Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1903 (f d 5B2501) SANNOLIKHETSLÄRA OCH STATISTIK FÖR 3-ÅRIG MEDIA MÅNDAGEN DEN 13 AUGUSTI 2012 KL 08.00–13.00.

Examinator: Gunnar Englund, tel. 073 321 37 45

Tillåtna hjälpmedel: Läroboken, Formel- och tabellsamling i Matematisk statistik. Räknare.

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 5 uppgifter. Varje korrekt lösning ger 10 poäng. Gränsen för godkänt är preliminärt 20 poäng. Möjlighet att komplettera ges för de tentander med 18–19 poäng. Tid och plats för komplettering kommer att anges på kursens hemsida. Det ankommer på dig själv att ta reda på om du har rätt att komplettera.

Tentamen kommer att finnas tillgänglig på elevexpeditionen sju veckor efter skrivningstillfället.

Uppgift 1

a) För en viss region gäller följande: 20% av befolkningen är rökare. Sannolikheten att en rökare får lungcancer är 10 gånger så hög som att en icke-rökare får lungcancer. Antag att sannolikheten att få lungcancer för en slumpmässigt vald person är 0.006. Vad är då sannolikheten att en person får lungcancer *givet* att personen i fråga är rökare? (5 p)

b) Beräkna väntevärde och varians för produkten

$$X_1 \cdot X_2 \cdots X_{10}$$

där alla X_i:na är oberoende och $R(0,2)$. (5 p)

Uppgift 2

Den stokastiska variabeln X har en täthet $f_X(x) = cx^5$ för $2 \leq x \leq 8$.

1. Bestäm normeringskonstanten c så att $f_X(x)$ är en sannolikhetstäthet. (3 p)

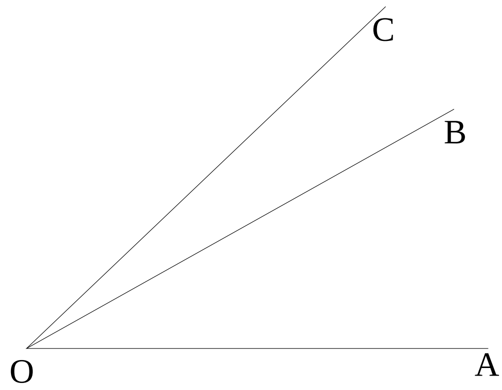
2. Beräkna $P(6 < X < 12)$. (2 p)

3. Låt $Y = \frac{1}{X^2}$. Bestäm $E(Y)$ (2 p)

   och $V(Y)$. (3 p)

Uppgift 3

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.

Uppgift 4

Vid undersökning av livslängden hos en viss sorts komponenter sattes 5 stycken under provning och livslängderna mättes. Resultat (veckor):
9.85 9.93 10.12 9.94 10.06
Man antar att livslängden av en komponent är $N(m, \sigma)$.

a) Ge ett 95 % konfidensintervall för medellivslängden m. (5 p)

b) Skatta på lämpligt sätt den s.k. L_{10}-livslängden, d.v.s. den livslängd som 10% av komponenterna kommer att underskrida. (5 p)

Uppgift 5

I SIFO-undersökningen från februari 2002 av partisympatierna fick Socialdemokraterna 43.8% av de 2500 tillfrågade som uppgivit partisympati medan motsvarande siffra var 39.8% av 4000 som angivit partisympati i december 2001.

a) Skatta förändringen för Socialdemokraterna samt medelfelet för denna skattning. (2 p)

b) Ge ett (approximativt) 95%-igt tvåsidigt konfidensintervall för förändringen för Socialdemokraterna mellan de två undersökningarna. (6 p)

c) Kan man anse att Socialdemokraternas uppgång är statistiskt säkerställd? Besvara frågan genom att utföra en (dubbelsidig) hypotesprövning på (den approximativa) nivån 5%. Formulera noggrant hypoteserna Du testar! (2 p)

Avd. Matematisk statistik

KTH Matematik

LÖSNINGAR TILL
TENTAMEN I SF1903 (f d 5B2501) SANNOLIKHETSLÄRA OCH STATISTIK FÖR 3-ÅRIG MEDIA MÅNDAGEN DEN 18 AUGUSTI 2008 KL 08.00–13.00.

Uppgift 1

a) Inför beteckningarna $R =$ ”personen är rökare” resp. $C =$ ”personen får cancer”.
Enligt texten är $P(R) = 0.2$, $P(C) = 0.006$ och $P(C|R) = 10 \cdot P(C|R^*)$.
Lagen om total sannolikhet ger

$$P(C) = P(C|R)P(R) + P(C|R^*)P(R^*)$$

dvs

$$0.006 = P(C|R) \cdot 0.2 + \frac{1}{10}P(C|R) \cdot (1 - 0.2) = 0.28 \cdot P(C|R)$$

och således

$$P(C|R) = \frac{0.006}{0.28} \approx 0.021.$$

b)

$$\begin{aligned} E(X_1X_2\cdots X_{10}) = &\ \{\text{oberoende}\}\ = E(X_1)E(X_2)\cdots E(X_{10}) \\ &= \{X_i \in R(0,2),\ \text{se FS}\ \} = 1^{10} = 1. \end{aligned}$$

$$\begin{aligned} V(X_1X_2\cdots X_{10}) &= E\left((X_1X_2\cdots X_{10})^2\right) - \left(E(X_1)E(X_2)\cdots E(X_{10})\right)^2 = \{\text{oberoende}\} \\ &= E(X_1^2)E(X_2^2)\cdots E(X_{10}^2) - 1^2 \\ &= \left(V(X_1) + E(X_1)^2\right)\left(V(X_2) + E(X_2)^2\right)\cdots\left(V(X_{10}) + E(X_{10})^2\right) - 1 \\ &= \{X_i \in R(0,2), \text{ se FS}\} = \left(\frac{1}{3} + 1\right)^{10} - 1 = \left(\frac{4}{3}\right)^{10} - 1 \approx 16.76. \end{aligned}$$

Uppgift 2

1. Eftersom den totala sannolikheten är 1 så skall

$$1 = \int_{-\infty}^{\infty} f_X(x)\,dx = \int_2^8 cx^5\,dx = \frac{c}{6}\left[x^6\right]_2^8 = 43680c$$

vilket ger $\underline{c = 1/43680}$.

2. Med det valet på c är

$$P(6 < X < 12) = \int_6^{12} f_X(x)\,dx = \int_6^8 cx^5\,dx = \frac{c}{6}\left[x^6\right]_6^8 = 35915c = \underline{0.8222}.$$

3. Med $Y = 1/X^2$ så är

$$\begin{aligned} E(Y) &= E\left(\frac{1}{X^2}\right) = \int_{-\infty}^{\infty} \frac{1}{x^2} f_X(x)\,dx = \int_2^8 \frac{1}{x^2} cx^5\,dx = \frac{c}{4}\left[x^4\right]_2^8 = 1020c \\ &= \underline{0.023352}. \end{aligned}$$

Vidare så är

$$\begin{aligned} E\left(Y^2\right) &= E\left(\frac{1}{X^4}\right) = \int_{-\infty}^{\infty} \frac{1}{x^4} f_X(x)\,dx = \int_2^8 \frac{1}{x^4} cx^5\,dx = \frac{c}{2}\left[x^2\right]_2^8 = 30c \\ &= \underline{0.00068681} \end{aligned}$$

så $V(Y) = E(Y^2) - (E(Y))^2 = \underline{0.00014151}$.

Uppgift 3

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

$$2\theta_1 + \theta_2 = x + z \qquad (1)$$
$$\theta_1 + 2\theta_2 = y + z. \qquad (2)$$

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

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

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

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

Uppgift 4

a) Vi får att medelvärdet $\bar{x} = 9.98$ och stickprovsvariansen är $s^2 = 0.01175$. Ett 95% konfidensintervall ges av $\bar{x} \pm t_{0.025}(4)s/\sqrt{5} = \underline{9.98 \pm 0.135}$.

b) Om X är livslängden har vi att $P(X < L_{10}) = 0.1$. Men $P(X < L_{10}) = \Phi\left(\frac{L_{10}-m}{\sigma}\right)$ vilket ger $\frac{L_{10}-m}{\sigma} = -\lambda_{0.1} = -1.2816$. Av detta erhålles $L_{10} = m - 1.2816\sigma$. m och σ skattas med $\bar{x}$ respektive s och vi erhåller skattningen $\underline{L_{10}^* = 9.84}$.

Uppgift 5

Antalet Socialdemokrater i de två undersökningarna är utfall av X som är $\text{Bin}(2500, p_1)$ respektive Y som är $\text{Bin}(4000, p_2)$ där p_1 och p_2 är deras verkliga andelar i februari respektive december. Vi skattar dessa med $p_1^* = 0.438$ respektive $p_2^* = 0.398$ och dessa kan ses som utfall av $X/2500$ respektive $Y/4000$.

a) Skattningen av förändringen är naturligtvis $0.438 - 0.398 = 0.04 = 4.0\%$.

Medelfel är ju en skattning av standardavvikelsen för skattningen. Vi skattar $p_1 - p_2$ med $X/2500 - Y/4000$ som har variansen

$$\frac{V(X)}{2500^2} + \frac{V(Y)}{4000^2} = \frac{p_1(1-p_1)}{2500} + \frac{p_2(1-p_2)}{4000}.$$

Medelfelet får vi om vi ersätter de okända p_1 och p_2 med skattningarna $p_1^* = 0.438$ respektive $p_2^* = 0.398$ och erhåller medelfelet

$$\sqrt{\frac{p_1^*(1-p_1^*)}{2500} + \frac{p_2^*(1-p_2^*)}{4000}} \approx \sqrt{\frac{0.438\cdot(1-0.438)}{2500} + \frac{0.398\cdot(1-0.398)}{4000}} \approx \underline{0.0126}$$

b) $X/2500$ och $Y/4000$ är approximativt normalfördelade $N(p_1, \sqrt{p_1(1-p_1)/2500})$ respektive $N(p_1, \sqrt{p_2(1-p_2)/4000})$ där det är tillåtet att göra normalapproximation ty (t ex) $2500p_1(1-p_1) \approx 2500\cdot 0.438(1-0.438) \geq 10$ med mycket bred marginal. Vi vill beräkna ett (approximativt) 95%-igt konfidensintervall för $p_1 - p_2$ och får med approximativa metoden enligt FS 11.3 intervallet

$$p_1^* - p_2^* \pm \lambda_{0.025}\sqrt{\frac{p_1^*(1-p_1^*)}{2500} + \frac{p_2^*(1-p_2^*)}{4000}} \approx$$

$$\approx 0.438 - 0.398 \pm 1.9600\cdot 0.0126 = 0.040 \pm 0.025 = \underline{(1.5\%, 6.5\%)}.$$

c) Vi testar nollhypotesen $H_0 : p_1 = p_2$ mot alternativhypotesen $H_1 : p_1 \neq p_2$ på nivån 5% och förkastar H_0 eftersom konfidensintervallet i a-delen ej innehöll 0. Uppgången är alltså statistiskt säkerställd!