STATISTIKA (ITK) 7. 5. 2014, teorija, drugi test s pojasnili

1. (2,5) Naključna spremenljivka T ima zvezno porazdelitveno funkcijo F, podano z

$$F(x) = \begin{cases} 0, & x \le -2, \\ ax+b, & -2 \le x \le 8, \\ 1, & 8 \le x. \end{cases}$$

$a = \ldots\ldots,\quad b = \ldots\ldots$

$x_{0,25} = \ldots\ldots,\quad x_{0,5} = \ldots\ldots$

Porazdelitev je . . . .

R.: $0 = F(-2) = -2a + b$ in $1 = F(8) = 8a + b$; razlika $1 = F(8) - F(-2) = 10a$; $a = 0,1$, $b = 0,2$. Porazdelitev je enakomerno zvezna na $[-2, 8]$, saj je za $-2 < x < 8$ gostota verjetnosti $p(x) = F'(x) = a = 1/10$ konstanta!

2. (2) Naključna spremenljivka U je porazdeljena normalno, $U \sim N(\mu, \sigma)$, pri čemer velja

$F_U(28) = 0,5 - \Phi(1)$ in $F_U(48) = 0,5 + \Phi(1,5)$.

$\mu =. \ldots\ldots \quad \sigma = \ldots\ldots$

Kako pa je porazdeljena $V = 100 - 2U$? $V \sim \ldots\ldots\ldots\ldots\ldots$

R.: Iz $F_U(28) = 0,5 - \Phi(1)$ sklepamo, da je $\mu - \sigma = 28$, analogno je $\mu + 1,5\sigma = 48$. Torej je razlika $20 = 48 - 28 = 2,5\sigma$ in zato $\sigma = 8$, $\mu = 28 + \sigma = 36$.
V je porazdeljena normalno s parametroma $E(V) = 100 - 2\mu = 28$ in $\sigma(V) = 2\sigma = 16$.

3. (2) Naključna spremenljivka H je porazdeljena po zakonu $\chi^2(9)$. Kako dobimo približka števil a in b, za kateri velja: $P(H \le a) = 0,01$ in $P(H \ge b) = 0,01$?

R.: Približka lahko preberemo v tabeli kritičnih vrednosti porazdelitve χ^2. Nahajata se v deveti vrstici, a v stolpcu, kjer je kumulativa $0,01$, b pa v stolpcu, kjer je kumulativa $0,99$. Boljša približka dobimo s pomočjo računalniških programov za statistiko, npr. v Excelu z uporabo inverzne funkcije k porazdelitveni funkciji porazdelitve tipa $\chi^2(9)$.

4. (3) Frekvenca K dogodka A v zaporedju n neodvisnih izvedb poskusa je za velike n porazdeljena približno normalno $N(140, 10)$. Določite n in $p = P(A)$, s funkcijo Φ pa izrazite približek za $P(130 \le K \le 140)$.

$n = \ldots\ldots \quad p = \ldots\ldots \quad P(130 \le K \le 140) \approx \ldots\ldots\ldots\ldots\ldots$

R.: Pri velikih n je $b(n, p)$ možno dobro aproksimirati z $N(np, \sqrt{npq})$.

Torej je $np = 140$ in $npq = 10^2$. Kvocient $q = \frac{npq}{np} = 10/14 = 5/7$, zato je $p = 1 - q = 2/7$, od koder sledi še $n = np/p = 140/(2/7) = 490$.

Laplaceov približek za $P(130 \le K \le 140)$ je $\Phi((140 + 1/2 - \mu)/\sigma) - \Phi((130 - 1/2 - \mu)/\sigma)$, kar je v danem primeru enako $\Phi(1/20) - \Phi(-21/20) = \Phi(1/20) + \Phi(21/20)$.

5. Opišite, kako na osnovi velikega števila metov igralne kocke:

a) (2) ocenimo verjetnost dogodka, da v posameznem metu pade 5 pik;

R.: a) s cenilko $\overline{p} = k/n$, kjer je k frekvenca dogodka in n število izvedenih metov.
b) z <u>intervalom zaupanja</u> (pri stopnji zaupanja $1 - \alpha$) $[\overline{p} - z_\alpha SE(\overline{p}),\ \overline{p} + z_\alpha SE(\overline{p})]$, kjer je $SE(\overline{p}) = \sqrt{\overline{p}(1 - \overline{p})/n}$, z_α pa kritična vrednost za α pri porazdelitvi $N(0, 1)$.

b) (2) opravimo preizkus značilnosti za domnevo, da je kocka poštena.

R.: Uporabimo prilagoditveni test (Pearsonov) χ^2, kjer imamo šest razredov s pričakovanimi verjetnostmi $1/6$; število prostostnih stopenj je 5. Testna statistika za n metov se glasi

$$\chi^2 = \sum\nolimits_{j=1}^{6} (n_j - n/6)^2/(n/6) = \frac{1}{6n} \sum\nolimits_{j=1}^{6} (6n_j - n)^2 \ ,$$

tu so n_j frekvence izidov z j pikami.

6. (3) Par spremenljivk (X, Y) je na dani populaciji porazdeljen po (dvorazsežnem) normalnem zakonu. Opišite vse možnosti, s katerimi lahko testiramo domnevo o njuni neodvisnosti.

R.: Kot za splošne porazdelitve lahko tudi v takšnem primeru uporabimo:

a) test s **kontingenčno** tabelo (razdelimo zalogi vrednosti v razrede, izračunamo pričakovane robne frekvence in uporabimo formulo za testno statistiko χ^2 za kontingenčno tabelo);

b) test s **Spearman**ovim koeficientom korelacije rangov R_S: potrebno je predhodno rangirati vrednosti.

Ker vemo, da je porazdelitev normalna, pa zadošča (in je v takšnem primeru bolje) testirati parametrično domnevo, da je korelacijski koeficient ρ enak 0. Testna statistika $T = \frac{R}{\sqrt{1-R^2}}\sqrt{n-2}$ je porazdeljena po $S(n-2)$ oz. približno po $N(0,1)$ za velike n.