Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1913 MATEMATISK STATISTIK FÖR IT OCH ME
ONSDAGEN DEN 1 JUNI 2011 KL 14.00–19.00.

Examinator: Camilla Landén , tel. 790 8466.

Tillåtna hjälpmedel: Formel- och tabellsamling i Matematisk statistik. Räknare. Extrablad om icke-parametriska test finns sist i tentamen.

Införda beteckningar skall förklaras och definieras. Resonemang och uträkningar skall vara så utförliga och väl motiverade att de är lätta att följa. Numeriska svar skall anges med minst två siffrors noggrannhet. Tentamen består av 6 uppgifter. Varje korrekt lösning ger 10 poäng. Gränsen för godkänt är preliminärt 24 poäng. Möjlighet att komplettera ges för de tentander med 22–23 poäng. Det ankommer på dig själv att ta reda på om du har rätt att komplettera.

Uppgift 1

a) En förenklad typ av spelautomat innehåller två roterande hjul som oberoende av varandra och oberoende varje gång hamnar i något av lägena A, B eller C. Sannolikheten att hamna i de tre lägena är 0.1, 0.3 respektive 0.6 för vardera hjulet. Automaten ger vinst om de båda hjulen kommer i samma läge.
Beräkna sannolikheten att det vid spel två gånger i automaten blir förlust vid minst ett tillfälle. (5 p)

b) Låt den stokastiska variablen X vara $N(5, 4)$. Beräkna $P(X \leq 8 \mid X \geq 3)$. (5 p)

Uppgift 2

Ett bibliotek har i genomsnitt 78 besökare en vanlig fredag. Bestäm (ev approximativt) sannolikheten att det en vanlig fredag kommer fler än 90 besökare. Ange vilka antaganden du gör. (10 p)

Uppgift 3

En matematiskt bevandrad taxichaufför har funderat lite över sin situation och kommit fram till att en körnings längd i kilometer kan betraktas som en stokastisk variabel med täthetsfunktionen

$$f(x) = \begin{cases} \dfrac{1}{7000}\left(\dfrac{1136}{3} + 10x - x^2\right), & 1 \leq x \leq 25, \\ 0, & \text{för övrigt} \end{cases}$$

Beräkna (approximativt) sannolikheten att han efter 100 körningar har kört mer än 1100 km. (10 p)

Uppgift 4

Man ville undersöka om livslängden för en viss typ av glödlampa kunde anses vara exponentialfördelad. Man lät därför hundra lampor brinna tills de gick sönder. Man fann då att den genomsnittliga livslängden var 906 timmar, 12 glödlampor hade en livslängd på mer än 1500 timmar, 37 hade en livslängd på mellan 750 och 1500 timmar och 51 hade en livslängd på färre än 750 timmar. Är det rimligt att utifrån dessa data anta att livslängden är exponentialfördelad? (10 p)

Uppgift 5

I en undersökning om belysningens effekt på den mänskliga förmågan att genomföra koncentrationskrävande arbetsuppgifter gjordes följande experiment. Nio personer med normal, men varierande, synförmåga fick som sin uppgift att så snabbt som möjligt dra en tunn tråd genom tio nålsögon. Detta gjordes både mot svart bakgrund under lägre ljusstyrka och mot vit bakgrund under högre ljusstyrka. Man mätte tiden att fullborda uppgiften. Resultatet blev:

Person nr	1	2	3	4	5	6	7	8	9	
Högre ljusstyrka	25.85	28.84	32.05	25.74	20.89	41.05	25.01	24.96	27.47	(sek)
Lägre ljusstyrka	18.23	20.84	22.96	19.68	19.50	24.98	16.61	16.07	24.59	(sek)

Vi antar som modell att observationerna är oberoende, normalfördelade med samma varians. Vi antar också att de förväntade tidsskillnaderna är desamma ($= \Delta$) för varje person.

a) Gör ett 99.9%:igt konfidensintervall för denna skillnad Δ. (3 p)

b) Kan vi på basis av dessa data anse att belysningen har en effekt på tiden att genomföra den givna uppgiften? Ge Ditt svar genom att utföra ett statistiskt test på nivån 0.1% av hypotesen

$$H_0 : \Delta = 0 \quad \text{(sek)}$$

mot

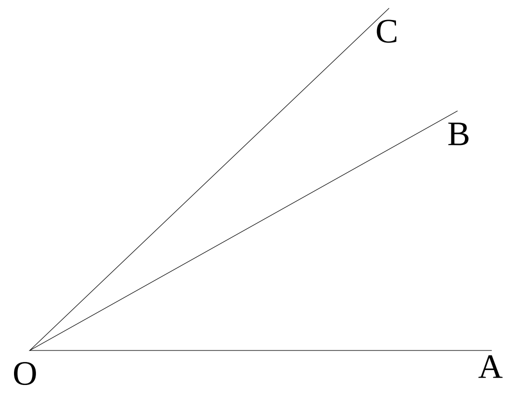
$$H_1 : \Delta \neq 0 \quad \text{(sek)}.$$

Det skall klart framgå om hypotesen förkastas eller ej. (2 p)

c) Antag att man är osäker på normalfördelningsantagandet. Genom att låta ytterligare 10 personer genomföra testet fann man att tidsdifferensen mellan högre och lägre ljusstyrka var positiv i 17 fall av 19 (ingen differens var lika med noll) och rangsumman för högre ljusstyrka var 543. Gör ett icke-parametriskt test, utgående från data för 19 personer, av om det på nivå 0.1% finns någon systematisk skillnad i förväntad tid vid olika ljusstyrkor. (5 p)

Uppgift 6

Från en punkt O har man mätt vinklarna AOB, BOC och AOC och därvid erhållit värdena: $43.2°, 29.2°$ och $70.1°$. Punkternas läge framgår av figuren.



De tre mätningarna kan uppfattas som utfall av oberoende normalfördelade stokastiska variabler vilkas väntevärden överensstämmer med de sanna vinklarna och vilkas standardavvikelse är $2°$. Detta svarar emot att mätutrustningen inte har något systematiskt fel och att den är beprövad så att man vet standardavvikelsen.

Skatta vinkeln AOB med MK-metoden samt bestäm ett 95% konfidensintervall för vinkeln AOB utgående från MK-skattningen. (10 p)

Observera: För att lösningen ska ge poäng måste alla mätningarna utnyttjas.

Icke-Parametriska Test

- **Teckentestet.** Låt $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ vara ett stickprov i par. Bilda differenserna mellan x-observationerna och y-observationerna och låt t vara antalet gånger differensen är strikt positiv. Då är t en observation av T som är $Bin(n_z, 0.5)$, under förutsättning att x_i och y_i är observationer ur samma fördelning. Med n_z avses antalet differenser som inte är noll.

- **Wilcoxons Rangsummetest.** Låt $x_1, x_2, \ldots, x_{n_1}$ och $y_1, y_2, \ldots, y_{n_2}$ vara två oberoende stickprov. Låt r vara rangsumman för x-observationerna, då x-observationerna och y-observationerna storleksordnats. Då gäller att r är en observation av R för vilken

$$E(R) = n_1 \frac{n_1 + n_2 + 1}{2} \quad \text{och} \quad V(R) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12},$$

under förutsättning att x-observationerna och y-observationerna kommer från samma fördelning. Förutom för små n_1 och n_2 är R approximativt normalfördelad.

Avd. Matematisk statistik

KTH Matematik

LÖSNINGAR TILL
TENTAMEN I SF1913 MATEMATISK STATISTIK FÖR IT OCH ME
ONSDAGEN DEN 1 JUNI 2011 KL 14.00–19.00.

Uppgift 1

a) Betrakta först en spelomgång i automaten.

$$P(\text{vinst}) = P(\text{bägge visar } A \,\cup\, \text{bägge visar } B \,\cup\, \text{bägge visar } C) = \{\text{disjunkta händelser}\}$$

$$= P(\text{bägge visar } A) + P(\text{bägge visar } B) + P(\text{bägge visar } C) = \{\text{hjulen oberoende}\}$$

$$= P(\text{hjul 1 visar } A)P(\text{hjul 2 visar } A) + P(\text{hjul 1 visar } B)P(\text{hjul 2 visar } B)$$

$$+P(\text{hjul 1 visar } C)P(\text{hjul 2 visar } C) = 0.1^2 + 0.3^2 + 0.6^2 = 0.46.$$

Vi får därför för bägge spelomgångar:

$$P(\text{ingen vinst vid någon omgång}) = 1 - P(\text{vinst i bägge omgångar})$$

$$= 1 - P(\text{vinst i första omgången } \cap \text{ vinst i andra omgången}) = \{\text{omgångarna oberonde}\}$$

$$= 1 - P(\text{vinst i första omgången})P(\text{vinst i andra omgången}) = 1 - 0.46^2 = 0.7884.$$

b)

$$P(X \leq 8 \mid X \geq 3) = \frac{P(X \leq 8 \cap X \geq 3)}{P(X \geq 3)} = \frac{P(3 \leq X \leq 8)}{P(X \geq 3)} = \frac{P(X \leq 8) - P(X \leq 3)}{P(X \geq 3)}$$

$$= \left(\Phi\left(\frac{8-5}{4}\right) - \Phi\left(\frac{3-5}{4}\right)\right)/\Phi\left(\frac{3-5}{4}\right) = \left(\Phi(0.75) - (1-\Phi(0.5))\right)/\Phi(0.5) \approx 0.6723.$$

Uppgift 2

Det är rimligt att anta att $X =$ antalet besökare är Po(78). Alltså är den sökta sannolikheten

$$P(X > 90) = 1 - P(X \leq 90) = 1 - \text{Poissoncdf}(78, 90) \approx 0.0810 = 8.10\%.$$

(Detta fås med TI:s miniräknare.)
Man kan som alternativ använda normalapproximation: X är approximativt $N(78, \sqrt{78})$, så

$$P(X > 90) = 1 - P(X \leq 90.5) = P\left(\frac{X-78}{\sqrt{78}} \leq \frac{90.5-78}{\sqrt{78}}\right) \approx 1 - \Phi(1.415) \approx 0.0785 = 7.85\%$$

(med halvkorrektion).

Uppgift 3

Vi bestämmer först väntevärde och varians för $X =$ en körnings längd (km, resp km^2.)

$$\begin{aligned} E[X] &= \int_1^{25} x f(x)\, dx \approx 10.367, \\ E[X^2] &= \int_1^{25} x^2 f(x)\, dx \approx 142.219, \\ V(X) &\approx 142.219 - 10.367^2 \approx 34.744. \end{aligned}$$

Vi använder nu CGS och approximerar totala körlängden av 100 körningar till $Y \in \mathrm{N}(1036.7, \sqrt{3474.4})$. Den efterfrågade sannolikheten är alltså

$$P(Y > 1100) = \text{normalcdf}(1100, E99, 1036.7, \sqrt{3474.4}) \approx 0.14 = 14\%.$$

Detta fås med TI:s miniräknare. Alternativt kan man naturligtvis använda tabellen för normalfördelningen.

Uppgift 4

Låt $X \in \mathrm{Exp}(\lambda)$. Då är $E[X] = 1/\lambda$, så en skattning av λ under hypotesen att $X =$ "livslängden för en glödlampa är $\mathrm{Exp}(\lambda)$" är $\lambda = 1/906$. Vi får då att $P(X > 1500) = e^{-1500/906} \approx 0.1910$, $P(750 < X \leq 1500) = e^{-750/906} - e^{-1500/906} \approx 0.2460$, samt $P(X \leq 1500) = 1 - e^{-750/906} \approx 0.5630$. Förväntat antal brinntider i dessa intervall är alltså 19.1, 24.6, resp 56.3. Vi gör ett χ^2-test:

$$Q = \frac{(12 - 19.1)^2}{19.1} + \frac{(37 - 24.6)^2}{24.6} + \frac{(51 - 56.3)^2}{56.3} \approx 9.39.$$

Antalet frihetsgrader är 1, eftersom vi har en skattad parameter. Med hjälp av miniräknaren får vi att $\chi^2\text{cdf}(9.39, E99, 1) \approx 0.0022$. Vi förkastar alltså hypotesen med felrisk 0.22% att brinntiderna är exponentialfördelade.
Alternativt kan vi naturligtvis se i tabellen över χ^2-kvantilerna att eftersom $9.39 > 7.88$ förkastar vi hypotesen på nivån 0.5%.

Uppgift 5

(a) Detta är hypotesprövning *med stickprov i par*. Vi betecknar tiderna under lägre ljusstyrka med x_i och tiderna under högre ljusstyrka med y_i. Vi bildar skillnaden mellan dessa som $z_i = y_i - x_i$, $i = 1, \ldots, 9$. Vi ser z_i som utfall av oberoende normalfördelade stokastiska variabler $Z_i \in N(\Delta, \sigma)$, $i = 1, \ldots, 9$, respektive. Vi bildar ett konfidensintervall för Δ med t -metoden.
Det aritmetiska medelvärdet är en punktskattning av Δ med

$$\overline{z} = \frac{1}{9} \sum_{i=1}^{9} z_i = 7.60.$$

Vi skattar σ med

$$s_z = \sqrt{\frac{1}{8} \sum_{i=1}^{9} (z_i - \overline{z})^2} = 4.1779.$$

(Vid behov kan räknehjälp fås ur formelsamlingen

$$\sum_{i=1}^{9}(z_i-\overline{z})^2=\sum_{i=1}^{9}z_i^2-9\cdot\overline{z}^2=139.64.)$$

Det sökta konfidensintervallet är av formen

$$\overline{z}\pm t_{0.0005}(8)\cdot\frac{s_z}{\sqrt{9}},$$

där $t_{0.0005}(8)$ är 0.0005-kvantilen för t-fördelningen med 8 frihetsgrader. Tabellslagning ger

$$t_{0.0005}(8)=5.04.$$

Med insättning fås

$$7.6\pm 5.04\cdot\frac{4.18}{\sqrt{9}}=[0.58,14.6].$$

SVAR: 99.9 %:igt konfidensintervall för Δ är $[0.58, 14.6]$.

(b) Värdet $\Delta=0$ ingår inte i det konfidensintervall, som tagits fram i uppgiftens (a)-del. Därav dras slutsatsen att nollhypotesen H_0 måste förkastas.
SVAR: Nollhypotesen $H_0:\Delta=0$ (sek) förkastas på signifikansnivån 0.001.

(c) Teckentest. Om de förväntade tiderna var lika stora borde T=antalet positiva differenser vara Bin$(19,1/2)$-fördelad. Vi har fått $t=17$ som utfall och beräknar sannolikheten att få lika extremt (eller extremare) utfall och gör detta ”dubbelsidigt”.

$$P=P(T\leq 2 \text{ eller } T\geq 17)=2P(T\leq 2)=2\cdot 0.00036=0.00072.$$

Eftersom denna sannolikhet $\leq 0.1\%$ kan
H_0 : ”de förväntade tiderna vid olika ljusstyrka är lika långa” förkastas på nivån 0.1%.

Uppgift 6

(a) Låt θ_1 vara vinkeln AOB och θ_2 vinkeln BOC. Detta innebär att $\theta_1+\theta_2$ är vinkeln AOC. Låt vidare x,y,z vara resp. mätvärden, dvs $x=43.2$, $y=29.2$ och $z=70.1$. Dessa uppfattas som utfall av oberoende stokastiska variabler X,Y,Z som är $N(\theta_1,2)$-, $N(\theta_2,2)$- och $N(\theta_1+\theta_2,2)$-fördelade. För att beräkna MK-skattningen av θ_1 betraktar vi

$$Q(\theta_1,\theta_2)=(x-\theta_1)^2+(y-\theta_2)^2+(z-\theta_1-\theta_2)^2.$$

Derivation ger

$$\frac{\partial Q(\theta_1,\theta_2)}{\partial\theta_1}=-2(x-\theta_1)-2(z-\theta_1-\theta_2)=-2(x+z-2\theta_1-\theta_2)$$

och

$$\frac{\partial Q(\theta_1,\theta_2)}{\partial\theta_2}=-2(y-\theta_2)-2(z-\theta_1-\theta_2)=-2(y+z-\theta_1-2\theta_2).$$

Sätter vi derivatorna $=0$ fås således ekvationssystemet

$$2\theta_1 + \theta_2 = x + z \tag{1}$$
$$\theta_1 + 2\theta_2 = y + z. \tag{2}$$

Bildar vi $2 \cdot (1) - (2)$ fås $3\theta_1 = 2x - y + z = 2x + (z - y)$ och således är MK-skattningen av θ_1 given av

$$\theta_1^* = \frac{2x + (z - y)}{3} = \frac{2 \cdot 43.2 + (70.1 - 29.2)}{3} = \frac{127.3}{3} = \underline{42.4^\circ}.$$

(b) Vi betraktar nu den till θ_1^* hörande stickprovsvariabeln $\theta_1^*(X, Y, Z)$. För denna gäller

$$E(\theta_1^*(X, Y, Z)) = E\left(\frac{2X + (Z - Y)}{3}\right) = \frac{2\theta_1 + (\theta_1 + \theta_2 - \theta_2)}{3} = \theta_1,$$

dvs skattningen är väntevärdesriktig. Vidare har vi

$$V(\theta_1^*(X, Y, Z)) = V\left(\frac{2X + (Z - Y)}{3}\right) = \frac{1}{9}(4V(X) + V(Z) + V(Y)) = \frac{(4 + 1 + 1) \cdot 2^2}{9} = \frac{24}{9}.$$

Ett 95% konfidensintervall fås således av

$$I_{\theta_1} = \theta_1^* \pm \lambda_{0.025}\sqrt{\frac{24}{9}} = 42.4 \pm 1.96 \cdot 1.63 = \underline{42.4^\circ \pm 3.2^\circ}.$$