1	2	3	Σ

FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Verjetnost in statistika 2010/11

Teorija

2. sept. 2011

Ime in priimek: ______________________ Vpisna št: | | | | | | | | |

Zaokroži rok na katerem si opravil(a) kolokvij: kol., jan., feb., avg. in ocena (v %): ______

Navodila

Pazljivo preberite besedila vprašanj, predno pričnete pisati odgovore. Čas pisanja je 45 minut. Možnih je 360=300+60 bonus točk. Za pozitivno končno oceno iz teorije je potrebno zbrati vsaj 150 točk (najmanj 30 pri vsakem vprašanju). *Veliko uspeha!*

1. Kaj veš o $F_X(x) := P(X < x)$, verjetnost pa računamo s seštevanjem $\{p_i\}$ v diskretnem primeru, in integr. $(-\infty, x)$ v zv. prim

(a) [30] porazdelitveni funkciji slučajne spremnljivke (definicija, kako z njo računamo verjetnost in naštej vsaj 4 osnovne lastnosti),

$F(x) \in [0,1]$ $F(x)$ je nepadajoča funkcija

$F(-\infty) = 0$ v diskretnem primeru je npr. stopničasta

$F(+\infty) = 1$

(b) [20] standardnemu odklonu,

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

in obstaja, če obstaja mat. upanje EX^2.

(c) [20] binomski porazdelitvi, $B(n,p)$ P_k = verjetnost, da se pri n ponovitvah dogodek z verj. p zgodi natanko k-krat $= \binom{n}{k} p^k (1-p)^{n-k}$

zaloga vrednosti: $\{0, 1, \ldots, n\}$

(d) [30] funkciji napake ali pa normalni porazdelitvi.

$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{1}{2}t^2} dt$; $N(\mu, \sigma)$: za $x \in \mathbb{R}$ (zaloga vrednosti) $p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ $\mu = EX$, $\sigma = \sqrt{DX}$

$N(0,1)$ je standardizirana normalna porazdelitev.

(e) [20] Podaj tudi Laplacov obrazec ali pa pojasni, kako lahko aproksimiramo binomsko porazdelitev z normalno.

Laplaceov točkovno obrazec: $P_n(k) \approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}$ ko ~~gre~~ $n \to \infty$ (vsaj 30) in p blizu $\frac{1}{2}$, kjer je $q = 1-p$.

Za velike n je $B(n,p) \approx N(np, \sqrt{npq})$ in lahko izračunamo naslednjo verj.:

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

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:

$Y \setminus X$	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

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

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

(a) [20] Definiraj slučajni vektor in njegovo porazdelitveno funkcijo?

Slučajni vektor $\vec{X} = (X_1, X_2, \ldots, X_n)$ je n-terica slučajnih spremenljivk tj. katerega komponente so slučajne spremenljivke

$P(X_1 < x_1, X_2 < x_2, \ldots, X_n < x_n) = P((X_1 < x_1) \cap \ldots \cap (X_n < x_n)) =: F_{\vec{X}}(x_1, \ldots, x_n)$

(b) [30] Poišči verjetnostno tabelo spremenljivke X in izračunaj njeno matematično upanje (seveda ga prej še definiraj tako za zvezne kot diskretne spremenljivke in povej kdaj obstaja).

$$EX = \begin{cases} \sum_{i=1}^{\infty} x_i p_i & \text{diskr.} \\ \int_{-\infty}^{\infty} x\, p(x)\, dx & \text{zv.} \end{cases}$$

Matematično upanje EX obstaja, če je v disk. pr. $\sum_{i=1}^{\infty} |x_i| p_i < \infty$ (vedno, če je $n \neq \infty$)

oz. v zv. primeru $\int_{-\infty}^{\infty} |x|\, p(x)\, dx < \infty$

(c) [20] Ali sta slučajni spremenljivki X in Y neodvisni?

Za neodvisni slučajni spremenljivki X in Y mora veljati $p_{ij} = p_i q_j$

V našem primeru npr. $p_{11} = 0 \neq 0{,}08 \cdot 0{,}40 = p_1 q_1$, kar pomeni, da sl. spr. X in Y nista neodvisni.

(d) [30] 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)?

Neodvisne slučajne spremenljivke so tudi nekorelirane (obratno ni nujno res), torej mora veljati $\mathrm{cov}(U,V) = E(UV) - EU \cdot EV = 0$,

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

(e) [20] Kdaj sta komponenti dvorazsežnega vektorja (X,Y) z normalno porazdelitvijo $N(\mu_x, \mu_y, \rho, \sigma_x, \sigma_y)$ neodvisni (odgovor utemelji!)?

V primeru $\vec{X} \sim N(\mu_x, \mu_y, \rho, \sigma_x, \sigma_y)$ je $r(X,Y) = \rho$, korelacijski koeficient, tj. $\frac{\mathrm{cov}(X,Y)}{\sigma X \sigma Y}$

Ker je $p(x,y) = \frac{1}{2\pi \sigma_x \sigma_y \sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}$

je $p(x,y) = p_X(x) \cdot p_Y(y)$ natanko tedaj, ko je $\rho = 0$.

robni porazdelitvi

3. (a) [40] Opiši splošni postopek preverjanja domneve.

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 odvisnosti od števila podatkov, tj. velikosti vzorca).

3. Določimo odločitveno pravilo: izberemo stopnjo značilnosti (α) in na osnovi nje ter porazdelitve statistike določimo kritično območje.

4. Zberemo/manipuliramo podatke ter na osnovi vzorčnih podatkov (neodvisni, naključno izbrani) izračunamo (eksperimentalno) vrednost tesne statistike (T.S.)

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

Pojasni tudi kaj je

(b) [15] zavrnitveni kriterij,
α je običajno 10%, 5%, 1%
določimo enostranski ali dvo-str. test: (a) $(-\infty, z_\alpha)$ ali (z_α, ∞)
(b) $(-\infty, -z_{\alpha/2}) \cup (z_{\alpha/2}, \infty)$ kritično območje (in ne interval. zaup.)

(c) [15] stopnja značilnosti, (signifikantnost)
je največji α, ki ga je vodja eksperimenta pripravljen sprejeti (tj. zgornja meja za napako 1. vrste).

(d) [15] moč testa ter
$1-\beta$ je verjetnost zavrnitve ničelne domneve v primeru, ko je le-ta v resnici napačna, kjer β označuje verjetnost napake 2. vrste.

(e) [15] razliko med napakama 1. in 2. vrste.

Pri napaki 1. vrste zavrnemo pravilno domnevo H_0
pri napaki 2. vrste pa "sprejmemo" napačno domnevo H_0.
(ne moremo zavrniti)

(f) [20] Definiraj cenilko (in seveda v ta namen še vzorčno statistiko).

Vzorčna statistika je poljubna simetrična funkcija (tj. njena vrednost je neodvisna od vrstnega reda argumentov) vzorca $Y = g(X_1, X_2, \dots, X_n)$.

Cenilka parametra ζ je vzorčna statistika $C = C(X_1, \dots, X_n)$, katere porazdelitveni zakon je odvisen le od parametra ζ, njene vrednosti pa ležijo v prostoru parametrov.