NMAI059 Pravděpodobnost a statistika 1
3. přednáška

Robert Šámal

Přehled

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

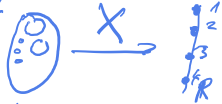
Definice

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

velké X   malé x ∈ ℝ

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

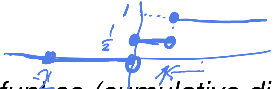
Ω

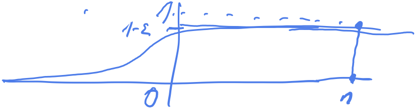


- $\sum_{x \in Im(X)} p_X(x) = 1$
- a to je jediné omezení
- X určuje diskrétní pravděpodobnostní prostor na $Im(X)$

osudí s míčky

Jiný popis – distribuční funkce





Definice

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

$\mathbb{R}$ ω $! \le !$

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

- F_X je neklesající funkce → Dk: $x < y \Rightarrow P(X \le x) \le P(X \le y)$
  Protože $A = \{\omega : X(\omega) \le x\}$, $B = \{\omega : X(\omega) \le y\}$
  $A \subseteq B \Rightarrow P(A) \le P(B)$ ■
- $\lim_{x\to-\infty} F_X(x) = 0$
- $\lim_{x\to+\infty} F_X(x) = 1$
- F_X je zprava spojitá

$\boxed{Dk}$ $A_n = \{X \le n\}$; platí $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$

Tak $\bigcup_{n=1}^{\infty} A_n = \Omega$

Podle Věty o spoj.: $\underset{1}{P(\Omega)} = \lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} F_X(n)$

$X = \begin{cases} 0 & \text{s pst. } \frac{1}{2} \\ 1 & \text{s pst. } \frac{1}{2} \end{cases}$

$\forall \varepsilon > 0\ \exists n : F_X(n) > 1 - \varepsilon$

$\forall x \ge n : F_X(x) > 1 - \varepsilon$ ■

Přehled

Bernoulliho/alternativní rozdělení

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

  → X má rozdělení ..... : popis pstní fce/distr. fce

- Dáno $p \in [0, 1]$.
- $p_X(1) = p$ $= P(X=1)$
- $p_X(0) = 1 - p$ $= P(X=0)$
- $p_X(k) = 0$ 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 $k \in \{0, 1, \dots, n\}$

$p_X(k) = 0$ jinak

P(orel padl když měl)

P(orel nepadl když neměl)

$n=8$

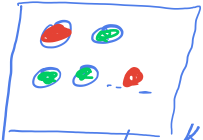
$k=3$

aby to bylo korektní rozdělení, je třeba

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = 1$$

$$= (p + (1-p))^n = 1^n$$

$Bin(n, K/N)$



bedna s míčky

K červ.

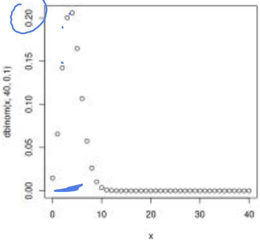
N-K zel.

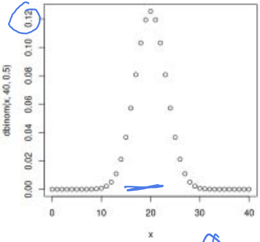
Y = # červ, když vytáhnu n míčků S VRACENÍM

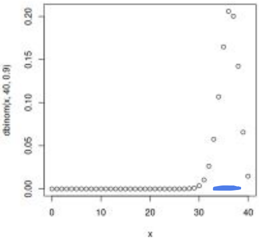
Binomiální rozdělení: pravděpodobnostní funkce

symetrie

$Bin(40, \frac{1}{10})$ $Bin(40, \frac{1}{2})$ $Bin(40, \frac{9}{10})$







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))
```

x n p

v jejích bodech akce hodnot pstní fce

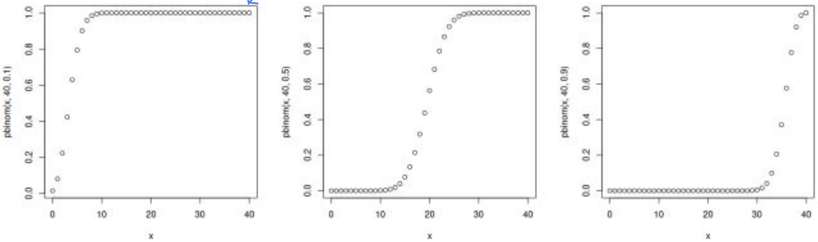
$p = \frac{1}{2}$

$\binom{n}{k} p^k (1-p)^{n-k} = \binom{n}{k} \frac{1}{2^n}$

binom. koef.

Binomiální rozdělení: distribuční funkce

=1



Vygenerováno následujícím kódem v R

```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

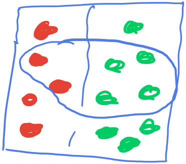
- X = počet vytažených červených míčků při n tazích, v osudí je K červených z N celkových míčků $\boxed{\text{BEZ VRACENÍ}}$
- Dáno n, N, K.
- Značíme $X \sim Hyper(N, K, n)$.

\# dobrých možností

K $\qquad N-K$

- $p_X(k) = P(X = k) = \dfrac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$

\# všech možností



$k = 2$

$n = 6$

$\doteq Bin\left(n, \frac{K}{N}\right)$ pokud $n \ll K, N$

$\left(\frac{N-1}{N} \doteq 1\right)$

$\doteq \binom{n}{k}\left(\frac{K}{N}\right)^n$

Mělo by platit $\sum_k \dfrac{\binom{K}{k}\binom{N-k}{n-k}}{\binom{N}{n}} = 1$

$\sum_k \binom{K}{k}\binom{N-k}{n-k} = \binom{N}{n}$

Poissonovo rozdělení

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

MA: $e^{\lambda} = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}$

- Značíme $X \sim Pois(\lambda)$.
- Dáno reálné $\lambda > 0$.
- $\boxed{p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}}$ ≥ 0, ≤ 1
- $Pois(\lambda)$ je limitou $Bin(n, \lambda/n) \sim X_n$ $\qquad$ λ pevné
- X popisuje např. počet emailů, které dostaneme za jednu hodinu.

$$P(X_n = k) = \binom{n}{k}\left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{n(n-1)\dots(n-k+1)}{k!\,n^k}\,\lambda^k \cdot \frac{\left(1-\frac{\lambda}{n}\right)^n}{} \cdot \frac{\left(1-\frac{\lambda}{n}\right)^{-k}}{}$$

$\to 1$, $\to e^{-\lambda}$, $\to 1$

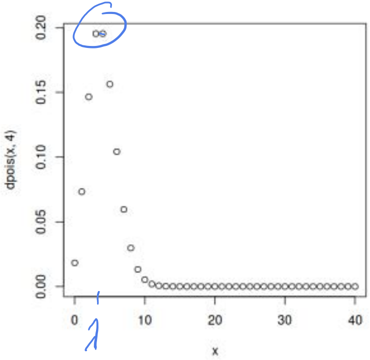
k pevné, $n \to \infty$

def. $Bin(n, \frac{\lambda}{n})$ $\longrightarrow 1 \cdot \frac{\lambda^k}{k!} \cdot 1 \cdot e^{-\lambda}$

$\frac{n-k}{n} \to 1$

Poissonovo rozdělení: pravděpodobnostní funkce

max . . . . . modus



Vygenerováno následujícím kódem v R

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

x = 0:40

Poissonovo paradigma

- $A_1, \dots, A_n$ 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
$$\sum_{i=1}^{n} I_{A_i} \sim Pois(\lambda).$$

jevů, které nastanou

Geometrické rozdělení

- X = kolikátým hodem mincí padl první orel.
- Značíme $X \sim Geom(p)$. $Geom_1(p)$
- Dáno $p \in [0, 1]$.
- $p_X(k) = \underline{(1-p)^{k-1}}\,\underline{p}$, 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.

$? \; Geom_0(p)$

chceme $\underbrace{\sum_{k=1}^{\infty} (1-p)^{k-1} p}$ = 1

$$= \frac{(1-p)^0 \cdot p}{1-(1-p)} = \frac{p}{p} = 1$$

Přehled

Střední hodnota

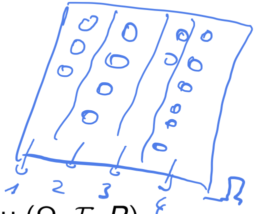
$\sum_{k=1}^{\infty} \frac{1}{2^k} = \frac{1}{2} + \frac{1}{4} + \dots = 1$ $\qquad \sum_{k=1}^{\infty} 1 = \infty$

$\sum_{k=1}^{\infty} (-1)^k$ nedef.

Definice

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

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



pokud součet má smysl.

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

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

dk že obě definice souhlasí !

vážený průměr

$$= \sum_{x \in Im X} \sum_{\substack{\omega \in \Omega \\ X(\omega) = x}} X(\omega) P(\{\omega\}) = \sum_{x \in Im X} \left(x \cdot P(\{\omega \in \Omega : X(\omega) = x\}) \right)$$

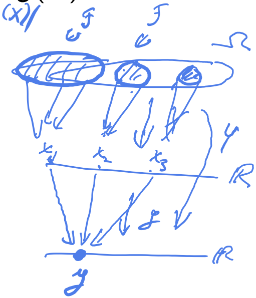
LOTUS

$g: \mathbb{R} \to \mathbb{R}$ (Law Of The Unconscious Statist.)

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

$Im(Y)$ spoč. $|Im(Y)| \leq |Im(X)|$

$\{Y=y\} \in \mathcal{F}$



Věta (LOTUS)

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

$Y=g(X)$

$$\mathbb{E}(g(X)) = \sum_{x\in Im(X)} g(x)\underbrace{P(X=x)}_{\text{p. fce } X}$$

pokud součet má smysl.

$$\mathbb{E}Y = \sum_{y\in Y} y \cdot \underbrace{P(Y=y)}_{\text{pravd. fce } Y} \quad \text{def.}$$

$$\sum_{x\in Im X,\ g(x)=y} P(X=x)$$

$$\{\omega : Y(\omega)=y\} = \bigcup_{\substack{x\in Im X\\ g(x)=y}} \underbrace{\{\omega : X(\omega)=x\}}_{\in \mathcal{F}}$$

spoč. sjedn.

podle předp. (X je d.n.v.)

$$\sum_{y\in Y} \sum_{\substack{x\in Im X\\ g(x)=y}} g(x) \cdot P(X=x) =$$

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

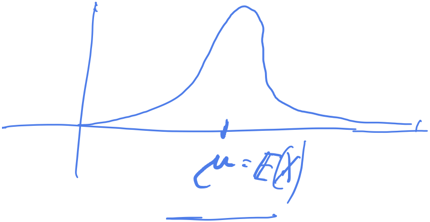
$\mathbb{E}(X-\mu)^2$

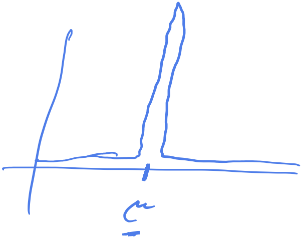
Definice

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 $B_1, B_2, \dots$ *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