Sveučilište u Zagrebu
PMF-Matematički odjel
Poslijediplomski stručni studij aktuarske matematike

ISPIT

VJEROJATNOST I MATEMATIČKA STATISTIKA

4. 10. 2004.

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. Portfelj se sastoji od 2000 istovrsnih polica osiguranja. Poznato je da se šteta po jednoj polici pojavljuje s vjerojatnosti 0.015 neovisno o ostalim policama. Po jednoj polici osiguranja moguća je najviše jedna šteta. Izračunajte (približno) vjerojatnost da će po tom portfelju biti više od 40 šteta.

(15 bodova)

2. Neka je X binomna slučajna varijabla s parametrima (n, θ).

(a) Pokažite da je funkcija izvodnica vjerojatnosti od X jednaka $G_X(t) = (1 - \theta + t\theta)^n$. (7 bodova)

(b) Izvedite pripadnu funkciju izvodnicu momenata. (8 bodova)

(ukupno 15 bodova)

3. 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)

4. Broj odlazaka aktuara s posla nakon redovnog radnog vremena tijekom radnog tjedna modelira se kao binomna slučajna varijabla X s parametrima (n, θ) gdje je $n = 5$, a $\theta = \frac{4}{5}$. Za uvjetnu razdiobu ukupnog vremena Y koje je aktuar proveo na poslu tijekom tjedna (u satima) ako je taj tjedan morao na poslu ostati dulje x dana, vrijedi:

$$\mathbb{E}[Y|X = x] = 4(x + 10), \quad \mathbb{V}\mathrm{ar}[Y|X = x] = x.$$

(a) Koliko iznosi matematičko očekivanje $\mathbb{E}[Y]$ ukupnog vremena koje aktuar provodi na poslu tijekom tjedna? (7 bodova)

(b) Izračunajte bezuvjetnu varijancu od Y (tj. varijancu $\mathbb{V}\mathrm{ar}[Y]$ marginalne razdiobe od Y). (8 bodova)

(ukupno 15 bodova)

5. Poduzeće želi procijeniti postotak (proporciju) svojih klijenata koji su voljni kupovati preko interneta. Preciznije, treba procijeniti 95% pouzdani interval za taj parametar.

(a) Pokažite da na osnovi slučajnog uzorka duljine 200 klijenata tog poduzeća, taj interval neće biti veći od 13.9% (tj. njegova duljina nije veća od 0.139). (8 bodova)

(b) Izračunajte kolika bi trebala biti duljina slučajnog uzorka da bi 95% pouzdani interval za navedeni parametar bio najviše 10%. (7 bodova)

Uputa: Izraz $p(1 - p)$ za $p \in [0, 1]$ ocijenite odozgo sa $1/4$. (15 bodova)

6. Za zadanih 12 vrijednosti x_1, x_2,..., x_{12} (nezavisne) varijable poticaja X izmjerene su pripadne vrijednosti y_1, y_2,..., y_{12} (zavisne) varijable odziva Y. Na taj način dobiven je uzorak (x_i, y_i), $i = 1, 2, \dots, 12$ za koji vrijedi

$$\sum_{i=1}^{12} x_i = 516.4, \qquad \sum_{i=1}^{12} x_i^2 = 22741.34$$

$$\sum_{i=1}^{12} y_i = 14821, \qquad \sum_{i=1}^{12} y_i^2 = 18695125$$

$$\sum_{i=1}^{12} x_i y_i = 650264.8.$$

(a) Uz pretpostavku da je model za vezu varijabli poticaja X i odziva Y jednostavni linearni regresijski model, procijenite pravac regresije. (7 bodova)

(b) Konstruirajte i procijenite 95%-pouzdan interval za koeficijent smjera regresijskog pravca. (5 bodova)

(c) Testirajte nulhipotezu da je koeficijent smjera jednak nuli u odnosu na alternativu da to nije tako, uz razinu značajnosti od 5%. (3 boda)

(d) Konstruirajte i procijenite 95%-pouzdan interval za očekivanu (srednju) vrijednost varijable Y ako je $X = 50$. (2 boda)

(e) Opišite linearni regresijski model i navedite koje ste sve pretpostavke na njega koristili u zadacima $(a - d)$. (3 boda)

(ukupno 20 bodova)

Sveučilište u Zagrebu
PMF-Matematički odjel
Poslijediplomski stručni studij aktuarske matematike

ISPIT

VJEROJATNOST I MATEMATIČKA STATISTIKA

4. 10. 2004.

RJEŠENJA ZADATAKA

1. Neka je X ukupanj broj šteta u portfelju. Tada X ima binomnu razdiobu s parametrima $n = 2000$ i $\theta = 0.015$. (4)
Vjerojatnost da će biti više od 40 šteta:

$$
\begin{aligned}
& \mathbb{P}(X > 40) = && && (3)\\
= \ & \mathbb{P}(X > 40.5) = && \text{(korekcija zbog neprekidnosti)} && (2)\\
= \ & \mathbb{P}\left(\tfrac{X-30}{\sqrt{29.55}} > \tfrac{10.5}{\sqrt{29.55}} = 1.93\right) = && && (2)\\
\approx \ & 1 - \Phi(1.93) = && \text{(CGT)} && (2)\\
= \ & 1 - 0.973 = && \text{(tablice)} && (1)\\
= \ & 0.027. && && (1)
\end{aligned}
$$

(15)

2. (*a*) Funkcija izvodnica vjerojatnosti od X:

$$
\begin{aligned}
& G_X(t) = \mathbb{E}[t^X] = && (3)\\
= \ & \textstyle\sum_{k=0}^{n} t^k \cdot \binom{n}{k} \theta^k (1-\theta)^{n-k} && (2)\\
= \ & \textstyle\sum_{k=0}^{n} \binom{n}{k} (t\theta)^k (1-\theta)^{n-k} = \text{(binomni teorem)} = (1 - \theta + t\theta)^n. && (2)
\end{aligned}
$$

(*b*) Funkcija izvodnica momenata od X:

$$
\begin{aligned}
& M_X(t) = \mathbb{E}[e^{tX}] = && && (3)\\
= \ & G_X(e^t) = && && (3)\\
= \ & (1 - \theta + \theta \cdot e^t)^n && \text{(prema}(a)). && (2)
\end{aligned}
$$

(15)

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

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

(*b*) Log-vjerodostojnost:

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

Budući da je MLE rješenje stacionarne jednadžbe:

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

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

$$\ell'(0.196551) = -0.008 \approx 0, \quad (1)$$

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.

4. $\mathbb{E}[X] = 5 \cdot \frac{4}{5} = 4$, $\mathbb{V}\text{ar}[X] = 5 \cdot \frac{4}{5} \cdot \frac{1}{5} = \frac{4}{5}$ (2+2)

(a) Očekivanje od Y:

$$\begin{aligned} \mathbb{E}[Y] &= \mathbb{E}[\mathbb{E}[Y|X]] = && (2)\\ &= \mathbb{E}[4(X+10)] = 4(\mathbb{E}[X]+10) = && (2)\\ &= 4 \cdot (4+10) = 56. && (1) \end{aligned}$$

(b) Varijanca od Y:

$$\begin{aligned} \mathbb{V}\text{ar}[Y] &= \mathbb{E}[\mathbb{V}\text{ar}[Y|X]] + \mathbb{V}\text{ar}[\mathbb{E}[Y|X]] = && (3)\\ &= \mathbb{E}[X] + \mathbb{V}\text{ar}[4(X+10)] = \mathbb{E}[X] + 16\mathbb{V}\text{ar}[X] = && (2)\\ &= 4 + 16 \cdot \tfrac{4}{5} = \tfrac{84}{5} = 16.8. && (1) \end{aligned}$$

(15)

5. (a) Budući da se radi o Bernoullijevom modelu, aproks. 95% pouzdani interval za parametar proporcije p je $\hat{p} \pm 1.96\sqrt{p(1-p)/200}$. (4)

Duljina tog intervala je $2 \cdot 1.96\sqrt{p(1-p)/200}$ (2)

što je najviše (ocjena iz upute) $1.96\sqrt{1/200} = 0.139 = 13.9\%$. (2)

(b) Traži se takav n da je $2 \cdot 1.96\sqrt{0.25/200} \leq 0.1$. (3)

Odavde slijedi da je $n \geq 385$. (4)

(15)

6. (*a*) Opažene vrijednosti pomoćnih statistika:

$$\begin{aligned} \bar{x} &= \tfrac{516.4}{12} = 43.3 & \bar{y} &= \tfrac{1481}{12} = 1235.083 & (\tfrac{1}{2}+\tfrac{1}{2}) \\ S_{xx} &= \textstyle\sum_{i=1}^{12} x_i^2 - 12\cdot\bar{x}^2 = & S_{yy} &= \textstyle\sum_{i=1}^{12} y_i^2 - 12\cdot\bar{y}^2 = & (\tfrac{1}{2}) \\ &= 522.37 & &= 389965 & (\tfrac{1}{2}) \\ S_{xy} &= \textstyle\sum_{i=1}^{12} x_i y_i - 12\cdot\bar{x}\bar{y} = & & & (\tfrac{1}{2}) \\ &= 12517.35 & & & (\tfrac{1}{2}) \end{aligned}$$

Parametri pravca:

$$\begin{aligned} \hat{\alpha} &= \tfrac{S_{xy}}{S_{xx}} = & \hat{\beta} &= \bar{y} - \hat{\alpha}\bar{x} = & (1+1) \\ &= 23.962612 & &= 203.9719 & (\tfrac{1}{2}+\tfrac{1}{2}) \end{aligned}$$

Regresijski pravac: $y = 23.963x + 203.97$ (1)

(*b*) 95% pouzdani interval za α: $\hat{\alpha} \pm t_{0.025}(10)\hat{\sigma}\sqrt{\frac{1}{S_{xx}}}$. (1)
Iz tablice za t-razdiobu: $t_{0.025}(10) = 2.228$. (1)

$$\begin{aligned} \text{SSE} &= S_{yy} - \hat{\alpha}^2 S_{xx} = 90016.61 & (\tfrac{1}{2}+\tfrac{1}{2}) \\ \hat{\sigma}^2 &= \tfrac{\text{SSE}}{n-2} = \tfrac{90016.61}{10} = 9001.661 \quad \Rightarrow \quad \hat{\sigma} = 94.877 & (\tfrac{1}{2}+\tfrac{1}{2}) \end{aligned}$$

Dakle, opaženi 95% pouzdani interval je:
$23.963 \pm 2.228 \cdot 94.88 \cdot \sqrt{\frac{1}{522.37}} = 23.963 \pm 9.249.$ (1)

(*c*) Budući da se nula ne nalazi u 95% poudanom intervalu za α, odbacujemo nulhipotezu $H_0 : \alpha = 0$ u odnosu na dvostranu alternativu uz razinu značajnosti od 5%. (3)

(*d*) Procjena za $\mathbb{E}[Y|50]$: $\hat{\mathbb{E}}[Y|50] = 23.963 \cdot 50 + 203.97 = 1402.10$.
95% pouzdani interval: $\hat{\mathbb{E}}[Y|50] \pm t_{0.025}(10)\hat{\sigma}\sqrt{\frac{1}{12} + \frac{(50-\bar{x})^2}{S_{xx}}}$. (1)
Opaženi 95% pouzdani interval: 1402.10 ± 88.77. (1)

(*e*) Model je

$$Y_i = \alpha x_i + \beta + \varepsilon_i, \quad i = 1, 2, \ldots, n,$$

gdje su α, β parametri modela, a ε_i, $i = 1, 2, \ldots, n$, su slučajne varijable (slučajne greške) (1)
za koje smo pretpostavili da su:

(*i*) centrirane ($\mathbb{E}[\varepsilon_i] = 0$ za sve i) $(\frac{1}{2})$

(*ii*) jednake varijance ($\mathbb{V}\text{ar}[\varepsilon_i] = \sigma^2$ za sve i) $(\frac{1}{2})$

(*iii*) nekorelirane ($\text{cov}[\varepsilon_i, \varepsilon_j] = 0$ za sve $i \neq j$) $(\frac{1}{2})$

(*iv*) normalno distribuirane. $(\frac{1}{2})$

(20)