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

ISPIT

VJEROJATNOST I MATEMATIČKA STATISTIKA

8. 5. 2006.

Vrijeme trajanja ispita: 120 minuta
Ukupan broj bodova: 100
Broj zadataka: 5
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. Pretpostavimo da se za skupinu polica osiguranja broj šteta po svakoj od polica može modelirati kao Poissonov proces s istim intenzitetom od λ šteta godišnje, te da su ti procesi nezavisni. Poznato je da je po 1500 polica iz te skupine ukupan broj šteta tijekom prošle godine bio jednak 183.

(a) Metodom maksimalne vjerodostojnosti procijenite parametar λ. (5 bodova)

(b) Procijenite vjerojatnost da po 10 (točno određenih) polica iz te skupine tijekom sljedećih šest mjeseci neće biti šteta. (5 bodova)

(c) Procijenite vjerojatnost da će po 250 (točno određenih) polica iz te skupine tijekom sljedeće godine ukupno biti više od 40 šteta. (5 bodova)

(ukupno 15 bodova)

2. Neka su X i Y nezavisne Poissonove slučajne varijable s očekivanjima λ i μ redom, te $S = X + Y$.

(a) Pokažite da S ima Poissonovu distribuciju s očekivanjem $\lambda + \mu$. (7 bodova)

(b) Pokažite da je uvjetna distribucija od X uz dano $S = s$ binomna. S kojim parametrima? (8 bodova)

(ukupno 15 bodova)

3. Broj šteta N po portfelju istovrsnih nezavisnih polica osiguranja ima Poissonovu razdiobu s parametrom očekivanja μ. Kada se šteta dogodi, njezin iznos X_i ($i = 1, 2, \ldots$) ima gama razdiobu $\Gamma(\alpha, 1/\lambda)$. Varijable X_i, $i = 1, 2, \ldots$, su nezavisne od N. Neka je

$$S = X_1 + X_2 + \cdots + X_N$$

ukupan iznos šteta po tom portfelju. Izračunajte matematičko očekivanje i varijancu od S kao funkciju parametara μ, α i λ. (15 bodova)

4. U tablici se nalaze podaci dobiveni mjerenjem stupnja zagađenja u 20 na slučajan način odabranih područja oko industrijskih zona. Zagađenje se mjerilo prije ugradnje pročistača (X) i godinu dana nakon ugradnje i puštanja počistača u pogon (Y).

područje	1	2	3	4	5	6	7	8	9	10
X	76	78	76	78	84	79	79	81	85	76
Y	74	74	76	79	83	76	76	81	84	76
područje	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
X	78	79	75	83	87	80	78	77	81	77
Y	81	77	74	83	89	78	77	72	79	78

(a) Izračunajte razliku u stupnju zagađenja $X - Y$ za svako područje, zatim tako dobivene podatke prikažite grafički pomoću dijagrama točaka i (na osnovi dijagrama) komentirajte oblik distribucije podataka. (5 bodova)

(b) Izračunajte 95% pouzdan interval za parametar razlike srednjih vrijednosti (očekivanja) od X i Y. (8 bodova)

(c) Sprovedite odgovarajući test da li se ugradnjom počistaća smanjio stupanj zagađenja. (5 bodova)

(b) Uzimajući u obzir zaključak iz (a) dijela zadatka, komentirajte opravdanost primijenjenog testa u (c) i pouzdanog intervala u (b). (2 boda)

(ukupno 20 bodova)

5. U tablici se nalaze frekvencije podataka o oboljelima od bolesti krvožilnog sustava po dobnim skupinama. U istoj tablici su prikazane srednje vrijednosti dobnih skupina (x) i pripadne opažene vrijednosti (y) varijable $Y = \log \frac{\hat{p}(x)}{1-\hat{p}(x)}$, gdje je $\hat{p}(x)$ proporcija bolesnika u dobnoj skupini sa srednjom dobi x.

dobna skupina	bolesni	zdravi	ukupno	x	y
20 - 29	1	9	10	25.0	-2.19722
30 - 34	2	13	15	32.5	-1.87180
35 - 39	3	9	12	37.5	-1.09861
40 - 44	5	10	15	42.5	-0.69315
45 - 49	6	7	13	47.5	-0.15415
50 - 54	5	3	8	52.5	0.51083
55 - 59	13	4	17	57.5	1.17865
60 - 69	8	2	10	65.0	1.38629

(a) Procijenite vjerojatnost oboljenja od bolesti krvožilnog sustava uz pretpostavku da se ta vjerojatnost ne razlikuje po dobnim skupinama. (3 boda)

(b) Sprovedite test homogenosti distribucije oboljenja u odnosu na dobne skupine. (12 bodova)

Pretpostavimo linearni regresijski model $Y = \alpha + \beta x + \varepsilon$ za x i Y. Zbirne statistike za tablične vrijednosti su:

$$\sum_{i=1}^{8} x_i = 360, \qquad \sum_{i=1}^{8} y_i = -2.9392,$$
$$\sum_{i=1}^{8} x_i^2 = 17437.5, \qquad \sum_{i=1}^{8} y_i^2 = 13.615 \qquad \sum_{i=1}^{8} x_i y_i = -9.0429$$

(c) Prikažite podatke u dijagramu raspršenja (skicirajte) i komentirajte prikladnost pretpostavljenog modela za Y u ovisnosti od x. (3 boda)

(d) Prilagodite pravac podacima metodom najmanjih kvadrata. (8 bodova)

(e) Uz pretpostavku da su slučajne pogreške normalno distribuirane i da zadovoljavaju Gauss-Markovljeve uvjete, procijenite njihovu zajedničku varijancu σ^2. (3 boda)

(f) Uz iste pretpostavke kao u (e), konstruirajte i procijenite 99% pouzdan interval za koeficijent smjera regresijskog pravca. (6 bodova)

(ukupno 35 bodova)

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

ISPIT

VJEROJATNOST I MATEMATIČKA STATISTIKA

8. 5. 2006.

RJEŠENJA ZADATAKA

1. Označimo sa $X_{i,t}$ ukupan broj šteta po polici i za razdoblje od t_0 do $t_0 + t$ godina.

(a) Budući da je opažena vrijednost od

$$Y_{1500,1} := \sum_{i=1}^{1500} X_{i,1} \sim P(1500 \cdot \lambda \cdot 1) \quad (1)$$

jednaka 183, vjerodostojnost od λ je:

$$\begin{aligned} L(\lambda) &= \tfrac{(1500\lambda)^{183}}{183!} e^{-1500\lambda} && (1) \\ \ell(\lambda) &= \ln L(\lambda) = 183 \ln \lambda - 1500\lambda + \text{const.} \\ \ell'(\lambda) &= \tfrac{183}{\lambda} - 1500 = 0 && (1) \\ \Rightarrow \quad & \hat{\lambda} = \tfrac{183}{1500} = 0.122. && (2) \end{aligned}$$

(b) Neka je

$$Y_{10,0.5} := \sum_{i=1}^{10} X_{i,0.5} \sim P(10 \cdot \lambda \cdot 0.5) \quad (2)$$

ukupan broj šteta po 10 polica tijekom sljedećih 6 mjeseci. Tada je, uz

$$\lambda \approx 0.122 \quad (1)$$

tražena vjerojatnost jednaka

$$\begin{aligned} \mathbb{P}(Y_{10,0.5} = 0) &= e^{-5\lambda} = && (1) \\ &\approx e^{-5 \cdot 0.122} = e^{-0.61} = 0.543. && (1) \end{aligned}$$

(c) Budući da je ukupan broj šteta po 250 polica tijekom sljedeće godine jednak

$$Y_{250,1} := \sum_{i=1}^{250} X_{i,1} \sim P(250 \cdot \lambda \cdot 1), \quad (1)$$

tražena vjerojatnost je jednaka

$$\begin{aligned} \mathbb{P}(Y_{250,1} > 40) &= 1 - \mathbb{P}(Y_{250,1} \leq 40) \approx && (1) \\ &\approx 1 - \mathbb{P}(Y^*_{250,1} < \tfrac{40.5 - 30.5}{\sqrt{30.5}}) = && (1) \\ &= 1 - \mathbb{P}(Y^*_{250,1} < 1.81) \approx 1 - \Phi(1.81) = && (1) \\ &= 1 - 0.96485 = 0.035. && (1) \end{aligned}$$

2. (*a*) Ako su $X \sim P(\lambda)$ i $Y \sim P(\mu)$, njihove f.i.v. su

$$G_X(t) = e^{\lambda(t-1)}, \quad G_Y(t) = e^{\mu(t-1)} \tag{2}$$

Zbog nezavisnosti X i Y, (1)
f.i.v. od $S = X + Y$ je

$$G_S(t) = G_X(t) \cdot G_Y(t) = e^{(\lambda+\mu)(t-1)}. \tag{2}$$

Budući da je zakon razdiobe brojaće slučajne varijable jednoznačno određen f.i.v., $S \sim P(\lambda + \mu)$. (2)

(*b*) Za $k \in \{0, 1, \ldots, s\}$ vrijedi

$$\begin{aligned}
\mathbb{P}(X = k|S = s) &= \tfrac{\mathbb{P}(X=k, X+Y=s)}{\mathbb{P}(S=s)} = & (2)\\
&= \text{(nezavisnost)} = \tfrac{\mathbb{P}(X=k)\cdot\mathbb{P}(Y=s-k)}{\mathbb{P}(S=s)} = & (1)\\
&= \tfrac{\frac{\lambda^k}{k!}e^{-\lambda}\cdot\frac{\mu^{s-k}}{(s-k)!}e^{-\mu}}{\frac{(\lambda+\mu)^s}{s!}e^{-(\lambda+\mu)}} = & (1)\\
&= \binom{s}{k}\left(\tfrac{\lambda}{\lambda+\mu}\right)^k\left(\tfrac{\mu}{\lambda+\mu}\right)^{s-k}. & (2)
\end{aligned}$$

Dakle, uvjetna razdioba od X uz dano $S = s$ je binomna $b(s, \frac{\lambda}{\lambda+\mu})$. (2)

3. Za slučajnu varijablu S, f.i.m. je

$$m_S(t) = G_N(m_X(t)), \tag{2}$$

gdje su

$$\begin{aligned}
G_N(t) &= e^{\mu(t-1)} & (1)\\
m_X(t) &= \left(\tfrac{\lambda}{\lambda-t}\right)^\alpha, \ \ t < \lambda, & (1)
\end{aligned}$$

f.i.v. Poissonove $P(\mu)$ i f.i.m. $\Gamma(\alpha, 1/\lambda)$ razdiobe. Prema tome,

$$m_S(t) = \exp\left(\mu\left(\left(\tfrac{\lambda}{\lambda-t}\right)^\alpha - 1\right)\right), \ \ t < \lambda. \tag{1}$$

Budući da je

$$\begin{aligned}
m_S'(t) &= m_S(t)\alpha\mu\tfrac{\lambda^\alpha}{(\lambda-t)^{\alpha+1}}, \ \ t < \lambda & (2)\\
m_S''(t) &= m_S'(t)\left(\tfrac{\alpha+1}{\lambda-t} + \tfrac{m_S'(t)}{m_S(t)}\right), \ \ t < \lambda, & (2)
\end{aligned}$$

slijedi da je

$$\begin{aligned}
\mathbb{E}S &= m_S'(0) = \tfrac{\alpha\mu}{\lambda}, & (2)\\
\mathbb{V}\text{ar}S &= \mathbb{E}[S^2] - (\mathbb{E}S)^2 = & (1)\\
&= m_S''(0) - \left(\tfrac{\alpha\mu}{\lambda}\right)^2 = \tfrac{\alpha+1+\alpha\mu}{\lambda^2}\alpha\mu - \left(\tfrac{\alpha\mu}{\lambda}\right)^2 = & (2)\\
&= \tfrac{\alpha\mu(\alpha+1)}{\lambda^2}. & (1)
\end{aligned}$$

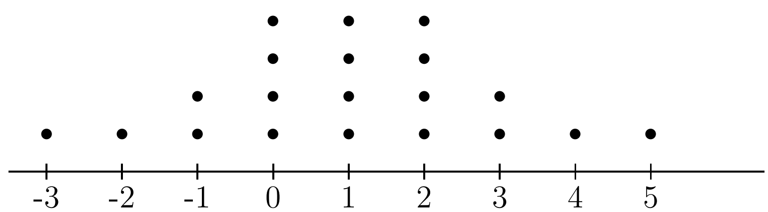
Napomena: Zadatak se može riješiti i tako da se matematičko očekivanje i varijanca od S izračunaju pomoću uvjetovanja na N.

4. Frekvencijaska tablica za $D = X - Y$ (s pomoćnim statistikama):

d_j	f_j	$f_j d_j$	$f_j d_j^2$
-3	1	-3	9
-2	1	-2	4
-1	2	-2	2
0	4	0	0
1	4	4	4
2	4	8	16
3	2	6	18
4	1	4	16
5	1	5	25
Σ	20	20	94

(1+1+1)

(a) Dijagram točaka: (2)



Distribucija podataka je simetrična, gotovo normalna. (2)

(b) Aritmetička sredina $\bar{d}$ i uzoračka standardna devijacija s podataka za D su:

$$\bar{d} = \tfrac{20}{20} = 1 \qquad (1)$$

$$s = \sqrt{\tfrac{94}{19} - \tfrac{20}{19} \cdot 1^2} = 1.97 \approx 2.0. \qquad (1)$$

95% pouzdan interval za $\mu_D := \mathbb{E}D$ je

$$\bar{d} \pm t_{0.025}(19)\tfrac{s}{\sqrt{20}} = 1 \pm 2.09 \cdot 0.44 = 1 \pm 0.9. \qquad (4)$$

(c) Hipoteze: $H_0 : \mu_D = 0$, $H_1 : \mu_D > 0$. (1)
Testna statistika: $T = \frac{\bar{D}}{S}\sqrt{20} \overset{H_0}{\sim} t(19)$. (1)
Opažena vrijednost: $t = \frac{1}{2} \cdot \sqrt{20} = 2.2$. (1)
p-vrijednost: $\mathbb{P}(T > 2.2 | H_0) < 0.025$. (1)
Budući da je p-vrijednost manja od 0.05, u srednjem se stupanj zagađenja značajno smanjio uz razinu značajnosti od 5%. (1)

(d) Budući da je osnovna pretpostavka za primjenu pouzdanog intervala u (b) i t-testa u (c) populacijska normalnost od D, podaci o D ne proturječe toj pretpostavci. (2)

5. (a) $\hat{p} = \dfrac{\text{ukupno bolesnih}}{\text{ukupno ispitanika}} = \frac{43}{100} = 0.43.$ (3)

(b) Frekvencijska tablica:

dobna skupina	opženi bolesni	opaženi zdravi	očekivani bolesni	očekivani zdravi	ukupno
20 - 29	1	9	4.30	5.70	10
30 - 34	2	13	6.45	8.55	15
35 - 39	3	9	5.16	6.84	12
40 - 44	5	10	6.45	8.55	15
45 - 49	6	7	5.59	7.41	13
50 - 54	5	3	3.44	4.56	8
55 - 59	13	4	7.31	9.69	17
60 - 69	8	2	4.30	5.70	10
ukupno	43	57	43	57	100

(2)

Broj stupnjeva slobode: $(8-1)\cdot(2-1) = 7.$ (2)

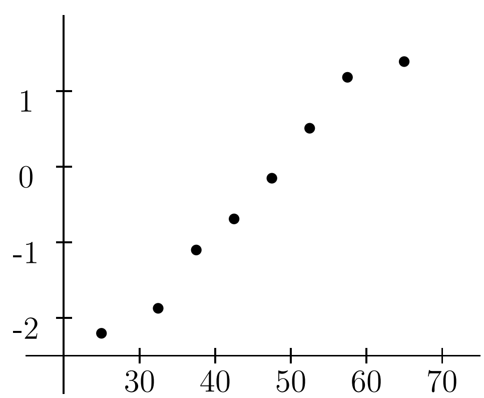
Testna statistika: $H \overset{H_0}{\sim} \chi^2(7).$ (2)

Opažena vrijednost od H: $h = \frac{(1-4.30)^2}{4.30} + \cdots + \frac{(2-5.7)^2}{5.7} = 26.6$ (2)

p-vrijednost: $\mathbb{P}(H > 26.6 | H_0) < 0.005.$ (2)

Podaci jako ne podržavaju H_0, dakle, distribucija bolesnih/zdravi po dobnim skupinama nije homogena. (2)

(b) Dijagram raspršenja: (1)



Osim možda krajnjih točaka, ostale točke se grupiraju oko pravca što upućuje na linearni regresijski model. (2)

(d) Procjena parametara pravca $y = \alpha + \beta x$:

$$\begin{aligned}
\hat{\beta} &= \tfrac{S_{xy}}{S_{xx}} = && (1)\\
&= \frac{\sum_{i=1}^{8} x_i y_i - (\sum_{i=1}^{8} x_i)(\sum_{i=1}^{8} y_i)/8}{\sum_{i=1}^{8} x_i^2 - (\sum_{i=1}^{8} x_i)^2/8} = && (2)\\
&= 0.0995726, && (1)\\
\hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x} = && (1)\\
&= (\textstyle\sum_{i=1}^{8} y_i - \hat{\beta}(\sum_{i=1}^{8} x_i))/8 = && (2)\\
&= -4.848167. && (1)
\end{aligned}$$

(e) $\hat{\sigma}^2 = \frac{SSE}{n-2} = \frac{S_{yy} - \hat{\beta}^2 S_{xx}}{6} = 0.044283$ (1+2)

(f) Budući da je $t_{0.005}(6) = 3.707$ (2)
99% pouzdani interval za β je

$$\hat{\beta} \pm t_{0.005}(6)\hat{\sigma}\sqrt{\tfrac{1}{S_{xx}}} = 0.0996 \pm 0.0222. \qquad (4)$$