Přednáška 08b

NÁHODNÉ VELIČINY

Opakování: pojem σ-algebry

Borelovské množiny v $\mathbb{R}^n$ je příklad σ-algebry.

Začneme borelovskými množinami v $\mathbb{R}$

$\mathcal{B}$ je systém podmnožin v $\mathbb{R}$, který je nejmenší σ-algebrou generovanou všemi otevřenými intervaly v $\mathbb{R}$, tj množinami $(-\infty,\infty)$, $(-\infty,a)$, (b,∞), (a,b) pro $a<b$.

Speciálně
- všechny otevřené množiny v $\mathbb{R}$ leží v $\mathcal{B}$ (každá je spočetným sjednocením otevřených intervalů)
- všechny uzavřené množiny v $\mathbb{R}$ leží v $\mathcal{B}$ (každá je doplňkem otevřené množiny)

Obdobně definujeme borelovské množiny v $\mathbb{R}^n$ je to systém $\mathcal{B}$ podmnožin v $\mathbb{R}^n$, který je generován otevřenými množinami tvaru

$$(a_1,b_1)\times(a_2,b_2)\times\dots\times(a_n,b_n)$$

jako σ-algebra. Speciálně otevřené a uzavřené množiny leží v $\mathcal{B}$.

Náhodné veličiny

Definice: Nechť $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor. **Náhodná veličina** X je funkce

$$X : \Omega \longrightarrow \mathbb{R}$$

taková, že $X^{-1}(B) \in \mathcal{A}$ pro každou borelovskou množinu B v $\mathbb{R}$. Funkce

$$P_X : \mathcal{B} \longrightarrow [0,1]$$

$$P_X(B) = P\, X^{-1}(B)$$

se nazývá **rozdělení** (pravděpodobnosti) **náhodné veličiny**.

Důležitá poznámka: jestliže $X^{-1}(-\infty, a] \in \mathcal{A}$, pro všechna $a \in \mathbb{R}$, pak $X^{-1}(B) \in \mathcal{A}$ pro všechny borelovské množiny, neboť systém $\mathcal{B}$ je generován množinami

$$(-\infty, a].$$

Stejně tak rozdělení náhodné veličiny P_X je určeno hodnotami funkce P_X na intervalech $(-\infty, a]$.

Náhodný vektor $X = (X_1, X_2, \dots, X_k)$ je k-tice náhodných veličin $X_j : \Omega \to \mathbb{R}$.

Protože σ-algebra borelovských množin v $\mathbb{R}^k$ je generována množinami

$$I = (-\infty, a_1] \times (-\infty, a_2] \times \dots \times (-\infty, a_k]$$

a $X^{-1}(I) = X_1^{-1}(-\infty, a_1] \cap X_2^{-1}(-\infty, a_2] \cap \dots \cap X_k^{-1}(-\infty, a_k]$ leží v $\mathcal{A}$, neboť $X_j^{-1}(-\infty, a_j]$ leží v $\mathcal{A}$; je rovněž $X^{-1}(B) \in \mathcal{A}$ pro libovolnou borelovskou podmnožinu B v $\mathbb{R}^k$.

Distribuční funkce

Definice: Distribuční funkce náhodné veličiny X je funkce

$$F_X : \mathbb{R} \to [0,1]$$

definovaná pro všechna $x \in \mathbb{R}$ jako

$$F_X(x) = P\left(X^{-1}(-\infty, x]\right) = P(X \le x)$$

Distribuční funkce náhodného vektoru $X : \Omega \to \mathbb{R}^k$ je funkce $F_X : \mathbb{R}^k \to [0,1]$ definovaná

$$F_X(x_1, x_2, \dots x_k) = P(X_1 \le x_1, X_2 \le x_2, \dots X_k \le x_k).$$

Důležité upozornění

Někdy se definuje

$$F_X(x) = P(X < x)$$

s ostrou nerovností (např. v textu prof. Slováka).

To má vliv význam pouze pro jedinou vlastnost distribuční funkce – podle naší definice je spojitá zprava v každém x, podle druhé definice je spojitá zleva. Časem ukážeme na příkladech.

Věta (Vlastnosti distribuční funkce)

Pro každou náhodnou veličinu má její distribuční funkce $F : \mathbb{R} \to [0,1]$ tyto vlastnosti:

(1) F je neklesající

(2) F je zprava spojitá

(3) F má limitu zleva

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

(5) $P(X = x) = F(x) - \lim_{y \to x-} F(y)$

(6) Distribuční funkce má nejvýše spočetně mnoho bodů nespojitosti.

Důkaz provádím na přednášce.

V knize prof. Slováka ho lze najít za větou 9.20 na straně 541. Vzhledem k odlišné definici distribuční funkce se liší tvrzení (2), (3) a (5). Modifikace důkazu pro naši definici by neměla činit potíže.

Speciální typy náhodných veličin
– diskrétní a spojité

Diskrétní náhodné veličiny

Jestliže náhodná veličina X nabývá pouze konečně mnoha nebo spočetně mnoha reálných hodnot, nazýváme ji diskrétní.

Definujeme pravděpodobnostní funkci $f: \mathbb{R} \to [0,1]$

$$f(x) = P(X=x).$$

Tedy f bude nenulová pouze na konečné nebo spočetné množině $x_1, x_2, x_3, \dots$

Platí
$$\sum_i f(x_i) = 1$$

a dále
$$F_X(x) = \sum_{x_i \le x} f(x_i)$$

$$P_X(B) = \sum_{x_i \in B} f(x_i)$$

Příklady diskrétních rozdělení

(1) Degenerované rozdělení

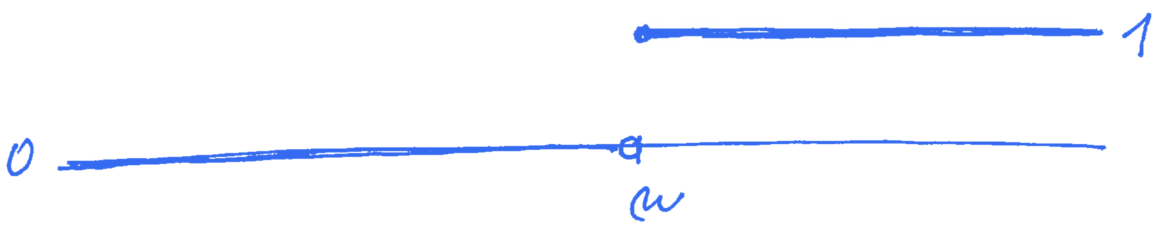
Náhodná veličina $X : \Omega \to \mathbb{R}$ nabývá jediné hodnoty μ. Její rozdělení se nazývá degenerované a značí se $Dg(\mu)$.

Pravděpodobnostní funkce je

$$f(x) = \begin{cases} 1 & x = \mu \\ 0 & x \neq 0 \end{cases}$$

Distribuční funkce je

$$F(x) = \begin{cases} 0 & \text{pro } x < \mu \\ 1 & \text{pro } x \geq \mu \end{cases}$$



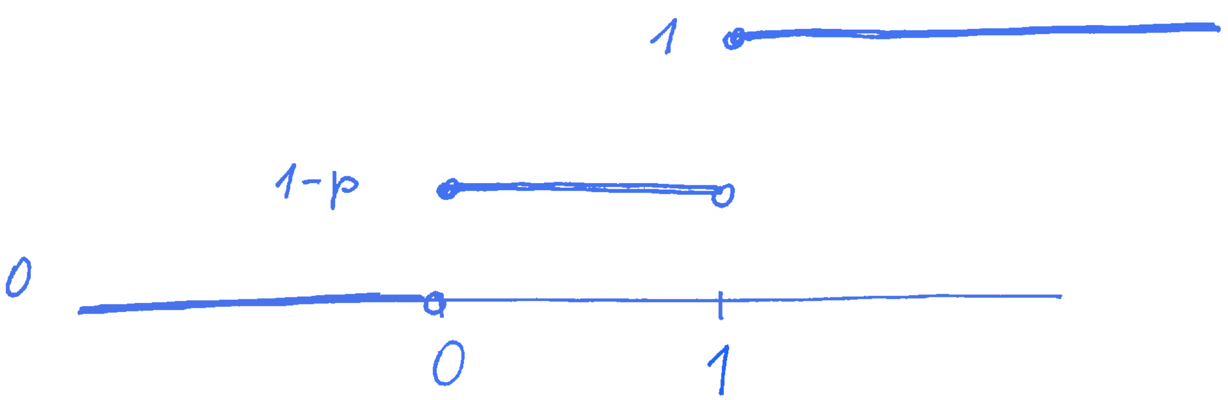
(2) Alternativní rozdělení

Náhodná veličina X nabývá hodnoty 0 s pravděpodobností $1-p$ a hodnoty 1 s pravděpodobností p. Rozdělení této náhodné veličiny značíme $A(p)$. Její pravděpodobnostní funkce je

$$f(x) = \begin{cases} 1-p & x = 0 \\ p & x = 1 \\ 0 & x \neq 0, x \neq 1 \end{cases}$$

a distribuční funkce

$$F(x) = \begin{cases} 0 & x < 0 \\ 1-p & x \in [0,1) \\ 1 & x \in [1,\infty) \end{cases}$$



(3) Binomické rozdělení

je rozdělení náhodné veličiny, která nabývá hodnot $0, 1, 2, \dots, n$ a má pravděpodobnostní funkci

$$f(t) = \begin{cases} \binom{n}{t} p^t (1-p)^{n-t} & \text{pro } t \in \{0, 1, \dots, n\} \\ 0 & \text{jinak} \end{cases}$$

Značíme $Bi(n,p)$

Binomické rozdělení a alternativní rozdělení

Pokus, který má výsledek 0 nebo 1, první s pravděpodobností $1-p$, druhé s pravděpodobností p je možné považovat za náhodnou veličinu s alternativním rozdělením $A(p)$. Provádějme tento pokus n krát a jako výsledek berme součet hodnot. Tedy dostáváme novou náhodnou

veličinu s možnými výsledky $0, 1, 2, \ldots n$.
Pravděpodobnost výsledku $t \in \{0, 1, 2, \ldots, n\}$
je
$$\binom{n}{t} p^t (1-p)^{n-t}$$
Tedy jde o veličinu s binomickým rozdělením $Bi(n, p)$.

(4) Geometrické rozdělení

Nechť X je náhodná veličina, která nabývá hodnot $0, 1, 2, \ldots$ a má pravděpodobnostní funkci
$$\begin{aligned} f(t) = P(X=t) &= p(1-p)^t \qquad t \in \{0,1,2,\ldots\} \\ &= 0 \qquad \text{jinak} \end{aligned}$$
Její rozdělení se nazývá geometrické $(Ge(p))$ (použití u tzv. Einsteinovy - Boseovy statistiky)

Uvažujme nezávislé prováděné pokusy s alternativním rozdělením $A(p)$. Budeme pokračovat tak dlouho dokud nenastane zdar.
Náhodná veličina popisující počet pokusů před tím, než nastane zdar má
$$\begin{aligned} f(0) &= p \\ f(1) &= (1-p)p \\ f(2) &= (1-p)^2 p \end{aligned}$$

tedy má geometrické rozdělení.

(5) Poissonovo rozdělení $Po(\lambda)$

je rozdělení náhodné veličiny, která nabývá hodnot $0, 1, 2, \dots$ s pravděpodobnostní funkcí

$$f(t) = P(X=t) = \begin{cases} \dfrac{\lambda^t}{t!} e^{-\lambda} & \text{pro } t \in \mathbb{N} \cup \{0\} \\ 0 & \text{jinak} \end{cases}$$

Model pro Poissonovo rozdělení

Náhodnou veličinu, která popisuje počet předmětů v pevně zvolené přihrádce. Uvažujme n přihrádek, r předmětů a pravděpodobnost, že se předmět dostane do vybrané přihrádky je $\frac{1}{n}$.

možné výsledky jsou $0, 1, 2, \dots, r$ s pravděpodobnostmi

$$P(X=k) = \binom{r}{k} \left(\frac{1}{n}\right)^k \left(1-\frac{1}{n}\right)^{r-k} = \binom{r}{k} \frac{(n-1)^{r-k}}{n^r}$$

jde o rozdělení $Bi(r, \frac{1}{n})$.

Co se bude dít, jestliže bude stoupat číslo n i číslo r tak, že poměr $\frac{r}{n}$ se bude blížit číslu $\lambda > 0$.

$$\lim_{n\to\infty} P(X_{n,r_n} = k) = \lim_{n\to\infty} \binom{r_n}{k} \frac{(n-1)^{r_n-k}}{n^{r_n}} =$$

$$= \lim_{n \to \infty} \frac{r_n(r_n-1)\dots(r_n-k+1)}{(n-1)^k} \frac{1}{k!} \left(1-\frac{1}{n}\right)^{r_n} =$$

$$= \frac{\lambda^k}{k!} \lim_{n\to\infty} \left(1 + \frac{-\frac{r_n}{n}}{r_n}\right)^{r_n} = \frac{\lambda^k}{k!} e^{-\lambda}$$

Tedy Poissonovo rozdělení $Po(\lambda)$ je dobře aproximováno binomickým rozdělením $Bi(n, p_n)$ s $np_n = \lambda$ na velká n.

Spojité náhodné veličiny

jsou náhodné veličiny, jejichž distribuční funkci lze psát

$$F(x) = \int_{-\infty}^{x} f(t)\,dt.$$

Funkce f se nazývá hustota pravděpodobnosti.

Distribuční funkce $F(x)$ je diferencovatelná a

$$F'(x) = f(x), \quad \text{je-li } f \text{ spojitá.}$$

Příklady spojitých rozdělení

(1) Rovnoměrné rozdělení

je rozdělení s následující hustotou a distribuční funkcí pro $-\infty < a < b < \infty$.

$$f(t) = \begin{cases} 0 & t < a \\ \frac{1}{b-a} & t \in [a,b) \\ 0 & t \geq b \end{cases} \qquad F(x) = \begin{cases} 0 & t < a \\ \frac{t-a}{b-a} & t \in [a,b) \\ 1 & t \geq b \end{cases}$$

(2) <u>Exponenciální rozdělení</u>

je rozdělení s hustotou

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & \text{pro } t \geq 0 \\ 0 & \text{pro } t < 0 \end{cases}$$

Označuje se $ex(\lambda)$.

(3) <u>Gama rozdělení</u> je rozdělení s hustotou tvaru $c\, x^{a-1} e^{-bx}$ pro $x > 0$ s danými konstantami $a > 0$, $b > 0$. c volíme tak, aby

$$c \int_0^\infty x^{a-1} e^{-bx}\, dx = 1$$

Tedy při substituci $t = bx$

$$1 = c \int_0^\infty x^{a-1} e^{-bx}\, dx = c \int_0^\infty \left(\frac{t}{b}\right)^{a-1} e^{-t} \frac{1}{b}\, dt =$$

$$= \frac{c}{b^a}\, \Gamma(a).$$ Proto $$c = \frac{b^a}{\Gamma(a)}$$

<u>Gama rozdělení $\Gamma(a, b)$</u> s parametry $a > 0$, $b > 0$ má hustotu

$$f(x) = \begin{cases} \dfrac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} & \text{pro } x > 0 \\ 0 & \text{pro } x \leq 0 \end{cases}$$

Exponenciální rozdělení je speciální případ Gama rozdělení $ex(\lambda) = \Gamma(1, \lambda)$.

(4) Normální rozdělení

je rozdělení náhodné veličiny s hustotou

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

a příslušnou distribuční funkcí

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt .$$

Označujeme $N(0,1)$ a hovoříme o standardizovaném normálním rozdělení.

Musíme ukázat, že

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

To lze následujícím „trikem"

$$\int_{\mathbb{R}^2} e^{-(x^2+y^2)/2}\, dx\, dy = \left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx\right)\left(\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy\right)$$

ale současně pomocí polárních souřadnic

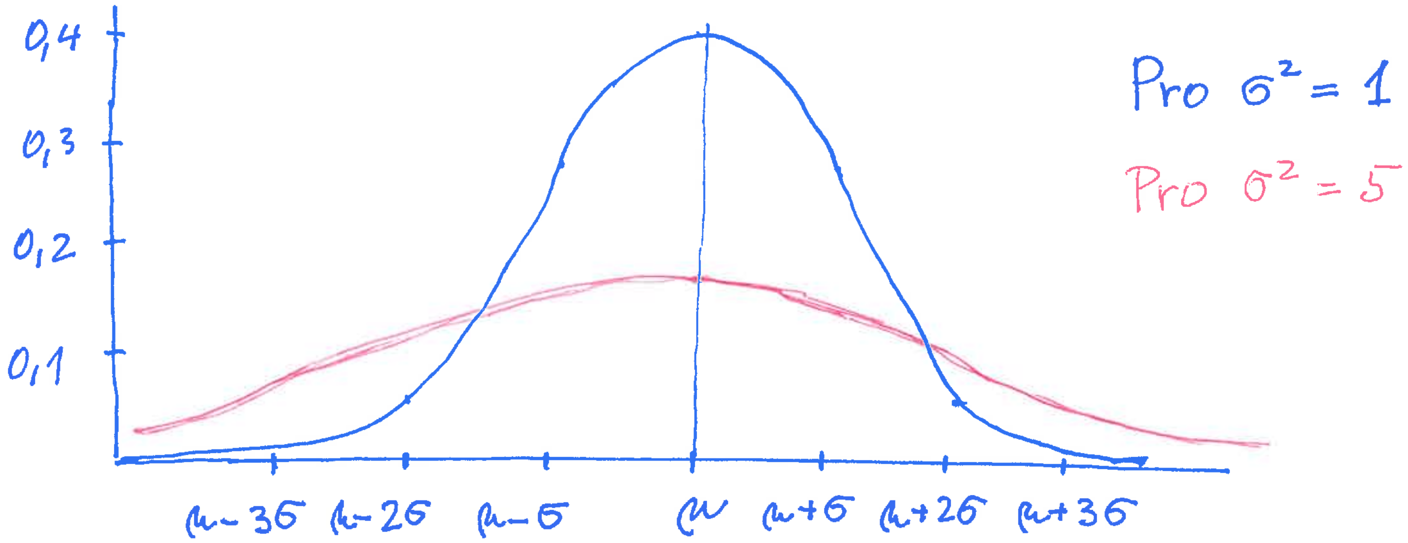
$$\int_{\mathbb{R}^2} e^{-(x^2+y^2)/2}\, dx\, dy = \int_0^{2\pi}\left(\int_0^{\infty} r\, e^{-r^2/2}\, dr\right) d\alpha$$

$$= \int_0^{2\pi}\left(\int_0^{\infty} e^{-t}\, dt\right) d\alpha = 2\pi$$

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

Zavedení dvou parametrů. Normální rozdělení $N(\mu, \sigma)$ je rozdělení s hustotou

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



Souvislost s binomickým rozdělením pro $n \to \infty$.