# NMAI059 Pravděpodobnost a statistika 1
## 2. přednáška

Robert Šámal

# Přehled

# Náhodná proměnná

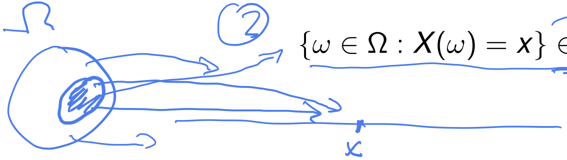
Často nás zajímá číslo dané výsledkem náhodného pokusu.

- Hodíme na terč a změříme vzdálenost od středu.
- Házíme kostkou, dokud nepadne šestka, ale pak si všimneme jenom toho, kolik hodů to trvalo.
- U quicksortu (algoritmus na třídění) měříme počet kroků (v závislosti na náhodné volbě pivotu).

## Definice

n.v.

(1) *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*




(2)

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

$X^{-1}(x)$

$\Longleftrightarrow P(X^{-1}(x))$ je def.

$\|$

$P(X = x)$

$\mathbb{R}$

obvykle $= 0$

# Pravděpodobnostní funkce

$P\left(\bigcup_{x\in \mathrm{Im}\,X}\{X=x\}\right) = \sum_{x\in \mathrm{Im}\,X} p_X(x) = 1$

## Definice

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

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

$X$ $\quad x \in \mathbb{R}$ $\qquad A \subseteq S \qquad P(X \in A) =$

- Označíme-li $S$ obor hodnot $X$ a $Q(A) := \sum_{x\in A} p_X(x)$, tak $(S, \mathcal{P}(S), Q)$ je diskrétní pravděpodobnostní prostor.
- Pro $S = \{s_i : i \in I\}$ spočetnou množinu reálných čísel a $c_i \in [0,1]$ splňující $\sum_{i\in I} c_i = 1$ existuje pravděpodobnostní prostor a diskrétní n.v. $X$ na něm taková, že $p_X(s_i) = c_i$ pro $i \in I$.

# Přehled

Diskrétní náhodné veličiny

Příklady diskrétních n.v.

Střední hodnota

Bonusy

# Bernoulliho/alternativní rozdělení (distribucí)

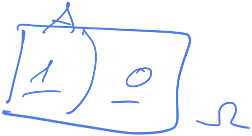
- $X =$ počet orlů při jednom hodu nespravedlivou mincí.
- Značíme $X \sim Bern(p)$.

  náhodná

  X má prav. fci danou výrazem

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

- 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.
- Značíme $X \sim Bin(n, p)$.

$p = \frac{1}{2}$ $\qquad \frac{\binom{n}{k}}{2^n}$

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

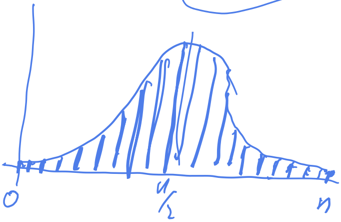
TODO: OBR.

$\sum p_X(k) = 1$ ?

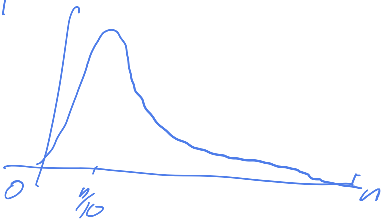
$= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n$

$= 1^n = 1$

$p = \frac{1}{2}$ $\qquad \sim e^{-x^2}$




$p = 0.1$



# 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.

TODO: OBR.

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

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

znáte z MA

$X_n$

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

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

$k$ pevné

$n \to \infty$

# 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

$$\sum_{i=1}^{n} I_{A_i} \sim Pois(\lambda).$$

$A_i = \{$ člověk $i$ nám pošle email dneska $\}$

$A_1 \dots A_n$ nezávislé, $p_i = \frac{\lambda}{n}$

$\sum I_{A_i} \sim Bin(n, \frac{\lambda}{n})$

# Geometrické rozdělení

geometrické řada

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

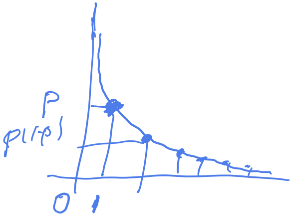
(k-1)x padla panna pak orel

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

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

$k = 1, 2, 3, \ldots$

- 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*

$$\mathbb{E}(X) = \sum_{x \in Im(X)} x \cdot P(X = x), \quad \left( = \sum_{\omega \in \Omega} X(\omega) \cdot P(\{\omega\}) \right)$$

*pokud součet má smysl.*

- $Im(X)$ je spočetná ... umíme sčítat
- součet by mohl být typu "$\infty - \infty$"

---

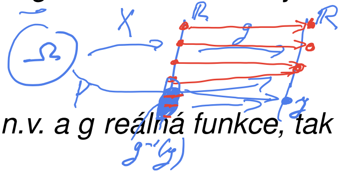
$X$ = výsledek hodu kostkou

$P(X=1) = \frac{1}{6}$ .... podíl "1" při opak. házení

$\mathbb{E}X$ odhad průměru při opak. házení kostkou

# LOTUS (Law of the Unconscious Statist.)

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




① $Y : \Omega \to \mathbb{R}$

$\operatorname{Im} Y = g(\operatorname{Im} X)$

$|\operatorname{Im} Y| \le |\operatorname{Im} X|$

## Věta (LOTUS)

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

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

② $Y^{-1}(y)$

$\left( = X^{-1}(g^{-1}(y)) = \bigcup_{\substack{g(x)=y \\ \& \ x \in \operatorname{Im} X}} X^{-1}(x) \right)$

spoč. sjedn.

$\in$

$X^{-1}(x) \in \mathcal{F}$

$Y = g(X)$, $I = \operatorname{Im}(X)$

*pokud součet má smysl.*

Def

$\operatorname{Im}(Y) = g(I)$

$$\mathbb{E}Y = \sum_{y \in g(I)} y \cdot P(Y = y)$$

$$= \sum_{y \in g(I)} y \cdot \sum_{\substack{x \in I \\ g(x) = y}} P(X = x) = \sum_{x \in I} \sum_{y = g(x)} \cancel{y}\, P(X = x)$$

$g(x)$

prohození pořadí sčítání

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

Dk 1) $EX = \sum_{x \in \mathcal{I}} x\, \underbrace{P(X = x)}_{\geq 0} = 0$

$x \geq 0$

$\forall x \quad x \cdot P(X = x) = 0$

3) $g(x) = ax + b$ & LOTUS

$\sum (ax+b) P(X = x)$

$a \underbrace{\sum x P(X=x)}_{EX} + b \underbrace{\sum P(X=x)}_{=1}$

4) později

# 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*

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

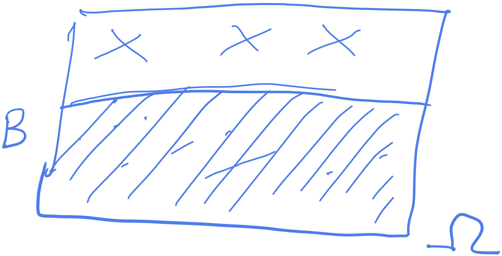
*pokud součet má smysl.*

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

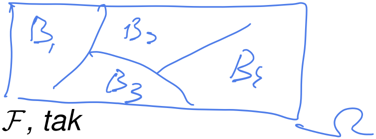
& ...

$X : B \to \mathbb{R}$

$\mathbb{E}X = \sum x \, P(X = x)$



# Rozbor všech možností




## Věta

*Pokud $B_1, B_2, \dots$ je rozklad $\Omega$ 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.)*

$$\text{Dk: } \cancel{\mathbb{E}X} = \sum_i \sum_{x \in \mathrm{Im}\, X} x \cdot P(X = x \mid B_i) P(B_i)$$

$$= \sum_{x \in \mathrm{Im}\, X} \Big( x \underbrace{\sum_i P(X = x \mid B_i) \cdot P(B_i)}_{P(X = x)} \Big) = \sum x P(X = x)$$

$$= \mathbb{E}X$$

# Rozbor všech možností

$X \sim Geom(p)$ ... čekání na úspěch

posl. $Bern(p)$ navzájem nezávislé!

$B_1$ = poprvé uspějeme

$B_2 = B_1^c$

$$EX = \underbrace{E(X/B_1)}_{=1} \underbrace{P(B_1)}_{= p} + \underbrace{E(X/B_2)}_{(1+EX)} \underbrace{P(B_2)}_{1-p}$$

$$= p + (1+EX)(1-p) = 1 + EX(1-p)$$

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

(aby bylo zcela korektní: $EX$ existuje

$X \geq 0$ ... další $EX$ má smysl)

# Přehled

# Bertrand's paradox

# Simpsons's paradox