ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

Fakulta dopravní

Pravděpodobnost a matematická statistika

RNDr. Jana Novovičová, CSc.

1999

Vydavatelství ČVUT

Lektor : Doc. Ing. Miloslav Vošvrda, CSc.

Obsah

Kapitola 4

Náhodná veličina

Dosud jsme se zabývali v podstatě jen otázkou, zda uvažované náhodné jevy nastanou nebo nenastanou. V mnoha případech je však takový kvalitativní výrok nepostačující, a je nutné i kvantitativní vyšetření. Jinými slovy, k popisu hromadných náhodných jevů budeme obecně potřebovat také číselné údaje; přitom tyto číselné údaje nejsou konstantní, ale vykazují náhodné výchylky. Takovou náhodnou číselnou hodnotou je například počet aut, které vlastní náhodně vybraná pražská domácnost, zrovna tak jako množství spotřebované elektřiny za měsíc ve vybrané domácnosti. Obě tyto veličiny jsou numerické a jejich hodnota závisí na tom, která domácnost byla vybraná.

Můžeme říci, že výsledek náhodného pokusu, daný reálným číslem, je hodnotou veličiny, kterou nazveme **náhodná veličina**. Jinak řečeno, náhodná veličina je veličina, jejíž hodnota je jednoznačně určena výsledkem náhodného pokusu.

Rozlišujeme dva základní typy náhodných veličin: *diskrétní* a *spojité*. **Diskrétní** (čili **nespojitá**) náhodná veličina může nabývat pouze konečně nebo spočetně nekonečně mnoha hodnot. Počet aut, které vlastní domácnost, je příklad diskrétní veličiny. **Spojitá** náhodná veličina může nabývat všech hodnot z nějakého konečného nebo nekonečného intervalu. Množství elektřiny spotřebované za měsíc je příklad spojité náhodné veličiny.

4.1 Náhodná veličina a její rozdělení

Nyní uvedeme matematickou definici náhodné veličiny.

Definice 4.1 NÁHODNÁ VELIČINA

Náhodná veličina je každé zobrazení $X\colon \Omega \to \mathbb{R}$ takové, že pro každé $x \in \mathbb{R}$ je

$$A = \{\omega | X(\omega) \leq x\} \in \mathcal{A}.$$

Jestliže $\mathcal{A}$ je systém všech podmnožin Ω, pak každá reálná funkce X definovaná na Ω je náhodná veličina.

Náhodné veličiny budeme označovat velkými písmeny z konce abecedy, např. X, Y, Z nebo $X_1, X_2, \cdots$. Jejich konkrétní hodnoty pak malými písmeny x, y, z nebo $x_1, x_2, \cdots$. Počet členů domácnosti v souboru pražských domácností je náhodná veličina např. X, zatímco v určité náhodně vybrané třeba čtyřčlenné domácnosti jde už o konkrétní hodnotu této náhodné veličiny, o konkrétní počet členů této domácnosti, tudíž $X = 4$. Označení $[X = 4]$

bude vyjadřovat jev, že vybraná domácnost má 4 členy, zatímco označení $P(X = 4)$ je zjednodušené označení pro pravděpodobnost tohoto jevu.

Náhodnou veličinu považujeme za danou, známe-li všechny její možné hodnoty a pravděpodobnosti výskytu každé z nich. Pravidlo, které každé hodnotě nebo množině hodnot z každého intervalu přiřazuje pravděpodobnost, že náhodná veličina nabude této hodnoty nebo hodnoty z určitého intervalu, se nazývá **zákon rozdělení náhodné veličiny** nebo krátce **rozdělení náhodné veličiny**.

4.1.1 Distribuční funkce a hustota

Základní formou popisu zákona rozdělení je *distribuční funkce.* **Distribuční funkce** náhodné veličiny udává pravděpodobnost, že náhodná veličina X nabude hodnoty menší nebo rovné než zvolené x. Značíme ji $F(x)$.

Definice 4.2 DISTRIBUČNÍ FUNKCE

Distribuční funkce náhodné veličiny X je funkce $F\colon \mathbb{R} \to \langle 0, 1\rangle$ definovaná vztahem

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

□ **Základní vlastnosti distribučních funkcí**

1. $F(x)$ je neklesající funkce, tj. pro každou dvojici $x_1 < x_2$ platí

$$F(x_1) \leq F(x_2).$$

2. $F(x)$ je zprava spojitá, tj. pro libovolnou distribuční funkci platí

$$\lim_{h \to 0_+} F(x + h) = F(x).$$

3. Pro každou distribuční funkci platí

$$\lim_{x \to -\infty} F(x) = 0 \quad \text{a} \quad \lim_{x \to \infty} F(x) = 1,$$

zkráceně

$$F(-\infty) = 0 \quad \text{a} \quad F(\infty) = 1.$$

Jestliže možné hodnoty náhodné veličiny X patří do intervalu (a, b) pak

$$F(a) = 0, \quad F(b) = 1.$$

Každou funkci, která má všechny vlastnosti 1.–3. můžeme pokládat za distribuční funkci.
Poznámka: Definujeme-li distribuční funkci vztahem $F(x) = P(X < x)$ (tj. vynecháme znaménko (=)), pak F je zleva spojitá.
Často se používá i další vlastnost distribučních funkcí: nechť $x_1 < x_2$, potom platí

$$P(x_1 < X \leq x_2) = P([X \leq x_2] \cap [X > x_1]) = P([X \leq x_2]) - P([X \leq x_1]) = F(x_2) - F(x_1).$$

Distribuční funkce nemusí být spojitá, ale bodů nespojitosti může mít nanejvýš spočetně mnoho. Dva nejdůležitější typy distribučních funkcí, které mají největší uplatnění v matematické statistice, jsou *diskrétní* distribuční funkce a *absolutně spojité* distribuční funkce.

Diskrétní distribuční funkce

Distribuční funkce $F(x)$ se nazývá **diskrétní**, existuje-li konečná nebo spočetná posloupnost bodů $\{x_n\}$ a posloupnost nezáporných čísel $\{p_n\}$ splňujících podmínku $\sum_n p_n = 1$ taková, že

$$F(x) = \sum_{\{n:x_n \leq x\}} p_n, \quad \text{pro} \quad x \in \mathbb{R}. \tag{4.1}$$

Diskrétní distribuční funkce má schodovitý tvar se skoky velikosti p_n v bodech x_n. Má-li náhodná veličina X diskrétní distribuční funkci (**??**), tj. $p_n = P(X = x_n)$, říkáme, že X má **diskrétní rozdělení pravděpodobností**, stručně **diskrétní rozdělení**. Grafu diskrétní distribuční funkce odpovídá v popisné statistice graf kumulativních četností.
Diskrétní zákon rozdělení lze vedle distribuční funkce popsat i tzv. **pravděpodobnostní funkcí**

$$P(x) = P(X = x), \tag{4.2}$$

která každému x přiřazuje jeho pravděpodobnost $P(x)$. Tyto pravděpodobnosti $P(x)$ splňují podmínku $\sum_x P(x) = 1$.

Pomocí pravděpodobnostní funkce $P(x)$ můžeme stanovit s použitím pravidla o sčítání pravděpodobností pro neslučitelné jevy pravděpodobnost, že náhodná veličina nabude hodnoty z intervalu $\langle x_1, x_2 \rangle$. Tato pravděpodobnost je rovna součtu pravděpodobností hodnot z tohoto intervalu

$$P(x_1 \leq X \leq x_2) = \sum_{x=x_1}^{x_2} P(x). \tag{4.3}$$

Specifikace diskrétního rozdělení náhodné veličiny X pomocí pravděpodobností $P(x)$ a pomocí distribuční funkce je rovnocenná. Ze známých pravděpodobností $P(x)$ je možno odvodit distribuční funkci $F(x)$ a naopak, jak vyplývá z definice **??**.
Pravděpodobnostní funkci odpovídají v popisné statistice relativní četnosti.

Příklad 4.1 *Diskrétní náhodná veličina, distribuční funkce*
Házíme-li třikrát po sobě mincí, dostaneme osm stejně možných výsledků jak ukazuje následující tabulka **??**

Tabulka 4.1 *Možné výsledky při třech hodech mincí*

Pokus	Házení 3krát jednou mincí							
Možné výsledky ω	LLL	LLR	LRL	RLL	LRR	RRL	RLR	RRR

Nechť X udává celkový počet líců při třech hodech jednou mincí. Pak X je náhodná veličina, která může nabývat hodnot 0, 1, 2 a 3.

a) Vyjádřete pomocí náhodné veličiny jev, že padly právě dva líce. Určete $P(X = 2)$, tj. pravděpodobnost, že padnou právě dva líce.

b) Najděte rozdělení náhodné veličiny X.

c) Vyjádřete pomocí náhodné veličiny jev, že padnou nejvýše dva líce. Vypočítejte $P(X \leq 2)$, tj. pravděpodobnost, že padnou nejvýše dva líce.

d) Určete distribuční funkci náhodné veličiny X.

e) Vyjádřete pomocí náhodné veličiny jev, že počet líců, které padnou, je nejvýše roven třem a větší než jedna. Vypočítejte $P(1 < X \leq 3)$.

Řešení:

a) Jev, že padnou právě dva líce lze vyjádřit $[X = 2]$. $P(X = 2)$ je pravděpodobnost, že padnou právě dva líce. Z tabulky ?? vidíme, že jsou tři způsoby jak dostat celkově dva líce a že je celkem osm možných výsledků. Tudíž podle klasického pravidla výpočtu pravděpodobností dostaneme

$$P(X = 2) = \frac{3}{8} = 0.375.$$

b) Zbývající pravděpodobnosti pro X jsou vypočítány stejným způsobem a jsou uvedeny v následující tabulce ??.

Tabulka 4.2 *Rozdělení veličiny X udávající počet líců při třech hodech mincí.*

Počet líců x	0	1	2	3
Pravděpodobnost $P(X = x)$	0.125	0.375	0.375	0.125

c) Jev $[X \leq 2]$, že padnou nejvýše dva líce lze vyjádřit jako

$$[X \leq 2] = ([X = 0] \cup [X = 1] \cup [X = 2]).$$

Protože tři jevy na pravé straně rovnice jsou vzájemně neslučitelné, dostaneme aplikací pravidla pro sčítání pravděpodobností a z tabulky ??

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0.125 + 0.375 + 0.375 = 0.875$$

Tudíž pravděpodobnost, že padnou nejvýše dva líce je rovna 0.875.

d) Distribuční funkci $F(x)$ vypočteme podle vzorce

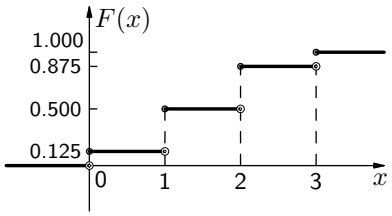
$$F(x) = \sum_{n=0}^{x} P(X = n) \text{ pro } x = 0, 1, 2, 3.$$

Hodnoty $F(x)$ jsou uvedeny v tabulce ??. a její graf na obrázku ??

Tabulka 4.3 *Distribuční funkce rozdělení počtu líců při 3 hodech mincí*

Počet líců x	0	1	2	3
Distribuční funfce $F(x)$	0.125	0.500	0.875	1.000

Obrázek 4.1 *Graf distribuční funkce*



Distribuční funkce má schodovitý tvar se skoky velikosti 0.375 v bodech $x = 1$ a $x = 2$ a se skoky velikosti 0.125 v bodech $x = 0$ a $x = 3$.

e) Jev, že padnou nejvýše tři líce a více než 1 líc může být vyjádřen jako

$$[1 < X \leq 3] = ([X \leq 3] \cap [X > 1]) = ([X \leq 3] - [X \leq 1]).$$

Protože, platí $[X \leq 1] \subset [X \leq 3]$ použijeme vlastnost 2. pravděpodobnosti (viz kapitola **??**) k výpočtu $P(1 < X \leq 3)$:

$$P(1 < X \leq 3) = P(X \leq 3) - P(X \leq 1) = 1.000 - 0.500 = 0.500.$$

Tudíž pravděpodopbnost, že padnou nejvýše tři líce a více než jeden líc je rovna 0.5. ■

Absolutně spojitá distribuční funkce

Zvláštní pozornost zasluhují distribuční funkce, které jsou nejen spojité, ale dokonce absolutně spojité. Distribuční funkce F se nazývá **absolutně spojitá**, jestliže existuje nezáporná funkce $f(x)$ taková, že platí

$$F(x) = \int_{-\infty}^{x} f(u)\,\mathrm{d}u \quad \text{pro každé } x \in \mathbb{R}. \tag{4.4}$$

Funkce $f(x)$ se nazývá **hustota rozdělení pravděpodobností**, definovaného distribuční funkcí $F(x)$, stručně hustota pravděpodobnosti nebo jen **hustota**. Má-li náhodná veličina X absolutně spojitou distribuční funkci, říkáme, že má **spojité rozdělení pravděpodobností**, zkráceně **spojité rozdělení**.
Hustota $f(x)$ splňuje rovnost

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1. \tag{4.5}$$

Existuje-li derivace F' distribuční funkce v bodě x, je $F'(x) = f(x)$. Tato hustota pravděpodobnosti je definována jako

$$f(x) = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{P(x < X \leq x + \Delta x)}{\Delta x},$$

tj. jako limita pravděpodobnosti, že veličina X padne do velmi malého intervalu $(x, x + \Delta x)$, vydělená délkou tohoto intervalu v případě, že se tato délka Δx blíží nule. Součin $\Delta x f(x)$ pak přibližně vyjadřuje pravděpodobnost, že náhodná veličina X padne do velmi malého intervalu $(x, x + \Delta x)$, a to tím přesněji, čím je Δx menší.
Pro $a, b \in \mathbb{R}, a < b$ platí

$$P(a < X \leq b) = \int_{a}^{b} f(x)\,\mathrm{d}x = F(b) - F(a).$$

Pravděpodobnost je tedy plocha pod křivkou hustoty. Odtud plyne, že pro náhodnou veličinu se spojitým rozdělením je $P(X = a) = 0$ pro libovolné $a \in \mathbb{R}$.

Příklad 4.2 *Distribuční funkce a hustota pravděpodobnosti spojitého rozdělení*
Funkce $F(x) = 1 - e^{-\lambda x}$ pro $x > 0$ a $F(x) = 0$ pro $x \leq 0$, kde $\lambda > 0$ je konstanta, splňuje základní vlastnosti 1. – 3. distribuční funkce a je distribuční funkcí nějaké náhodné veličiny X se spojitým rozdělením. Odpovídající hustota je $f(x) = \lambda e^{-\lambda x}$ pro $x > 0$ a $f(x) = 0$ pro $x \leq 0$.
$P(1 < X \leq 2) = \lambda \int_1^2 e^{-\lambda x} dx = 1 - e^{-2\lambda} - 1 + e^{-\lambda} = e^{-\lambda}(1 - e^{-\lambda})$. ■

4.1.2 Vícerozměrná rozdělení pravděpodobností

Často se neomezujeme pouze na jednu náhodnou veličinu, ale zkoumáme celý systém náhodných veličin, tak zvanou *vícerozměrnou* přesněji *n-rozměrnou náhodnou veličinu*.
Vícerozměrnou náhodnou veličinou $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ budeme nazývat n-rozměrný vektor, jehož všechny složky X_i jsou náhodné veličiny. Pro vícerozměrnou náhodnou veličinu se také používá název **náhodný vektor**. Nadále budeme podle potřeby používat obou názvů.

Všimneme si podrobněji dvourozměrné náhodné veličiny (X, Y). Zákon rozdělení této náhodné veličiny může být dán ve formě **sdružené (simultánní) distribuční funkce** $F(x, y)$, která je definovaná jako pravděpodobnost, že náhodná veličina X, nabude hodnoty menší než x a současně náhodná veličina Y nabude hodnoty menší než y.

Definice 4.3 SDRUŽENÁ DISTRIBUČNÍ FUNKCE NÁHODNÉHO VEKTORU (X, Y)

Sdružená distribuční funkce náhodného vektoru (X, Y) je funkce definovaná vztahem

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

pro každé $x \in \mathbb{R}$, $y \in \mathbb{R}$.

□ **Základní vlastnosti distribuční funkce $F(x, y)$**

1. $F(x, y)$ je neklesající v každé své proměnné.
2. $\lim_{x,y\to\infty} F(x, y) = 1$.
3. $\lim_{x\to-\infty} F(x, y) = 0, \quad \lim_{y\to-\infty} F(x, y) = 0$.
4. $F(x, y)$ je zprava spojitá v každé proměnné.

Kromě těchto triviálních vlastností má každá dvourozměrná distribuční funkce jednu další charakterizující vlastnost, kterou je možné vyjádřit ve tvaru

$$P(x_1 < X \leq x_2, y_1 < Y \leq y_2) = F(x_1, y_1) - F(x_1, y_2) - F(x_2, y_1) + F(x_2, y_2)$$

pro každé $x_1 < x_2,\ y_1 < y_2$.

Sdružená distribuční funkce $F(x, y)$ se nazývá **diskrétní**, jestliže

$$F(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} P(X = x_i, Y = y_j), \tag{4.6}$$

kde $\{x_i\}$ respektive $\{y_j\}$ jsou konečné nebo spočetné posloupnosti všech hodnot, kterých nabývá X respektive Y. Pravděpodobnosti $P(X = x_i, Y = y_j)$ se nazývají **sdružené pravděpodobnosti** a platí

$$\sum_{x_i} \sum_{y_j} P(X = x_i, Y = y_j) = 1.$$

Náhodný vektor (X, Y) s diskrétní distribuční funkcí má **diskrétní sdružené rozdělení** (diskrétní rozdělení). Součty sdružených pravděpodobností

$$P_X(x_i) = \sum_{y_j} P(X = x_i, Y = y_j) \text{ resp. } P_Y(y_j) = \sum_{x_i} P(X = x_i, Y = y_j)$$

se nazývají **marginální pravděpodobnosti** náhodné veličiny X respektive Y a vyjadřují pravděpodobnosti různých hodnot jedné z veličin bez ohledu na hodnotu veličiny druhé. Zákon rozdělení, který popisují, se nazývá **marginální zákon rozdělení**.

Omezíme-li se na dvě diskrétní náhodné veličiny X a Y, můžeme pravděpodobnosti současného výskytu různých kombinací dvojic hodnot $(x_i, y_j),\ i = 1, 2, \cdots, r,\ j = 1, 2, \cdots, s$ obou veličin uspořádat do dvourozměrné **kombinační tabulky??**.

Tabulka 4.4 *Kombinační tabulka*

$X \setminus Y$	y_1	$\cdots$	y_j	$\cdots$	y_s	$P_X(x_i)$
x_1	$P(x_1, y_1)$	$\cdots$	$P(x_1, y_j)$	$\cdots$	$P(x_1, y_s)$	$P_X(x_1)$
$\cdot$						
x_i	$P(x_i, y_1)$	$\cdots$	$P(x_i, y_j)$	$\cdots$	$P(x_i, y_s)$	$P_X(x_i)$
$\cdot$						
x_r	$P(x_r, y_1)$	$\cdots$	$P(x_r, y_j)$	$\cdots$	$P(x_r, y_s)$	$P_X(x_r)$
$P_Y(y_j)$	$P_Y(y_1)$	$\cdots$	$P_Y(y_j)$	$\cdots$	$P_Y(y_s)$	1

Distribuční funkce $F(x, y)$ se nazývá **absolutně spojitá**, jestliže existuje nezáporná funkce $f(x, y)$ nazývaná **sdružená hustota pravděpodobnosti** taková, že

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\, \mathrm{d}u\mathrm{d}v. \tag{4.7}$$

Hustota sdruženého rozdělení má tyto základní vlastnosti:

1. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, \mathrm{d}x\, \mathrm{d}y = 1.$
2. $\dfrac{\partial^2 F(x, y)}{\partial x \partial y} = f(x, y)$ pokud derivace funkce F existuje.
3. $P(x_1 < X \le x_2, y_1 < Y \le y_2) = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f(x, y)\, \mathrm{d}x\, \mathrm{d}y \quad \text{pro } x_1 < x_2,\ y_1 < y_2.$

Náhodný vektor (X, Y) s absolutně spojitou distribuční funkcí má **spojité sdružené rozdělení**. Z distribuční funkce $F(x, y)$ můžeme odvodit **marginální distribuční funkce** náhodné veličiny X respektive Y

$$F_X(x) = P(X \le x) = \lim_{y \to \infty} F(x, y), \ \text{ resp. } \ F_Y(y) = P(Y \le y) = \lim_{x \to \infty} F(x, y). \tag{4.8}$$

Podobně z hustoty pravděpodobnosti $f(x, y)$ můžeme odvodit **marginální hustoty rozdělení pravděpodobností** náhodné veličiny X respektive Y

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, \mathrm{d}y, \ \text{ resp. } \ f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, \mathrm{d}x. \tag{4.9}$$

4.1.3 Nezávislost náhodných veličin

Budeme říkat, že náhodné veličiny X a Y jsou **nezávislé**, jestliže pro všechna x,$y \in \mathbb{R}$ platí

$$P(X \le x, Y \le y) = P(X \le x) P(Y \le y),$$

tj. jestliže se dvourozměrná distribuční funkce náhodných veličin X a Y rovná součinu distribučních funkcí náhodné veličiny X a náhodné veličiny Y. Pro diskrétní rozdělení to znamená totéž jako

$$P(X = x_i, Y = y_j) = P_X(x_i)P_Y(y_j), \qquad i = 1, 2, \cdots, r, \quad j = 1, 2, \cdots, s$$

a pro rozdělení s hustotou $f(x, y)$

$$f(x, y) = f_X(x)f_Y(y)$$

pro všechna $x, y \in \mathbb{R}$.

Nezávislost více náhodných veličin je možno definovat obdobně. Náhodné veličiny $X_1, X_2, \cdots, X_n$ jsou **nezávislé**, jestliže pro každou n-tici $x_1, x_2, \cdots, x_n$ reálných čísel platí

$$P(X_1 \leq x_1, \cdots, X_n \leq x_n) = \prod_{i=1}^{n} P(X_i \leq x_i).$$

Pro nezávislé náhodné veličiny platí:

1. Jestliže $X_1, X_2, \cdots, X_n$ jsou nezávislé náhodné veličiny, a $h_k(x),\ k = 1, 2, \cdots, n$ funkce reálné proměnné, pak náhodné veličiny $Y_k = h_k(X),\ k = 1, 2, \cdots, n$ jsou také nezávislé.

2. Jestliže náhodné veličiny $X_1, X_2, \cdots, X_n$ jsou nezávislé, a každá z nich má hustotu, pak platí

$$f(x_1, \cdots, x_n) = \prod_{i=1}^{n} f_i(x_i), \tag{4.10}$$

kde $f_i(x_i)$ je hustota náhodné veličiny $X_i,\ i = 1, 2, \cdots, n$ a $f(x_1, \cdots, x_n)$ je hustota n-rozměrné náhodné veličiny $(X_1, X_2, \cdots, X_n)$. Ze vztahu (**??**) plyne naopak nezávislost náhodných veličin $X_1, X_2, \cdots, X_n$.

4.2 Charakteristiky náhodných veličin

Distribuční funkce podává o náhodné veličině úplnou informaci. Známe-li tuto funkci, víme jakých hodnot může uvažovaná náhodná veličina nabývat a jaké jsou pravděpodobnosti jednotlivých hodnot. V praxi často potřebujeme koncentrovanější a přehlednější vyjádření této informace. K tomu používáme podobně jako v popisné statistice, číselné hodnoty, které nazýváme **charakteristiky náhodných veličin**. Nejčastěji používanými charakteristikami jsou *střední hodnota*, která popisuje polohu (úroveň) náhodné veličiny, a *rozptyl* který popisuje variabilitu (rozptýlenost) náhodné veličiny. Stručně se zmíníme i o dalších charakteristikách.

4.2.1 Střední hodnota

Nechť X je náhodná veličina s distribuční funkcí $F(x)$. Pak máme následující definice střední hodnoty náhodné veličiny X s diskrétním respektive spojitým rozdělením. Budeme ji značit $E(X)$.

Definice 4.4 STŘEDNÍ HODNOTA NÁHODNÉ VELIČINY

Střední hodnota náhodné veličiny X s diskrétním rozdělením daným pravděpodobnostní funkcí $P(x)$ je definována vztahem

$$E(X) = \sum_x xP(x).$$

Střední hodnota náhodné veličiny se spojitým rozdělením s hustotou $f(x)$ je definována vztahem

$$E(X) = \int_{-\infty}^{\infty} xf(x)\,\mathrm{d}x.$$

V diskrétním případě jde v podstatě o jakýsi vážený průměr možných hodnot veličiny X s vahami odpovídajícími jednotlivým pravděpodobnostem. Ve spojitém případě je střední hodnota náhodné veličiny X definována obdobně (součet je nahrazen integrálem).
Poznámka: V dalším textu budeme označovat střední hodnotu náhodné veličiny X také symbolem μ_x.

Střední hodnota se někdy nazývá první obecný moment. Obecně, **k-tý obecný moment** $E(X^k)$ náhodné veličiny X je definován jako

$$E(X^k) = \begin{cases} \sum_x x^k P(x) & \text{pro diskrétní rozdělení} \\ \\ \int_{-\infty}^{\infty} x^k f(x)\,\mathrm{d}x & \text{pro spojité rozdělení.} \end{cases}$$

Pro práci se středními hodnotami jsou důležité některé její matematické vlastnosti, které uvedeme.

Základní vlastnosti střední hodnoty

1. Střední hodnota konstanty je rovna konstantě, $E(c) = c$.
2. Střední hodnota součinu konstanty a náhodné veličiny je rovna součinu této konstanty a střední hodnoty dané veličiny, $E(cX) = cE(X)$.
3. Střední hodnota součtu n náhodných veličin je rovna součtu jejich středních hodnot,

$$E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i).$$

Pojem střední hodnoty zobecníme na nějakou funkci $h(X)$ náhodné veličiny X

$$E(h(X)) = \sum_j h(x_j)P(x_j), \qquad \text{resp.} \qquad E(h(X)) = \int_{-\infty}^{\infty} h(x)f(x)\,\mathrm{d}x.$$

4.2.2 Rozptyl

Rozptyl je mírou variability náhodné veličiny.

Definice 4.5 ROZPTYL NÁHODNÉ VELIČINY

Rozptyl náhodné veličiny s diskrétním rozdělením s pravděpodobnostní funkcí $P(x)$ je definován vztahem

$$D(X) = \sum_x (x - E(X))^2 P(x).$$

Rozptyl náhodné veličiny se spojitým rozdělením s hustotou $f(x)$ je definován vztahem

$$D(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x)\,\mathrm{d}x.$$

Rozptyl se také nazývá druhý centrální moment.
Obecně, **k-tý centrální moment** $E(X - \mu_x)^k$ náhodné veličiny X je definován jako

$$E((X-\mu_x)^k) = \begin{cases} \sum_x (x-\mu_x)^k P(x) & \text{pro diskrétní rozdělení} \\ \int_{-\infty}^{\infty} (x-\mu_x)^k f(x)\,\mathrm{d}x & \text{pro spojité rozdělení.} \end{cases}$$

Rozptyl lze počítat podle vzorce

$$D(X) = E(X - E(X))^2 = E(X^2 - 2XE(X) + (E(X))^2) = E(X^2) - [E(X)]^2. \qquad (4.11)$$

Poznámka: V dalším textu budeme označovat rozptyl náhodné veličiny X také symbolem σ_x.

Měrné jednotky, ve kterých je vyjádřen rozptyl $D(X)$ jsou čtverce jednotek náhodné veličiny X. V původních jednotkách měří variabilitu odmocnina rozptylu, kterou nazýváme **směrodatnou odchylkou** a značíme $\sigma_x = \sqrt{D(X)}$.

Základní vlastnosti rozptylu

1. Rozptyl konstanty je rovna nule, $D(c) = 0$.
2. Rozptyl součinu konstanty a náhodné veličiny je roven součinu čtverce této konstanty a rozptylu dané veličiny, $D(cX) = c^2 D(X)$.
3. Rozptyl součtu *nezávislých* náhodných veličin je roven součtu rozptylů těchto náhodných veličin,
$$D(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} D(X_i).$$

4.2.3 Kvantily

Vedle uvedených charakteristik náhodné veličiny se při popisu spojité náhodné veličiny velmi často používají *kvantily*. S tímto pojmem jsme se již seznámili v popisné statistice v části **??**. Nyní tuto charakteristiku uvedeme do souvislosti se spojitou náhodnou veličinou.

Definice 4.6 KVANTIL

Nechť X je náhodná veličina s distribuční funkcí $F(x)$ a hustotou pravděpodobnosti $f(x)$. *p-kvantilem náhodné veličiny* X nebo $100p$ *procentním kvantilem* je číslo Q_p, pro které platí

$$P(X \leq Q_p) = F(Q_p) = \int_{-\infty}^{Q_p} f(x)\,\mathrm{d}x = p, \quad 0 < p < 1.$$

50% kvantil nazýváme **medián**. Medián $Q_{0.5}$ náhodné veličiny je jednoznačně určen podmínkou $F(Q_{0.5}) = \frac{1}{2}$.

Příklad 4.3 *Střední hodnota a rozptyl diskrétního rozdělení*
Určete $E(X)$ a $D(X)$ náhodné veličiny, která nabývá hodnot z množiny $\{0, 1\}$ s pravděpodobnostní funkcí $P(X = 1) = p, \quad P(X = 0) = 1 - p, \quad 0 < p < 1$.
Řešení: $E(X) = 1p + 0(1 - p) = p$ a $D(X) = (1 - p)^2 p + (0 - p)^2(1 - p) = p(1 - p)$ ■

Příklad 4.4 *Střední hodnota, rozptyl a medián spojitého rozdělení*
Uvažujme náhodnou veličinu z příkladu **??**. Určete střední hodnotu, rozptyl a medián této veličiny.
Řešení: K výpočtu použijeme gama funkci :
$\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}dx, \ a > 0, \ \Gamma(a+1) = a\Gamma(a), \ \Gamma(1) = 1.$
$E(X) = \lambda \int_0^\infty xe^{-\lambda x}dx = \frac{1}{\lambda}\int_0^\infty ue^{-u}du = \frac{\Gamma(2)}{\lambda} = \frac{1}{\lambda}.$
Rozptyl vypočítáme pomocí vzorce (**??**), tudíž musíme spočítat $E(X^2)$.
$E(X^2) = \lambda \int_0^\infty x^2e^{-\lambda x}dx = \frac{1}{\lambda^2}\int_0^\infty u^2e^{-u}du = \frac{\Gamma(3)}{\lambda^2} = \frac{2}{\lambda^2}. \quad D(X) = \frac{2}{\lambda^2} - (\frac{1}{\lambda})^2 = \frac{1}{\lambda^2}.$
Medián $Q_{0.5}$ se nalezne řešením rovnice $1 - e^{-\lambda Q_{0.5}} = 0.5$, z níž dostaneme $Q_{0.5} = \frac{1}{\lambda}\ln 2$. ■

4.2.4 Kovariance a korelace

Kovariance a korelační koeficient (koeficient korelace) patří mezi nejčastěji používané charakteristiky sdruženého rozdělení dvou náhodných veličin. *Kovariance* je střední hodnota součinu odchylek obou náhodných veličin X a Y od jejich středních hodnot.

Definice 4.7 KOVARIANCE

Kovariance σ_{xy} dvou náhodných veličin X a Y se středními hodnotami μ_x a μ_y je definována vztahem

$$\sigma_{xy} = E(X - \mu_x)(Y - \mu_y).$$

K výpočtu kovariance veličin X a Y lze použít střední hodnotu $E(XY)$ nazývanou **smíšený obecný moment** a definovou vztahem :

$$E(XY) = \begin{cases} \sum\limits_{x,y} xyP(X = x, Y = y) & \text{pro diskrétní rozdělení} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x, y)\,\mathrm{d}x\mathrm{d}y & \text{pro spojitá rozdělení.} \end{cases} \tag{4.12}$$

Z definice **??** a z (**??**) plyne, že

$$\sigma_{xy} = E(XY) - \mu_x \mu_y. \tag{4.13}$$

Z definice nezávislých náhodných veličin a ze vztahu (**??**) plyne, že pro nezávislé náhodné veličiny platí $E(XY) = E(X)E(Y)$. Kovariance dvou nezávislých náhodných veličin je tudíž rovna nule.

Pomocí kovariance můžeme výjádřit rozptyl součtu dvou náhodných veličin X a Y. Je roven součtu rozptylů obou náhodných veličin a dvojnásobku kovariance obou veličin.

$$\begin{aligned} D(X+Y) &= E(X+Y-\mu_x-\mu_y)^2 = E(X-\mu_x)^2 + E(Y-\mu_y)^2 + 2E(X-\mu_x)(Y-\mu_y) \\ &= D(X) + D(Y) + 2\sigma_{xy}. \end{aligned} \tag{4.14}$$

Korelační koeficient dává určitou informaci o stupni závislosti dvou náhodných veličin. Je definován jako poměr kovariance k součinu směrodatných odchylek obou náhodných veličin.

Definice 4.8 KORELAČNÍ KOEFICIENT

Korelační koeficient ρ_{xy} dvou náhodných veličin X a Y s rozptyly $\sigma_x^2 > 0$ a $\sigma_y^2 > 0$ je definován vztahem

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}.$$

Je-li $\sigma_x^2 = 0$ nebo $\sigma_y^2 = 0$ pokládáme $\rho_{xy} = 0$.

Pro korelační koeficient platí:

1. Hodnota korelačního koeficientu je číslo z intervalu $\langle -1, 1\rangle$, tj. $-1 \le \rho_{xy} \le 1$.
2. Jsou-li X a Y nezávislé, je $\rho_{xy} = 0$.
   Poznámka: Opačné tvrzení neplatí. Ze vztahu $\rho_{xy} = 0$ obecně nevyplývá, že veličiny X a Y jsou nezávislé. Je-li $\rho_{xy} = 0$, říkáme, že náhodné veličiny X a Y jsou **nekorelované**.
3. $|\rho_{xy}| = 1$ právě tehdy, když s pravděpodobností 1 platí $Y = a + bX$, kde $a, b,\ b \neq 0$ jsou reálné konstanty. Přitom je $\rho_{xy} = 1$ nebo -1 podle toho, je-li $b > 0$ nebo $b < 0$.

S interpretací a výpočtem korelačního koeficientu se podrobněji seznámíme v kapitole o regresi a korelaci.

4.2.5 Vektor středních hodnot, kovarianční matice

Z charakteristik n-rozměrného náhodného vektoru $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ jsou nejdůležitější střední hodnoty jednotlivých veličin X_i

$$\mu_i = E(X_i), \quad i = 1, 2, \cdots, n,$$

dále jejich rozptyly

$$\sigma_i^2 = D(X_i), \quad i = 1, 2, \cdots, n$$

a konečně kovariance dvojic veličin

$$\sigma_{ij} = E(X_i - \mu_i)(X_j - \mu_j), \quad i = 1, 2, \cdots, n; \ \ i \neq j.$$

Střední hodnoty zapisujeme často ve formě **vektoru středních hodnot**

$$\mu = (\mu_1, \mu_2, \cdots, \mu_n)^T$$

a kovariance spolu s rozptyly ve formě **kovarianční matice**

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \dots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \dots & \sigma_n^2 \end{pmatrix}.$$

Kovarianční matice je symetrická a positivně definitní.

4.3 Některá rozdělení pravděpodobností

Rozdělení jednorozměrných i vícerozměrných náhodných veličin se používají jako pravděpodobnostní modely při popisu konkrétních praktických problémů. V této části se seznámíme s nejčastěji používanými pravděpodobnostními rozděleními.

4.3.1 Diskrétní rozdělení

□ **Alternativní rozdělení $\mathcal{A}(p)$**

Rozdělení pravděpodobností na $\Omega = \{0, 1\}$ s pravděpodobnostní funkcí

$$P(x) = p^x(1-p)^{1-x}, \tag{4.15}$$

kde $p \in (0, 1)$ se nazývá alternativní rozdělení s parametrem p.
Střední hodnota tohoto rozdělení je $E(X) = p$ a rozptyl $D(X) = p(1-p)$.
Interpretace: Uvažujme náhodný pokus. Nastane-li sledovaný náhodný jev A, nabude náhodná veličina X hodnoty $x = 1$, nenastane-li tento jev A, nabude náhodná veličina X hodnoty $x = 0$. Náhodná veličina X tedy vyjadřuje, kolikrát jev A v pokusu nastane.

□ **Binomické rozdělení $\mathcal{B}(n, p)$**

Rozdělení pravděpodobností na $\Omega = \{0, 1, ..., n\}$ s pravděpodobnostní funkcí

$$P(x) = \begin{pmatrix} n \\ x \end{pmatrix} p^x(1-p)^{n-x} \tag{4.16}$$

pro $p \in (0, 1)$ a $n \in \mathbb{N}_+$ se nazývá binomické rozdělení s parametry n a p.

Střední hodnota je $E(X) = np$ a rozptyl $D(X) = np(1-p)$.

Binomické rozdělení je obecně nesymetrické. S růstem n ($n \to \infty$) nebo přibližováním p k hodnotě 0.5 se stává postupně symetričtějším. Pro $p = 0.5$ je symetrické. Pro $n = 1$ dostaneme $\mathcal{A}(p)$-rozdělení.
Interpretace: Předpokládejme, že provádíme n nezávislých pokusů, při nichž může nastat jev A s pravděpodobností p a nenastat s pravděpodobností $q = 1 - p$. Pravděpodobnost, že se v takové sérii pokusů objeví jev A právě x-krát, je dána výrazem (**??**).

Pravděpodobnosti jednotlivých hodnot náhodné veličiny s binomickým rozdělením jsou obecným členem binomického rozvoje

$$(p+q)^n = \sum_{x=1}^{n} \left(\begin{array}{c} n \\ x \end{array} \right) p^x (1-p)^{n-x}.$$

Hypergeometrické rozdělení $\mathcal{H}g(N, M, n)$

Rozdělení pravděpodobností s $\Omega = \{0, 1, ..., min\{M, n\}\}$ a pravděpodobnostní funkcí

$$P(x) = \frac{\left(\begin{array}{c} M \\ x \end{array} \right) \left(\begin{array}{c} N-M \\ n-x \end{array} \right)}{\left(\begin{array}{c} N \\ n \end{array} \right)}, \quad \max(n-N+M,\ 0) \leq x \leq \min(M,\ n) \tag{4.17}$$

se nazývá hypergeometrické rozdělení s parametry N, M, n.
Střední hodnota je $E(X) = n\frac{M}{N}$, a rozptyl $D(X) = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\left(\frac{N-n}{N-1}\right)$.
Interpretace: Uvažujme situaci, kdy v souboru N prvků je jich M $(N \geq M)$ s určitou vlastností a zbylých $N - M$ tuto vlastnost nemá. Postupně vybereme ze souboru n prvků, z nichž žádný nevracíme zpět. Počet prvků se sledovanou vlastností mezi n vybranými prvky je náhodná veličina X mající hypergeometrické rozdělení.

Jestliže N je velké a n a $\frac{M}{N}$ se nemění, blíží se hypergeometrické rozdělení binomickému. To znamená, že můžeme pro velká N zanedbat rozdíl mezi výběrem bez vracení a s vracením. Prakticky postupujeme tak, že vypočítáme poměr $\frac{n}{N}$ a je-li tento poměr větší než 0.05, lze hypergeometrické rozdělení nahradit rozdělením binomickým s parametry n a $\frac{M}{N}$.
Aplikace: Hypergeometrické rozdělení se vyskytuje například ve statistické kontrole jakosti v případech, kdy zkoumáme jakost malého počtu výrobků nebo když kontrola má charakter destrukční zkoušky, tj. výrobek je při zkoušce zničen. Dále jako pravděpodobnostní model některých her jako Sportky.

Geometrické rozdělení $\mathcal{G}(p)$

Rozdělení pravděpodobností na $\mathbb{N}_+$ s pravděpodobnostní funkcí

$$P(x) = p(1-p)^{x-1} = pq^{x-1} \tag{4.18}$$

pro $p \in (0, 1)$ se nazývá geometrické rozdělení s parametrem p.
Střední hodnotu vypočítáme:

$$E(X) = \sum_{x=1}^{\infty} xpq^{x-1} = p\sum_{x=1}^{\infty} xq^{x-1} = p\sum_{x=1}^{\infty} \frac{dq^x}{dq} = p\frac{d}{dq}\sum_{x=0}^{\infty} q^x = p\frac{d}{dq}\frac{1}{1-q} = \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}.$$

Rozptyl tohoto rozdělení je $D(X) = \frac{1-p}{p}$. Medián leží mezi 0 a 1 pro $p < 0.5$ a je roven nule pro $p \geq 0.5$.
Interpretace: Provádějme pokus se dvěma možnými výsledky, které nazveme „úspěch“ a „neúspěch“. Pravděpodobnost úspěchu nechť je p. Počet nezávislých opakování pokusů do prvního úspěchu je náhodná veličina, která má geometrické rozdělení. $P(x)$ udává pravděpodobnost, že prvních x pokusů bude neúspěšných a že k úspěchu dojde teprve v $(x+1)$-ním pokusu.

Příklad 4.5 *Geometrické rozdělení*

Mezi N výrobky je M vadných. Provádíme výběr s vracením. Nechť X značí náhodnou veličinu, že prvních x výrobků bude dobrých a v $(x+1)$-ním tahu jsme vytáhli vadný výrobek. Pak má náhodná veličina X geometrické rozdělení s parametrem $p = \frac{M}{N}$. ■

Poissonovo rozdělení $\mathcal{P}(\lambda)$

Rozdělení pravděpodobností na $\mathbb{N}$ *s pravděpodobnostní funkcí*

$$p(x) = e^{-\lambda}\frac{\lambda^x}{x!}, \tag{4.19}$$

kde $\lambda > 0$ *je konstanta, se nazývá Poissonovo rozdělení s parametrem* λ.

Střední hodnotu vypočítáme následujícím způsobem:

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

Rozptyl $D(X) = \lambda$.

Jestliže je počet pokusů n dosti velký (prakticky stačí $n > 30$) a $p \to 0$ (prakticky $p \leq 0.01$), pak lze binomické rozdělení aproximovat Poissonovým rozdělením s parametrem $\lambda = np$.
Aplikace: Toto rozdělení pravděpodobností se často užívá k modelování četností s jakou určitá událost nastane během určitého časového úseku. Na příklad počet telefonních volání v určitém časovém intervalu, počet zákazníků obsloužených za jednotku času u pokladny v obchodě, počet poruch nějakého zařízení za časovou jednotku, počet vad na výrobku.

Příklad 4.6 *Poissonovo rozdělení*

Předpokládejte, že počet telefonických hovorů došlých během 1 hodiny na ústřednu v jedné malé firmě, má Poissonovo rozdělení s parametrem $\lambda = 5.2$. Vypočítejte pravděpodobnost, že během jedné hodiny přijdou na ústřednu *a)* právě dva hovory; *b)* nejvýše šest a nejméně 4 hovory;
c) aspoň jeden hovor. *d)* Jaký je průměrný počet hovorů za jednu hodinu?
Řešení:
a) Protože $\lambda = 5.2$ je podle (**??**) $P(X = 2) = e^{-5.2}\frac{(5.2)^2}{2!} = 0.0746$.
b) $P(4 < X \leq 6) = P(X \leq 6) - P(X \leq 4) = 0.7323 - 0.4060 = 0.3263$.
c) $P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-5.2} = 0.994$.
d) Průměrný počet hovorů za jednu hodinu je roven střední hodnotě Poissonova rozdělení s parametrem $\lambda = 5.2$, tudíž je roven 5.2.

■

Diskrétní rovnoměrné rozdělení $D\mathcal{U}(m)$

Rozdělení pravděpodobností na $\mathbb{N}_m$, *kde* $m \in \mathbb{N}_+$, *s pravděpodobnostní funkcí*

$$p(x) = \frac{1}{m}, \tag{4.20}$$

se nazývá diskrétní rovnoměrné rozdělení nebo $D\mathcal{U}(m)$*-rozdělení.*

Distribuční funkce

$$F(x) = \begin{cases} 0 & \text{pro } x < 1 \\ \frac{x}{m} & \text{pro } 1 \leq x < m \\ 1 & \text{pro } x \geq m. \end{cases}$$

Střední hodnota $E(X) = \frac{m+1}{2}$, rozptyl $D(X) = \frac{m^2-1}{12}$, medián $Q_{0.5} = [\frac{m}{2}] + 1$ pro m liché a $Q_{0.5} = [\frac{m+1}{2}]$ pro m sudé.

4.3.2 Spojitá rozdělení

V dalším výkladu se zaměříme na některá spojitá rozdělení.

Rovnoměrné rozdělení $\mathcal{U}(a,b)$

Rovnoměrné rozdělení na reálném intervalu (a,b) má hustotu

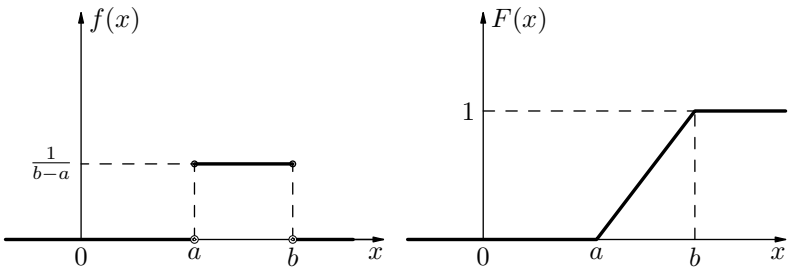
$$f(x) = \begin{cases} 0 & \text{pro } x < a \text{ a pro } b < x \\ \frac{1}{b-a} & \text{pro } a < x < b. \end{cases} \tag{4.21}$$

Pro příslušnou distribuční funkci platí

$$F(x) = \begin{cases} 0 & \text{pro } x < a \\ \frac{x-a}{b-a} & \text{pro } a \le x < b \\ 1 & \text{pro } x \ge b. \end{cases} \tag{4.22}$$

Základní charakteristiky $\mathcal{U}(a,b)$-rozdělení jsou střední hodnota $E(X) = \frac{a+b}{2}$, rozptyl $D(X) = \frac{1}{12}(b-a)^2$ a medián $Q_{0.5} = \frac{b+a}{2}$.

Obrázek 4.2 *Hustota a distribuční funkce $\mathcal{U}(a,b)$-rozdělení*



(a) hustota (b) distribuční funkce

Interpretace: Rovnoměrným rozdělením se řídí takové náhodné veličiny, které mají stejnou možnost nabýt kterékoliv hodnoty z nějakého intervalu. Jsou to např. chyby při zaokrouhlování čísel, chyby při odečítání údajů z lineárních stupnic měřících přístrojů, doby čekání na uskutečnění jevu opakujícího se v pravidelných časových intervalech.

Příklad 4.7 *Rovnoměrné rozdělení*

Určitým místem výrobní linky prochází každých 5 minut polotovar. Pracovník technické kontroly odebírá několikrát za den jeden polotovar, aby ho vyzkoušel. Pravděpodobnost příchodu pracovníka k lince je pro každý časový okamžik stejná. Jaká je pravděpodobnost, že bude čekat na polotovar nejvýše jednu minutu?

Řešení: Požadovanou pravděpodobnost udává distribuční funkce (??), přičemž $a = 0,\ b = 5$. $P(X \le 1) = F(1) = \frac{1}{5}$.

■

Normované normální rozdělení $\mathcal{N}(0,1)$

Rozdělení pravděpodobností na $\mathbb{R}$ s hustotou

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right), \tag{4.23}$$

se nazývá normované normální (Gaussovo) rozdělení nebo $\mathcal{N}(0,1)$-rozdělení. Náhodná veličina s $\mathcal{N}(0,1)$-rozdělením se nazývá normovaná normální náhodná veličina. Hustota $\mathcal{N}(0,1)$-rozdělení má tvar zvonovité křivky a nazývá se *normovaná normální (Gaussova, gaussovská) křivka.*

Základní vlastnosti $\mathcal{N}(0,1)$-rozdělení

1. Platí $\lim_{z\to\pm\infty}\varphi(z)=0$.
   To znamená, že pro $z\to\pm\infty$ se normovaná normální křivka asymptoticky přibližuje k nule.
2. Hustota $\varphi(z)$ je sudá funkce: $\varphi(-z)=\varphi(z)$.
   Tudíž normovaná normální křivka je symetrická kolem 0. Hustota $\mathcal{N}(0,1)$-rozdělení nabývá svého maxima pro $z=0$.
3. $E(Z)=0$, $D(Z)=1$, $Q_{0.5}=0$.
   Střední hodnota tohoto rozdělení charakterizující polohu rozdělení je rovna nule, a rozptyl charakterizující rozptýlení hodnot kolem nuly je roven jedné.
4. $P(-3<Z\leq 3)\approx 0.997$. To znamená, že většina plochy pod normovanou normální křivkou leží mezi -3 a $+3$.

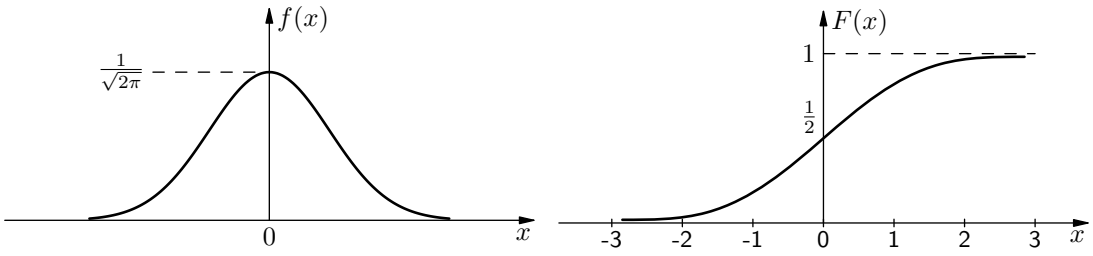
Distribuční funkce $\mathcal{N}(0,1)$-rozdělení se obvykle značí Φ

$$\Phi(z)=\int_{-\infty}^{z}\varphi(u)\,\mathrm{d}u,\qquad z\in\mathbb{R} \tag{4.24}$$

a bývá tabelována pouze pro hodnoty $z>0$. Protože však hustota φ je sudá, platí

$$\Phi(-z)=1-\Phi(z). \tag{4.25}$$

Obrázek 4.3 *Hustota a distribuční funkce $\mathcal{N}(0,1)$-rozdělení*



(a) hustota (b) distribuční funkce

Zároveň lze dokázat, že pro kvantily Q_p normovaného normálního rozdělení platí:

$$Q_p=-Q_{1-p} \tag{4.26}$$

Symbolem z_α budeme značit hodnotu pro kterou platí:

$$\alpha=\int_{z_\alpha}^{\infty}\varphi(z)\,\mathrm{d}z. \tag{4.27}$$

Normální rozdělení $\mathcal{N}(\mu, \sigma^2)$

Rozdělení pravděpodobností na $\mathbb{R}$ se nazývá normální (Gaussovo) rozdělení se střední hodnotou μ a rozptylem σ^2 nebo $\mathcal{N}(\mu, \sigma^2)$-rozdělení, jestliže má hustotu

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad \mu \in \mathbb{R},\ \sigma^2 \in \mathbb{R}_+. \tag{4.28}$$

Normální rozdělení má tvar zvonovité křivky, která nabývá maxima v bodě $x = \mu$ a při $n \to \pm\infty$ se přibližuje k ose x.

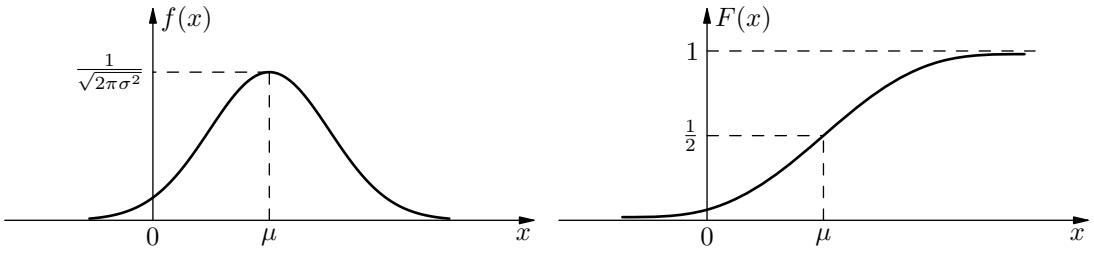
Výpočet distribuční funkce tohoto rozdělení je obtížný. Proto transformujeme náhodnou veličinu X na **normovanou normální veličinu** Z, kde

$$Z = \frac{X - \mu}{\sigma}. \tag{4.29}$$

Veličina Z má pak $\mathcal{N}(0, 1)$-rozdělení. Distribuční funkci $F(x)$ lze vyjádřit pomocí distribuční funkce $\mathcal{N}(0, 1)$-rozdělení

$$F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right).$$

Obrázek 4.4 *Hustota a distribuční funkce $\mathcal{N}(\mu, \sigma^2)$-rozdělení*



(a) hustota (b) distribuční funkce

Empirické pravidlo pro normálně rozdělené náhodné veličiny
Pro každou normálně rozdělenou náhodnou veličinu X platí:

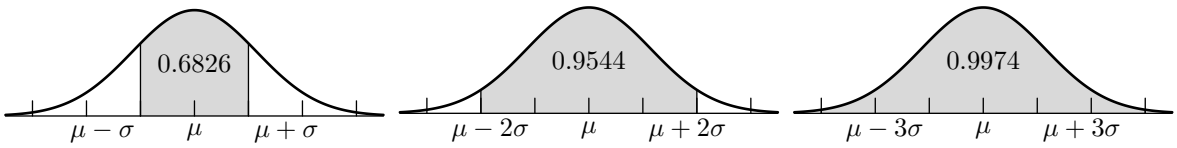
(a) $P(\mu - \sigma < X < \mu + \sigma) = 0.6826$,

(b) $P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.9544$,

(c) $P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9974$.

Tyto vlastnosti jsou graficky znázorněny na obr. **??**.

Obrázek 4.5 *Empirická pravidla pro normálně rozdělenou náhodnou veličinu*



Aplikace: Normální rozdělení má v teorii pravděpodobnosti mimořádný význam. Slouží jako pravděpodobnostní model chování velkého množství náhodných jevů v technice, přírodních

vědách a v ekonomii. Mnoho náhodných veličin vyskytujících se v praktických aplikacích má alespoň přibližně normální rozdělení. Normální rozdělení bývá někdy nazýváno „zákonem chyb“. Při opakovaném měření téže veličiny za stejných podmínek způsobují náhodné vlivy odchylky od skutečné hodnoty měřené veličiny. Tyto náhodné chyby mají často normální rozdělení. Velký význam normálního rozdělení spočívá také v tom, že za určitých podmínek lze pomocí něj aproximovat řadu diskrétních i spojitých rozdělení.

Příklad 4.8 *Normální rozdělení*

Doba potřebná na vypracování testu na vysoké škole má normální rozdělení se střední hodnotou 110 minut a směrodatnou odchylkou 20 minut.

a) Kolik procent studentů dokončí test do dvou hodin? *b)* Jak dlouho by měl test trvat, aby ho dokončilo právě 90% studentů?

Řešení: Nechť X značí dobu potřebnou na vypracování testu. Pak $X \sim \mathcal{N}(110, 400)$.

a) $P(X \leq 120) = F(120) = \Phi(\frac{120-110}{20}) = \Phi(\frac{10}{20}) = \Phi(0.5) = 0.6915$. Pouze 69.15% studentů dokončí test do dvou hodin. *b)* $P(X \leq t) = F(t) = \Phi(\frac{t-110}{20}) = 0.90$. V tabulkách najdeme, že pro $z = 1.28$ je $P(X \leq 1.28) = 0.90$. Tudíž $\frac{t-110}{20} = 1.28$ a z toho dostaneme $t = 135.6$. Doba potřebná k tomu, aby test dokončilo právě 90% studentů je 2hodiny a 15 minut. ■

□ Exponenciální rozdělení $\mathcal{E}(\lambda)$

Rozdělení pravděpodobností na $\mathbb{R}_+$ se nazývá exponenciální rozdělení s parametrem $\lambda > 0$ nebo $\mathcal{E}(\lambda)$-rozdělení, jestliže má hustotu

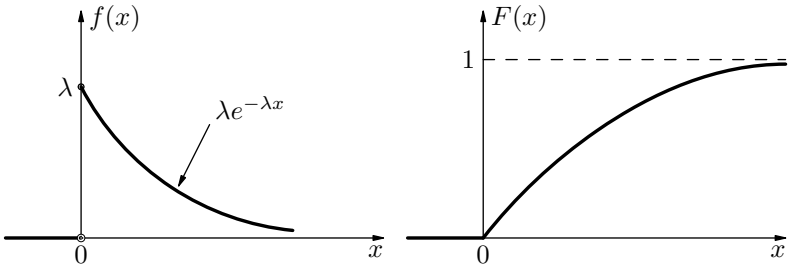
$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{pro } x > 0 \\ 0 & \text{pro } x \leq 0. \end{cases} \tag{4.30}$$

Distribuční funkce je

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{pro } x > 0 \\ 0 & \text{pro } x \leq 0. \end{cases} \tag{4.31}$$

Střední hodnota tohoto rozdělení $E(X) = 1/\lambda$, rozptyl $E(X) = 1/\lambda^2$ a medián $Q_{0.5} = \ln 2/\lambda$.

Obrázek 4.6 *Hustota a distribuční funkce $\mathcal{E}(\lambda)$-rozdělení*



(a) hustota (b) distribuční funkce

Aplikace: Toto rozdělení má uplatnění v teorii spolehlivosti a v teorii hromadné obsluhy, zejména při výpočtu pravděpodobnosti životnosti výrobků a zařízení. Typický příklad náhodné veličiny s $\mathcal{E}(\lambda)$-rozdělením je doba mezi výskytem dvou po sobě následujících náhodných jevů. Ve fyzice je hodnota mediánu $Q_{0.5} = 1/\lambda \ln 2$ známá jako poločas rozpadu radioaktivního prvku.

Příklad 4.9 *Exponenciální rozdělení*

Průměrná doba čekání zákazníka na obsluhu v určité prodejně je 50 sekund, přičemž doba čekání se řídí exponenciálním rozdělením. Jaká je pravděpodobnost, že náhodný zákazník bude obsloužen za dobu ne delší než 30 sekund?

Řešení: Protože $\lambda = 1/50 = 0.02$ je $P(X \leq 30) = 1 - e^{-(0.02).30} = 1 - e^{-0.6} \approx 0.451$.

■

S normálním rozdělením jsou spjata některá další důležitá rozdělení, která budeme používat v dalších kapitolách. Jejich hustotu zde nebudeme uvádět.

□ **chí-kvadrát rozdělení** $\chi^2(n)$

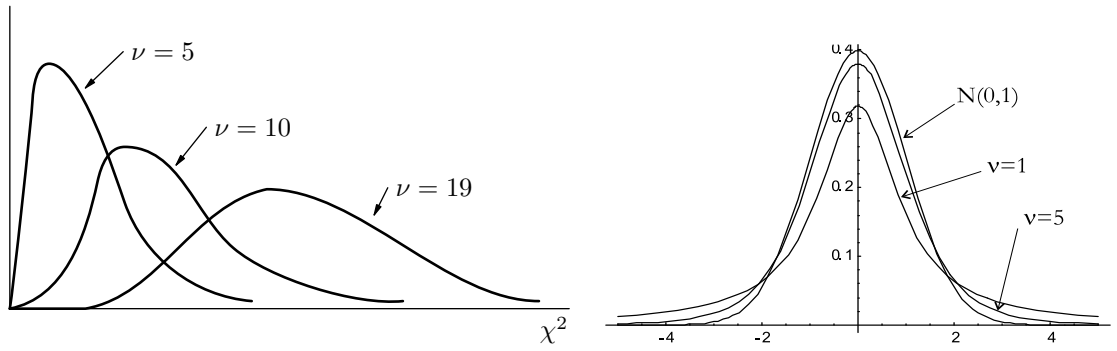
Jestliže $Z_1, Z_2, \cdots, Z_n$ je posloupnost nezávislých náhodných veličin, z nichž každá má $\mathcal{N}(0,1)$-rozdělení, pak součet čtverců těchto veličin, tj. veličina

$$\chi^2 = \sum_{i=1}^{n} Z_i^2,$$

má **chí–kvadrát rozdělení s** n **stupni volnosti**. Počtem stupňů volnosti se rozumí počet nezávislých sčítanců. Je jediným parametrem rozdělení.

Střední hodnota tohoto rozdělení je $E(\chi^2) = n$ a rozptyl $D(\chi^2) = 2n$. Pro různé počty stupňů volnosti ν jsou tabelovány hodnoty χ^2_α, splňující vztah $P(\chi^2 > \chi^2_\alpha) = \alpha, \quad 0 < \alpha < 1$. Se vzrůstajícím počtem stupňů volnosti se χ^2-rozdělení blíží normálnímu rozdělení.

Obrázek 4.7 *Hustota χ^2-rozdělení a t-rozdělení*



(a) χ^2-rozdělení (b) t-rozdělení

□ **Studentovo** t**-rozdělení** $t(n)$

Jestliže Z a χ^2 jsou dvě nezávislé náhodné veličiny takové, že Z má $\mathcal{N}(0,1)$-rozdělení a χ^2 má $\chi^2(n)$-rozdělení, pak veličina

$$T = \frac{Z}{\sqrt{\chi^2}}\sqrt{n}$$

má **Studentovo** t**-rozdělení s** n **stupni volnosti**. Počet stupňů volnosti je jediný parametr tohoto rozdělení. Pro $n \to \infty$ se t-rozdělení blíží normovanému normálnímu rozdělení. Při praktických aplikacích pro $n > 30$ považujeme rozdělení již za normální.

Základní vlastnosti t-rozdělení s n stupni volnosti

1. Hustota $g_n(t)$ je sudá funkce: $g_n(t) = g_n(-t)$.
2. Distribuční funkce splňuje podmínku $G_n(t) = 1 - G_n(-t)$.
3. Pro kvantily platí $Q_p(n) = -Q_{1-p}(n),\ n = 1, 2, \cdots,\ 0 < p < 1$.

Dvourozměrné normální rozdělení

Náhodný vektor (X, Y) má dvourozměrné normální rozdělení s vektorem středních hodnot μ, a kovarianční maticí $\mathbf{\Sigma}$

$$\mu = (\mu_x, \mu_y)^T, \qquad \mathbf{\Sigma} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix},$$

jestliže jeho hustota $f(x, y)$ má tvar

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right)\right\},$$

kde $(x, y) \in \mathbb{R}^2$, a $\rho = \sigma_{xy}/\sigma_x\sigma_y$ je korelační koeficient složek X a Y náhodného vektoru (X, Y). Pro $|\rho| = 1$ není hustota definována. Jestliže $\rho = 0$, pak veličiny X a Y jsou nekorelované, ale v tomto případě také i nezávislé.

4.4 Některé limitní věty

Limitní věty teorie pravděpodobnosti se zabývají chováním posloupností náhodných veličin. Jsou důležité pro popis pravděpodobnostních modelů v případě rostoucího počtu náhodných pokusů.

V tomto odstavci zformulujeme zákon velkých čísel a centrální limitní věty jen v jejich nejjednodušší podobě bez formálního důkazu, pouze s ohledem na jejich věcný obsah.

4.4.1 Zákon velkých čísel

Obecné znění zákona velkých čísel je možné zformulovat takto: Jestliže zvětšujeme počet nezávislých pokusů, přibližuje se empiricky zjištěná charakteristika, popisující výsledky těchto pokusů, charakteristice teoretické. Podmínky působení tohoto zákona specifikují dílčí věty, z nichž nejdůležitější uvedeme. Dílčí věty se dokazují pomocí tzv. Čebyševovy nerovnosti.

Čebyševova nerovnost.

Nechť X je náhodná veličina se střední hodnotou $E(X)$ a rozptylem $D(X)$. Pak pro každé reálné číslo $\epsilon > 0$ platí

$$P(|\ X - E(X)\ | \geq \epsilon) \leq \frac{D(X)}{\epsilon^2}. \tag{4.32}$$

Příklad 4.10 *Ilustrace Čebyševovy nerovnosti*

Nechť náhodná veličina X má libovolné rozdělení se střední hodnotou $\mu = 2$ a rozptylem $\sigma^2 = 1$. Určete pravděpodobnost, že náhodná veličina nabude hodnoty, která se bude lišit od

μ o méně než ± 2.

Řešení: V tomto případě je $\epsilon = 2$. Požadovaná pravděpodobnost je

$$P(\mid X - 2 \mid < 2) = 1 - P(\mid X - 2 \mid \geq 2) \geq 1 - 1/4 = 0.75.$$

■

Přistoupíme nyní k jedné z dílčích vět zákona velkých čísel, a sice k Bernoulliho větě.

Bernoulliho věta (Bernoulliho zákon velkých čísel). Nechť $X_1, X_2, \cdots$ je posloupnost nezávislých stejně rozdělených náhodných veličin s alternativním rozdělením $\mathcal{A}(p)$. Označme $S_n = \sum_{i=1}^{n} X_i$. Pak pro každé $\epsilon > 0$ platí:

$$\lim_{n\to\infty} P\left(\mid \frac{S_n}{n} - p \mid > \epsilon\right) = 0.$$

Bernoulliho věta je jednoduchým důsledkem Čebyševovy nerovnosti.

Výraz S_n/n v předchozí větě je relativní četnost jevu $A = [X_i = 1]$ v n nezávislých opakováních pokusu. Zákon velkých čísel potvrzuje, že pro $n \to \infty$ konverguje relativní četnost ke konstantě a sice k pravděpodobnosti p jevu A. Pojem konvergence posloupnosti náhodných veličin lze definovat různým způsobem, v Bernoulliho větě jde o konvergenci podle pravděpodobnosti.

Řekneme, že posloupnost $X_1, X_2, \cdots$ náhodných veličin **konverguje podle pravděpodobnosti** ke konstantě c, jestliže pro každé $\epsilon > 0$ platí

$$\lim_{n\to\infty} P(\mid X_n - c \mid > \epsilon) = 0.$$

Bernoulliho větu můžeme nyní pomocí pojmu konvergence podle pravděpodobnosti formulovat takto: *Relativní četnost sledovaného jevu v posloupnosti nezávislých pokusů konverguje podle pravděpodobnosti k pravděpodobnosti sledovaného jevu, roste-li počet pokusů nade všechny meze.* Jinak řečeno, při dostatečně velkém počtu nezávislých pokusů velké odchylky relativní četnosti od pravděpodobnosti jsou velmi nepravděpodobné.

Praktický význam této věty spočívá mimo jiné v možnosti experimentálně odhadovat neznámou pravděpodobnost pomocí napozorované relativní četnosti.

Příklad 4.11 *Ilustrace Bernoulliho věty*

Z 2500 nezávisle vyrobených výrobků při určitém procesu výroby jich bylo 100 vadných. Podíl $100/2500 = 0.04$ je blízký číslu p, které vyjadřuje neznámou pravděpodobnost vyrobení vadného výrobku při daném procesu výroby.

■

Následující věta říká, že aritmetický průměr konverguje ke střední hodnotě. To je zobecnění Bernoulliho věty, neboť relativní četnost je průměrem veličin s alternativním rozdělením a pravděpodobnost jevu A je jejich střední hodnotou.

Chinčinova věta

Nechť $X_1, X_2, \cdots$ je posloupnost nezávislých stejně rozdělených náhodných veličin se střední hodnotou μ. Pak pro každé $\epsilon > 0$ platí

$$\lim_{n\to\infty} P\left(\mid \frac{1}{n}\sum_{i=1}^{n} X_i - \mu \mid > \epsilon\right) = 0.$$

Podle zákona velkých čísel můžeme vypočtením relativní četnosti respektive aritmetického průměru (pokud se vztahují k dostatečně velkému počtu pozorování) získat velmi přesnou informaci o pravděpodobnosti nějakého jevu respektive o střední hodnotě nějaké náhodné veličiny.

Příklad 4.12 *Ilustrace Chinčinovy věty*
Nechť doba životnosti X určitého výrobku má $\mathcal{P}(\lambda)$-rozdělení. Potom průměrná doba životnosti $\overline{X} = \sum_{i=1}^{n} X_i$ nezávisle vyrobených výrobků se jen velmi málo liší od neznámé doby životnosti $1/\lambda$. ■

4.4.2 Centrální limitní věty

Centrální limitní věty tvrdí, že součty a tedy i průměry velkého počtu nezávislých náhodných veličin mají za velmi obecných podmínek přibližně normální rozdělení. Tyto věty vysvětlují, proč se v různých oborech setkáváme tak často s normálním nebo přibližně normálním rozdělením.

Typickým příkladem jsou nepřesnosti při měření; výsledná chyba měření je složena z mnoha různých malých chyb. Centrální limitní věty nám umožňují předpokládat, že rozdělení chyb měření je normální. Proto se normálnímu zákonu rozdělení říká zákon chyb. Zmínili jsme se o tom již v odstavci **??**, kde jsme uváděli definici a vlastnosti normálního rozdělení.

Poznámka: O náhodných veličinách, jejichž limitním zákonem je normální rozdělení říkáme, že mají **asymptoticky normální rozdělení**.

Nejjednodušší případ centrální limitní věty je tzv. Moivreova-Laplaceova věta, která vyjadřuje konvergenci binomického rozdělení k rozdělení normálnímu a dává tak možnost aproximovat binomické rozdělení rozdělením normálním.

Moivreova-Laplaceova věta. Nechť $X_1, X_2, \cdots$ je posloupnost nezávislých stejně rozdělených náhodných veličin s alternativním rozdělením $\mathcal{A}(p)$. Položme $S_n = \sum_{i=1}^{n} X_i$ a $Z_n = (S_n - np)/\sqrt{np(1-p)}$. Potom platí

$$\lim_{n\to\infty} P(Z_n \le x) = \Phi(x), \ x \in \mathbb{R}.$$

Příklad 4.13 *Aproximace binomického rozdělení normálním rozdělením*
Student se podrobí zkoušce ve formě testu s 10 otázkami, na které odpovídá *ano* nebo *ne*. Student hádá odpovědi na všechny otázky. Užijte binomické rozdělení ke stanovení přesné pravděpodobnosti, že student odpoví na 7 nebo 8 otázek správně. Pak použijte aproximaci binomického rozdělení normálním rozdělením.
Řešení: Nechť S_{10} je počet správných odpovědí na 10 otázek. Protože student hádá odpovědi, je pravděpodobnost správné odpovědi $p = 0.5$, $S_{10} \sim \mathcal{B}(10, 0.5)$. Z tabulky binomického rozdělení nebo přímým výpočtem dostaneme

$$P(S_{10} = 7 \vee 8) = P(7) + P(8) = 0.1172 + 0.0439 = 0.1611.$$

($X = 7 \vee 8$ označuje výrok X *se rovná 7 nebo 8*). $E(S_{10}) = np = 10 \cdot 0.5 = 5$ a $D(S_n) = \sqrt{np(1-p)} = 1.58$. Protože n není příliš vysoké, je třeba při použití normální aproximace provést

korekci pro nahrazení diskrétního rozdělení spojitým, tzv. *korekci na spojitost.* Úlohu lze totiž formulovat jako určení $P(6.5 \leq S_{10} \leq 8.5)$, neboť platí

$$\begin{aligned} P(6.5 \leq S_{10} \leq 8.5) &= P(S_{10} \leq 8.5) - P(S_{10} < 6.5) = P(S_{10} \leq 8) - P(S_{10} \leq 6) \\ &= P(S_{10} = 8) + P(S_{10} = 7). \end{aligned}$$

Použitím Moivreova-Laplaceovy věty dostaneme

$$\begin{aligned} P\left(\frac{6.5-5}{1.58} \leq Z_{10} \leq \frac{8.5-5}{1.58}\right) &= P(0.95 \leq Z_{10} \leq 2.22) = \Phi(2.22) - \Phi(0.95) \\ &= 0.9868 - 0.8289 = 0.1579. \end{aligned}$$

Porovnáním této hodnoty s hodnotou $P(S_{10} = 7 \vee 8)$ vidíme, že normální aproximace je velice dobrou aproximací binomického rozdělení. ■

Centrální limitní větu, která je přímým zobecněním Moivreovy-Laplaceovy věty, lze vyslovit takto:

Linderbergova-Lévyho věta

Nechť $X_1, X_2, \cdots$ jsou nezávislé náhodné veličiny se stejným rozdělením, které mají konečnou střední hodnotu μ a rozptyl σ^2. Položme $Y_n = \sum_{i=1}^{n} X_i$ a $Z_n = (Y_n - n\mu)/\sigma\sqrt{n}$. Potom platí

$$\lim_{n\to\infty} P(Z_n \leq x) = \Phi(x), \quad x \in \mathbb{R}.$$

Podle této věty konverguje distribuční funkce normovaných součtů k distribuční funkci $\mathcal{N}(0,1)$-rozdělení pro libovolné výchozí rozdělení s konečnou střední hodnotou a konečným rozptylem. Jinak řečeno součet a tím i průměr n nezávislých náhodných veličin, které mají stejné (libovolné) rozdělení s konečnou střední hodnotou a konečným rozptylem má pro dosti velké n přibližně normální rozdělení.

Příklad 4.14 *Ilustrace Linderbergovy-Lévyho věty*
Nechť doba životnosti X určitého výrobku má $\mathcal{P}(\lambda)$-rozdělení. Potom normovaný tvar průměru $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ dob životnosti $X_1, X_2, \cdots, X_n$ nezávisle vyráběných výrobků je

$$Z_n = \frac{\overline{X} - 1/\lambda}{1/\lambda\sqrt{n}}.$$

Z_n se dá pro dostatečně velké n aproximovat rozdělením $\mathcal{N}(0,1)$. ■

Kapitola 5

Náhodný výběr

V předcházejících kapitolách jsme se zabývali popisnou statistikou, pravděpodobností, náhodnými veličinami, některými rozděleními pravděpodobností a limitními větami. Nyní si ukážeme, že tyto zdánlivě různé pojmy jsou základem inferenční statistiky.

Zavedeme pojem náhodný výběr z rozdělení, který má v matematické statistice ústřední postavení a spojuje většinu teoretických výsledků s praktickými situacemi.

5.1 Pojem náhodného výběru

Uvažujme náhodný pokus, jehož výsledkem je hodnota x jednorozměrné náhodné veličiny X, která má distribuční funkci $F(x)$. Opakujeme-li náhodný pokus nezávisle n krát, dostaneme hodnoty $x_1, x_2, \cdots, x_n$. Přitom x_i, $i = 1, 2, \cdots, n$ lze považovat za hodnotu náhodné veličiny X_i. Protože n uvažovaných pokusů je n nezávislých opakování téhož pokusu, jsou náhodné veličiny $X_1, X_2, \cdots, X_n$ vzájemně nezávislé a všechny mají stejné rozdělení, jaké má náhodná veličina X (tj. všechny mají tutéž distribuční funkci $F(x)$, jakou má náhodná veličina X).

Posloupnost nezávislých a stejně rozdělených náhodných veličin $X_1, X_2, \cdots, X_n$ nazýváme **náhodným výběrem o rozsahu** n z rozdělení, které má každá uvažovaná náhodná veličina $X_1, X_2, \cdots, X_n$ (tj. z rozdělení majícího distribuční funkci $F(x)$; místo distribuční funkcí $F(x)$ můžeme ovšem diskrétní rozdělení popsat pravděpodobnostmi $P(x)$ a spojitá rozdělení hustotou pravděpodobnosti $f(x)$). Náhodný výběr budeme značit $\mathbf{X} = (X_1, X_2, \cdots, X_n)$. Posloupnost hodnot $x_1, x_2, \cdots, x_n$, které nabývají náhodné veličiny $X_1, X_2, \cdots, X_n$ nazveme **výběrovými hodnotami** nebo **realizací náhodného výběru**. Množina V hodnot, které nabývají náhodné veličiny $X_1, X_2, \cdots, X_n$, se nazývá **výběrovým prostorem**. Výběrový prostor V je podmnožinou $\mathbb{R}^n$.

Protože náhodné veličiny $X_1, X_2, \cdots, X_n$ jsou vzájemně nezávislé a mají stejné rozdělení, platí pro distribuční funkci $H(\mathbf{x})$ náhodného výběru

$$H(\mathbf{x}) = F(x_1)F(x_2)...F(x_n), \quad x_i \in \mathbb{R}.$$

Příklad 5.1 *Distribuční funkce náhodného výběru*
Nechť $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ je náhodný výběr ze spojitého rovnoměrného rozdělení na intervalu (0,1). Určete distribuční funkci $H(\mathbf{x})$ náhodného výběru $\mathbf{X}$.
Řešení: $X_i \sim \mathcal{U}(0, 1)$

$$H(\mathbf{x}) = H(x_1, x_2, \cdots, x_n) = x_1 \cdot x_2 \cdots x_n.$$

■

Pravděpodobnostní funkce $q(\mathbf{x})$ náhodného výběru v případě diskrétního rozdělení náhodných veličin $X_1, X_2, \cdots, X_n$ je

$$q(\mathbf{x}) = P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n) = p(x_1)p(x_2)\cdots p(x_n)$$

Příklad 5.2 *Pravděpodobnostní funkce náhodného výběru*
Nechť $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ je náhodný výběr z Poissonova rozdělení s parametrem λ. Určete pravděpodobnostní funkci $q(\mathbf{x})$.
Řešení: $X_i \sim \mathcal{P}(\lambda), \quad f(x_i) = \frac{\lambda^{x_i}}{x_i!}e^{-\lambda}, \quad x_i = 0, 1\cdots, \quad i = 1, 2, \cdots, n$

$$q(\mathbf{x}) = \lambda^{\sum_{i=1}^n x_i} e^{-n\lambda} \frac{1}{x_1!x_2!...x_n!}\ .$$

■

Hustota rozdělení $h(\mathbf{x})$ náhodného výběru z rozdělení s hustotou $f(x)$ je

$$h(\mathbf{x}) = h(x_1, x_2, \cdots, x_n) = f(x_1)f(x_2)\cdot f(x_n), \quad x_i \in \mathbb{R}, \quad i = 1, 2, \cdots, n.$$

Příklad 5.3 *Hustota rozdělení náhodného výběru*
Nechť $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ je náhodný výběr z normálního rozdělení $\mathcal{N}(\mu, \sigma^2)$. Najděte hustotu $h(\mathbf{x})$.
Řešení: $X_i \sim \mathcal{N}(\mu, \sigma^2)$

$$h(\mathbf{x}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{1}{2}(\frac{x_i - \mu}{\sigma})^2\} = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2\}, \quad x_i \in \mathbb{R}.$$

■

5.2 Výběrové charakteristiky

Jak již víme, statistický soubor lze popsat pomocí různých popisných charakteristik. Mezi nejdůležitější charakteristiky patří aritmetický průměr, rozptyl a relativní četnost. U spočetných statistických souborů bychom měli spíše hovořit o parametrech rozdělení sledovaného znaku. K těmto charakteristikám a parametrům můžeme najít ve výběrovém souboru příslušné protějšky, tj. **výběrové charakteristiky** neboli **statistiky**.

Zatímco charakteristiky základního souboru a parametry rozdělení sledovaného znaku jsou pevné hodnoty, statistiky se mění od jednoho náhodného výběru ke druhému. Z pravděpodobnostního hlediska mají charakter náhodných veličin, neboť jsou vypočteny z hodnot náhodného výběru, které jsou samy hodnotami náhodných veličin. Tyto náhodné veličiny neobsahují parametry rozdělení. Příklady výběrových charakteristik jsou: *výběrový průměr*, *výběrový rozptyl* a *výběrový podíl*.

5.3 Rozdělení výběrových charakteristik

Chceme-li na základě výběrové charakteristiky dělat závěry o charakteristice základního souboru nebo o parametru rozdělení, je nutné vždy znát pravděpodobnostní rozdělení výběrové charakteristiky, které se nazývá **výběrové rozdělení**.

Výběrová rozdělení jsou teoretickým základem pro zpracování výsledků výběrových šetření, jejich poznání je rozhodujícím krokem, který teprve umožňuje aplikovat zákonitosti počtu pravděpodobnosti na hodnocení kvality úsudků opírajících se o náhodný výběr.

V této části uvedeme výběrová rozdělení statistik, na jejichž základě budeme v kapitole 6 odhadovat neznámé parametry rozdělení pravděpodobností a v kapitole **??** testovat hypotézy o těchto parametrech.

5.3.1 Rozdělení výběrového průměru

Je-li $(X_1, X_2, \cdots, X_n)$ náhodný výběr o rozsahu n, pak **výběrový průměr** (nebo také výběrový 1. obecný moment) je statistika definovaná jako

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{5.1}$$

Obecně, **výběrový k-tý obecný moment** je statistika

$$M_k' = \frac{1}{n}\sum_{i=1}^{n} X_i^k. \tag{5.2}$$

Nechť $(X_1, X_2, \cdots, X_n)$ je náhodný výběr o rozsahu n z rozdělení se střední hodnotou μ a rozptylem σ^2, pak pro střední hodnotu $\mu_{\bar{x}}$ a rozptyl $\sigma^2_{\bar{x}}$ výběrového průměru $\overline{X}$ platí

$$\mu_{\bar{x}} = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu \tag{5.3}$$

$$\sigma^2_{\bar{x}} = D(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) = \frac{1}{n}\sigma^2. \tag{5.4}$$

Známe-li rozdělení, z něhož náhodný výběr pochází, můžeme stanovit rozdělení výběrového průměru jako rozdělení lineární funkce náhodných veličin. Je-li např. $(X_1, X_2, \cdots, X_n)$ náhodný výběr z $\mathcal{N}(\mu, \sigma^2)$-rozdělení, pak $\overline{X} \sim \mathcal{N}(\mu, \sigma^2/n)$.

Pokud náhodný výběr nepochází z normálního rozdělení, pak z centrální limitní věty (viz odst. **??**) vyplývá, že náhodná veličina $\overline{X}$ má přibližně normální rozdělení za předpokladu, že rozsah výběru je relativně velký. Všeobecně vzato, čím více se rozdělení, z něhož výběr pochází, liší od normálního, tím větší rozsah výběru potřebujeme pro adekvátní aproximaci rozdělení výběrového průměru. Na základě experimentálních výsledků se doporučuje, aby rozsah výběru n byl alespoň 30. Tudíž máme následující poznatek.

Tvrzení 5.1 ROZDĚLENÍ VÝBĚROVÉHO PRŮMĚRU

Předpokládejme, že máme náhodný výběr o rozsahu $n \geq 30$ z rozdělení se střední hodnotou μ, a rozptylem σ^2. Pak bez ohledu na rozdělení, z něhož výběr pochází, má náhodná veličina $\overline{X}$ přibližně normální rozdělení se střední hodnotou $\mu_{\bar{x}} = \mu$ a rozptylem $\sigma^2_{\bar{x}} = \sigma^2/n$.

V kapitolách 6 a **??** budeme používat normovaný tvar náhodné veličiny $\overline{X}$, to je veličinu

$$Z = \frac{\overline{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}, \tag{5.5}$$

která má v důsledku centrální limitní věty rozdělení specifikované při různých podmínkách ve následujícím tvrzení.

Tvrzení 5.2 ROZDĚLENÍ NORMOVANÉHO TVARU VÝBĚROVÉHO PRŮMĚRU

Předpokládejme, že máme náhodný výběr o rozsahu n z rozdělení se střední hodnotou μ a směrodatnou odchylkou σ^2. Pak *normovaný tvar výběrového průměru* $\overline{X}$

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

1. má bez ohledu na rozsah výběru normované normální rozdělení, pokud výběr pochází z normálního rozdělení;
2. má pro $n \geq 30$ přibližně normované normální rozdělení bez ohledu na rozdělení, z něhož výběr pochází.

5.3.2 Rozdělení výběrového rozptylu

Je-li $(X_1, X_2, \cdots, X_n)$ náhodný výběr o rozsahu n, pak **výběrový rozptyl** je statistika definovaná jako

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2. \tag{5.6}$$

Poznámka : **Výběrový k-tý centrální moment** je statistika

$$M_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k. \tag{5.7}$$

Podobně jako v případě výběrového průměru, chceme-li získat informaci o rozptylu rozdělení prostřednictvím výběrového rozptylu, musíme znát jeho rozdělení.

Tvrzení 5.3 ROZDĚLENÍ VÝBĚROVÉHO ROZPTYLU

Předpokládejme, že máme náhodný výběr o rozsahu n z normálního rozdělení s rozptylem σ^2. Pak náhodná veličina

$$\chi^2 = \frac{n-1}{\sigma^2}S^2$$

má χ^2-rozdělení s $n-1$ stupni volnosti.

Nyní předpokládejme, že máme náhodný výběr o rozsahu n z normálního rozdělení se střední hodnotou μ a s neznámým rozptylem. Jelikož náhodná veličina $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$ a veličina $\chi^2 = \frac{n-1}{\sigma}^2 S^2 \sim \chi^2(n-1)$, pak z definice t-rozdělení vyplývá že náhodná veličina $Z/\sqrt{\chi 2/n-1}$ má t-rozdělení s $n-1$ stupni volnosti. Vzhledem k tomu, že platí relace

$$\frac{Z}{\sqrt{\chi^2/n-1}} = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \cdot \frac{\sqrt{n-1}}{\sqrt{\frac{n-1}{\sigma^2}S^2}} = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \cdot \frac{\sigma}{S} = \frac{\overline{X}-\mu}{S/\sqrt{n}}$$

dostáváme pro statistiku

$$T = \frac{\overline{X}-\mu}{S/\sqrt{n}},$$

kterou budeme nazývat **t-statistikou**, následující tvrzení.

Tvrzení 5.4 ROZDĚLENÍ t-STATISTIKY

Mějme náhodný výběr o rozsahu n z normálního rozdělení se střední hodnotou μ. Pak má náhodná veličina

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

t-rozdělení s $n - 1$ stupni volnosti.

5.3.3 Rozdělení výběrového podílu

Uvažujme náhodný výběr ze základního souboru, v němž sledovaný statistický znak nebo sledovaná náhodná veličina nabývá pouze hodnot nula a jedna. V tomto případě mluvíme o výběru z alternativního rozdělení. Tímto rozdělením kvantifikujeme například takové situace, kdy hodnotě statistického znaku, který nás zajímá, přiřadíme číselnou hodnotu 1 a všem dalším číselnou hodnotu 0 a zajímá nás, jaké procento statistických jednotek ze základního souboru má určitou sledovanou vlastnost. Jde o tzv. **dvoukategoriální** základní soubor. Například, jestliže základní soubor o rozsahu N, který uvažujeme, tvoří všechny domácnosti v ČR, sledovaná vlastnost je „vlastnictví osobního počítače“, (1 – domácnost má osobní počítač, 0 – domácnost nemá osobní počítač), počet domácností vlastnících osobní počítač je N_v, pak **podíl základního souboru** je podíl všech domácností v ČR, které vlastní osobní počítač, tj. N_v/N.

Předpokládejme, že rozdělení v základním souboru je alternativní a že p značí buď relativní četnost hodnoty 1 (podíl statistických jednotek s hodnotou sledovaného znaku 1) v konečném základním souboru, nebo pravděpodobnost hodnoty 1, uvažujeme-li nekonečný základní soubor. Může-li sledovaný znak nebo sledovaná náhodná veličina nabývat pouze hodnot 0 a 1, pak také výběrovými hodnotami $x_1, x_2, \cdots, x_n$ mohou být buď jedničky nebo nuly. Protože výběr je náhodný, je počet jedniček x ve výběru hodnotou náhodné veličiny X, která se nazývá **výběrovou absolutní četností**. Podíl $\hat{p} = x/n$, kde x značí počet jednotek výběru majících specifikovanou vlastnost (nazývaný často „počet úspěchů“ a $n - x$ „počet neúspěchů“) a n je rozsah výběru, je pak hodnotou náhodné veličiny

$$\hat{P} = \frac{X}{n},$$

která se nazývá **výběrovou relativní četností** nebo častěji **výběrovým podílem**. Z toho, co bylo řečeno je zřejmé, že výběrový podíl je roven výběrovému průměru náhodného výběru z alternativního rozdělení.
Poznámka: V dalším textu budeme používat stejné označení $\hat{p}$ pro náhodnou veličinu $\hat{P}$ i její hodnotu $\hat{p}$.

Podobně jako v případě střední hodnoty, musíme znát **výběrové rozdělení podílu**, (pravděpodobnostní rozdělení náhodné veličiny $\hat{p}$) , abychom mohli dělat závěry o podílu p. Z Moivreovy-Laplaceovy limitní věty (viz odst. **??**) vyplývá následující tvrzení.

Tvrzení 5.5 ROZDĚLENÍ VÝBĚROVÉHO PODÍLU

Předpokládejme, že máme náhodný výběr velkého rozsahu n z alternativního rozdělení s podílem p. Pak náhodná veličina $\hat{p}$ má přibližně normální rozdělení se střední hodnotou $\mu_{\hat{p}} = p$ a směrodatnou odchylkou $\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$.

Z tvrzení 5.4 lze odvodit, že normovaná náhodná veličina

$$Z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \tag{5.8}$$

má pro velká n přibližně normované normální rozdělení.

Přesnost normální aproximace závisí na n a p. Pro p blízké 0.5 je aproximace dostatečně přesná pro rozumné n. Čím se p více liší od 0.5, tím větší n potřebujeme k tomu, aby aproximace byla přesná. Bývá zvykem používat aproximaci normálním rozdělením, pokud $np \geq 5$ a zároveň $n(1-p) \geq 5$, neboli $\min(np, n(1-p)) \geq 5$.

5.4 Nezávislé náhodné výběry

Některé metody, kterými se budeme v kapitole **??** zabývat, nevyžadují pouze, aby výběry byly náhodné, ale také aby byly *nezávislé*, zhruba řečeno, aby výběr z jednoho rozdělení neměl žádný vliv na výběr z jiného rozdělení.

Nechť $\mathbf{X_1} = (X_{11}, X_{12}, \cdots, X_{1n_1})$ je náhodný výběr rozsahu n_1 z rozdělení s distribuční funkcí $F_1(x)$ a $\mathbf{X_2} = (X_{21}, X_{22}, \cdots, X_{2n_2})$ je náhodný výběr rozsahu n_2 z rozdělení s distribuční funkcí $F_2(x)$. Náhodné výběry $\mathbf{X_1}$ a $\mathbf{X_2}$ jsou **nezávislé**, jestliže náhodné veličiny $X_{11}, X_{12}, \cdots, X_{1n_1}, X_{21}, X_{22}, \cdots, X_{2n_2}$ jsou nezávislé, přičemž veličiny $X_{11}, X_{12}, \cdots, X_{1n}$ mají distribuční funkcí $F_1(x)$ a X_{21},X_{22},$\cdots, X_{2n}$ mají distribuční funkcí $F_2(x)$ (viz odst. **??**). Jsou-li distribuční funkce $F_1(x)$ a $F_2(x)$ identické, jedná se o dva nezávislé výběry z téhož rozdělení.

5.4.1 Dva nezávislé výběry z normálního rozdělení nebo velké rozsahy výběrů

Mějme náhodný výběr $\mathbf{X_1} = (X_{11}, X_{12}, \cdots, X_{1n_1})$ rozsahu n_1 z rozdělení $\mathcal{N}(\mu_1, \sigma_1^2)$ a náhodný výběr $\mathbf{X_2} = (X_{21}, X_{22}, \cdots, X_{2n_2})$ rozsahu n_2 z rozdělení $\mathcal{N}(\mu_2, \sigma_2^2)$. Nechť výběry $\mathbf{X_1}$ a $\mathbf{X_2}$ jsou nezávislé. Potom statistiky $\overline{X_1}$ a $\overline{X_2}$ jsou nezávislé (viz odstavec **??**), $\overline{X_1} \sim \mathcal{N}(\mu_1, \sigma_1^2/n_1)$, $\overline{X_2} \sim \mathcal{N}(\mu_2, \sigma_2^2/n_2)$ a statistika $\overline{X_1} - \overline{X_2}$ má rozdělení $\mathcal{N}(\mu_1 - \mu_2, \sigma_1^2/n_1 + \sigma_2^2/n_2)$ (viz odstavec 5.3.1). Bezprostředním důsledkem je následující tvrzení.

Tvrzení 5.6 ROZDĚLENÍ ROZDÍLU VÝBĚROVÝCH PRŮMĚRŮ (NEZÁVISLÉ VÝBĚRY)

Předpokládejme, že máme dva nezávislé náhodné výběry o rozsazích n_1 a n_2 z rozdělení se středními hodnotami μ_1 a μ_2 a směrodatnými odchylkami σ_1 a σ_2. Dále předpokládejme, že buď obě rozdělení jsou normální nebo oba výběry mají velký rozsah. Pak náhodná veličina $\overline{X}_1 - \overline{X}_2$ má (přibližně) normální rozdělení se střední hodnotou $\mu_{(\bar{x}_1 - \bar{x}_2)} = \mu_1 - \mu_2$ a směrodatnou odchylkou $\sigma_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\sigma_1/n_1 + \sigma_2/n_2}$. Tudíž normovaná náhodná veličina

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}} \tag{5.9}$$

má alespoň přibližně normované normální rozdělení.

Toto tvrzení tvoří teoretický základ pro odvození statistických indukčních metod pro porovnání středních hodnot dvou základních souborů.

☐ Dva nezávislé výběry z rozdělení se shodnými rozptyly

Nyní předpokládejme, že $\sigma_1^2 = \sigma_2^2 = \sigma^2$ a rozptyl σ^2 není znám, což je obvyklé v praktických případech. Dosazením hodnoty σ^2 za σ_1^2 a σ_2^2 do definice náhodné veličiny Z ve vztahu (5.9) dostaneme náhodnou veličinu

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{(1/n_1) + (1/n_2)}}. \tag{5.10}$$

Výběrové rozptyly S_1^2 a S_2^2 použijeme k sestrojení tzv. **sdruženého výběrového rozptylu** S_P^2

$$S_P^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}. \tag{5.11}$$

Sdružený výběrový rozptyl můžeme chápat jako vážený rozptyl, ve kterém jednotlivé výběrové rozptyly S_1^2 a S_2^2 jsou váženy odpovídajícími stupni volnosti. (Index „P“ pochází z anglického termínu „pooled sample variance“, který znamená sdružený výběrový rozptyl). Nahrazením neznámého rozptylu σ^2 v rovnici (5.10) sdruženým výběrovým rozptylem S_P^2, dostaneme náhodnou veličinu

$$\frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_P\sqrt{(1/n_1) + (1/n_2)}}, \tag{5.12}$$

která na rozdíl od náhodné veličiny definované v (5.10), nemá normované normální rozdělení, ale t-rozdělení. Náhodnou veličinu definovanou v (5.12) budeme nazývat **sdružená t-statistika**. Její rozdělení specifikuje následující tvrzení.

Tvrzení 5.7 ROZDĚLENÍ SDRUŽENÉ t-STATISTIKY

Předpokládejme, že máme dva nezávislé náhodné výběry o rozsazích n_1 a n_2 z rozdělení se středními hodnotami μ_1 a μ_2. Dále předpokládejme, že směrodatné odchylky obou rozdělení jsou shodné. Pak náhodná veličina

$$T = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{S_P\sqrt{1/n_1 + 1/n_2}},$$

kde S_P je definováno v (5.11), má t-rozdělení s $n_1 + n_2 - 2$ stupni volnosti.

☐ Dva nezávislé výběry z rozdělení s různými rozptyly

Podobně jako v případě diskutovaném výše budeme předpokládat, že standardní odchylky v obou výběrech jsou neznámé. Nahradíme σ_1 a σ_2 výběrovými směrodatnými odchylkami S_1 a S_2 a dostaneme náhodnou veličinu,

$$\frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{(S_1^2/n_1) + (S_2^2/n_2)}}, \tag{5.13}$$

která již nemá normované normální rozdělení, ale má přibližně t-rozdělení. Tuto statistiku budeme nazývat **nesdružená t-statistika** .

Tvrzení 5.8 ROZDĚLENÍ NESDRUŽENÉ t-STATISTIKY

Předpokládejme, že máme dva nezávislé výběry o rozsahu n_1 a n_2 z normálních rozdělení se středními hodnotami μ_1 a μ_2. Pak má náhodná veličina

$$T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{(S_1^2/n_1) + (S_2^2/n_2)}}$$

přibližně t-rozdělení s počtem stupňů volnosti δ, kde

$$\delta = \frac{[(s_1^2/n_1) + (s_2^2/n_2)]^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}},$$

zaokrouhleno dolů na nejbližší celé číslo.

5.4.2 Dva nezávislé výběry z alternativního rozdělení

Máme-li dva nezávislé náhodné výběry o rozsahu n_1 a n_2 z alternativních rozdělení s parametry (podíly) p_1 a p_2, pak je výběrový podíl $\hat{p}_i$, $i = 1, 2$ roven výběrovému průměru $\overline{X_i}$. Z tvrzení 5.5 a 5.6 plyne následující tvrzení 5.9, které tvoří teoretický základ nutný pro odvození statistických indukčních metod pro porovnání dvou dvoukategoriálních základních souborů.

Tvrzení 5.9 ROZDĚLENÍ ROZDÍLU DVOU VÝBĚROVÝCH PODÍLŮ (NEZÁVISLÉ VÝBĚRY)

Předpokládejme, že máme dva nezávislé náhodné výběry o rozsazích n_1 a n_2 z alternativních rozdělení s podíly p_1 a p_2. Pak pro velké výběry má náhodná veličina $\hat{p}_1 - \hat{p}_2$ přibližně normální rozdělení se střední hodnotou $\mu_{(\hat{p}_1-\hat{p}_2)} = p_1 - p_2$ a směrodatnou odchylkou $\sigma_{(\hat{p}_1-\hat{p}_2)} = \sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}$, kde $\hat{p}_i = x_i/n_i$ je výběrový podíl i-té populace, x_i je počet úspěchů v i-té populaci, $i = 1, 2$. Tudíž normovaná náhodná veličina

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}}$$

má přibližně normované normální rozdělení.

5.5 Párové náhodné výběry

Nechť $\mathbf{X_1} = (X_{11}, X_{12}, \cdots, X_{1n})$ je náhodný výběr rozsahu n z rozdělení se střední hodnotou μ_1 a rozptylem σ_1^2, a $\mathbf{X_2} = (X_{21}, X_{22}, \cdots, X_{2n})$ je náhodný výběr stejného rozsahu n z rozdělení se střední hodnotou μ_2 a rozptylem σ_2^2. Z těchto dvou výběrů utvoříme výběr n dvojic $(X_{11}, X_{21}), (X_{12}, X_{22}), ..., (X_{1n}, X_{2n})$. Každé dvojici veličin (X_{1j}, X_{2j}), $j = 1, 2, \cdots, n$ přiřadíme náhodnou veličinu $D_j = X_{1j} - X_{2j}$, $j = 1, 2, \cdots, n$, tzv. **párovou diferenci**, kterou získáme odečtením příslušné párové hodnoty v druhém výběru od párové hodnoty v prvním výběru. Na posloupnost párových diferencí $D_1, D_2, \cdots, D_n$ náhodně vybraných n dvojic se můžeme dívat jako na náhodný výběr z rozdělení všech možných párových diferencí. Označme střední hodnotu takového rozdělení párových diferencí μ_d.

Pak lze ukázat, že

$$\mu_d = \mu_1 - \mu_2. \tag{5.14}$$

O vztahu rozptylu σ_d^2 rozdělení párových diferencí k rozptylům σ_1^2 a σ_2^2 nemůžeme vzhledem k možné závislosti veličin nic předpokládat. Označme $\overline{D}$ výběrový průměr párových diferencí, tudíž $\overline{D} = \overline{X}_1 - \overline{X}_2$, kde $\overline{X}_i$ je výběrový průměr náhodného výběru z i-tého rozdělení, $i = 1, 2$. Dále označme S_d výběrovou směrodatnou odchylku párových diferencí pro kterou platí

$$S_d = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(D_j - \overline{D})^2}. \tag{5.15}$$

Je-li rozdělení párových diferencí normální, pak můžeme aplikovat tvrzení 5.3, použít rovnost (5.14) a dostaneme následující výsledek.

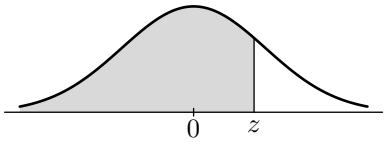
Tvrzení 5.10 ROZDĚLENÍ PÁROVÉ t-STATISTIKY

Předpokládejme, že máme náhodný výběr n dvojic z rozdělení se středními hodnotami μ_1 a μ_2. Dále předpokládejme, že rozdělení všech párových dvojic je normální. Pak náhodná veličina

$$T = \frac{\overline{D} - (\mu_1 - \mu_2)}{S_d/\sqrt{n}}$$

má t-rozdělení s $n - 1$ stupni volnosti.

Tabulka I: *Distribuční funkce normovaného normálního rozdělení* $\mathcal{N}(0,1)$



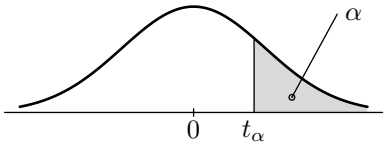
Pro $z < 0.0$ použijte vztah $\Phi(z) = 1 - \Phi(-z)$.

z	0.000	0.010	0.020	0.030	0.040	0.050	0.060	0.070	0.080	0.090	z
0.0	0.500	0.504	0.508	0.512	0.516	0.520	0.524	0.528	0.532	0.536	0.0
0.1	0.540	0.544	0.548	0.552	0.556	0.560	0.564	0.567	0.571	0.575	0.1
0.2	0.579	0.583	0.587	0.591	0.595	0.599	0.603	0.606	0.610	0.614	0.2
0.3	0.618	0.622	0.626	0.629	0.633	0.637	0.641	0.644	0.648	0.652	0.3
0.4	0.655	0.659	0.663	0.666	0.670	0.674	0.677	0.681	0.684	0.688	0.4
0.5	0.691	0.695	0.698	0.702	0.705	0.709	0.712	0.716	0.719	0.722	0.5
0.6	0.726	0.729	0.732	0.736	0.739	0.742	0.745	0.749	0.752	0.755	0.6
0.7	0.758	0.761	0.764	0.767	0.770	0.773	0.776	0.779	0.782	0.785	0.7
0.8	0.788	0.791	0.794	0.797	0.800	0.802	0.805	0.808	0.811	0.813	0.8
0.9	0.816	0.819	0.821	0.824	0.826	0.829	0.831	0.834	0.836	0.839	0.9
1.0	0.841	0.844	0.846	0.848	0.851	0.853	0.855	0.858	0.860	0.862	1.0
1.1	0.864	0.867	0.869	0.871	0.873	0.875	0.877	0.879	0.881	0.883	1.1
1.2	0.885	0.887	0.889	0.891	0.893	0.894	0.896	0.898	0.900	0.901	1.2
1.3	0.903	0.905	0.907	0.908	0.910	0.911	0.913	0.915	0.916	0.918	1.3
1.4	0.919	0.921	0.922	0.924	0.925	0.926	0.928	0.929	0.931	0.932	1.4
1.5	0.933	0.934	0.936	0.937	0.938	0.939	0.941	0.942	0.943	0.944	1.5
1.6	0.945	0.946	0.947	0.948	0.949	0.951	0.952	0.953	0.954	0.954	1.6
1.7	0.955	0.956	0.957	0.958	0.959	0.960	0.961	0.962	0.962	0.963	1.7
1.8	0.964	0.965	0.966	0.966	0.967	0.968	0.969	0.969	0.970	0.971	1.8
1.9	0.971	0.972	0.973	0.973	0.974	0.974	0.975	0.976	0.976	0.977	1.9
2.0	0.977	0.978	0.978	0.979	0.979	0.980	0.980	0.981	0.981	0.982	2.0
2.1	0.982	0.983	0.983	0.983	0.984	0.984	0.985	0.985	0.985	0.986	2.1
2.2	0.986	0.986	0.987	0.987	0.987	0.988	0.988	0.988	0.989	0.989	2.2
2.3	0.989	0.990	0.990	0.990	0.990	0.991	0.991	0.991	0.991	0.992	2.3
2.4	0.992	0.992	0.992	0.992	0.993	0.993	0.993	0.993	0.993	0.994	2.4
2.5	0.994	0.994	0.994	0.994	0.994	0.995	0.995	0.995	0.995	0.995	2.5
2.6	0.995	0.995	0.996	0.996	0.996	0.996	0.996	0.996	0.996	0.996	2.6
2.7	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.997	2.7
2.8	0.997	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998	2.8
2.9	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.999	0.999	0.999	2.9

Tabulka II: *Kritické hodnoty normovaného normálního rozdělení* $\mathcal{N}(0,1)$

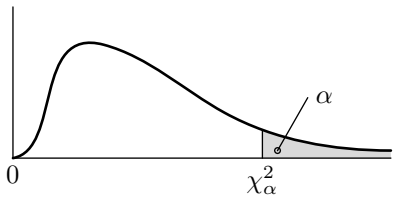
α	0.2	0.1	0.05	0.025	0.01	0.005	0.0025	0.001
z_α	0.842	1.282	1.645	1.960	2.326	2.576	2.807	3.090

Tabulka III: *Kritické hodnoty t-rozdělení*



ν	$t_{0.2}$	$t_{0.1}$	$t_{0.05}$	$t_{0.025}$	$t_{0.01}$	$t_{0.005}$	$t_{0.0025}$	$t_{0.001}$	ν
1	1.376	3.078	6.314	12.706	31.821	63.656	127.321	318.289	1
2	1.061	1.886	2.920	4.303	6.965	9.925	14.089	22.328	2
3	0.978	1.638	2.353	3.182	4.541	5.841	7.453	10.214	3
4	0.941	1.533	2.132	2.776	3.747	4.604	5.598	7.173	4
5	0.920	1.476	2.015	2.571	3.365	4.032	4.773	5.894	5
6	0.906	1.440	1.943	2.447	3.143	3.707	4.317	5.208	6
7	0.896	1.415	1.895	2.365	2.998	3.499	4.029	4.785	7
8	0.889	1.397	1.860	2.306	2.896	3.355	3.833	4.501	8
9	0.883	1.383	1.833	2.262	2.821	3.250	3.690	4.297	9
10	0.879	1.372	1.812	2.228	2.764	3.169	3.581	4.144	10
11	0.876	1.363	1.796	2.201	2.718	3.106	3.497	4.025	11
12	0.873	1.356	1.782	2.179	2.681	3.055	3.428	3.930	12
13	0.870	1.350	1.771	2.160	2.650	3.012	3.372	3.852	13
14	0.868	1.345	1.761	2.145	2.624	2.977	3.326	3.787	14
15	0.866	1.341	1.753	2.131	2.602	2.947	3.286	3.733	15
16	0.865	1.337	1.746	2.120	2.583	2.921	3.252	3.686	16
17	0.863	1.333	1.740	2.110	2.567	2.898	3.222	3.646	17
18	0.862	1.330	1.734	2.101	2.552	2.878	3.197	3.610	18
19	0.861	1.328	1.729	2.093	2.539	2.861	3.174	3.579	19
20	0.860	1.325	1.725	2.086	2.528	2.845	3.153	3.552	20
21	0.859	1.323	1.721	2.080	2.518	2.831	3.135	3.527	21
22	0.858	1.321	1.717	2.074	2.508	2.819	3.119	3.505	22
23	0.858	1.319	1.714	2.069	2.500	2.807	3.104	3.485	23
24	0.857	1.318	1.711	2.064	2.492	2.797	3.091	3.467	24
25	0.856	1.316	1.708	2.060	2.485	2.787	3.078	3.450	25
26	0.856	1.315	1.706	2.056	2.479	2.779	3.067	3.435	26
27	0.855	1.314	1.703	2.052	2.473	2.771	3.057	3.421	27
28	0.855	1.313	1.701	2.048	2.467	2.763	3.047	3.408	28
29	0.854	1.311	1.699	2.045	2.462	2.756	3.038	3.396	29
30	0.854	1.310	1.697	2.042	2.457	2.750	3.030	3.385	30
40	0.851	1.303	1.684	2.021	2.423	2.704	2.971	3.307	40
50	0.849	1.299	1.676	2.009	2.403	2.678	2.937	3.261	50
60	0.848	1.296	1.671	2.000	2.390	2.660	2.915	3.232	60
70	0.847	1.294	1.667	1.994	2.381	2.648	2.899	3.211	70
80	0.846	1.292	1.664	1.990	2.374	2.639	2.887	3.195	80
90	0.846	1.291	1.662	1.987	2.368	2.632	2.878	3.183	90
100	0.845	1.290	1.660	1.984	2.364	2.626	2.871	3.174	100

Tabulka IV: *Kritické hodnoty χ^2-rozdělení*



ν	$\chi^2_{0.995}$	$\chi^2_{0.99}$	$\chi^2_{0.975}$	$\chi^2_{0.95}$	$\chi^2_{0.9}$	ν
1	0.000	0.000	0.001	0.004	0.016	1
2	0.010	0.020	0.051	0.103	0.211	2
3	0.072	0.115	0.216	0.352	0.584	3
4	0.207	0.297	0.484	0.711	1.064	4
5	0.412	0.554	0.831	1.145	1.610	5
6	0.676	0.872	1.237	1.635	2.204	6
7	0.989	1.239	1.690	2.167	2.833	7
8	1.344	1.647	2.180	2.733	3.490	8
9	1.735	2.088	2.700	3.325	4.168	9
10	2.156	2.558	3.247	3.940	4.865	10
11	2.603	3.053	3.816	4.575	5.578	11
12	3.074	3.571	4.404	5.226	6.304	12
13	3.565	4.107	5.009	5.892	7.041	13
14	4.075	4.660	5.629	6.571	7.790	14
15	4.601	5.229	6.262	7.261	8.547	15
16	5.142	5.812	6.908	7.962	9.312	16
17	5.697	6.408	7.564	8.672	10.085	17
18	6.265	7.015	8.231	9.390	10.865	18
19	6.844	7.633	8.907	10.117	11.651	19
20	7.434	8.260	9.591	10.851	12.443	20
21	8.034	8.897	10.283	11.591	13.240	21
22	8.643	9.542	10.982	12.338	14.041	22
23	9.260	10.196	11.689	13.091	14.848	23
24	9.886	10.856	12.401	13.848	15.659	24
25	10.520	11.524	13.120	14.611	16.473	25
26	11.160	12.198	13.844	15.379	17.292	26
27	11.808	12.878	14.573	16.151	18.114	27
28	12.461	13.565	15.308	16.928	18.939	28
29	13.121	14.256	16.047	17.708	19.768	29
30	13.787	14.953	16.791	18.493	20.599	30
40	20.707	22.164	24.433	26.509	29.051	40
50	27.991	29.707	32.357	34.764	37.689	50
60	35.534	37.485	40.482	43.188	46.459	60
70	43.275	45.442	48.758	51.739	55.329	70
80	51.172	53.540	57.153	60.391	64.278	80
90	59.196	61.754	65.647	69.126	73.291	90
100	67.328	70.065	74.222	77.929	82.358	100

Tabulka IV: *Kritické hodnoty χ^2-rozdělení (pokračování)*

ν	$\chi^2_{0.1}$	$\chi^2_{0.05}$	$\chi^2_{0.025}$	$\chi^2_{0.01}$	$\chi^2_{0.005}$	ν
1	2.706	3.841	5.024	6.635	7.879	1
2	4.605	5.991	7.378	9.210	10.597	2
3	6.251	7.815	9.348	11.345	12.838	3
4	7.779	9.488	11.143	13.277	14.860	4
5	9.236	11.070	12.832	15.086	16.750	5
6	10.645	12.592	14.449	16.812	18.548	6
7	12.017	14.067	16.013	18.475	20.278	7
8	13.362	15.507	17.535	20.090	21.955	8
9	14.684	16.919	19.023	21.666	23.589	9
10	15.987	18.307	20.483	23.209	25.188	10
11	17.275	19.675	21.920	24.725	26.757	11
12	18.549	21.026	23.337	26.217	28.300	12
13	19.812	22.362	24.736	27.688	29.819	13
14	21.064	23.685	26.119	29.141	31.319	14
15	22.307	24.996	27.488	30.578	32.801	15
16	23.542	26.296	28.845	32.000	34.267	16
17	24.769	27.587	30.191	33.409	35.718	17
18	25.989	28.869	31.526	34.805	37.156	18
19	27.204	30.144	32.852	36.191	38.582	19
20	28.412	31.410	34.170	37.566	39.997	20
21	29.615	32.671	35.479	38.932	41.401	21
22	30.813	33.924	36.781	40.289	42.796	22
23	32.007	35.172	38.076	41.638	44.181	23
24	33.196	36.415	39.364	42.980	45.558	24
25	34.382	37.652	40.646	44.314	46.928	25
26	35.563	38.885	41.923	45.642	48.290	26
27	36.741	40.113	43.195	46.963	49.645	27
28	37.916	41.337	44.461	48.278	50.994	28
29	39.087	42.557	45.722	49.588	52.335	29
30	40.256	43.773	46.979	50.892	53.672	30
40	51.805	55.758	59.342	63.691	66.766	40
60	74.397	79.082	83.298	88.379	91.952	60
50	63.167	67.505	71.420	76.154	79.490	50
70	85.527	90.531	95.023	100.425	104.215	70
80	96.578	101.879	106.629	112.329	116.321	80
90	107.565	113.145	118.136	124.116	128.299	90
100	118.498	124.342	129.561	135.807	140.170	100

Literatura

[1] M. Aldrin (1995). A statistical approach to the modelling of daily car traffic. *Traffic Engineering and Control*, Vol. 36, Nb. 3, pp. 489–493.

[2] J.Anděl (1985). Matematická statistika. SNTL, Alfa.

[3] V. Beneš, G. Dohnal (1993). Pravděpodobnost a matematická statistika. Vydavatelství ČVUT.

[4] P. Brémaud (1994). An Introduction to Probabilistic Modeling. Springer Verlag, New York.

[5] J.Hátle, J. Likeš (1972). Základy počtu pravděpodobnosti a matematické statistiky. SNTL/Alfa, Praha

[6] A.Rényi (1972). Teorie pravděpodobnosti. Academia, Praha.

[7] J. Seger, R. Hindls (1995). Statistické metody v tržním hospodářství. Victoria Publishing, Praha.

[8] J.Štěpán (1987). Teorie pravděpodobnosti. Matematické základy. Akademia, Praha.

[9] N.A. Weiss (1996). Elementary Statistics, Addison-Wesley Publishing Company.

[10] T.H. Wonnacott, R.J. Wonnacott (1995). Statistika pro obchod a hospodářství. (překlad z amerického originálu Introductory Statistics for Business and Economics), J. Wiley & Sons, New York.