

# Drsná matematika IV – 7. přednáška

## Jak na statistiku?

Jan Slovák

Masarykova univerzita  
Fakulta informatiky

2. 4. 2012

# Obsah přednášky

- 1 Literatura
- 2 Co je statistika?
- 3 Popisná statistika
  - Míry polohy statistických znaků
  - Míry variability statistických znaků
- 4 Pravděpodobnost
- 5 Náhodné veličiny

# Plán přednášky

- 1 Literatura
- 2 Co je statistika?
- 3 Popisná statistika
  - Míry polohy statistických znaků
  - Míry variability statistických znaků
- 4 Pravděpodobnost
- 5 Náhodné veličiny

## Kde je dobré číst?

- vlastní poznámky, texty současného přednášejícího, GOOGLE, atd.
- Karel Zvára, Josef Štěpán, Pravděpodobnost a matematická pravděpodobnost statistika, Matfyzpress, 2006, 230pp.
- Marie Budíková, Štěpán Mikoláš, Pavel Osecký, Teorie pravděpodobnosti a matematická statistika (sbírka příkladů), Masarykova univerzita, 3. vydání, 2004, 117 stran, ISBN 80-210-3313-4.
- Marie Budíková, Tomáš Lerch, Štěpán Mikoláš, Základní statistické metody, Masarykova univerzita, 2005, 170 stran, ISBN 80-210-3886-1.
- Riley, K.F., Hobson, M.P., Bence, S.J. Mathematical Methods for Physics and Engineering, second edition, Cambridge University Press, Cambridge 2004, ISBN 0 521 89067 5, xxiii + 1232 pp.

# Plán přednášky

- 1 Literatura
- 2 Co je statistika?
- 3 Popisná statistika
  - Míry polohy statistických znaků
  - Míry variability statistických znaků
- 4 Pravděpodobnost
- 5 Náhodné veličiny

Statistika v širším slova smyslu je jakékoliv zpracování číselných dat o nějakém souboru objektů a jejich více či méně přehledná prezentace.

Statistika v širším slova smyslu je jakékoliv zpracování číselných dat o nějakém souboru objektů a jejich více či méně přehledná prezentace.

Podstatou **matematické statistiky** je pro daná data zjišťovat, jaké vlastnosti mají objekty, které jsou daty popisovány. Zároveň jde ale o věrohodnost odvozených výsledků. Zpravidla jde o sběr dat o části souboru objektů, jejich následnou analýzu a konečně o vyslovení důsledků pozorování pro celý soubor.

Statistika v širším slova smyslu je jakékoliv zpracování číselných dat o nějakém souboru objektů a jejich více či méně přehledná prezentace.

Podstatou **matematické statistiky** je pro daná data zjišťovat, jaké vlastnosti mají objekty, které jsou daty popisovány. Zároveň jde ale o věrohodnost odvozených výsledků. Zpravidla jde o sběr dat o části souboru objektů, jejich následnou analýzu a konečně o vyslovení důsledků pozorování pro celý soubor.

Výsledkem práce matematického statistika je sdělení o velkém souboru objektů na základě studia malé (cíleně nebo náhodně vybrané) části z nich, společně s kvalitativním odhadem věrohodnosti výsledného sdělení.



Statistika v širším slova smyslu je jakékoliv zpracování číselných dat o nějakém souboru objektů a jejich více či méně přehledná prezentace.

Podstatou **matematické statistiky** je pro daná data zjišťovat, jaké vlastnosti mají objekty, které jsou daty popisovány. Zároveň jde ale o věrohodnost odvozených výsledků. Zpravidla jde o sběr dat o části souboru objektů, jejich následnou analýzu a konečně o vyslovení důsledků pozorování pro celý soubor.

Výsledkem práce matematického statistika je sdělení o velkém souboru objektů na základě studia malé (cíleně nebo náhodně vybrané) části z nich, společně s kvalitativním odhadem věrohodnosti výsledného sdělení.

Teorie pravděpodobnosti studuje modely popisující chování abstraktních souborů (pravděpodobnost jevů z jevového pole), statistika studuje skutečné náhodné výběry z nějakého základního souboru a zdůvodňuje výběr teoretického pravděpodobnostního modelu, resp. kvalitativní informace o jeho parametrech.

## Example

Za soubor objektů vezměme všechny studenty této přednášky „Drsná matematika“, jako číselný údaj můžeme uvažovat

- 1 „průměrný počet bodů“ dosažený při hodnocení tohoto předmětu v minulém semestru,
- 2 průměrnou známku dosaženou u zkoušky z tohoto a z jiných pevně vybraných předmětů,
- 3 číselná data vypovídající o historii dřívějšího studia,
- 4 počet pracovních hodin týdně odpracovaných mimo fakultu.

## Example

Za soubor objektů vezměme všechny studenty této přednášky „Drsná matematika“, jako číselný údaj můžeme uvažovat

- 1 „průměrný počet bodů“ dosažený při hodnocení tohoto předmětu v minulém semestru,
- 2 průměrnou známku dosaženou u zkoušky z tohoto a z jiných pevně vybraných předmětů,
- 3 číselná data vypovídající o historii dřívějšího studia,
- 4 počet pracovních hodin týdně odpracovaných mimo fakultu.

Zastavme se u prvního údaje. Samotný aritmetický průměr bodů nám mnoho neřekne ani o kvalitě přednášky ani o kvalitě přednášejícího ani o samotném hodnocení. Zajímá nás také hodnota, která bude „uprostřed souboru“, tj. počet bodů, pro které je stejně studentů pod ní a nad ní.

## Example

Za soubor objektů vezměme všechny studenty této přednášky „Drsná matematika“, jako číselný údaj můžeme uvažovat

- 1 „průměrný počet bodů“ dosažený při hodnocení tohoto předmětu v minulém semestru,
- 2 průměrnou známku dosaženou u zkoušky z tohoto a z jiných pevně vybraných předmětů,
- 3 číselná data vypovídající o historii dřívějšího studia,
- 4 počet pracovních hodin týdně odpracovaných mimo fakultu.

Zastavme se u prvního údaje. Samotný aritmetický průměr bodů nám mnoho neřekne ani o kvalitě přednášky ani o kvalitě přednášejícího ani o samotném hodnocení. Zajímá nás také hodnota, která bude „uprostřed souboru“, tj. počet bodů, pro které je stejně studentů pod ní a nad ní. Obdobně první a poslední čtvrtina, desetina apod. Všem takovým údajům říkáme **statistiky** posuzované veličiny. V uvedených příkladech se jim říká **medián**, **kvartil**, **decil** apod.

Z obecné zkušenosti nebo jako výsledek úvah mimo matematiku víme, jakou „strukturu“ by měla mít sledovaná data. Např. víme, že rozumné hodnocení studentů by mělo mít tzv. **normální rozdělení**. Tento pojem patří do teorie pravděpodobnosti a k jeho zavedení budeme potřebovat poměrně dost matematiky.

Z obecné zkušenosti nebo jako výsledek úvah mimo matematiku víme, jakou „strukturu“ by měla mít sledovaná data. Např. víme, že rozumné hodnocení studentů by mělo mít tzv. **normální rozdělení**. Tento pojem patří do teorie pravděpodobnosti a k jeho zavedení budeme potřebovat poměrně dost matematiky.

Pokud je naše představa oprávněná, pak porovnáním výsledku třeba i docela malého náhodného výběru studentů s teoretickým modelem můžeme zjistit odhad parametrů takového rozdělení a činit závěry, zda je hodnocení „skutečně rozumné“. Zároveň lze popsat věrohodnost našich závěrů, ale to už k teoretickému ověření vyžaduje opravdu hodně matematiky.

Daleko zajímavější vývody ovšem můžeme činit, když porovnáním statistik pro různé veličiny budeme moci dovozovat informace o souvislostech. Pokud např. neexistuje žádná doložitelná souvislost mezi historií předchozího studia a výsledky v dané přednášce, je jedním z možných vysvětlení vývod, že je přednáška prostě špatná.

Daleko zajímavější vývody ovšem můžeme činit, když porovnáním statistik pro různé veličiny budeme moci dovozovat informace o souvislostech. Pokud např. neexistuje žádná doložitelná souvislost mezi historií předchozího studia a výsledky v dané přednášce, je jedním z možných vysvětlení vývod, že je přednáška prostě špatná.

### Závěr úvodních úvah:

- V matematice pracujeme s abstraktním matematickým popisem pravděpodobnosti.
- Vývody pro konkrétní soubory dat, pro které je zvolený model relevantní dává matematická statistika.
- Názor, zda je takový popis adekvátní pro konkrétní výběr dat, je také možné podpořit nebo zavrhnout pomocí metod matematické statistiky.



# Plán přednášky

- 1 Literatura
- 2 Co je statistika?
- 3 Popisná statistika**
  - Míry polohy statistických znaků
  - Míry variability statistických znaků
- 4 Pravděpodobnost
- 5 Náhodné veličiny

Popisná statistika není matematická disciplína ...

Jde o dlouho řadu zvyklostí, jak zpracovávat a prezentovat data, a názvů pro jednotlivé typy sestav dat.

Popisná statistika není matematická disciplína ...

Jde o dlouho řadu zvyklostí, jak zpracovávat a prezentovat data, a názvů pro jednotlivé typy sestav dat.

Zpravidla pracujeme se **statistickým souborem**, který je sestaven ze **statistických jednotek**. Na statistických jednotkách se pak měří (zjišťují) jednotlivé **statistické znaky**.

Popisná statistika není matematická disciplína ...

Jde o dlouho řadu zvyklostí, jak zpracovávat a prezentovat data, a názvů pro jednotlivé typy sestav dat.

Zpravidla pracujeme se **statistickým souborem**, který je sestaven ze **statistických jednotek**. Na statistických jednotkách se pak měří (zjišťují) jednotlivé **statistické znaky**.

Např. souborem mohou být všichni studenti MU, každý zvlášť je pak **statistickou jednotkou**. O těchto jednotkách pak můžeme schraňovat mnoho znaků – např. všechny číselné hodnoty zjistitelné z ISu, jakou mají nejraději barvu, co snědli večer před poslední písemkou, atd.

Popisná statistika není matematická disciplína ...

Jde o dlouho řadu zvyklostí, jak zpracovávat a prezentovat data, a názvů pro jednotlivé typy sestav dat.

Zpravidla pracujeme se **statistickým souborem**, který je sestaven ze **statistických jednotek**. Na statistických jednotkách se pak měří (zjišťují) jednotlivé **statistické znaky**.

Např. souborem mohou být všichni studenti MU, každý zvlášť je pak **statistickou jednotkou**. O těchto jednotkách pak můžeme schraňovat mnoho znaků – např. všechny číselné hodnoty zjistitelné z ISu, jakou mají nejraději barvu, co snědli večer před poslední písemkou, atd.

Základním objektem pro zkoumání jednotlivých znaků je pak **soubor hodnot**. Zpravidla jej máme ve formě uspořádaných hodnot. Uspořádání je buď dáno přirozeně (když jsou hodnotami např. reálná čísla) nebo je můžeme zavést pro určitost (třeba když budeme sledovat barvy, tak je můžeme vyjádřovat v RGB standardu a řadit podle tohoto příznaku).

Protože smyslem statistického popisu je srozumitelně a přehledně sdělit něco o celém souboru, budeme jistě chtít umět jednotlivé hodnoty nějak porovnávat a poměřovat. Je tedy podstatné mít k tomu dispozici nějaké **měřítko**. Podle toho jakého charakteru jsou hodnoty, hovoříme měřítku:

- **nominálním** (mezi hodnotami není žádný vztah, jde pouze o četnosti možných hodnot, např. politická strana v ČR nebo učitelé MU při zkoumání oblíbenosti);
- **ordinální** (totéž jako předchozí, ale s přidáním uspořádáním, např. počet hvězdiček u hotelu v bedekrech);
- **intervalové** (jde o číselné hodnoty, ale jde o porovnání velikostí, nikoliv absolutní hodnotu, např. u měření teplot je poloha nuly dohodnuta, ale není podstatná);
- **poměrové** (máme pevně stanovené měřítko a nulu, např. většina fyzikálních veličin).

V dalším budeme pracovat se **souborem hodnot**  $x_1, x_2, \dots, x_n$  (které vznikly měřením na  $n$  statistických jednotkách) a uspořádáme je do **uspořádaného souboru hodnot**

$$x_{(1)}, x_{(2)}, \dots, x_{(n)}.$$

Číslo  $n$  nazýváme **rozsah souboru**.

Nejjednodušší je u rozsáhlých souborů znaků, které ale připouští jen málo hodnot uvádět pouze četnosti. Např. při průzkumu preferencí politických stran nebo u prezentace kvality hotelové sítě uvádíme u každé možné hodnoty počet jejích výskytů.

Pokud je i možných hodnot více (nebo dokonce připouštíme kontinuální reálné hodnoty), dělíme často možný rozsah hodnot na vhodný počet intervalů a o statistickém znaku uvádíme četnost hodnot v daných intervalech. Intervalům se často říká **třídy** a počtu znaku ve třídě pak **třídní četnost**.

Používáme také **kumulativní třídní četnosti**, které vznikají prostým součtem třídních četností s hodnotami nejvýše jako má daná třída.



Pokud je  $i$  možných hodnot více (nebo dokonce připoustíme kontinuální reálné hodnoty), dělíme často možný rozsah hodnot na vhodný počet intervalů a o statistickém znaku uvádíme četnost hodnot v daných intervalech. Intervalům se často říká **třídy** a počtu znaku ve třídě pak **třídní četnost**.

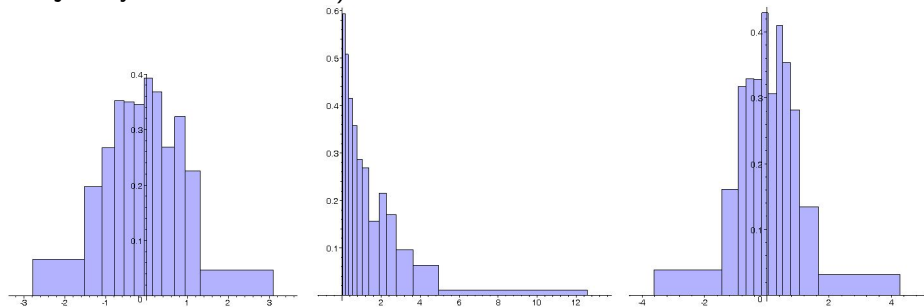
Používáme také **kumulativní třídní četnosti**, které vznikají prostým součtem třídních četností s hodnotami nejvýše jako má daná třída.

Nejčastěji pak uvažujeme střed  $a_i$  dané třídy za hodnotu, která ji reprezentuje a hodnota  $a_i n_i$ , kde  $n_i$  je četnost výskytu této třídy představuje celkový příspěvek této třídy. Velmi často také místo četností zobrazujeme relativní četnosti  $a_i/n$ , resp. relativní kumulativní četnosti.

Graf, který na jedné ose vynáší intervaly jednotlivých tříd a nad nimi obdélníky s výškou rovnou četnosti se nazývá **histogram**.

Obdobně se znázorňuje kumulativní četnost.

Na obrázku jsou histogramy souborů o rozsahu  $n = 500$ , které vznikly náhodným generováním dat s rozdělením normálním,  $\chi^2$  a studentovým (postupně zepředu dozadu, dívejte se na obrázek jakoby zleva shora ... :-).



# Míry polohy statistických znaků

Chceme-li velikost hodnot, kolem kterých se jednotlivá pozorování znaků shromažďují používáme většinou následující:

## Definition

Nechť  $(x_1, \dots, x_n)$  je soubor hodnot měřeného znaku.

- **Průměr** (nebo také výběrový průměr) je dán

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \sum_{j=1}^m n_j a_j;$$

- **Geometrický průměr** je dán

$$\bar{x}_G = \sqrt[n]{x_1 x_2 \cdots x_n}$$

a má smysl pouze u kladných hodnot znaků.

## Definition (pokračování ...)

- Harmonický průměr je dán

$$\bar{x}_H = \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{x_i} \right)^{-1}$$

a je také definován jen pro kladné hodnoty znaků.

## Definition (pokračování ...)

- Harmonický průměr je dán

$$\bar{x}_H = \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{x_i} \right)^{-1}$$

a je také definován jen pro kladné hodnoty znaků.

Výběrový průměr je jediný invariantní vůči afinním transformacím, tj. pro libovolné skaláry  $a$ ,  $b$  platí  $\overline{(a + b \cdot x)} = a + b \cdot \bar{x}$ . Ostatní průměry jsou proto nevhodné pro intervalová měřítka.

## Definition (pokračování ...)

- Harmonický průměr je dán

$$\bar{x}_H = \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{x_i} \right)^{-1}$$

a je také definován jen pro kladné hodnoty znaků.

Výběrový průměr je jediný invariantní vůči afinním transformacím, tj. pro libovolné skaláry  $a$ ,  $b$  platí  $\overline{(a + b \cdot x)} = a + b \cdot \bar{x}$ . Ostatní průměry jsou proto nevhodné pro intervalová měřítka.

Logaritmus harmonického průměru je obyčejný průměr logaritmů znaků. Je obzvláště vhodný pro znaky, které se kumulují multiplikativně, např. úrokové míry. Je-li totiž úroková míra  $v$  jednotlivých časových jednotkách  $x_i$ %, bude za celé období výsledek takový, jakoby byla konstantní úroková míra  $\bar{x}$ %.

## Definition (pokračování ...)

- Harmonický průměr je dán

$$\bar{x}_H = \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{x_i} \right)^{-1}$$

a je také definován jen pro kladné hodnoty znaků.

Výběrový průměr je jediný invariantní vůči afinním transformacím, tj. pro libovolné skaláry  $a, b$  platí  $\overline{(a + b \cdot x)} = a + b \cdot \bar{x}$ . Ostatní průměry jsou proto nevhodné pro intervalová měřítka.

Logaritmus harmonického průměru je obyčejný průměr logaritmů znaků. Je obzvláště vhodný pro znaky, které se kumulují multiplikativně, např. úrokové míry. Je-li totiž úroková míra  $v$  jednotlivých časových jednotkách  $x_i\%$ , bude za celé období výsledek takový, jakoby byla konstantní úroková míra  $\bar{x}\%$ .

Platí  $\bar{x}_H \leq \bar{x}_G \leq \bar{x}$ .

# Medián, kvartil, decil, percentil, ...

Jiný způsob vyjádření míry, jakou hodnotu nabývají znaky je najít pro číslo  $\alpha$  mezi nulou a jedničkou takovou hodnotu  $x_\alpha$ , aby 100 $\alpha$ % hodnot znaku bylo nejvýše  $x_\alpha$  a zbylé byly alespoň  $x_\alpha$ . Pokud takový znak není určen jednoznačně, volíme zpravidla průměr mezi dvěmi možnými hodnotami. Nejobvyklejší jsou:

- **medián** (často také výběrový medián) definovaný vztahem  $\tilde{x} = x_{(\frac{n+1}{2})}$  pro liché  $n$  a  $\tilde{x} = \frac{1}{2}(x_{(n/2)} + x_{(n/2+1)})$ ;
- **dolní a horní kvartil**  $Q_1 = x_{0,25}$  a  $Q_3 = x_{0,75}$ ;
- **$p$ -tý kvantil** (též výběrový kvantil nebo percentil)  $x_p$ , kde  $0 < p < 1$  (zpravidla zadaný na dvě desetinná místa).

Lze se setkat také s hodnotou **modus**, která udává hodnotu znaku s největší četností.



# Míry variability statistických znaků

Rozumným požadavkem na jakoukoliv míru variability je její invariance vůči konstantním posunutím.

## Definition

- **Rozptyl** souboru znaků  $x$  je definován vztahem

$$s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_i)^2 = \frac{1}{n} \sum_{j=1}^m n_j (a_j - \bar{x})^2$$

případně v jmenovateli zlomku používáme  $(n - 1)$ .

- **Směrodatná odchylka** je dána jako odmocnina z výběrového rozptylu.
- **Rozpětí výběru** je  $R = x_{(n)} - x_{(1)}$ , **kvartilové rozpětí** je  $Q = Q_3 - Q_1$ .

Rozptyl je vlastně zprůměrovaný kvadrát standardní euklidovské vzdálenosti vektoru výběrových hodnot od jejich střední hodnoty. Díky této definici se chová velice přirozeně a budeme se s ním často potkávat.

Používá se také tzv. **průměrná odchylka**

$$d_x = \frac{1}{n} \sum_{i=1}^n |x_i - \tilde{x}|.$$

Všimněme si, že tady jde o skutečný průměr vzdáleností hodnot znaků, ovšem od mediánu!

Rozptyl je vlastně zprůměrovaný kvadrát standardní euklidovské vzdálenosti vektoru výběrových hodnot od jejich střední hodnoty. Díky této definici se chová velice přirozeně a budeme se s ním často potkávat.

Používá se také tzv. **průměrná odchylka**

$$d_x = \frac{1}{n} \sum_{i=1}^n |x_i - \tilde{x}|.$$

Všimněme si, že tady jde o skutečný průměr vzdáleností hodnot znaků, ovšem od mediánu!

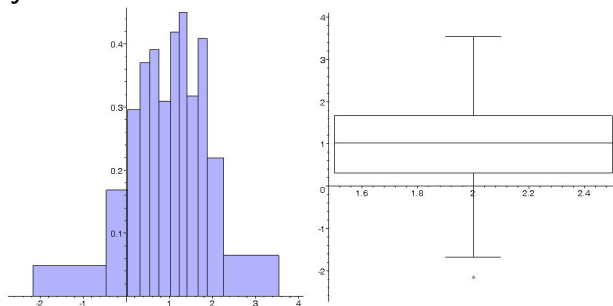
Následující věta říká, proč zrovna tyto míry volíme:

### Theorem

- Funkce  $S(t) = (1/n) \sum_{i=1}^n (x_i - t)^2$  nabývá svého minima pro  $t = \bar{x}$ , tj. pro výběrový průměr.
- Funkce  $D(t) = (1/n) \sum_{i=1}^n |x_i - t|$  nabývá svého minima pro  $t = \tilde{x}$ , tj. pro medián.

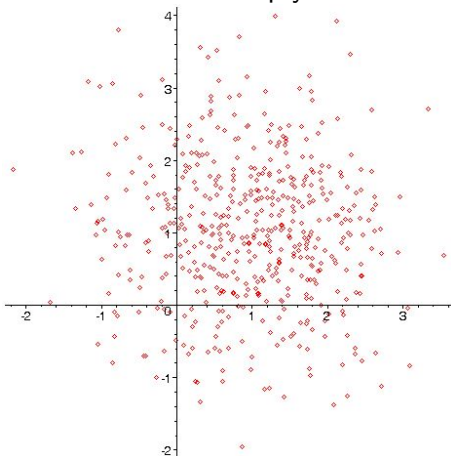
# Diagramy

Pro rychlé vstřebávání složitější strukturovaných informací je člověk skvěle vybaven zrakově. Proto se pro zobrazení statistiky jednotlivých znaků nebo jejich korelací používá mnoho standardizovaných nástrojů. Jedním z nich jsou tzv. **krabicové diagramy**.



Střední linka je medián, kraje boxu jsou kvartily, "packy" ukazují 1,5 kvartilového rozsahu, ne však víc než kraje rozsahu výběru, případné hodnoty mimo jsou přímo naznačeny body.

Běžné zobrazovací nástroje nám umožňují dobře vidět případné závislosti dvou výběrů zjištěných znaků. Např. na obrázku jsou za souřadnice voleny hodnoty ze dvou nezávislých výběrů z normálních rozdělení se střední hodnotou 1 a rozptylem 1.



# Plán přednášky

- 1 Literatura
- 2 Co je statistika?
- 3 Popisná statistika
  - Míry polohy statistických znaků
  - Míry variability statistických znaků
- 4 Pravděpodobnost
- 5 Náhodné veličiny

Připomeneme (a trochu zobecníme) pojmy a výsledky z druhé přednášky prvního semestru.

### Definition (Náhodné jevy)

Budeme pracovat s neprázdnou pevně zvolenou množinou  $\Omega$  všech možných výsledků, kterou nazýváme **základní prostor**.

Připomeneme (a trochu zobecníme) pojmy a výsledky z druhé přednášky prvního semestru.

### Definition (Náhodné jevy)

Budeme pracovat s neprázdnou pevně zvolenou množinou  $\Omega$  všech možných výsledků, kterou nazýváme **základní prostor**. Prvky  $\omega \in \Omega$  představují jednotlivé **možné výsledky**.



Připomeneme (a trochu zobecníme) pojmy a výsledky z druhé přednášky prvního semestru.

### Definition (Náhodné jevy)

Budeme pracovat s neprázdnou pevně zvolenou množinou  $\Omega$  všech možných výsledků, kterou nazýváme **základní prostor**.

Prvky  $\omega \in \Omega$  představují jednotlivé **možné výsledky**.

Systém podmnožin  $\mathcal{A}$  základního prostoru se nazývá **jevové pole** a jeho prvky se nazývají **jevy**, jestliže

- $\Omega \in \mathcal{A}$ , tj. základní prostor, je jevem,
- je-li  $A, B \in \mathcal{A}$ , pak  $A \setminus B \in \mathcal{A}$ , tj. pro každé dva jevy je jevem i jejich množinový rozdíl,
- je-li  $A_i \in \mathcal{A}$ ,  $i \in I$  nejvýše spočetný systém jevů, pak také jejich sjednocení je jevem, tj.  $\cup_{i \in I} A_i \in \mathcal{A}$ .

- Komplement  $A^c = \Omega \setminus A$  jevu  $A$  je jevem, který nazýváme *opačný jev* k jevu  $A$ .

- Komplement  $A^c = \Omega \setminus A$  jevu  $A$  je jevem, který nazýváme *opačný jev* k jevu  $A$ .
- Průnik dvou jevů opět jevem, protože pro každé dvě podmnožiny  $A, B \subset \Omega$  platí

$$A \setminus (\Omega \setminus B) = A \cap B.$$

- Komplement  $A^c = \Omega \setminus A$  jevu  $A$  je jevem, který nazýváme *opačný jev* k jevu  $A$ .
- Průnik dvou jevů opět jevem, protože pro každé dvě podmnožiny  $A, B \subset \Omega$  platí

$$A \setminus (\Omega \setminus B) = A \cap B.$$

Jevové pole je tedy systém podmnožin základního prostoru uzavřený na konečné průniky, spočetná sjednocení a množinové rozdíly. Jednotlivé množiny  $A \in \mathcal{A}$  nazýváme **náhodné jevy** (vzhledem k  $\mathcal{A}$ ).

Terminologie připomíná souvislosti s popisem skutečných jevů a jejich statistickým popisem:

- celý základní prostor  $\Omega$  se nazývá **jistý jev**, prázdná podmnožina  $\emptyset \in \mathcal{A}$  se nazývá **nemožný jev**,

Terminologie připomíná souvislosti s popisem skutečných jevů a jejich statistickým popisem:

- celý základní prostor  $\Omega$  se nazývá **jistý jev**, prázdná podmnožina  $\emptyset \in \mathcal{A}$  se nazývá **nemožný jev**,
- jednoprvkové podmnožiny  $\{\omega\} \in \Omega$  se nazývají **elementární jevy**,

Terminologie připomíná souvislosti s popisem skutečných jevů a jejich statistickým popisem:

- celý základní prostor  $\Omega$  se nazývá **jistý jev**, prázdná podmnožina  $\emptyset \in \mathcal{A}$  se nazývá **nemožný jev**,
- jednoprvkové podmnožiny  $\{\omega\} \in \Omega$  se nazývají **elementární jevy**,
- **společné nastoupení jevů**  $A_i, i \in I$ , odpovídá jevu  $\bigcap_{i \in I} A_i$ ,  
**nastoupení alespoň jednoho z jevů**  $A_i, i \in I$ , odpovídá jevu  $\bigcup_{i \in I} A_i$ ,

Terminologie připomíná souvislosti s popisem skutečných jevů a jejich statistickým popisem:

- celý základní prostor  $\Omega$  se nazývá **jistý jev**, prázdná podmnožina  $\emptyset \in \mathcal{A}$  se nazývá **nemožný jev**,
- jednoprvkové podmnožiny  $\{\omega\} \in \Omega$  se nazývají **elementární jevy**,
- **společné nastoupení jevů**  $A_i, i \in I$ , odpovídá jevu  $\bigcap_{i \in I} A_i$ ,  
**nastoupení alespoň jednoho z jevů**  $A_i, i \in I$ , odpovídá jevu  $\bigcup_{i \in I} A_i$ ,
- $A, B \in \mathcal{A}$  jsou **neslučitelné jevy**, je-li  $A \cap B = \emptyset$ ,



Terminologie připomíná souvislosti s popisem skutečných jevů a jejich statistickým popisem:

- celý základní prostor  $\Omega$  se nazývá **jistý jev**, prázdná podmnožina  $\emptyset \in \mathcal{A}$  se nazývá **nemožný jev**,
- jednoprvkové podmnožiny  $\{\omega\} \in \Omega$  se nazývají **elementární jevy**,
- **společné nastoupení jevů**  $A_i, i \in I$ , odpovídá jevu  $\bigcap_{i \in I} A_i$ , **nastoupení alespoň jednoho z jevů**  $A_i, i \in I$ , odpovídá jevu  $\bigcup_{i \in I} A_i$ ,
- $A, B \in \mathcal{A}$  jsou **neslučitelné jevy**, je-li  $A \cap B = \emptyset$ ,
- jev  $A$  má za **důsledek** jev  $B$ , když  $A \subset B$ ,

Terminologie připomíná souvislosti s popisem skutečných jevů a jejich statistickým popisem:

- celý základní prostor  $\Omega$  se nazývá **jistý jev**, prázdná podmnožina  $\emptyset \in \mathcal{A}$  se nazývá **nemožný jev**,
- jednoprvkové podmnožiny  $\{\omega\} \in \Omega$  se nazývají **elementární jevy**,
- **společné nastoupení jevů**  $A_i, i \in I$ , odpovídá jevu  $\bigcap_{i \in I} A_i$ , **nastoupení alespoň jednoho z jevů**  $A_i, i \in I$ , odpovídá jevu  $\bigcup_{i \in I} A_i$ ,
- $A, B \in \mathcal{A}$  jsou **neslučitelné jevy**, je-li  $A \cap B = \emptyset$ ,
- jev  $A$  má za **důsledek** jev  $B$ , když  $A \subset B$ ,
- je-li  $A \in \mathcal{A}$ , pak se jev  $B = \Omega \setminus A$  nazývá **opačný jev k jevu**  $A$ , píšeme  $B = A^c$ .

## Definition (Pravděpodobnost)

**Pravděpodobnostní prostor** je jevové pole  $\mathcal{A}$  podmnožin (konečného) základního prostoru  $\Omega$ , na kterém je definována skalární funkce  $P : \mathcal{A} \rightarrow \mathbb{R}$  s následujícími vlastnosti:

- je nezáporná, tj.  $P(A) \geq 0$  pro všechny jevy  $A$ ,
- je aditivní, tj.  $P(\cup_{i \in I} A_i) = \sum_{i \in I} P(A_i)$ , pro každý nejvýše spočetný systém po dvou disjunktních jevů,
- pravděpodobnost jistého jevu je 1.

Funkci  $P$  nazýváme **pravděpodobností** na jevovém poli  $(\Omega, \mathcal{A})$ .

## Definition (Pravděpodobnost)

**Pravděpodobnostní prostor** je jevové pole  $\mathcal{A}$  podmnožin (konečného) základního prostoru  $\Omega$ , na kterém je definována skalární funkce  $P : \mathcal{A} \rightarrow \mathbb{R}$  s následujícími vlastnosti:

- je nezáporná, tj.  $P(A) \geq 0$  pro všechny jevy  $A$ ,
- je aditivní, tj.  $P(\cup_{i \in I} A_i) = \sum_{i \in I} P(A_i)$ , pro každý nejvýše spočetný systém po dvou disjunktních jevů,
- pravděpodobnost jistého jevu je 1.

Funkci  $P$  nazýváme **pravděpodobností** na jevovém poli  $(\Omega, \mathcal{A})$ .

## Důsledky

Pro všechny jevy platí  $P(A^c) = 1 - P(A)$ .

## Definition (Pravděpodobnost)

**Pravděpodobnostní prostor** je jevové pole  $\mathcal{A}$  podmnožin (konečného) základního prostoru  $\Omega$ , na kterém je definována skalární funkce  $P : \mathcal{A} \rightarrow \mathbb{R}$  s následujícími vlastnosti:

- je nezáporná, tj.  $P(A) \geq 0$  pro všechny jevy  $A$ ,
- je aditivní, tj.  $P(\cup_{i \in I} A_i) = \sum_{i \in I} P(A_i)$ , pro každý nejvýše spočetný systém po dvou disjunktních jevů,
- pravděpodobnost jistého jevu je 1.

Funkci  $P$  nazýváme **pravděpodobností** na jevovém poli  $(\Omega, \mathcal{A})$ .

## Důsledky

Pro všechny jevy platí  $P(A^c) = 1 - P(A)$ .

Additivnost platí pro jakýkoliv spočetný počet neslučitelných jevů  $A_i \subset \Omega$ ,  $i \in I$ , tj.

$$P(\cup_{i \in I} A_i) = \sum_{i \in I} P(A_i), \text{ kdykoliv je } A_i \cap A_j = \emptyset, i \neq j, i, j \in I.$$

Připomeňme si klasickou konečnou pravděpodobnost.

Připomeňme si klasickou konečnou pravděpodobnost.

### Definition

Nechť  $\Omega$  je konečný základní prostor a necht' jevové pole  $\mathcal{A}$  je právě systém všech podmnožin v  $\Omega$ . **Klasická pravděpodobnost** je pravděpodobnostní prostor  $(\Omega, \mathcal{A}, P)$  s pravděpodobnostní funkcí  $P : \mathcal{A} \rightarrow \mathbb{R}$ ,

$$P(A) = \frac{|A|}{|\Omega|}.$$

Zjevně takto zadaná funkce skutečně definuje pravděpodobnost.

# Peterburgský paradox (Bernoulli, 1738)

Typický příklad klasické pravděpodobnosti jsou jevy související s házením mincí. Představme si následující pravidla kasina:



## Peterburgský paradox (Bernoulli, 1738)

Typický příklad klasické pravděpodobnosti jsou jevy související s házením mincí. Představme si následující pravidla kasina: Návštěvník zaplatí vklad  $C$  a poté hází mincí. Je-li  $T$  počet hodů potřebných k první hlavě, pak obdrží výhru  $2^T$ . Jaká je „fér hodnota“ pro vklad  $C$ ?

# Peterburgský paradox (Bernoulli, 1738)

Typický příklad klasické pravděpodobnosti jsou jevy související s házením mincí. Představme si následující pravidla kasina:

Návštěvník zaplatí vklad  $C$  a poté hází mincí. Je-li  $T$  počet hodů potřebných k první hlavě, pak obdrží výhru  $2^T$ . Jaká je „fér hodnota“ pro vklad  $C$ ?

Pravděpodobnost, že padne hlava je u férové mince  $1/2$ , je proto  $P(T = k) = 2^{-k}$ . Sečteme-li všechny pravděpodobnosti výsledků vynásobených výhrami  $2^k$ , dostaneme  $\sum_1^{\infty} 1 = \infty$ . Zdá se proto, že se vyplatí vložit i velký vklad...

# Peterburgský paradox (Bernoulli, 1738)

Typický příklad klasické pravděpodobnosti jsou jevy související s házením mincí. Představme si následující pravidla kasina:

Návštěvník zaplatí vklad  $C$  a poté hází mincí. Je-li  $T$  počet hodů potřebných k první hlavě, pak obdrží výhru  $2^T$ . Jaká je „fér hodnota“ pro vklad  $C$ ?

Pravděpodobnost, že padne hlava je u férové mince  $1/2$ , je proto  $P(T = k) = 2^{-k}$ . Sečteme-li všechny pravděpodobnosti výsledků vynásobených výhrami  $2^k$ , dostaneme  $\sum_1^{\infty} 1 = \infty$ . Zdá se proto, že se vyplatí vložit i velký vklad. . .

Ve skutečnosti simulací hry zjistíme, že nezávisle na počtu pokusů se prakticky všechny výhry budou pohybovat v rozmezí  $T$  do 6.

Důvodem je, že vysoké výhry jsou velice nepravděpodobné a proto je při reálných úvahách nelze brát vážně.

# Podmíněná pravděpodobnost

Obvyklé je také klást dotazy s dodatečnou podmínkou. Např. „jaká je pravděpodobnost, že při hodu dvěma kostkami padly dvě pětky, je-li součet hodnot deset?“. Připomeneme, že formalizovat takové úvahy umíme následovně.

## Definition

Nechť  $H$  je jev s nenulovou pravděpodobností v jevovém poli  $\mathcal{A}$  v pravděpodobnostním prostoru  $(\Omega, \mathcal{A}, P)$ . **Podmíněná pravděpodobnost**  $P(A|H)$  jevu  $A \in \mathcal{A}$  vzhledem k hypotéze  $H$  je definována vztahem

$$P(A|H) = \frac{P(A \cap H)}{P(H)}.$$

Definice odpovídá požadavku, že jevy  $A$  a  $H$  nastanou zároveň, za předpokladu, že  $A$  nastal s pravděpodobností  $P(A \cap H)/P(A)$ .

# Podmíněná pravděpodobnost

Obvyklé je také klást dotazy s dodatečnou podmínkou. Např. „jaká je pravděpodobnost, že při hodu dvěma kostkami padly dvě pětky, je-li součet hodnot deset?“. Připomeneme, že formalizovat takové úvahy umíme následovně.

## Definition

Nechť  $H$  je jev s nenulovou pravděpodobností v jevovém poli  $\mathcal{A}$  v pravděpodobnostním prostoru  $(\Omega, \mathcal{A}, P)$ . **Podmíněná pravděpodobnost**  $P(A|H)$  jevu  $A \in \mathcal{A}$  vzhledem k hypotéze  $H$  je definována vztahem

$$P(A|H) = \frac{P(A \cap H)}{P(H)}.$$

Definice odpovídá požadavku, že jevy  $A$  a  $H$  nastanou zároveň, za předpokladu, že  $A$  nastal s pravděpodobností  $P(A \cap H)/P(A)$ . Je také vidět přímo z definice, hypotéza  $H$  a jev  $A$  jsou nezávislé tehdy a jen tehdy, je-li  $P(A) = P(A|H)$ .

# Bayesovy věty

Přepsáním formule pro podmíněnou pravděpodobnost dostáváme

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

## Theorem (Bayesovy věty)

*Pro pravděpodobnost jevů  $A$  a  $B$  platí*

- 1  $P(A|B) = \frac{P(A)P(B|A)}{P(B)}.$
- 2  $P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A)+P(A')P(B|A')}.$

# Bayesovy věty

Přepsáním formule pro podmíněnou pravděpodobnost dostáváme

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

## Theorem (Bayesovy věty)

*Pro pravděpodobnost jevů  $A$  a  $B$  platí*

- 1  $P(A|B) = \frac{P(A)P(B|A)}{P(B)}.$
- 2  $P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A')P(B|A')}.$

## Důkaz.

První tvrzení je přepsáním předchozí formule, druhé z prvního plyne dosazením  $P(B) = P(A)P(B|A) + P(A')P(B|A')$ . □

## Příklad – preventivní screening

Předpokládejme, že krevní test na HIV pozitivní osoby má 99% správnost v případě osoby skutečně HIV pozitivní. Zároveň předpokládejme, že u HIV negativní osoby dopadně test pozitivně v 0.2% případů.



## Příklad – preventivní screening

Předpokládejme, že krevní test na HIV pozitivní osoby má 99% správnost v případě osoby skutečně HIV pozitivní. Zároveň předpokládejme, že u HIV negativní osoby dopadne test pozitivně v 0.2% případů.

Náhodně z populace vyberem osobu a otestujeme pozitivně. S jakou pravděpodobností je skutečně HIV pozitivní, jestliže četnost výskytu HIV v populaci je  $p$  promile (tj.  $p$  osob z tisíce je skutečně HIV pozitivní).

## Příklad – preventivní screening

Předpokládejme, že krevní test na HIV pozitivní osoby má 99% správnost v případě osoby skutečně HIV pozitivní. Zároveň předpokládejme, že u HIV negativní osoby dopadne test pozitivně v 0.2% případech.

Náhodně z populace vyberem osobu a otestujeme pozitivně. S jakou pravděpodobností je skutečně HIV pozitivní, jestliže četnost výskytu HIV v populaci je  $p$  promile (tj.  $p$  osob z tisíce je skutečně HIV pozitivní).

Označme  $A$  jev, že je daná osoba HIV pozitivní, a  $B$  jev, že daná osoba má pozitivní test. Dle druhé Bayesovy věty je hledaná pravděpodobnost

$$P(A|B) = \frac{p/1000 \cdot 99/100}{p/1000 \cdot 99/100 + (1000 - p)/1000 \cdot 2/1000}$$

Jestliže zvolíme za  $p$  nějaké konkrétní četnosti, dostaneme příslušné očekávatelné spolehlivosti testu. V následující tabulce je spočten výsledek pro několik  $p$ :

$p$	100	10	1	0.1
$P(A B)$	0.982	0.8333	0.3313	0.0471

Výsledek asi neodpovídá naší intuici a může se zdát šokující ve vztahu k použití takovýchto testů.

Výsledek asi neodpovídá naší intuici a může se zdát šokující ve vztahu k použití takovýchto testů.

Evidentně prostý výběr náhodné osoby a použití jediného testu, byť velmi citlivého, specifického a účinného, nejsou vhodné ani na otestování skutečného stavu populace, ani na preventivní vyšetření jednotlivců, pokud nemáme další podpůrné informace a lepší nástroje.

Výsledek asi neodpovídá naší intuici a může se zdát šokující ve vztahu k použití takovýchto testů.

Evidentně prostý výběr náhodné osoby a použití jediného testu, byť velmi citlivého, specifického a účinného, nejsou vhodné ani na otestování skutečného stavu populace, ani na preventivní vyšetření jednotlivců, pokud nemáme další podpůrné informace a lepší nástroje.

Právě matematická statistika dává nástroje na kvalifikovanější postupy v medicínské i průmyslové diagnostice, ekonomických modelech, vyhodnocování experimentálních dat atd.

# Plán přednášky

- 1 Literatura
- 2 Co je statistika?
- 3 Popisná statistika
  - Míry polohy statistických znaků
  - Míry variability statistických znaků
- 4 Pravděpodobnost
- 5 **Náhodné veličiny**

Vraťme se k jednoduchému a názornému příkladu statistik kolem výsledků studentů<sup>1</sup> v daném předmětu. Je a není podobný klasické pravděpodobnosti a s ní související statistice při házení kostkou.

---

<sup>1</sup>Myslíme samozřejmě na „studenty a studentky“, pro zestručnění textu ale používám podobně jako v legislativních textech bezpohlavní označení „student“



Vraťme se k jednoduchému a názornému příkladu statistik kolem výsledků studentů<sup>1</sup> v daném předmětu. Je a není podobný klasické pravděpodobnosti a s ní související statistice při házení kostkou. Na jedné straně jsme připustili pouze konečný počet možných bodových hodnocení (celá čísla od 0 do 20), zároveň ale není patrně vhodné představovat si výsledky jednotlivých studentů jako analogii nezávislého házení kostkou (to by byla skutečně divně vedená přednáška).

---

<sup>1</sup>Myslíme samozřejmě na „studenty a studentky“, pro zestručnění textu ale používám podobně jako v legislativních textech bezpohlavní označení „student“

Vraťme se k jednoduchému a názornému příkladu statistik kolem výsledků studentů<sup>1</sup> v daném předmětu. Je a není podobný klasické pravděpodobnosti a s ní související statistice při házení kostkou. Na jedné straně jsme připustili pouze konečný počet možných bodových hodnocení (celá čísla od 0 do 20), zároveň ale není patrně vhodné představovat si výsledky jednotlivých studentů jako analogii nezávislého házení kostkou (to by byla skutečně divně vedená přednáška).

Místo toho máme na základním prostoru  $\Omega$  všech studentů definovanou funkci bodového ohodnocení  $X : \Omega \rightarrow \mathbb{R}$ . Je to typický příklad **náhodné veličiny**.

S každou náhodnou veličinou potřebujeme umět pracovat s vhodnou množinou jevů. Zpravidla požadujeme, abychom mohli pracovat s pravděpodobnostmi příslušnosti hodnoty  $X$  do předem zadaného intervalu.

---

<sup>1</sup>Myslíme samozřejmě na „studenty a studentky“, pro zestručnění textu ale používám podobně jako v legislativních textech bezpohlavní označení „student“

Na prostoru  $\mathbb{R}^k$  uvažujme nejmenší jevové pole  $\mathcal{B}$  obsahující všechny  $k$ -rozměrné intervaly. Množinám v  $\mathcal{B}$  říkáme **Borelovské množiny** na  $\mathbb{R}^k$ .

Na prostoru  $\mathbb{R}^k$  uvažujme nejmenší jevové pole  $\mathcal{B}$  obsahující všechny  $k$ -rozměrné intervaly. Množinám v  $\mathcal{B}$  říkáme **Borelovské množiny** na  $\mathbb{R}^k$ .

### Definition (Náhodné veličiny a distribuční funkce)

**Náhodná veličina**  $X$  na pravděpodobnostním prostoru  $(\Omega, \mathcal{A}, P)$  je taková funkce  $X : \Omega \rightarrow \mathbb{R}$ , že vzor  $X^{-1}(B)$  patří do  $\mathcal{A}$  pro každou Borelovskou množinu  $B \in \mathcal{B}$  na  $\mathbb{R}$ .

Na prostoru  $\mathbb{R}^k$  uvažujme nejmenší jevové pole  $\mathcal{B}$  obsahující všechny  $k$ -rozměrné intervaly. Množinám v  $\mathcal{B}$  říkáme **Borelovské množiny** na  $\mathbb{R}^k$ .

### Definition (Náhodné veličiny a distribuční funkce)

**Náhodná veličina**  $X$  na pravděpodobnostním prostoru  $(\Omega, \mathcal{A}, P)$  je taková funkce  $X : \Omega \rightarrow \mathbb{R}$ , že vzor  $X^{-1}(B)$  patří do  $\mathcal{A}$  pro každou Borelovskou množinu  $B \in \mathcal{B}$  na  $\mathbb{R}$ .

**Náhodný vektor**  $(X_1, \dots, X_k)$  na  $(\Omega, \mathcal{A}, P)$  je  $k$ -tice náhodných veličin.

Definice náhodné veličiny zajišťuje, že pro všechny  $-\infty \leq a \leq b \leq \infty$  existuje pravděpodobnost  $P(a < X \leq b)$ , kde používáme stručné značení pro jev  $A = (\omega \in \Omega; a < X(\omega) \leq b)$ .

### Definition

**Distribuční funkcí** náhodné veličiny  $X$  je funkce  $F : \mathbb{R} \rightarrow \mathbb{R}$  definovaná pro všechny  $x \in \mathbb{R}$  vztahem

$$F(x) = P(X \leq x).$$

Definice náhodné veličiny zajišťuje, že pro všechny  $-\infty \leq a \leq b \leq \infty$  existuje pravděpodobnost  $P(a < X \leq b)$ , kde používáme stručné značení pro jev  $A = (\omega \in \Omega; a < X(\omega) \leq b)$ .

### Definition

**Distribuční funkcí** náhodné veličiny  $X$  je funkce  $F : \mathbb{R} \rightarrow \mathbb{R}$  definovaná pro všechny  $x \in \mathbb{R}$  vztahem

$$F(x) = P(X \leq x).$$

**Distribuční funkcí** náhodného vektoru  $(X_1, \dots, X_k)$  je funkce  $F : \mathbb{R}^k \rightarrow \mathbb{R}$  definovaná pro všechny  $(x_1, \dots, x_k) \in \mathbb{R}^k$  vztahem

$$F(x) = P(X_1 \leq x_1 \wedge \dots \wedge X_k \leq x_k).$$

# Diskrétní náhodné veličiny

Předpokládejme, že pro náhodná veličina  $X$  na pravděpodobnostním prostoru  $(\Omega, \mathcal{A}, P)$  nabývá jen konečně mnoha hodnot  $x_1, x_2, \dots, x_n \in \mathbb{R}$ . Pak existuje tzv. **pravděpodobnostní funkce**  $f(x)$  taková, že

$$f(x) = \begin{cases} P(X = x_i) & x = x_i \\ 0 & \text{jinak.} \end{cases}$$

Evidentně  $\sum_1^n f(x_i) = 1$ .



# Diskrétní náhodné veličiny

Předpokládejme, že pro náhodná veličina  $X$  na pravděpodobnostním prostoru  $(\Omega, \mathcal{A}, P)$  nabývá jen konečně mnoha hodnot  $x_1, x_2, \dots, x_n \in \mathbb{R}$ . Pak existuje tzv. **pravděpodobnostní funkce**  $f(x)$  taková, že

$$f(x) = \begin{cases} P(X = x_i) & x = x_i \\ 0 & \text{jinak.} \end{cases}$$

Evidentně  $\sum_1^n f(x_i) = 1$ .

Takové náhodné veličině se říká **diskrétní**.

# Diskrétní náhodné veličiny

Předpokládejme, že pro náhodná veličina  $X$  na pravděpodobnostním prostoru  $(\Omega, \mathcal{A}, P)$  nabývá jen konečně mnoha hodnot  $x_1, x_2, \dots, x_n \in \mathbb{R}$ . Pak existuje tzv. **pravděpodobnostní funkce**  $f(x)$  taková, že

$$f(x) = \begin{cases} P(X = x_i) & x = x_i \\ 0 & \text{jinak.} \end{cases}$$

Evidentně  $\sum_1^n f(x_i) = 1$ .

Takové náhodné veličině se říká **diskrétní**.

Každá náhodná veličina definovaná pro klasickou pravděpodobnost je diskrétní.

# Diskrétní náhodné veličiny

Předpokládejme, že pro náhodná veličina  $X$  na pravděpodobnostním prostoru  $(\Omega, \mathcal{A}, P)$  nabývá jen konečně mnoha hodnot  $x_1, x_2, \dots, x_n \in \mathbb{R}$ . Pak existuje tzv. **pravděpodobnostní funkce**  $f(x)$  taková, že

$$f(x) = \begin{cases} P(X = x_i) & x = x_i \\ 0 & \text{jinak.} \end{cases}$$

Evidentně  $\sum_1^n f(x_i) = 1$ .

Takové náhodné veličině se říká **diskrétní**.

Každá náhodná veličina definovaná pro klasickou pravděpodobnost je diskrétní. Obdobně lze definici pravděpodobnostní funkce rozšířit na veličiny se spočetně mnoha hodnotami (pracujeme pak s nekonečnými řadami :-)

# Spojité náhodné veličiny

I když hodnoty náhodné veličiny  $X$  nejsou diskrétní, můžeme postupovat podobně s užitím ideí diferenciálního a integrálního počtu. Intuitivně lze uvažovat takto: **hustotu**  $f(x)$  **pravděpodobnosti** pro  $X$  si představíme jako

$$P(x < X \leq x + dx) = f(x)dx.$$

# Spojité náhodné veličiny

I když hodnoty náhodné veličiny  $X$  nejsou diskrétní, můžeme postupovat podobně s užitím ideí diferenciálního a integrálního počtu. Intuitivně lze uvažovat takto: **hustotu**  $f(x)$  **pravděpodobnosti** pro  $X$  si představíme jako

$$P(x < X \leq x + dx) = f(x)dx.$$

To znamená, že chceme pro  $-\infty \leq a \leq b \leq \infty$

$$P(a < X \leq b) = \int_a^b f(x)dx. \quad (*)$$

# Spojité náhodné veličiny

I když hodnoty náhodné veličiny  $X$  nejsou diskrétní, můžeme postupovat podobně s užitím ideí diferenciálního a integrálního počtu. Intuitivně lze uvažovat takto: **hustotu**  $f(x)$  **pravděpodobnosti** pro  $X$  si představíme jako

$$P(x < X \leq x + dx) = f(x)dx.$$

To znamená, že chceme pro  $-\infty \leq a \leq b \leq \infty$

$$P(a < X \leq b) = \int_a^b f(x)dx. \quad (*)$$

## Definition

Náhodná veličina  $X$ , pro kterou existuje její **hustota pravděpodobnosti** splňující (\*), se nazývá **spojitá**.

## Theorem

*Nechť  $X$  je náhodná veličina,  $F(x)$  je její distribuční funkce.*

- 1  *$F$  je zprava spojitá,  $\lim_{x \rightarrow -\infty} F(x) = 0$  a  $\lim_{x \rightarrow \infty} F(x) = 1$ .*
- 2 *Je-li  $X$  diskrétní s hodnotami  $x_1, \dots, x_n$ , pak je  $F(x)$  po částech konstantní,  $F(x) = \sum_{x_i \leq x} P(X = x_i)$  a  $F(x) = 1$  kdykoliv  $x \geq x_n$ .*
- 3 *Je-li  $X$  spojitá, pak je  $F(x)$  diferencovatelná a její derivace se rovná hustotě  $X$ , tj. platí  $F'(x) = f(x)$ .*