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

ISPIT

VJEROJATNOST I MATEMATIČKA STATISTIKA

2. 6. 2003.

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. Neko je osiguravajuće društvo (OD) za neživotna osiguranja prodalo 1000 istovrsnih polica osiguranja. Po svakoj polici tog osiguranja, šteta u iznosu od 100 kn nastupa s vjerojatnošću 0.05, a šteta od 500 kn s vjerojatnošću 0.01. Ostali iznosi šteta nisu mogući (osim da nema štete po toj polici).

(a) Kolika je vjerojatnost da po jednoj polici osiguranja neće biti štete? (2 boda)

(b) Premija osiguranja jednaka je matematičkom očekivanju iznosa šteta po polici tog osiguranja. Koliko iznosi premija za navedeno osiguranje? (2 boda)

(c) Kažemo da će OD poslovati nesolventno po navedenom portfelju od 1000 polica ako će ukupni iznos šteta po tom portfelju (po isteku razdoblja osiguranja) biti veći od zbroja ukupnog prihoda od premija i iznosa pričuve. Kolika je vjerojatnost nesolventnosti OD-a po tom portfelju ako je OD osiguralo pričuvu u iznosu od 5000 kn? (6 bodova)

(ukupno 10 bodova)

2. Neka su X i Y nezavisne Γ-distribuirane slučajne varijable, $X \sim \Gamma(\alpha, 1/\lambda)$, $Y \sim \Gamma(\beta, 1/\lambda)$ ($\alpha > 0$, $\beta > 0$, $\lambda > 0$). Pomoću funkcije izvodnice momenata:

(a) izračunajte prvi, drugi i treći moment od X, te pomoću njih koeficijent asimetrije $\alpha_3(X)$ od X,

$$\alpha_3(X) = \mathbb{E}[\left(\frac{X - \mathbb{E}[X]}{\sigma(X)}\right)^3];$$

(10 bodova)

(b) pokažite da i $X+Y$ ima Γ-razdiobu i odredite joj parametre. (5 bodova)

(ukupno 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	$\mathbb{P}(C = c)$	$\mathbb{E}[N \mid C = c]$	$\mathbb{V}\mathrm{ar}[N \mid C = c]$
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. $\mathbb{V}\mathrm{ar}[N]$. (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. Agencija za istraživanje tržišta sprovela je 1998. g. sljedeće istraživanje. Na slučajan način odabrano je 400 kućanstava iz jednog velikog grada. U tom uzorku opaženo je točno 68 kućanstava u kojima je barem jedan član učlanjen u neki *health/fitness* klub.

(a) Na osnovi opaženog uzorka procijenite 95% pouzdani interval za parametar proporcije onih kućanstava u populaciji kućanstava tog grada za koje vrijedi da je barem jedan njihov član učlanjen u neki *health/fitness* klub. (5 bodova)

(b) Ista agencija je 1999. g. sprovela slično istraživanje na slučajnom uzorku iste veličine (400). Tada je opaženo točno 80 kućanstava u kojima je barem jedan član učlanjen u neki *health/fitness* klub. Sprovedite prikladni jednostrani test hipoteze da je proporcija kućanstava sa opažanim svojstvom u populaciji kućanstava istoga (promatranog) grada porasla od 1998. do 1999. g. Procijenite p-vrijednost i na osnovi nje ocijenite održivost postavljene hipoteze. (10 bodova)

(ukupno 15 bodova)

6. Kao dio istraživanja uspješnosti predloženog državnog plana privatizacije velikih poduzeća, sprovedeno je sljedeće mjerenje dvodimenzionalnog vektora (Z, P) na zaposlenicima. Komponente vektora (Z, P) su varijable:

Z = zadovoljstvo zaposlenika predloženim planom
P = predanost zaposlenika poduzeću.

Na slučajan način odabrano je deset velikih poduzeća. Zatim se iz svakog poduzeća birao slučajan uzorak zaposlenika (svi uzorci su jednakih duljina). Svaki zaposlenik morao je rangirati svoje zadovoljstvo planom i svoju predanost poduzeću na skali od 1 (najniži stupanj zadovoljstva/predanosti) do 10 (najviši stupanj zadovoljstva/predanosti). Srednje vrijednosti zadovoljstva planom i predanosti poduzeću zaposlenika nekog poduzeća čine pokazatelje zadovoljstva planom (X) i predanosti (Y) zaposlenika tog poduzeća. Rezultati mjerenja (opažene vrijednosti (x, y) od (X, Y)) prikazani su u tablici:

poduzeće	A	B	C	D	E	F	G	H	I	J
x	5.05	4.12	5.38	4.17	3.81	4.47	5.41	4.88	4.64	5.19
y	5.36	4.59	5.42	4.35	4.03	5.34	5.64	4.89	4.52	5.88

Zbirno:

$$\sum_{i=1}^{10} x_i = 47.12, \qquad \sum_{i=1}^{10} y_i = 50.02,$$
$$\sum_{i=1}^{10} x_i^2 = 224.8554, \quad \sum_{i=1}^{12} y_i^2 = 253.5796 \quad \sum_{i=1}^{10} x_i y_i = 238.3676.$$

(a) Prikažite podatke u dijagramu raspršenja (skicirajte) i komentirajte kakva je priroda veze varijabli X i Y, te kako su asocirane. Zatim izračunajte Pearsonov koeficijent korelacije i interpretirajte ga.
(8 bodova)

(b) Prilagodite jednostavni linearni regresijski model (pravac regresije) podacima uz interpretaciju varijable X kao varijable poticaja i Y kao varijable odziva. (3 boda)

(c) Izračunajte koeficijent determinacije R^2 i interpretirajte ga. (2 boda)

(d) Uz pretpostavku da su slučajne pogreške normalno distribuirane i da zadovoljavaju Gauss-Markovljeve uvjete, procijenite njihovu zajedničku varijancu σ^2 i izračunajte njezin 95% pouzdan interval. (4 boda)

(e) Uz iste pretpostavke kao u (d), konstruirajte i procijenite 95%-pouzdan interval za koeficijent smjera regresijskog pravca. (4 boda)

(f) Za poduzeća čiji je pokazatelj zadovoljstva zaposlenika planom jednak 5.0, procijenite očekivanu (srednju) vrijednost pokazatelja predanosti zaposlenika tome poduzeću i procijenite 95%-pouzdan interval za tu vrijednost. (4 boda)

(ukupno 25 bodova)

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

ISPIT

VJEROJATNOST I MATEMATIČKA STATISTIKA

2. 6. 2003.

RJEŠENJA ZADATAKA

1. Neka je X slučajan iznos štete po jednoj polici osiguranja, a $S = X_1 + X_2 + \cdots + X_{1000}$ ukupan iznos šteta po portfelju od 1000 istovrsnih polica osiguranja ($X_1, X_2, \ldots, X_{1000}$ su n.j.d. kao X).

(a) Vjerojatnost da neće biti štete po polici osiguranja:

$$\begin{aligned} \mathbb{P}(X=0) &= 1-\mathbb{P}(X=100)-\mathbb{P}(X=500) = & (1)\\ &= 1-0.05-0.01 = 0.94. & (1) \end{aligned}$$

(b) Premija za policu osiguranja:

$$\begin{aligned} \mathbb{E}[X] &= 0\cdot\mathbb{P}(X=0)+100\cdot\mathbb{P}(X=100)+500\cdot\mathbb{P}(X=500) = & (1)\\ &= 0+100\cdot 0.05+500\cdot 0.01 = 10. & (1) \end{aligned}$$

(c) Za ukupan iznos šteta S vrijedi:

$$\begin{aligned} \mathbb{E}[S] &= 1000\cdot\mathbb{E}[X] = 10000 & (1)\\ \mathbb{V}\mathrm{ar}[X] &= \mathbb{E}[X^2]-\mathbb{E}[X]^2 = 3000-100 = 2900 & (1)\\ \Rightarrow \quad \mathbb{V}\mathrm{ar}[S] &= 1000\cdot\mathbb{V}\mathrm{ar}[X] = 2900000. & (1) \end{aligned}$$

Vjerojatnost nesolventosti:

$$\begin{aligned} \mathbb{P}(S>10000+5000) &= \mathbb{P}\left(\tfrac{S-10000}{\sqrt{2900000}} > \tfrac{50}{\sqrt{290}}\right) = & (1)\\ &= \mathbb{P}(S^* > 2.936) \overset{\text{CGT}}{\approx} 1-\Phi(2.936) = & (1)\\ &= 0.00166. & (1) \end{aligned}$$

(10)

2. (*a*) Momenti:

$$\begin{array}{rcllcll}
M_X(t) & = & \mathbb{E}[e^{tX}] = (1-t/\lambda)^{-\alpha}, & & \mathbb{E}[X^k] & = & M_X^{(k)}(0) & (2)\\
M_X'(t) & = & M_X(t)\cdot\frac{\alpha}{\lambda-t} & \Rightarrow & \mu=\mathbb{E}[X] & = & \frac{\alpha}{\lambda} & (1)\\
M_X''(t) & = & M_X(t)\cdot\frac{\alpha(\alpha+1)}{(\lambda-t)^2} & \Rightarrow & \mathbb{E}[X^2] & = & \frac{\alpha(\alpha+1)}{\lambda^2} & (1)\\
M_X'''(t) & = & M_X(t)\cdot\frac{\alpha(\alpha+1)(\alpha+2)}{(\lambda-t)^3} & \Rightarrow & \mathbb{E}[X^3] & = & \frac{\alpha(\alpha+1)(\alpha+2)}{\lambda^3}. & (1)
\end{array}$$

Standardna devijacija:

$$\sigma=\sqrt{\mathbb{V}\mathrm{ar}[X]}=\sqrt{\frac{\alpha(\alpha+1)}{\lambda^2}-\frac{\alpha^2}{\lambda^2}}=\frac{\sqrt{\alpha}}{\lambda}. \qquad (2)$$

Koeficijent asimetrije:

$$\begin{array}{rcll}
\alpha_3(X) & = & \mathbb{E}[\left(\frac{X-\mu}{\sigma}\right)^3] = & \\
 & = & \frac{1}{\sigma^3}(\mathbb{E}[X^3]-3\mu\mathbb{E}[X^2]+3\mu^2\mathbb{E}[X]-\mu^3) = & (1)\\
 & = & \frac{2}{\sqrt{\alpha}}. & (2)
\end{array}$$

(*b*) F.i.m. od $X+Y$:

$$\begin{array}{rcll}
M_{X+Y}(t) & = & M_X(t)\cdot M_Y(t) \qquad \text{(zbog nezavisnosti)} & (1)\\
 & = & (1-t/\lambda)^{-\alpha}\cdot(1-t/\lambda)^{-\beta} = & (2)\\
 & = & (1-t/\lambda)^{-(\alpha+\beta)} & (1)
\end{array}$$

što je f.i.m. $\Gamma(\alpha+\beta,1/\lambda)$-razdiobe. Dakle, $X+Y$ ima $\Gamma(\alpha+\beta,1/\lambda)$-razdiobu. (1)

(15)

3. (a) Matematičko očekivanje od N:

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

(b) Račun za varijancu od N:

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

(15)

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

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

(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.

5. Neka p_1 označava traženu proporciju kućanstava u populaciji za godinu 1998., a p_2 za godinu 1999.

(a) Procjena od p_1: $\hat{p}_1 = \frac{68}{400} = 0.17$. (1)
95% pouzdan interval za p_1:

$$\hat{p}_1 \pm 1.96 \cdot \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{400}} = 0.17 \pm 0.037 \qquad (2+1)$$

(b) Testiramo:

$$H_0: \quad p_2 = p_1 \qquad (1)$$
$$H_1: \quad p_2 > p_1 \qquad (1)$$

Testna statistika je:

$$Z = \frac{\hat{p}_2 - \hat{p}_1}{\sqrt{\hat{p}(1-\hat{p})(2/400)}} \overset{H_0}{\approx} N(0,1), \qquad (2+1)$$

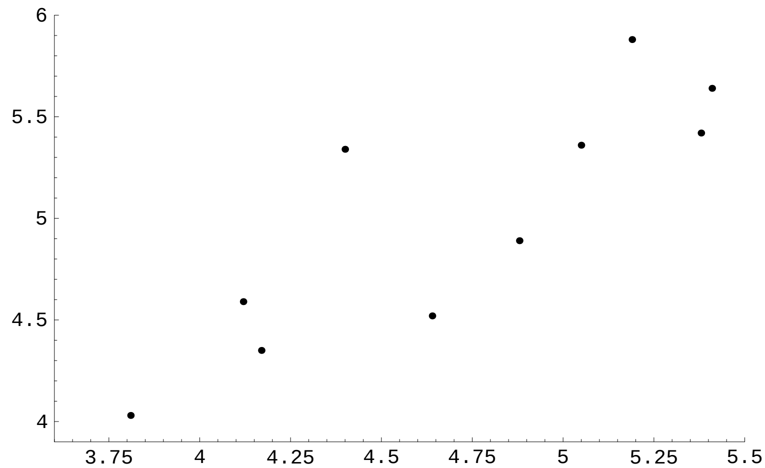
gdje su $\hat{p}_2 = \frac{80}{400} = 0.2$ i $\hat{p} = \frac{68+70}{800} = 0.185$. (1)
Opažena vrijednost od Z je $z = 1.09$. (1)
p-vrijednost:

$$\mathbb{P}(Z > 1.09 | H_0) = 1 - \Phi(1.09) = 0.14, \qquad (3)$$

odakle zaključujemo da nemamo jake dokaze za odbacivanje H_0, iako bi H_0 odbacili u korist alternative uz razinu značajnosti od 5%. (1)

(15)

6. (a) Skica grafa: (1)



Dijagram raspršenja *upućuje* na linearnu vezu između X i Y (1)
uz izrazitu pozitivnu koreliranost. (1)

Opažene vrijednosti pomoćnih statistika:

$$
\begin{aligned}
\bar{x} &= \tfrac{47.12}{10} = 4.712 & \bar{y} &= \tfrac{50.02}{10} = 5.002 \qquad (1)\\
S_{xx} &= \textstyle\sum_{i=1}^{10} x_i^2 - 10 \cdot \bar{x}^2 = & S_{yy} &= \textstyle\sum_{i=1}^{10} y_i^2 - 10 \cdot \bar{y}^2 = \\
&= 2.82596 & &= 3.37956 \qquad (1)\\
S_{xy} &= \textstyle\sum_{i=1}^{02} x_i y_i - 10 \cdot \bar{x}\bar{y} = \\
&= 2.67336 & & \qquad (1)
\end{aligned}
$$

Pearsonov koeficijent korelacije

$$r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}} = 0.86, \qquad (1)$$

pokazuje da je koreliranost između X i Y u danom uzorku pozitivna i visoka. (1)

(b) Parametri pravca:

$$
\begin{aligned}
\hat{\alpha} &= \tfrac{S_{xy}}{S_{xx}} = & \hat{\beta} &= \bar{y} - \hat{\alpha}\bar{x} = \qquad (1)\\
&= 0.946 & &= 0.544 \qquad (1)
\end{aligned}
$$

Regresijski pravac: $y = 0.94 \cdot x + 0.544$ (1)

(c) Koeficijent determinacije:

$$R^2 = \frac{\text{SSR}}{S_{yy}} = r^2 = 0.748 = 74.8\% \qquad (1)$$

pokazuje da je 74.8% ukupne varijabilnosti u Y opisano linearnim modelom. (1)

(d) Procjena za σ^2:

$$\text{SSE} = S_{yy} - \hat{\alpha}^2 S_{xx} = 0.85 \quad \Rightarrow \quad \hat{\sigma}^2 = \frac{\text{SSE}}{n-2} = \frac{0.85}{8} = 0.1063. \qquad (1)$$

Budući da je $(n-2)\hat{\sigma}^2/\sigma^2 \sim \chi^2(n-2)$, (1)
95% pouzdani interval za σ^2 je:

$$\left[\frac{8 \cdot \hat{\sigma}^2}{\chi^2_{0.025}(8)}, \frac{8 \cdot \hat{\sigma}^2}{\chi^2_{0.975}(8)}\right] = [0.0485, 03902]. \qquad (2)$$

(e) 95% pouzdani interval za α: $\hat{\alpha} \pm t_{0.025}(8) \cdot \hat{\sigma} \cdot \sqrt{\frac{1}{S_{xx}}}$. (2)
Iz tablice za t-razdiobu: $t_{0.025}(8) = 2.306$. (1)
Dakle, opaženi 95% pouzdani interval je:
0.946 ± 0.447. (1)

(f) Procjena za $\mathbb{E}[Y|5.0]$: $\hat{\mathbb{E}}[Y|5.0] = 0.946 \cdot 5.0 + 0.544 = 5.274$. (1)
95% pouzdani interval: $\hat{\mathbb{E}}[Y|5.0] \pm t_{0.025}(8) \cdot \hat{\sigma} \cdot \sqrt{\frac{1}{10} + \frac{(5.0-\bar{x})^2}{S_{xx}}}$. (2)
Opaženi 95% pouzdani interval: 5.24 ± 0.270. (1)

(25)