Základy biostatistiky

(MD710P09)

ak. rok 2008/2009

Karel Zvára

karel.zvara@mff.cuni.cz
http://www.karlin.mff.cuni.cz/∼zvara
katedra pravděpodobnosti a matematické statistiky MFF UK

(naposledy upraveno 17. března 2009)

sdružené rozdělení

- abychom mohli popsat **závislost** náhodných veličin, zajímáme se o **společné** chování dvojice (trojice,...) náhodných veličin, tedy chování **náhodného vektoru**
- příklad **rodina**
  - X – počet děvčat v rodině s třemi dětmi
  - Y – počet děvčat mezi dvěma staršími dětmi
  - Z – počet hochů v rodině s třemi dětmi
- zajímá nás rozdělení náhodného vektoru (X, Y)
- proč nemá smysl vyšetřovat **vektor** (X, Z)?
  (protože Z je určeno X jednoznačně: $Z = 3 - X$)

sdružené rozdělení

- abychom mohli popsat **závislost** náhodných veličin, zajímáme se o **společné** chování dvojice (trojice,. . .) náhodných veličin, tedy chování **náhodného vektoru**
- příklad **rodina**
  - X – počet děvčat v rodině s třemi dětmi
  - Y – počet děvčat mezi dvěma staršími dětmi
  - Z – počet hochů v rodině s třemi dětmi
- zajímá nás rozdělení náhodného vektoru (X, Y)
- proč nemá smysl vyšetřovat **vektor** (X, Z)?
  (protože Z je určeno X jednoznačně: $Z = 3 - X$)

sdružené rozdělení

- abychom mohli popsat **závislost** náhodných veličin, zajímáme se o **společné** chování dvojice (trojice,...) náhodných veličin, tedy chování **náhodného vektoru**
- příklad **rodina**
  - X – počet děvčat v rodině s třemi dětmi
  - Y – počet děvčat mezi dvěma staršími dětmi
  - Z – počet hochů v rodině s třemi dětmi
- zajímá nás rozdělení náhodného vektoru (X, Y)
- proč nemá smysl vyšetřovat **vektor** (X, Z)? (protože Z je určeno X jednoznačně: $Z = 3 - X$)

sdružené rozdělení

- abychom mohli popsat **závislost** náhodných veličin, zajímáme se o **společné** chování dvojice (trojice,. . .) náhodných veličin, tedy chování **náhodného vektoru**
- příklad **rodina**
  - X – počet děvčat v rodině s třemi dětmi
  - Y – počet děvčat mezi dvěma staršími dětmi
  - Z – počet hochů v rodině s třemi dětmi
- zajímá nás rozdělení náhodného vektoru (X, Y)
- proč nemá smysl vyšetřovat **vektor** (X, Z)?
  (protože Z je určeno X jednoznačně: $Z = 3 - X$)

sdružené rozdělení

- abychom mohli popsat **závislost** náhodných veličin, zajímáme se o **společné** chování dvojice (trojice,...) náhodných veličin, tedy chování **náhodného vektoru**
- příklad **rodina**
  - X – počet děvčat v rodině s třemi dětmi
  - Y – počet děvčat mezi dvěma staršími dětmi
  - Z – počet hochů v rodině s třemi dětmi
- zajímá nás rozdělení náhodného vektoru (X, Y)
- proč nemá smysl vyšetřovat **vektor** (X, Z)?
  (protože Z je určeno X jednoznačně: $Z = 3 - X$)

sdružené rozdělení

- abychom mohli popsat **závislost** náhodných veličin, zajímáme se o **společné** chování dvojice (trojice,...) náhodných veličin, tedy chování **náhodného vektoru**
- příklad **rodina**
  - X – počet děvčat v rodině s třemi dětmi
  - Y – počet děvčat mezi dvěma staršími dětmi
  - Z – počet hochů v rodině s třemi dětmi
- zajímá nás rozdělení náhodného vektoru (X, Y)
- proč nemá smysl vyšetřovat **vektor** (X, Z)?
  (protože Z je určeno X jednoznačně: $Z = 3 - X$)

sdružené rozdělení

- abychom mohli popsat **závislost** náhodných veličin, zajímáme se o **společné** chování dvojice (trojice,...) náhodných veličin, tedy chování **náhodného vektoru**
- příklad **rodina**
  - X – počet děvčat v rodině s třemi dětmi
  - Y – počet děvčat mezi dvěma staršími dětmi
  - Z – počet hochů v rodině s třemi dětmi
- zajímá nás rozdělení náhodného vektoru (X, Y)
- proč nemá smysl vyšetřovat **vektor** (X, Z)? (protože Z je určeno X jednoznačně: $Z = 3 - X$)

sdružené, marginální a podmíněné rozdělení

sdružené rozdělení – popisuje **společné chování** X, Y

$$\boxed{\mathrm{P}(X = x_i^*, Y = y_j^*)} \text{ resp. } \boxed{f_{X,Y}(x,y)}$$

marginální rozdělení: chování jedné bez ohledu na hodnotu druhé

$$\boxed{\mathrm{P}(X = x_i^*) = \sum_j \mathrm{P}(X = x_i^*, Y = y_j^*) \qquad \forall x_i^*}$$

$$\boxed{\mathrm{P}(Y = y_j^*) = \sum_i \mathrm{P}(X = x_i^*, Y = y_j^*) \qquad \forall y_j^*}$$

podmíněné rozdělení: chování Y při **dané** hodnotě X

$$\mathrm{P}(Y = y_j^* | X = x_i^*) = \frac{\mathrm{P}(X = x_i^*, Y = y_j^*)}{\mathrm{P}(X = x_i^*)}$$

příklad rodina

X počet děvčat, *Y* počet děvčat mezi dvěma staršími dětmi

sdružené, marginální a podmíněné rozdělení

ω_i	x_i	y_i
(m, m, m)	0	0
(m, m, f)	1	1
(m, f, m)	1	1
(f, m, m)	1	0
(f, f, m)	2	1
(f, m, f)	2	1
(m, f, f)	2	2
(f, f, f)	3	2

x_i^*	y_j^*			celkem
	0	1	2	
0	1/8	0	0	1/8
1	1/8	2/8	0	3/8
2	0	2/8	1/8	3/8
3	0	0	1/8	1/8
	2/8	4/8	2/8	1

x_i^*	y_j^*			celkem
	0	1	2	
0	1	0	0	1
1	1/3	2/3	0	1
2	0	2/3	1/3	1
3	0	0	1	1

kovariance

protějšek s_{xy}, [covariance]

kovariance vyjadřuje vzájemnou závislost náhodných veličin:

$$\boxed{\sigma_{X,Y} = \mathrm{E}\,(X-\mu_X)(Y-\mu_Y)}$$

$$\boxed{\sigma_{X,Y} = \sum_i \sum_j (x_i^* - \mu_X)(y_j^* - \mu_Y)\mathrm{P}(X = x_i^*, Y = y_j^*)}$$

označení metody výpočtu: $\mathrm{cov}(X, Y)$

zřejmě platí $\boxed{\mathrm{cov}(X,X) = \mathrm{var}\,X}$ tj. $\boxed{\sigma_{X,X} = \sigma_X^2}$

náhodné veličiny jsou **nezávislé** právě tehdy, když platí **(ze znalosti hodnoty jedné nic nevíme o druhé)**

$$\boxed{\mathrm{P}(X = x_i^*, Y = y_j^*) = \mathrm{P}(X = x_i^*) \cdot \mathrm{P}(Y = y_j^*), \qquad \forall (x_i^*, y_j^*)}$$

jsou-li X, Y – nezávislé $\Rightarrow$ $\boxed{\sigma_{X,Y} = 0}$ (nikoliv obrácená implikace)

příklad rodina

x_i^*	y_j^* 0	1	2	celkem
0	0,125	0	0	0,125
1	0,125	0,250	0	0,375
2	0	0,250	0,125	0,375
3	0	0	0,125	0,125
celkem	0,250	0,500	0,250	1,000

$$\mu_X = 0 \cdot 0{,}125 + 1 \cdot 0{,}375 + 2 \cdot 0{,}375 + 3 \cdot 0{,}125 = 1{,}5$$
$$\mu_Y = 0 \cdot 0{,}250 + 1 \cdot 0{,}500 + 2 \cdot 0{,}250 = 1$$
$$\sigma_X^2 = (0 - 1{,}5)^2 \cdot 0{,}125 + \ldots + (3 - 1{,}5)^2 \cdot 0{,}125 = 0{,}75$$
$$\sigma_Y^2 = (0 - 1)^2 \cdot 0{,}25 + (1 - 1)^2 \cdot 0{,}5 + (2 - 1)^2 \cdot 0{,}25 = 0{,}5$$
$$\sigma_{XY} = (0 - 1{,}5) \cdot (0 - 1) \cdot 0{,}125 + \ldots = 0{,}5$$

X, Y jsou závislé, neboť např. $0{,}25 \cdot 0{,}125 \neq 0{,}125$

shrnutí vlastností populačního průměru a rozptylu

srovnej s požadavky na míry polohy a míry variability

$$\mu_{\alpha+X} = \alpha + \mu_X, \qquad \mu_{\beta X} = \beta \cdot \mu_X,$$
$$\sigma^2_{\alpha+X} = \sigma^2_X, \qquad \sigma^2_{\beta X} = \beta^2 \cdot \sigma^2_X,$$
$$\sigma_{\alpha+X} = \sigma_X, \qquad \sigma_{\beta X} = |\beta| \cdot \sigma_X,$$

pro součet náhodných veličin $X + Y$ dále platí

$$\mu_{X+Y} = \mu_X + \mu_Y$$
$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2\sigma_{XY} \qquad \text{obecně}$$
$$\sigma_{X,Y} = 0 \qquad \text{pro nezávislé } X, Y$$
$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y \qquad \text{pro nezávislé } X, Y$$

ukázka důkazu

$$\begin{aligned}
\mu_{\alpha+\beta X} &= \mathsf{E}\,(\alpha+\beta X) \\
&= \sum_i (\alpha+\beta x_i^*)\mathsf{P}(X=x_i^*) \\
&= \sum_i \alpha \mathsf{P}(X=x_i^*) + \sum_i \beta x_i^* \mathsf{P}(X=x_i^*) \\
&= \alpha \sum_i \mathsf{P}(X=x_i^*) + \beta \sum_i x_i^* \mathsf{P}(X=x_i^*) \\
&= \alpha + \beta \cdot \mathsf{E}\,X = \alpha + \beta \cdot \mu_X
\end{aligned}$$

normování náhodné veličiny X (populační obdoba z-skórů)

$$Z = \frac{X-\mu_X}{\sigma_X} \qquad \text{(bezrozměrné!)}$$

$$\Rightarrow \qquad \mu_Z = 0, \qquad \sigma_Z = 1$$

charakteristiky založené na normované verzi

charakteristiky X nezávislé na μ_X a σ_X, protějšky popisných statistik

- (populační) **korelační koeficient** [correlation coefficient]

$$\rho_{XY} = \mathrm{cov}\left(\frac{X-\mu_X}{\sigma_X}, \frac{Y-\mu_Y}{\sigma_Y}\right) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

- (populační) **šikmost** náhodné veličiny X [skewness]

$$\gamma_1 = \mathrm{E}\left(\frac{X-\mu_X}{\sigma_X}\right)^3 = \frac{\mathrm{E}\,(X-\mu_X)^3}{\sigma_X^3}$$

- (populační) **špičatost** náhodné veličiny X (někdy bez -3) [kurtosis]

$$\gamma_2 = \mathrm{E}\left(\frac{X-\mu_X}{\sigma_X}\right)^4 - 3 = \frac{\mathrm{E}\,(X-\mu_X)^4}{\sigma_X^4} - 3$$

charakteristiky založené na normované verzi

charakteristiky X nezávislé na μ_X a σ_X, protějšky popisných statistik

- (populační) **korelační koeficient** [correlation coefficient]

$$\rho_{XY} = \operatorname{cov}\left(\frac{X-\mu_X}{\sigma_X}, \frac{Y-\mu_Y}{\sigma_Y}\right) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

- (populační) **šikmost** náhodné veličiny X [skewness]

$$\gamma_1 = \mathrm{E}\left(\frac{X-\mu_X}{\sigma_X}\right)^3 = \frac{\mathrm{E}\,(X-\mu_X)^3}{\sigma_X^3}$$

- (populační) **špičatost** náhodné veličiny X (někdy bez -3) [kurtosis]

$$\gamma_2 = \mathrm{E}\left(\frac{X-\mu_X}{\sigma_X}\right)^4 - 3 = \frac{\mathrm{E}\,(X-\mu_X)^4}{\sigma_X^4} - 3$$

charakteristiky založené na normované verzi

charakteristiky X nezávislé na μ_X a σ_X, protějšky popisných statistik

- (populační) **korelační koeficient** [correlation coefficient]

$$\rho_{XY} = \operatorname{cov}\left(\frac{X-\mu_X}{\sigma_X}, \frac{Y-\mu_Y}{\sigma_Y}\right) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

- (populační) **šikmost** náhodné veličiny X [skewness]

$$\gamma_1 = \mathsf{E}\left(\frac{X-\mu_X}{\sigma_X}\right)^3 = \frac{\mathsf{E}\,(X-\mu_X)^3}{\sigma_X^3}$$

- (populační) **špičatost** náhodné veličiny X (někdy bez -3) [kurtosis]

$$\gamma_2 = \mathsf{E}\left(\frac{X-\mu_X}{\sigma_X}\right)^4 - 3 = \frac{\mathsf{E}\,(X-\mu_X)^4}{\sigma_X^4} - 3$$

alternativní rozdělení

nula-jedničkové, Bernoulliovo

- pouze dvě možné hodnoty: 1 (zdar), 0 (nezdar)
- $\mathrm{P}(X = 1) = \pi, \mathrm{P}(X = 0) = 1 - \pi$
- π je jediný parametr, $0 < \pi < 1$
- X – počet zdarů v jednom pokusu, kde pst zdaru je π
- $X \sim \mathrm{alt}(\pi)$
- $\mu_X = \mathrm{E}\,X = 1 \cdot \pi + 0 \cdot (1 - \pi) = \pi$
- $\sigma_X^2 = \mathrm{var}\,X = (1 - \pi)^2 \cdot \pi + (0 - \pi)^2 \cdot (1 - \pi) = \pi(1 - \pi)$

alternativní rozdělení

nula-jedničkové, Bernoulliovo

- pouze dvě možné hodnoty: 1 (zdar), 0 (nezdar)
- $\mathrm{P}(X = 1) = \pi, \mathrm{P}(X = 0) = 1 - \pi$
- π je jediný parametr, $0 < \pi < 1$
- X – počet zdarů v jednom pokusu, kde pst zdaru je π
- $X \sim \mathrm{alt}(\pi)$
- $\mu_X = \mathrm{E}\,X = 1 \cdot \pi + 0 \cdot (1 - \pi) = \pi$
- $\sigma_X^2 = \mathrm{var}\,X = (1 - \pi)^2 \cdot \pi + (0 - \pi)^2 \cdot (1 - \pi) = \pi(1 - \pi)$

alternativní rozdělení
nula-jedničkové, Bernoulliovo

- pouze dvě možné hodnoty: 1 (zdar), 0 (nezdar)
- $\mathrm{P}(X = 1) = \pi, \mathrm{P}(X = 0) = 1 - \pi$
- π je jediný parametr, $0 < \pi < 1$
- X – počet zdarů v jednom pokusu, kde pst zdaru je π
- $X \sim \mathrm{alt}(\pi)$
- $\mu_X = \mathrm{E}\,X = 1 \cdot \pi + 0 \cdot (1 - \pi) = \pi$
- $\sigma_X^2 = \mathrm{var}\,X = (1 - \pi)^2 \cdot \pi + (0 - \pi)^2 \cdot (1 - \pi) = \pi(1 - \pi)$

alternativní rozdělení

nula-jedničkové, Bernoulliovo

- pouze dvě možné hodnoty: 1 (zdar), 0 (nezdar)
- $\mathrm{P}(X = 1) = \pi, \mathrm{P}(X = 0) = 1 - \pi$
- π je jediný parametr, $0 < \pi < 1$
- X – počet zdarů v jednom pokusu, kde pst zdaru je π
- $X \sim \mathrm{alt}(\pi)$
- $\mu_X = \mathrm{E}\,X = 1 \cdot \pi + 0 \cdot (1 - \pi) = \pi$
- $\sigma_X^2 = \mathrm{var}\,X = (1 - \pi)^2 \cdot \pi + (0 - \pi)^2 \cdot (1 - \pi) = \pi(1 - \pi)$

alternativní rozdělení

nula-jedničkové, Bernoulliovo

- pouze dvě možné hodnoty: 1 (zdar), 0 (nezdar)
- $\mathrm{P}(X = 1) = \pi, \mathrm{P}(X = 0) = 1 - \pi$
- π je jediný parametr, $0 < \pi < 1$
- X – počet zdarů v jednom pokusu, kde pst zdaru je π
- $X \sim \mathrm{alt}(\pi)$
- $\mu_X = \mathrm{E}\,X = 1 \cdot \pi + 0 \cdot (1 - \pi) = \pi$
- $\sigma_X^2 = \mathrm{var}\,X = (1 - \pi)^2 \cdot \pi + (0 - \pi)^2 \cdot (1 - \pi) = \pi(1 - \pi)$

alternativní rozdělení

nula-jedničkové, Bernoulliovo

- pouze dvě možné hodnoty: 1 (zdar), 0 (nezdar)
- $\mathrm{P}(X = 1) = \pi, \mathrm{P}(X = 0) = 1 - \pi$
- π je jediný parametr, $0 < \pi < 1$
- X – počet zdarů v jednom pokusu, kde pst zdaru je π
- $X \sim \mathrm{alt}(\pi)$
- $\mu_X = \mathrm{E}\, X = 1 \cdot \pi + 0 \cdot (1 - \pi) = \pi$
- $\sigma_X^2 = \mathrm{var}\, X = (1 - \pi)^2 \cdot \pi + (0 - \pi)^2 \cdot (1 - \pi) = \pi(1 - \pi)$

alternativní rozdělení

nula-jedničkové, Bernoulliovo

- pouze dvě možné hodnoty: 1 (zdar), 0 (nezdar)
- $\mathrm{P}(X = 1) = \pi, \mathrm{P}(X = 0) = 1 - \pi$
- π je jediný parametr, $0 < \pi < 1$
- X – počet zdarů v jednom pokusu, kde pst zdaru je π
- $X \sim \mathrm{alt}(\pi)$
- $\mu_X = \mathrm{E}\, X = 1 \cdot \pi + 0 \cdot (1 - \pi) = \pi$
- $\sigma_X^2 = \mathrm{var}\, X = (1 - \pi)^2 \cdot \pi + (0 - \pi)^2 \cdot (1 - \pi) = \pi(1 - \pi)$

binomické rozdělení
[binomial distribution]

- $Y \sim \mathrm{bi}(n, \pi)$
- n **nezávislých** pokusů takových, že
- $\mathrm{P}(\text{zdar}) = \pi, \mathrm{P}(\text{nezdar}) = 1 - \pi, \ (0 < \pi < 1)$
- Y je **počet zdarů** v těchto pokusech
- $$\boxed{\mathrm{P}(Y = k) = \binom{n}{k} \pi^k (1-\pi)^{n-k}, \quad k = 0, 1, \ldots, n}$$
  [dbinom(k,n,pst)]
- např. ze 7 vajíček se vylíhne Y slepiček, $Y \sim \mathrm{bi}(7, 1/2)$
- např. při 60 hodech kostkou padlo Y šestek, $Y \sim \mathrm{bi}(60, 1/6)$
- předem nevíme, kolik bude slepiček (šestek), ale v dlouhodobém průměru je relativní četnost blízká 1/2 (1/6)

binomické rozdělení
[binomial distribution]

- $Y \sim \mathrm{bi}(n, \pi)$
- n **nezávislých** pokusů takových, že
- $\mathrm{P}(\text{zdar}) = \pi, \mathrm{P}(\text{nezdar}) = 1 - \pi$, $(0 < \pi < 1)$
- Y je **počet zdarů** v těchto pokusech
- $\boxed{\mathrm{P}(Y = k) = \binom{n}{k} \pi^k (1 - \pi)^{n-k}, \quad k = 0, 1, \ldots, n}$

  [dbinom(k,n,pst)]
- např. ze 7 vajíček se vylíhne Y slepiček, $Y \sim \mathrm{bi}(7, 1/2)$
- např. při 60 hodech kostkou padlo Y šestek, $Y \sim \mathrm{bi}(60, 1/6)$
- předem nevíme, kolik bude slepiček (šestek), ale v dlouhodobém průměru je relativní četnost blízká 1/2 (1/6)

binomické rozdělení

[binomial distribution]

- $Y \sim \text{bi}(n, \pi)$
- n **nezávislých** pokusů takových, že
- $\text{P}(\text{zdar}) = \pi, \text{P}(\text{nezdar}) = 1 - \pi$, $(0 < \pi < 1)$
- Y je **počet zdarů** v těchto pokusech
- $$\boxed{\text{P}(Y = k) = \binom{n}{k} \pi^k (1 - \pi)^{n-k}, \quad k = 0, 1, \ldots, n}$$

  [dbinom(k,n,pst)]
- např. ze 7 vajíček se vylíhne Y slepiček, $Y \sim \text{bi}(7, 1/2)$
- např. při 60 hodech kostkou padlo Y šestek, $Y \sim \text{bi}(60, 1/6)$
- předem nevíme, kolik bude slepiček (šestek), ale v dlouhodobém průměru je relativní četnost blízká 1/2 (1/6)

binomické rozdělení
[binomial distribution]

- $Y \sim \mathrm{bi}(n, \pi)$
- n **nezávislých** pokusů takových, že
- $\mathrm{P}(\text{zdar}) = \pi, \mathrm{P}(\text{nezdar}) = 1 - \pi$, $(0 < \pi < 1)$
- Y je **počet zdarů** v těchto pokusech
- $$\boxed{\mathrm{P}(Y = k) = \binom{n}{k} \pi^k (1-\pi)^{n-k}, \quad k = 0, 1, \ldots, n}$$
  [dbinom(k,n,pst)]
- např. ze 7 vajíček se vylíhne Y slepiček, $Y \sim \mathrm{bi}(7, 1/2)$
- např. při 60 hodech kostkou padlo Y šestek, $Y \sim \mathrm{bi}(60, 1/6)$
- předem nevíme, kolik bude slepiček (šestek), ale v dlouhodobém průměru je relativní četnost blízká 1/2 (1/6)

binomické rozdělení

[binomial distribution]

- $Y \sim \mathrm{bi}(n, \pi)$
- n **nezávislých** pokusů takových, že
- $\mathrm{P}(\mathrm{zdar}) = \pi, \mathrm{P}(\mathrm{nezdar}) = 1 - \pi,\ (0 < \pi < 1)$
- Y je **počet zdarů** v těchto pokusech
- $$\boxed{\mathrm{P}(Y = k) = \binom{n}{k} \pi^k (1 - \pi)^{n-k}, \quad k = 0, 1, \ldots, n}$$

  [dbinom(k,n,pst)]
- např. ze 7 vajíček se vylíhne Y slepiček, $Y \sim \mathrm{bi}(7, 1/2)$
- např. při 60 hodech kostkou padlo Y šestek, $Y \sim \mathrm{bi}(60, 1/6)$
- předem nevíme, kolik bude slepiček (šestek), ale v dlouhodobém průměru je relativní četnost blízká 1/2 (1/6)

binomické rozdělení

[binomial distribution]

- $Y \sim \mathrm{bi}(n, \pi)$
- n **nezávislých** pokusů takových, že
- $\mathrm{P}(\text{zdar}) = \pi, \mathrm{P}(\text{nezdar}) = 1 - \pi$, $(0 < \pi < 1)$
- Y je **počet zdarů** v těchto pokusech
- $$\boxed{\mathrm{P}(Y = k) = \binom{n}{k} \pi^k (1-\pi)^{n-k}, \quad k = 0, 1, \ldots, n}$$

  [dbinom(k,n,pst)]
- např. ze 7 vajíček se vylíhne Y slepiček, $Y \sim \mathrm{bi}(7, 1/2)$
- např. při 60 hodech kostkou padlo Y šestek, $Y \sim \mathrm{bi}(60, 1/6)$
- předem nevíme, kolik bude slepiček (šestek), ale v dlouhodobém průměru je relativní četnost blízká 1/2 (1/6)

binomické rozdělení

[binomial distribution]

- $Y \sim \mathrm{bi}(n, \pi)$
- n **nezávislých** pokusů takových, že
- $\mathrm{P}(\text{zdar}) = \pi, \mathrm{P}(\text{nezdar}) = 1 - \pi$, $(0 < \pi < 1)$
- Y je **počet zdarů** v těchto pokusech
- $\boxed{\mathrm{P}(Y = k) = \binom{n}{k} \pi^k (1 - \pi)^{n-k}, \quad k = 0, 1, \ldots, n}$
  [dbinom(k,n,pst)]
- např. ze 7 vajíček se vylíhne Y slepiček, $Y \sim \mathrm{bi}(7, 1/2)$
- např. při 60 hodech kostkou padlo Y šestek, $Y \sim \mathrm{bi}(60, 1/6)$
- předem nevíme, kolik bude slepiček (šestek), ale v dlouhodobém průměru je relativní četnost blízká 1/2 (1/6)

binomické rozdělení
[binomial distribution]

- $Y \sim \mathrm{bi}(n, \pi)$
- n **nezávislých** pokusů takových, že
- $\mathrm{P}(\text{zdar}) = \pi, \mathrm{P}(\text{nezdar}) = 1 - \pi$, $(0 < \pi < 1)$
- Y je **počet zdarů** v těchto pokusech
- $$\boxed{\mathrm{P}(Y = k) = \binom{n}{k} \pi^k (1 - \pi)^{n-k}, \quad k = 0, 1, \ldots, n}$$
  [dbinom(k,n,pst)]
- např. ze 7 vajíček se vylíhne Y slepiček, $Y \sim \mathrm{bi}(7, 1/2)$
- např. při 60 hodech kostkou padlo Y šestek, $Y \sim \mathrm{bi}(60, 1/6)$
- předem nevíme, kolik bude slepiček (šestek), ale v dlouhodobém průměru je relativní četnost blízká 1/2 (1/6)

binomické rozdělení pomocí alternativního

- $Y \sim \text{bi}(n, \pi)$
- Y je celkový počet zdarů v n pokusech, tedy
- $Y = X_1 + X_2 + \ldots + X_n = \sum_{i=1}^{n} X_i$,
  kde X_i je počet zdarů v i-tém pokusu
- z vlastností střední hodnoty (očekávaný počet zdarů)

$$\mu_Y = \mathrm{E}\, Y = \mathrm{E} \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} \mathrm{E}\, X_i = \sum_{i=1}^{n} \pi = n\pi$$

- protože jsou pokusy nezávislé

$$\sigma_Y^2 = \operatorname{var} \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} \operatorname{var} X_i = \sum_{i=1}^{n} \pi(1-\pi) = n\pi(1-\pi)$$

binomické rozdělení pomocí alternativního

- $Y \sim \mathrm{bi}(n, \pi)$
- Y je celkový počet zdarů v n pokusech, tedy
- $Y = X_1 + X_2 + \ldots + X_n = \sum_{i=1}^{n} X_i$,
  kde X_i je počet zdarů v i-tém pokusu
- z vlastností střední hodnoty (očekávaný počet zdarů)

$$\mu_Y = \mathrm{E}\, Y = \mathrm{E} \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} \mathrm{E}\, X_i = \sum_{i=1}^{n} \pi = n\pi$$

- protože jsou pokusy nezávislé

$$\sigma_Y^2 = \mathrm{var} \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} \mathrm{var}\, X_i = \sum_{i=1}^{n} \pi(1-\pi) = n\pi(1-\pi)$$

binomické rozdělení pomocí alternativního

- $Y \sim \mathrm{bi}(n, \pi)$
- Y je celkový počet zdarů v n pokusech, tedy
- $Y = X_1 + X_2 + \ldots + X_n = \sum_{i=1}^{n} X_i$,
  kde X_i je počet zdarů v i-tém pokusu
- z vlastností střední hodnoty (očekávaný počet zdarů)

$$\mu_Y = \mathrm{E}\, Y = \mathrm{E} \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} \mathrm{E}\, X_i = \sum_{i=1}^{n} \pi = n\pi$$

- protože jsou pokusy nezávislé

$$\sigma_Y^2 = \mathrm{var} \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} \mathrm{var}\, X_i = \sum_{i=1}^{n} \pi(1-\pi) = n\pi(1-\pi)$$

binomické rozdělení pomocí alternativního

- $Y \sim \text{bi}(n, \pi)$
- Y je celkový počet zdarů v n pokusech, tedy
- $Y = X_1 + X_2 + \ldots + X_n = \sum_{i=1}^{n} X_i$,
  kde X_i je počet zdarů v i-tém pokusu
- z vlastností střední hodnoty (očekávaný počet zdarů)

$$\mu_Y = \mathrm{E}\, Y = \mathrm{E} \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} \mathrm{E}\, X_i = \sum_{i=1}^{n} \pi = n\pi$$

- protože jsou pokusy nezávislé

$$\sigma_Y^2 = \mathrm{var} \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} \mathrm{var}\, X_i = \sum_{i=1}^{n} \pi(1-\pi) = n\pi(1-\pi)$$

binomické rozdělení pomocí alternativního

- $Y \sim \mathrm{bi}(n, \pi)$
- Y je celkový počet zdarů v n pokusech, tedy
- $Y = X_1 + X_2 + \ldots + X_n = \sum_{i=1}^{n} X_i$,
  kde X_i je počet zdarů v i-tém pokusu
- z vlastností střední hodnoty (očekávaný počet zdarů)

$$\mu_Y = \mathrm{E}\, Y = \mathrm{E} \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} \mathrm{E}\, X_i = \sum_{i=1}^{n} \pi = n\pi$$

- protože jsou pokusy nezávislé

$$\sigma_Y^2 = \mathrm{var} \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} \mathrm{var}\, X_i = \sum_{i=1}^{n} \pi(1-\pi) = n\pi(1-\pi)$$

příklad: kuřáci

- mezi dvacetiletými muži je (řekněme) 35 % kuřáků ($\pi = 0{,}35$)
- je-li dvacetiletých 70 tisíc ($m = 70\,000$), pak je kuřáků asi $m\pi = 70\,000 \cdot 0{,}35 = 24\,500$, ale nevíme, kteří to jsou
- vyberme náhodně $n = 60$ dvacetiletých mužů, označme jako Y počet kuřáků mezi nimi, je tedy $Y \sim \mathrm{bi}(60, 0{,}35)$
- střední hodnota (očekávaný počet), rozptyl

$$\mu_Y = 60 \cdot 0{,}35 = 21 \qquad \sigma_Y^2 = 60 \cdot 0{,}35 \cdot 0{,}65 = 13{,}65 \doteq (3{,}7)^2$$

- ukázky pravděpodobností možných hodnot

k	15	17	19	21	23	25
$\mathrm{P}(Y = k)$	0,029	0,062	0,095	0,107	0,091	0,059

- psti počítány pomocí [dbinom(0:60,60,0.35)]

příklad: kuřáci

- mezi dvacetiletými muži je (řekněme) 35 % kuřáků ($\pi = 0{,}35$)
- je-li dvacetiletých 70 tisíc ($m = 70\,000$), pak je kuřáků asi $m\pi = 70\,000 \cdot 0{,}35 = 24\,500$, ale nevíme, kteří to jsou
- vyberme náhodně $n = 60$ dvacetiletých mužů, označme jako Y počet kuřáků mezi nimi, je tedy $Y \sim \mathrm{bi}(60, 0{,}35)$
- střední hodnota (očekávaný počet), rozptyl

$$\mu_Y = 60 \cdot 0{,}35 = 21 \qquad \sigma_Y^2 = 60 \cdot 0{,}35 \cdot 0{,}65 = 13{,}65 \doteq (3{,}7)^2$$

- ukázky pravděpodobností možných hodnot

k	15	17	19	21	23	25
$P(Y = k)$	0,029	0,062	0,095	0,107	0,091	0,059

- psti počítány pomocí [dbinom(0:60,60,0.35)]

příklad: kuřáci

- mezi dvacetiletými muži je (řekněme) 35 % kuřáků ($\pi = 0{,}35$)
- je-li dvacetiletých 70 tisíc ($m = 70\,000$), pak je kuřáků asi $m\pi = 70\,000 \cdot 0{,}35 \; = 24\,500$, ale nevíme, kteří to jsou
- vyberme náhodně $n = 60$ dvacetiletých mužů, označme jako Y počet kuřáků mezi nimi, je tedy $Y \sim \mathrm{bi}(60, 0{,}35)$
- střední hodnota (očekávaný počet), rozptyl

$$\mu_Y = 60 \cdot 0{,}35 = 21 \qquad \sigma_Y^2 = 60 \cdot 0{,}35 \cdot 0{,}65 = 13{,}65 \doteq (3{,}7)^2$$

- ukázky pravděpodobností možných hodnot

k	15	17	19	21	23	25
$P(Y = k)$	0,029	0,062	0,095	0,107	0,091	0,059

- psti počítány pomocí [dbinom(0:60,60,0.35)]

příklad: kuřáci

- mezi dvacetiletými muži je (řekněme) 35 % kuřáků ($\pi = 0{,}35$)
- je-li dvacetiletých 70 tisíc ($m = 70\,000$), pak je kuřáků asi $m\pi = 70\,000 \cdot 0{,}35 = 24\,500$, ale nevíme, kteří to jsou
- vyberme náhodně $n = 60$ dvacetiletých mužů, označme jako Y počet kuřáků mezi nimi, je tedy $Y \sim \mathrm{bi}(60, 0{,}35)$
- střední hodnota (očekávaný počet), rozptyl

$$\mu_Y = 60 \cdot 0{,}35 = 21 \qquad \sigma_Y^2 = 60 \cdot 0{,}35 \cdot 0{,}65 = 13{,}65 \doteq (3{,}7)^2$$

- ukázky pravděpodobností možných hodnot

k	15	17	19	21	23	25
$\mathrm{P}(Y = k)$	0,029	0,062	0,095	0,107	0,091	0,059

- psti počítány pomocí [dbinom(0:60,60,0.35)]

příklad: kuřáci

- mezi dvacetiletými muži je (řekněme) 35 % kuřáků ($\pi = 0{,}35$)
- je-li dvacetiletých 70 tisíc ($m = 70\,000$), pak je kuřáků asi $m\pi = 70\,000 \cdot 0{,}35 = 24\,500$, ale nevíme, kteří to jsou
- vyberme náhodně $n = 60$ dvacetiletých mužů, označme jako Y počet kuřáků mezi nimi, je tedy $Y \sim \mathrm{bi}(60, 0{,}35)$
- střední hodnota (očekávaný počet), rozptyl

$$\mu_Y = 60 \cdot 0{,}35 = 21 \qquad \sigma_Y^2 = 60 \cdot 0{,}35 \cdot 0{,}65 = 13{,}65 \doteq (3{,}7)^2$$

- ukázky pravděpodobností možných hodnot

k	15	17	19	21	23	25
$P(Y = k)$	0,029	0,062	0,095	0,107	0,091	0,059

- psti počítány pomocí [dbinom(0:60,60,0.35)]

příklad: kuřáci

- mezi dvacetiletými muži je (řekněme) 35 % kuřáků ($\pi = 0{,}35$)
- je-li dvacetiletých 70 tisíc ($m = 70\,000$), pak je kuřáků asi $m\pi = 70\,000 \cdot 0{,}35 \ = 24\,500$, ale nevíme, kteří to jsou
- vyberme náhodně $n = 60$ dvacetiletých mužů, označme jako Y počet kuřáků mezi nimi, je tedy $Y \sim \mathrm{bi}(60, 0{,}35)$
- střední hodnota (očekávaný počet), rozptyl

$$\mu_Y = 60 \cdot 0{,}35 = 21 \qquad \sigma_Y^2 = 60 \cdot 0{,}35 \cdot 0{,}65 = 13{,}65 \doteq (3{,}7)^2$$

- ukázky pravděpodobností možných hodnot

k	15	17	19	21	23	25
$\mathrm{P}(Y = k)$	0,029	0,062	0,095	0,107	0,091	0,059

- psti počítány pomocí [dbinom(0:60,60,0.35)]

Poissonovo rozdělení

[Poisson distribution]

- $X \sim \text{Po}(\lambda) \qquad (\lambda > 0)$
- zákon vzácných (řídkých) jevů
- kolikrát nastal jev během jednotkového časového intervalu, na jednotkové ploše, v jednotkovém objemu ...
- $\boxed{P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}}, \qquad k = 0, 1, \dots$
- $\mu_X = \lambda$, $\sigma_X^2 = \lambda$
- pro velké n a malé π lze rozdělení bi(n, π) aproximovat pomocí rozdělení Po$(n\pi)$
- např. počet kolonií na Petriho misce

Poissonovo rozdělení

[Poisson distribution]

- $X \sim \mathrm{Po}(\lambda) \qquad (\lambda > 0)$
- zákon vzácných (řídkých) jevů
- kolikrát nastal jev během jednotkového časového intervalu, na jednotkové ploše, v jednotkovém objemu ...
- $\boxed{\mathrm{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}}, \qquad k = 0, 1, \ldots$
- $\mu_X = \lambda$, $\sigma_X^2 = \lambda$
- pro velké n a malé π lze rozdělení bi(n, π) aproximovat pomocí rozdělení Po$(n\pi)$
- např. počet kolonií na Petriho misce

Poissonovo rozdělení
[Poisson distribution]

- $X \sim \text{Po}(\lambda) \qquad (\lambda > 0)$
- zákon vzácných (řídkých) jevů
- kolikrát nastal jev během jednotkového časového intervalu, na jednotkové ploše, v jednotkovém objemu ...
- $\boxed{\text{P}(X = k) = \frac{\lambda^k}{k!}\text{e}^{-\lambda}}, \qquad k = 0, 1, \ldots$
- $\mu_X = \lambda$, $\sigma_X^2 = \lambda$
- pro velké n a malé π lze rozdělení bi(n, π) aproximovat pomocí rozdělení Po$(n\pi)$
- např. počet kolonií na Petriho misce

Poissonovo rozdělení

[Poisson distribution]

- $X \sim \mathrm{Po}(\lambda) \qquad (\lambda > 0)$
- zákon vzácných (řídkých) jevů
- kolikrát nastal jev během jednotkového časového intervalu, na jednotkové ploše, v jednotkovém objemu ...
- $\boxed{\mathrm{P}(X = k) = \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda}}, \qquad k = 0, 1, \ldots$
- $\mu_X = \lambda$, $\sigma_X^2 = \lambda$
- pro velké n a malé π lze rozdělení $\mathrm{bi}(n, \pi)$ aproximovat pomocí rozdělení $\mathrm{Po}(n\pi)$
- např. počet kolonií na Petriho misce

Poissonovo rozdělení

[Poisson distribution]

- $X \sim \mathrm{Po}(\lambda) \qquad (\lambda > 0)$
- zákon vzácných (řídkých) jevů
- kolikrát nastal jev během jednotkového časového intervalu, na jednotkové ploše, v jednotkovém objemu ...
- $\boxed{\mathrm{P}(X = k) = \dfrac{\lambda^k}{k!}\mathrm{e}^{-\lambda}}, \qquad k = 0, 1, \ldots$
- $\mu_X = \lambda$, $\sigma_X^2 = \lambda$
- pro velké n a malé π lze rozdělení $\mathrm{bi}(n, \pi)$ aproximovat pomocí rozdělení $\mathrm{Po}(n\pi)$
- např. počet kolonií na Petriho misce

Poissonovo rozdělení

[Poisson distribution]

- $X \sim \mathrm{Po}(\lambda) \qquad (\lambda > 0)$
- zákon vzácných (řídkých) jevů
- kolikrát nastal jev během jednotkového časového intervalu, na jednotkové ploše, v jednotkovém objemu ...
- $\boxed{\mathrm{P}(X = k) = \dfrac{\lambda^k}{k!}\mathrm{e}^{-\lambda}}, \qquad k = 0, 1, \ldots$
- $\mu_X = \lambda$, $\sigma_X^2 = \lambda$
- pro velké n a malé π lze rozdělení $\mathrm{bi}(n, \pi)$ aproximovat pomocí rozdělení $\mathrm{Po}(n\pi)$
- např. počet kolonií na Petriho misce

Poissonovo rozdělení
[Poisson distribution]

- $X \sim \mathrm{Po}(\lambda) \qquad (\lambda > 0)$
- zákon vzácných (řídkých) jevů
- kolikrát nastal jev během jednotkového časového intervalu, na jednotkové ploše, v jednotkovém objemu ...
- $\boxed{\mathrm{P}(X = k) = \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}}, \qquad k = 0, 1, \dots$
- $\mu_X = \lambda$, $\sigma_X^2 = \lambda$
- pro velké n a malé π lze rozdělení $\mathrm{bi}(n, \pi)$ aproximovat pomocí rozdělení $\mathrm{Po}(n\pi)$
- např. počet kolonií na Petriho misce

příklad

s jakou pstí udělá 5 z 55 stejně připravených studentů zkoušku na výbornou, je-li pst jedničky 0,1?

- binomické rozdělení $Y \sim \mathrm{bi}(55, 0{,}1)$ [dbinom(5,55,0.1)]

$$\mathrm{P}(Y=5) = \binom{55}{5} \cdot 0{,}1^5 \cdot 0{,}9^{50} = 0{,}179$$

- aproximace Poissonovým rozdělením $Y \sim \mathrm{Po}(55 \cdot 0{,}1) = \mathrm{Po}(5{,}5)$ [dpois(5, 5.5)]

$$\mathrm{P}(Y=5) = \frac{5{,}5^5}{5!} e^{-5{,}5} = 0{,}171$$

příklad

s jakou pstí udělá 5 z 55 stejně připravených studentů zkoušku na výbornou, je-li pst jedničky 0,1?

- binomické rozdělení $Y \sim \mathrm{bi}(55, 0{,}1)$ [dbinom(5,55,0.1)]

$$P(Y = 5) = \binom{55}{5} \cdot 0{,}1^5 \cdot 0{,}9^{50} = 0{,}179$$

- aproximace Poissonovým rozdělením $Y \sim \mathrm{Po}(55 \cdot 0{,}1) = \mathrm{Po}(5{,}5)$ [dpois(5, 5.5)]

$$P(Y = 5) = \frac{5{,}5^5}{5!} e^{-5{,}5} = 0{,}171$$

normální (Gaussovo) rozdělení $\mathrm{N}(\mu, \sigma^2)$

[normal (Gaussian) distribution]

- $\boxed{\mu_X = \mu,\ \sigma_X^2 = \sigma^2}$
- spojité rozdělení, symetrické okolo střední hodnoty μ
- maximální hodnota hustoty přibližně $0{,}4/\sigma$
- $\mathrm{N}(0,1)$ (normované normální rozdělení):
  $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ (hustota),
  $\Phi(x) = \int_{-\infty}^{x} \varphi(t)\mathrm{d}t$ (distr. fce)
- $X \sim \mathrm{N}(\mu, \sigma^2)$, pak $Z = \frac{X-\mu}{\sigma} \sim \mathrm{N}(0,1)$

$$\boxed{\mathrm{P}(a < X < b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}$$

- model vzniku: součet velkého počtu nepatrných příspěvků
- velmi často modeluje znaky v poměrovém měřítku

normální (Gaussovo) rozdělení $\mathrm{N}(\mu, \sigma^2)$

[normal (Gaussian) distribution]

- $\boxed{\mu_X = \mu,\ \sigma_X^2 = \sigma^2}$
- spojité rozdělení, symetrické okolo střední hodnoty μ
- maximální hodnota hustoty přibližně $0{,}4/\sigma$
- $\mathrm{N}(0,1)$ (normované normální rozdělení):
  $\varphi(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}$ (hustota),
  $\Phi(x) = \int_{-\infty}^{x} \varphi(t)\mathrm{d}t$ (distr. fce)
- $X \sim \mathrm{N}(\mu, \sigma^2)$, pak $Z = \frac{X-\mu}{\sigma} \sim \mathrm{N}(0,1)$

$$\boxed{\mathrm{P}(a < X < b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}$$

- model vzniku: součet velkého počtu nepatrných příspěvků
- velmi často modeluje znaky v poměrovém měřítku

normální (Gaussovo) rozdělení $\mathrm{N}(\mu, \sigma^2)$

[normal (Gaussian) distribution]

- $\boxed{\mu_X = \mu,\ \sigma_X^2 = \sigma^2}$
- spojité rozdělení, symetrické okolo střední hodnoty μ
- maximální hodnota hustoty přibližně $0{,}4/\sigma$
- $\mathrm{N}(0,1)$ (normované normální rozdělení):
  $\varphi(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}$ (hustota),
  $\Phi(x) = \int_{-\infty}^{x} \varphi(t)\mathrm{d}t$ (distr. fce)
- $X \sim \mathrm{N}(\mu, \sigma^2)$, pak $Z = \frac{X-\mu}{\sigma} \sim \mathrm{N}(0,1)$

$$\boxed{\mathrm{P}(a < X < b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}$$

- model vzniku: součet velkého počtu nepatrných příspěvků
- velmi často modeluje znaky v poměrovém měřítku

normální (Gaussovo) rozdělení $\mathrm{N}(\mu, \sigma^2)$

[normal (Gaussian) distribution]

- $\boxed{\mu_X = \mu,\ \sigma_X^2 = \sigma^2}$
- spojité rozdělení, symetrické okolo střední hodnoty μ
- maximální hodnota hustoty přibližně $0{,}4/\sigma$
- $\mathrm{N}(0,1)$ (normované normální rozdělení):
  $\varphi(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}$ (hustota),
  $\Phi(x) = \int_{-\infty}^{x} \varphi(t)\mathrm{d}t$ (distr. fce)
- $X \sim \mathrm{N}(\mu, \sigma^2)$, pak $Z = \frac{X-\mu}{\sigma} \sim \mathrm{N}(0,1)$

$$\boxed{\mathrm{P}(a < X < b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}$$

- model vzniku: součet velkého počtu nepatrných příspěvků
- velmi často modeluje znaky v poměrovém měřítku

normální (Gaussovo) rozdělení $\mathrm{N}(\mu, \sigma^2)$

[normal (Gaussian) distribution]

- $\boxed{\mu_X = \mu,\ \sigma_X^2 = \sigma^2}$
- spojité rozdělení, symetrické okolo střední hodnoty μ
- maximální hodnota hustoty přibližně $0{,}4/\sigma$
- $\mathrm{N}(0,1)$ (normované normální rozdělení):
  $\varphi(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}$ (hustota),
  $\Phi(x) = \int_{-\infty}^{x} \varphi(t)\mathrm{d}t$ (distr. fce)
- $X \sim \mathrm{N}(\mu, \sigma^2)$, pak $Z = \frac{X-\mu}{\sigma} \sim \mathrm{N}(0,1)$

$$\boxed{\mathrm{P}(a < X < b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}$$

- model vzniku: součet velkého počtu nepatrných příspěvků
- velmi často modeluje znaky v poměrovém měřítku

normální (Gaussovo) rozdělení $\mathrm{N}(\mu, \sigma^2)$

[normal (Gaussian) distribution]

- $\boxed{\mu_X = \mu,\ \sigma_X^2 = \sigma^2}$
- spojité rozdělení, symetrické okolo střední hodnoty μ
- maximální hodnota hustoty přibližně $0{,}4/\sigma$
- $\mathrm{N}(0, 1)$ (normované normální rozdělení):
  $\varphi(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}$ (hustota),
  $\Phi(x) = \int_{-\infty}^{x} \varphi(t)\mathrm{d}t$ (distr. fce)
- $X \sim \mathrm{N}(\mu, \sigma^2)$, pak $Z = \frac{X-\mu}{\sigma} \sim \mathrm{N}(0, 1)$

$$\boxed{\mathrm{P}(a < X < b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}$$

- model vzniku: součet velkého počtu nepatrných příspěvků
- velmi často modeluje znaky v poměrovém měřítku

normální (Gaussovo) rozdělení $\mathrm{N}\left(\mu, \sigma^2\right)$

[normal (Gaussian) distribution]

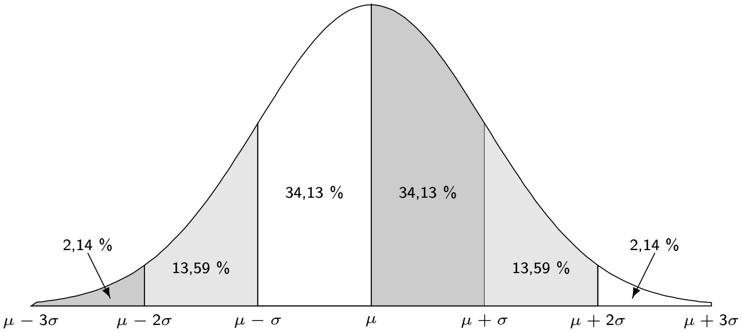
- $\boxed{\mu_X = \mu,\ \sigma_X^2 = \sigma^2}$
- spojité rozdělení, symetrické okolo střední hodnoty μ
- maximální hodnota hustoty přibližně 0,4$/\sigma$
- $\mathrm{N}(0,1)$ (normované normální rozdělení):
  $\varphi(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}$ (hustota),
  $\Phi(x) = \int_{-\infty}^{x} \varphi(t)\mathrm{d}t$ (distr. fce)
- $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$, pak $Z = \frac{X-\mu}{\sigma} \sim \mathrm{N}(0,1)$

$$\boxed{\mathrm{P}(a < X < b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}$$

- model vzniku: součet velkého počtu nepatrných příspěvků
- velmi často modeluje znaky v poměrovém měřítku

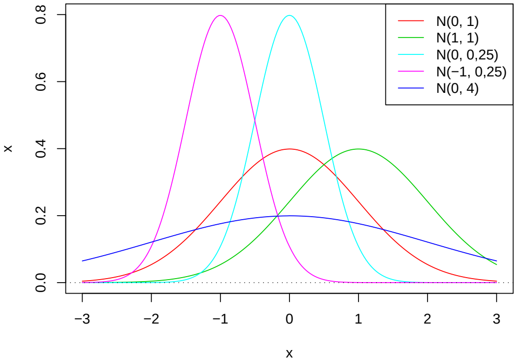
hustota $\mathrm{N}(\mu, \sigma^2)$

[dnorm(x,mu,sigma)]



normální (Gaussovo) rozdělení $N(\mu, \sigma^2)$

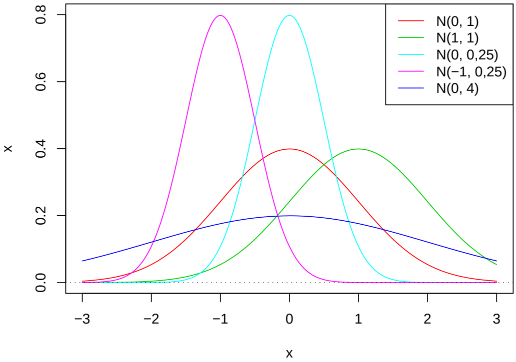
význam parametrů



- spojité rozdělení, symetrické okolo střední hodnoty μ
- maximální hodnota hustoty přibližně $0{,}4/\sigma$
- model vzniku: součet velkého počtu nepatrných příspěvků

normální (Gaussovo) rozdělení N(μ, σ^2)

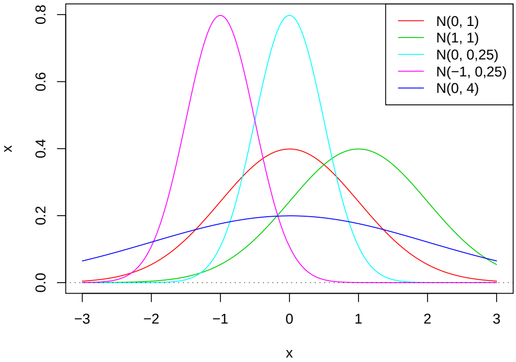
význam parametrů



- spojité rozdělení, symetrické okolo střední hodnoty μ
- maximální hodnota hustoty přibližně 0,4$/\sigma$
- model vzniku: součet velkého počtu nepatrných příspěvků

normální (Gaussovo) rozdělení N(μ, σ^2)

význam parametrů



- spojité rozdělení, symetrické okolo střední hodnoty μ
- maximální hodnota hustoty přibližně 0,4$/\sigma$
- model vzniku: součet velkého počtu nepatrných příspěvků

výpočet pravděpodobnosti, že $a < X < b$

použije distribuční funkci $\mathrm{N}(0,1)$

$\mathrm{P}(a < X < b) = F_X(b) - F_X(a)$ platí obecně pro spoj. rozděl.

$$X \sim \mathrm{N}(\mu, \sigma^2) \Rightarrow Z = \frac{X-\mu}{\sigma} \sim \mathrm{N}(0,1)$$

$$\mathrm{P}(X \leq x) = \mathrm{P}\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = \mathrm{P}\left(Z \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

$$\boxed{\mathrm{P}(a < X < b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}$$

[pnorm((b-mu)/sigma)-pnorm((a-mu)/sigma)]

v programu R je distribuční funkce $\mathrm{N}(\mu, \sigma^2)$ s obecnými parametry:

[pnorm(b,mu,sigma)-pnorm(a,mu,sigma)]

příklad

- u jakého dílu populace desetiletých hochů naměříme výšku od 135 do 140 cm, když pro výšku desetiletých platí $X \sim \mathrm{N}\left(136{,}1, 6{,}4^2\right)$
- předpokládáme zaokrouhlování na celá čísla při měření:

$$\mathrm{P}(134{,}5 < X < 140{,}5) = \Phi\left(\frac{140{,}5 - 136{,}1}{6{,}4}\right) - \Phi\left(\frac{134{,}5 - 136{,}1}{6{,}4}\right)$$
$$= 0{,}754 - 0{,}401 = 0{,}353$$

[pnorm((140.5-136.1)/6.4)-pnorm((134.5-136.1)/6.4)]

- pomocí distribuční fce s obecnými parametry [pnorm(140.5,136.1,6.4)-pnorm(134.5,136.1,6.4)]

příklad

- u jakého dílu populace desetiletých hochů naměříme výšku od 135 do 140 cm, když pro výšku desetiletých platí $X \sim \mathrm{N}\big(136{,}1, 6{,}4^2\big)$
- předpokládáme zaokrouhlování na celá čísla při měření:

$$\mathrm{P}(134{,}5 < X < 140{,}5) = \Phi\left(\frac{140{,}5 - 136{,}1}{6{,}4}\right) - \Phi\left(\frac{134{,}5 - 136{,}1}{6{,}4}\right) = 0{,}754 - 0{,}401 = 0{,}353$$

[pnorm((140.5-136.1)/6.4)-pnorm((134.5-136.1)/6.4)]

- pomocí distribuční fce s obecnými parametry [pnorm(140.5,136.1,6.4)-pnorm(134.5,136.1,6.4)]

příklad

- u jakého dílu populace desetiletých hochů naměříme výšku od 135 do 140 cm, když pro výšku desetiletých platí $X \sim \mathrm{N}\big(136{,}1, 6{,}4^2\big)$
- předpokládáme zaokrouhlování na celá čísla při měření:

$$\mathrm{P}(134{,}5 < X < 140{,}5) = \Phi\left(\frac{140{,}5 - 136{,}1}{6{,}4}\right) - \Phi\left(\frac{134{,}5 - 136{,}1}{6{,}4}\right)$$
$$= 0{,}754 - 0{,}401 = 0{,}353$$

[pnorm((140.5-136.1)/6.4)-pnorm((134.5-136.1)/6.4)]

- pomocí distribuční fce s obecnými parametry [pnorm(140.5,136.1,6.4)-pnorm(134.5,136.1,6.4)]