# Rozpoznávanie obrazcov - 3. cvicečenie Štatistika II.

Viktor Kocur viktor.kocur@fmph.uniba.sk

DAI FMFI UK

2.3.2020

#### Bernoulliho schéma

#### Bernoulliho schéma

Uvažujeme n na sebe nezávislých pokusov. Pradvdepodobnosť úspechu v každom z nich je p. Potom pre premennú X ktora označuje počet úspečných pokusov platí:

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{1}$$

Študent má vypracovať test, ktorý obsahuje 10 otázok a ku každej z nich sú 4 odpovede, pričom práve jedna je správna. Aké sú pravdepodobnosti, že študent, ktorý látku vôbec nepozná a volí odpovede náhodne, zodpovie správne a) aspoň 5 otázok b) najviac 5 otázok Riešenie a)

Študent má vypracovať test, ktorý obsahuje 10 otázok a ku každej z nich sú 4 odpovede, pričom práve jedna je správna. Aké sú pravdepodobnosti, že študent, ktorý látku vôbec nepozná a volí odpovede náhodne, zodpovie správne a) aspoň 5 otázok b) najviac 5 otázok

$$P(A) = P(X = 5) + P(X = 6) + P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10)$$

Študent má vypracovať test, ktorý obsahuje 10 otázok a ku každej z nich sú 4 odpovede, pričom práve jedna je správna. Aké sú pravdepodobnosti, že študent, ktorý látku vôbec nepozná a volí odpovede náhodne, zodpovie správne a) aspoň 5 otázok b) najviac 5 otázok

$$P(A) = P(X = 5) + P(X = 6) + P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10)$$

$$P(X=5) = \binom{10}{5} \cdot 0.75^5$$

Študent má vypracovať test, ktorý obsahuje 10 otázok a ku každej z nich sú 4 odpovede, pričom práve jedna je správna. Aké sú pravdepodobnosti, že študent, ktorý látku vôbec nepozná a volí odpovede náhodne, zodpovie správne a) aspoň 5 otázok b) najviac 5 otázok

$$P(A) = P(X = 5) + P(X = 6) + P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10)$$

$$P(X=5) = \binom{10}{5} \cdot 0.75^5$$

$$P(A) = \sum_{k=5}^{10} {10 \choose k} 0.25^k \cdot 0.75^{10-k}$$

$$P(A) = P(X = 17) + P(X = 18) + P(X = 19) + P(X = 20)$$

$$P(A) = P(X = 17) + P(X = 18) + P(X = 19) + P(X = 20)$$

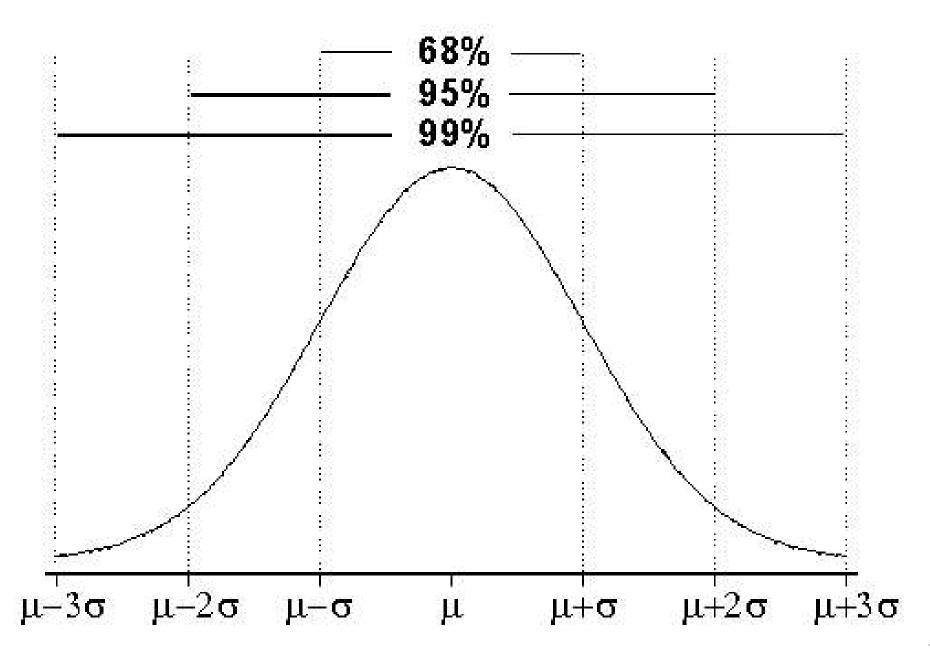
$$P(X=20) = \binom{20}{20} \cdot 0.75^{20} \cdot 0.25^0 = 0.75^{20}$$

$$P(A) = P(X = 17) + P(X = 18) + P(X = 19) + P(X = 20)$$

$$P(X = 20) = \binom{20}{20} \cdot 0.75^{20} \cdot 0.25^0 = 0.75^{20}$$

$$P(A) = \sum_{k=17}^{20} {10 \choose k} 0.75^k \cdot 0.25^{10-k}$$

# Štandardná odchylka



# Odhad parametrov rozdelení

#### Výberový priemer

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

#### Výberový rozptyl

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

#### Smerodajná odchýlka

$$S = \sqrt{S^2}$$

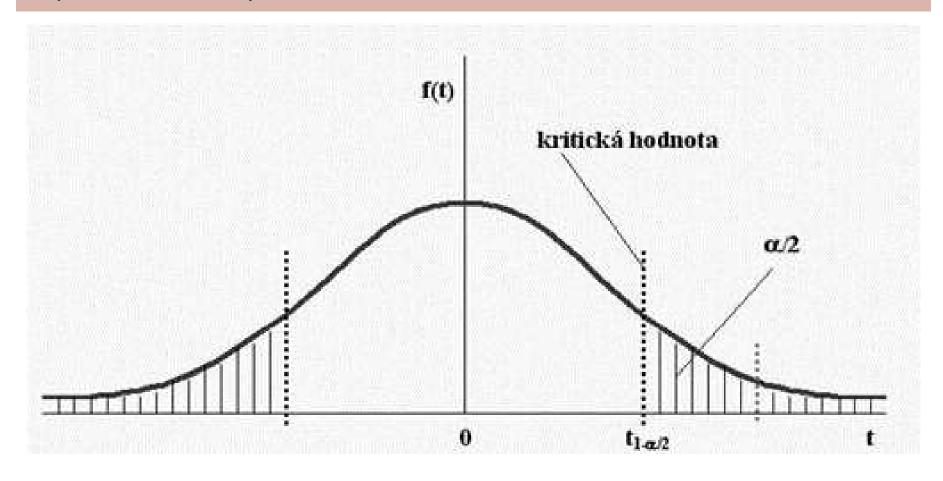
#### Výberová kovariancia

$$S_{XY} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})$$

# Odhad parametrov rozdelení

## Intervalový odhad spoľahlivosti

$$P(G_D < \theta < G_H) = 1 - \alpha$$



# Odhad parametrov rozdelení

$\alpha$	0.01	0.02	0.05	0.1	0.2
$u_{\alpha/2}$	2.5758	2.3263	1.9599	1.6448	1.299

$$X \sim N(0,1)P(|X| > u_{\alpha/2}) = \alpha$$

$$1 - \alpha = P(-u_{\alpha/2} < U < u_{\alpha/2})$$

$$= P(-u_{\alpha/2} < \frac{\overline{X} - \mu}{\sigma} \cdot \sqrt{n} < u_{\alpha/2})$$

$$= P(\overline{X} - u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}) < \mu < \overline{X} + u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}})$$

$$G_D$$

## Testovacie štatistiky

Ak poznáme hodnotu  $\sigma$  originálnej distribúcie. Potom používame normálnu distribúciu:

$$u = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

Ak ju nepoznáme, tak pre n > 30 použijeme:

$$u = \frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$

Ak je n < 30, tak použieje Študentovu distribúciu pre:

$$t = rac{\overline{X} - \mu_0}{rac{S}{\sqrt{n}}}$$
 (stupeň voľnosti = n-1)

# Testovacie štatistiky - Matlab

#### Hodnoty $u_{\alpha}$ a $t_{\alpha}$

Kritické hodnotenia je možé zistiť z tabuliek. My to budeme robiť pomocou matlabu.

#### norminv

norminv(alpha) - vráti kritickú hodnotu pre hladinu alpha pre normálne rozdelenie

#### tinv

tinv(alpha, n) - vráti kritickú hodnotu pre hladinu alpha pre Študentovo rozdelenie pre výberový súbor s n stupňami volnosti.

#### Poznámka

Ak chceme napr. obojstranný interval spoľahlivosti 0.95, tak ako alpha použijeme 0.975, resp. [0.025, 0.0975].

$$n = 15, \sigma = 6.253, \overline{X} = 139.13$$

$$n = 15, \sigma = 6.253, \overline{X} = 139.13$$

$$1 - \alpha = P(\overline{X} - u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}})$$

$$n = 15, \sigma = 6.253, \overline{X} = 139.13$$

$$1 - \alpha = P(\overline{X} - u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}})$$

$$139.13 \pm 2.5758 \cdot \frac{6.253}{\sqrt{15}}$$

$$n = 15, \sigma = 6.253, \overline{X} = 139.13$$

$$1 - \alpha = P(\overline{X} - u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}})$$

$$139.13 \pm 2.5758 \cdot \frac{6.253}{\sqrt{15}}$$

■ 
$$134.97 \le \mu \le 143.28$$

$$n = 20, S = 5, \overline{X} = 112$$

$$n = 20, S = 5, \overline{X} = 112$$

$$1 - \alpha = P(\overline{X} - t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}} \le \mu \le \overline{X} + t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}} )$$

$$n = 20, S = 5, \overline{X} = 112$$

$$1 - \alpha = P(\overline{X} - t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}} \le \mu \le \overline{X} + t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}} )$$

$$\blacksquare 112 \pm 2.093 \cdot \frac{5}{\sqrt{20}}$$

$$n = 20, S = 5, \overline{X} = 112$$

$$1 - \alpha = P(\overline{X} - t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}} \le \mu \le \overline{X} + t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}} )$$

■ 
$$112 \pm 2.093 \cdot \frac{5}{\sqrt{20}}$$

■ 
$$109.65 \le \mu \le 114.34$$

$$n = 10, S = 10.319, \overline{X} = 9.5$$

$$n = 10, S = 10.319, \overline{X} = 9.5$$

$$1 - \alpha = P(\overline{X} - t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}} \le \mu \le \overline{X} + t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}} )$$

$$n = 10, S = 10.319, \overline{X} = 9.5$$

$$1 - \alpha = P(\overline{X} - t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}} \le \mu \le \overline{X} + t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}} )$$

$$9.5 \pm 2.262 \cdot \frac{10.319}{\sqrt{10}}$$

$$n = 10, S = 10.319, \overline{X} = 9.5$$

$$1 - \alpha = P(\overline{X} - t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}} \le \mu \le \overline{X} + t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}} )$$

$$9.5 \pm 2.262 \cdot \frac{10.319}{\sqrt{10}}$$

$$2.118 \le \mu \le 16.881$$

■ 
$$n = 7, \sigma = 0.245, \overline{X} = 1.4$$

■ 
$$n = 7, \sigma = 0.245, \overline{X} = 1.4$$

$$1 - \alpha = P(\overline{X} - u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}})$$

$$n = 7, \sigma = 0.245, \overline{X} = 1.4$$

$$1 - \alpha = P(\overline{X} - u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}})$$

■ 
$$1.4 \pm 1.9599 \cdot \frac{0.245}{\sqrt{7}}$$

$$n = 7, \sigma = 0.245, \overline{X} = 1.4$$

$$1 - \alpha = P(\overline{X} - u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}})$$

■ 
$$1.4 \pm 1.9599 \cdot \frac{0.245}{\sqrt{7}}$$

■ 
$$1.218 \le \mu \le 1.581$$

## Testovanie hypotéz - Matlab

#### ztest

[h, p, ci] = ztest(X, m, sigma, 'Alpha', alpha) - vráti výsledok testu pre hypotézu, že dáta vo vektore X sú z normálnej distribúcie so strednou hodnotou m a štandardnou odchýlkou sigma. h obsahuje 1 ak sa hypotéza nepotrvdí pre danú úroveň významnosti alpha, inak 0, ci obsahuje interval spoľahlivosti.

#### ttest

[h, p, ci] = ttest(X, m, 'Alpha', alpha', alpha) - vráti výsledok testu pre hypotézu, že dáta vo vektore <math>X sú z normálnej distribúcie so trednou hodnotou m a a neznámou štandardnou odchýlkou. h obsahuje 1 ak sa hypotéza nepotrvdí pre danú úroveň významnosti alpha, inak 0, ci obsahuje interval spoľahlivosti.

#### Úloha

Otestuite ztest funkciu na 16. a 17. príklad.

Tvrdíme, že guľôčky vyrobené automatickým sústruhom, majú strednú hodnotu polomeru 10mm. Použitím obojstranného testu významnosti na hladine 0.05 otestujte túto hypotézu ak vo výbere zo 16 guľôčok je priemer ich polomerov 10.3 mm a a)  $\sigma^2=1$ , b)  $S^2=1.21$ . Riešenie a)

Tvrdíme, že guľôčky vyrobené automatickým sústruhom, majú strednú hodnotu polomeru 10mm. Použitím obojstranného testu významnosti na hladine 0.05 otestujte túto hypotézu ak vo výbere zo 16 guľôčok je priemer ich polomerov 10.3 mm a a)  $\sigma^2=1$ , b)  $S^2=1.21$ .

■ 
$$n = 16, \sigma = 1, \overline{X} = 10.3$$

Tvrdíme, že guľôčky vyrobené automatickým sústruhom, majú strednú hodnotu polomeru 10mm. Použitím obojstranného testu významnosti na hladine 0.05 otestujte túto hypotézu ak vo výbere zo 16 guľôčok je priemer ich polomerov 10.3 mm a a)  $\sigma^2=1$ , b)  $S^2=1.21$ .

$$n = 16, \sigma = 1, \overline{X} = 10.3$$

$$1 - \alpha = P(\overline{X} - u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}})$$

Tvrdíme, že guľôčky vyrobené automatickým sústruhom, majú strednú hodnotu polomeru 10mm. Použitím obojstranného testu významnosti na hladine 0.05 otestujte túto hypotézu ak vo výbere zo 16 guľôčok je priemer ich polomerov 10.3 mm a a)  $\sigma^2=1$ , b)  $S^2=1.21$ .

$$n = 16, \sigma = 1, \overline{X} = 10.3$$

$$1 - \alpha = P(\overline{X} - u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}})$$

■ 
$$10.3 \pm 1.9599 \cdot \frac{1}{\sqrt{16}}$$

Tvrdíme, že guľôčky vyrobené automatickým sústruhom, majú strednú hodnotu polomeru 10mm. Použitím obojstranného testu významnosti na hladine 0.05 otestujte túto hypotézu ak vo výbere zo 16 guľôčok je priemer ich polomerov 10.3 mm a a)  $\sigma^2=1$ , b)  $S^2=1.21$ .

Riešenie a)

$$n = 16, \sigma = 1, \overline{X} = 10.3$$

$$1 - \alpha = P(\overline{X} - u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}})$$

■ 
$$10.3 \pm 1.9599 \cdot \frac{1}{\sqrt{16}}$$

■ 
$$9.81 \le \mu \le 10.789$$

Hypotézu nezamietneme

Tvrdíme, že guľôčky vyrobené automatickým sústruhom, majú strednú hodnotu polomeru 10mm. Použitím obojstranného testu významnosti na hladine 0.05 otestujte túto hypotézu ak vo výbere zo 16 guľôčok je priemer ich polomerov 10.3 mm a a)  $\sigma^2=1$ , b)  $S^2=1.21$ . Riešenie b)

Tvrdíme, že guľôčky vyrobené automatickým sústruhom, majú strednú hodnotu polomeru 10mm. Použitím obojstranného testu významnosti na hladine 0.05 otestujte túto hypotézu ak vo výbere zo 16 guľôčok je priemer ich polomerov 10.3 mm a a)  $\sigma^2=1$ , b)  $S^2=1.21$ .

$$n = 16, S = 1.1, \overline{X} = 10.3$$

Tvrdíme, že guľôčky vyrobené automatickým sústruhom, majú strednú hodnotu polomeru 10mm. Použitím obojstranného testu významnosti na hladine 0.05 otestujte túto hypotézu ak vo výbere zo 16 guľôčok je priemer ich polomerov 10.3 mm a a)  $\sigma^2=1$ , b)  $S^2=1.21$ .

$$n = 16, S = 1.1, \overline{X} = 10.3$$

$$1 - \alpha = P(\overline{X} - t_{\alpha/2, n-1} \cdot \frac{s}{\sqrt{n}} \le \mu \le \overline{X} + t_{\alpha/2, n-1} \cdot \frac{s}{\sqrt{n}} )$$

Tvrdíme, že guľôčky vyrobené automatickým sústruhom, majú strednú hodnotu polomeru 10mm. Použitím obojstranného testu významnosti na hladine 0.05 otestujte túto hypotézu ak vo výbere zo 16 guľôčok je priemer ich polomerov 10.3 mm a a)  $\sigma^2=1$ , b)  $S^2=1.21$ .

$$n = 16, S = 1.1, \overline{X} = 10.3$$

$$1 - \alpha = P(\overline{X} - t_{\alpha/2, n-1} \cdot \frac{s}{\sqrt{n}} \le \mu \le \overline{X} + t_{\alpha/2, n-1} \cdot \frac{s}{\sqrt{n}} )$$

$$\blacksquare$$
 10.3  $\pm$  2.131  $\cdot \frac{1.1}{\sqrt{16}}$ 

Tvrdíme, že guľôčky vyrobené automatickým sústruhom, majú strednú hodnotu polomeru 10mm. Použitím obojstranného testu významnosti na hladine 0.05 otestujte túto hypotézu ak vo výbere zo 16 guľôčok je priemer ich polomerov 10.3 mm a a)  $\sigma^2=1$ , b)  $S^2=1.21$ .

Riešenie b)

$$n = 16, S = 1.1, \overline{X} = 10.3$$

$$1 - \alpha = P(\overline{X} - t_{\alpha/2, n-1} \cdot \frac{s}{\sqrt{n}} \le \mu \le \overline{X} + t_{\alpha/2, n-1} \cdot \frac{s}{\sqrt{n}} )$$

$$\blacksquare$$
 10.3  $\pm$  2.131  $\cdot \frac{1.1}{\sqrt{16}}$ 

■ 
$$9.71 \le \mu \le 10.88$$

Hypotézu nezamietneme