehledné a hodně složité pojmy. Číselné charakteristiky umožňují snazší porovnávání a interpretaci náhodných veličin.

Definice 18 (Střední hodnota, matematická definice). Buď X náhodná veličina definovaná na pravděpodobnostním prostoru (Ω, \mathcal{F}, P) . Hodnotu

$$E X := \int_{\Omega} X(\omega) dP(\omega)$$

nazveme *střední hodnotou* X , má-li integrál vpravo smysl.

V praxi se však střední hodnota počítá pomocí rozdělení náhodné veličiny.

Věta 14 (Střední hodnota diskrétní náhodné veličiny). Buď X diskrétní náhodná veličina s hodnotami v \mathbb{S} a s hustotou p_X . Pak

$$E X = \sum_{s \in \mathbb{S}} s P[X = s] = \sum_{s \in \mathbb{S}} s p_X(s),$$

má-li součet vpravo smysl.

Důkaz. Využijeme definici: $E X$. Množiny $A_s := \{\omega : X(\omega) = s\}$ tvoří disjunktní rozklad Ω . Proto

$$\begin{aligned} E X &= \int_{\bigcup A_s} X(\omega) dP(\omega) = \sum_{s \in \mathbb{S}} \int_{A_s} X(\omega) dP(\omega) \\ &= \sum_{s \in \mathbb{S}} s \int_{A_s} dP(\omega) = \sum_{s \in \mathbb{S}} s P(A_s) \\ &= \sum_{s \in \mathbb{S}} s P[X = s]. \end{aligned}$$

□

Věta 15 (Střední hodnota spojitě náhodné veličiny). Buď X spojitá náhodná veličina s hustotou f_X . Pak

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

má-li integrál vpravo smysl.

Poznámka. Střední hodnota **má smysl**, i když je nekonečná. Vyjde-li tedy integrál v definici 18 nekonečný ($\pm\infty$), má náhodná veličina X nekonečnou střední hodnotu. Integrál nemá smysl, je-li typu „ $\infty - \infty$ “. Tedy $\int_{\mathbb{R}} 1 dx$ smysl má (i když je integrál nekonečný), zatímco $\int_{\mathbb{R}} x^{-1} dx$ smysl nemá.

Rozmyslete. Zkuste najít funkci $p_X(s)$, $s \in \mathbb{Z}$ takovou, že:

- (1) $p_X(s) \geq 0$
- (2) $\sum_{s \in \mathbb{Z}} p_X(s) = 1$
- (3) $\sum_{s > 0} s p_X(s) = \infty$, $\sum_{s < 0} s p_X(s) = -\infty$

Pro tuto funkci neexistuje střední hodnota, jelikož máme nedefinovanou sumu $(\infty - \infty)$. Najít hustotu f_X tak, aby neexistovala střední hodnota je podobně snadné.

Poznámka (Terminologie a triviální pozorování).

- Pokud $E X$ existuje a je konečná, pak mluvíme o náhodné veličině s konečnou střední hodnotou.
- Pokud $P[X \geq b] = 1$ pro b nějakou konečnou konstantu, pak $E X$ existuje a $E X > b$. Speciálně, nezáporná náhodná veličina má nezápornou střední hodnotu. (Analogicky pro opačnou nerovnost.)

- Pokud $\exists a, b$ konečné a $P[a \leq X \leq b] = 1$, pak $E X$ existuje konečná a $a \leq E X \leq b$.
- Definujme $E|X| = \int_{\Omega} |X(\omega)| dP(\omega)$, která existuje vždy. Pokud $E|X| < \infty$, pak $X \in L_1(P)$ a v tom případě existuje i konečná střední hodnota X , tedy $|E X| < \infty$.

Můžeme definovat i momenty funkce náhodné veličiny a momenty vyšších řádů.

Definice 19 (Obecné momenty náhodné veličiny). Buď X náhodná veličina a $g : \mathbb{R} \rightarrow \mathbb{R}$. Pak $Eg(X) = \int_{\Omega} g(X(\omega)) dP(\omega)$, má-li integrál vpravo smysl.

Věta 16 (Výpočet střední hodnoty). Buď X diskrétní náhodná veličina s hodnotami v \mathbb{S} a s hustotou $p_X s$, buď $g : \mathbb{R} \rightarrow \mathbb{R}$. Pak

$$Eg(X) = \sum_{s \in \mathbb{S}} g(s)P[X = s] = \sum_{s \in \mathbb{S}} g(s)p_X(s),$$

má-li součet vpravo smysl.

Buď X spojitá náhodná veličina s hustotou $f_X x$. Pak

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

má-li integrál vpravo smysl.

Věta 17 (Linearita střední hodnoty). Buď X náhodná veličina s konečnou střední hodnotou. Pak

$$E(a + bX) = a + bE X, \forall a, b \in \mathbb{R}.$$

Důkaz. Plyne přímo z definice střední hodnoty a jejího výpočtu. \square

Věta 18 (Jensenova nerovnost). Buď X náhodná veličina s konečnou střední hodnotou $E X$. Buď φ konvexní funkce. Pak $E\varphi(X) \geq \varphi(E X)$.

Důkaz. Protože je funkce φ konvexní, pro každé zvolené a existuje konstanta λ taková, že $\varphi(x) \geq \varphi(a) + \lambda(x - a)$. Volbou $a = E X$ tak dostáváme

$$\varphi(X) \geq \varphi(E X) + \lambda(X - E X).$$

Vzhledem k tomu, že střední hodnota nezáporné náhodné veličiny je nezáporná a $E(X - E X) = 0$, tvrzení věty již snadno platí. \square

Uvědomme si, že (reálná) funkce náhodné veličiny je opět náhodnou veličinou, jen pokud splňuje požadavky z definice 9. To ale běžně používané funkce splňují, v jazyce teorie míry požadujeme, aby uvažovaná funkce byla měřitelná.

Definice 20 (Významné momenty a momentová vytvořující funkce). Buď X náhodná veličina a uvažujme $r \in \mathbb{N}$. Pak

- (1) $E X^r$ je r -tý moment X .
- (2) $E|X|^r$ je r -tý absolutní moment X .
- (3) pro $r \in \mathbb{N}$ $E(X - E X)^r$ je r -tý centrální moment X .
- (4) pro $r = 2$ máme $\text{var } X = E(X - E X)^2$ rozptyl (variance) X .
- (5) $\mu_3 = \frac{E(X - E X)^3}{(E(X - E X)^2)^{\frac{3}{2}}}$ je šikmost rozdělení, měří asymetrii.
- (6) $\Psi_X(t) = E e^{tX}$ je momentová vytvořující funkce definovaná pro ta t , pro která má smysl pravé strana. Vždy platí $\Psi_X(0) = 1$.

Všechny uvedené charakteristiky jsou definovány tehdy, mají-li příslušné integrály a součty smysl.

Poznámka (Význam a základní vlastnosti momentů). Momenty shrnují některé vlastnosti náhodných veličin a jejich rozdělení. Existují mezi nimi i některé základní vztahy.

- Střední hodnota $E X$ je charakteristikou polohy náhodné veličiny.
- Rozptyl je charakteristikou variability, jde o střední čtvercovou odchylku od střední hodnoty.
- Jinou charakteristikou variability je například první absolutní centrální moment $E|X - E X|$.

- Buďte $0 < s < r$. Jestliže $E|X|^r < \infty$, pak $E|X|^s < \infty$ a $|EX^r| < \infty$ má-li druhý výraz smysl (například pro $r \in \mathbb{N}$). Dokonce platí $(E|X|^s)^{1/s} \leq (E|X|^r)^{1/r}$, speciálně $E|X| \leq (E|X|^2)^{1/2} \leq (E|X|^3)^{1/3}$.
- Existuje-li $\delta > 0$ takové, že momentová vytvořující funkce $\psi_X(t)$ existuje konečná $\forall |t| < \delta$. Pak $\forall r \in \mathbb{N} EX^r = \psi_X^{(r)}(0)$ (r -tá derivace $\psi_X(t)$ v bodě 0).

Věta 19 (Výpočet a vlastnosti rozptylu). *Buď X náhodná veličina s konečným rozptylem. Pak*

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

Pro jakékoliv konstanty a a b platí

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

Důkaz. Přímý výpočet. □

Poznámka. Rozptyl je střední hodnotou nezáporné funkce, čili je nezáporný. Proto musí být $EX^2 \geq (EX)^2$. To plyne také z Jensenovy nerovnosti.

Nyní přidáme momenty náhodného vektoru. V podstatě je důležitý jen jeden, ostatní již máme připravené.

Definice 21 (Střední hodnota náhodného vektoru). *Buď X náhodný vektor. Potom definujeme:*

- (1) $EX = (EX_1, \dots, EX_d)^T$
- (2) Nechť $g : \mathbb{R}^d \rightarrow \mathbb{R}$ (taková, že $g(X)$ je náhodná veličina). Pak definujeme $EG(X) = \int_{\Omega} g(X(\omega)) dP(\omega)$, pokud všechny integrály existují.

Věta 20. *Je-li X diskrétní s hodnotami v \mathbb{N}_0^d , pak $EG(X) = \sum_{z \in \mathbb{N}_0^d} g(z)P[X = z]$, existuje-li řada napravo.*

Je-li X spojitý s hodnotami v \mathbb{R}^d a s hustotou f_X , pak $EG(X) = \int_{\mathbb{R}^d} g(z)f_X(z)dz$, existuje-li integrál napravo.

V obou uvedených případech jde o násobné sčítání či integrování, které se provádí postupně po složkách z .

Věta 21 (Linearita střední hodnoty II). *Buď $X = (X_1, \dots, X_d)$ náhodný vektor a konstanty $a \in \mathbb{R}, b_1, \dots, b_d \in \mathbb{R}$ a $E|X_i| < \infty \forall i$. Pak*

$$E\left(a + \sum_{i=1}^d b_i X_i\right) = a + \sum_{i=1}^d b_i EX_i.$$

Důkaz. Jen pro diskrétní náhodný vektor, indukci.

Vezměme si (X_1, X_2) a a, b_1, b_2 . Potom

$$\begin{aligned} E(a + b_1 X_1 + b_2 X_2) &= \sum_{z \in \mathbb{N}_0^2} (a + b_1 z_1 + b_2 z_2) P[X_1 = z_1, X_2 = z_2] \\ &= \sum_{z \in \mathbb{N}_0^2} a P[X_1 = z_1, X_2 = z_2] + \sum_{z \in \mathbb{N}_0^2} b_1 z_1 P[X_1 = z_1, X_2 = z_2] + \sum_{z \in \mathbb{N}_0^2} b_2 z_2 P[X_1 = z_1, X_2 = z_2] \\ &= a \underbrace{\sum_{z \in \mathbb{N}_0^2} P[X_1 = z_1, X_2 = z_2]}_1 + b_1 \sum_{z_1 \in \mathbb{N}_0} z_1 \underbrace{\sum_{z_2 \in \mathbb{N}_0} P[X_1 = z_1, X_2 = z_2]}_{P[X_1 = z_1]} + b_2 \sum_{z_2 \in \mathbb{N}_0} z_2 \underbrace{\sum_{z_1 \in \mathbb{N}_0} P[X_1 = z_1, X_2 = z_2]}_{P[X_2 = z_2]} \\ &= a + b_1 \sum_{z_1 \in \mathbb{N}_0} z_1 P[X_1 = z_1] + b_2 \sum_{z_2 \in \mathbb{N}_0} z_2 P[X_2 = z_2] = a + b_1 EX_1 + b_2 EX_2. \end{aligned}$$

Takto můžeme pokračovat indukci. □

Definice 22 (Kovariance a korelace). *Buď (X_1, X_2) náhodný vektor takový, že $\text{var } X_1 < \infty, \text{var } X_2 < \infty$.*

- (1) *Kovariance* X_1 a X_2 je definována $\text{cov}(X_1, X_2) = E(X_1 - E X_1)(X_2 - E X_2)$.
 (2) *Korelace* X_1 a X_2 je definována

$$\text{corr}(X_1, X_2) = \frac{\text{cov}(X_1, X_2)}{\sqrt{\text{var } X_1 \text{ var } X_2}}$$

Význam kovariance: $X_1 - E X_1, X_2 - E X_2$ jsou odchylky od příslušných středních hodnot. Proto součin $(X_1 - E X_1)(X_2 - E X_2)$ je kladný, když násobíme odchylky stejného znaménka. Součin je naopak záporný, když odchylky mají různé znaménko. Takže střední hodnota je kladná, pokud „častěji“ mají odchylky od středních hodnot shodné znaménko a naopak. Ale pozor, kovariance je jen indikátor, rozdělení náhodného vektoru je složité a jedna hodnota nemůže postihnout všechny možnosti závislosti X_1 a X_2 . Korelaci zavádíme proto, abychom získali hodnotu nezávislou na „jednotkách“. Platí totiž

$$\text{cov}(aX, Y) = a \text{cov}(X, Y), \text{ ale } \text{corr}(aX, Y) = \text{sign}(a) \text{corr}(X, Y).$$

Rozmyslete (Výpočet kovariance). K výpočtu kovariance můžeme využít přímou definici. Ta však nemusí být úplně vhodná.

Podobně jako pro rozptyl můžeme tento vzoreček zjednodušit:

$$E((X_1 - E X_1)(X_2 - E X_2)) = E(X_1 X_2 - X_1 E X_2 - X_2 E X_1 + E X_1 E X_2) = E(X_1 X_2) - E X_1 E X_2$$

Věta 22 (Nezávislost a korelace). *Nechť náhodné veličiny X a Y jsou nezávislé a mají konečný rozptyl. Pak $\text{cov}(X, Y) = \text{corr}(X, Y) = 0$, neboli X a Y jsou nekorelované.*

Opačná implikace NEPLATÍ!!!

Důkaz. Uvedeme pro diskrétní náhodný vektor (X, Y) , pro spojitý jde o analogický postup. Přímou z výpočtu střední hodnoty funkce $g(X, Y) = XY$ a z nezávislosti X a Y plyne

$$\begin{aligned} E XY &= \sum_m \sum_n mn P[X = m, Y = n] = \sum_m \sum_n mn P[X = m] P[Y = n] \\ &= \sum_m m P[X = m] \sum_n n P[Y = n] = E X E Y. \end{aligned}$$

Kovariance je tedy nulová a proto i korelace je nulová. □

Věta 23 (Rozptyl součtu náhodných veličin). *Buď $\mathbf{X} = (X_1, \dots, X_d)$ náhodný vektor s konečnými rozptyly. Pak*

$$\text{var} \left(\sum_{i=1}^d X_i \right) = \sum_{i=1}^d \text{var } X_i + \sum_{1 \leq i \neq j \leq d} \text{cov}(X_i, X_j) = \sum_{i=1}^d \text{var } X_i + 2 \sum_{1 \leq i < j \leq d} \text{cov}(X_i, X_j).$$

Speciálně, jsou-li X_1, \dots, X_d nezávislé, pak

$$\text{var} \left(\sum_{i=1}^d X_i \right) = \sum_{i=1}^d \text{var } X_i.$$

Důkaz. Podle definice rozepíšeme a postupnou úpravou:

$$\begin{aligned} &\text{var} \left(\sum_{i=1}^d X_i \right) \\ &= E \left(\sum_{i=1}^d X_i - E \left(\sum_{i=1}^d X_i \right) \right)^2 = E \left(\sum_{i=1}^d (X_i - E X_i)^2 + 2 \sum_{1 \leq i < j \leq d} (X_i - E X_i)(X_j - E X_j) \right) \\ &= \sum_{i=1}^d \text{var } X_i + 2 \sum_{1 \leq i < j \leq d} \text{cov}(X_i, X_j). \end{aligned}$$

□

Definice 23 (Varianční a korelační matice). Buď $\mathbf{X} = (X_1, \dots, X_d)$ náhodný vektor, $\text{var } X_i < \infty \forall i = 1, \dots, d$. Potom varianční matice je $\text{Var } \mathbf{X} = \{\text{cov}(X_i, X_j)\}_{i,j}$ a korelační matice je $\text{Corr } \mathbf{X} = \{\text{corr}(X_i, X_j)\}_{i,j}$.

Varianční matice $\text{Var } \mathbf{X}$ má díky definici na diagonále právě $\text{var } X_i$. Podobně korelační matice má na diagonále samé jedničky.

Věta 24 (Vlastnosti kovariance a korelace). Buďte \mathbf{X} náhodný vektor s varianční maticí $\text{Var } \mathbf{X}$ (která existuje a má konečné prvky). Dále mějme X, Y náhodné veličiny s konečným rozptylem. Potom:

- (1) $-1 \leq \text{corr}(X, Y) \leq 1, \text{corr}(X, X) = 1$
- (2) $|\text{corr}(X, Y)| = 1 \Leftrightarrow \exists a \neq 0, b \in \mathbb{R} : X = aY + b$ (Korelace měří míru lineární závislosti X a Y .)
- (3) $\text{cov}(aX + c, bY + d) = ab \text{cov}(X, Y)$
- (4) $\text{corr}(aX + c, bY + d) = \text{sign}(ab) \text{corr}(X, Y)$
- (5) $\text{Var } \mathbf{X}, \text{Corr } \mathbf{X}$ jsou symetrické pozitivně semidefinitní
- (6) $\forall A \in \mathbb{R}^{l \times d}, B \in \mathbb{R}^l, \mathbf{X} \in \mathbb{R}^d : \text{Var}(A\mathbf{X} + B) = A(\text{Var } \mathbf{X})A^T$

Podíváme-li se na vzorec pro rozptyl součtu náhodných veličin, vidíme, že platí

$$\text{var} \sum X_i = \mathbf{1}^T \text{Var } \mathbf{X} \mathbf{1},$$

kde $\mathbf{1}$ je vektor složený ze samých jedniček. Ve stejném duchu je možné psát maticově i

$$\text{var} \sum a_i X_i = \mathbf{a}^T \text{Var } \mathbf{X} \mathbf{a},$$

pro libovolný vektor \mathbf{a} . Protože výraz na levé straně je nezáporný, plyne odtud i pozitivní semidefinitnost matice $\text{Var } \mathbf{X}$.

Odtud již snadno plyne závěr věty pro $|\text{corr}(X, Y)| = 1$. V tomto případě platí, že $\text{cov}^2(X, Y) = \text{var } X \text{var } Y$. Proto

$$\text{Var}(X, Y) = \begin{pmatrix} \text{var } X & \pm \sqrt{\text{var } X \text{var } Y} \\ \pm \sqrt{\text{var } X \text{var } Y} & \text{var } Y \end{pmatrix},$$

je symetrická matice, znaménko na vedlejší diagonále odpovídá znaménku korelace a tato matice je singulární. Existuje tedy vektor, konkrétně například $\mathbf{a}^T = (\text{var } Y, \mp \sqrt{\text{var } X \text{var } Y})$, takový, že

$$\mathbf{a}^T \text{Var}(X, Y) \mathbf{a} = 0.$$

To ovšem znamená, že náhodná veličina $a_1 X + a_2 Y$ má nulový rozptyl a proto je skoro jistě konstantní.

4. KRÁTKÁ ODBOČKA KE KONVERGENCI

Než pokročíme dále, bude dobré zadefinovat dva typy konvergence náhodných veličin. Připomeňme, že náhodná veličina je zobrazení $X : \Omega \rightarrow \mathbb{R}$ a tedy pojem konvergence toto musí zohlednit.

Definice 24 (Konvergence v pravděpodobnosti). Buďte X náhodná veličina a X_1, X_2, \dots posloupnost náhodných veličin definovaných na tomtéž pravděpodobnostním prostoru. Řekneme, že X_n konverguje v pravděpodobnosti k X , pokud

$$\forall \varepsilon > 0 \mathbb{P}[|X_n - X| > \varepsilon] \xrightarrow{n \rightarrow \infty} 0.$$

Tuto konvergenci značíme $X_n \xrightarrow{\text{P}} X$.

Poznámka. Limitní náhodná veličina X v předchozí definici může být (a často bude) konstantou.

Zapišeme-li předchozí konvergenci s využitím zamlčených proměnných ω , dostaneme

$$\mathbb{P}\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon\} \xrightarrow{n \rightarrow \infty} 0.$$

To znamená, že se zvětšujícím se n se „zmenšuje“ (alespoň v pravděpodobnostním smyslu) množina těch elementárních jevů, pro něž se X_n a X liší o více než „zanedbatelnou“ hodnotu. Uvědomme si však, že množina $\Omega_\varepsilon := \{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon\}$ závisí na n a obecně tyto množiny netvoří monotónní do sebe vnořenou posoupnost.

Definice 25 (Konvergence v rozdělení). Buďte X náhodná veličina a X_1, X_2, \dots posloupnost náhodných veličin definovaných na totéž pravděpodobnostním prostoru. Řekneme, že X_n *konverguje v rozdělení* (v distribuci) k X , pokud

$$F_{X_n}(x) = P[X_n \leq x] \xrightarrow{n \rightarrow \infty} P[X \leq x] = F_X(x) \text{ pro všechna } x, \text{ kde } F_X \text{ je spojitá.}$$

Tuto konvergenci značíme $X_n \xrightarrow{d} X$.

Zde vlastně nekonvergují X_n , ale „jen“ jejich rozdělení. Pravděpodobnostní chování náhodných veličin X_n se blíží chování náhodné veličiny X .

5. ODHADY MOMENTŮ A ODHADY PARAMETRŮ MOMENTOVOU METODOU

Rozdělení náhodného vektoru obvykle neznáme. Dají se rozlišit dvě základní situace (nedělejme si ovšem iluze o úplnosti tohoto výčtu):

- (1) **Parametrické modelování:** O rozdělení máme nějakou představu, obvykle kladenou na hustotu rozdělení. Předpokládáme, že náhodná veličina má rozdělení, které patří do nějaké třídy popsatelné až na konečný (malý) počet parametrů. Příkladem může být počet úspěchů v nějakém počtu pokusů. Model binomického rozdělení je naprosto vyhovující, můžeme-li považovat úspěchy v jednotlivých pokusech za nezávislé a nastávající se stejnou pravděpodobností. V tomto případě obvykle potřebujeme *něco říci o neznámých parametrech*.
- (2) **Neparametrický přístup:** O rozdělení (hustotě) předpokládáme jen celkem málo. Například existence některých momentů a spojitost rozdělení jsou typické slabé požadavky kladené na rozdělení. V tomto případě obvykle chceme *něco říci o charakteristikách rozdělení*, jeho konkrétní tvar nás moc nezajímá. Častou otázkou tohoto typu je, zda dvě náhodné veličiny mají stejné střední hodnoty (rozptyl, ...), nebo zda jsou nezávislé (nekorelované).

5.1. Náhodný výběr a odhad. Nejčastějším nástrojem k hledání odpovědí na otázky týkající se náhodných veličin a vektorů (parametricky i neparametricky) je *náhodný výběr*. Opakovaným pozorováním nezávislých výsledků experimentu (výsledky náhodného výběru) dostáváme *data*, ve kterých máme ukrytou dostupnou informaci o rozdělení. Již víme, jak odhadnout distribuční funkci pomocí empirické distribuční funkce. Nyní budeme odhadovat momenty a parametry.

Definice 26 (Bodový odhad). Buď X_1, X_2, \dots, X_n náhodný výběr z rozdělení s distribuční funkcí F_X . Funkci $g_n : \mathbb{R}^n \rightarrow \Theta$, jejíž předpis nezávisí na rozdělení F_X (ani jeho parametrech), nazveme *bodový odhad*.

Poznámka. Bodový odhad, ale čeho? V zásadě čehokoliv, parametru, momentu, hustoty, pravděpodobnosti, distribuční funkce. Důležité je, že předpis g_n nezávisí na neznámém—rozdělení či parametru, ale je funkcí jen hodnot náhodného výběru. Všimněme si, že předpis g_n obvykle závisí na rozsahu n náhodného výběru (připomeňte si definici 17). Z toho ovšem plyne důležitý závěr: *odhad sám je náhodnou veličinou*.

Odhadovat můžeme buď nějaký parametr (například p v binomickém rozdělení, μ v normálním rozdělení), nebo jinou hodnotu svázanou s rozdělením (pravděpodobnost nějaké hodnoty či intervalu, střední hodnotu, distribuční funkci či hustotu v nějakém bodě). V souladu se zvyklostmi označíme symbolem θ *hodnotu, kterou odhadujeme*, pokud nebude specifikována přesněji. Odhad hodnoty θ založený na náhodném výběru o rozsahu n se obvykle značí $\hat{\theta}_n$.

Poznámka. Čtenář v tuto chvíli může být zmaten a ptát se „Co je tedy vlastně odhad, jak jej mám spočítat, jaký je vzorec?“. Podstatné je *co* odhadujeme a také jak *posuzujeme chybu* odhadu—a k té zákonitě dochází, neboť odhad je náhodná veličina a odhadujeme obvykle reálné číslo (vektor).

5.2. Kvalita odhadu. Jednou z žádaných vlastností odhadu je jeho nestrannost. Tedy aby se *v průměru rovnal* odhadované hodnotě.

Definice 27 (Nestrannost odhadu). Buď X_1, X_2, \dots, X_n náhodný výběr z rozdělení F_X a buď θ odhadovaný parametr. Bodový odhad $\hat{\theta}_n$ nazveme nestranným, pokud $\forall \theta \in \Theta$ platí $E\hat{\theta}_n = \theta$ je-li θ skutečnou hodnotou parametru.

Chceme tedy, aby střední hodnota odhadu byla onou neznámou odhadovanou hodnotou. Je jistě žádoucí, aby odhad netrpěl systematickou chybou.

Druhý dobrým požadavkem je, aby se odhad *blížil odhadované hodnotě*. Tomu se říká konzistence.

Definice 28 (Konzistence odhadu). Buď X_1, X_2, \dots, X_n náhodný výběr z rozdělení F_X a buď θ odhadovaný parametr. Bodový odhad $\hat{\theta}_n$ nazveme konzistentním, pokud $\forall \theta \in \Theta$ platí $\hat{\theta}_n \xrightarrow{P} \theta$ je-li θ skutečnou hodnotou parametru.

Konečně posledním častým požadavkem je asymptotická normalita, tedy aby se chování vhodně upraveného odhadu blížilo standardnímu normálnímu rozdělení. O tom si však povíme až později

Zvědavý čtenář si jistě klade spoustu otázek: jak máme zaručeno, že existuje nekonečně mnoho nezávislých náhodných veličin se stejným rozdělením a na stejném pravděpodobnostním prostoru? Spokojme se s tím, že takový prostor lze bezesporně zkonstruovat a tedy je to možné, pokud dokážeme zkonstruovat nekonečně rozměrný součinný prostor a na něm součinnou míru.

Jinou otázkou je, co znamená „pro každý parametr θ , je-li θ skutečný parametr“, když přeci θ neznáme. Znamená to, že předpis pro $\hat{\theta}_n$ musí být takový, že nezávisí na θ , ale jen na hodnotách náhodného výběru (obvykle na hodnotách v \mathbb{R}^n , kde n je rozsah výběru (tedy pro každé n máme trochu jiný předpis pro $\hat{\theta}_n$). Jelikož rozdělení odhadu $\hat{\theta}_n$ je odvozené od rozdělení náhodného výběru, závisí toto rozdělení také na θ . Proto dosadíme-li *libovolné* θ do rozdělení $\hat{\theta}_n$, musí nám vyjít to, co chceme.

5.3. Výběrové momenty. V této sturčné kapitole jen ukážeme jak počítat výběrové momenty a proč.

Definice 29 (Výběrový průměr a výběrový rozptyl). Buď X_1, \dots, X_n náhodný výběr. Výběrovou střední hodnotou nazveme \bar{X}_n a výběrovým průměrem S_n^2 definované

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i, \quad S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

Věta 25 (Nestrannost výběrového průměru, rozptylu a empirické distribuční funkce). Je-li X_1, \dots, X_n náhodný výběr z rozdělení s distribuční funkcí F_X , pak $\bar{F}_n(x)$ je nestranným odhadem $F(x)$ v každém bodě x .

Je-li X_1, \dots, X_n náhodný výběr z rozdělení s konečnou střední hodnotou $E X$, pak \bar{X}_n je nestranným odhadem $E X$.

Je-li X_1, \dots, X_n náhodný výběr z rozdělení s konečným rozptylem $\text{var } X$, pak S_n^2 je nestranným odhadem $\text{var } X$.

Definice 30 (Výběrové momenty). Buď X_1, \dots, X_n náhodný výběr. Pak

$$\widehat{E X^r} = \frac{1}{n} \sum_{i=1}^n X_i^r$$

se nazývá r -tý výběrový moment.

Věta 26 (Nestrannost výběrových momentů). Je-li X_1, \dots, X_n náhodný výběr z rozdělení s konečným r -tým momentem $E X^r$, pak $\widehat{E X^r}$ je nestranným odhadem $E X^r$.

5.4. Odhady metodou momentů. Abychom si přeci jen ukázali nějaký způsob konstrukce odhadu, ukažme momentovou metodu, která nepotřebuje žádné další teoretické kapitoly.

Předpokládejme, že distribuční funkce F (rozdělení P) závisí na neznámém parametru. Pak obvykle i momenty náhodné veličiny s tímto rozdělením závisí na těchto parametrech (vzpomeňte si na příklady jednotlivých rozdělení).

Předpokládejme, že chceme odhadnout neznámý parametr θ , jehož hodnoty se nacházejí v množině $\Theta \subset \mathbb{R}^d$ (často $d = 1, 2$). Necht momenty náhodné veličiny X existují alespoň do řádu d a závisí na parametru θ , tedy nechť platí

$$E X^r = \mu_r(\theta), r = 1, 2, \dots, d,$$

kde známe funkce $\mu_r(\theta)$. Levou stranu nahradíme výběrovými momenty a řešíme soustavu rovnic

$$\widehat{E X^r} = \mu_r(\hat{\theta}_n), r = 1, 2, \dots, d,$$

čili hledáme $\hat{\theta}_n$ splňující tyto rovnice. Obvykle je pro d -rozměrný parametr potřeba právě d těchto *momentových rovnic*, někdy však může stačit méně.

Obecně odhady momentovou metodou nemusí být nestranné, někdy je možné je na nestranné upravit. Hlavní výhodou této metody je jednoduchost a pochopitelnost.

6. LIMITNÍ VĚTY

V této důležité kapitole si představíme některé nerovnosti a limitní věty pro pravěpodobnost a pro náhodný výběr.

Věta 27 (Markovova nerovnost I). *Buď X nezáporná náhodná veličina. Potom*

$$P[X \geq \varepsilon] \leq \frac{E X}{\varepsilon} \quad \forall \varepsilon > 0$$

Důkaz. Ukážeme si pro X spojitou náhodnou veličinu s hustotou f , ale lze i obecně.

$$P[X \geq \varepsilon] = \int_{x \geq \varepsilon} 1 f(x) dx \leq \int_{x \geq \varepsilon} \underbrace{\frac{x}{\varepsilon}}_{\text{vždy kladné}} f(x) dx \leq \int_0^\infty \frac{x}{\varepsilon} f(x) dx = \frac{E X}{\varepsilon}.$$

□

Věta 28 (Markovova nerovnost II). *Buď X nezáporná náhodná veličina. Potom*

$$P[X \geq \varepsilon] \leq \frac{E(X^k)}{\varepsilon^k} \quad \forall \varepsilon > 0, k > 0$$

Věta 29 (Čebyševova nerovnost). *Buď X náhodná veličina a $E|X| < \infty$. Pak*

$$P[|X - E X| \geq \varepsilon] \leq \frac{\text{var } X}{\varepsilon^2}$$

Důkaz. Věta 28 použitá na $|X - E X|$ a $k = 2$.

□

Čebyševova nerovnost může být zajímavě zesílena Kolmogorovovou nerovností.

Věta 30 (Kolmogorovova nerovnost). *Buďte X_1, X_2, \dots nezávislé náhodné veličiny a $E|X_i| < \infty \forall i$. Pak*

$$P\left[\max_{1 \leq k \leq n} \left|\sum_{i=1}^k (X_i - E X_i)\right| \geq \varepsilon\right] \leq \frac{\sum_{i=1}^n \text{var } X_i}{\varepsilon^2}.$$

Důkaz. Označme si $S_k = \sum_{i=1}^k (X_i - E X_i)$. Potom $E S_k = 0$ a $\text{var } S_k = \sum_{i=1}^k \text{var } X_i$ díky nezávislosti. Hledáme tedy pravděpodobnost $P[\max_{1 \leq k \leq n} |S_k| \geq \varepsilon]$.

Zadefinujme si množinu $A_k = \{\omega : |S_j(\omega)| < \varepsilon, j < k, |S_k(\omega)| \geq \varepsilon\}$, kde $S_0 = 0$. Můžeme si všimnout, že A_k jsou po dvou disjunktní, jde tedy o disjunktní rozklad jevu $[\max_{1 \leq k \leq n} |S_k| \geq \varepsilon]$.

Nyní předpokládejme, že náhodné veličiny X jsou spojitě, tedy i $|S_k|$ jsou spojitě s hustotou s_k (opět lze dokázat obecně). Platí

$$P[\max_{1 \leq k \leq n} |S_k| \geq \varepsilon] = \sum_{k=1}^n P[A_k] \leq \sum_{k=1}^n \frac{1}{\varepsilon^2} E \chi(A_k) S_k^2,$$

kde k poslední nerovnosti využijeme Markovovu nerovnost. Použijme dále identitu

$E \chi(A_k) S_n^2 = E \chi(A_k) (S_n - S_k + S_k)^2 = E \chi(A_k) (S_n - S_k)^2 + 2 E \chi(A_k) S_k (S_n - S_k) + E \chi(A_k) (S_k)^2$
a zaměříme se na střední člen. Díky nezávislosti X_i dostáváme nezávislost $\chi(A_k) S_k$ a $(S_n - S_k)$, z čehož plyne

$$E \chi(A_k) 2(S_n - S_k) \cdot S_k = E(S_n - S_k) E S_k \chi_{A_k} = 0.$$

Celkově $E \chi(A_k) (S_k)^2 \leq E \chi(A_k) S_n^2$ a dosazením do předchozího dostaneme

$$\sum_{k=1}^n P[A_k] \leq \sum_{k=1}^n \frac{1}{\varepsilon^2} E \chi(A_k) S_k^2 \leq \sum_{k=1}^n \frac{1}{\varepsilon^2} E \chi(A_k) S_n^2 \leq \frac{1}{\varepsilon^2} E S_n^2.$$

Postačí si uvědomit, že $E S_n^2 = \text{var } S_n$, protože $E S_n = 0$. \square

Nyní lze ukázat, že odhady momentů i odhad distribuční funkce pomocí empirické distribuční funkce jsou konzistentní. Nejprve použijeme vhodné nerovnosti na součty náhodných veličin a odvodíme obecný výsledek, který pak budeme používat na jednotlivé případy.

Věta 31 ((Slabý) Zákon velkých čísel). *Buďte X_1, X_2, \dots iid náhodné veličiny s konečnou střední hodnotou $E X_1 = \mu$ a s konečným rozptylem $\text{var } X_1 = \sigma^2$. Pro libovolné $\varepsilon > 0$*

$$P \left[\left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| > \varepsilon \right] \xrightarrow{n \rightarrow \infty} 0.$$

Důkaz. Z věty 26 víme, že aritmetický průměr je nestranným odhadem své střední hodnoty μ , existuje-li tato. S využitím Čebyševovy nerovnosti

$$P[|\overline{X}_n - \mu| > \varepsilon] \leq \frac{\text{var } \overline{X}_n}{\varepsilon^2},$$

což spolu se skutečností

$$\text{var } \overline{X}_n = \text{var } \frac{1}{n} \sum X_i = \frac{1}{n^2} \text{var } \sum X_i = \frac{1}{n^2} n \text{var } X_1 = \frac{\text{var } X_1}{n} \xrightarrow{n \rightarrow \infty} 0,$$

kde předposlední rovnost platí díky nezávislosti a stejnému rozdělení X_i , dává požadovanou konvergenci. \square

Důsledek 1. *Buďte X_1, X_2, \dots iid náhodné veličiny s konečným $2r$ -tým momentem $E X_1^{2r}$. Pak pro libovolné $\varepsilon > 0$*

$$P \left[\left| \frac{1}{n} \sum_{i=1}^n X_i^r - E X_1^r \right| > \varepsilon \right] \xrightarrow{n \rightarrow \infty} 0.$$

Ve skutečnosti lze předchozí větu dokázat i silněji. Zprv nepotřebujeme nezávislost a stačí nekorelovanost, protože v tom případě také platí, že

$$\text{var } \sum_i X_i = \sum_i \text{var } X_i + \underbrace{\sum_{i \neq j} \text{cov}(X_i, X_j)}_{=0} = \sum_i \text{var } X_i.$$

Dokonce stačí požadovat ještě méně a to aby kovariance byly dostatečně rychle mizející se vzdáleností indexů i a j , aby $n^{-2} \text{var } \sum_i X_i$ konvergovalo k nule při n rostoucím do nekonečna. Zadrugé, pokud si *ponecháme* nezávislost, můžeme naopak požadavek na existenci „druhého“ momentu nahradit požadavkem $E |X| < \infty$ (ve větě 31, případně $E |X|^r < \infty$ v důsledku) a věta zůstane v platnosti. Dokonce lze i zesílit smysl konvergence na konvergenci ve *všech* ω až na množinu pravděpodobnosti 0. Takovému výsledku pak říkáme *silný* zákon velkých čísel. Praktici často považují rozdíl mezi silným a slabým ZVČ za zanedbatelný a z hlediska používání a interpretace nezajímavý, teoretici naopak velmi lpějí na jejich rozlišování a jejich snahou je pokud možno dokázat silný zákon (konvergenci skoro všude). V našem textu však nemáme techniky pro důkaz silného zákona, zájemci si laskavě vyhledají příslušné výsledky v bohaté dostupné literatuře.

Důsledek 2. *Bud' $\hat{F}_n(x)$ empirická distribuční funkce. Pak platí $\hat{F}_n(x) \xrightarrow[n \rightarrow \infty]{P} F(x)$ pro každé $x \in \mathbb{R}$.*

Dokázali jsme tedy konzistenci odhadů momentů náhodných veličin a odhadu distribuční funkce. Pokud jsou parametry odhadované pomocí metody momentů spojitými funkcemi momentů (a jednoznačně určenými), pak ze spojitosti a konvergence dostaneme i konzistenci odhadů vytvořených momentovou metodou (tedy třeba i takto odhadnuté distribuční funkce). Základem může být jednoduché tvrzení, které uvedeme bez čísla a bez důkazu.

Věta (Konvergence spojitě transformace). *Bud' X_1, X_2, \dots posloupnost náhodných veličin a a konstanta taková, že $P[|X_n - a| > \varepsilon] \rightarrow 0$ pro libovolné $\varepsilon > 0$. Bud' ϕ spojitá funkce. Pak*

$$P[|\phi(X_n) - \phi(a)| > \varepsilon] \xrightarrow[n \rightarrow \infty]{} 0 \quad \forall \varepsilon > 0.$$

Další možnou modifikací zákona velkých čísel je vynechání požadavku na stejné rozdělení náhodných veličin. Pak je samozřejmě otázkou, k jaké hodnotě konverguje aritmetický průměr a za jakých podmínek. Opět se vyskytuje více variant odpovědi, ale nám bude stačit jen jednoduchá varianta.

Věta. *Bud' X_1, X_2, \dots iid náhodné veličiny s konečnou střední hodnotou $E X_i = \mu_i$ a s omezeným rozptylem $\text{var } X_i = \sigma_i^2 \leq K$. Pak*

$$P \left[\left| \frac{1}{n} \sum_{i=1}^n (X_i - \mu_i) \right| > \varepsilon \right] \xrightarrow[n \rightarrow \infty]{} 0.$$

Pokud $n^{-1} \sum \mu_i$ konverguje k hodnotě μ při n rostoucím nade všechny meze, pak platí

$$P \left[\left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| > \varepsilon \right] \xrightarrow[n \rightarrow \infty]{} 0.$$

Teď již můžeme u každého odhadu rozhodovat o jeho nestrannosti (asymptotické nestrannosti) a konzistenci. Mnohem hlubší výsledek nám poskytuje *centrální limitní věta* (CLV). Její důkaz však je buď velice zdlouhavý, nebo vyžaduje pokročilejší techniky, takže se spokojíme se zněním.

Věta 32 (Centrální limitní věta). *Bud' X_1, X_2, \dots nezávislé a stejně rozdělené náhodné veličiny s konečnou střední hodnotou $E X_1 = \mu$ a konečným kladným rozptylem $\text{var } X_1 = \sigma^2 > 0$. Pak*

$$P \left[\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \leq x \right] \xrightarrow[n \rightarrow \infty]{} \Phi(x), \quad \forall x \in \mathbb{R},$$

kde Φ je distribuční funkce standardního normálního rozdělení $N(0, 1)$.

Ekvivalentně můžeme říct

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{d} N(0, 1) \text{ nebo } \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} N(0, 1) \text{ nebo } \sqrt{n} (\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2).$$

Použití CLV jsou všude. Využijeme je ke konstrukci (asymptotických) intervalových odhadů, asymptotických predikčních intervalů, testování, ... Zde se naplno ukazuje význam postavení normálního rozdělení: mnoho navzájem nezávislých vlivů se posčítá, extrémy se vyruší a výsledkem je normální (Gaussovo) rozdělení.

I zde je možné zeslabit požadavky a tak není třeba ani stejné rozdělení, ani úplná nezávislost. Jde však o velmi specializovaná tvrzení, tak se jim vyhneme a naučíme se používat CLV alespoň v tomto základním tvaru.

Příklad (Příklady použití CLV). Centrální limitní věta se obvykle používá k určení pravděpodobnosti, k určení intervalu či k určení potřebného počtu pozorování. Jednotlivé úlohy v podstatě odpovídají hledání jedné neznámé v rovnici odvozené pomocí CLV.

- (1) Chceme vědět, s jakou pravděpodobností bude počet úspěchů v Bernoulliových pokusech vyšší, než a . Máme X_1, \dots, X_n nezávislé stejně rozdělené s alternativním rozdělením s parametrem p . Kolik bude $P[\sum X_i > a]$?

Použijeme CLV a stejné úpravy na obě strany. Cílem je získat jeden z tvarů uvedených ve větě 32.

$$P[\sum X_i > a] = P \left[\frac{\sum X_i - np}{\sqrt{np(1-p)}} > \frac{a - np}{\sqrt{np(1-p)}} \right] = 1 - P[\dots \leq \dots] \approx 1 - \Phi \left(\frac{a - np}{\sqrt{np(1-p)}} \right)$$

- (2) Určení počtu pokusů. Chceme vědět, kolik pokusů máme udělat, aby se relativní počet úspěchů v Bernoulliiovských pokusech (tedy podíl úspěchů k počtu pokusů) lišil od teoretické pravděpodobnosti nejvýše o ε s pravděpodobností nejvýše α . Jinými slovy hledáme n tak, aby $P[|\bar{X}_n - p| > \varepsilon] \leq \alpha$.

Znova využijeme CLV a upravujeme. Obdobnými úpravami jako výše dostaneme

$$P\left[\sqrt{n} \frac{|\bar{X}_n - p|}{\sqrt{p(1-p)}} > a\right] = 1 - P\left[\sqrt{n} \frac{\bar{X}_n - p}{\sqrt{p(1-p)}} > a\right] + P\left[\sqrt{n} \frac{\bar{X}_n - p}{\sqrt{p(1-p)}} < -a\right] = 1 - \Phi(a) + \Phi(-a).$$

S využitím symetrie normálního rozdělení, jmenovitě $\Phi(a) = 1 - \Phi(-a)$, poslední výraz upravíme na $2(1 - \Phi(a))$. To znamená, že hledáme a tvaru $\frac{\sqrt{n}\varepsilon}{\sqrt{p(1-p)}}$ tak, aby $2(1 - \Phi(a)) \leq \alpha$, tedy $a \geq \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$. Z tohoto a potom vypočteme potřebné n .

- (3) Určení intervalu. Každý z n účastníků korespondenčního semináře dokončí všechny úkoly—a dostane dárek—s pravděpodobností p (nezávisle na ostatních). Kolik musí mít pořadatelé v zásobě dáreků, chtějí-li s pravděpodobností 0,99 mít dárek pro každého úspěšného řešitele?

Triviální odpověď je n . Pak máme *jistotu*, že se dostane na každého. Ale zároveň s pravděpodobností $1 - p^n$ (a ta může být *hodně* velká) nám nějaký dárek zbyde.

Označíme-li X_i náhodnou veličinu „účastník uspěl“, pak $\sum X_i$ je počet dáreků, které potřebujeme. Okamžitě vidíme, že můžeme využít CLV a po několika úpravách dospět k asymptotické rovnosti

$$P\left[\sum X_i \leq a\right] = P\left[\frac{\sum X_i - np}{\sqrt{np(1-p)}} \leq \frac{a - np}{\sqrt{np(1-p)}}\right] \doteq \Phi\left(\frac{a - np}{\sqrt{np(1-p)}}\right).$$

Poslední výraz má být nejméně 0,99, z čehož pak plyne nerovnice

$$\frac{a - np}{\sqrt{np(1-p)}} \geq \Phi^{-1}(0,99) \Rightarrow a \geq np + \sqrt{np(1-p)}\Phi^{-1}(0,99).$$

Alespoň tolik dáreků musíme mít připraveno.

Již jsme viděli, že konvergence v pravděpodobnosti „přežije“ transformaci spojitou funkcí. U centrální limitní věty máme podobný výsledek, jen je třeba správně určit asymptotický rozptyl. K tomu slouží takzvaná *delta-metoda*.

Věta 33 (Delta). *Bud' Y_1, Y_2, \dots posloupnost náhodných veličin takových, že $\sqrt{n}(Y_n - \mu) \xrightarrow{d} N(0, \sigma^2)$ a diferencovatelnou funkcí g . Pak*

$$\sqrt{n}(g(Y_n) - g(\mu)) \xrightarrow{d} N(0, (g'(\mu))^2 \sigma^2)$$

Delta věta dává velmi univerzální a mocný nástroj. Ukažme si její použití na příkladu Poissonova rozdělení.

Příklad. Předpokládejme, že máme Poissonovo rozdělení s parametrem λ a náhodný výběr X_1, \dots, X_n z tohoto rozdělení. Pak víme, že $EX = \text{var } X = \lambda$. Z CLV víme, že $\sqrt{n}(\bar{X}_n - \lambda) \xrightarrow{d} N(0, \lambda)$. Dále víme, že $P[X = 0] = e^{-\lambda} = g(\lambda)$.

Delta věta nám dává asymptotické normální rozdělení pro $P[X = 0]$ ve tvaru $\sqrt{n}(e^{-\bar{X}_n} - e^{-\lambda}) \xrightarrow{d} N(0, \lambda(g'(\lambda))^2) = N(0, \lambda e^{-2\lambda})$.

7. ASYMPTOTICKÉ INTERVALOVÉ ODHADY

Předchozí příklady nás mohou snadno navést na myšlenku například pro Bernoulliho pokusy *přibližně určit* i neznámou pravděpodobnost p . Víme-li, kolik pokusů n jsme učinili a kolik úspěchů u jsme zaznamenali, pak pro každé p víme, jaké je pravděpodobnost, že náhodná veličina U (počet úspěchů) bude „blízko“ zaznamenané hodnoty u . Co to však znamená *blízko*?

Zvolíme-li pevně délku $2d$ intervalu $(a - d, a + d)$, pak v normálním rozdělení $N(0, 1)$ se jako nejpravděpodobnější ukazuje být interval $(-d, d)$ s pravděpodobností $\Phi(d) - \Phi(-d) = 2\Phi(d) - 1$.

Má-li tento interval mít předepsanou pravděpodobnost alespoň $1 - \alpha$, pak d zvolíme s ohledem na tento požadavek, tedy $d = \Phi^{-1}(1 - \alpha/2)$. Asymptoticky dostáváme

$$P \left[-d \leq \frac{U - np}{\sqrt{np(1-p)}} \leq d \right] \rightarrow 1 - \alpha \Rightarrow P \left[\left(\frac{U}{n} - p \right)^2 \leq \frac{d^2}{n} p(1-p) \right] \rightarrow 1 - \alpha$$

Odtud dostáváme po obvyklých úpravách konečný výsledek

$$P \left[\frac{U}{n + d^2} \leq p \leq \frac{U + d^2}{n + d^2} \right] \rightarrow 1 - \alpha.$$

Uvědomme si hlavně, že náhodnou veličinou je zde U , čili náhodné jsou *meze intervalu* $[U/(n + d^2), (U + d^2)/(n + d^2)]$. Tento interval v sobě tedy obsahuje neznámou hodnotu p s pravděpodobností blížící se $1 - \alpha$, jde tedy o nějaký *intervalový odhad*.

Zde jsme použili znalost centrální limitní věty pro alternativní rozdělení a z toho plynoucí znalost rozptylu. To však není vždy možné. Nejprve

Definice 31 (Intervalový odhad). Buď X_1, \dots, X_n náhodný výběr z P_θ , θ neznámý parametr a $\Theta \subset \mathbb{R}$. Intervalovým odhadem nazveme dvojici funkcí $L(X_1, \dots, X_n) \rightarrow \mathbb{R}$ a $U(X_1, \dots, X_n) \rightarrow \mathbb{R}$, jejichž předpis nezávisí na θ a které splňují $P[L(X_1, \dots, X_n) \leq \theta \leq U(X_1, \dots, X_n)] \geq 1 - \alpha$ (Interval spolehlivosti s hladinou spolehlivosti $1 - \alpha$) pro každé θ , je-li θ skutečnou hodnotou parametru.

Najít obecně intervalový odhad není snadné. My si ukážeme jen asymptotickou cestu k odhadu střední hodnoty.

Obecný postup k získání intervalového odhadu asi popsat nejde. Jeden z možných postupů je tento:

- (1) Najdeme funkci $H : (X_1, \dots, X_n, \theta)$, jejíž rozdělení nezávisí na θ . (nebo alespoň nezávisí asymptoticky). Obvykle pomůže Centrální limitní věta. Příkladem je X_1, \dots, X_n s alternativním rozdělením závislém na p , odhadem může být $\frac{\sum x_i - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0, 1)$ nezávislé na p .
- (2) Najdeme kvantily $l(\alpha), u(\alpha)$ rozdělení H takové, že $P[l(\alpha) \leq H \leq u(\alpha)] \geq 1 - \alpha$. Je-li (asymptotické) rozdělení H normální $N(0, 1)$, pak použijeme kvantily $l(\alpha) = q(\alpha/2) = \Phi^{-1}(\frac{\alpha}{2})$ a $u(\alpha) = q_{1-\alpha/2} = \Phi^{-1}(1 - \frac{\alpha}{2})$.
- (3) Osvobození parametru různými algebraickými úpravami dojdeme k nerovnosti $l(\alpha) \leq H \leq u(\alpha)$ právě, když $L(X_1, \dots, X_n, l, u) \leq \theta \leq U(X_1, \dots, X_n, l, u)$ je-li to možné. Avšak někdy ještě musíme zapojit Cramérovu—Sluckého větu.

Uvažujme náhodný výběr X_1, \dots, X_n z rozdělení s konečným rozptylem σ^2 a se střední hodnotou $E X$. Vyjdeme-li s CLV, pak dostáváme

$$P \left[|\bar{X}_n - E X| \leq \sqrt{\frac{\sigma^2}{n}} \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \right] \rightarrow 1 - \alpha$$

odkud po úpravách plyne

$$P \left[\bar{X}_n - \sqrt{\frac{\sigma^2}{n}} \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \leq E X \leq \bar{X}_n + \sqrt{\frac{\sigma^2}{n}} \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \right] \rightarrow 1 - \alpha.$$

Tento intervalový odhad ale závisí na *neznámé* hodnotě rozptylu σ^2 . V některých případech je rozptyl jen funkcí střední hodnoty (známe Bernoulliovo, Poissonovo, exponenciální, geometrické rozdělení, kde tomu tak je) a nerovnosti lze řešit, nebo někdy alespoň numericky řešit. V obecném případě však $E X$ a σ^2 nemusejí souviset. Naštěstí máme dobrý odhad pro σ^2 a nástroj v podobě Sluckého (Cramérovu—Sluckého) věty.

Věta 34 (Sluckého věta). Buď $X_n \xrightarrow{d} N(0, 1)$, $U_n \xrightarrow{P} a$ a $Z_n \xrightarrow{P} s > 0$. Pak

$$Z_n X_n + U_n \xrightarrow{d} N(a, s^2)$$

Důkaz. Ukážeme si základní myšlenku důkazu jen pro $U_n, a = 0$. Pro libovolné $\varepsilon > 0$ platí

$$P[Z_n X_n \leq x] = P[Z_n X_n \leq x, |Z_n - s| \leq \varepsilon] + P[Z_n X_n \leq x, |Z_n - s| > \varepsilon].$$

Druhá pravděpodobnost s roztoucím n jde k nule. První pravděpodobnost je omezena

$$P[(1 + \varepsilon)X_n \leq x, |Z_n - s| \leq \varepsilon] \leq P[Z_n X_n \leq x, |Z_n - s| \leq \varepsilon] \leq P[(1 - \varepsilon)X_n \leq x, |Z_n - s| \leq \varepsilon]$$

přičemž krajní pravděpodobnosti se blíží $\Phi(x/(1+\varepsilon))$, případně $\Phi(x/(1-\varepsilon))$. Důkaz dokončíme poukazem na skutečnosti, že $\varepsilon > 0$ může být libovolně malé a distribuční funkce Φ je spojitá. \square

Tato věta se nejčastěji používá, když neznáme rozptyl nebo je práce s rozptylem příliš obtížná. V tom případě potom víme, že díky CLV $\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{d} N(0, 1)$ a také ze zákona velkých čísel $\frac{S_n}{\sigma} \xrightarrow{P} 1$ pokud $0 < \sigma^2 < \infty$.

Díky Sluckého větě pak $\frac{(\sqrt{n} \cdot \bar{X}_n - \mu)/\sigma}{S_n/\sigma} = \sqrt{n} \frac{\bar{X}_n - \mu}{S_n} \xrightarrow{d} N(0, 1)$, tedy jsme nahradili rozptyl jeho výběrovým odhadem.

Příklad. Mějme náhodný výběr X_1, \dots, X_n z neznámého rozdělení s konečným rozptylem. Chceme najít intervalový odhad o spolehlivosti $1 - \alpha$ pro neznámou střední hodnotu μ .

Kombinací CLV a Sluckého větě dostáváme

$$P \left[\frac{\bar{X}_n - \mu}{S_n} \sqrt{n} \leq x \right] \rightarrow \Phi(x).$$

Podobně jako výše tedy můžeme po nahrazení neznámého rozptylu jeho *konzistentním* odhadem odvodit

$$P \left[\bar{X}_n - \sqrt{\frac{S_n^2}{n}} \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \leq E X \leq \bar{X}_n + \sqrt{\frac{S_n^2}{n}} \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \right] \rightarrow 1 - \alpha.$$

7.1. Porovnání dvou středních hodnot. Jednou z nejdůležitějších základních aplikací CLV a Sluckého vět, případně delta-věty je porovnání středních hodnot dvou náhodných výběrů.

Obvyklé předpoklady a značení jsou:

- (1) Existují dvě populace (skupiny) u nichž se zajímáme o nějakou hodnotu—náhodnou veličinu. Například kvalitu piva, rychlost výpočtu, rozměr, krevní tlak, ...
- (2) Náhodný výběr z populace 1 označíme X_1, \dots, X_n , náhodný výběr z populace 2 označíme Y_1, \dots, Y_n . Obecně lze uvažovat i rozdílné rozsahy výměrů n a m , ale toto vede k určitému ztížení úlohy.
- (3) Předpokládáme, že oba výběry pochází ze stejného typu rozdělení, lišícího se jen svými parametry. Značíme μ_X střední hodnotu X_1 a μ_Y střední hodnotu Y_1 , podobně σ_X^2 rozptyl X_1 a σ_Y^2 rozptyl Y_1 . Obvyklý předpoklad je, že odlišné mohou být pouze střední hodnoty, ale i tyto předpoklady je možné omezovat.

Víme, že $X - Y$ je náhodná veličina se střední hodnotou $\mu_X - \mu_Y$ a s rozptylem $\sigma_X^2 + \sigma_Y^2$. Zároveň platí $\sum (X_i - Y_i) = \sum X_i - \sum Y_i$. Použitím CLV dostaneme

$$P \left[\frac{\sum (X_i - Y_i) - (\mu_X - \mu_Y)}{\sqrt{n(\sigma_X^2 + \sigma_Y^2)}} \leq x \right] = P \left[\frac{\bar{X}_n - \bar{Y}_n - (\mu_X - \mu_Y)}{\sqrt{\sigma_X^2 + \sigma_Y^2}} \sqrt{n} \leq x \right] \rightarrow \Phi(x).$$

Neznáme-li (a obvykle to neznáme) hodnotu rozptylů σ_X^2 a σ_Y^2 , pak můžeme použít jejich konzistentní odhady S_X^2 a S_Y^2 a Sluckého větu a dostaneme

$$P \left[\frac{\bar{X}_n - \bar{Y}_n - (\mu_X - \mu_Y)}{\sqrt{S_X^2 + S_Y^2}} \sqrt{n} \leq x \right] \rightarrow \Phi(x).$$

Odtud již snadno odvodíme intervalový odhad na hladině spolehlivosti $1 - \alpha$ pro rozdíl středních hodnot $\mu_X - \mu_Y$

$$P \left[\bar{X}_n - \bar{Y}_n - \sqrt{\frac{S_X^2 + S_Y^2}{n}} \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \leq \mu_X - \mu_Y \leq \bar{X}_n - \bar{Y}_n + \sqrt{\frac{S_X^2 + S_Y^2}{n}} \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \right] \rightarrow 1 - \alpha.$$

Jak tento odhad použít? Jedna z možností je při rozhodování o shodě středních hodnot.

Příklad (Statistické rozhodnutí). Uvažujme dvě města. Máme rozhodnout, zda děti z obou měst mají stejnou šanci zvládnout test. Nášim předpokladem je, že ano a tento předpoklad chceme „potvrdit“. Protože víme, že naše rozhodnutí je založené na náhodných veličinách, stanovíme si hladinu α maximální přípustné pravděpodobnosti chybného rozhodnutí. Hodnotu α obvykle volíme 0,05 (případně v rozmezí 0,01 – 0,1, ale volba α je samostatný problém).

Vybereme n dětí z jednoho města a n dětí z druhého města. Musí jít o náhodný výběr. Všechny děti budou psát test. V prvním městě dostaneme výsledky X_1, \dots, X_n (body za test), ve druhém městě Y_1, \dots, Y_n . Rozptyl výsledků neznáme a musíme jej odhadnout. Odhadneme též střední bodové zisky v obou městech:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i, S_X^2 = \frac{1}{n-1} \sum (X_i - \bar{X}_n)^2, \quad \bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i, S_Y^2 = \frac{1}{n-1} \sum (Y_i - \bar{Y}_n)^2.$$

Spočítáme intervalový odhad pro $\mu_X - \mu_Y$ na hladině spolehlivosti $1 - \alpha$. Již víme, co to znamená. Naší hypotézou je, že $\mu_X - \mu_Y = 0$. Nepokrývá-li tedy intervalový odhad hodnotu 0, pak zavrhneme naši hypotézu a řekneme, že děti z těchto měst nemají šanci zvládnout test stejně dobře (a dodáváme „se statistickou významností α “, čímž sdělujeme podstoupené riziko chybného zavržení předpokladu $\mu_X - \mu_Y = 0$).

Rozmyslete. Samozřejmě můžeme rozhodovat o libovolném rozdílu středních hodnot. Budeme-li předpokládat, že $\mu_X - \mu_Y = \delta$, pak tuto hypotézu zamítneme s hladinou významnosti α , pokud v intervalovém odhadu pro $\mu_X - \mu_Y$ není obsažena hodnota δ .

7.2. Odhady pravděpodobností. Chceme-li odhadnout v podstatě libovolnou pravděpodobnost, lze použít empirické četnosti, tedy obdobu empirické distribuční funkce. Tím se vlastně problém redukuje na odhad pravděpodobnosti v Bernoulliových pokusech. Lze však postupovat i jinak, máme-li parametrický model pro rozdělení.

Příklad. Chceme odhadnout pravděpodobnost $p_1 = P[X > 1]$ a odvodit asymptotický intervalový odhad pro p_1 .

Víme, že $p_1 = 1 - F(1)$, tedy že můžeme použít empirickou distribuční funkci $\hat{F}_n(1)$, která je konzistentním a nestranným odhadem. Navíc ale také platí asymptotická normalita e.d.f.

Protože $\hat{F}_n(x) = 1/n \sum \chi(X_i \leq x)$, jde o aritmetický průměr nezávislých stejně rozdělených náhodných veličin. A pro ten platí CLV, pokud máme zaručen nenulový a konečný rozptyl, což zde triviálně platí, neboť charakteristická funkce množiny χ nabývá jen hodnot 0 a 1. S pomocí věty 13 víme, že rozptyl $\text{var } \hat{F}_n(x) = 1/n F(x)(1 - F(x))$. Tím máme vše připraveno pro použití CLV.

$$P \left[\frac{\hat{F}_n(x) - F(x)}{\sqrt{F(x)(1 - F(x))}} \sqrt{n} \leq x \right] \rightarrow \Phi(x),$$

což s pomocí snadných operací a s využitím Sluckého věty po nahrazení rozptylu $F(x)(1 - F(x))$ jeho konzistentním odhadem $\hat{F}_n(x)(1 - \hat{F}_n(x))$ vede k asymptotickému intervalovému odhadu

$$P \left[\hat{F}_n(x) - \sqrt{\frac{\hat{F}_n(x)(1 - \hat{F}_n(x))}{n}} \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \leq F(x) \leq \hat{F}_n(x) + \sqrt{\frac{\hat{F}_n(x)(1 - \hat{F}_n(x))}{n}} \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \right] \rightarrow 1 - \alpha.$$

Snadnou úpravou pak dostaneme asymptotický intervalový odhad pro $1 - F(1)$.

Přidejme předpoklad, že distribuční funkce F odpovídá exponenciálnímu rozdělení. Pak víme, že $1 - F(1) = e^{-\lambda}$. Víme, že střední hodnota $E X = 1/\lambda$ a můžeme tedy uvažovat odhad s využitím asymptoticky normálního odhadu pro $1/\lambda$. Platí

$$P \left[\frac{\bar{X}_n - \lambda^{-1}}{\sqrt{\lambda^{-2}}} \sqrt{n} \leq x \right] \rightarrow \Phi(x) \Rightarrow \sqrt{n}(\bar{X}_n - \lambda^{-1}) \xrightarrow{d} N(0, \lambda^{-2}).$$

S využitím delta věty na funkci $g(u) = \exp(-u^{-1})$ dostaneme

$$\sqrt{n}(\exp(-1/\bar{X}_n) - e^{-\lambda}) \xrightarrow{d} N(0, \lambda^2 e^{-2\lambda}),$$

neboť $(g'(u))^2 = u^{-4} \exp(-u^{-1})$.

Rozmyslete. Dobře si však rozmyslete, že metoda odhadu pravděpodobnosti uvedená výše je použitelná jen pro pravděpodobnost jediného jevu. Důvod je ten, že náhodné veličiny $\chi(X \in A)$ a $\chi(X \in B)$ nejsou nezávislé.

8. BORELŮV A CANTELLIŮV ZÁKON, CHERNOFFOVY MEZE

Definice 32 (Množinový lim sup a lim inf). Buď $A_n, n = 1, 2, \dots$ množiny (náhodné jevy). Definujme

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i \quad \liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{i=n}^{\infty} A_i$$

Jestliže $a \in \limsup_{n \rightarrow \infty} A_n$, potom existuje nekonečně mnoho množin A_i , kde $a \in A_i$. Jestliže $b \in \liminf_{n \rightarrow \infty} A_n$, potom existuje nejvýše konečně mnoho množin A_i , kde $b \notin A_i$.

Věta 35 (Borelův-Cantelliův 0-1 zákon). Předpokládejme, že máme nekonečný soubor náhodných jevů A_1, A_2, \dots .

(1) Jestliže

$$\sum_{i=1}^{\infty} P(A_i) < \infty, \text{ pak } P\left(\limsup_{n \rightarrow \infty} A_n\right) = 0.$$

(2) Buď $A_n, n = 1, 2, \dots$ navíc nezávislé. Potom

$$\sum_{i=1}^{\infty} P(A_i) = \infty \Rightarrow P\left(\limsup_{n \rightarrow \infty} A_n\right) = 1.$$

Věta 36 (Chernoffovy meze). Buď X náhodná veličina. Označme $\psi_X(t) = Ee^{tX}$ její momentovou vytvořující funkci. Pak

$$P[X \geq a] \leq \min_{t>0} \frac{\psi_X(t)}{e^{ta}}, \quad P[X \leq a] \leq \min_{t<0} \frac{\psi_X(t)}{e^{ta}}$$

Důkaz. První mez plyne z $P[X \geq a] = P[e^{tX} \geq e^{ta}]$ pro $t > 0$. Dostali jsme nezápornou veličinu, využijeme Markovovu nerovnost, tedy $P[e^{tX} \geq e^{ta}] \leq \frac{Ee^{tx}}{e^{ta}} \forall t > 0$. Poté vezmeme minimum přes všechna t .

Druhá mez je dokázána analogicky. $P[X \leq a] = P[e^{tX} \geq e^{ta}] \leq \frac{Ee^{tx}}{e^{ta}}$ pro $t < 0$. \square

Definice 33 (Poissonovské pokusy). Buďte X_1, X_2, \dots nezávislé náhodné veličiny takové, že $P[X_i = 1] = p_i = 1 - P[X_i = 0], p_i \in (0, 1)$. Takové pokusy se nazývají nezávislé poissonovské pokusy.

Jestliže u poissonovských pokusů bude každé $p_i = p$ pro nějaké konstantní p , dostaneme bernoulliovské pokusy.

Věta 37 (Horní Chernoffovy meze pro poissonovské pokusy). Buďte X_1, X_2, \dots nezávislé poissonovské pokusy s parametry p_1, p_2, \dots . Označme $S_n = \sum_{i=1}^n X_i, \mu_S = \sum_{i=1}^n p_i$.

(1) Pro $\delta > 0$:

$$P[S_n \geq (1 + \delta)\mu_s] \leq \left(\frac{e^\delta}{(1 + \delta)^{(1+\delta)}} \right)^{\mu_s}$$

(2) Pro $0 < \delta \leq 1$:

$$P[S_n \geq (1 + \delta)\mu_s] \leq e^{-\mu_s \delta^2 / 3}$$

(3) Pro $\delta > 6\mu_s$:

$$P[S_n > \delta] \leq 2^{-\delta}$$

Důkaz. Jen první. $P[S_n \geq (1 + \delta)\mu_s] \leq \frac{\psi_{S_n}(t)}{e^{t(1+\delta)\mu_s}}$.

Z nezávislosti dostáváme $\psi_{S_n} = \prod_{i=1}^n \psi_{x_i}(t)$ a $\psi_{x_i}(t) = Ee^{tX_i} = p_i e^t + (1 - p_i) = 1 + p_i(e^t - 1)$.

Dále $\prod_{i=1}^n (1 + p_i(e^t - 1)) = \exp\{\sum_{i=1}^n \log(1 + p_i(e^t - 1))\} \leq \exp\{(e^t - 1) \sum_{i=1}^n p_i\} = \exp\{(e^t - 1)\mu_s\}$.

Dosazením do vzorce, dostáváme $\frac{\prod_{i=1}^n (1 + p_i(e^t - 1))}{e^{t(1+\delta)\mu_s}} \leq \frac{\exp\{(e^t - 1)\mu_s\}}{\exp t(1+\delta)\mu_s} = \exp \mu_s(e^t - 1 - t(1 + \delta))$.

Nakonec $P[S_n \geq (1 + \delta)\mu_s] \leq \frac{\prod_{i=1}^n (1 + p_i(e^t - 1))}{e^{t(1+\delta)\mu_s}} = \exp \mu_s[\delta - \log(1 + \delta)^{1+\delta}] =$, což je hledaný výraz. \square

Věta 38 (Dolní Chernoffovy meze pro poissonovské pokusy). *Bud' X_1, \dots, X_n nezávislé poissonovské pokusy s pravděpodobnostmi $P[X_i = 1] = p_i$. Bud' $S_n = \sum_{i=1}^n X_i$ a $\mu_s = \sum_{i=1}^n p_i$. Pak*

(1) Pro $0 < \delta < 1$

$$P[S_n \leq (1 - \delta)\mu_s] \leq \left(\frac{e^{-\delta}}{(1 - \delta)^{1-\delta}} \right)^{\mu_s}.$$

(2) Pro $0 < \delta \leq 1$

$$P[S_n \leq (1 - \delta)\mu_s] \leq e^{-\mu_s \delta^2 / 2}.$$

Rozmyslete. Uvažujte n nezávislých Bernouliových pokusů X_1, \dots, X_n s (neznámou) pravděpodobností úspěchu $p \in (0, 1)$. Označme

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i$$

nejlepší odhad p . Pro zadané $\alpha \in (0, 1)$ použijte Čebyševovu a Markovovu nerovnost a Chernoffovy meze pro nalezení nejmenší hodnoty δ takové, že

$$P[p \in (\hat{p} - \delta, \hat{p} + \delta)] \geq 1 - \alpha.$$

9. ROZDĚLENÍ TRANSFORMACE, SOUČTU A SOUČINU

Věta 39 (Rozdělení součtu a součinu diskrétních náhodných veličin). *Bud' X, Y diskrétní náhodné veličiny, pak:*

- (1) $Z = X + Y$ je také diskrétní náhodná veličina a platí $P[Z = v] = \sum_u P[X = u, Y = v - u]$.
- (2) Pokud jsou X, Y kladné (tedy $P[X > 0] = P[Y > 0] = 1$), pak $V = XY$ je také kladná diskrétní náhodná veličina a $P[V = z] = \sum_u P[X = u, Y = \frac{z}{u}]$. pro obecné d.n.v je třeba dát pozor na $ab = -(a)(-b)$ a $0b = a0 = 0$.

Důkaz. Podle věty o úplné pravděpodobnosti $P[Z = v] = \sum_u P[Z = v, X = u] = \sum_u P[X = u, X + Y = v] = \sum_u P[X = u, Y = v - u]$. Druhý bod analogicky. Pro více, než dvě veličiny můžeme postupovat indukcí. \square

Věta 40 (Konvoluce). *Bud' (X, Y) spojitý náhodný vektor s hustotou $f_{(X,Y)}$. Pak $Z = X + Y$ je spojitá náhodná veličina a platí*

$$f_Z(t) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, t - x) dx$$

Důkaz. Chceme najít $P[Z \leq z] = P[X + Y \leq z] = P[(X, Y) \in H]$ kde H je polorovina rozdělená přímkou $z = x + y$. To se rovná

$$\iint_H f_{(X,Y)}(u, v) dv du = \int_{-\infty}^{\infty} \int_{-\infty}^{z-u} f_{(X,Y)}(u, v) dv du = F_Z(z).$$

Ze vztahu distribuční funkce a hustoty: $F_Z(z) = \int_{-\infty}^z f(t) dt$ víme, že platí $f_Z(z) = F'_Z(z)$ (až na „nemnoho“ bodů) v nichž můžeme f definovat podle potřeby.

Potřebujeme tedy najít parciální derivaci

$$\frac{\partial}{\partial z} \int_{-\infty}^{\infty} \int_{-\infty}^{z-u} f_{(X,Y)}(u, v) dv du = \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial z} \int_{-\infty}^{z-u} f_{(X,Y)}(u, v) dv \right) du = \int_{-\infty}^{\infty} f_{(X,Y)}(u, z - u) du.$$

Dostáváme tedy hledaný výraz. \square

Věta 41 (Hustota podílu a součinu). *Bud' (X, Y) spojitý náhodný vektor s hustotou $f_{(X,Y)}$ a necht' $P[Y > 0] = 1$. Pak $V = \frac{X}{Y}$ je spojitá náhodná veličina s hustotou*

$$f_V(v) = \int_{-\infty}^{\infty} f_{(X,Y)}(vu, u) \cdot u \, du.$$

Náhodná veličina $W = XY$ je také spojitá s hustotou

$$f_W(w) = \int_{-\infty}^{\infty} f_{(X,Y)}\left(\frac{w}{y}, y\right) \cdot \frac{1}{y} \, dy.$$

Důkaz. Opět nejprve spočítáme $P[V \leq v] = P[\frac{X}{Y} \leq v] = P[X \leq vY] = P[(X, Y) \in H]$ kde H je polorovina rozdělená přímkou $x = vy$. Postupujeme tedy podobně, jako u předchozího důkazu.

Dostaneme dvojný integrál

$$\iint_H f_{(X,Y)}(u, t) \, dt \, du = \int_{-\infty}^{\infty} \int_{-\infty}^{vu} f_{(X,Y)}(u, t) \, dt \, du = F_V(v).$$

Při derivování podle v si musíme uvědomit, že v horní mezi vnitřního integrálu není v , ale vu a i tento výraz musíme zderivovat. Tím dostaneme požadovaný výsledek.

Hustotu součinu dostaneme přímým dosazením obrácené hodnoty Y do podílu. \square

Věta (Momentová vytvořující funkce součtu). *Budte X_1, \dots, X_n nezávislé náhodné veličiny a Y jejich součet. Pak*

$$\psi_Y(t) = \prod_{i=1}^n \psi_{X_i}(t).$$

Důkaz. Z definice plyne, že $\psi_Y(t) = E e^{tY} = E e^{t \sum X_i} = E \prod e^{tX_i}$. Z nezávislosti poté máme $\prod \psi_{X_i}(t)$. \square

10. PODMÍNĚNÉ ROZDĚLENÍ A PODMÍNĚNÁ STŘEDNÍ HODNOTA

Definice 34 (Podmíněné rozdělení diskrétního náhodného vektoru). *Bud' $X = (Y, Z)$ diskrétní náhodný vektor (s hodnotami v \mathbb{N}_0^2). Pak pro každé $n \in \mathbb{N}_0$ takové, že $P[Z = n] > 0$ definujeme podmíněné rozdělení Y za podmínky $Z = n$:*

$$P[Y = k | Z = n] = \frac{P[Y = k, Z = n]}{P[Z = n]}.$$

Podmínka $Z = n$ může přidat další informaci také pro náhodnou veličinu Y . Samozřejmě, pokud Y a Z jsou nezávislé, pak informace $Z = n$ žádné upřesnění pro Y nepřináší.

Příklad. Mějme 6 mincí a jednu kostku. Hodíme kostkou a podle výsledku hodíme odpovídající počet mincí. Jaká je pravděpodobnost, že nepadne žádný líc?

Jaké je rozdělení počtu líců? Víme, že počet líců je $0, \dots, 6$. To však nedokážeme spočítat přímo, musíme vždy spočítat rozdělení za podmínky hození kostkou.

Definice 35 (Podmíněná střední hodnota). *Bud' $X = (Y, Z)$ diskrétní náhodný vektor takový, že*

- (1) $P[Z = n] > 0$
- (2) $E g(Y, n) < \infty$ pro zvolenou funkci g .

Pak definujeme $E g(Y, Z | Z = n) = \sum_{k=0}^{\infty} g(k, n) P[Y = k | Z = n]$.

Speciálně, pokud $E|Y| < \infty$, potom $E(Y | Z = n) = \sum_{k=0}^{\infty} k P[Y = k | Z = n]$, což nazveme *podmíněnou střední hodnotou* $g(Y, Z)$ za podmínky $Z = n$.

Všimneme si, že podmíněná střední hodnota *závisí* na hodnotě n , ale pochopitelně *nezávisí* na hodnotě Y .

Věta 42 (O úplné střední hodnotě). *Bud' $X = (Y, Z)$ diskrétní náhodný vektor a EY existuje konečná. Pak*

$$EY = \sum_n E(Y | Z = n) P[Z = n] \text{ pokud } P[Z = n] > 0.$$

Důkaz. Pro výraz na pravé straně platí

$$\sum_n E[Y|Z = n]P[Z = n] = \sum_n \sum_{k=0}^{\infty} kP[Y = k|Z = n]P[Z = n] = \sum_n \sum_{k=0}^{\infty} kP[Y = k, Z = n].$$

Jelikož předpokládáme existenci konečné střední hodnoty EY , můžeme zaměnit pořadí sčítání a získáme tak

$$\sum_{k=0}^{\infty} k \sum_n P[Y = k, Z = n] = \sum_{k=0}^{\infty} kP[Y = k] = EY.$$

□

Všimněme si, že podmíněná střední hodnota je konstanta závislá na n a jejich součet přes všechna n vážený pomocí pravděpodobností $P[Z = n]$ dává (marginální) střední hodnotu EY , která již na n pochopitelně nezávisí.

Můžeme si také všimnout, že podmíněná pravděpodobnost rozděluje (faktorizuje) množinu Ω podle hodnoty náhodné veličiny Z . To nám umožní nahlédnout podmíněnou střední hodnotu jako náhodnou veličinu, funkci Z .

Definice 36 (Podmíněná střední hodnota jako náhodná veličina). Buď $X = (Y, Z)$ diskretní náhodný vektor, $E|Y| < \infty$. Definujeme náhodnou veličinu $E(Y|Z)$ předpisem $E(Y|Z)(\omega) = E(Y|Z = n) \forall \omega \in \{\omega : Z(\omega) = n\}$.

Snadno nahlédneme, že rozdělení $E(Y|Z)$ je diskretní a platí $P[E(Y|Z) = E(Y|Z = n)] = P[Z = n]$.

Střední hodnota EY minimalizuje střední čtvercovou vzdálenost Y od konstanty, neboli: $E(Y - EY)^2 = \min_{a \in \mathbb{R}} E(Y - a)^2$. Střední hodnota je tedy nejlepší predikce neznámé hodnoty Y pomocí konstanty, chceme-li nejmenší čtvercovou chybu. Na množině $\{\omega : Z(\omega) = n\}$ platí $E(Y - E(Y|Z = n))^2 = \min_{b \in \mathbb{R}} E(Y - b)^2$. Na této podmnožině jsme tedy našli lepší odhad. Dáme-li si tyto skutečnosti dohromady, pak vidíme, že $E(Y|Z)$ je mezi všemi funkcemi Z tím nejlepším odhadem Y s ohledem na střední čtvercovou chybu. Z definice také vyplývá, že $E(g(Y)|Y) = g(Y)$ pro každou (měřitelnou) funkci g . Co je v souladu s interpretací: známe-li hodnotu Y , pak nejlepším odhadem $g(Y)$ je dosažení známého Y do g .

Věta 43. Buď $X = (Y, Z)$ diskretní náhodný vektor, $E|Y| < \infty$. Pak $E(E(Y|Z)) = EY$.

Důkaz. Přímý důsledek věty 42. □

Poznámka (Podmíněný rozptyl). Podobně jako podmíněnou střední hodnotu můžeme definovat i podmíněný rozptyl.

$$\text{var}(Y|Z = n) = E\left((Y - E(Y|Z = n))^2 \mid Z = n\right).$$

Snadným výpočtem zjistíme, že

$$\text{var}(Y|Z) = E(Y^2|Z) - (E(Y|Z))^2, \quad \text{var } Y = E \text{var}(Y|Z) + \text{var } E(Y|Z).$$

Podstatně složitější situace nastává, chceme-li podmiňovat spojitou náhodnou veličinou. Pravděpodobnosti $P[Z = z]$ jsou nulové pro všechna z a jak víme, nulou nelze dělit. Přesto i zde je možné dobře definovat podmíněné rozdělení pomocí podmíněné hustoty při zachování všech vlastností, které rozdělení má mít.

Definice 37 (Podmíněná hustota). Buď $X = (Y, Z)$ absolutně spojitý náhodný vektor se sdruženou hustotou $f_{Y,Z}$. Definujme $f_{Y|Z}(v|w)$ předpisem:

$$f_{Y|Z}(v|w) = \begin{cases} \frac{f_{Y,Z}(v,w)}{f_Z(w)} & f_Z(w) > 0 \\ 0 & \text{jinak} \end{cases}$$

a tuto funkci (proměnné v indexovanou z) nazveme hustotou podmíněného rozdělení Y za podmínky $Z = w$.

Nyní pojd' me ověřit, že takto definovaná hustota je opravdu hustotou.

Marginální hustota je $f_Z = \int_{-\infty}^{\infty} f_{Y,Z}(v, w) dv$. Pokud je rovna nule, potom nutně $f_{Y,Z}(v, w) =$

0. Potom

$$\int_{-\infty}^{\infty} f_{Y|Z}(v|w) dv = \int \frac{f_{Y,Z}(v, w)}{f_Z(w)} dv = \frac{\int_{-\infty}^{\infty} f_{Y,Z}(v, w) dv}{f_Z(w)} = \frac{f_Z(w)}{f_Z(w)}$$

kdykoliv $f_Z(w) > 0$.

Podívejme se ještě na pravděpodobnost $P[Y \in A, Z \in B]$. Z definice víme, že

$$P[Y \in A, Z \in B] = \int_A \int_B f_{Y,Z}(v, w) dw dv = \int_A \int_B f_{Y|Z}(v|w) f_Z(w) dw dv = \int_B \left(\int_A f_{Y|Z}(v|w) dv \right) f_Z(w) dw.$$

Zde vidíme základ věty o úplně pravděpodobnosti pro absolutně spojitý náhodný vektor.

Definice 38 (Podmíněná střední hodnota). Buď $X = (Y, Z)$ absolutně spojitý náhodný vektor, $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ taková, že $E|g(Y, w)| < \infty$. Pak definujeme $E(g(Y, Z)|Z = w) = \int_{-\infty}^{\infty} g(v, w) f_{Y|Z}(v|w) dv$ a $E(Y|Z = w) = \int_{-\infty}^{\infty} v f_{Y|Z}(v|w) dv$.