Pravděpodobnost & Statistika 1

definice

Prostor jevů

$\mathcal{F} \subseteq \mathcal{P}(\Omega)$ je prostor jevů (též σ-algebra), pokud:

1) $\emptyset \in \mathcal{F}$ a $\Omega \in \mathcal{F}$

2) $A \in \mathcal{F} \Rightarrow \Omega / A \in \mathcal{F}$ $= A^c$ neboli A-complement

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

Axiomy pravděpodobnosti

$P : \mathcal{F} \to [0,1]$ se nazývá pravděpodobnost, pokud:

1) $P(\emptyset) = 0$

2) $P(\Omega) = 1$

3) $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ pro libovolnou posloupnost po-dvou disjunktních jevů

Pravděpodobnostní prostor

je trojice $(\Omega, \mathcal{F}, P)$ taková, že:

1) $\Omega \neq \emptyset$ je libovolná množina
2) $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ je prostor jevů
3) P je pravděpodobnost

Podmíněná pravděpodobnost

Pokud $A, B \in \mathcal{F}$ a $P(B) > 0$, definujeme podmíněnou pravděpodobnost A při B jako:

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

Platí, že pro $\underbrace{Q(A) := P(A \mid B)}_{\text{„podmíněnou pstní fci"}}$ je $(\Omega, \mathcal{F}, Q)$ pravděpodobnostní prostor.

Rozklad Ω

"rozbor všech možností"

Spočetný systém množin $B_1, B_2, \ldots \in \mathcal{F}$ je rozklad Ω, pokud:

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

2) $\bigcup B_i = \Omega$

Nezávislost jevů

Jevy $A, B \in \mathcal{F}$ jsou nezávislé, pokud:

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

Pak, pokud je $P(B) > 0$, platí:

$$P(A|B) = P(A)$$

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

Nezávislost více jevů

Jevy $\{A_i : i \in I\}$ jsou vzájemně nezávislé, pokud pro každou konečnou množinu $J \subseteq I$ platí:

$$P\Big(\bigcap_{i \in J} A_i\Big) = \prod_{i \in J} P(A_i)$$

(*) odkaz na def. nez. jevů
„průnik ∀ se spočítá vynásobením ∀

Pokud podmínka platí jen pro dvouprvkové množiny J, nazýváme jevy $\{A_i\}$ po dvou nezávislé.

Náhodná veličina

Mějme pravděpodobnostní prostor $(\Omega, \mathcal{F}, P)$. Funkci $X : \Omega \to \mathbb{R}$ nazveme diskrétní náhodná veličina, pokud $\mathrm{Im}(X)$ = obor hodnot X je spočetná množina a pokud pro ∀ $x \in \mathbb{R}$ platí:

aby ta veličina byla diskrétní

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

Ω je vlastně $\{X^{-1}(x)\}$, $P(X^{-1}(x))$ ale používáme zapsanou jako $P(X = x)$

Pravděpodobnostní funkce

P. fce. diskrétní náhodné veličiny X je funkce $P_X : \mathbb{R} \to [0,1]$ taková, že:

$$P_X(x) = P(X=x)$$ = v podstatě hustotní fce pro d.n.v.

1) $\sum_{\forall x \in \mathrm{Im}(X)} P_X(x) = 1$

$A \subseteq \mathbb{R},\ f: A \to [0,1],\ \sum_{a \in \mathrm{Im}(A)} f(a) = 1$

2) pstní prostor dokážeme vyrobit z lib. podmnožiny $\mathbb{R}$ a vhodné fce

---

Střední hodnota

„popisuje očekávanou hodnotu n.v."

Pokud je X diskrétní náhodná veličina, tak je její střední hodnota označována $\mathbb{E}(X)$ a definovaná:

$$\mathbb{E}(X) := \underbrace{\sum_{x \in \mathrm{Im}(X)} x \cdot P(X=x)}_{\text{pokud má součet smysl.}} = \sum_{\omega \in \Omega} X(\omega) \cdot P(\{\omega\})$$

pokud má součet smysl.

(*) $\mathrm{Im}(X)$ musí být spočetná, jinak hrozí („$+\infty - \infty$")

Podmíněná střední hodnota

Pokud je X diskrétní náhodná veličina a $P(B) > 0$, tak podmíněná střední hodnota X za předpokladu B je:

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

pokud má součet smysl

Rozptyl

"popisuje odchylku hodnoty n.v. od její střední hodnoty"

Variance n.v. X nazveme číslo: "aka střední kvadratická odchylka"

$$\mathrm{var}(X) = \mathbb{E}\left((X - \mathbb{E}X)^2\right)$$

musí platit: $\mathrm{var}(X) \geq 0$ $\mathrm{var}(X) = 0 \iff P(X \neq \mathbb{E}(X)) = 0$

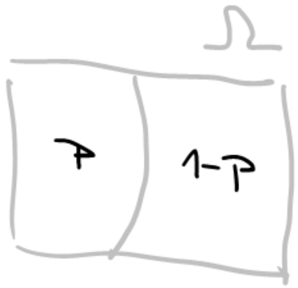
Bernoulliho (alt.) rozdělení

"X = počet orlů při jednom hodu nespravedlivou mincí"

$$X \sim \mathrm{Bern}(p) \qquad P_X(1) = p$$

$$\mathbb{E}(X) = p \qquad P_X(0) = 1 - p$$

$$\mathrm{var}(X) = p(1-p) \qquad P_X(k) = 0 \quad \forall k \neq 0, 1$$



---

pro libovolný jev $A \in \mathcal{F}$ definujeme indikátorovou n.v. I_A:

$$I_A(\omega) = \begin{cases} 1 & \text{pokud } \omega \in A \\ 0 & \text{jinak} \end{cases}$$

$$I_A \sim \mathrm{Bern}(P(A))$$

Binomiální rozdělení

„X = počet orlů při n hodech nespravedlivou mincí"

$X \sim Bin(n,p)$ n := počet hodů, p := pravd. že padne orel

→ $p \in [0,1]$

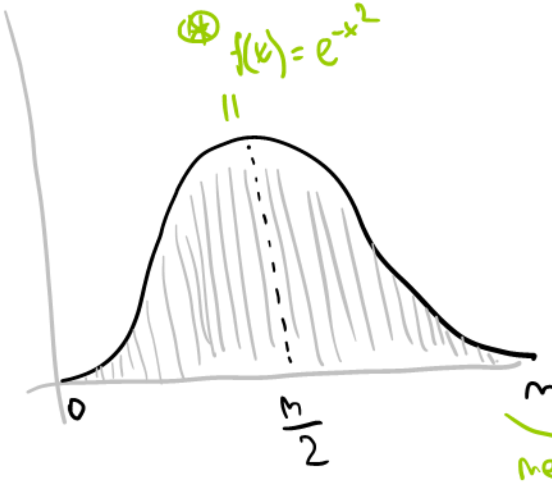
$$p_X(k) = \underbrace{\binom{n}{k} p^k}_{\text{\# způsobů jak uspět } * \text{ P(že k-krát uspěju)}} \cdot \underbrace{(1-p)^{n-k}}_{\text{P(že (n-k)-krát neuspěju)}}$$ pro $k \in \{0, 1, \dots, n\}$

⊛ pravd. fce: z n hodů k orlů

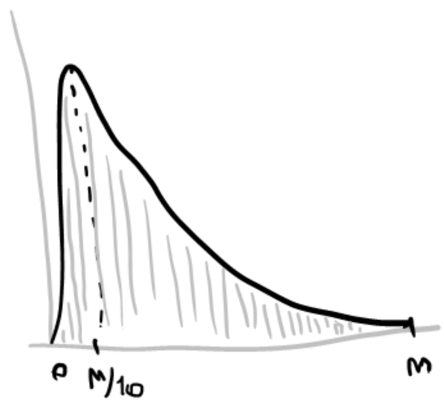
$$\mathbb{E}(X) = np$$

$$var(X) = np(1-p)$$

$p = \frac{1}{2}$:



$p = \frac{1}{10}$:



nejsou spojité, ale líp se to nakreslí...

Poissonovo rozdělení

"X = počet emailů, které dostaneme za časový úsek"

$X \sim \text{Pois}(\lambda)$ $\quad \lambda \in \mathbb{R}, \lambda > 0$

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

$$\mathbb{E}(X) = \lambda$$

$$\text{var}(X) = \lambda$$

(*) $\text{Pois}(\lambda)$ je limitou:

$$\text{Bin}\left(n, \frac{\lambda}{n}\right)$$

máme-li $X_n \sim \text{Bin}(n, \frac{\lambda}{n})$ pak:

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

$$\underbrace{\frac{\lambda^k}{k!}}_{\text{konst}} \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\text{konv} \to e^{-\lambda}} \cdot \underbrace{\frac{n(n-1)\dots(n-k+1)}{n^k}}_{\text{konv.} \to 1} \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-k}}_{\text{konv} \to 1}$$

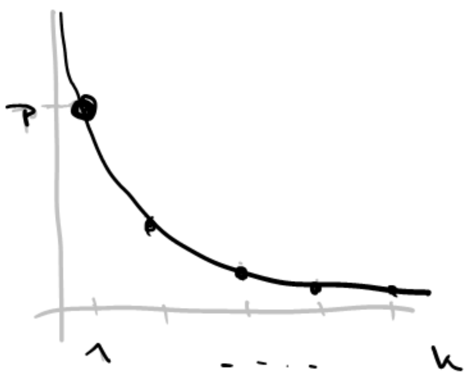
Geometrické rozdělení „X = orel poprvé padl k-tým hodem"

$X \sim Geom(p)$ $\quad p \in [0,1]$

$P_X(k) = (1-p)^{k-1} \cdot p$ $\quad k = 1, 2, 3, \ldots$ $\quad$ * pro $p = \frac{1}{2}$ je $P_X(k) = p^k$

$\mathbb{E}(X) = \frac{1}{p}$

$var(X) = \frac{1}{p}\left(\frac{1-p}{p}\right)$

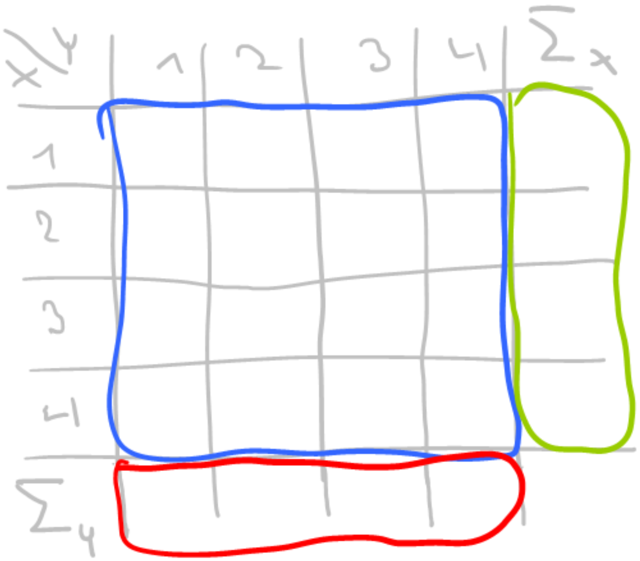


Sdružené rozdělení

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

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

(*) definice funguje i pro více než dvě náhodné veličiny: $P_{X_1,\dots,X_n}(x_1,\dots,x_n)$



(*) sdružené rozdělení $P_{X,Y}$

(*) marginální rozdělení P_X

(*) marginální rozdělení P_Y

$\Rightarrow$

1) z $P_{X,Y}$ lze získat P_X a P_Y

2) z P_X, P_Y nelze získat $P_{X,Y}$

Nezávislost náh. veličin

Diskrétní m.v. X, Y jsou nezávislé, pokud pro $\forall x, y \in \mathbb{R}$ jsou jevy $\{X=x\}$ a $\{Y=y\}$ nezávislé, což platí právě když:

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

(*) v případě nezávislých X, Y lze def. sdružené rozd. $P_{X,Y} \leftarrow P_X, P_Y$ právě jako jejich součin

---

Směrodatná odchylka

Pro m.v. X definujeme její směrodatnou odchylku jako:

$$\sigma_X = \sqrt{var(X)}$$

Obecná náhodná veličina

Náhodná veličina na (Ω, F, P) je zobrazení $X: \Omega \to \mathbb{R}$, které pro $\forall x \in \mathbb{R}$ splní:

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

(*) diskrétní m.v. je obecná m.v.

Distribuční funkce

"kumulativní" protože se kumuluje (nasčítá) na $\underline{\underline{1}}$

Dist. fce. náhodné veličiny X je funkce:

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

- F_X je neklesající fce
- $\lim_{x \to -\infty} F_X(x) = 0$
- $\lim_{x \to +\infty} F_X(x) = 1$
- F_X je zprava spojitá

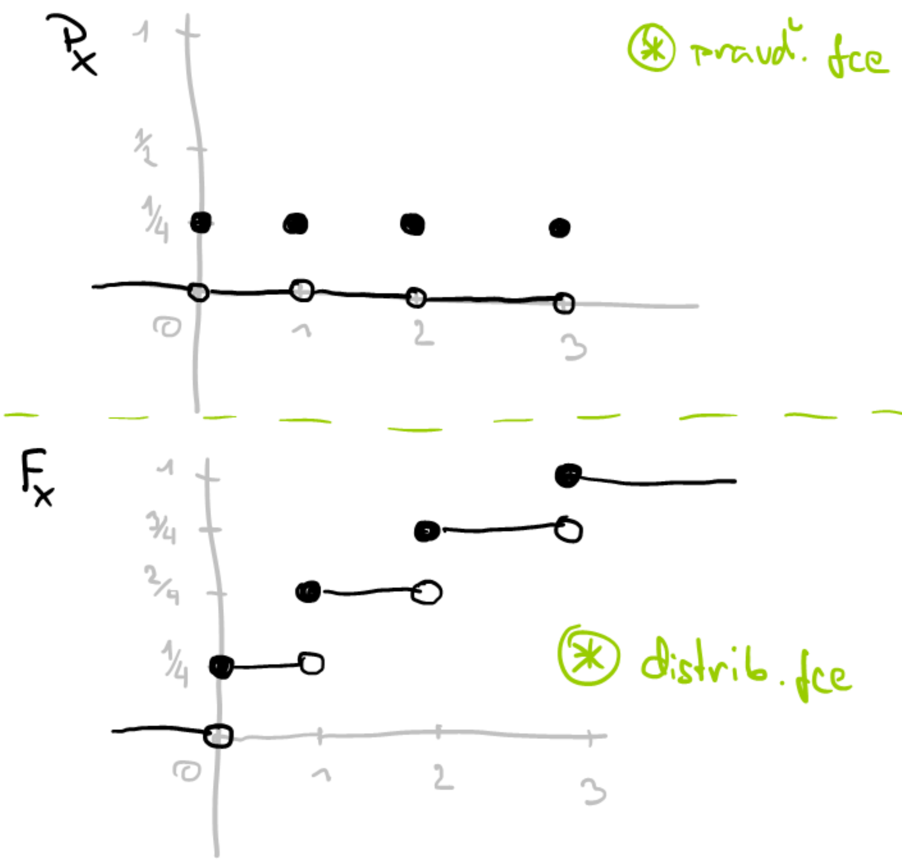
protože jdeme zprava a dojdeme až do bodu, když zleva, v bodě skok

$P(X \le x + \varepsilon) \to P(X \le x)$ pro $\varepsilon \to 0^+$

$x \to +\infty$ do Ω přibývají $\forall$ body $\Rightarrow P(\Omega) = 1$

$x \to -\infty$ z Ω mizí body $\Rightarrow P(\emptyset) = 0$

příklad: $X \sim$ uniformní $(\{0,1,2,3\})$



Spojitá náhodná veličina

n.v. X se nazývá spojitá, pokud existuje nezáporná reálná funkce f_X tak, že:

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

(*) P_X pravděpodobnostní fce
F_X distribuční fce
f_X hustotní fce

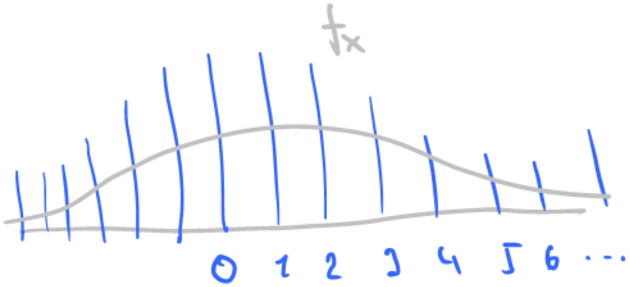
funkce f_X se nazývá HUSTOTA

Střední hodnota spojité n.v.

Nechť spojitá n.v. X má hustotu f_X. Pak její střední hodnota je označována $E(X)$ a definována:

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

má-li integrál smysl (*) tedy není „$+\infty - \infty$"



$$E(X) = \sum_{n} \int_{n}^{n+1} x \cdot f_X(x)$$ -> platí rovnost protože integrujeme přes celou R, ale po částech

$$\doteq \sum_{n} n \cdot \int_{n}^{n+1} f_X(x) = \sum_{n} n \cdot P(n \le X \le n+1)$$ -> n.v. zaokrouhlená dolů na celé číslo

=> pro kroky v Σ -> 0,00....1 -> se blížíme k $\sum x \cdot P(X=x)$

Rozptyl spojité n.v.

Víme:

"1. moment" ... $\mathbb{E}(X) = \int_{-\infty}^{+\infty} x \cdot f_X(x)\,dx$

"2. moment" ... $\mathbb{E}(X^2) = \int_{-\infty}^{+\infty} x^2 \cdot f_X(x)\,dx$ ... spojitý LOTUS pro $g(X) = x^2$

Proto označme-li $\mu = \mathbb{E}(X)$, tak:

$$\mathrm{var}(X) \overset{def}{=} \int_{-\infty}^{+\infty} (x-\mu)^2 \, f_X(x)\,dx = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

$$\mathbb{E}(X-\mu)^2 = \mathbb{E}(X-\mathbb{E}(X))^2$$

$$\mathbb{E}(aX) = a\,\mathbb{E}(X)$$

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

(*) rozptyl počítáme "jako u diskrétních n.v.", nejpohodlnější bývá

"k-tý moment" ... $\mathbb{E}(X^k) = \int_{-\infty}^{+\infty} x^k \cdot f_X(x)\,dx$

Uniformní rozdělení

$$X \sim U(a,b)$$

$$F_X = \begin{cases} 0 & \text{pro } x \leq a \\ \frac{x-a}{b-a} & \text{pro } a < x < b \\ 1 & \text{pro } x \geq b \end{cases}$$

$$f_X = \frac{1}{b-a} \quad \text{pro } x \in [a,b]$$

$$\mathbb{E}(X) = \frac{a+b}{2}$$

$$= \int_a^b x \cdot \frac{1}{b-a} = \left[\frac{x^2}{2}\right]_a^b \cdot \frac{1}{b-a} = \left(\frac{b^2}{2} - \frac{a^2}{2}\right) \cdot \frac{1}{b-a} = \frac{\overbrace{b^2 - a^2}^{(b-a)(b+a)}}{2(b-a)} = \underline{\underline{\frac{a+b}{2}}}$$

$$\text{var}(X) = \frac{(b-a)^2}{12} \qquad \sigma_X = \frac{b-a}{2} \cdot \frac{1}{\sqrt{3}} = \sqrt{\text{var}(X)}$$

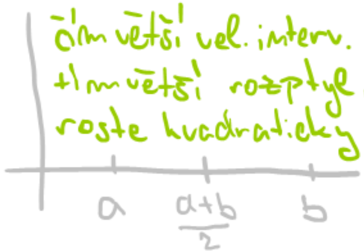
$$= \int_{-\infty}^{+\infty} (x-\mu)^2 \cdot f_X(x) \cdot dx \quad \ldots\ldots \quad \mathbb{E}(X) = \mu$$

$$= \int_a^b \underbrace{\left(x - \frac{b-a}{2}\right)^2}_{y = \left(x - \frac{a+b}{2}\right)^2} \cdot \frac{1}{b-a} \cdot dx = \int_{-\frac{b-a}{2}}^{\frac{b-a}{2}} y^2 \cdot \frac{1}{b-a} dy = \left[\frac{y^3}{3}\right]_{-\frac{b-a}{2}}^{\frac{b-a}{2}} \cdot \frac{1}{b-a} = \left(\frac{(b-a)^3}{3\cdot 8} + \frac{(b-a)^3}{3\cdot 8}\right) \cdot \frac{1}{b-a} = \frac{2(b-a)^3}{24} \cdot \frac{1}{b-a}$$

$$= \underline{\frac{(b-a)^2}{12}}$$

(*)

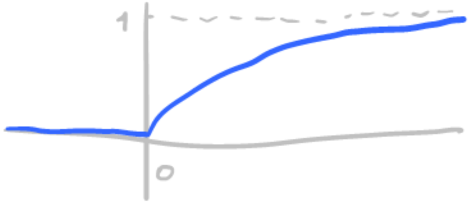
čím větší vel. interv. tím větší rozptyl, roste kvadraticky



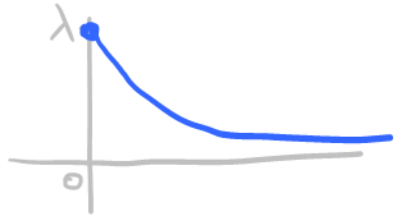
Exponenciální rozdělení

$X \sim Exp(\lambda)$ pro $\lambda > 0$

$$F_X(x) = \begin{cases} 0 & \text{pro } x \le 0 \\ 1 - e^{-\lambda \cdot x} & \text{pro } x \ge 0 \end{cases}$$



$$f_X(x) = \begin{cases} 0 & \text{pro } x \le 0 \\ F_X' = \lambda \cdot e^{-\lambda x} & \text{pro } x \ge 0 \end{cases}$$



jak se počítá: $\int f_X = \int_0^\infty \lambda e^{-\lambda x} = \left[-e^{-\lambda x}\right]_0^\infty = 0 - (-1) = 1$

$$\mathbb{E}(X) = \frac{1}{\lambda}$$

PER PARTES

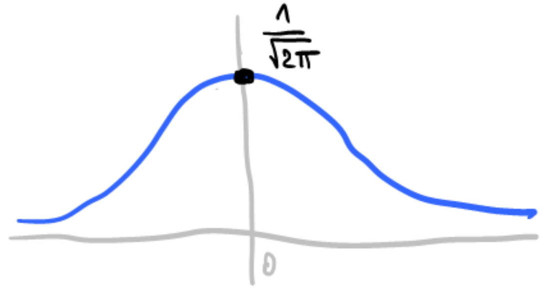
$$= \int_0^\infty \underbrace{x}_{\text{der.}} \cdot \underbrace{\lambda e^{-\lambda x}}_{\text{integ.}} = \underbrace{\left[x \cdot (-e^{-\lambda x})\right]_0^\infty}_{0} - \int_0^\infty 1 \cdot (-e^{-\lambda x}) = \int_0^\infty e^{-\lambda x} = \frac{1}{\lambda}$$

$$\mathbb{E}(X^2) = \int_0^\infty x^2 \cdot \lambda e^{-\lambda x} = \left[x^2 \cdot (-e^{-\lambda x})\right]_0^\infty - \int_0^\infty 2x \cdot (-e^{-\lambda x}) = \int_0^\infty 2x \cdot e^{-\lambda x} = \frac{2}{\lambda} \cdot \int_0^\infty x \cdot \lambda e^{-\lambda x} = \frac{2}{\lambda^2}$$

$$\mathrm{var}(X) = \frac{1}{\lambda^2} = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

$$\sigma_X = \frac{1}{\lambda}$$

Normální rozdělení



⊛ sudá fce

$$Z \sim \text{Norm}(\overset{\mathbb{E}(Z)}{0}, \overset{var(Z)}{1})$$

$F_Z(z) = \Phi(z)$ = primitivní fce k $\varphi(z)$, nejde vyjádřit pomocí elem. fcí

$f_Z(z) = \varphi(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}}$ proč je $\varphi(z)$ hust. fce? $\int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}} = C = \sqrt{2\pi} \Rightarrow \int_{-\infty}^{+\infty} \varphi(z) = 1$

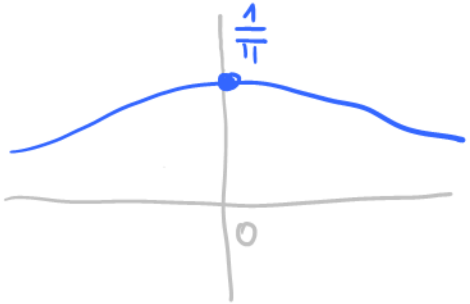
$\mathbb{E}(Z) = 0 = \int_{-\infty}^{+\infty} \underbrace{z \cdot \overbrace{\varphi(z)}^{\text{sudá}}}_{\text{lichá}} = 0$

???

$var(Z) = 1 = \mathbb{E}(Z^2) - \underbrace{\mathbb{E}(Z)^2}_{0} = \mathbb{E}(Z^2) = \int_{-\infty}^{+\infty} z \cdot \varphi(z) = \int_{-\infty}^{+\infty} z \cdot \frac{1}{\sqrt{2\pi}} \cdot z \cdot e^{-\frac{z^2}{2}} = \underbrace{\left[\frac{z}{\sqrt{2\pi}} \cdot -e^{-\frac{z^2}{2}}\right]_{-\infty}^{+\infty}}_{=0} + \int_{-\infty}^{+\infty} 1 \cdot \varphi(z) = 1$

Cauchyho rozdělení

$$f_X(x) = \frac{1}{\pi(1+x^2)}$$



$\mathbb{E}(X)$ neexistuje   nelze zintegrovat „$\infty - \infty$"

Gamma rozdělení, Beta rozdělení, χ^2 rozdělení s n°volnosti, Studentovo t-rozdělení

(✱) budu dělat až ve statistice

Sdružená distribuční fce

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

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

píšeme $F_{X,Y} = P(X \leq x \ \& \ Y \leq y)$

- Dá se definovat: pro více než dvě n.v.

$$F_{X_1,\dots,X_n}(x_1,\dots,x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$$

Sdružená hustota

- sdruženou distribuční funkci můžeme často psát jako integrál přes nezápornou fci $f_{X,Y}$:

$$F_{X,Y}(x,y) = \underbrace{\int_{-\infty}^{x}}_{s=-\infty,\dots,x} \underbrace{\int_{-\infty}^{y}}_{t=-\infty,\dots,y} f_{X,Y}(s,t)\,.dt\,.ds$$

(*) tedy vyjadřujeme distrib. fci jako dvojitý integrál přes hustotní fci

- Potom nazýváme n.v. X,Y sdružené spojité. Funkce $f_{X,Y}$ je jejich sdružená hust. fce.
- $f_{X,Y} > 1$ je validní

(**) Fubiniho věta ⇒ mohu (jsou-li splněny podm.) prohodit integrály

$$F_{X,Y}(x,y) = \int_{t=-\infty}^{y} \int_{s=-\infty}^{x} f_{X,Y}(s,t)\cdot ds\,.dt$$

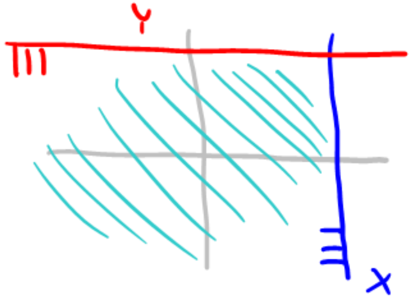
Nezávislost spojitých m.v.

Libovolné m.v. nazveme nezávislé, pokud jevy $\{X \le x\}$ a $\{Y \le y\}$ jsou nezávislé pro libovolná $x, y \in \mathbb{R}$. Ekvivalentně:

dva způsoby jak zapsat totéž

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

$$F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y)$$



Podmiňování

Platí věta, že distrib. fce je primitivní funkcí k hustotní fci. Proto

X je m.v. na $(\Omega, \mathcal{F}, P)$ a $B \in \mathcal{F}$

$$F_{X|B}(x) := P(X \le x \mid B) = \frac{P(X \le x \,\&\, B)}{P(B)} = \int_{-\infty}^{x} f_{X|B}(s)$$

Podmíněná hustota

chceme pro $D = \{Y = y\}$

umět: $f_{X|D}$ ale $P(D) = 0$

proto musíme jinak...

Pro spojité n.v. X,Y definujeme podmíněnou hustotu předpisem:

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

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

$$= \frac{f_{X|Y}(x,y)}{\int_{x\in\mathbb{R}} f_{X|Y}(x,y)\,dx}$$

analogie s diskrétními veličinami, vydělením $f_Y(y)$ "zafixujeme" y a "sčítáme" tak přes x

Kovariance

$$X - \mathbb{E}X = \begin{cases} + & \dots X > \mathbb{E}X \\ - & \dots X < \mathbb{E}X \end{cases}$$

& $Y - \mathbb{E}Y$ analogicky

Pro n.v. X,Y definujeme jejich kovarianci předpisem:

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

Korelace

Korelace n.v. X, Y je definována předpisem:

$$\operatorname{corr}(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var}(X)\cdot\operatorname{var}(Y)}}$$

někdy je značí $\rho(X,Y)$

korelace je přenormovaná kovariance tak, aby:

$$-1 \leq \operatorname{corr}(X,Y) \leq 1$$

víme:

$$\operatorname{corr}(X,Y) = \begin{cases} 1 & \text{pro } X = Y \\ (-1,1) & \text{jinak} \\ -1 & \text{pro } X = -Y \end{cases}$$