Přednáška 09a+b

Lepší model pro Poissonovo rozdělení

Ve stanoveném zápase padne v průměru λ gólů ($\lambda > 0$). Náhodná veličina X = počet gólů v daném zápase má Poissonovo rozdělení.

Rozdělme zápas na λn stejně dlouhých částí. Pravděpodobnost, že v jedné části padne gól je $\frac{1}{n}$. Pravděpodobnost, že v zápase padne k gólů je

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

Spočítejme limitu pro $n \to \infty$:

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

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

$$= \frac{1}{k!}\lambda^k e^{-\lambda}\cdot 1 = \frac{\lambda^k}{k!}e^{-\lambda}$$

Rozdělení náhodných vektorů

Uvažujme náhodný vektor $X = (X_1, X_2, \dots, X_n)$
Jeho distribuční funkce je
$$F_X : \mathbb{R}^n \longrightarrow [0,1]$$
definovaná
$$F_X(x_1, x_2, \dots x_n) = P(X_1 \leq x_1 \wedge X_2 \leq x_2 \wedge \dots \dots X_n \leq x_n)$$
Pro diskrétní náhodný vektor, který nabývá ~~nenulových~~ pouze konečně nebo spočetně mnoha hodnot je pravděpodobnostní funkce
$$f(x_1, x_2, \dots x_n) = P(X_1 = x_1 \wedge X_2 = x_2 \wedge \dots)$$
Spojitý náhodný vektor je takový, že
$$F_X(x_1, x_2, \dots x_n) = \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_1} f(t_1, t_2, \dots t_n)\, dt_1\, dt_2 \dots dt_n$$
pro nějakou funkci f, kterou nazýváme hustotou pravděpodobnosti.

<u>Marginální rozložení</u> obdržíme z rozdělení náhodného vektoru tak, že přes jednu složku provedeme sčítání nebo integrování.
Např. u diskrétního náhodného vektoru (X, Y) dostáváme
$$P(X = x_i) = \sum_{j=1}^{\infty} P(X = x_i \wedge Y = y_j)$$

Nezávislost náhodných veličin

Náhodné veličiny jsou statisticky nezávislé, jestliže pro libovolná čísla $x_1, x_2, \dots, x_n$ jsou jevy $X_1 \le x_1, X_2 \le x_2, \dots, X_n \le x_n$ nezávislé.

Tj:

$$F_X(x_1, \dots x_n) = F_{X_1}(x_1)\, F_{X_2}(x_2) \dots F_{X_n}(x_n)$$

pro distribuční funkce.

Pro diskrétní náhodné veličiny to znamená

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

$$= \left(\sum_{x_i \le x} f(x_i) \right) \left(\sum_{y_j \le y} g(y_j) \right)$$

Tedy pravděpodobnostní funkce h náhodného vektoru je

$$h(x_i, y_j) = f(x_i)\, g(y_j)$$

kde f je pravděpodobnostní funkce X a g je pravděpodobnostní funkce Y.

Obdobně pro spojité náhodné veličiny a jejich hustoty pro stoch. nezávislé veličiny platí

$$f_{X_1 \dots X_n}(x_1, x_2, \dots, x_n) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdot \dots f_{X_n}(x_n)$$

Důkaz: Pro $f_{(X,Y)}$ spojitou platí

$$f_{(X,Y)}(x,y) = \frac{\partial^2}{\partial y\, \partial x}\left\{F_{(X,Y)}(x,y)\right\} =$$

$$= \frac{\partial^2}{\partial y\, \partial x}\left(F_X(x)\cdot F_Y(y)\right) =$$

$$= \left(\frac{\partial}{\partial x} F_X(x)\right)\cdot\left(\frac{\partial}{\partial y} F_Y(y)\right) =$$

$$= f_X(x)\cdot f_Y(y). \qquad \blacksquare$$

Příklad náhodných veličin, které nejsou stochasticky nezávislé.

(X,Y) má spojité rozložení s hustotou

$$f(x,y) = \begin{cases} \frac{1}{\pi} & x^2+y^2 \le 1 \\ 0 & \text{jinak} \end{cases}$$

Kdyby Pravděpodobnost, že

$$(*) \qquad (X,Y) \in [-1,1]\times[-1,1] \setminus \{(x,y)\in\mathbb{R}^2;\ x^2+y^2\le 1\}$$

je evidentně 0. Hustota náhodných veličin X a Y je nenulová (a samozřejmě kladná) na intervalu $(-1,1)$. Kdyby $f_{X,Y}(x,y) = f_X(x)\cdot f_Y(y)$, pak by ovšem pravděpodobnost jevu $(*)$ byla integrálem z $f_X(x)\cdot f_Y(y)$ přes množinu, kde je součin kladný, tedy > 0. To je spor.

Funkce náhodných veličin

Je-li $\varphi : \mathbb{R} \to \mathbb{R}$ spojitá a $X : \Omega \to \mathbb{R}$ je náhodná veličina, pak $\varphi \circ X$ je opět náhodná veličina $Y = \varphi(X)$.

Obdobně pro náhodný vektor $(X_1, \dots X_n) : \Omega \to \mathbb{R}$ a spojitou $\varphi : \mathbb{R}^n \to \mathbb{R}$ je

$$Y = \varphi(X_1, \dots, X_n)$$

náhodná veličina.

Afinní závislost

$$\varphi(X) = aX + b$$

Funkce náhodného vektoru

$$Y = X_1 + X_2 + \dots + X_n$$

$X_1, X_2, \dots X_n$ nezávislé náhodné veličiny s alt. rozložením

$$f(0) = 1-p$$
$$f(1) = p$$

Pak

$$f_Y(k) = \sum_{\sum x_i = k} f(x_1) f(x_2) \cdots f(x_n) =$$

$$= \binom{n}{k} p^k (1-p)^{n-k}$$

je binomické rozložení.

Normální rozložení

Nechť Z je náhodný vektor s normálním rozložením $N(0,1)$. Náhodná veličina

$$Y = \mu + \sigma Z$$

kde má náhodné rozdělení $N(\mu,\sigma)$, neboť

$$F_Y(y) = P(Y \le y) = P(\mu + \sigma Z < y) =$$

$$= P\left(Z \le \frac{y-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{y-\mu}{\sigma}} e^{-\frac{z^2}{2}} dz$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{y} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Provedli jsme substituci $z = \frac{x-\mu}{\sigma}$.

Součet dvou ~~[illegible]~~ náhodných veličin

X a Y s hustotami f_X a f_Y. Spočteme distribuční funkci náhodné veličiny $V = X+Y$.

$$F_V(v) = \int_{x+y \le v} f_X(x) f_Y(y)\, dx\, dy =$$

$$= \int_{-\infty}^{v} \int_{-\infty}^{\infty} f_X(x) f_Y(v-x)\, dx\, dv$$

Tedy hustota náhodné veličiny V je

$$f_V(v) = \int_{-\infty}^{\infty} f_X(x) f_Y(v-x)\, dx = f_X * f_Y$$

konvoluce hustot f_X a f_Y.

STŘEDNÍ HODNOTA

Střední hodnota náhodné veličiny X je číslo EX definované jako

$$EX = \begin{cases} \sum_i x_i f(x_i) & \text{pro diskrétní } X \\ \int_{-\infty}^{\infty} x f(x) dx & \text{pro spojitou } X \end{cases}$$

pokud součty a integrály konvergují absolutně. Pokud nekonvergují absolutně, říkáme, že náhodná veličina střední hodnotu nemá.

Střední hodnota náhodného vektoru je vektor středních hodnot jeho složek.

$$E(X_1, X_2, \ldots, X_n) = (EX_1, EX_2, \ldots, EX_n)$$

Střední hodnota náhodné veličiny $Y = \varphi(X)$

Diskrétní případ

$$EY = \sum_j y_j \; P(Y = y_j) = \sum y_j \Big(\sum_{\varphi(x_i) = y_j} P(X = x_i) \Big)$$
$$= \sum_i \varphi(x_i) \; P(X = x_i)$$

Spojitý případ – předpokládejme, že ψ je diferencovatelná a rostoucí a platí

$$F_X'(x) = f_X(x)$$

Potom

$$F_Y(y) = F_X(\psi^{-1}(y)) = P(X \leq \psi^{-1}(y))$$

$$\frac{\partial F_Y}{\partial y}(y) = f_Y(y)$$

$$\|$$

$$\frac{\partial F_X}{\partial x}(\psi^{-1}(y)) \cdot \frac{1}{\psi'(x)} = f_X(x) \cdot \frac{1}{\psi'(x)}$$

Proto

$$EY = \int_{-\infty}^{\infty} y\, f_Y(y)\, dy = \int_{-\infty}^{\infty} y\, f_X(x) \frac{1}{\psi'(x)}\, dy$$

$$= \int_{-\infty}^{\infty} \psi(x)\, f_X(x)\, dx.$$

Příklady

① Střední hodnota alternativního rozdělení $X = A(p)$ je

$$EX = 0 \cdot (1-p) + 1 \cdot p = p.$$

② Střední hodnota $aX+b$

$$E(aX+b) = aEX + b$$

③ Střední hodnota součinu dvou náhodných veličin $X \cdot Y$, které jsou nezávislé:

$$E(XY) = \sum_i \sum_j x_i y_j f_X(x_i) f_Y(y_j) =$$

$$= \left(\sum_i x_i f_X(x_i)\right) \left(\sum_j y_j f_Y(y_j)\right)$$

$$= EX \cdot EY$$

Stejně pro spojité náhodné veličiny.

④ Střední hodnota součtu $X+Y$

$$E(X+Y) = \sum_i \sum_j (x_i + y_j) \, P(X = x_i, Y = y_j)$$

$$= \sum_i x_i \left(\sum_j P(X = x_i, Y = y_j)\right) +$$

$$+ \sum_j y_j \left(\sum_i P(X = x_i, Y = y_j)\right) =$$

$$= \sum_i x_i P(X = x_i) + \sum_j y_j P(Y = y_j)$$

$$= EX + EY$$

Pro spojité náhodné veličiny

$$E(X+Y) = \int_{-\infty}^{\infty} z \, (f_X * f_Y)(z) \, dz =$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z f_X(x) f_Y(z-x) \, dx \, dz =$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (z-x) f_X(x) f_Y(z-x)\, dx\, dz +$$

$$+ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_X(x) f_Y(z-x)\, dx\, dz =$$

$$= \int_{-\infty}^{\infty} u f_Y(u)\, du \int_{-\infty}^{\infty} f_X(x)\, dx + \int_{-\infty}^{\infty} x f_X(x)\, dx \int_{-\infty}^{\infty} f_Y(y)\, dy$$

$$= EX + EY$$

(5) Střední hodnota binomického rozdělení

$$B(n,p) = A(p) + A(p) + \dots + A(p)$$

je $EB(n,p) = n\, EA(p) = np$.

(6) Střední hodnota Poissonova rozdělení

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

je

$$EX = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} = \lambda$$

Viz příklad s pohhaleným zápasem.

(7) Střední hodnota normálního rozložení $N(0,1)$

je $EN(0,1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x\, e^{-\frac{x^2}{2}}\, dx = 0$

Střední hodnota normálního rozložení $N(\mu,\sigma)$

je $EN(\mu,\sigma) = \mu + \sigma\, EN(0,1) = \mu$.

8 Příklady náhodných veličin, které nemají střední hodnotu

Diskrétní náhodná veličina
- určená pravděpodobnostní funkcí $f : \mathbb{N} \to \mathbb{R}$ takovou, že $\sum_{i=1}^{\infty} f(i) = 1$, ale $\sum_{i=1}^{\infty} i f(i)$ diverguje.

Např. $\sum_{i=1}^{\infty} \frac{1}{i(i+1)} = 1$, ale $\sum_{i=1}^{\infty} \frac{1}{i+1} = \infty$.

Spojitá náhodná veličina určená hustotou takovou, že $\int_{-\infty}^{\infty} f(t)\,dt = 1$, ale $\int_{-\infty}^{\infty} t f(t)\,dt$ diverguje. Např.

$$f(t) = 0 \text{ pro } t < 1$$
$$f(t) = \frac{1}{t^2} \text{ pro } t \geq 1$$

$$\int_1^{\infty} \frac{1}{t^2}\,dt = \left[-\frac{1}{t}\right]_1^{\infty} = 1 \qquad \int_1^{\infty} t\,\frac{1}{t^2}\,dt = \int_1^{\infty} \frac{1}{t}\,dt = \infty .$$

ROZPTYL

Nechť X je náhodná veličina s konečnou střední hodnotou. **Rozptyl** definujeme

$$\operatorname{var} X = E(X - EX)^2 ,$$

pokud taková konečná hodnota existuje.

Směrodatná odchylka je $\sqrt{\operatorname{var} X}$.

Věta Výpočet ~~variance~~ rozptylu pro spojité a diskrétní náhodné veličiny.

(1) Nechť X je spojitá náhodná veličina s hustotou f a střední hodnotou μ.

Pak
$$\operatorname{var} X = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$$

(2) Je-li X diskrétní náhodná veličina s ~~hustotou~~ pravděpodobnostní funkcí f, pak
$$\operatorname{var} X = \sum_i (x_i-\mu)^2 f(x_i).$$

Důkaz: Plyne z toho, že
$$\operatorname{var} X = E\,(X-EX)^2$$
Tedy X dosadíme do funkce $\psi(x) = (x-\mu)^2$ a použijeme
$$E\psi(X) = \int_{-\infty}^{\infty} \psi(x) f(x)\,dx$$
ve spojitém případě, analogicky v diskrétním případě.

(1) <u>Rozptyl alternativního rozdělení</u> z definice

X nabývá hodnot 0 a 1 s pravd. $1-p$ a p.

$X-EX$ nabývá hodnot $-p$ a $1-p$ s pravd. $1-p$ a p.

$(X-EX)^2$ nabývá hodnot p^2 a $(1-p)^2$ s pravd. $1-p$ a p.

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

(2) Střední hodnota degenerovaného rozložení $\deg(\mu)$ je evidentně μ. Jeho rozptyl je 0, neboť

$$X-EX$$

je degenerované rozložení s hodnotou 0.

K výpočtu rozptylu je někdy lepší použít jeho vlastnosti.

<u>Věta (Vlastnosti rozptylu)</u>

Jestliže má náhodná veličina X rozptyl, pak platí

(1) $\operatorname{var} X = EX^2 - (EX)^2$

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

(3) $\sqrt{\operatorname{var}(a+bX)} = |b|\sqrt{\operatorname{var} X}$

<u>Důkaz</u>

(1) $$\operatorname{var} X = E(X-EX)^2 = E\left(X^2-2X\cdot EX+(EX)^2\right)$$
$$= E(X^2)-2(EX)(EX)+(EX)^2 = E(X^2)-(EX)^2$$

Použili jsme
$$E(X+Y) = EX+EY$$
a
$$E(aX+b) = aEX+b.$$

(2) $$\operatorname{var}(a+bX) = E(a+bX)^2-\left(E(a+bX)\right)^2 =$$
$$= E(a^2+2abX+b^2X^2)-(a+bEX)^2 =$$
$$= a^2+2abEX+b^2EX^2-a^2-2abEX-b^2(EX)^2$$
$$= b^2\left(EX^2-(EX)^2\right) = b^2 \operatorname{var} X \qquad \blacksquare$$

K náhodné veličině X někdy přiřazujeme náhodnou veličinu Z
$$Z = \frac{X-EX}{\sqrt{\operatorname{var} X}}.$$

Ta má $$EZ = \frac{EX-EX}{\sqrt{\operatorname{var} X}} = 0$$ a
$$\operatorname{var} Z = \frac{1}{\operatorname{var} X}\operatorname{var} X = 1.$$

Normální rozdělení

Nechť náhodná veličina X má náhodné rozdělení $N(0,1)$. Píšeme $X \sim N(0,1)$. Ukážeme, že její rozptyl je 1. K tomu potřebujeme spočítat hustotu náhodné veličiny X^2. Její distribuční funkce je

$$F_{X^2}(x) = P(X^2 \le x) = \int_{-\sqrt{x}}^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

$$\overset{x \ge 0}{=} 2 \int_0^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \underset{z^2 = t}{=} 2 \int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t}{2}} \frac{dt}{2z}$$

$$= \int_0^x \frac{1}{\sqrt{2\pi}} t^{-\frac{1}{2}} e^{-\frac{t}{2}} dt$$

Tedy veličina X^2 má hustotu

$$f_{X^2}(x) = \begin{cases} 0 & \text{pro } x \le 0 \\ \frac{1}{\sqrt{2\pi}} x^{-1/2} e^{-x/2} & \text{pro } x > 0 \end{cases}$$

Tomuto rozdělení říkáme chí kvadrát - značení $\chi^2(1)$, tedy $X^2 \sim \chi^2(1)$.

Jeho střední hodnota dává rovněž rozptyl normálního rozložení, neboť

$$\operatorname{var} X = E(X^2) - (EX)^2 = E(X^2) - 0.$$

$$EX^2 = \int_0^\infty \frac{1}{\sqrt{2\pi}}\, t\cdot t^{-1/2} e^{-\frac{t}{2}}\, dt = \int_0^\infty \frac{1}{\sqrt{2\pi}}\, t^{1/2} e^{-\frac{t}{2}}\, dt$$

$$\overset{t=z^2}{=} \int_0^\infty \frac{1}{\sqrt{2\pi}}\, z\cdot e^{-\frac{z^2}{2}}\, 2z\, dz = 2\int_0^\infty \frac{1}{\sqrt{2\pi}}\, z^2 e^{-\frac{z^2}{2}}\, dz =$$

$$= -2\int_0^\infty \frac{1}{\sqrt{2\pi}}\, z\cdot \left(e^{-\frac{z^2}{2}}\right)' dz \overset{\text{per partes}}{=}$$

$$= 2\int_0^\infty \frac{1}{\sqrt{2\pi}}\, (z)'\cdot e^{-\frac{z^2}{2}}\, dz = 2\int_0^\infty \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}}\, dz =$$

$$= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}}\, dz = 1.$$

Variace normální rozdělení $N(\mu,\sigma)$ reprezentované veličinou $Z \sim N(\mu,\sigma)$

$$Z = \mu + \sigma X \quad \text{kde } X \sim N(0,1)$$

je

$$\operatorname{var} Z = \sigma^2 \operatorname{var} X = \sigma^2. \qquad \blacksquare$$

KVANTILOVÁ FUNKCE

Nechť X je náhodná veličina s distribuční funkcí $F_X : \mathbb{R} \to [0,1]$. Kvantilová funkce této náhodné veličiny je „inverzní" funkcí k F_X: $F^{-1}(0,1) \to \mathbb{R}$

$$F^{-1}(\alpha) = \inf\{x \in \mathbb{R},\ F(x) \geq \alpha\}$$

Hodnota $F^{-1}(\alpha)$ se nazývá α-kvantil.

ČEBYŠEVOVA NEROVNOST

Předpokládejme, že náhodná veličina má konečný rozptyl a uvažujme $\varepsilon > 0$. Potom platí

$$P(|X - EX| \ge \varepsilon) \le \frac{\operatorname{var} X}{\varepsilon^2}.$$

Důkaz: pro spojitou náhodnou veličinu X. Položme $\mu = EX$ a počítejme podle definice

$$\operatorname{var} X = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx = \int_{|x-\mu|\ge\varepsilon} (x-\mu)^2 f(x)\,dx$$

$$+ \int_{|x-\mu|<\varepsilon} (x-\mu)^2 f(x)\,dx \ge \int_{|x-\mu|\ge\varepsilon} \varepsilon^2 f(x)\,dx = \varepsilon^2 P(|X-\mu| \ge \varepsilon)$$

∎

Důsledky Pro $\varepsilon = \sigma = \sqrt{\operatorname{var} X}$ je

$$P(|X-EX| \ge \sigma) \le 1 \quad , \text{ což je zřejmé}$$

Pro $\varepsilon = k\sigma$, $\sigma = \sqrt{\operatorname{var} X}$ je

$$P(|X-EX| \ge k\sigma) \le \frac{1}{k^2}.$$