1	2	3	Σ

FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

OSNOVE VERJETNOSTI IN STATISTIKE 20010/2011

TEORIJA (8. JULIJ 2011)

IME IN PRIIMEK: ______________________ VPISNA ŠT: [| | | | | | | |]

NAVODILA

Pazljivo preberite besedila vprašanj, predno pričnete pisati odgovore. Čas pisanja je 30 minut. Možnih točk je 30, za pozitivno oceno je potrebno zbrati vsaj polovico (najmanj po 3 pri vsaki nalogi). Veliko uspeha!

(a)

1. Kaj veš o porazdelitveni funkciji slučajne spremnljivke (definicija, kako z njo računamo verjetnost in naštej vsaj 4 osnovne lastnosti), (b) standardnemu odklonu in (c) binomski porazdelitvi (d) (omeni tudi Laplacov obrazec ali pa kako lahko aproksimiramo binomsko porazdelitev z normalno, a pozor, v ta namen moraš definirati bodisi funkcijo napake ali pa normalno porazdelitev)?

(a) $F_X(x) := P(X < x)$, verjetnost pa računamo s seštevanjem $\{p_i\}$ v primeru diskr. in integriranjem $(-\infty, x)$ v primeru zvezne.

[3]

Lastnosti: $F(x) \in [0,1]$, $F(-\infty) = 0$, $F(\infty) = 1$, $F(x)$ je nepadajoča funkcij (v diskretnem primeru je npr. stopničas

[2] (b) $\sigma(X) = \sqrt{DX}$, kjer je $DX = EX^2 - (EX)^2$

(c) Binomska porazdelitev $X \sim B(n,p)$ p_k = verj. da se pri n ponovitvah (neodvi dogodek z verj. p zgodi k-krat

[2]

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

(d)

[3]

$P(k_1 < X < k_2) = \Phi\left(\frac{k_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - np}{\sqrt{npq}}\right)$, kjer je $\Phi(x)$ funkcija napa

$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{1}{2}t^2} dt$. Upoštevali smo, da je za velike n $B(n,p) \approx N(np,$ (in p blizu $\frac{1}{2}$) Laplaceov obrazec

2. Naj slučajna spremenljivka X predstavlja število naprav, ki so na voljo, slučajna spremenljivka Y pa število zaporednih operacij, ki jih moramo opraviti za procesiranje kosa materiala. Verjetnostna funkcija $P(X = x, Y = y) = p(x, y)$ je definirana z naslednjo tabelo:

[3]

$Y \setminus X$	1	2	3	4	q_i
0	0	0·10	0·20	0·10	0,40
1	0·03	0·07	0·10	0·05	0,25
2	0·05	0·10	0·05	0	0,20
3	0	0·10	0·05	0	0,15
p_i	0,08	0,37	0,40	0,15	1

a
$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0{,}08 & 0{,}37 & 0{,}40 & 0{,}15 \end{pmatrix} \sim X$$

$$EX = \begin{matrix} 0{,}08 \\ 0{,}74 \\ 1{,}20 \\ 0{,}60 \end{matrix} = 2{,}62$$

$= \sum_{i=1}^{n} p_i x_i$, pri čemer v primeru $n = \infty$ preverimo $\sum_{i=1}^{\infty} |x_i| p_i < \infty$

(Mimogrede: d bi znal definirati tudi slučajni vektor in njegovo porazdelitveno funkcijo?) a Poišči verjetnostno tabelo spremenljivke X in izračunaj njeno matematično upanje. b Ali sta slučajni spremenljivki X in Y neodvisni? c Za neki drugi slučajni spremnljivki U in V vemo, da je $EU = 2$, $EV = 4$ in $E(UV) = 6$ in nas zanima ali sta lahko neodvisni (odgovor utemelji - kot povsod)?

[2] d

$\vec{X} = (X_1, \ldots, X_n)$ (n-terica sl. spremenljivk X_i)

Slučajni vektor je vektor katerega komponente so slučajne spr.

$$P(X_1 \leq x_1, \ldots, X_n \leq x_n) = P((X_1 \leq x_1) \cap \ldots \cap (X_n \leq x_n)) = F_{\vec{X}}(x_1, \ldots, x_n)$$

[2] b Za neodvisni diskr. slučajni spremenljivki X in Y mora veljati $p_{ij} = p_i \cdot q_j$

vendar v našem primeru npr. $p_{11} = 0 \neq 0{,}08 * 0{,}40 = p_1 \cdot q_1$

[3] c Neodvisne slučajne spremenljivke so tudi nekorelirane (obratno ni nujno res), torej mora veljati $Kov(U,V) = E(UV) - EU \cdot EV = 0$

v našem primeru pa to ne drži: $6 \neq 8 = 2 \cdot 4$

3. a Opiši splošni postopek preverjanja domneve. b Pojasni tudi kaj je zavrnitveni kriterij, c stopnja značilnosti ter d razliko med napakama 1. in 2. vrste.

[4] a

1. Postavimo domnevo (običajno o parametrih): ničelno H_0 in osnovno/alternativno H_1/H_a (ki jo želimo preveriti)
2. Za parameter poiščemo kar se da dobro cenilko (npr. nepristransko) in njeno porazdelitev ali porazdelitev ustrezne statistike (tudi v odv. od števila podatkov/vel. vzorca)
3. Določimo odločitveno pravilo: izberemo stopnjo značilnosti (α) in na osnovi nje ter porazdelitve statistike določimo kritično obmoçje
4. Zberemo/manipuliramo podatke ter na vzorčnih (naključno, neodvisno) podatkih izračunamo (eksperimentalno) vrednost testne statistike (TS.)
5. Primerjamo in naredimo zaključek: če kritično območje eksperimentalno vrednost
(a) vsebuje, ničelno domnevo zavrnemo in sprejmemo osnovno
(b) ne vsebuje, pa pravimo da vzorčni podatki kažejo na statistično neznačilne razlike med domnevno vrednostjo parametra in vzorčno oceno.

b [2]

Zavrnitveni kriterij: α je običajno 10% 5% 1% določimo enostranski (a) ali dvostranski test (b)

(a) $(-\infty, z_\alpha)$ ali (z_α, ∞) kritično

(b) $(-\infty, z_{\alpha/2}) \cup (z_{\alpha/2}, \infty)$ območje (in ne interval)

Stopnja značilnosti c [2] (signifikantnost) je največji α, ki ga je vodja ek pripravljen sprejeti (tj. zgornja meja za napako 1. vr

Pri napaki 1. vrste d [2] zavrnemo pravilno domnevo H_0, pri napaki 2. vrste pa sprejmemo napačno domnevo H_0