Sveučilište u Zagrebu
PMF-Matematički odjel
Poslijediplomski specijalistički sveučilišni studij aktuarske matematike

ISPIT

VJEROJATNOST I MATEMATIČKA STATISTIKA

20. 11. 2006.

Vrijeme trajanja ispita: 120 minuta
Ukupan broj bodova: 100
Broj zadataka: 6
Naznačeno je koliko bodova donosi svaki točan odgovor.
Odgovore i rješenja zadataka pišite na dobivenim papirima. Na svakom korištenom papiru naznačite na koji se zadatak odnosi i čitko se potpišite.
Dozvoljeno je korištenje džepnog kalkulatora, statističkih tablica, vlastitih formula ili *Formulae and Tables for Actuarial Examinations* (Institute of Actuaries).

---

1. Iz portfelja istovrsnih polica na slučajan način je izabrano njih 500. Poznato je da se šteta po jednoj polici tijekom godine pojavljuje s vjerojatnosti 0.04 neovisno o ostalim policama. Po jednoj polici osiguranja moguća je najviše jedna šteta. Izračunajte (približno) vjerojatnost da na kraju godine u uzorku neće biti više od 30 šteta.

(15 bodova)

2. Funkcija izvodnica kumulanata neke slučajne varijable X je:

$$C_X(t) = 2\left(\frac{1}{(1-t)^{10}} - 1\right).$$

Izračunajte matematičko očekivanje i varijancu od X.

(15 bodova)

3. Navedenom tablicom zadan je model za broj šteta po polici osiguranja autoodgovornosti nekog osiguravajućeg društva u ovisnosti o faktorima dob vozača (mlađi od 25 godina (< 25) ili stariji ($25+$)) i vrsta automobila (obiteljski ili sportski). Po tom modelu slučajan broj šteta N ovisi o četiri razine (zajedničkog) faktora C: vozač je dobi $25+$ i vozi obiteljski auto, dobi je $25+$ i vozi sportski auto, dobi je < 25 i vozi obiteljski auto, ili je dobi < 25 i vozi sportski auto. Uz zakon razdiobe faktora C u populaciji osiguranika od autoodgovornosti, u tablici su navedena uvjetna očekivanja i uvjetne varijance broja šteta uz uvjet pripadnosti vozača kategoriji c.

kategorija c	$\mathbb{P}(C = c)$	$\mathbb{E}[N \vert C = c]$	$\mathbb{V}\text{ar}[N \vert C = c]$
$25+$, obiteljski a.	0.47	0.15	0.1547
$25+$, sportski a.	0.11	0.19	0.2170
< 25, obiteljski a.	0.29	0.21	0.2095
< 25, sportski a.	0.13	0.28	0.2849

Na osnovi zadanog modela, izračunajte:

(a) očekivani broj šteta po polici autoodgovornosti, tj. $\mathbb{E}[N]$; (5 bodova)

(b) varijancu broja šteta po polici autoodgovornosti, tj. $\mathbb{V}\text{ar}[N]$. (10 bodova)

(ukupno 15 bodova)

4. Podaci o opaženim brojevima šteta po 4000 polica osiguranja koje su bile pod rizikom točno godinu dana, prikazani su u frekvencijskoj tablici:

broj šteta i	frekvencija f_i
0	3288
1	642
2	66
≥ 3	4
ukupno	4000

Redak u tablici označen sa "≥ 3" odnosi se na police po kojima je bilo 3 i više šteta. Pretpostavlja se da se slučajan broj šteta X po polici osiguranja ponaša po Poissonovom zakonu razdiobe $P(\lambda)$, pri čemu je parametar λ nepoznat.

(a) Odredite vjerodostojnost parametra λ na osnovi navedenog uzorka. (5 bodova)

(b) Provjerite da je $\hat{\lambda} = 0.196551$ procjena od λ maksimalne vjerodostojnosti na osnovi navedenog uzorka. (6 bodova)

(c) Sprovedite χ^2-test prilagodbe Poissonovog modela navedenim podacima. Procijenite p-vrijednost i na osnovi nje ocijenite prihvatljivost nulhipoteze. (9 bodova)

(ukupno 20 bodova)

5. Za realizaciju x_1, x_2,..., x_{16} slučajnog uzorka duljine 16 iz normalno distribuirane populacije vrijedi da je

$$\sum_{i=1}^{16} x_i = 51.2, \quad \sum_{i=1}^{16} x_i^2 = 243.19.$$

(a) Procijenite 95% pouzdan interval za populacijsku srednju vrijednost. (8 bodova)

(b) Koristeći procjenu populacijske standardne devijacije dobivenu pomoću zadane realizacije slučajnog uzorka, odredite koliko veliki uzorak bi trebali uzeti da uz 95% pouzdanosti populacijsku srednju vrijednost procijenimo s točnosti od $\varepsilon = 0.5$. (7 bodova)

(ukupno 15 bodova)

6. Na slučajan način odabrano je tridesetoro zaposlenika jednog poduzeća. Zatim su na slučajan način ti zaposlenici raspoređeni u tri jednake grupe. Prva grupa je pohađala tečaj A, druga tečaj B, a treća je bila kontrolna skupina, dakle nije pohađala nikakav tečaj. Oba tečaja su osposobljavala za isti radni proces. Nakon završetka oba tečaja zaposlenici svih skupina podvrgnuti su istovrsnim testovima. Tijekom tog procesa iz raznih razloga troje zaposlenika je odustalo od projekta. Ukupan uspjeh izražen je brojem bodova od 0 do 100 i prikazan je u tablici:

skupina (i)											$\bar{x}_i$	s_i^2
kontrola	55	74	64	62	37	78	50	44	–	–	58.0	202.571
A	63	79	60	75	89	58	75	72	84	69	72.4	103.156
B	64	55	57	73	51	60	62	78	68	–	63.1	75.611

(a) Pomoću usporednih linijskih dijagrama grafički usporedite opažene podatke (po skupinama). (3 boda)

(b) Ispišite ANOVA tablicu i sprovedite test nulhipoteze da nema razlike u distribuciji rezultata testa između tri navedena skupine. Komentirajte. (17 bodova)

(ukupno 20 bodova)

Sveučilište u Zagrebu
PMF-Matematički odjel
Poslijediplomski specijalistički sveučilišni studij aktuarske matematike

ISPIT

VJEROJATNOST I MATEMATIČKA STATISTIKA

20. 11. 2006.

RJEŠENJA ZADATAKA

1. Neka je X ukupanj broj šteta u uzorku. Tada X ima binomnu razdiobu s parametrima $n = 500$ i $\theta = 0.04$. (4)

Vjerojatnost da neće biti više od 30 šteta:

$$\begin{aligned}
& \mathbb{P}(X \leq 30) = && && (3)\\
= \; & \mathbb{P}(X < 30.5) = && \text{(korekcija zbog neprekidnosti)} && (2)\\
= \; & \mathbb{P}\left(\tfrac{X-20}{\sqrt{19.2}} < \tfrac{10.5}{4.38} = 2.40\right) = && && (2)\\
\approx \; & \Phi(2.40) = && \text{(CGT)} && (2)\\
= \; & 0.9918 && \text{(tablice)} && (2)
\end{aligned}$$

(15)

2. Za f.i. kumulanata vrijedi:

$$\begin{aligned}
C'_X(t) &= \tfrac{20}{(1-t)^{11}} && (2)\\
\Rightarrow \quad \mathbb{E}[X] &= C'_X(0) = && (3)\\
&= 20 && (2)\\
C''_X(t) &= \tfrac{220}{(1-t)^{12}} && (2)\\
\Rightarrow \quad \mathbb{V}\mathrm{ar}[X] &= C''_X(0) = && (4)\\
&= 220. && (2)
\end{aligned}$$

(15)

3. (*a*) Matematičko očekivanje od N:

$$\begin{aligned}
\mathbb{E}[N] &= \mathbb{E}[\mathbb{E}[N|C]] = && (2)\\
&= \textstyle\sum_c \mathbb{E}[N|C=c]\cdot\mathbb{P}(C=c) = && (1)\\
&= 0.15\cdot 0.47 + 0.19\cdot 0.11 + 0.21\cdot 0.29 + 0.28\cdot 0.13 = 0.1887. && (1)
\end{aligned}$$

(*b*) Račun za varijancu od N:

$$\begin{aligned}
\mathbb{E}[\mathbb{V}\mathrm{ar}[N|C]] &= \textstyle\sum_c \mathbb{V}\mathrm{ar}[N|C=c]\cdot\mathbb{P}(C=c) = && (1)\\
&= 0.1547\cdot 0.47 + 0.2170\cdot 0.11 + 0.2095\cdot 0.29 + \\
&\quad +0.2849\cdot 0.13 = 0.1944, && (1)\\
\mathbb{E}[\mathbb{E}[N|C]^2] &= \textstyle\sum_c \mathbb{E}[N|C=c]^2\cdot\mathbb{P}(C=c) = && (1)\\
&= 0.15^2\cdot 0.47 + 0.19^2\cdot 0.11 + 0.21^2\cdot 0.29 + \\
&\quad +0.28^2\cdot 0.13 = 0.0375 && (1)\\
\Rightarrow\quad \mathbb{V}\mathrm{ar}[E[N|C]] &= \mathbb{E}[\mathbb{E}[N|C]^2] - \mathbb{E}[\mathbb{E}[N|C]]^2 = && (1)\\
&= \mathbb{E}[\mathbb{E}[N|C]^2] - \mathbb{E}[N]^2 = && (1)\\
&= 0.0375 - 0.1887^2 = 0.0019, && (1)\\
\Rightarrow\quad \mathbb{V}\mathrm{ar}[N] &= \mathbb{E}[\mathbb{V}\mathrm{ar}[N|C]] + \mathbb{V}\mathrm{ar}[\mathbb{E}[N|C]] = && (3)\\
&= 0.1944 + 0.0019 = 0.1963. && (1)
\end{aligned}$$

(15)

4. (*a*) Vjerodostojnost od λ (do na konstantu):

$$L(\lambda) = (e^{-\lambda})^{3288} \cdot (\lambda e^{-\lambda})^{642} \cdot (\frac{\lambda^2}{2} e^{-\lambda})^{66} \cdot (1 - e^{-\lambda}(1 + \lambda + \frac{\lambda^2}{2}))^4. \tag{5}$$

(*b*) Log-vjerodostojnost:

$$\ell(\lambda) = -4000\lambda + 774 \log \lambda + 4 \log(e^{\lambda} - (1 + \lambda + \frac{\lambda^2}{2})). \tag{2}$$

Budući da je MLE rješenje stacionarne jednadžbe:

$$\ell'(\lambda) = 0 \quad \Leftrightarrow \quad -4000 + \frac{774}{\lambda} + 4 \cdot \frac{e^{\lambda} - (1 + \lambda)}{e^{\lambda} - (1 + \lambda + \frac{\lambda^2}{2})} = 0, \tag{2}$$

dovoljno je provjeriti da ju $\hat{\lambda} = 0.196551$ zadovoljava:

$$\ell'(0.196551) = -0.008 \approx 0, \tag{1}$$

što je zadovoljeno (do na zadovoljavajuću numeričku točnost). (1)

(*c*) Testitamo nulhipotezu $H_0 : X \sim P(\lambda)$ u odnosu na alternativu da to nije tako. (1)
Tablica:

i	f_i		p_i	e_i		$\frac{(f_i - e_i)^2}{e_i}$
0	3288		0.8216	3286.2		0.0009
1	642		0.1615	645.9		0.0237
2	66	70	0.0159	63.5	67.9	0.0682
≥ 3	4		0.0011	4.4		
Σ	4000		1.0001	4000.0		0.0928

(2 + 1)

Broj stupnjeva slobode: $3 - 1 - 1 = 1$. (1)
Testna statistika: $H \overset{H_0}{\sim} \chi^2(1)$, opažena vrijednost: $h = 0.0928$. (2)
p-vrijednost: $\mathbb{P}(H > 0.0928 | H_0) = 0.76$. (1)
Nemamo dokaza u korist odbacivanja H_0, dakle, model je vrlo dobro prilagođen. (1)

(20)

Napomena: U ovom primjeru moguće je test sprovesti i bez združivanja zadnja dva razreda. U tom slučaju imamo ukupno dva stupnja slobode i sličan zaključak testa.

5. Procjena aritmetičke sredine i uzoračke standardne devijacije je

$$
\begin{aligned}
\bar{x} &= \frac{\sum_{i=1}^{16}}{16} = \frac{51.2}{16} = && (1)\\
&= 3.2 && (1)\\
s &= \sqrt{\frac{1}{15}\sum_{i=1}^{16} x_i^2 - \frac{16}{15}\bar{x}^2} = \sqrt{5.29} = && (2)\\
&= 2.3 && (1)
\end{aligned}
$$

Neka je μ populacijska srednja vrijednost.

(a) Procjena 95% pouzdanog intervala za μ:

$$\bar{x} \pm t_{0.025}(15) \cdot \frac{s}{\sqrt{16}} = 3.2 \pm 2.131 \cdot \frac{2.3}{4} = 3.2 \pm 1.2 \qquad (2+2+1)$$

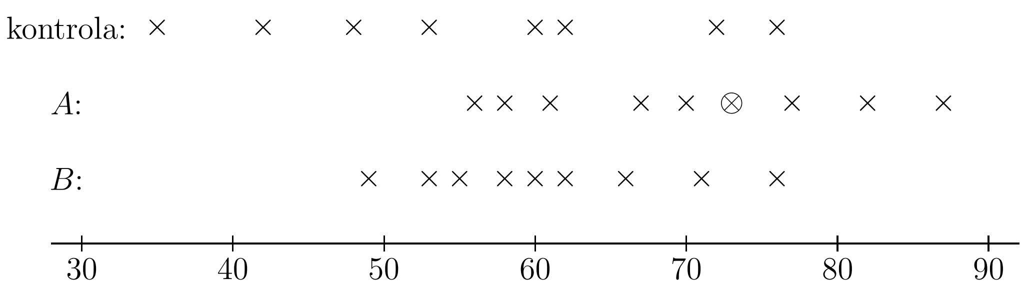
(b) Duljina n traženog uzorka zadovoljava nejednadžbu:

$$
\begin{aligned}
& 2 \cdot \tfrac{s}{\sqrt{n}} \leq \varepsilon && (2)\\
\Leftrightarrow\ & n \geq 84.64. && (2)
\end{aligned}
$$

Dakle, $n \geq 85$. (1)

(15)

6. (*a*) Usporedni linijski dijagrami: (3)



(*b*) Statistike:

$$
\begin{aligned}
\bar{x} &= \tfrac{n_c\bar{x}_c + n_A\bar{x}_A + n_B\bar{x}_B}{n} = & (1)\\
&= \tfrac{8\cdot 58.0 + 10\cdot 72.4 + 9\cdot 63.1}{27} = 65.037 & (1)\\
\text{SSE} &= (n_c-1)s_c^2 + (n_A-1)s_A^2 + (n_B-1)s_B^2 = & (1)\\
&= 7\cdot 202.571 + 9\cdot 103.156 + 8\cdot 75.611 = 2951.29 & (1)\\
\text{MSE} &= \tfrac{\text{SSE}}{n-k} = & (1)\\
&= \tfrac{2951.29}{27-3} = 122.97 & (1)\\
\text{SST} &= n_c(\bar{x}_c-\bar{x})^2 + n_A(\bar{x}_A-\bar{x})^2 + n_B(\bar{x}_B-\bar{x})^2 = & (1)\\
&= 971.674 & (1)\\
\text{MST} &= \tfrac{\text{SST}}{k-1} = & (1)\\
&= \tfrac{971.674}{2} = 485.837 & (1)\\
f &= \tfrac{\text{MST}}{\text{MSE}} = & (1)\\
&= \tfrac{485.837}{122.97} = 3.95085 & (1)
\end{aligned}
$$

ANOVA-tablica:

izvor varijabilnosti	stupnjevi slobode	sume kvadrata	srednji kvadrati	testna statist.
skupina	2	971.67	485.84	3.95
greške	24	2951.29	122.97	—
ukupno	26	3922.96	—	—

Neka je sa μ_i označena srednja populacijska vrijednost rezultata testa za skupinu i.
Za test nulhipoteze $H_0 : \mu_c = \mu_A = \mu_B$, (1)
testna statistika je: $F = \frac{\text{MST}}{\text{MSE}} \overset{H_0}{\sim} F(2, 24)$. (1)
p-vrijednost: $\mathbb{P}(F > 3.95 | H_0) = 0.033$. (1)
Dakle, ima značajne razlike u distribuciji rezultata testa između tri opažane skupine. (2)

(20)