Rozpoznávanie obrazcov - 3. cvicečenie
Štatistika II.

Viktor Kocur
viktor.kocur@fmph.uniba.sk

DAI FMFI UK

5.3.2019

Náhodná premenná

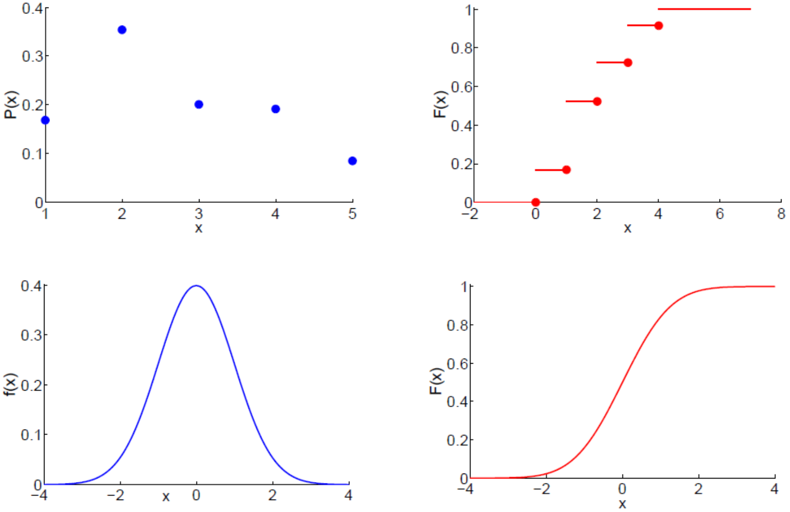
Náhodná premenná

Funkcia, ktorej hodnota je určená výsledkom náhodného pokusu. Priraďuje číselnú hodnotu každému javu.

Distribučná funkcia

Distribučná funkcia - opisuje rozdelenie pravdepodobnosti náhodnej premennej definovanej na pravdepodobnostnom priestore.

Distribučná funkcia



Bernoulliho schéma

Bernoulliho schéma

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

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

12. Príklad

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)

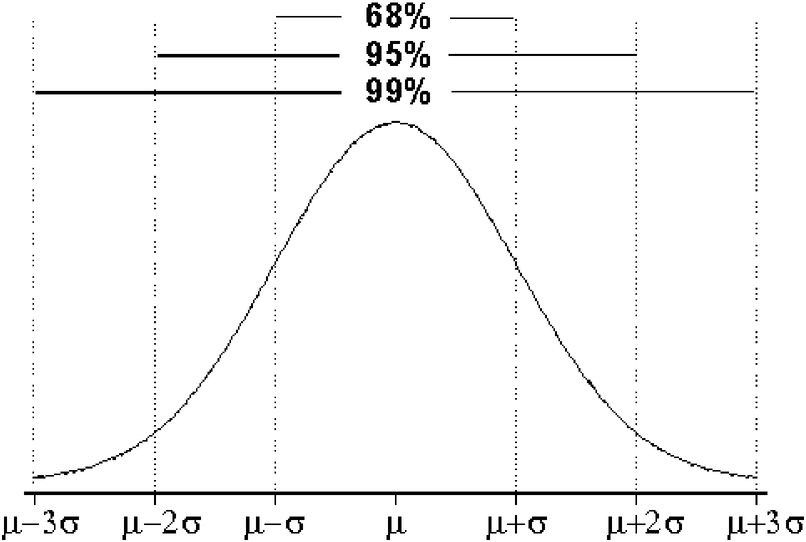
- $P(A) = P(X = 5) + P(X = 6) + P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10)$
- $P(X = 5) = \binom{10}{5} 0.25^5 \cdot 0.75^5$
- $P(A) = \sum_{k=5}^{10} \binom{10}{k} 0.25^k \cdot 0.75^{10-k}$

13. Príklad

Asi 75% zahraničným turistom chutia naše bryndzové halušky. Aká je pravdepodobnosť, že z 20 zahraničných hostí si a) aspoň 17 na haluškách pochutí, b) že halušky budú chutiť všetkým hosťom?

- $P(A) = P(X = 17) + P(X = 18) + P(X = 19) + P(X = 20)$
- $P(X = 20) = \binom{20}{20} 0.75^{20} \cdot 0.25^{0} = 0.75^{20}$
- $P(A) = \sum_{k=17}^{20} \binom{10}{k} 0.75^{k} \cdot 0.25^{10-k}$

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

$S = \sqrt{S^2}$

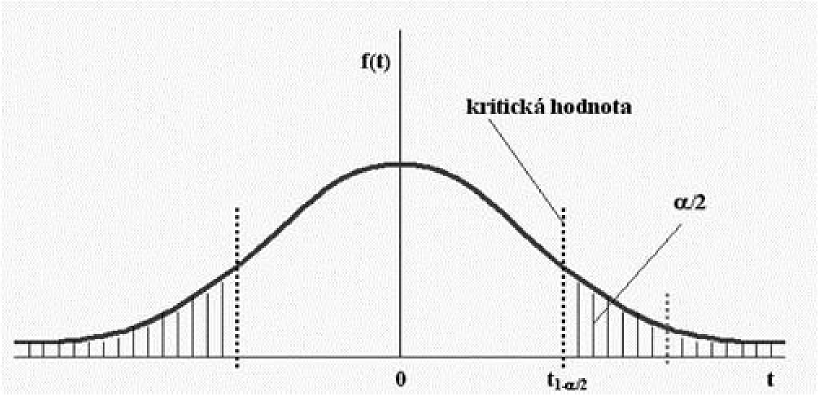
Výberová kovariancia

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

Odhad parametrov rozdelení

Intervalový odhad spoľahlivosti

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



Odhad parametrov rozdelení

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

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

$$\begin{aligned} 1-\alpha &= P(-u_{\alpha/2} < U < u_{\alpha/2}) \\ &= P(-u_{\alpha/2} < \frac{\overline{X}-\mu}{\sigma} < u_{\alpha/2}) \\ &= P(\overline{X} - u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}) \end{aligned}$$

Testovacie štatistiky

Ak poznáme hodnotu σ 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:

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

Testovacie štatistiky - Matlab

Hodnoty u_α a t_α

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].

14. Príklad

Predpokladáme, že výška chlapcov vo veku 9-10 rokov má normálne rozdelenie s neznámou strednou hodnotou a odchýlko $\sigma^2 = 39.112$. Zmerali sme výšku 15 chlapcov a určili výberovú strednú hodnotu 139.13 cm. Určite 99% obojstranný interval spolahlivosti.

- $n = 15, \sigma = 6.253, \overline{X} = 139.13$
- $1 - \alpha = P(\overline{X} - u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}})$
- $139.13 \pm 2.5758 \cdot \frac{6.253}{\sqrt{15}}$
- $134.97 \leq \mu \leq 143.28$

15. Príklad

Letecká spoločnosť odhaduje priemerný počet cestujúcich. V priebehu 20 dní bol priemerný počet cestujúcich 112 s výberovým rozptylom 25. Nájdite 95%-ný obojstranný interval spoľahlivosti pre priemerný počet cestujúcich μ.

- $n = 20, S = 5, \overline{X} = 112$
- $1 - \alpha = P(\overline{X} - t_{\alpha/2,n-1} \cdot \frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{\alpha/2,n-1} \cdot \frac{S}{\sqrt{n}})$
- $112 \pm 2.093 \cdot \frac{5}{\sqrt{20}}$
- $109.65 \leq \mu \leq 114.34$

16. Príklad

Náhodná premenná X má normálne rozdelenie, kde stredná hodnota aj odchýlka sú neznáme. Namerané realizácie X sú 27, 15, -3, -6, 12, 20, 13, 0, 7, 10. Zistite 95%-ný obojstranný interval spoľahlivosti pre strednú hodnotu.

- $n = 10, S = 10.319, \overline{X} = 9.5$
- $1 - \alpha = P(\overline{X} - t_{\alpha/2,n-1} \cdot \frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{\alpha/2,n-1} \cdot \frac{S}{\sqrt{n}})$
- $9.5 \pm 2.262 \cdot \frac{10.319}{\sqrt{10}}$
- $2.118 \leq \mu \leq 16.881$

17. Príklad

Zo základného súboru s normálnym rozdelením, kde je známy rozptyl $\sigma^2 = 0.06$ sme urobili náhodný výber s prvkami 1.3, 1.8, 1.4, 1.2, 0.9, 1.5, 1.7. Zistite 95% interval spoľahlivosti pre strednú hodnotu μ.

- $n = 7, \sigma = 0.245, \overline{X} = 1.4$
- $1 - \alpha = P(\overline{X} - u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}})$
- $1.4 \pm 1.9599 \cdot \frac{0.245}{\sqrt{7}}$
- $1.218 \leq \mu \leq 1.581$

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) - vráti výsledok testu pre hypotézu, že dáta vo vektore 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

Otestujte ztest funkciu na 16. a 17. príklad.

18. 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)

- $n = 16, \sigma = 1, \overline{X} = 10.3$
- $1 - \alpha = P(\overline{X} - u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}})$
- $10.3 \pm 1.9599 \cdot \frac{1}{\sqrt{16}}$
- $9.81 \leq \mu \leq 10.789$
- Hypotézu nezamietneme

18. 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 b)

- $n = 16, S = 1.1, \overline{X} = 10.3$
- $1 - \alpha = P(\overline{X} - t_{\alpha/2,n-1} \cdot \frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{\alpha/2,n-1} \cdot \frac{S}{\sqrt{n}})$
- $10.3 \pm 2.131 \cdot \frac{1.1}{\sqrt{16}}$
- $9.71 \leq \mu \leq 10.88$
- Hypotézu nezamietneme