NMAI059 Pravděpodobnost a statistika 1
3. přednáška

Robert Šámal

Přehled

Diskrétní náhodné veličiny

Příklady diskrétních n.v.

Střední hodnota

Parametry náhodných veličin

Náhodné vektory

Co už víme

Definice

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

Definice

Pravděpodobnostní funkce (probability mass function, pmf) diskrétní náhodné veličiny X *je funkce* $p_X : \mathbb{R} \to [0, 1]$ *taková, že*

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

- $\sum_{x \in Im(X)} p_X(x) = 1$
- a to je jediné omezení
- X určuje diskrétní pravděpodobnostní prostor na $Im(X)$

Jiný popis – distribuční funkce

Definice

Distribuční funkce (cumulative distribution function, CDF) n.v. X je funkce

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

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

Přehled

Bernoulliho/alternativní rozdělení

- $X =$ počet orlů při jednom hodu nespravedlivou mincí.
- Značíme $X \sim Bern(p)$. (Někdy se značí $Alt(p)$.)

- Dáno $p \in [0, 1]$.
- $p_X(1) = p$
- $p_X(0) = 1 - p$
- $p_X(k) = 0$ pro $k \neq 0, 1$

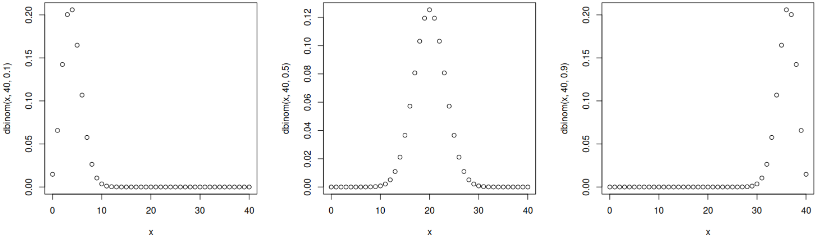
- Pro libovolný jev $A \in \mathcal{F}$ definujeme *indikátorovou n.v.* I_A:
- $I_A(\omega) = 1$ pokud $\omega \in A$, $I_A(\omega) = 0$ jinak.
- $I_A \sim Bern(P(A))$

Binomiální rozdělení

- $X =$ počet orlů při n hodech nespravedlivou mincí.
- Dáno $p \in [0, 1]$ – pravděpodobnost orla při jednom hodu.
- Značíme $X \sim Bin(n, p)$.

- $X = \sum_{i=1}^{n} X_i$ pro nezávislé n.v. $X_1, \dots, X_n \sim Bern(p)$.
- $p_X(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ pro $k \in \{0, 1, \dots, n\}$

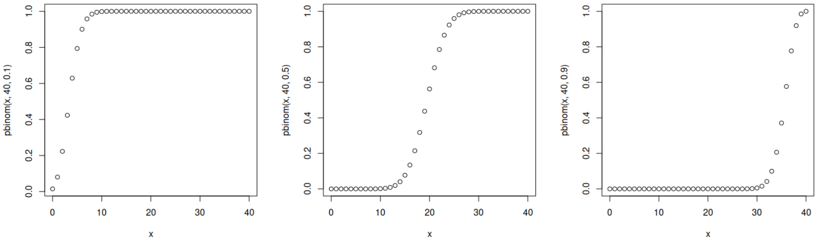
Binomiální rozdělení: pravděpodobnostní funkce



Vygenerováno následujícím kódem v R

```
x <- 0:40
plot(x,dbinom(x,40,0.1))
plot(x,dbinom(x,40,0.5))
plot(x,dbinom(x,40,0.9))
```

Binomiální rozdělení: distribuční funkce



Vygenerováno následujícím kódem v R

```
x <- 0:40
plot(x,pbinom(x,40,0.1))
plot(x,pbinom(x,40,0.5))
plot(x,pbinom(x,40,0.9))
```

Hypergeometrické rozdělení

- $X =$ počet vytažených červených míčků při n tazích, v osudí je K červených z N celkových míčků
- Dáno n, N, K.
- Značíme $X \sim Hyper(N, K, n)$.
- $p_X(k) = P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$

Poissonovo rozdělení

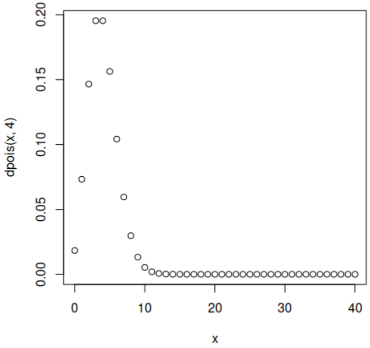
- Značíme $X \sim Pois(\lambda)$.

- Dáno reálné $\lambda > 0$.
- $p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}$

- $Pois(\lambda)$ je limitou $Bin(n, \lambda/n)$

- X popisuje např. počet emailů, které dostaneme za jednu hodinu.

Poissonovo rozdělení: pravděpodobnostní funkce



Vygenerováno následujícím kódem v R

```
x <- seq(0,40,by=1)
plot(x,dpois(x,4))
```

Poissonovo paradigma

- $A_1, \ldots, A_n$ jsou (skoro-)nezávislé jevy s $P(A_i) = p_i$, $\lambda = \sum_i p_i$. Nechť n je velké, každé z p_i malé. Pak přibližně platí
$$\sum_{i=1}^{n} I_{A_i} \sim Pois(\lambda).$$

Geometrické rozdělení

- $X =$ kolikátým hodem mincí padl první orel.
- Značíme $X \sim Geom(p)$.

- Dáno $p \in [0, 1]$.
- $p_X(k) = (1 - p)^{k-1}p$, pro $k = 1, 2, \ldots$

- Někdy se tomuto rozdělení říká posunuté geometrické, a za normální geometrické se považuje rozdělení $X - 1$, tj. počet neúspěšných hodů.

Přehled

Střední hodnota

Definice

Pokud X je diskrétní n.v., tak její střední hodnota (expectation) je označována $\mathbb{E}(X)$ a definována

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

pokud součet má smysl.

- Nechť X je definována na diskrétním prostoru $(\Omega, \mathcal{F}, P)$. Pak střední hodnotu lze také definovat

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}).$$

LOTUS

- Pro reálnou funkci g a diskrétní n.v. X je $Y = g(X)$ také diskrétní n.v.

Věta (LOTUS)

Pokud X je diskrétní n.v. a g reálná funkce, tak

$$\mathbb{E}(g(X)) = \sum_{x \in Im(X)} g(x)P(X = x)$$

pokud součet má smysl.

Vlastnosti $\mathbb{E}$

Věta

Nechť X, Y jsou diskrétní n.v. a $a, b \in \mathbb{R}$.

1. *Pokud $P(X \geq 0) = 1$ a $\mathbb{E}(X) = 0$, tak $P(X = 0) = 1$.*
2. *Pokud $\mathbb{E}(X) \geq 0$ tak $P(X \geq 0) > 0$.*
3. $\mathbb{E}(a \cdot X + b) = a \cdot \mathbb{E}(X) + b$.
4. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$.

Rozptyl

Definice

Rozptyl (variance) n.v. X nazveme číslo $\mathbb{E}((X - \mathbb{E}X)^2)$. Značíme jej $var(X)$.

Věta

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

Podmíněná střední hodnota

Definice

Pokud X je diskrétní n.v. a $P(B) > 0$, tak podmíněná střední hodnota X za předpokladu B (conditional expectation of X given B) je

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

pokud součet má smysl.

Rozbor všech možností

Věta

Pokud $B_1, B_2, \ldots$ je rozklad Ω a $A \in \mathcal{F}$, tak

$$\mathbb{E}(X) = \sum_i \mathbb{E}(X \mid B_i) P(B_i),$$

kdykoliv má součet smysl. (Sčítance s $P(B_i) = 0$ považujeme za 0.)

Rozbor všech možností

Přehled

Parametry rozdělení – Bernoulliho

Pro $X \sim Bern(p)$ je

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

Parametry rozdělení – binomické

Pro $X \sim Bin(n, p)$ je

- $\mathbb{E}(X) = np$
- $var(X) = np(1 - p)$

Parametry rozdělení – geometrické

Pro $X \sim Geom(p)$ je

- $\mathbb{E}(X) = 1/p$
- $var(X) = \frac{1-p}{p^2}$

Parametry rozdělení – Poissonovo

Pro $X \sim Pois(\lambda)$ je

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

Přehled

Základní popis náhodných vektorů

- X, Y – náhodné veličiny na stejném pravděpodobnostním prostoru $(\Omega, \mathcal{F}, P)$.
- Budeme chtít uvažovat (X, Y) jako jeden objekt – náhodný vektor.
- Jak to udělat?
- Příklad: házíme dvakrát čtyřstěnnou kostkou, $X =$ první hod, $Y =$ druhý hod.

Sdružené rozdělení

Definice

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

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

- Mohli bychom definovat i pro více než dvě n.v. $p_{X_1,\dots,X_n}(x_1,\dots,x_n)$.

Marginální rozdělení

- Máme-li dáno $p_{X,Y}$, jak zjistit rozdělení jednotlivých složek, tj. p_X a p_Y?

Funkce náhodného vektoru

Věta

Nechť X, Y jsou n.v. na $(\Omega, \mathcal{F}, P)$, nechť $g : \mathbb{R}^2 \to \mathbb{R}$ je funkce.

- *Pak $Z = g(X, Y)$ je n.v. na $(\Omega, \mathcal{F}, P)$*
- *a platí pro ni*

$$\mathbb{E}(g(X, Y)) = \sum_{x \in ImX} \sum_{y \in ImY} g(x, y) P(X = x, Y = y).$$

Věta

Pro X, Y n.v. a $a, b \in \mathbb{R}$ platí

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$

Důkaz věty o rozptylu

Nezávislost náhodných veličin

Definice

Diskrétní n.v. X, Y jsou nezávislé (independent) pokud pro každé $x, y \in \mathbb{R}$ jsou jevy $\{X = x\}$ a $\{Y = y\}$ nezávislé. To nastane, právě když

$$P(X = x, Y = y) = P(X = x)P(Y = y).$$

Součin nezávislých n.v.

Věta

Pro nezávislé diskrétní n.v. X, Y platí

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

Součet nezávislých n.v.

- Máme-li dáno $p_{X,Y}$, jak zjistit rozdělení součtu, $Z = X + Y$?

Součet nezávislých n.v. – konvoluce

Věta

Pokud X, Y jsou diskrétní nezávislé náhodné veličiny (zkráceně n.n.v.), tak jejich součet $Z = X + Y$ má pravděpodobnostní funkci

$$P(Z = z) = \sum_{x \in ImX} P(X = x)P(Y = z - x).$$