Zápisky z pravděpodobnosti a statistiky FEL ČVUT

Přednášky Mgr. Matěj Novotný, Ph. D.

František Boháček

LS 2021, verze 200521

Obsah

1 Úvod, upozornění

Tento dokument obsahuje mé vlastní zápisky z přednášek z kurzu B0B01PST na ČVUT FEL (Letní semestr 2021). Neručím za žádné chyby, co se v dokumentu mohou objevit. Jakékoliv připomínky mi posílejte na Discord `Rutherther#8497`.

V dokumentu nejsou všechny příklady z přednášek. Zvláště ne ty, pro které je potřeba nakreslit obrázek.

Dokument by měl obsahovat všechny probrané definice, tvrzení, věty, ale může se stát, že něco chybí. V takovém případě mi to můžete také nahlásit na můj Discord.

2 Poděkování

Chtěl bych poděkovat zejména panu doktorovi Matěji Novotnému za opravu spousty chyb nebo překlepů v těchto zápiscích. Dále také Martinu Mrázovi za doručení těchto chyb ke mně.

3 Seznam klíčových pojmů

- 3.1 Pravděpodobnostní prostor

  Prostor elementárních jevů, množina všech jevů a pravděpodobnost. Elementární jev, jev. Příklady.

- 4.2 Nezávislost jevů

  Nezávislost dvou jevů, nezávislost n jevů, po dvou nezávislé jevy. Příklady.

- 5.1 Podmíněná pravděpodobnost

  Podmíněná pravděpodobnost vs. nezávislost. Příklady.

- 6.2 Náhodná veličina

  7.1 Diskrétní náhodná veličina, 7.3 Spojitá náhodná veličina. 7.4 Distribuční funkce náhodné veličiny a její existence. Příklady.

- Charakteristiky náhodných veličiny

  Střední hodnota diskrétní/spojité náhodné veličiny. Rozptyl náhodné veličiny, směrodatná odchylka.

- Příklady pravděpodobnostních rozdělení

  Alternativní rozdělení, binomické rozdělení, rovnoměrné rozdělení na intervalu. Gaussovo (normální) rozdělení, normované normální rozdělení, Paretovo rozdělení. Jejich charakteristiky, příklady.

- Náhodné vektory

  Náhodný vektor, sdružená distribuční funkce. Spojitě rozdělený n. vektor, diskrétně rozdělený n. vektor. Marginální rozdělení.

- Charakteristiky náhodných vektorů

  Střední hodnota n. vektoru, kovariance veličin, korelace veličin. Varianční matice vektoru, korelační matice vektoru. Vztah kovariance a rozptylu.

- Nezávislost náhodných veličin

  Nezávislost n náhodných veličin, souvislost s distribučními funkcemi. Nezávislost (spočetně) mnoha náhodných veličin.

- Náhodný výběr

  Náhodný výběr, realizace náhodného výběru.

- Bodový odhad parametru

  Výběrový průměr, výběrový rozptyl, výběrová směrodatná odchylka. Nestranný odhad parametru.

- Intervalový odhad

  Intervalový odhad, spolehlivost odhadu.

- Testování hypotéz

  Test nulové hypotézy proti alternativní, chyba 1. druhu, chyba 2. druhu, kritický obor. Jednovýběrový T-test.

Přednáška 1: Základy pravděpodobnosti

17. 2. 2021

Příklad. Dva hráči hrají hru, u níž mají oba stejnou šanci na výhru. Hrají do té doby, než jeden z nich získá 5 výher. Přijde bouře, hra musí skončit za stavu 4:3. V jakém poměru by si měli rozdělit odměnu určenou tomu, kdo první vyhraje 5x.

Šance na výhru 1 … a

Šance na výhru 2 … b

Přednáška 2: Definice pravděpodobnosti

18. 2. 2021

Definice 2.1. Sigma algebra

Mějme množinu Ω. Množinu $\mathcal{A} \subseteq 2^\Omega$ nazveme σ-algebrou, pokud:

- $\emptyset \in \mathcal{A}$
- $A \in \mathcal{A} \implies \overline{A} \in \mathcal{A}$
- $A_1, A_2, A_3, \ldots \in \mathcal{A} \implies P\left(\cup_{n=1}^{\infty} \in \mathcal{A}\right)$

Definice 2.2. Buďte Ω neprázdná množina, $\mathcal{A}$ σ-algebra na Ω.

Funkci P: $\mathcal{A} \to [0,1]$, splňující:

1. $P(A) \geq 0, A \in \mathcal{A}$
2. pro systém navzájem disjunktních množin $A_1, A_2, A_3, \ldots, A_n \in \mathcal{A}$ platí, že $P(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$
3. $P(\Omega) = 1$
4. $(P(\emptyset) = 0)$ lze vypustit, vyplývá z předchozích

nazveme pravděpodobností na Ω.

Příklad. Hodíme kostkou

$\Omega = \{1, 2, 3, 4, 5, 6\}$ (množina elementárních jevů, každému prvku Ω říkáme elementární jev)

$\mathcal{A} = 2^\Omega$ je množina jevů (např. $\{2, 4, 6\}, \{1, 2\}, \{1\}$)

$A \in \mathcal{A}, P(A) = \frac{|A|}{|\Omega|}$

Přednáška 3: Pravděpodobnostní prostor

24. 2. 2021

Definice 3.1. Pravděpodobnostní prostor

Mějme neprázdnou množinu Ω, σ-algebrou $\mathcal{A}$ na množině Ω a pravděpodobnost $P: \mathcal{A} \to \mathbb{R}$.

Uspořádanou trojici $(\Omega, \mathcal{A}, P)$ nazýváme pravděpodobnostní prostor, množinu Ω množina elementárních jevů a $\mathcal{A}$ nazýváme množinou jevů, speciálně $\emptyset$ je jev nemožný, Ω je jev jistý.

Věta 3.2. Nechť $(\Omega, \mathcal{A}, P)$ je pstní prostor, potom:

1. $A \in \mathcal{A}: \; P(A) + P(\overline{A}) = 1$
2. $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots, A_i \in \mathcal{A} \implies P(\cup_{n=1}^{\infty} A_n) = \lim_{n\to\infty} P(A_n)$

3. $A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots, A_i \in \mathcal{A} \implies P(\cap_{n=1}^{\infty} A_n) = \lim_{n\to\infty} P(A_n)$

Přednáška 4: Princip inkluze a exkluze, nezávislost

25. 2. 2021

Definice 4.1. Princip inkluze a exkluze

Mějme jevy $A_1, \ldots, A_n$ na postním prostoru $(\Omega, \mathcal{A}, P)$.

Potom platí, že pravděpodobnost

$$\begin{aligned} P(\cup_{k=1}^{n} A_k) = &\sum_{k=1}^{n} P(A_k) \\ &- \sum_{1\le i<k\le n} P(A_i \cap A_k) \\ &+ \sum_{1\le i<k<j\le n} P(A_i \cap A_k \cap A_j) \\ &- \ldots \\ &+ (-1)^{n-1} P(\cap_{k=1}^{n} A_k) \end{aligned}$$

Definice 4.2. Nezávislost jevů

Nechť A, B jsou jevy na $(\Omega, \mathcal{A}, P)$. Říkáme, že A, B jsou nezávislé, platí-li

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

Tvrzení 4.3. Tvrzení o nezávislých jevech

Nechť A, B jsou jevy na $(\Omega, \mathcal{A}, P)$.

Potom platí:

1. $A \cap B = \emptyset \implies P(A) = 0 \vee P(B) = 0 \vee A, B$ jsou závislé.
2. Dvojice $(\Omega, A), (\emptyset, A)$ jsou nezávislé $\forall A \in \mathcal{A}$.
3. A, B jsou nezávislé $\implies A, \overline{B}$ nezávislé.
4. A, B nezávislé $\implies \overline{A}, \overline{B}$ nezávislé.

Přednáška 5: Podmíněná pravděpodobnost, VoÚP

3. 3. 2021

Definice 5.1. Podmíněná pravděpodobnost

Nechť $(\Omega, \mathcal{A}, P)$ je pstní prostor. Buďte $A, B \in \mathcal{A}, P(B) > 0$.

Podmíněnou pst jevu A za podmínky B značíme $P(A|B)$ a definujeme

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

Definice 5.2. Úplný systém jevů

Buď $(\Omega, \mathcal{A}, P)$ pstní prostor. Systém jevů $A_1, A_2, A_3, \ldots, A_i \in \mathcal{A}$ nazvu úplný systém jevů, pokud platí:

1. $P(A_i \cap A_j) = 0$ pro $i \neq j$
2. $P(\cup_{i=1}^{\infty} A_i) = 1$.

Věta 5.3. Věta o úplné pravděpodobnosti

Mějme $B = \{B_1, B_2, B_3, \ldots\}$, úplný systém jevů na $(\Omega, \mathcal{A}, P)$. Nechť $P(B_i) > 0$ pro každé $i \in \mathbb{N}$.

Potom pro každý jev $A \in \mathcal{A}$ platí

$$P(A) = \sum_{i=1}^{\infty} P(A|B_i) \cdot P(B_i)$$

Přednáška 6: Bayesův vzorec a náhodné veličiny

4. 3. 2021

Tvrzení 6.1. Bayesův vzorec

Mějme $B_1, B_2, \ldots, B_n$, jevy na $(\Omega, \mathcal{A}, P)$ tvořící úplný systém jevů s nenulovou pstí $P(B_i) > 0, i \in \mathbb{N}$.

Potom každý jev $A \in \mathcal{A} > 0$, platí:

$$\forall i \in \mathbb{N}: \; P(B_i|A) = \frac{P(A|B_i) \cdot P(B_i)}{\sum_{j=1}^{\infty} P(A|B_j) \cdot P(B_j)}$$

Definice 6.2. Náhodná veličina

Mějme pstní prostor $(\Omega, \mathcal{A}, P)$. Reálnou funkci $X : \Omega \to \mathbb{R}$ nazveme náhodnou veličinou, pokud pro každý interval $I \in R$ platí: $X^{-1}(I) \in \mathcal{A}$ (podmínka měřitelnosti)

Přednáška 7: Náhodné veličiny

10. 3. 2021

Definice 7.1. Diskrétní náhodná veličina

Náhodnou veličinu X nazveme diskrétní, nabývá-li nejvýše spočetně mnoha hodnot.

Definice 7.2. Pravděpodobnostní funkce

Funkce $p_X : \mathbb{R} \to \mathbb{R}$, $p_X(a) = P(X = a)$ se nazývá pravděpodobnostní funkce diskrétní veličiny X, případně pravděpodobnostní rozdělení veličiny X. Lze definovat pouze na oboru hodnot X.

Povšimněme si.

$$\sum_{a \in \mathbb{R}} p_X(a) = \sum_{a \in H_X} p_X(a) = 1,$$

kde H_X označuje obor hodnot X.

Příklad. Indikátor jevu. Hození kostkou.

$$X = \begin{cases} 0 & \text{nepadlo 6} \\ 1 & \text{padlo 6} \end{cases}.$$

$$X(\omega) = \begin{cases} 0 & \omega \notin A \\ 1 & \omega \in A \end{cases}.$$

Určete p_X. $X \in \{0, 1\}$

$$p_X(0) = P(X = 0) = \frac{5}{6} \tag{1}$$

$$p_X(1) = P(X = 1) = \frac{1}{6}. \tag{2}$$

Příklad. Házíme mincí než padne panna.

X ... počet hodů, které jsme provedli

$$X \in \{1, 2, 3, 4, \ldots\}.$$

$$p_X(i) = P(X = i) = 2^{-i}, i \in \mathbb{N}.$$

Povšimněme si. $\forall I \ \subseteq \mathbb{R}$, I interval

$$\sum_{a \in I} p_X(a) = \sum_{a \in I} P(X = a) = P(X \in I).$$

Definice 7.3. Spojitá náhodná veličina

Náhodná veličina X se nazývá spojitě rozdělená (spojitá), pokud existuje integrovatelná nezáporná funkce $f : \mathbb{R} \to \mathbb{R}$, pro kterou platí, že pro každý interval $I \subseteq \mathbb{R}, P(X \in I) = \int_I f\,dx$.

Tj. pokud $I = [a, b], P(a \leq X \leq b) = \int_a^b f(x)\,dx$.

Funkce f se nazývá hustota náhodné veličiny X.

Definice 7.4. Distribuční funkce

Nechť X je náhodná veličina.

Funkce $F_X : \mathbb{R} \to [0, 1]$ dána předpisem $F_X(u) = P(X \leq u), u \in \mathbb{R}$ se nazývá distribuční funkce náhodné veličiny X.

F_X vždy existuje.

Povšimněme si.

$$F_X(u) = \int_{-\infty}^{u} f_X(t)\,dt, F'_X(u) = f_X(u).$$

pro X spojité.

Přednáška 8: Náhodné veličiny

11. 3. 2021

Věta 8.1. Vlastnosti F_X

Nechť $F_X : \mathbb{R} \to [0, 1]$ je dist. fce. n. v. X.

Potom platí:

1. F_X je neklesající
2. $\lim_{n\to\infty} F_X(u) = 1, \lim_{n\to-\infty}(u) = 0$
3. F_X je zprava spojitá

Věta 8.2 (Distribuční funkce a náhodná veličina)**.** Kdykoliv $F : \mathbb{R} \to [0, 1]$ je funkce splňující 1. - 3., existuje náhodná veličina X taková, že F je distribuční funkce X.

Příklad. Pro získání náhodné veličiny z distribuční funkce stačí vzít $\Omega = \mathbb{R}$.

$\mathcal{A}$ = měřitelné podmnožiny

P bude definovaná skrze hodnoty na intervalech typu $(a, b]$, kde $P((a, b]) = F(b) - F(a)$.

$X : \mathbb{R} \to \mathbb{R}, X(t) = t$

Důsledek 8.2.1. Není třeba dále pracovat s pravděpodobnostními prostory Ω. Stačí pracovat s rozdělením pravděpdobnosti hodnot X na $\mathbb{R}$. To je jednoznačně určeno pomocí F_X.

> Když tímto způsobem definujeme pravděpodobnost na $\mathbb{R}$, dostáváme pstní prostor.
>
> $$\forall I \subseteq \mathbb{R} : P(I) := P(X^{-1}(I)).$$

Věta 8.3. Nechť X je diskrétně rozdělená náhodná veličina a Y je spojitě rozdělená náhodná veličina. Potom platí pro libovolné $t \in \mathbb{R}$: $F_X(t) = \sum_{a \leq t} p_X(a), F_Y(t) = \int_{-\infty}^{t} f(a)\, da$

Příklad. Náhodná veličina X je dána jako počet panen při hodu dvěma mincemi. Nalezněte F_X.

$$F_X(u) = P(X \leq u) = \begin{cases} 0 & u < 0 \\ \frac{1}{4} & u \in [0, 1) \\ \frac{3}{4} & u \in [1, 2) \\ 1 & u \geq 2 \end{cases}.$$

Příklad. X je dána jako vzdálenost náhodně vybraného bodu z kruhu o poloměru $a > 0$ od kraje kruhu. Určete F_X a f_X.

$$F_X(u) = P(X \leq u) = \begin{cases} 0 & u < 0 \\ 1 - \frac{\pi(a-u)^2}{\pi a^2} = 1 - \left(\frac{a-u}{a}\right)^2 & u \in [0, a) \\ 1 & u \geq a \end{cases}.$$

$$f_X = F'_X(u) = \begin{cases} \frac{-2u}{a^2} + \frac{2}{a} & u \in [0, a) \\ 0 & \text{jinak} \end{cases}.$$

Definice 8.4 (Střední hodnota)**.** Nechť X je n. v. Pokud je X diskrétní, potom definujeme její střední hodnotu jako

$$EX = \sum_{a \in \mathbb{R}} a \cdot p_X(a).$$

Pokud je X spojitá, potom definujeme její střední hodnotu jako

$$EX = \int_{-\infty}^{\infty} x \cdot f(x)\, dx,$$

kde f je hustota

Přednáška 9: Náhodné veličiny

17. 3. 2021

Věta 9.1. Nechť X je náhodná veličina taková, že její distribuční funkce F_X je spojitá a má konečnou derivaci všude až na spočetně mnoho bodů. Potom je náhodná veličina X spojitě rozdělená.

Důkaz. nástin důkazu: definuji $M = \{u | u \in \mathbb{R}, F_X$ nemá konečnou derivaci v u$\}$. Zřejmě M je nejvýše spočetná, a proto $\forall I = (a, b] \subseteq \mathbb{R}$, platí $\int_{I \setminus M} 1 = \lambda(I) = b - a$.

chci $\exists f : \mathbb{R} \to \mathbb{R}, f \geq 0,$ integrovatelná, že $\forall I \; P(X \in I) = \int_I f(x) dx.$. Vezmu

$$f(u) = \begin{cases} F_X'(u) & u \notin M \\ 0 & u \in M \end{cases}.$$

Vezmu $I = (a, b]$. Potom dle definice

$$\begin{aligned}(P(X \in I) &= F_X(b) - F_X(a) = F_X(b^-) - F_X(a^+) \\ &= [F_X(u)]_{u=a}^{u=b} = \int_{I \setminus M} F_X'(u)\, du = \int_I f(u)\, du^{\cdot}\end{aligned}$$

Integrovatelnost f plyne z

$$\int_{\mathbb{R}} |f(x)|\, dx = \int_{\mathbb{R}} f(x)\, dx = \int_{-\infty}^{\infty} f(x)\, dx = F_X(\infty^-) - F_X(-\infty^+) = 1 - 0 = 1 < \infty.$$

Kde první úprava je možná, neboť $f(x) = F'(x)$ a F je neklesající. □

Příklad (Devil's staircase)**.** Odstrašující příklad. Fce, která je spojitá, ale nemá v nespočetně mnoha bodech derivaci a navíc má množina těchto bodů Lebesgueovu míru 0. (tedy ta fce má skoro všude derivaci)

Derivace je nulová skoro všude, tedy integrál bude nulový. Viz Wikipedia (EN).

K definici střední hodnoty. Nechť X je náhodná veličina, $X : \Omega \to \mathbb{R}$. Obecně se definuje střední hodnota X jako

$$EX = \int_\Omega X(\omega)\, dP(\omega) = \int_\mathbb{R} u\, dF(u).$$

Kde dF je pst. na $\mathbb{R}$, kterou získáme přenesením P z Ω na $\mathbb{R}$ pomocí X.

$$B \subseteq \mathbb{R} : dF(B) = P(X \in B) =$$
$$P(\{\omega | \omega \in \Omega, X(\omega) \in B\}) = P(X^{-1}(B))$$

Vlastnosti střední hodnoty

Nechť X je n. v. Y je n. v. $X : \Omega \to \mathbb{R}$

1. $\forall c \in \mathbb{R} \ : Ec = c$
2. $\forall c \in \mathbb{R} \ : E(cX) = cEX$
3. E(X+Y) $= EX + EY$
4. $\forall g : \mathbb{R} \to \mathbb{R}$, g měřitelná

   $$Eg(X) = \int_\Omega g(X)\, dP = \int_\mathbb{R} g(u)\, dF(u).$$

   Speciální případ X je spojitá

   $$Eg(X) = \int_{-\infty}^{\infty} g(u) f(u)\, du.$$

   Speciální případ X je diskrétní

   $$Eg(X) = \sum_{a \in \mathbb{R}} g(a) p_X(a).$$

5. X, Y pro které $X \leq Y$ skoro všude, potom $EX \leq EY$.
6. Jensenova nerovnost. Je-li $g : \mathbb{R} \to \mathbb{R}$ konvexní, potom $g(EX) \leq Eg(X)$

Důkaz. 1. $c \in \mathbb{R}$ chápeme jako konstantní veličinu. c je deterministická. $X : \Omega \to \mathbb{R}, X(\omega) = c$, $X \in \{c\} \implies$ X je diskrétní

$$EX = c \cdot 1 = c.$$

2.
$$E(cX) = \int_{\mathbb{R}} c \cdot x \, dF(x) = c \cdot \int_{R} x \, dF(x) = cEX.$$

Speciálně spojitá

$$E(cX) = \int_{\mathbb{R}} cx f_X(x) \, dx = c \int_{\mathbb{R}} x f_X(x) \, dx = cEX.$$

diskrétní

$$E(cX) = \sum_{a \in \mathbb{R}} a p_{cX}(a) = \sum_{a \in \mathbb{R}} a p(cX = a) = \sum_{a \in \mathbb{R}} cap(X = a).$$

3.
$$E(X+Y) = \int_{\Omega} (X+Y) \, dP = \int_{\Omega} X \, dP + \int_{\Omega} Y \, dP = EX + EY.$$

4.
$$\begin{aligned} Eg(X) = E(g \circ X) = \int_{\Omega} g \circ X(\omega) \, dP(\omega) = \int_{-\infty}^{\infty} a \, dF_{g \circ X}(a) = \\ \int_{-\infty}^{\infty} g(u) \, dF_X(u) \end{aligned}.$$

Spojitý případ

$$\int_{-\infty}^{\infty} g(u) f_X(u) \, du.$$

Diskrétní

$$\int_{-\infty}^{\infty} g(a) \cdot d \sum_i x_i \delta_{a_i} = \sum_i = g(a_i) \cdot x_i = \sum_i g(a_i) \cdot P(X = a_i) = \sum_{a \in \mathbb{R}} g(a) \cdot p_X(a).$$

5.
$$EX = \int_{\Omega} X \, dP \leq \int_{\Omega} Y \, dP = EY,$$

neboť $P(B) \geq 0, B \subseteq \mathcal{A}$

6. Zřejmé.

$\square$

Definice 9.2 (n-tý (centrální) moment)**.** Nechť X je náhodná veličina a $n \in \mathbb{N}$

- Hodnotu EX^n nazýváme n-tý moment n. v. X.
- $E|X|^n$ nazýváme n-tý absolutní moment n. v. X.
- $E|X - EX|^n$ nazýváme n-tý (absolutní) centrální moment n. v. X.

Speciálně:

- $E|X - EX|$ nazýváme absolutní odchylku od průměru
- $\operatorname{var} X = DX = E|X - EX|^2$ nazýváme rozptyl n. v. X
- $\sqrt{DX} = (E|X - EX|^2)^{\frac{1}{2}}$ nazýváme směrodatná odchylka. Často značíme σ_x

Přednáška 10: Rozdělení náhodných veličin

18. 3. 2021

Příklad. Spočtěte $E|X - EX|$ pro $X \sim U[0,1]$ (rovnoměrné rozdělení)

$$F_X(u) = P(X \leq u) = \begin{cases} 0 & u < 0 \\ u & u \in [0,1) \\ 1 & u \geq 1 \end{cases}.$$

platí:

$$F'_X(u) = f(u),$$

až na spočetně mnoho bodů.

$$f(u) = \begin{cases} 0 & u \notin (0,1) \\ 1 & u \in (0,1) \end{cases}.$$

$$EX = \int_{-\infty}^{\infty} u \cdot f(u)\,du = \int_{-\infty}^{0} u \cdot 0\,du + \int_0^1 u \cdot 1\,du + \int_1^{\infty} u \cdot 0\,du = \left[\frac{u^2}{2}\right]_0^1 = \frac{1}{2}.$$

$$E|X - EX| = E|X - \frac{1}{2}| = \int_{-\infty}^{\infty} |u - \frac{1}{2}| \cdot f(u)\,du = \int_0^1 |u - \frac{1}{2}|\,du = \frac{1}{4}.$$

Věta 10.1. Nechť X je n. v. s konečným 2. momentem. Potom:

1. $E(X - EX) = 0$
2. (rozptyl) $DX = E(X - EX)^2 = EX^2 - (EX)^2$
3. $\forall a, b \in \mathbb{R} \;\; D(aX + b) = a^2 DX$

4. $MAD(aX + b) = |a| \cdot MAD(X)$

Důkaz. Důkaz předchozí věty.

1. $E(X - EX) = EX - E(EX) = EX - EX = 0$
2. $DX = E(X-EX)^2 = E(X^2-2X\cdot EX+(EX)^2) = EX^2-E(2X\cdot EX)+E(EX)^2 = EX^2 - 2EX \cdot EX + (EX)^2 = EX^2 - 2(EX)^2 + (EX)^2 = EX^2 - (EX)^2$
3. nechť $a, b \in \mathbb{R}$.

   $D(aX+b) = E(aX+b-E(aX+b))^2 = E(aX+b-aEX-b)^2 = E(aX-aEX)^2 = E(a(X - EX))^2 = a^2E(X - EX)^2 = a^2DX$
4. $MAD(aX + B) = E|aX + b - E(aX + b)| = E|a(X - EX)| = |a|MAD(X)$

□

Některé typy pstních rozdělení

Když se veličina X řídí rozdělením B s parametry $(\alpha, \beta, \gamma, \ldots)$, značíme to $X \sim B(\alpha, \beta, \gamma, \ldots)$.

10.1 Alternativní rozdělení (nula-jedničkové, indikátor jevu)

parametr p:

$$X \sim Alt(p).$$

$$P(X = 0) = 1 - p.$$

$$P(X = 1) = p.$$

Charakteristiky

- $X \sim Alt(p)$
- $EX = 0(1 - p) + 1p = p$
- $EX^n = 0^n(1 - p) + 1^np = p$
- $DX = EX^2 - (EX)^2 = p - p^2 = p(1 - p)$

Příklad. Počet panen v jednom hodu mincí, kde panna má pst p.

10.2 Binomické rozdělení

parametry $n \in \mathbb{N}, p \in [0,1]$

značíme $X \sim \text{Bi}(n,p)$, pokud $X \in \{0,1,2,\ldots,n\}$, $\forall k \in \{0,1,\ldots,n\} : P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$

Charakteristiky

- $EX = np$
- $DX = np(1-p)$

Příklad. Počet panen v hodu n-mincemi, kde každá má pst panny p. (odpovídá součtu u nezávislých veličin s Alt rozdělením s parametrem p)

10.3 Geometrické rozdělení

parametr $p \in (0,1]$

značíme $X \sim Geom(p)$

$P(X = k) = p(1-p)^{k-1}$

Charakteristiky

- $EX = \sum_{k=1}^{\infty} k \cdot p \cdot (1-p)^{k-1} = p \cdot \sum_{k=0}^{\infty} k(1-p)^{k-1} = p \cdot \frac{1}{p^2} = \frac{1}{p}$

  $\varphi'(k) = \sum_{k=0}^{\infty} k(1-p)^{k-1} = \frac{1}{p^2}$

  $\varphi(k) = \sum_{k=0}^{\infty} -(1-p)^k = \frac{-1}{1-(1-p)} = -\frac{1}{p}$
- $DX = EX^2 - (EX)^2 = EX^2 - \frac{1}{p^2} = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$

  $EX^2 = \sum_{k=1}^{\infty} k^2 p(1-p)^{k-1} = \sum_{k=1}^{\infty} k(k+1)p(1-p)^{k-1} - \sum_{k=1}^{\infty} kp(1-p)^{k-1} = \frac{2-p}{p^2}$

Příklad. Odpovídá počtu nezávislých pokusů do úspěchu, v každém pokusu je pst úspěchu p.

Házíme minci dokud nehodíme pannu.

10.4 Rovnoměrné rozdělení na intervalu

parametry $a < b, a, b \in \mathbb{R}$

značíme $X \sim U((a,b))$

X nabývá hodnot z (a,b)

hustota:

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

Charakteristiky

- $EX = \int_{-\infty}^{\infty} u \cdot f(u)\,du = \int_a^b \frac{u}{b-a} = \left[\frac{u^2}{2}\frac{1}{b-a}\right]_a^b = \frac{a+b}{2}$
- $DX = EX^2 - (EX)^2 = EX^2 - \left(\frac{a+b}{2}\right)^2 = \frac{b^2-2ab+a^2}{12} = \frac{(b-a)^2}{12}$

  $EX^2 = \int_{-\infty}^{\infty} u^2 \cdot f(u)\,du = \frac{b^3a^3}{3(b-a)} = \frac{b^2-ab+a^2}{3}$

Přednáška 11: Rozdělení náhodných veličin

24. 3. 2021

11.1 Poissonovo rozdělení (rozdělení řídkých jevů)

parametr $\lambda, \lambda > 0$

$X \in \{0, 1, 2, 3, 4, \dots\}$

$p_X(n) = P(X = n) = e^{-\lambda} \cdot \frac{\lambda^n}{n!}, n \in \mathbb{N}_0$

Charakteristiky

- $EX = \sum_{n=0}^{\infty} n \cdot e^{-\lambda} \cdot \frac{\lambda^n}{n!} = e^{-\lambda}\lambda \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$
- $DX = EX^2 + (EX)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$ $EX^2 = e^{-\lambda} \sum_{n=0}^{\infty} n^2 \frac{\lambda^n}{n!} = \ldots = \lambda^2 + \lambda$

Příklad. Počet dopisů, které dostaneme za rok.

Poissonovo rozdělení lze získat jako limitní případ binomického rozdělení, kde $n \to \infty$, $p \to 0$, $np \to \lambda$

n lidí, z nichž každý mi pošle dopis, s pstí p, posílají nezávisle.

Příklad (špatný příklad (školácký))**.** Počet příchozích hovorů telefonní ústředny za den. (nesplňuje předpoklad nezávislosti: př. ústředna O2, jev . . . vypadne internet v Praze)

11.2 Exponenciální rozdělení

modelujeme pomocí něj dobu mezi dvěma událostmi, jejichž počet se řídí Paissonovým rozdělením.

parametr $\lambda, \lambda > 0$

$x \in \mathbb{R}$

hustota $X \sim Exp(\lambda)$

$$f_X(u) = \begin{cases} \lambda e^{-\lambda u} & u > 0 \\ 0 & u \leq 0 \end{cases}$$

Charakteristiky

- $EX = \int_{-\infty}^{\infty} x f_X(x) = \int_0^{\infty} x\lambda e^{-\lambda x}\, dx = \left[x \frac{e^{-\lambda x}}{-\lambda}\lambda\right]_0^{\infty} - \int_0^{\infty} \lambda \frac{e^{-\lambda x}}{-\lambda}\, dx = 0 + \int_0^{\infty} e^{-\lambda x} = \left[\frac{-e^{-\lambda x}}{\lambda}\right]_0^{\infty} = \frac{1}{\lambda}$
- $EX^2 = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx = \int_0^{\infty} x^2 \lambda e^{-\lambda x}\, dx = \left[x^2 \frac{e^{-\lambda x}}{-1}\right]_0^{i} nfty + 2\int_0^{\infty} x \frac{e^{-\lambda x}}{1}\, dx = 0 + \frac{2}{\lambda^2} = \frac{2}{\lambda^2}$

  $DX = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$

Vlastnost (absence paměti)

$$\forall a, b > 0, X \sim Exp(\lambda):\ P(X > a + b | X > a) = P(X > b).$$

Příklad. Podobné jako Poissonovo rozdělení, doba čekání do dalšího dopisu.

Příklad. Počet hodin, které žárovka vydrží svítit.

11.3 Gaussovo (normální) rozdělení

parametry $\mu, \sigma^2, \mu \in \mathbb{R}, \sigma^2 > 0$ (případně $\sigma = 0$)

hustota $X \sim N(\mu, \sigma^2)$:

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

Charakteristiky

- $EX = \ldots = \mu$ ... viz přednáška 06a
- $DX = EX^2 - (EX)^2 = \sigma^2$

Příklad. Výška jedince z jedné skupiny/populace,

hmotnost ...

(Téměř libovolný fyzický atribut v rámci jednoho druhu.)

Pevný bod Fourier transformace.

$$\psi(x) = e^{\frac{-x^2}{2}} \frac{1}{\sqrt{2\pi}}.$$

Její primitivní funkce není elementární.

$$\int_{-\infty}^{\infty} e^{\frac{-x^2}{2}}\, dx = \sqrt{2\pi}.$$

$X \sim N(\mu, \sigma^2) \ldots$ X má normální rozdělení se střední hodnotou μ a rozptylem σ^2 .

$X \sim N(0,1)$, normované normální rozdělení

Příklad. $X \sim N(\mu, \sigma^2)$

$$P(\mu - \sigma \leq X \leq \mu + \sigma) = 68\%.$$

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 95\%.$$

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 99.7\%.$$

Věta 11.1. Nechť $X \sim N(\mu, \sigma^2), \sigma^2, \mu \in \mathbb{R}$

Potom pro každé dvě čísla $a, b \in \mathbb{R}$ platí $aX + b \sim N(a\mu + b, a^2\sigma^2)$

Důkaz. Nechť $X \sim N(\mu, \sigma^2)$

Potom $P(X \leq u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\mu-x)^2}{2\sigma^2}}\, dx$

Pro a kladné

$$\begin{aligned} P(aX + b \leq u) = P(X \leq \frac{u-b}{a}) &= \int_{-\infty}^{\frac{u-b}{a}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(\mu-x)^2}{2\sigma^2}}\, dx \\ &= \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{\left(\mu - \frac{y-b}{a}\right)^2}{2\sigma^2}) \frac{dy}{a} \\ &= \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi a^2\sigma^2} \exp(\frac{(\mu a + b - y)^2}{2a^2\sigma^2})}\, dy \end{aligned}$$

$$\sigma_1^2 = a^2\sigma^2, \mu_1 = a\mu + b$$

Pro a záporné

$$\int_{\infty}^{u} \frac{1}{\sqrt{2\pi\sigma^2}\exp(-\frac{(a\mu+b-y)^2}{2a^2\sigma^2})}\,dy$$
$$= 1 - \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(a\nu+b-y)^2}{2a^2\sigma^2})\,dy$$

Pro $a = 0$

$$X \sim N(\mu, \sigma^2), aX + b \sim N(b, 0),$$

$aX + b$ je deterministické, konstanta b.

Zkonverguje na Diraca v bodě b. Veličina je konstantní. $a = 0, aX + b = b$ $\square$

Důsledek 11.1.1. Nechť $X \sim N(\mu, \sigma^2), \sigma^2 > 0$.

Potom normalizovaná veličina $U = \frac{x-\mu}{\sigma}$ má rozdělení $U \sim N(0, 1)$

$$E\left(\frac{X-\nu}{\sigma}\right) = \frac{EX-\nu}{\sigma} = \frac{\mu-\mu}{\sigma} = 0$$

$$D\left(\frac{X-\nu}{\sigma}\right) = \ldots = 1$$

Přednáška 12: Rozdělení náhodných veličin

25. 3. 2021

12.1 Paretovo rozdělení (někdy také 80:20 rozdělení)

parametr (pro nás stačí jeden) $\alpha, \alpha > 0$

distribuční funkce

$$F_X = \begin{cases} 1 - \frac{1}{u^\alpha} & u \geq 1 \\ 0 & u < 1 \end{cases}$$

hustota

$$f_X(u) = \begin{cases} \frac{\alpha}{u^{\alpha+1}} & u \geq 1 \\ 0 & u < 1 \end{cases}$$

Charakteristiky

- Střední hodnota

$$EX = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_{1}^{\infty} \alpha x^{-\alpha}\,dx = \begin{cases} \int_1^\infty \frac{1}{x}\,dx = [\log x]_1^\infty = \infty & \alpha = 1 \\ \int_1^\infty \left[\frac{x^{1-\alpha}}{-\alpha+1}\right]_1^\infty = \infty & \alpha < 1 \\ \left[\frac{x^{1-\alpha}\alpha}{-\alpha+1}\right]_1^\infty = \frac{\alpha}{\alpha-1} & \alpha > 1 \end{cases}$$

$\Rightarrow \alpha \leq 1 \ldots EX = \infty$

$\Rightarrow \alpha = 2 \ldots EX = 2$

$\Rightarrow \alpha = \frac{6}{5} \ldots EX = \frac{6}{5}\frac{5}{1} = 6$

$\Rightarrow \alpha = \frac{11}{10} \ldots EX \frac{\frac{11}{10}}{\frac{1}{10}} = 11$

- Pokud je střední hodnota nekonečná, rozptyl neexistuje. Rozptyl definujeme jen pro $\alpha \geq 1$

$$DX = E(X - EX)^2 = EX^2 - (EX)^2 = \frac{\alpha}{\alpha - 2} - \left(\frac{\alpha}{\alpha - 1}\right)^2,$$

pokud $\alpha > 2$

$$EX^2 = \int_{-\infty}^{\infty} u^2 f_X(u)\, du = \int_1^{\infty} u^2 \frac{\alpha}{u^{\alpha+1}}\, du = \begin{cases} \int_1^{\infty} \frac{\alpha}{u}\, du = [\alpha \log u]_1^{\infty} = \infty & \alpha = 2 \\ \left[\frac{\alpha}{-\alpha+2} u^{-\alpha+2}\right]_1^{\infty} = \infty & \alpha \leq 2 \\ \left[\frac{\alpha}{-\alpha+2} u^{-\alpha+2}\right]_1^{\infty} = \frac{\alpha}{\alpha-2} & \alpha > 2 \end{cases}$$

Příklad. motivace: Italský účetní Pareto: "20 % farmářů vlastní 80 % země (pozemků)"

Přednáška 13: Náhodné vektory

31. 3. 2021

Definice 13.1. Nechť $(\Omega, \mathcal{A}, P)$ je pstní prostor, $n \in \mathbb{N}$.

Zobrazení $X : \Omega \to \mathbb{R}^n$ nazvu (n-rozměrný) náhodný vektor, pokud pro každý kvádr $K = I_1 \times I_2 \times \ldots \times I_n \subseteq \mathbb{R}_n$, I_i jsou libovolné reálné intervaly, platí $X^{-1}(K) \in \mathcal{A}$.

Příklad. náhodně vyberu jedince z populace

$$X = (\text{výška}, \text{váha}, \text{indikátor pohlaví})$$

Příklad.

$$X = (X_1, \ldots, X_n)$$

X_i cena námi stanovené i-té akcie při otevření trhu.

Definice 13.2. Nechť X n-rozměrný n. vekt. Funkce $F : \mathbb{R}^n \to [0, 1]$ definovaná vztahem

$$F(x) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n), x \in \mathbb{R}^n, X = (X_1, \ldots, X_n)$$

se nazývá (sdružená) distribuční funkce n. vektoru X.

Tvrzení 13.3. Pokud $(\Omega, \mathcal{A}, P)$ je pstní prostor, $u \in \mathbb{R}$ a $X : \Omega \to R^n$ je n. vektor, potom na $\mathbb{R}^n$ lze definovat pst P^* vzorcem

$$P^*(B) = P(X^{-1}(B)), B \subseteq \mathbb{R}^n,$$

B měřitelná.

Tuto pst P^* budeme někdy značit dF, kde F je dist. fce vektoru X.

Vlastnosti 13.4. Vlastnosti F_X.

Nechť X je n-rozměrný náhodný vektor s dist. fcí F. Potom platí:

1. F je v každé proměnné neklesající.
2. Pokud $x \in \mathbb{R}^n, y_i \in \mathbb{R}^n, x < y_i, i \in \{1, \ldots, n\}$

$$\lim_{k \to \infty} F(y^k) = F(x)$$

3. $\forall k \in \{1, \ldots, n\}, x \in \mathbb{R}^n$:

$$\lim_{u \to -\infty} F((x_1, x_2, \ldots, x_{k-1}, u, x_{k+1}, \ldots, x_n)) = 0$$

$$\lim_{x_1 \to \infty} \lim_{x_2 \to \infty} \ldots \lim_{x_n \to \infty} F_x(x) = 1$$

Definice 13.5. Buď X n-rozměrný n. vektor. X nazvu diskrétně rozdělený, existuje-li nejvýše spočetná množina $A = \{a_1, a_2, \ldots\} \subseteq \mathbb{R}^n$ a nezáporná čísla $p_1, p_2, p_3, \ldots$ pro něž platí

$$\sum_{i=1}^{\infty} p_i = 1, \forall i \in \mathbb{N} : \; P(X = a_i) = p_i.$$

Pozn. Je-li X diskrétně rozdělený vektor, potom $X_1, \ldots, X_n$ jsou diskrétně rozdělené a naopak

Definice 13.6. X nazýváme spojitě rozdělený, pokud existuje nezáporná integrovatelná (a tedy měřitelná) fce $f : \mathbb{R}^n \to \mathbb{R}$ splňující

$$P(X \in K) = \int_K f \, dx,$$

pro každý n-rozměrný kvádr K.

Definice 13.7. Buď $X = (X_1, \dots, X_n)$ náhodný vektor. Rozdělení veličin X_i často označujeme jako marginální rozdělení. Distribuční funkce n. vektoru X se často nazývá sdružená dist. f. (analogicky f je sdružená hustota)

Věta 13.8. Pro n vektor $X = (X_1, \dots, X_n)$ a jeho dist. fci F lze marginální distribuční funkce veličiny X_k určit jako

$$\begin{aligned}
&\lim_{x_1\to\infty}\lim_{x_2\to\infty}\dots\lim_{x_n\to\infty} F((x_1, x_2, x_{k-1}, \dots, u, x_{k+1}, \dots, x_n)) \\
&= P(X \in (-\infty,\infty) \times (-\infty,\infty) \times \dots \times (-\infty, u], \times(-\infty,\infty)\dots(-\infty,\infty))\cdot \\
&= P(X_k \in (-\infty, u]) = F_{X_k}(u)
\end{aligned}$$

Tvrzení 13.9. Nechť X je spojitý, n-rozměrný n. vektor se sdruženou hustotou f a sdruženou distribuční funkcí F.

Potom F je spojitá funkce mající skoro všude derivaci a platí 1)

$$\frac{\partial^n F}{\partial x_1 \partial x_2 \dots \partial x_n}(u) = f(u),$$

(pro λ_n-skoro všechna u)

Platí 2)

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\dots\int_{-\infty}^{x_n} f(u_1, \dots, u_n)\, du_n \dots du_1$$

Důkaz. Tvrzení

1. spojitost integrálu vůči mezím a 1) $\implies$ F je spojitá.
2. Vezmeme kvádr $k = (-\infty, x_1] \times \dots \times (-\infty, x_n]$

$$F(x_1, \dots, x_n) = P(X \in k) = \int_K f\, du \dots$$

□

Přednáška 14: Příklady na náhodné vektory, jejich charakteristiky 1. 4. 2021

Příklady na náhodné vektory viz přednáška na Youtube.

Příklad. N. vektor (X, Y) má rozdělení hodnot dané tabulkou

X \ Y	0	1
0	0.2	0.3
1	0.2	0.3

$$(X,Y) \in \{(0,0),(0,1),(1,0),(1,1)\}$$

Pro představu $p_{X,Y}(0,0) = 0.2, p_{X,Y}(0,1) = 0.3$.

Určete marginální rozdělení veličin X a Y.

$$X \in \{0,1\}$$

$$P(X=0) = P(X=0 \wedge Y=0) + P(X=0 \wedge Y=1) = 0.2 + 0.3 = 0.5$$

$$P(X=1) = 1 - P(X=0) = 0.5$$

$$P(Y=0) = P(Y=0 \wedge X=0) + P(Y=0 \wedge X=1) = 0.2 + 0.2 = 0.4$$

$$P(Y=1) = 1 - P(Y=0) = 0.6$$

Veličiny mají alternativní rozdělení.

14.1 Charakteristiky náhodných vektorů

Definice 14.1. Buď $X = (X_1, \dots, X_n), n \in \mathbb{N}$ náhodný vektor.

$$EX = (EX_1, \dots, EX_n)$$

nazýváme střední hodnotu.

Pozn. geometricky odpovídá poloze těžiště.

Definice 14.2.

$$\operatorname{var} X = \{cov(X_i, X_j)\}_{i,j=1}^{n},$$

kde $cov(Y,Z) = E(Y - EY)(Z - EZ)$ je kovariance veličin Y a Z.

Definice 14.3.

$$\rho_X = \{\rho_{x_i,x_j}\}_{i,j=1}^{n},$$

kde

$$\rho_{Y,Z} = \frac{cov(Y,Z)}{\sqrt{DY \cdot DZ}}$$

je korelační koeficient veličin Y a Z.

Tvrzení 14.4. Nechť X, Y jsou n. v. s konečným druhým momentem.

Potom platí:

1. $cov(X,Y) = cov(Y,X)$
2. $cov(X,X) = DX$
3. $cov(X,Y) = EXY - (EX)(EY)$
4. $\rho_{X,Y} \in [-1,1]$
5. $a,b \in \mathbb{R}:\ cov(aX+b,Y) = E(aX+b)Y - (aEX+b)EY = aEXY + bEY - aEXEY - bEY = a \cdot cov(X,Y)$

Důkaz. 1. $cov(X,Y) = E(X-EX)(Y-EY) = E(Y-EY)(X-EX) = cov(Y,X)$ $\Longrightarrow$ variační matice je vždy symetrická

2. $cov(X,X) = E(X-EX)(X-EX) = E(X-EX)^2 = DX$
3. $cov(X,Y) = E(X-EX)(Y-EY) = E(XY - YEX - XEY + EXEY) = E(XY) - EY \cdot EX - EX \cdot EY + EXEY = EXY - EXEY$
4. Cauchy-Schwarz
$$\left| \frac{< u, v >}{||u|| \cdot ||v||} \right| \leq 1$$

   ... zbytek na přednášce

□

Definice 14.5. Nechť $X_1, \dots, X_n$ jsou náhodné veličiny, $n \in \mathbb{N}$. Říkáme, že $X_1, \dots, X_n$ jsou nezávislé, pokud pro každé $(u_1, \dots, u_n) = u \in \mathbb{R}^n$ jsou jevy $\{X_1 \leq u_1\}, \dots, \{X_n \leq u_n\}$ nezávislé.

Tvrzení 14.6. Nechť X je n. vektor se sdruženou dist. fcí F $F: \mathbb{R}^n \to [0,1]$. Veličiny $X_1, \dots, X_n$ jsou nezávislé $\iff$ $\forall u \in \mathbb{R}^n$

$$F(u) = F_{X_1}(u_1) \cdot F_{X_2}(u_2) \dots \cdot F_{X_n}(u_n),$$

kde F_{X_i} je dist. fce. veličiny X_i.

Pro X spojitý vektor: $X_1, \dots, X_n$ nezávislé $\iff$ $f_X = f_{x_1} \cdot f_{x_2} \cdot \dots \cdot f_{x_n}$

Pro X diskrétní: $X_1, \dots, X_n$ nezávislé $\iff$ $p_X = p_{X_1} \cdot \dots \cdot p_{X_n}$

Přednáška 15: Náhodné vektory

7. 4. 2021

Věta 15.1. Nechť $X_1, \dots, X_n$ jsou náhodné veličiny.

1. pokud $X_1, \dots, X_n$ jsou diskrétní, potom jsou nezávislé právě tehdy, je-li vektor X diskrétní a platí
$$\forall a \in \mathbb{R}^n: \; p_{X_1, X_2, \dots, X_n}(a) = p_{X_1}(a_1) \cdot p_{X_2}(a_2) \cdot \dots \cdot p_{X_n}(a_n)$$
2. Jsou-li $X_1, \dots, X_n$ spojité náhodné veličiny, potom $X_1, \dots, X_n$ jsou nezávislé $\iff$ vektor X je spojitě rozdělený a platí $f_X = f_{x_1} \cdot f_{x_2} \cdot \dots f_{x_n}$ až na množiny dimenze $< n$.

Pozn. pokud $cov(X, Y) = 0$, potom říkáme, že jsou veličiny X, Y nekorelované. (korelace je nulová)

Příklad. Korelované veličiny

1. Výška a váha jedince (pokud je někdo vyšší, pravděpodobně bude vážit víc) nezaměňovat korelaci a kauzalitu vyšší $\iff$ těžší. Ale pokud budu jíst a tloustnout, nevyrostu!
2. velikost reprodukčního indexu r v čase t a počet úmrtí v čase $t + 20$

Věta 15.2. Nechť X, Y jsou nezávislé a $EX^2, EY^2 < \infty$, potom platí $EXY = EXEY$

Důsledek 15.2.1. Nechť X, Y jsou n. v., $EX^2, EY^2 < \infty$, nezávislé, potom $cov(X, Y) = 0 a \rho_{X,Y} = 0$

Příklad. Příklad na cov.

X/Y	-1	0	1	
-1	0.1	0.3	0.1	0.5
1	0.25	0	0.25	0.5
	0.35	0.3	0.35	

X, Y nejsou nezávislé.

$P(X = 1, Y = 0) = 0 \neq 0.5 \cdot 0.3 = P(X = 1) \cdot P(Y = 0)$ $EY = 0$ $EXY = 0$

$cov(X, Y) = EXY - EXEY = 0 - 0 = 0$

Příklad. $X \sim U[-1, 1], Y = X^2$

$EX = 0, \; EX^3 = \int_{-1}^{1} x^3 \frac{1}{2}\, dx = 0$

$cov(X, Y) = 0$

Zjevně X a Y nejsou nezávislé (Y je definována pomocí X)

Přednáška 16: Náhodné vektory

8. 4. 2021

Tvrzení 16.1. Nechť X, Y jsou n. v. a $EX^2, EY^2 < \infty$.

Potom platí $D(X+Y) = DX + DY + 2cov(X,Y)$.

Důsledek 16.1.1.

1. Kdykoliv X, Y jsou nezávislé (nekorelované) veličiny s $EX^2, EY^2 < \infty$, máme $D(X+Y) = DX + DY$
2. Induktivně pokud $X_1, \ldots, X_n$ jsou nezávislé n. v. s $EX_i^2 < \infty$, platí $D(\sum_{i=1}^n X_i) = \sum_{i=1}^n DX_i$

Příklad. Nechť $X \sim Bi(n,p)$, ukažte, že $DX = np(1-p)$

$$X \sim Bi(n,p) \implies X = \sum_{i=1}^n X_i,$$

kde

$$X_i \sim Alt(p), X_1, \ldots, X_n \text{nezávislé}.$$

$$DX_i = p(1-p) \implies DX = D(\sum_{i=1}^n X_i) = \sum_{i=1}^n DX_i = n \cdot p(1-p).$$

Příklad. Nechť $X_1, \ldots, X_n$ jsou n. v. se stejnou střední hodnotou a rozptylem. ($\sigma^2 < \infty$). Nechť $X_1, \ldots, X_n$ jsou nezávislé.

Potom je

$$D\overline{X} = D\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2} D\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\sum_{i=1}^n DX_i = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

Příklad. Hrajeme hru.

1. hodíme mincí, naše výhra

   panna … 100 $

   orel … 0 $
2. hodíme 100x mincí, za každou pannu dostaneme dolar, za orla nic.

Poplatek za hru je 45 $.

1. moje výhra X - 45

$$X = \begin{cases} 100 & P(X=100) = \frac{1}{2} \\ 0 & P(X=0) = \frac{1}{2} \end{cases}$$

$E(X-45) = EX - 45 = 50 - 45 = 5$

$$\begin{aligned} D(X-45) = D(X) &= EX^2 - (EX)^2 \\ &= (100^2\frac{1}{2} + 0^2\frac{1}{2} - (50)^2 = \\ 5000 - 2500 &= 2500, \sqrt{DX} = 50 \end{aligned}$$

2. moje výhra X - 45

   $X_1, \ldots, X_{100}, X_1 \sim Alt(\frac{1}{2})$

$$E\sum_{i=1}^{100} X_i = 100 \cdot \frac{1}{2} = 50$$

$$Y = \sum_{i=1}^{100} X_i | E(Y-45) = 50 - 45 = 5$$

$$DY = D\left(\sum_{i=1}^{100} X_i\right) = \sum_{i=1}^{n} DX_i = 100\frac{1}{4} = 25, \sqrt{DY} = 5$$

Tvrzení 16.2. Nechť X, Y jsou nezávislé, spojitě rozdělené veličiny.

Potom $X + Y$ je spojitě rozdělená s hustotou

$$f_Z(u) = \int_{-\infty}^{\infty} f_X(t) \cdot f_y(u-t)\, dt = f_X * f_Y(u)$$

Důkaz.

$$\begin{aligned} F_Z(u) = P(X + Y \le u) &= \iint_{\{(x,y)\in\mathbb{R}^2 | x+y\le u\}} f_{X,Y}(x,y)\, dx\, dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{u-x} f_{X,Y}(x,y)\, dy\, dx \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{u-x} f_X(x) f_Y(y)\, dy\, dx \\ &= \int_{-\infty}^{\infty} f_X(x)\left(\int_{-\infty}^{u-x} f_Y(y)\, dy\right) dx \\ &= \int_{-\infty}^{\infty} f_X(x) \cdot F_Y(u-x)\, dx \end{aligned}$$

$$F'_Z(u) = \frac{\partial}{\partial u}\int_{-\infty}^{\infty} f_X(x) \cdot F_Y(u-x)\, dx = \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(u-x)\, dx$$

□

Příklad. Nechť $X_1, \dots, X_n$ jsou nezávislé n. v., $X_i \sim N(\nu_i, \sigma_i^2)$, potom

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

.

Protože (důkaz pro $X, Y \sim N(0,1)$)

$$f_X(u) = f_Y(u) = \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{-u^2}{2}}$$

$$\begin{aligned}
F_Z(u) &= \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(u-x)\, dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(u-x)^2}{2}}\, dx \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi} \exp\left(-\frac{1}{2}\left(x^2 + u^2 - 2xu + x^2\right)\right) dx \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{4}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{4} + xu - x^2}\, dx \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{4}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-(x-\frac{u}{2})^2}\, dx \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{4}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi 2}} e^{-\frac{y^2}{2}}\, dy \\
&= \frac{1}{\sqrt{4\pi}} e^{-\frac{u^2}{4}}
\end{aligned}$$

Přednáška 17: Limitní věty

14. 4. 2021

Definice 17.1. Nechť $X_1, X_2, \dots$ je posloupnost náhodných veličin.

Říkáme, že veličiny $X_1, X_2, X_3, \dots$ jsou nezávislé, pokud pro libovolnou n-tici $X_{i_1}, \dots, X_{i_n}, i_1 < i_2 < \dots < i_n$ platí, že veličiny $X_{i_1}, \dots, X_{i_n}$ jsou nezávislé.

Věta 17.2. Čebyševova nerovnost

1. Nechť X je náhodná veličina s $E|X|^n < \infty$ pro jisté $n \in \mathbb{N}$. Potom pro každé $\delta > 0$ platí

$$P(|X| > \delta) \leq \frac{E|X|^n}{\delta^n}$$

2. Pokud X je n. v. a $EX^2 < \infty$, potom pro každé $\delta > 0$ platí:

$$P(|X - EX| \geq \delta) \leq \frac{DX}{\delta^2}$$

Důkaz. Předchozí věty

1. Nechť $E|X|^n < \infty$

   (důkaz pouze pro spojitou X)

$$\begin{aligned} E|X|^n &= \int_{-\infty}^{\infty} |x|^n \cdot f_X(x)\,dx \\ &\geq \int_{\{x\in\mathbb{R}||x|\geq\delta\}} |x|^n \cdot f_X(x)\,dx \\ &\geq \int_{\{x\in\mathbb{R}||x|\geq\delta\}} \delta^n \cdot f_X(x)\,dx \\ &= \delta^n \cdot \int_{\{x\in\mathbb{R}||x|\geq\delta\}} f_X(x)\,dx \\ &= \delta^n P(|X| \geq \delta) \end{aligned}$$

   A odtud

$$P(|X| \geq \delta) \leq \frac{E|X|^n}{\delta^n}$$

2. Nechť $EX^2 < \infty$, $Y = X - EX$, $n = 2$. $EY^2 < \infty$ Potom pro každé $\delta > 0$ platí

$$\begin{aligned} P(|X - EX| \geq \delta) &= P(|Y| \geq \delta) \\ &\leq \frac{EY^2}{\delta^2} = \frac{E(X-EX)^2}{\delta^2} = \frac{DX}{\delta^2} \end{aligned}$$

□

Příklad. Odhadněte pravděpodobnost se kterou bude součet 100 hodů kostkou v rozpětí 340 - 360.

$X_1, \ldots, X_{100}$

$X_i \ldots$ hod i-tou kostkou

$EX_i = 3.5, DX_i = EX_i^2 - (EX_i)^2 = \frac{91}{6} - (3.5)^2 = \frac{35}{12}$

$Y = \sum_{i=1}^{100} X_i$

$EY = 100 \cdot 3.5 = 350, DY = 100 \cdot \frac{35}{12} = \frac{25\cdot35}{3} = \frac{875}{3}$

$$P(|Y - 350| < 10) = 1 - P(|Y - 350| \geq 10) \geq 1 - \frac{\frac{875}{3}}{100} \geq -1.92$$

$\Longrightarrow$ Čebyšev nic neříká. :D

Příklad. Odhadněte pravděpodobnost se kterou bude součet 100 hodů kostkou v rozpětí 320 - 380.

$X_1, \dots, X_{100}$

$X_i \dots$ hod i-tou kostkou

$EX_i = 3.5, DX_i = EX_i^2 - (EX_i)^2 = \frac{91}{6} - (3.5)^2 = \frac{35}{12}$

$Y = \sum_{i=1}^{100} X_i$

$EY = 100 \cdot 3.5 = 350, DY = 100 \cdot \frac{35}{12} = \frac{25 \cdot 35}{3} = \frac{875}{3}$

$$P(|Y - 350| < 30) = 1 - P(|Y - 350| \geq 30) \geq 1 - \frac{\frac{875}{3}}{900} \geq 0.67$$

...

Věta 17.3. (Slabý) zákon velkých čísel

Nechť $X_1, X_2, X_3, \dots$ je posloupnost nezávislých n. veličin se stejnou střední hodnotou $\mu < \infty$ a se stejným rozptylem $\sigma^2 < \infty$.

Potom pro každé $\delta > 0$ platí

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| \geq \delta) = 0 \iff \lim_{n\to\infty} P(|\overline{X}_n - \mu| < \delta) = 1,$$

kde $\overline{X}_i = \frac{1}{n} \sum_{i=1}^{n} X_i$ je průměr n veličin.

Důkaz. (Slabého) zákona velkých čísel

Důkaz v zápiscích chybí. □

Věta 17.4. (Silný) zákon velkých čísel

Nechť $X_1, X_2, \dots$ je posloupnost nezávislých n. veličin se stejnou střední hodnotou $\mu < \infty$. Potom platí

$$P(\lim_{n\to\infty} \overline{X}_n = \mu) = 1.$$

Přednáška 18: Centrální limitní věta

15. 4. 2021

Věta 18.1. Centrální limitní věta

Nechť $X_1, X_2, X_3, \dots$ je posloupnost nezávislých, stejně rozdělených (independent, identically distributed, i. i. d.) náhodných veličin s konečnou střední hodnotou $\infty < \mu < \infty$ a rozptylem $\sigma^2 < \infty$.

Označme

$$U_n = \frac{\sum_{i=1}^n X_i n \cdot \mu}{\sqrt{n \cdot \sigma^2}}.$$

Potom pro každé $x \in \mathbb{R}$ platí

$$\lim_{n\to\infty} F_{U_n}(x) \to \Phi(x),$$

kde Φ je dist. fce $N(0,1)$.

Pozn.

$$U_n = \frac{\sum_{i=1}^n X_i - n \cdot \mu}{\sqrt{n \cdot \sigma^2}} = \frac{\frac{1}{n}\sum_{i=1}^n X_i - \mu}{\sqrt{\frac{\sigma^2}{n}}}.$$

Povšimněme si. dle zákona velkých čísel

$$\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{n\to\infty} \mu \text{ skoro jistě}$$

$$D(U_n) = D\left(\frac{\sum X_i}{\sqrt{n\sigma^2}}\right) = \frac{1}{n\sigma^2} \cdot \sum_{i=1}^n DX_i = \frac{1}{n\sigma^2} n\sigma^2 = 1$$

Příklad. Mějme minci, kde panna padá s pravděpodobností 3/4. Jaká je pst, že při 100 hodech panna padne alespoň 70x.

$X_1, \dots, X_{100}$

$X_i \begin{cases} 0 & \text{orel v i-tém hodě} \\ 1 & \text{panna} \end{cases}$

$X_1, \dots, X_{100}$ nezávislé

$X_i \sim Alt(3/4)$

$EX_i = 3/4$

$DX_i = 3/4 \cdot \frac{1}{4} = \frac{3}{16} < \infty$

$$\begin{aligned} P\left(\sum_{i=1}^{100} X_i \geq 70\right) &= P\left(\sum_{i=100}^{100} X_i - 75 \geq 70 - 75\right) \\ &= P\left(\frac{\sum_{i=1}^{100} X_i - 75}{\sqrt{100 \cdot \frac{3}{16}}} \geq \frac{70-75}{\sqrt{100 \cdot \frac{3}{16}}}\right) \\ &= P\left(U_{100} \geq -\frac{-2}{\sqrt{3}}\right) = 1 - P\left(U_{100} < \frac{-2}{\sqrt{3}}\right) \end{aligned}$$

U_{100} podle CLV má rozdělení blízké $N(0,1)$

$$1 - P\left(U_{100} < \frac{-2}{\sqrt{3}}\right) = 1 - \Phi\left(\frac{-2}{\sqrt{3}}\right) = 1 - \left(1 - \Phi\left(\frac{2}{\sqrt{3}}\right)\right) = \Phi\left(\frac{2}{\sqrt{3}}\right) = 0.875$$

Příklad. Kolik nejméně musím vytáhnout čísel z $[0,1]$, aby pravděpodobnost, že jejich průměr byl menší než $\frac{7}{12}$ byla alespoň 95 %.

$X_1, \ldots, X_n \sim U[0,1]$, nezávislé.

$EX_i = \frac{1}{2}$

$DX_i = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$

$$\begin{aligned}
P\left(\frac{1}{n}\sum_{i=1}^{n} X_i \leq \frac{7}{12}\right) &\geq 0.95 \\
P\left(\frac{\overline{X}_n - \frac{1}{2}}{\sqrt{\frac{1}{12n}}} \leq \frac{7/12 - \frac{1}{2}}{\sqrt{\frac{1}{12n}}}\right) &\geq 0.95 \\
TODOdopsatdkazP\left(U_n \leq \sqrt{\frac{n}{12}}\right) &\geq 0.95 \\
\Phi\left(\sqrt{\frac{n}{12}}\right) &\geq 0.95 \\
\sqrt{\frac{n}{12}} &\geq \Phi^{-1}(0.95) \\
\sqrt{\frac{n}{12}} &\geq 1.65 \\
\sqrt{n} &\geq 1.65 \cdot \sqrt{12} \\
n &\geq 32.67 \\
n &\geq 33
\end{aligned}$$

Lze si dovolit aproximaci pro $\sum_{i=1}^{33} X_i$ pro $n = 33$? Ano. (Jde o případ rovnoměrného rozdělení na intervalu, které lze použít pro $n \geq 12$ pod 1 %)

Příklad. Protipříklady na CLV.

$X_1, X_2, X_3, \ldots$

$X_i \sim Pa(1.4)$

jak vypadá $\sum_{i=1}^{n} X_i$

$EX_i = \frac{1.4}{0.4} = 3.5$

$EX_i^2 = \infty \implies DX_i = \infty$

Nemůžeme použít CLV.

Přednáška 19: Důkaz CLV

21. 4. 2021

Definice 19.1 (Charakteristická funkce (Inverzní Fourierova transformace))**.** n. v. X je definována jako $\phi_X(t) = E(e^{itX})$, $t \in \mathbb{R}$.

Pro spojitě rozdělenou X:

$$\phi_X(t) = \int_{-\infty}^{\infty} e^{itx} \cdot f_X(x)\,dx$$

Vlastnosti 19.2 (charakteristické funkce)**.**

1. $\phi_X(0) = E(e^{i\cdot 0\cdot X}) = 1$
2. $|\phi_X(t)| = |E(e^{itX})| \leq E|e^{itX}| \leq 1$
3. $\phi_{cX+d}(t) = E(e^{it(cX+d)}) = e^{idt} \cdot E(e^{ictX}) = e^{idt}\phi_X(ct)$
4.

$$\begin{aligned} \frac{d^n}{dt^n}\phi_X(t)\bigg|_{t=0} &= \frac{d^n}{dt^n}\bigg|_{t=0} Ee^{itX} \\ &= \int_{-\infty}^{\infty} \frac{d^n}{dt^n}\bigg|_{t=0} e^{itx}\,dF(x) \\ &= \int_{-\infty}^{\infty} i^n x^n e^{ix\cdot 0}\,dF(x) \\ &= i^n \int_{-\infty}^{\infty} x^n\,dF(x) \\ &= i^n \cdot E(X^n) \end{aligned}$$

5. pokud $X_1, \dots, X_n$ jsou nezávislé n. v., potom

$$\phi_{\sum_{k=1}^n X_k(t)} = Ee^{it\sum_{k=1}^n X_k} = E\prod_{k=1}^n e^{itX_k} = \prod_{k=1}^n Ee^{itX_k} = \prod_{k=1}^n \phi_{X_k}(t)$$

Věta 19.3 (o spojitosti)**.** Nechť F je dist. fce veličiny s char. fcí ϕ. Potom pro libovolnou posloupnost n. v. $X_1, X_2, X_3, \dots$ s char. fcemi $\phi_1, \phi_2, \dots$ a dist. fcemi $F_1, F_2, \dots$ platí, že $F_n \to F$ v každém bodě spojitosti F právě tehdy, když $\phi_n \to \phi$ bodově.

Tvrzení 19.4 (čemu je rovna charakteristická funkce normálního rozdělení)**.** $X \sim N(\mu, \sigma^2) \implies \phi_X(t) = e^{(it\mu - \frac{\sigma^2 t^2}{2})}$

Důkaz.

$$\begin{aligned}
\phi_X(t) &= Ee^{itX} = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{itx} \cdot e^{-\frac{x^2}{2}}\, dx \\
\phi'_X(t) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} ix \cdot e^{itx} \cdot e^{-\frac{x^2}{2}}\, dx \\
&= -\frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} \cdot \frac{de^{-x^2/2}}{dx}\, dx \\
&= \frac{-i}{\sqrt{2\pi}} \left[e^{itx} e^{-x^2/2} \right]_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} ite^{itx} \cdot e^{x^2/2}\, dx \\
&= 0 - t \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{x^2/2} \cdot e^{itx}\, dx \\
&= -t\phi_X(t)
\end{aligned}$$

Víme, že $\phi'(t) = -t\phi(t)$, z toho $\frac{\phi'(t)}{\phi(t)} = -t$, pak integrací:

$$\begin{aligned}
\log \phi(t) &= -\frac{t^2}{2} + c \\
\phi(t) &= e^{-t^2/2} \cdot k \\
\phi(0) &= 1 \implies k = 1
\end{aligned}$$

□

Důkaz. Důkaz CLV.

Postup důkazu: vezmeme veličiny $X_1, X_2, X_3, \ldots$ provedeme "F. transformaci" na veličinu $U_n = \frac{\sum X_i - \mu n}{\sqrt{n\sigma^2}}$ sérií úprav zjistíme, že konverguje k char. fci rozdělení $N(0,1)$. Z jednoznačnosti F. transformace $\implies U_n \to N(0,1)$ v distribuci.

$$U_n = \frac{\sum_{k=1}^{n} X_k - \mu \cdot n}{\sqrt{n\sigma^2}} = \frac{\overline{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} Z_k,$$

kde $Z_k = \frac{X_k - \mu}{\sigma}$, $EZ_k = 0, DZ_k = 1$.

$\phi'_{Z_k}(0) = i \cdot EZ_k = 0$, $\phi(Z_k)''(0) = (-1)EZ_k^2 = -1$

Taylor pro ϕ_{Z_1} okolo 0:

$$
\begin{aligned}
&\phi_{Z_1}\left(\frac{t}{\sqrt{n}}\right)\\
&= \phi_{Z_1}(0) + \phi'_{Z_1}(0)\frac{t}{\sqrt{n}} + \phi''_{Z_1}(0)\cdot\left(\frac{t}{\sqrt{n}}\right)^2\frac{1}{2} + o\left(\left(\frac{t}{\sqrt{n}}\right)^2\right)\\
&= 1 + 0 - \frac{t^2}{2n} + o(\frac{t^2}{n})\\
&= 1 - \frac{t^2}{2n} + o(\frac{t^2}{n})
\end{aligned}
$$

$$
\begin{aligned}
\phi_{U_n}(t) &= \phi_{\frac{1}{\sqrt{n}}\sum_{k=1}^n Z_k}(t)\\
&= \phi_{\sum_{k=1}^n Z_k}\left(\frac{t}{\sqrt{n}}\right)\\
&= \prod_{k=1}^n \phi_{Z_k}\left(\frac{t}{\sqrt{n}}\right)\\
&= \left(\phi_{Z_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n\\
&= \left(1 - \frac{t^2}{2n} + o(\frac{t^2}{n})\right)^n\\
&= e^{n\cdot\log\left(1-\frac{t^2}{2n}+o\left(\frac{t^2}{n}\right)\right)} \xrightarrow{n\to\infty} e^{-t^2/2} = \phi_X(t)
\end{aligned}
$$

Z toho: $X \sim N(0,1)$ □

Přednáška 20: Konvergence CLV pro různá rozdělení, teorie odhadu

22. 4. 2021

Konvergence CLV pro různá rozdělení přednáška 10b.

Definice 20.1 (Náhodný výběr)**.** Nezávislé, stejně rozdělené veličiny $X_1, \dots, X_n, n \in \mathbb{N}$ nazýváme náhodný výběr o rozsahu n. (Pokud mají např. dist. fci F, mluvíme o náhodném výběru z rozdělení F)

Libovolné hodnotě $(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ vektoru $(X_1, \dots, X_n)$ říkáme realizace n výběru.

To odpovídá tomu, že máme sérii n nezávislých experimentů/pozorování, které jsou za stejných podmínek.

Příklad. Hodíme n-krát mincí. $X_i = \begin{cases} 0 & \text{v i-tém hodu padne orel} \\ 1 & \text{v i-tém hodu padne panna} \end{cases}$ Veličiny $X_1, \dots, X_n$ tvoří náhodný výběr.

Realizací tohoto výběru může být např. $(1, 0, 0, 0, \dots, 0)$

Příklad. Máme minci, která hází pannu s pstí p. Chceme odhadnout p, (nám bude stačit stanovit, jestli $p = \frac{1}{2}$).

<u>Postup:</u> hodíme 1000x mincí. Realita: panna padla 523x.

Otázka: jak je pravděpodobné, že $p = \frac{1}{2}$? Budeme postupovat podle CLV.

Za předpokladu, že $p = \frac{1}{2}$, $EX_i = \frac{1}{2}$, $DX_i = \frac{1}{4}$.

$$\begin{aligned} P\left(\sum_{i=1}^{1000} X_i \geq 523\right) &= P\left(\frac{\sum_{n=1}^{1000} X_i - 5000}{\sqrt{250}} \geq \frac{23}{\sqrt{250}}\right) \\ = P\left(U_{100} \geq \frac{23}{\sqrt{250}}\right) &= 1 - \Phi\left(\frac{23}{\sqrt{250}}\right) = 1 - 0.93 = 0.07. \end{aligned}$$

V praxi často chceme odhadovat hodnoty parametrů rozdělení nebo jejich charakteristiky, funkce.

Definice 20.2 (bodový odhad parametru)**.** Nechť $X_1, \dots, X_n$ je n. výběr z rozdělení s parametrem (charakteristikou) θ.

Libovolná funkce $g(X_1, \dots, X_n)$ náhodného výběru $X_1, \dots, X_n$, která není závislá na θ se nazývá bodový odhad parametru θ.

Příklad. Máme výběr z rozdělení $Alt(p), p > 0$. Utvořte různé odhady parametru p. Jak jsou dobré?

1. Výběrový průměr
$$g(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}_n \text{ ... značíme výběrový průměr.}$$
2. Ne úplně dobrý odhad
$$g(X_1, \dots, X_n) = \frac{\max X_i - \min X_i}{2}.$$
3. Velmi špatný příklad odhadu
$$g(X_1, \dots, X_n) = 0.$$
Je také bodový odhad.

Definice 20.3 (nestranný odhad)**.** Odhad parametru θ se nazývá nestranný, pokud pro každou hodnotu θ je

$$E_\theta g(X_1, \dots, X_n) = \theta$$

Příklad. Výběrový průměr $\overline{X}_n$ je nestranný odhad střední hodnoty EX, pro $EX < \infty$.

Důkaz.

$$E\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n}E\left(\sum_{i=1}^n X_i\right) = \frac{1}{n} \cdot n \cdot EX = EX$$

□

Přednáška 21: Výběrový rozptyl, metoda maximální věrohodnosti

28. 4. 2021

Mějme $X_1, \dots, X_n$ náhodný výběr.

Realizace $(x_1, x_2, \dots, x_n)$. Snažíme se určit parametr rozdělení.

Bodový odhad $g : (X_1, \dots, X_n) \to \mathbb{R}$ pro k-rozměrný parametr. Bodový odhad je nezávislý na parametru.

Víme: Výběrový průměr $\overline{X}$ je nestranný odhad střední hodnoty. (pokud střední hodnota existuje)

Definice 21.1 (výběrový rozptyl)**.** Nechť $X_1, \dots, X_n$ je n. výběr. Veličinu

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

nazýváme výběrový rozptyl a $S_n = \sqrt{S_n^2}$ výběrová směrodatná odchylka.

Tvrzení 21.2. Nechť $X_1, \dots, X_n$ je n. výběr z rozdělení s rozptylem $\sigma^2 < \infty$. Potom platí $ES_n^2 = \sigma^2$ (S_n^2 nestranný odhad σ^2)

Důkaz.

$$\begin{aligned}
ES_n^2 &= \frac{1}{n-1} E \sum_{i=1}^{n} (X_i^2 - 2\overline{X} \cdot X_i + \overline{X}^2) \\
&= \frac{1}{n-1} E \left(\sum_{i=1}^{n} X_i^2 - 2\overline{X}^2 n + n\overline{X}^2 \right) \\
&= \frac{1}{n-1} E \left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2 \right)
\end{aligned}$$

$$\begin{aligned}
ES_n^2 &= \frac{1}{n-1} E \left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2 \right) \\
&= \frac{1}{n-1} \left(\sum_{i=1}^{n} EX_i^2 - nE\overline{X}^2 \right) \\
&= \frac{1}{n-1} \left(n \cdot EX_1^2 - (EX_1^2 + (n-1)(EX_1)^2 \right) \\
&= \frac{1}{n-1} \left((n-1)(EX_1^2 - (EX_1)^2) \right) = EX_1^2 - (EX_i)^2 = DX = \sigma^2
\end{aligned}$$

$$\begin{aligned}
E\overline{X}^2 &= E \left(\frac{1}{n} \sum_{i=1}^{n} X_i \right)^2 \\
&= \frac{1}{n^2} E \left(\sum_{i=1}^{n} X_i \right)^2 \\
&= \frac{1}{n^2} \left(\sum_{i=1}^{n} EX_i^2 + 2 \sum_{i<j} EX_i X_j \right) \\
&= \frac{1}{n^2} \left(\sum_{i=1}^{n} EX_i^2 + 2 \sum_{i<j} EX_i \cdot EX_j \right) \\
&= \frac{1}{n^2} \left(\sum_{i=1}^{n} EX_i^2 + 2 \binom{n}{2} (EX_i)^2 \right) \\
&= \frac{EX_1^2 + (n-1)(EX_i^2)}{n}
\end{aligned}$$

□

21.1 Metoda maximální věrohodnosti (Maximum likelihood method)

Příklad. Máme minci, která hází pannu s pstí $p \in (0,1)$. Provedeme sérii hodů s výsledky

hody	n
panna	k

Určíme pst výsledku: P(Při n hodech padla k-krát panna) $= \binom{n}{k}p^k(1-p)^{n-k}$.

Definuji si věrohodnostní funkci $L(p) = P$(realizovaný výsledek experimentu). Hledám p tak, aby $L(p)$ měla co nejvyšší hodnotu.

Hledám maximum fce $L(p) = \binom{n}{k}p^k(1-p)^{n-k}$

Můžeme zlogaritmovat (body extrémů se nezmění)

$$\log(L(p)) = \log\binom{n}{k} + k\log p + (n-k)\log(1-p)$$

$$\frac{\partial}{\partial p}\log(L(p)) = 0 + \frac{k}{p} + \frac{n-k}{1-p}\cdot(-1) = 0$$

$$\frac{k}{p} = \frac{n-k}{1-p}$$
$$k - kp = pn - pk \implies p = \frac{k}{n}$$

Najdu maximum L, potom maximálně věrohodný odhad MLE(p) = bod maxima L = $\frac{k}{n}$

Obecný postup pro diskrétní rozdělení

Mějme náhodný výběr s realizací $(x_1, \dots, x_n)$.

1. Najdeme věrohodnostní funkci

$$L(\theta) = P((X_1, \dots, X_n) = (x_1, \dots, x_n)) = \prod_{i=1}^{n} P(X_i = x_i) = \prod_{i=1}^{n} P_X(x_i)$$

2. Hledáme max $L(\theta)$ (obvykle přes zlogaritmování)
3. Pokud θ_0 je bod maxima L, potom MLE(θ) = θ_0

Obecný postup pro spojité rozdělení

Mějme náhodný výběr s realizací $(x_1, \dots, x_n)$.

Jako věrohodnostní funkci bereme

$$L(\theta) = \prod_{i=1}^{n} f_X^{\theta}(x_i)$$

Jinak postup stejný jako pro diskrétní rozdělení.

Příklady

Příklad. Nechť $(x_1, x_2, \dots, x_n)$ je realizací n. výběru z rozdělení $U[0, \alpha]$, kde $\alpha > 0$ je parametr. MMV odhadněte α.

$$f_X(u) = \begin{cases} \frac{1}{\alpha} & u \in [0, \alpha] \\ 0 & \text{jinak} \end{cases}$$

$$L(\alpha) = \prod_{i=1}^{n} f_X^{\theta}(x_i) = \begin{cases} \left(\frac{1}{\alpha}\right)^n & (1)\, \alpha \geq x_i, \forall i \in \{1, \dots, n\} \\ 0 & (2)\, \exists i \in \{1, \dots, n\} : \alpha < x_i \end{cases}$$

Zřejmě L nebude maximální, pokud platí podmínka (2). Tedy (1) platí. Maximalizuji fci $L_1(\alpha) = \left(\frac{1}{\alpha}\right)^n$, $\alpha \geq \max x_i$. $\left(\frac{1}{\alpha}\right)^n$ je klesající v $\alpha \implies$ max L se nabývá v krajním bodě def. oboru => MLE(α) = $\max x_i$.

Přednáška 22: Příklad MMV pro Pa(), Odhady

29. 4. 2021

Příklad. Mějme výběr z $Pa(\alpha)$ o rozsahu $n \in \mathbb{N}$, s realizacemi $(x_1, \dots, x_n)$. Odhadněte střední hodnotu tohoto rozdělení

1. Výběrový průměr

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

2. Metodou MV

Víme $X \sim Pa(\alpha)$, $\alpha > 1 \implies EX = \frac{\alpha}{\alpha - 1}$

$$L(\alpha) = \prod_{i=1}^{n} \frac{\alpha}{x_i^{\alpha+1}}$$

$$\log L(\alpha) = \sum_{i=1}^{n} \log \frac{\alpha}{x_i^{\alpha+1}} = \sum_{i=1}^{n} (\log \alpha - (\alpha + 1) \log x_i) = n \log \alpha - (\alpha + 1) \sum_{i=1}^{n} \log x_i$$

$$\frac{\partial}{\partial \alpha} log(L(\alpha)) = \frac{n}{\alpha} - \sum_{i=1}^{n} \log(x_i) = 0$$

$$\frac{n}{\alpha} = \sum_{i=1}^{n} \log(x_i)$$

$$\alpha = \frac{n}{\sum_{i=1}^{n} \log(x_i)}$$

Kvantilová funkce

$$F_X(u) = \begin{cases} 0 \\ 1 - \frac{1}{u^\alpha} \end{cases}$$

$$\begin{aligned} \frac{1}{u^\alpha} &= 1 - F_X(u) \\ u^\alpha &= \frac{1}{1 - F_X(u)} \\ u &= \frac{1}{(1-F)^{1/\alpha}} \\ Q(u) &= \frac{1}{(1-u)^{1/\alpha}} \end{aligned}$$

Pokus o konkrétní hodnoty Vezmeme hodnoty rovnoměrného rozdělení na $[0, 1]$. Poté spočteme $Pa(1, 16)$ pomocí kvantilové funkce aplikované na hodnoty rovnoměrného rozdělení. A z těchto hodnot pak aritmetický průměr a průměr logaritmů - z něj pak střední hodnotu.

Definice 22.1 (intervalové odhady)**.** Nechť $X_1, \ldots, X_n$ je n. výběr z rozdělení s parametrem θ.

Dvojice fcí (statistik) $\{\psi_L(X_1, \ldots, X_n), \psi_U(X_1, \ldots, X_n)\}$ nazýváme intervalový odhad parametru θ o spolehlivosti $1 - \alpha$, pokud platí že $\forall\, \theta$:

$$P_\theta(\psi_L(X_1, \ldots, X_n) < \theta < \psi_U(X_1, \ldots, X_n)) = 1 - \alpha$$

Intervalu $(\psi_L(X_1, \ldots, X_n), \psi_U(X_1, \ldots, X_n))$ říkáme konfidenční interval.

Definice 22.2. Statistiky $\psi_L(X_1, \ldots, X_n)$, $\psi_U(X_1, \ldots, X_n)$ se nazývají dolní, resp. horní odhad parametru θ o spolehlivosti $1 - \alpha$, pokud platí

$$P_\theta(\psi_L(X_1, \ldots, X_n) < \theta) = 1 - \alpha,$$

resp.

$$P_\theta(\psi_U(X_1, \ldots, X_n) > \theta) = 1 - \alpha.$$

Interval $(-\infty, \psi_U(X_1, \ldots, X_n)$, resp. $(\psi_L(X_1, \ldots, X_n), \infty)$ se nazývá dolní (nebo levostranný), resp. horní (nebo pravostranný) interval spolehlivosti.

Definice 22.3 (β-kvantil)**.** Nechť distribuční funkce F je spojitá a rostoucí. Nechť $\beta \in (0,1)$.

Hodnotu $F^{-1}(\beta) = Q(\beta)$ nazveme β-kvantil rozdělení $\sim$ F. Q je kvantilová funkce.

Pozn. pro $X \sim N(0,1)$ značíme často jako β-kvantil u_β.

Obecně značíme $q_F(\beta)$ jako β-kvantil rozdělení F.

$X \sim F$

$$P(q_F(\frac{\beta}{2}) < X < q_F(1-\frac{\beta}{2})) = F_X(q_F(1-\frac{\beta}{2})^-) - F_X(q_F(\frac{\beta}{2}) = 1 - \frac{\beta}{2} - \frac{\beta}{2} = 1-\beta$$

Věta 22.4 (odhad μ pro $N(\mu,\sigma^2)$, σ^2 známe)**.** Nechť $X_1,\ldots,X_n$ je výběr z $N(\mu,\sigma^2)$, kde $\sigma^2 > 0$ je známo.

Potom $\{\overline{X} - u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \overline{X} + u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\}$ je intervalový odhad μ o spolehlivosti $1-\alpha$.

Důkaz. Platí $X_1,\ldots,X_n$ nezávislé.

Prvé odvození: $X \sim N(\mu,\sigma^2)$

$$\frac{X-\mu}{\sigma} \sim N(0,1)$$

$$\frac{\frac{1}{n}\sum X_i - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

(Díky nezávislosti)

$$\frac{\overline{X}-\mu}{\sigma} \cdot \sqrt{n} \sim N(0,1)$$

$$P\left(u_{\alpha/2} < \frac{\overline{X}-\mu}{\sigma} \cdot \sqrt{n} < u_{1-\alpha/2}\right) = 1-\alpha$$
$$P\left(\frac{u_{\alpha/2}\sigma}{\sqrt{n}} < \overline{X}-\mu < \frac{u_{1-\alpha/2}\sigma}{\sqrt{n}}\right) = 1-\alpha$$
$$P\left(\overline{X} - \frac{u_{1-\alpha/2}\sigma}{\sqrt{n}} < \mu < \overline{X} - \frac{u_{\alpha/2}\sigma}{\sqrt{n}}\right) = 1-\alpha$$
$$P\left(\overline{X} - \frac{u_{1-\alpha/2}\sigma}{\sqrt{n}} < \mu < \overline{X} + \frac{u_{1-\alpha/2}\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$\square$

Příklad. Hmotnost zajíce v revíru se řídí rozdělením $N(\mu, 2)$. Odchycením 6 zajíců jsme naměřili hmotnosti: $4.23, 3.81, 6.15, 5.40, 4.73, 3.9$

Sestrojte intervalový odhad pro průměrnou hmotnost zajíce v revíru. Pracujte se spolehlivostí 95 %.

$X_1, \ldots, X_6 \sim N(\mu, 2)$

$$I\left(\overline{X} - u(0.975)\frac{\sigma}{\sqrt{n}}, \overline{X} + u(0.975)\frac{\sigma}{\sqrt{n}}\right)$$

Průměr z dat je 4.7, $u(0.975) = 1.96$

$L = 3.572, U = 5.835$

$\overline{X} = 4.703$

$I = (3.572, 5.835)$

Přednáška 23: Chí kvadrát, Studentovo rozdělení

5. 5. 2021

Definice 23.1 (Funkce Gamma)**.** Funkce $\Gamma : \mathbb{R}^+ \to \mathbb{R}$ (Gamma) je definována následovně

$$\Gamma(a) = \int_0^\infty e^{-x} \cdot x^{a-1}\, dx$$

Víme (per parte): $\Gamma(n) = (n-1)!, n \in \mathbb{N}$

Věta 23.2 (chí-kvadrát o n stupních volnosti)**.** Nechť $X_1, \ldots, X_n$ jsou i. i. d, $X_i \sim N(0,1)$. Potom má veličina $Y = \sum_{i=1}^n X_i^2$ hustotu

$$g_n(t) = \frac{t^{n/2-1}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \cdot e^{-t/2}, t > 0$$

Rozdělení Y se nazývá χ_n^2 chí-kvadrát o n stupních volnosti.

Pozn. $EY = n$, $DX = 2n$

Věta 23.3. Nechť $X_1, \ldots, X_n$ je n. výběr z $N(\mu, \sigma^2)$. Potom platí:

1. $\overline{X}$ a S_n^2 jsou nezávislé.
2. $\frac{(n-1)}{\sigma^2} \cdot S_n^2$ má rozdělení χ_{n-1}^2.

Důkaz. Důkaz v zápiscích chybí. Je v přednášce. $\square$

Důsledek 23.3.1. Nechť $X_1, \dots, X_n$ je n. výběr z rozdělení $N(\mu, \sigma^2)$. Potom je interval

$$\left(\frac{(n-1)S_n^2}{q_{\chi^2_{n-1}}(1-\frac{\alpha}{2})}, \frac{(n-1)S_n^2}{q_{\chi^2_{n-1}}(\frac{\alpha}{2})} \right)$$

konfidenčním intervalem pro rozptyl σ^2 se spolehlivostí $1-\alpha$.

Důkaz. $X_1, \dots, X_n \sim N(\nu, \sigma^2)$

vím:

$$P\left(q_{\chi^2_{n-1}}(\alpha/2) < \frac{(n-1)S_n^2}{\sigma^2} < q_{\chi^2_{n-1}}(1-\alpha/2) \right) = 1-\alpha$$

$$P\left(\frac{1}{q_{\chi^2_{n-1}}(1-\alpha/2)} < \frac{\sigma^2}{(n-1)S_n^2} < \frac{1}{q_{\chi^2_{n-1}}(\alpha/2)} \right) = 1-\alpha$$

$$P\left(\frac{(n-1)S_n^2}{q_{\chi^2_{n-1}}(1-\alpha/2)} < \sigma^2 < \frac{(n-1)S_n^2}{q_{\chi^2_{n-1}}(\alpha/2)} \right) = 1-\alpha$$

□

Věta 23.4. Nechť $X \sim N(0,1)$ a Y_n χ^2_n jsou nezávislé n. v.

Potom má veličina

$$Z_n = \frac{X}{\sqrt{Y_n/n}}$$

rozdělení s hustotou

$$h_n(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n} \cdot \Gamma\left(\frac{n}{2}\right)} \cdot \left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}} = k_n \cdot \left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}, t \in \mathbb{R}$$

Tomuto rozdělení se říká Studentovo t-rozdělení o n stupních volnosti.

Student. t-rozdělení je symetrické kolem 0. $EZ_n = 0$

Přednáška 24: Studentovo t-rozdělení, testování hypotéz

6. 5. 2021

Pro vysoké hodnoty $|t| = \mathrm{h_n}(t) \sim t^{-(n+1)} \cdot c$ polynomiální pokles.

$$\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

Lze rozdělit na:

$$\left(1+\frac{t^2}{n}\right)^{-1/2} \to 1,$$

$$\left(1-\frac{\frac{t^2}{2}}{-\frac{n}{2}}\right)^{-n/2} \to e^{-t^2/2}$$

.

Celé jde k

$$e^{t^2/2}$$

.

Tedy pro $n \to \infty$ platí $t_n \to N(0,1)$ (v distribuci).

V praxi se student. t-dist tabeluje např. do n = 100, poté se kvantily nahrazují kvantily $N(0,1)$.

Věta 24.1. Nechť $X_1, \ldots, X_n$ je n. výběr z $N(\mu, \sigma^2)$. $\sigma^2 > 0$. Potom má veličina

$$T = \frac{\overline{X} - \mu}{S_n}\sqrt{n}$$

Studentovo t-rozdělení o $(n-1)$ stupních volnosti.

Důkaz. Víme $X' \sim N(0,1)$, $Y_n \sim X_n^2$ nezávislé.

Potom

$$\frac{X'}{\sqrt{\frac{Y_n}{n}}} \sim t_n$$

$$X' = \frac{\overline{X} - \mu}{\sigma}\sqrt{n} \ldots \text{opravdu má N(0, 1)}$$

$$Y = \frac{S_n^2(n-1)}{\sigma^2} \ldots \text{odpovídá } \chi^2_{n-1}$$

X', Y nezávislé, neboť $\overline{X}$ a S_n^2 jsou nezávislé.

$$t_{n-1} \sim \frac{X'}{\sqrt{\frac{Y}{n-1}}} = \frac{\frac{\overline{X}-\mu}{\sigma}\sqrt{n}}{\sqrt{\frac{S_n^2(n-1)}{\sigma^2(n-1)}}} = \frac{\overline{X} - \mu}{S_n}\sqrt{n} = T$$

$\square$

Důsledek 24.1.1. Nechť $X_1, \ldots, X_n$ je n. výběr z $N(\mu, \sigma^2)$, kde ani jeden z parametrů není znám. Potom

$$\left\{\overline{X} - q_{t_{n-1}}\left(1-\frac{\alpha}{2}\right)\cdot\frac{S_n}{\sqrt{n}}, \overline{X} + q_{t_{n-1}}\left(1-\frac{\alpha}{2}\right)\cdot\frac{S_n}{\sqrt{n}}\right\}$$

je int. odhad μ o spolehlivosti $1-\alpha$.

Důkaz. Ze symetrie t-rozdělení platí $q_{t_{n-1}}(\alpha/2) = -q_{t_{n-1}}(1-\frac{\alpha}{2}$.

Platí

$$P\left(q_{n-1}(\alpha/2) < T < q_{t_{n-1}}(1-\alpha/2)\right) = 1-\alpha$$
$$P\left(-q_{n-1}(1-\frac{\alpha}{2})\cdot\frac{S_n}{\sqrt{n}} < \overline{X} - \mu < q_{t_{n-1}}(1-\frac{\alpha}{2})\cdot\frac{S_n}{\sqrt{n}}\right) = 1-\alpha$$
$$P\left(\overline{X} - q_{t_{n-1}}\left(1-\frac{\alpha}{2}\right)\frac{S_n}{\sqrt{n}} < \mu < \overline{X} + q_{t_{n-1}}\left(1-\frac{\alpha}{2}\right)\cdot\frac{S_n}{\sqrt{n}}\right) = 1-\alpha$$

□

24.1 Testování hypotéz

Máme výběr $X_1, \ldots, X_n$ z rozdělení F s parametrem θ a chceme otestovat dvě proti sobě jdoucí hypotézy o θ.

1. $\theta \in \Omega_0$,
2. $\theta \in \Omega_1$,

$\Omega_0 \cap \Omega_1 = \emptyset$

Typicky si stanovíme tzv. nulovou hypotézu $H_0 : \theta \in \Omega_0$ a tzv. alternativní hypotézu $H_1 : \theta \in \Omega_1$. Obvykle $\Omega_1 = \overline{\Omega}_0$.

Testem nulové hypotézy proti alternativní rozumíme rozhodovací proces založený na realizaci výběru $x_1, \ldots, x_n$, na jehož základě zamítneme nebo nezamítneme H_0. (Když zamítneme H_0, je to ve prospěch H_1.)

Možné 4 výsledky

1. platí H_0 a my nezamítáme H_0.
2. zamítneme H_0 i když H_0 platí. To je chyba 1. druhu.
3. nezamítneme H_0, ale H_0 neplatí. To je chyba 2. druhu.
4. zamítneme H_0 a H_0 neplatí.

V praxi chceme kontrolovat chybu 1. druhu (čili $\leq \alpha$) za těchto podmínek minimalizovat chybu 2. druhu.

α se nazývá hladina významnosti testu.

Postupujeme tak, že v závislosti na α stanovíme tzv. kritický obor $W \subseteq \mathbb{R}^n$, což je množina, pro kterou platí $(x_1, \ldots, x_n) \in W$, potom H_0 zamítáme. (Kritický obor)

Pro W platí $P_\theta((X_1, \ldots, X_n) \in W) \leq \alpha$, $\forall \theta \in \Omega_0$. Dále chceme, aby $P_\theta((X_1, \ldots, X_n) \notin W)$, $\theta \in \Omega_1$ bylo minimální.

Příklad. Máme výběr z $N(\mu, 1)$ a chceme testovat $\mu = \mu_0$ (pro jisté dané μ_0) proti $\mu > \mu_0$, na hladině významnosti α.

Tedy: $H_0 : \mu = \mu_0$, $H_1 : \mu > \mu_0$

Sestavíme testovací statistiku

$$U = \frac{\overline{X} - \mu}{\sigma}\sqrt{n} = \frac{\overline{X} - \mu}{1}\sqrt{n}$$

Pokud platí H_0, potom $U \sim N(0, 1)$

Hodnotu $U(x_1, \ldots, x_n) > u_{1-\alpha}$, potom zamítám H_0.

Pokud platí H_0 (tj. $\mu = \mu_0$, potom $P(\text{zamítáme } H_0) = \alpha$. Kritický obor je roven $W\{(x_1, \ldots, x_n) \in \mathbb{R}^n | U(x_1, \ldots, x_n) > u_{1-\alpha}\}$

Příklad. Odpor rezistoru je n. veličina s rozdělením $N(\mu, 0.01)$ v Ohmech. Provedli jsme 5 měření, jejichž hodnoty byly: $3.1, 2.95, 3.23, 3.01, 2.98$.

Testujte na hladině $\alpha = 0.05$, že průměrný odpor rezistoru je roven 3Ω.

Hypotézy: $H_0 : \mu = 3$, $H_1 : \mu \neq 3$

Za platnosti H_0 má veličina

$$U = \frac{\overline{X} - \mu_0}{\sigma}\sqrt{n} = \frac{\overline{X} - 3}{0.1}\sqrt{5}$$

rozdělení $N(0, 1)$. Dosadím hodnoty realizací.

$$\overline{X} = \frac{15.27}{5} = 3.054,$$

$$U = \frac{0.054}{0.1}\sqrt{5} = 0.54\sqrt{5} = 1.207.$$

$$u_{0.95} = 1.64$$

Na základě naměřených dat na hladině významnosti 0.05 nelze zamítnout H_0.

Přednáška 25: Statistické testy 13. 5. 2021

25.1 Testy normálního rozdělení

Situace: Máme výběr $X_1, \dots, X_n$ z rozdělení $N(\mu, \sigma^2)$.

1. Chceme testovat $\mu = \mu_0$ a je známa hodnota $\sigma^2 > 0$. (Test střední hodnoty n. rozdělení při známém rozptylu μ_0 je předem stanovená.

   $H_0 : \mu = \mu_0$

   $H_1 : \mu \neq \mu_0$ (oboustranný test)

   případně $\mu > \mu_0$ nebo $\mu < \mu_0$

   Statistika

   $$U = \frac{\overline{X} - \mu_0}{\sigma} \sqrt{n} \text{ má za platnosti } H_0 \text{ rozdělení } N(0,1)$$

   Z toho vyplývá, že kritický obor pro test na hladině α:

   Pro oboustrannou variantu

   $$W_\alpha = \{x \in \mathbb{R}^n | |U(z)| > u_{1-\frac{\alpha}{2}}\}$$

   Pro $H_1 : \mu > \mu_0$

   $$W_\alpha = \{x \in \mathbb{R}^n | U(x) > u_{1-\alpha}\}$$

   Pro $H_1 : \mu < \mu_0$

   $$W_\alpha = \{x \in \mathbb{R}^n | U(x) < -u_{1-\alpha}\}$$

2. Chceme testovat σ^2 při neznámých parametrech $\mu, \sigma^2 > 0$

   $H_0 : \sigma = \sigma_0$, pro dané σ_0

   $H_1 : \sigma \neq \sigma_0$

   případně $\sigma > \sigma_0, \sigma < \sigma_0$

   Za platnosti H_0 má veličina Y

   $$Y = \frac{S_n^2(n-1)}{\sigma_0^2} \text{ rozdělení } \chi^2_{n-1}$$

   Pro oboustrannou variantu

   $$W_\alpha = \{x \in \mathbb{R}^n | Y(x) < q_{\chi^2_{n-1}}(\alpha/2) \vee Y(x) > q_{\chi^2_{n-1}}\left(1 - \frac{\alpha}{2}\right)\}$$

   Pro jednostrannou variantu $\sigma > \sigma_0$

   $$W_\alpha = \{x \in \mathbb{R}^n | Y(x) > q_{\chi^2_{n-1}}(1-\alpha)\}$$

   Analogicky $\sigma < \sigma_0$

3. Chceme testovat μ a neznáme $\sigma^2 > 0$ (test střední hodnoty normálního rozdělení s neznámým rozptylem)

   Tento test se nazývá (jednovýběrový) T-test.

   $H_0: \mu = \mu_0$

   $H_1: \mu \neq \mu_0$

   případně $\mu > \mu_0$ nebo $\mu < \mu_0$

   Za platnosti H_0 má veličina

   $$T = \frac{\overline{X} - \mu_0}{S_n}\sqrt{n} \text{ Studentovo t-rozdělení s n-1 stupni volnosti}$$

   Pro oboustrannou variantu

   $$W_\alpha = \{x \in \mathbb{R}^n || T(x)| > q_{t_{n-1}}\left(1 - \frac{\alpha}{2}\right)\}$$

   Pro jednostrannou variantu $\mu > \mu_0$

   $$W_\alpha = \{x \in \mathbb{R}^n | T(x) > q_{t_{n-1}}(1 - \alpha)\}$$

   pro $\mu < \mu_0$ analogicky

25.2 Párový T-test

Situace: máme výběr $(Y_1, Z_1), \ldots, (Y_n, Z_n)$ z nějakého dvourozměrného rozdělení se střední hodnotou $(\mu_1, \mu_2) \in \mathbb{R}^2$. Předpokládejme $Y_i - Z_i \sim N(\mu_1 - \mu_2, \sigma^2), \sigma^2 > 0$. Chceme testovat

$H_0: \mu_1 - \mu_2 = k$ pro předem dané $k \in \mathbb{R}$ (často $k = 0$)

$H_1: \mu_1 - \mu_2 \neq k$

pozn. v praxi bývají Y, Z závislé, často 2 atributy jednoho objektu.

Převedeme na jednovýběrový T-test

$$X_i = Y_i - Z_i \sim N(\mu_1 - \mu_2, \sigma^2)$$

Z toho plyne $X_1, \ldots, X_n$ je výběr z $N(\mu_1 - \mu_2, \sigma^2)$, kde parametry nejsou známy.

$H_0: EX_1 = k$

$H_1: EX_1 \neq k$

$$W_\alpha = \{x \in \mathbb{R}^n || T(x)| > q_{t_{n-1}}\left(1 - \frac{\alpha}{2}\right)\}$$

Příklad. Má se rozhodnout, zda se u osobního automobilu určité značky i typu při správném seřízení geometrie vozu sjíždějí obě přední pneu stejně rychle. Bylo vybráno 6 nových vozů a po určité době bylo zjištěno, o kolik mm se sjely pravé a levé přední pneumatiky

vozidlo	1	2	3	4	5	6
L	1.8	1.0	2.2	0.9	1.5	1.6
P	1.5	1.1	2.0	1.4	1.4	1.4
rozdíl	0.3	-0.1	0.2	-0.2	0.1	0.2

$\alpha = 0.05$

$X = L - P$

$X_i \sim N(\mu, \sigma^2)$

$$T = \frac{\overline{X} - 0}{S_n}\sqrt{n} \sim t_{n-1}$$

Hypotézu, že $H_0 :$ pneu se sjíždějí průměrně stejně $H_1 : \neg H_0$

H_0 zamítne, pokud $|T| > q_{t_5}\left(1 - \frac{\alpha}{2}\right)$

$\overline{X} = 0.0833$

$S^2 = 0.0377$

$T = 1.0518$

$q_{t_5}(0.975) = 2.571$

To znamená, že nelze zamítnout hypotézu o tom, že se pravá a levá pneu sjíždějí stejně.

Věta 25.1. Nechť $(X_1, \ldots, X_n)$ je výběr z $N(\mu_1, \sigma^2))$, $\sigma > 0$ a $Y_1, \ldots, Y_m)$ je výběr z $N(\mu_2, \sigma^2)$. Nechť jsou tyto výběry nezávislé.

Potom má veličina

$$T = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{S_X^2(n-1)S_Y^2(m-1)}}\sqrt{\frac{(n+m-2)nm}{n+m}}$$

rozdělení t_{n+m-2}

Důkaz. (Náznak)

víme $W \sim N(0,1)$, $Z \sim \chi_n^2$

Z toho

$$\frac{W}{\sqrt{\frac{Z}{n}}} \sim t_n$$

$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}} \sim N(0,1) \ (1)$$

$$\frac{X_X^2(n-1)}{\sigma^2} \sim \chi^2_{n-1}, \frac{S_Y^2(m-1)}{\sigma^2} \sim \chi^2_{m-1}$$

$$\frac{S_X^2(n-1) + S_Y^2(m-1)}{\sigma^2} \sim \chi^2_{n+m-2} \ (2)$$

Dáme dohromady (1) a (2)

$$\frac{\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sigma^2(\frac{m+n}{mn})}}{\sqrt{\frac{S_X^2(n-1)+S_Y^2(m-1)}{\sigma^2(n+m-2)}}}$$

Což nám dá veličinu T z věty. □

Přednáška 26: Statistické testy

19. 5. 2021

26.1 Dvouvýběrový T-test

situace: Máme výběr $(X_1, \ldots, X_n)$ z $N(\mu_1, \sigma^2)$, $n \in \mathbb{N}$ a výběr $(Y_1, \ldots, Y_m)$, $m \in \mathbb{N}$.

Tyto výběry (předpokládáme) jsou nezávislé.

Chceme testovat shodnost středních hodnot za předpokladu shody rozptylů.

$H_0 : \mu_1 - \mu_2 = c$ (často volíme $c = 0$.) (případně jednostranné varianty.)

Pokud platí H_0, potom T

$$T = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{(m-1)S_Y^2 + (n-1)S_X^2 E}} \sqrt{\frac{mn(m+n-2)}{m+n}}$$

kritický obor:

$$W_\alpha = \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m | T(X,Y) > q_{t_{m+n-2}}\left(1 - \frac{\alpha}{2}\right)\}$$

Příklad. Máme dvě naleziště zlata. Stejnou metodou zpracováváme rudu. Chceme zjistit, na kterém nalezišti je výhodnější pracovat, máme-li omezené zdroje. Máme data (g zlata na kilogram rudy) k oběma nalezištím (6 vzorků z A, 8 vzorků z B)

A . . . 16.1, 18.3, 12, 17.6, 23.4, 20.2

B . . . 14.8, 16.2, 24.6, 20.4, 19.8, 16.6, 17.7, 15.3

$X_1, \dots, X_6 \sim N(\mu_1, \sigma^2)$

$Y_1, \dots, X_8 \sim N(\mu_2, \sigma^2)$

$\overline{X} = 17.93$

$\overline{Y} = 18.05$

$S_X^2 = 12.3$

$S_Y^2 = 10$

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

Za předpokladu H_0 má statistika

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{7 \cdot S_Y^2 + 5 \cdot S_X^2}} \sqrt{\frac{48(6+8-2)}{6+8}}$$

Studentovo t-rozdělení.

$$T = \frac{-0.12}{\sqrt{131.5}} \sqrt{\frac{288}{7}} = -0.067$$

$|T| = 0.067 < q_{t_{1/2}}(0.975)$

Na hladině 0.05 nelze zamítnout hypotézu o tom, že obě naleziště jsou stejně výnosná v obsahu zlata na kilogram rudy.

Párový T-test vs. dvouvýběrový T-test

Párový T-test

- rozsahy výběru jsou stejné
- výběry nejsou nezávislé, naopak jde o 2 atributy jednoho objektu

Dvouvýběrový T-test

- rozsahy výběrů se můžou lišit
- výběry jsou nezávislé

26.2 Test shodnosti rozptylů (norm. rozdělení)

Máme 2 nezávislé výběry $X_1, \dots, X_n$ z $N(\mu_1, \sigma_1^2)$, $Y_1, \dots, Z_m$ z $N(\mu_2, \sigma_2^2)$

chceme testovat $H_0 : \sigma_1 = \sigma_2$

$H_1 : \sigma_1 \neq \sigma_2$

Za platnosti H_0 má statistika

$$F = \frac{S_X^2}{S_Y^2}$$

rozdělení F (Fisherovo) o $(n-1)$ a $(m-1)$ stupních volnosti.

V praxi za $X_1, \dots, X_n$ volím ten výběr s $S_X^2 > S_Y^2$. Potom pokud $F > q_{F_{n-1,m-i}}$, potom zamítáme $H_0(1-\alpha)$

Fisherovo rozdělení

$X \sim \chi_n^2$, X, Y nezávislé

$Y \sim \chi_m^2$

Potom

$$Z = \frac{\frac{X}{n}}{\frac{Y}{m}}$$

má rozdělení F (Fisherovo) o n, m stupních volnosti.

26.3 Test homogenity 2 binomických rozdělení

Máme dva nezávislé výběry z $Alt(p_1)$ a z $Alt(p_2)$. Čili $X_1, \dots, X_n \sim Alt(p_1)$. $Y_1, \dots, Y_m \sim Alt(p_2)$

$$X = \sum_{i=1}^{n} X_i \sim Bi(n, p_1)$$

$$Y = \sum_{i=1}^{n} Y_i \sim Bi(m, p_2)$$

$H_0: p_1 = p_2$

$H_1: p_1 \neq p_2$

$$x = \frac{X}{n} \dots \text{ odhad } p_1$$

$$y = \frac{Y}{m} \dots \text{ odhad } p_2$$

$$X \sim N(np_1, np_1(1-p_1)), Y \sim N(mp_2, mp_2(1-p_2))$$

$$Ex = p_1, Ey = p_2$$

$$Dx = \frac{p_1(1-p_1)}{n}, Dy = \frac{p_2(1-p_2)}{m}$$

$$z = \frac{X+Y}{n+m}$$

$$U_1 = \frac{x-y}{\sqrt{\frac{x(1-x)}{n} + \frac{y(1-y)}{m}}} \sim N(0,1), \text{ pokud platí} H_0$$

$$U_2 = \frac{x-y}{\sqrt{\frac{z(1-1)}{m+n}}} \sim N(0,1), \text{ pokud platí} H_0$$

H_0 zamítáme, pokud $|U_1| > u_{1-\frac{\alpha}{2}}$, v případě prvního testu.

H_0 zamítáme, pokud $|U_2| > u_{1-\frac{\alpha}{2}}$, v případě druhého testu.

Přednáška 27: Chí kvadrát testy, testy dobré shody

20. 5. 2021

Definice 27.1. Máme $k \in \mathbb{N}$, $k \geq 2$ navzájem se vylučujících jevů.

$A_1, \dots, A_k$, pro které platí $P(\cup_{i=1}^k A_i = 1$

Označme jejich psti $P(A_i) = p_i$ (tedy $\sum_{i=1}^k p_i = 1$)

Provádíme experiment, u kterého jsou možné pouze výsledky $A_1, \dots, A_k$, experiment n-krát opakujeme.

Tedy výsledky testu lze zapsat do tabulky

	A_1	A_2	$\dots$	A_k	celkem
četnost jevu	l_1	l_2	$\dots$	l_k	n

Rozdělení s parametry $n, p_1, p_2, \dots, p_k$ nazýváme multinomické.

Formálně lze říci, že $X \sim Multi(n, p_1, \dots, p_k)$, pokud $X = (X_1, \dots, X_k)$ je n. vektor $X_i \in \{0, \dots, n\}$

$\sum_{i=1}^k X_i = n$

Pstní fce vektoru X s $Multi(n, p_1, \dots, p_k)$ je rovna

$$p_X(a_1, \dots, a_k) = P(X_1 = a_1, X_2 = a_2, \dots, X_k = a_k) = \frac{n!}{a_1! \cdot a_2! \cdot \dots \cdot a_n} p_1^{a_1} \cdot \dots \cdot p_k^{a_k},$$

pro $a = (a_1, \dots, a_k) \in \{0, \dots, n\}^k$

Příklad. n-krát hodíme kostkou. Vektor $X = (X_1, \dots, X_6)$ označující počty jedniček, $\dots$, šestek má multinomické rozdělení o rozsahy n a psti $p_1 = p_2 = \dots = p_6 = \frac{1}{2}$

$X \sim Multi\left(n, \frac{1}{6}, \dots, \frac{1}{6}\right)$

Poznámka. pro $k = 2$ máme $A_1 = \overline{A_2}$

$$\implies p_X(a_1, n - a_1) = \frac{n!}{a_1! \cdot (n - a_1)!} p_1^{a_1}(1 - p_1)^{n-a_1} = \binom{n}{a_1} \cdot p_1^{a_1}(1 - p_1)^{n-a_1}$$

$\implies$ pro k = 2 jde o binomické rozdělení

27.1 Test parametrů multinomického rozdělení

Máme experiment, jehož výsledek lze chápat jako n. vektor $X = (X_1, \dots, X_k) \sim Mult(n, p_1, \dots, p_k)$, kde $p_i > 0, \forall i \in \{1, \dots, k\}$

Chceme testovat, jestli $p_1 = p_1^0, \dots, p_k = p_k^0$ pro jistý vektor pstí $p_1^0, \dots, p_k^0$

$H_0 : \forall i \in \{1, \dots, k\} : p_i = p_i^0$

$H_1 : \forall i \in \{1, \dots, k\} : p_i = p_i^0$

Testování statistiky

$$\chi^2 = \sum_{i=1}^{k} \frac{X_i - n \cdot p_i^0}{n \cdot p_i^0}$$

má za platnosti H_0 rozdělení χ^2_{k-1}. Hypotézu H_0 zamítáme, pokud $\chi^2 > q_{\chi^2_{k-1}}(1 - \alpha)$.

np_i^0 jsou teoretické četnosti

X_i jsou pozorované četnosti

Příklad. Chceme testovat, jestli kostka je symetrická/hází férově. Provedeme 600 hodů, jejichž souhrnné výsledky jsou v tabulce.

číslo	1	2	3	4	5	6	celkem
četnost	96	91	107	102	93	111	600
p_i	1/6	1/6	1/6	1/6	1/6	1/6	1
teoretické četnosti np_i	100	100	100	100	100	100	600
χ^2	16/100	81/100	49/100	4/100	49/100	121/100	

H_0 : kostka je férová (tj. $p_i = 1/6, \forall i \in \{1, \dots, 6\}$

$H_1 : \neg H_0$

$X_i \dots O_i$ observed

$np_i \dots E_i$ expected

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

$$\chi^2 = \sum_{i=1}^{k} \frac{(X_i - np_i)^2}{np_i} = \sum_{i=1}^{6} \frac{(X_i - 100)^2}{100} = \frac{320}{100} = 3.2 < 11.1 = q_{\chi^2_5(0.95)}$$

Nelze na hladině $\alpha = 0.05$ zamítnout hypotézu o tom, že je kostka symetrická.

Příklad. V jisté oblasti zkoumáme výskyt daného dravce. Oblast jsme rozdělili na nestejně velké části. Tabulka uvádí počty pozorovaných dravců v oblastech. Testujte, zda se dravec preferuje nějakou konkrétní oblast.

oblast	A	B	C	D	celkem
rozlohy	47 km^2	91 km^2	32 km^2	80 km^2	250 km^2
počet výskytů	4	7	5	12	25
p_i	47/250	91/250	32/250	80/250	
np_i	4.7	9.1	3.2	8	

$$\chi^2 = \sum_{i=1}^{4} \frac{(np_i - X_i)^2}{np_i} = \frac{1.7^2}{4.7} + \frac{2.1^2}{9.1} + \frac{0.8^2}{3.2} + \frac{3^2}{8} = 2.43$$

Na hladině $\alpha = 0.05$ nelze zamítnout hypotézu o tom, že výskyt dravce v regionu je rovnoměrný.