6 Výpočet číselných charakteristik náhodných veličin, aplikace Moivreovy – Laplaceovy věty

- $F(x)$, $p(x)$, $f(x)$ ... funkcionální charakteristiky
  - obsahují veškerou informaci o chování náh.veličiny
- někdy nás zajímají pouze rysy chování náh.veličiny $\rightarrow$ *číselné charakteristiky*
  - kvantily ($x_{0.25}$, $x_{0.5}$, $x_{0.75}$, apod.)
  - střední hodnota
  - rozptyl / směrodatná odchylka
  - kovariance
  - korelace

Kvantily vybraných spojitých rozdělení; α-kvantil

- α-kvantil náh.veličiny X ... x_α
- obdoba α-kvantilu v popisné statistice
- křivka hustoty:
  - plocha pod křivkou ... pst ... $= 1$
  - tuto plochu rozdělíme na 2 části
    * tmavá plocha α
    * světlá plocha $1 - \alpha$

- α-kvantil ... číslo, takové, že $\Pr(X \leq x_\alpha) = \alpha$
- pst, že náhodná veličina X je menší nebo rovna x_α je rovna α
- speciální kvantily
  - medián ... $x_{0.5}$
  - 1.kvartil ... $x_{0.25}$
  - 3.kvartil ... $x_{0.75}$
- **Standardizované normální rozdělení**
  -
    * $X \sim N(0,1)$
    * α-kvantil ... $u(\alpha)$
    * symetrické ... $u(\alpha) = -u(1-\alpha)$

  * qnorm(alpha)

- χ^2 **rozdělení** s n stupni volnosti
  - (Pearsonovo rozdělení)
  - $X \sim \chi^2(n)$
  - α-kvantil ... $\chi^2_n(\alpha)$
  - nesymetrické
  - qchisq(alpha,n)
- **Studentovo rozdělení** s n stupni volnosti
  - $X \sim t(n)$
  - α-kvantil ... $t_n(\alpha)$
  - symetrické ... $t_n(\alpha) = -t_n(1-\alpha)$
  - qt(alpha,n)
- **Fisherovo rozdělení** s n_1 a n_2 stupni volnosti
  - (Fisherovo-Snedecorovo rozdělení)
  - $X \sim F(n_1, n_2)$
  - α-kvantil ... $F_{n_1,n_2}(\alpha)$
  - nesymetrické, ale $F_{n_1,n_2}(\alpha) = \dfrac{1}{F_{n_1,n_2}(1-\alpha)}$
  - qf(alpha, n1, n2)

Příklad 6.1. Najděte medián a horní a dolní kvartil náhodné veličiny $U \sim N(0,1)$.

```
qnorm(0.5)
qnorm(0.25)
qnorm(0.75)
```

Příklad 6.2. Najděte dolní kvartil náhodné veličiny $X \sim N(3,5)$.

```
qnorm(0.25, 3, sqrt(5))
```

Příklad 6.3. Určete kvantil $\chi^2_{25}(0.025)$.

```
qchisq(0.025, 25)
```

Příklad 6.4. Určete kvantily $t_{30}(0.99)$ a $t_{14}(0.05)$.

```
qt(0.99, 30)
qt(0.05, 14)
```

Příklad 6.5. Určete kvantily $F_{5,20}(0.975)$ a $F_{2,10}(0.05)$.

```
qf(0.975, 5,20)
qf(0.05,  2,10)
```

Střední hodnota EX, μ

- 'idealizovaný' průměr
- diskrétní náhodná veličina $\rightarrow$

$$EX = \sum_{-\infty}^{\infty} xp(x)$$

- diskrétní transformovaná náhodná veličina $\rightarrow$

$$E(Y) = E(g(X)) = \sum_{\infty}^{\infty} g(x)p(x)$$

- náh. veličina X nemusí mít EX
- Vlastnosti
  1. $E(a) = a$
  2. $E(a + bX) = a + bE(X)$
  3. $E(X_1 + X_2) = EX_1 + EX_2$

Rozptyl $D(X)$, σ^2

- jakou má náh. vel. X tendenci realizovat se poblíž/daleko centrální polohy
-

$$DX = E([X - EX]^2)$$

- diskrétní náh. veličina

$$DX = E([X - EX]^2) = \sum_{-\infty}^{\infty} [x - EX]^2 p(x)$$

- čím více se hodnoty v souboru navzájem liší, tím je hodnota rozptylu vyšší
- $\sqrt{DX}$ ... směrodatná odchylka
- Vlastnosti
  1. $D(X) = E(X^2) - (EX)^2$
  2. $D(a) = 0$
  3. $D(a + bX) = b^2 D(X)$
  4. X_1, X_2 stoch.nezáv $\rightarrow D(X_1 + X_2) = D(X_1) + D(X_2)$

Příklad 6.6. Náhodná veličina X udává počet ok při hodu kostkou. Vypočtěte její střední hodnotu a rozptyl.

- X ... diskrétní n.v. $\rightarrow E(X) = \sum_{-\infty}^{\infty} xp(x) \rightarrow$ potřebujeme $p(x)$ pro $x = 1, \dots x = 6$

- $P(X=1)=1/6$, $P(X=2)=1/6$, $P(X=3)=1/6$, $P(X=4)=1/6$, $P(X=5)=1/6$, $P(X=6)=1/6$

$$E(X)=1\frac{1}{6}+2\frac{1}{6}+3\frac{1}{6}+4\frac{1}{6}+5\frac{1}{6}+6\frac{1}{6}=3.5$$

- $DX=\sum_{-\infty}^{\infty}[X-EX]^2p(x)$
  -
  $$\begin{aligned}DX&=(1-3.5)^2\frac{1}{6}+(2-3.5)^2\frac{1}{6}+(3-3.5)^2\frac{1}{6}+(4-3.5)^2\frac{1}{6}+(5-3.5)^2\frac{1}{6}+(6-3.5)^2\frac{1}{6}\\&=\frac{6.25}{6}+\frac{2.25}{6}+\frac{0.25}{6}+\frac{0.25}{6}+\frac{2.25}{6}+\frac{6.25}{6}=2.917\end{aligned}$$

- $DX=E(X^2)-(EX)^2$
  - EX máme; $E(X^2)$ vypočteme podle: $E(Y)=E(g(X))=\sum_{\infty}^{\infty}g(x)p(x)$, kde $Y=g(X)=X^2$

$$\begin{aligned}E(X^2)&=1^2\frac{1}{6}+2^2\frac{1}{6}+3^2\frac{1}{6}+4^2\frac{1}{6}+5^2\frac{1}{6}+6^2\frac{1}{6}\\&=15.1\end{aligned}$$

  - $DX=E(X^2)-(EX)^2=15.1-3.5^2=2.917.$

```
x <- 1:6
pi <- rep(1/6, 6)
EX <- sum(x*pi)
DX <- sum(x^2*pi)-EX^2
```

standardizovaná náh.veličina

-
$$\frac{X-E(X)}{\sqrt{D(X)}}$$

- centrování, škálování, standardizace

Kovariance $C(X,Y)$

- střední hodnota součinu centrovaných náh.veličin X,Y
- diskrétní náhodná veličina:
  $C(X,Y)=E([X-EX][Y-EY])=$

$$\sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty}[x-EX][y-EY]p(x,y)$$

  kde $p(x,y)$ je simultánní pstní fce.

- znaménko kovariance: $+$ ... *přímý*; $-$ ... *nepřímý* vztah

Korelace $R(X,Y)$

- střední hodnota součinu standardizovaných veličin X,Y
- $$R(X,Y) = E\left(\frac{X-EX}{\sqrt{DX}}\frac{Y-EY}{\sqrt{DY}}\right) = \frac{C(X,Y)}{\sqrt{DX}\sqrt{DY}}$$
- charakterizuje těsnost LINEÁRNÍHO vztahu mezi X a Y

Objektem zájmu rozsáhlé studie bylo sledování pohřebního rituálu dnes již vymřelého ale v minulosti velmi dlouho přetrvávajícího a rozsáhlého jihoamerického kmene. Součástí pohřebního rituálu tohoto kmene bylo odsekávání článků prstů na rukou a nohou zemřelého a jejich následné obětování bohům jako dar, aby zemřelého přijali mezi sebe. Zemřelému tak byl na ruce odetnut buď jeden nebo dva prsty a na noze tři nebo čtyři prsty.

Dále bylo zjištěno, že domorodci odtínali jeden prst na ruce a tři prsty na noze zemřelého s pravděpodobností 0.1, dva prsty na ruce a tři prsty na noze s pravděpodobností 0.3, jeden prst na ruce a čtyři prsty na noze s pravděpodobností 0.35 a dva prsty na ruce a čtyři prsty na noze s pravděpodobností 0.25. Určete korelaci znaků X – počet odetnutých prstů na rukou a Y – počet odetnutých prstů na nohou.
0.1, 0.3, 0.35 a 0.25 jsou simultánní psti $p(x_i, y_j)$.
Data můžeme uspořádat do přehledné tabulky:

	N3	N4	p(x)
R1	0.1	0.3	0.4
R2	0.35	0.25	0.6
p(y)	0.45	0.55	1

$$R(X,Y) = \frac{C(X,Y)}{\sqrt{DX}\sqrt{DY}}$$

$$C(X,Y) = \sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty}[x - EX][y - EY]p(x,y)$$

- $EX = 1 * 0.4 + 2 * 0.6 = 1.6$
- $EY = 3 * 0.45 + 4 * 0.55 = 3.55$

$$\begin{aligned} C(X,Y) &= (1-1.6)(3-3.55)0.1 + (1-1.6)(4-3.55)0.3 + \\ &+ (2-1.6)(3-3.55)0.35 + (2-1.6)(4-3.55)0.25 \\ &= 0.33 * 0.1 - 0.27 * 0.3 - 0.22 * 0.35 + 0.18 * 0.25 = -0.08 \end{aligned}$$

$E(X^2) = 1^2 0.4 + 2^2 0.6 = 2.8$
$E(Y^2) = 3^2 0.45 + 4^2 0.55 = 12.85$
$DX = E(X^2) - (EX)^2 = 2.8 - 1.6^2 = 0.24$
$DY = E(Y^2) - (EY)^2 = 12.85 - 3.55^2 - 1.6^2 = 0.2475$
$R(X,Y) = \frac{C(X,Y)}{\sqrt{DX}\sqrt{DY}} = \frac{-0.08}{\sqrt{0.24}\sqrt{0.2475}} = -0.328.$

```
x <- c( 1, 2)
y <- c( 3, 4)
n <- length(x)
pi <- data.frame(Tri= c(0.1,  0.35),
                 Ctyri=   c(0.3, 0.25),
                 row.names=c('Jeden', 'Dva'))
pix <- apply(pi, 1, sum)
piy <- apply(pi, 2, sum)

EX <- sum(x*pix)
```

```
EY <- sum(y*piy)
DX <- sum(x^2*pix)-EX^2
DY <- sum(y^2*piy)-EY^2
CXY <- sum(c((x-EX)*(y-EY)[1], (x-EX)*(y-EY)[2])*c(as.matrix(pi)))
RXY <- CXY/(sqrt(DX)*sqrt(DY))
(Tab <- round(data.frame(EX=EX, EY=EY, DX=DX, DY=DY, CXY=CXY, RXY=RXY, row.names=
   'Charakteristiky'), digits=4))
```

6.1 Aplikace Moivrovy a Laplaceovy věty

- $X_1, \ldots, X_n$ jsou stochasticky nezáv. náh. veličiny, $X_1 \sim Alt(\theta), \ldots X_n \sim Alt(\theta)$. Pak jejich součet $Y_n = \sum_{i=1}^n X_i$ má binomické rozdělení $\mathrm{Bin}(n, \theta)$. Střední hodnota veličiny Y_n je $EY_n = n\theta$, rozptyl $DY_n = n\theta(1-\theta)$. Podle centrální limitní věty se standardizovaná náhodná veličina $\dfrac{Y_n - n\theta}{\sqrt{n\theta(1-\theta)}}$ asymptoticky řídí standardizovaným normálním rozldělením $Y_n \sim N(0,1)$.

Příklad 6.11. Pravděpodobnost úspěchu při jednom pokusu je 0.3. S jakou pravděpodobností lze tvrdit, že počet úspěchů ve 100 pokusech bude v mezích od 20 do 40? Výpočet proveďte

a) přesně;

b) pomocí aproximace normálním rozložením.

```
# a)
sum(dbinom(20:40, 100, 0.3))
pbinom(40, 100, 0.3)-pbinom(19, 100, 0.3)
# b)
pnorm(40, 100*0.3, sqrt(100*0.3*0.7))-pnorm(19, 100*0.3, sqrt(100*0.3*0.7))
```

Příklad 6.10. Pravděpodobnost, že zakoupený elektrospotřebič bude vyžadovat opravu během záruční doby, je rovna 0.2. Jaká je pravděpodobnost, že během záruční doby bude nutno ze 400 prodaných spotřebičů opravit více než 96? Výpočet proveďte

a) přesně;

b) pomocí aproximace normálním rozložením.

```
# a)
1-pbinom(96, 400, 0.2)
# b)
1-pnorm(96, 400*0.2, sqrt(400*0.2*0.8))
```