**Matematički fakultet u Beogradu**

# Matematička statistika

*Ispit, 6. jul 2015*

Ispit traje 180 minuta. Isključiti mobilne telefone. Nije dozvoljena upotreba beležaka ili drugih pomoćnih sredstava.

---

**1.** Dat je nezavisan uzorak $\mathbf{X} = (X_1, \ldots, X_n)$ iz raspodele sa funkcijom gustine (ili zakonom raspodele) $f(x \,|\, \theta)$, gde je $\theta \in \Theta$ skalarni parametar, a $\Theta$ je interval. Označimo istim slovom $f$ i odgovarajuću funkciju za vektor $\mathbf{X}$: $f(\mathbf{x} \,|\, \theta) = \prod_{j=1}^{n} f(x_j \,|\, \theta)$. Neka je $l(\theta) = l(\mathbf{X} \,|\, \theta)$ logaritam funkcije verodostojnosti vektora $\mathbf{X}$. Uvedimo sledeće oznake:

$$Q(\theta) = l'(\theta), \qquad V_j(\theta) = (\log f(X_j \,|\, \theta))', \ j = 1, \ldots, n, \qquad i(\theta) = \mathrm{E}_\theta V_1^2(\theta),$$

pri čemu su izvodi po parametru $\theta$. Pretpostaviti da važe uslovi regularnosti. Neka je $\hat{\theta}$ ocena maksimalne verodostojnosti za $\theta$.

U ovom zadatku možete da koristite činjenicu da da slučajna promenljiva $\sqrt{n \cdot i(\theta)}(\hat{\theta} - \theta)$ ima asimptotski $\mathcal{N}(0, 1)$ raspodelu kad $n \to +\infty$ (ovo ne treba dokazivati).

**a)** Neka je $\theta = \theta_0$ stvarna vrednost parametra $\theta$ u modelu kao u postavci zadatka. Posmatrajmo
[10] statistike nad uzorkom kao u postavci zadatka, definisane sa

$$D = -2(l(\theta_0) - l(\hat{\theta}), \qquad D^* = -\frac{l''(\theta_0)}{n} \left( (\hat{\theta} - \theta_0)\sqrt{n} \right)^2.$$

Primenom Tejlorovog razvoja funkcije $\theta \mapsto l(\theta)$ za $\theta = \theta_0$ u okolini $\hat{\theta}$ pokazati (zanemarujući ostatak u Tejlorovoj formuli) da statistike $D$ i $D^*$ imaju istu asimptotsku raspodelu kad $n \to +\infty$. Zatim pokazati da je to $\chi^2(1)$ raspodela.

***Uputstvo:*** Za kompletno rešenje videti dokaz Teoreme 8 na str. 34–35, prvi deo teksta predavanja.

**b)** Za fiksirani vektor $\mathbf{X}$ i $c \in (0, 1)$, definišimo oblast verodostojnosti reda $c \in (0, 1)$:
[5]

$$S_c = \Big\{ \theta \in \Theta : \quad \frac{f(\mathbf{X} \,|\, \theta)}{f(\mathbf{X} \,|\, \hat{\theta})} \geq c \Big\}.$$

Pretpostavljajući da je $S_c$ interval za svako $c$, na osnovu rezultata pod a) naći $c$ tako da $S_c$ bude 95% interval poverenja za $\theta$.

***Uputstvo:*** Na strani 36, prvi deo teksta predavanja, primenom rezultata pod a), pokazano je da je $c = 0.147$ u ovom slučaju.

**c)** Izvodi se $n$ Bernulijevih opita sa nepoznatom verovatnoćom uspeha $\theta \in (0, 1)$. Definisati
[10] relevantne slučajne promenljive $X_j$, $j = 1, \ldots, n$, naći ocenu maksimalne verodostojnosti $\hat{\theta}$. Ako se u $n = 100$ opita dogodilo 30 uspeha, naći 95% dvostrani interval poverenja za $\theta$, polazeći od asimptotske raspodele pomenute u postavci zadatka (ne koristeći CGT direktno).

Napisati jednačinu iz koje bi se dobile granice intervala verodostojnosti koji bi bio takođe 95% interval poverenja (koristeći rezultat pod b)). Skicirati grafik funkcije verodostojnosti u zavisnosti od $\theta$ i grafički prikazati interval verodostojnosti.

***Rešenje:*** Definišemo $X_i = 1$ ako se u $i$-tom opitu dogodio uspeh i $X_i = 0$ za slučaj neuspeha. Na taj način dobijamo nezavisan uzorak $\mathbf{X} = (X_1, \ldots, X_n)$ iz Bernulijeve raspodele sa parametrom $\theta$ čiji je zakon raspodele dat sa $f(1 \,|\, \theta) = \theta$ i $f(0 \,|\, \theta) = 1 - \theta$. U ovom modelu, imamo da je

$$f(\mathbf{X} \,|\, \theta) = \theta^S (1 - \theta)^{n - S}, \qquad S = X_1 + \cdots + X_n,$$

$$l(\mathbf{X} \,|\, \theta) = S \cdot \log \theta + (n - S) \cdot \log(1 - \theta),$$

i ocena maksimalne verodostojnosti je $\hat{\theta} = S/n$. Dalje, imamo da je $V_j(\theta) = \log \theta$ za $X = 1$ i $V_j(\theta) = \log(1 - \theta)$ za $X = 0$, odakle se dobija da je

$$i(\theta) = \frac{1}{\theta^2} \cdot \theta + \frac{1}{(1 - \theta)^2} \cdot (1 - \theta) = \frac{1}{\theta(1 - \theta)}.$$

Odavde izlazi da je (prema uputstvu datom u postavci zadatka)

$$\sqrt{n \cdot i(\theta)}(\hat{\theta} - \theta) = \frac{\hat{\theta} - \theta}{\sqrt{\theta(1-\theta)}}\sqrt{n} \sim \mathcal{N}(0,1) \qquad ,$$

asimptotski kad $n \to +\infty$. S obzirom da OMV konvergira u verovatnoći stvarnoj vrednosti parametra (Teorema 1 na strani 20, prvi deo teksta predavanja) možemo $\theta$ u imeniocu da zamenimo sa $\hat{\theta}$ i da zaključimo da je

$$\frac{\hat{\theta} - \theta}{\sqrt{\hat{\theta}(1-\hat{\theta})}}\sqrt{n} \sim \mathcal{N}(0,1) \qquad (n \to +\infty).$$

Odavde se dobija klasičan asimptotski dvostrani interval poverenja reda $1-\alpha$ čije su granice date sa

$$(*) \qquad \hat{\theta} \pm K_{1-\alpha/2}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}},$$

gde je $K_{1-\alpha/2}$ kvantil standardne normalne raspodele reda $1-\alpha$.

Alternativno, s obzirom da se radi o konačnom uzorku, možemo da $\hat{\theta}(1-\hat{\theta})$ zamenimo sa maksimumom tog izraza a to je $1/2$, i da tako dobijemo sigurniji interval sa granicama

$$(**) \qquad \hat{\theta} \pm K_{1-\alpha/2}\frac{1}{2\sqrt{n}}$$

Za $n = 100$ i $S = 30$, $\alpha = 0.05$, imamo da je $\hat{\theta} = 0.3$ i $K = 1.96$, odakle se iz (*) dobija 95% interval poverenja $(0.210, 0.390)$, a iz (**) dobijamo $(0.202, 0.398)$.

Skiciranjem grafika funkcije $f(\theta) = \theta^{30}(1-\theta)^{70}$, $0 < \theta < 1$, možemo se uveriti da je skup tačaka $\frac{f(\theta)}{f(0.3)} \geq c$ neprazan interval za svako $c \in (0,1)$.

Granice intervala verodostojnosti dobijaju se kao rešenja jednačine

$$\theta^{30}(1-\theta)^{70} = 0.147 \cdot \frac{30^{30} \cdot 70^{70}}{100^{100}}$$

ili u logaritamskom obliku

$$30\log\theta + 70\log(1-\theta) = \log(0.147) + 30\log 30 + 70\log 70 - 100\log 100$$

Ova jednačina se može rešiti numerički, dobija se interval verodostojnosti $(0.2161, 0.394)$ (Maple). Primetimo da su prva dva dobijena intervala centrirana u $\hat{\theta} = 0.3$, dok je sredina intervala verodostojnosti u $\theta = 0.3050$.

(Rešenje je znatno opširnije nego što se traži od studenata za maksimalni broj poena)

**2.** U svakom od $n$ nezavisnih eksperimenata mogući ishodi su događaji $A_j$ sa verovatnoćama $p_j > 0$, $j = 1, \ldots, r$, $\sum p_j = 1$. Neka je $N_j$ broj ostvarivanja događaja $A_j$ ($j = 1, \ldots, r$) u $n$ eksperimenata. Raspodela vektora $(N_1, \ldots, N_r)$ zove se multinomna (ili polinomna).

**a)** Uvođenjem slučajnih vektora $\mathbf{X}_i$ dimenzije $r$ , $i = 1, \ldots n$, gde $X_{ij} \in \{0,1\}$ na odgovarajući
[10] način, naći vektor matematičkog očekivanja i matricu kovarijanse za slučajni vektor $(N_1, \ldots, N_{r-1})$.

**b)** Imamo nezavisan uzorak $Y_1, \ldots, Y_n$ sa nepoznatom raspodelom, $Y \sim F$, gde je $F$ funkcija
[5] raspodele. Testira se hipoteza $H_0 : F = F_0$, gde je $F_0$ raspodela bez nepoznatih parametara, protiv komplementarne alternativne hipoteze. Objasniti kako se ovaj test zamenjuje testiranjem hipoteza o multinomnoj raspodeli.

***Uputstvo:*** Pitanja pod a) i b) su obrađena na stranama 31-35 drugi deo teksta predavanja.

**c)** Radi ispitivanja kvaliteta programa za simulaciju standardne normalne raspodele, uzeto je
[10] 50 generisanih vrednosti. Rezultati su prikazani u tabeli:

| | $(-\infty, -1)$ | $(-1, 0)$ | $(0, 1)$ | $(1, +\infty)$ |
|---|---|---|---|---|
| $N_j$ | 6 | 17 | 16 | 11 |

Primenom Hi kvadrat testa sa pragom značajnosti od 0.05 testirati da li su podaci saglasni sa standardnom normalnom raspodelom. Primenom priložene tablice naći interval u kome se nalazi značajnost ($p$-vrednost) podataka.

***Rešenje:*** Broj stepeni slobode je 3, vrednost Pearson-ove hi kvadrat statistike je 1.723. Kako je kvantil reda 0.95 jednak 7.815 ne odbacujemo hipotezu da su podaci saglasni sa normalnom raspodelom. U ovom slučaju, značajnost ($p$-vrednost) je $1 - F(1.723)$, gde je $F$ funkcija raspodele za $\chi^2(3)$. Iz prilozene tablice može se zaključiti samo da je $p$-vrednost u intervalu $(0.05, 0.95)$ - ukoliko ste imali detaljniju tablicu, mogli ste da dobijete uži interval. Numeričkim izračunavanjem dobija se $p$ vrednost od 0.632. Drugim rečima, nultu hipotezu sa datim podacima bismo odbacili samo ako dopustimo da je verovatnoća greške jednaka 0.632 ili veća, odnosno da primenimo pravilo koje bi u preko 60% slučajeva bilo pogrešno.

**3.** U Bajesovskoj teoriji odlučivanja polazi se od sledeće postavke: Imamo skup stanja (parametara) $\Theta$, skup odluka (akcija) $\mathcal{A}$, funkciju gubitka $L(\theta, a)$ i podatke $X$ koji pripadaju nekom prostoru $\mathcal{X}$. Pravilo odlučivanja $\delta(x)$ je merljiva funkcija koja preslikava $\mathcal{X}$ u $\mathcal{A}$.

**a)** Definisati i objasniti pojmove: funkcija rizika, dopustivo pravilo, randomizirano pravilo, [10] Bajesovo pravilo.

***Uputstvo:*** Videti strane 43-45 u tekstu predavanja, deo 2.

**b)** Investitor treba da odluči da li da kupi rizične obveznice. Ako ih kupi, zaradiće 500 evra. U [10] slučaju stečaja kompanije koja je izdala obveznice, investitor gubi početni ulog od 1000 evra. Ako stavi novac u banku, u istom periodu zaradiće sigurnih 300 evra. Investitor ima informaciju da su šanse da dođe do stečaja 10%. Naći Bajesovo pravilo, a zatim naći i minimaks pravilo. Obrazložiti detaljno.

***Uputstvo:*** Ovo je zadatak iz teksta predavanja, deo 2, postavljen na strani 44, sa rešenjem na strani 52.

**Mala tablica normalne raspodele: vrednosti funkcije** $\Phi(x) = P(Z \leq x), X \sim \mathcal{N}(0,1)$
U tablici su date decimale iza nule; na primer, $\Phi(1.62) = 0.9474$.

| $x$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 5000 | 5040 | 5080 | 5120 | 5160 | 5199 | 5239 | 5279 | 5319 | 5359 |
| 0.1 | 5398 | 5438 | 5478 | 5517 | 5557 | 5596 | 5636 | 5675 | 5714 | 5753 |
| 0.2 | 5793 | 5832 | 5871 | 5910 | 5948 | 5987 | 6026 | 6064 | 6103 | 6141 |
| 0.3 | 6179 | 6217 | 6255 | 6293 | 6331 | 6368 | 6406 | 6443 | 6480 | 6517 |
| 0.4 | 6554 | 6591 | 6628 | 6664 | 6700 | 6736 | 6772 | 6808 | 6844 | 6879 |
| 0.5 | 6915 | 6950 | 6985 | 7019 | 7054 | 7088 | 7123 | 7157 | 7190 | 7224 |
| 0.6 | 7257 | 7291 | 7324 | 7357 | 7389 | 7422 | 7454 | 7486 | 7517 | 7549 |
| 0.7 | 7580 | 7611 | 7642 | 7673 | 7704 | 7734 | 7764 | 7794 | 7823 | 7852 |
| 0.8 | 7881 | 7910 | 7939 | 7967 | 7995 | 8023 | 8051 | 8078 | 8106 | 8133 |
| 0.9 | 8159 | 8186 | 8212 | 8238 | 8264 | 8289 | 8315 | 8340 | 8365 | 8389 |
| 1.0 | 8413 | 8438 | 8461 | 8485 | 8508 | 8531 | 8554 | 8577 | 8599 | 8621 |
| 1.6 | 9452 | 9463 | 9474 | 9484 | 9495 | 9505 | 9515 | 9525 | 9535 | 9545 |
| 1.7 | 9554 | 9564 | 9573 | 9582 | 9591 | 9599 | 9608 | 9616 | 9625 | 9633 |
| 1.8 | 9641 | 9649 | 9656 | 9664 | 9671 | 9678 | 9686 | 9693 | 9699 | 9706 |
| 1.9 | 9713 | 9719 | 9726 | 9732 | 9738 | 9744 | 9750 | 9756 | 9761 | 9767 |

**Kvantili** $\varepsilon_u$ **hi kvadrat raspodele** $\chi^2(n)$

| | $u$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $n$ | 0.005 | 0.01 | 0.025 | 0.05 | 0.95 | 0.975 | 0.99 | 0.995 |
| 1 | 0.00004 | 0.00016 | 0.00098 | 0.00393 | 3.841 | 5.024 | 6.635 | 7.879 |
| 2 | 0.010 | 0.0201 | 0.0506 | 0.103 | 5.991 | 7.378 | 9.210 | 10.597 |
| 3 | 0.072 | 0.115 | 0.216 | 0.352 | 7.815 | 9.348 | 11.345 | 12.838 |
| 4 | 0.207 | 0.297 | 0.484 | 0.711 | 9.488 | 11.143 | 13.277 | 14.860 |
| 5 | 0.412 | 0.554 | 0.831 | 1.145 | 11.070 | 12.832 | 13.086 | 16.750 |