



Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1901,SF1907,SF1908,SF1913 SANNOLIKHETSTEORI OCH STATISTIK
TORSDAGEN DEN 7:E JUNI 2012 KL 14.00–19.00.

Examinator: Gunnar Englund, tel. 073 321 37 45

Tillåtna hjälpmedel: Formel- och tabellsamling i Matematisk statistik, Mathematics Handbook (Beta), 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 6 uppgifter. Varje korrekt lösning ger 10 poäng. Gränsen för godkänt är preliminärt 24 poäng. Möjlighet att komplettera ges för tentander med 22–23 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 vara rättad inom tre arbetsveckor från skrivningstillfället och kommer att finnas tillgänglig på studentexpeditionen minst sju veckor efter skrivningstillfället.

Uppgift 1

$X_1, X_2, \dots, X_{100}$ är oberoende stokastiska variabler, alla med en tthet given av

$$f(x) = \frac{2}{x^2}, \quad 1 < x < 2.$$

Låt $Y = X_1 + X_2 + \dots + X_{100}$. Bestäm approximativt sannolikheten $P(Y > 140)$. (10 p)

Uppgift 2

I en byggnad sitter ett brandlarm monterat. Under en tidsperiod är sannolikheten att larmet går 3%. Man vet att 98% av alla larm är falsklarm, dvs brandlarmet signalerar trots att det inte är någon eldsvåda. Eldsvådor uppstår under samma tidsperiod med sannolikhet 0.001.

a) Bestäm sannolikheten att larmet går om en eldsvåda bryter ut. (5 p)

b) Brandlarmet fungerar inte med sannolikhet 0.04 (ingen strömförsörjning, trasig siren, trasig detektor etc.). Ett trasigt brandlarm kan inte larma. Eldsvådor uppstår oberoende av om brandlarmet fungerar eller ej.

Bestäm sannolikheten att larmet går om en eldsvåda bryter ut betingat att larmet fungerar. (5 p)

Uppgift 3

Enligt SIFOs opinionsmätning i maj 2011 var opinionen så här (%):

V	S	MP	C	FP	M	KD	SD	Övr
4,6	33,4	8,8	5,4	6,6	30,4	3,6	6,8	0.4

1910 personer var tillfrågade.

I januari var motsvarande siffror

V	S	MP	C	FP	M	KD	SD	Övr
4,3	27,8	10,3	4,7	7,1	33,9	4,4	6,9	0.6

1890 personer var tillfrågade.

Vi är intresserade av opinionen för MP. Kan vi utifrån dessa data dra någon slutsats om huruvida sympatierna för MP har ökat, minskat eller är oförändrade från januari till maj? Använd felrisken 5%. (10 p)

Uppgift 4

Vissa problem i partikelfysik ger upphov till följande sannolikhetsstäthet

$$f(x) = \begin{cases} \frac{1}{2}(1 + \theta x) & -1 \leq x \leq 1, \\ 0 & \text{för övrigt,} \end{cases}$$

där $-1 \leq \theta \leq 1$ är en okänd parameter.

a) Varför måste vi kräva att olikheterna $-1 \leq \theta \leq 1$ skall gälla ? (2 p)

b) Antag att $X_1, \dots, X_n$ är oberoende stokastiska variabler som alla har denna fördelning. En i statistikfrågor välbevandlad vän till Dig föreslår en skattningsfunktion för θ av formen

$$\theta^* = c \cdot \frac{1}{n} (X_1 + \dots + X_n).$$

Hur bör konstanten c väljas för att θ_{obs}^* blir en väntevärdesriktig punktskattning ? (3 p)

c) Antag att $X_1, \dots, X_n$ är oberoende stokastiska variabler som alla har denna fördelning. Bestäm MK-skattningen θ_{MK}^* (Minsta kvadrat-skattningen) av θ . (5 p)

Uppgift 5

Låt $x_1, \dots, x_n$ vara oberoende stickprov ur täthetsfunktionen $f_X(x)$ given av

$$f_X(x) = \begin{cases} \frac{\theta}{2\sqrt{x}} e^{-\theta\sqrt{x}} & x > 0 \\ 0 & \text{för övrigt,} \end{cases}$$

där $\theta > 0$ är en okänd parameter.

a) Härled formeln för maximum likelihoodskattningen θ_{obs}^* av θ på basis av $x_1, \dots, x_n$. (8 p)

b) Med $n = 4$ har vi stickproven $x_1 = 6.2, x_2 = 7.0, x_3 = 2.5, x_4 = 4.2$. Beräkna värdet på maximum likelihoodskattningen θ_{obs}^* för dessa stickprov. (2 p)

Uppgift 6

Två mäklare, A och B, bedömde marknadspriset på 12 stycken villor, 1,...,12. Resultatet blev så här (priser i miljoner kronor)

villa	1	2	3	4	5	6	7	8	9	10	11	12
A:	2,4	3,0	4,1	1,95	0,7	5,0	4,5	2,8	1,6	3,1	1,9	2,5
B:	2,8	3,4	4,3	2,2	1,0	4,4	4,6	3,0	1,5	3,3	2,3	2,8

Avgör, med en lämplig metod, på konfidens-nivån 95 %, om det finns någon systematisk skillnad i de två mäklarnas bedömningar. (10 p)



Avd. Matematisk statistik

KTH Matematik

LÖSNINGAR TILL

TENTAMEN I SF1901, SF1907, SF1908 samt SF1913 SANNOLIKHETSLÄRA OCH STATISTIK,
TORSDAGEN DEN 7:E JUNI 2012

Uppgift 1

Eftersom $X_1, \dots, X_{100}$ är oberoende, likafördelade och många till antalet kan man använda centrala gränsvärdessatsen (CGS) för att approximera fördelningen för deras summa. Enligt CGS är

$$Y \stackrel{\text{appr}}{\in} N(n\mu, \sigma\sqrt{n})$$

där $\mu = E(X_1)$, $\sigma = D(X_1)$ och $n = 100$ här. Beräkna μ och σ .

$$\mu = E(X_1) = \int_{-\infty}^{\infty} x f(x) dx = \int_1^2 x \frac{2}{x^2} dx = \left[2 \ln |x| \right]_1^2 = 2 \ln 2.$$

$$E(X_1^2) = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_1^2 x^2 \frac{2}{x^2} dx = \int_1^2 2 dx = 2.$$

$$V(X_1) = E(X_1^2) - (E(X_1))^2 = 2 - (2 \ln 2)^2 = 0.0782 \Rightarrow \sigma = D(X_1) = \sqrt{0.0782} = 0.2796$$

Alltså är fördelningen till Y approximativt $N(100 \cdot 2 \ln 2, 0.2796\sqrt{100})$ eller $N(138.63, 2.796)$. Den sökta sannolikheten är därmed

$$P(Y > 140) = 1 - P(Y \leq 140) = 1 - \Phi\left(\frac{140 - 138.63}{2.796}\right) = 1 - \Phi(0.490) = 1 - 0.6879 = 0.3121$$

Den sökta sannolikheten är 0.31.

Uppgift 2

Låt L vara händelsen att larmet går och E händelsen att en eldsvåda bryter ut. Då är

$$P(L|E) = \frac{P(L \cap E)}{P(E)} = \frac{P(E|L) P(L)}{P(E)} = \frac{0.02 \cdot 0.03}{0.001} = \underline{0.600}$$

Låt B vara händelsen att brandlarmet fungerar. Eftersom $L \subseteq B$ är $P(L \cap B \cap E) = P(L \cap E) = 0.0006$ enligt tidigare. Vidare så är $P(B \cap E) = P(B) P(E) = 0.96 \cdot 0.001$ på grund av oberoendet så

$$P(L|B \cap E) = \frac{P(L \cap B \cap E)}{P(E \cap B)} = \frac{0.0006}{0.96 \cdot 0.001} = \underline{0.625}$$

Uppgift 3

De enda data som är relevanta är givetvis de för MP. I maj-mätningen var 168 av 1910 personer MP-sympatisörer, i januari-mätningen 195 av 1890. Det gäller alltså att 168 är en observation

ur en $\text{Bin}(1910, p_1)$ -fördelning, och att 195 är en observation ur en $\text{Bin}(1890, p_2)$ -fördelning, och nollhypotesen är nu att $P_1 = p_2$.

En möjlig lösning är att göra ett konfidensintervall för skillnaden $p_1 - p_2$ på vanligt sätt och se om noll ligger i intervallet. Intervallet blir $-0.0339 \leq p_1 - p_2 \leq 0.00348$, så vi kan inte dra någon slutsats om förändringen av sympatier för MP med felrisken 5%.

En annan möjlighet är att göra ett homogenitetstest. Detta ger $Q = 2.546$ och antalet frihetsgrader är 1. χ^2 -kvantilen för 5% är $3.84 > 2.546$, så resultatet är detsamma som ovan.

Uppgift 4

a) I sannolikhetskalkylen krävs av en sannolikhetstäthet att för alla x

$$f(x) \geq 0.$$

För tätheten i denna uppgift innebär detta att

$$\frac{1}{2}(1 + \theta x) \geq 0 \Leftrightarrow 1 + \theta x \geq 0 \Leftrightarrow \theta x \geq -1.$$

Om alltså $\theta x \geq -1$ gäller för alla $x \in [-1, 1]$, så fås för $x = 1$ att $\theta \cdot 1 \geq -1$, dvs. $\theta \geq -1$. Nu gäller igen för alla $x \in [-1, 1]$ att

$$\theta x \geq -1 \Leftrightarrow -\theta x \leq 1$$

och med $x = -1$ fås $-\theta \cdot (-1) \leq 1$ dvs. $\theta \leq 1$. Således har vi sett att $-1 \leq \theta \leq 1$ för att $f(x) \geq 0$ skall gälla.

b) Definitionsmässigt är θ_{obs}^* en väntevärdesriktig punktskattning, om

$$E[\theta^*] = \theta.$$

Med avseende på den kunniga vännens förslag noterar vi att

$$\begin{aligned} E[\theta^*] &= c \cdot \frac{1}{n} (E[X_1] + \dots + E[X_n]) \\ &= c \cdot \frac{1}{n} (E[X_1] + \dots + E[X_n]). \end{aligned}$$

Eftersom $X_1, \dots, X_n$ har samma fördelning, gäller $E[X_1] = \dots = E[X_n]$. Då fås för $i = 1, 2, \dots, n$

$$\begin{aligned} E(X_i) &= \int_{-\infty}^{\infty} x \cdot f(x) dx = \frac{1}{2} \int_{-1}^1 x \cdot (1 + \theta x) dx \\ &= \frac{1}{2} \left[\int_{-1}^1 x \cdot (1 + \theta x) dx \right] = \frac{1}{2} \left[\int_{-1}^1 x dx + \theta \int_{-1}^1 x^2 dx \right] \\ &= \frac{1}{2} \left[\left[\frac{x^2}{2} \right]_{-1}^1 + \theta \left[\frac{x^3}{3} \right]_{-1}^1 \right] = \frac{1}{2} \left[\left(\frac{1}{2} - \frac{1}{2} \right) + \theta \left(\frac{1}{3} - \left(-\frac{1}{3} \right) \right) \right] = \frac{1}{2} \cdot \theta \frac{2}{3} = \frac{\theta}{3}. \end{aligned}$$

Med andra ord

$$E[\theta^*] = c \cdot \frac{1}{n} \cdot n \cdot \frac{\theta}{3} = c \cdot \frac{\theta}{3}.$$

Således gäller $E[\theta^*] = \theta$ om och endast om $c = 3$.

SVAR: c = 3.

c) MK-skattningen fås som det varde på θ som minimerar

$$Q(\theta) = \sum_{i=1}^n (x_i - E[X_i])^2.$$

Från uppgiftens del b) har vi att $E[X_i] = \frac{\theta}{3}$.

Således vill vi minimera

$$Q(\theta) = \sum_{i=1}^n \left(x_i - \frac{\theta}{3}\right)^2.$$

Vi deriverar med avseende på θ och sätter derivatan till noll,

$$Q'(\theta) = -2 \cdot \frac{1}{3} \cdot \sum_{i=1}^n \left(x_i - \frac{\theta}{3}\right) = 0,$$

vilket ger

$$\sum_{i=1}^n x_i - \sum_{i=1}^n \frac{\theta}{3} = 0,$$

eller

$$\frac{n \cdot \theta}{3} = \sum_{i=1}^n x_i.$$

Detta ger

$$\theta_{obs}^* = \frac{3}{n} \sum_{i=1}^n x_i,$$

vilket är den väntevärdesriktiga punktskattningen i del b).

SVAR: $\theta_{obs}^* = \frac{3}{n} \sum_{i=1}^n x_i$.

Uppgift 5

För de oberoende stickproven $x_1, \dots, x_n$ definieras likelihoodfunktionen $L(\theta)$ som

$$\begin{aligned} L(\theta) &= f_X(x_1) \cdot f_X(x_2) \cdots f_X(x_n) \\ &= \frac{\theta}{2\sqrt{x_1}} e^{-\theta\sqrt{x_1}} \cdot \frac{\theta}{2\sqrt{x_2}} e^{-\theta\sqrt{x_2}} \cdots \frac{\theta}{2\sqrt{x_n}} e^{-\theta\sqrt{x_n}}. \\ &= \frac{\theta^n}{2^n \sqrt{x_1} \cdot \sqrt{x_2} \cdots \sqrt{x_n}} e^{-\theta \cdot (\sqrt{x_1} + \dots + \sqrt{x_n})} \\ &= \frac{\theta^n}{2^n \prod_{i=1}^n \sqrt{x_i}} e^{-\theta \cdot (\sum_{i=1}^n \sqrt{x_i})}. \end{aligned}$$

Det är praktiskt att maximera $L(\theta)$ genom att ekvivalent maximera dess naturliga logaritm $\ln L(\theta)$. Vi logaritmerar och erhåller

$$\ln L(\theta) = n \ln \theta - n \ln 2 - \ln \prod_{i=1}^n \sqrt{x_i} - \theta \sum_{i=1}^n \sqrt{x_i}.$$

Låt oss även observera att $\ln \theta$ är definierad p.g.a att $\theta > 0$. Vi deriverar med avseende på θ och får

$$\frac{d}{d\theta} \ln L(\theta) = \frac{n}{\theta} - \sum_{i=1}^n \sqrt{x_i}.$$

Vi sätter $\frac{d}{d\theta} \ln L(\theta) = 0$, vilket ger

$$\frac{n}{\theta} - \sum_{i=1}^n \sqrt{x_i} = 0 \Leftrightarrow \frac{n}{\theta} = \sum_{i=1}^n \sqrt{x_i}.$$

Om vi löser den sista ekvationen m.a.p. θ får vi maximum likelihoodskattningen θ_{obs}^* av θ på basis av $x_1, \dots, x_n$ som

$$\theta_{\text{obs}}^* = \frac{n}{\sum_{i=1}^n \sqrt{x_i}}.$$

Eftersom alla $x_i > 0$, är summan i nämnaren > 0 .

SVAR a): $\theta_{\text{obs}}^* = \frac{n}{\sum_{i=1}^n \sqrt{x_i}}.$

b) Insättning av stickproven $x_1 = 6.2, x_2 = 7.0, x_3 = 2.5, x_4 = 4.2$ och $n = 4$ i den i del a) ovan härledda formeln ger

$$\begin{aligned} \theta_{\text{obs}}^* &= \frac{4}{\sum_{i=1}^4 \sqrt{x_i}} = \frac{4}{\sqrt{6.2} + \sqrt{7.0} + \sqrt{2.5} + \sqrt{4.2}} \\ &= \frac{4}{2.49 + 2.6548 + 1.5811 + 2.0494} = \frac{4}{8.7663} = 0.4563. \end{aligned}$$

SVAR b): $\underline{\theta_{\text{obs}}^* = 0.46}.$

Uppgift 6

Observationer i par. Den bästa modellen är att modellera skillnaderna $\ln(X_B) - \ln(x_A)$ som observationer från en normalfördelning. Men vi utgår från modellen att $X_A - X_B$ är observationer från en normalfördelning och gör ett t-test för att denna fördelnings väntevärde är noll.

Vi får att medelvärdet av de parvisa skillnaderna är $\bar{x} = 0.1708$ och stickprovs-standardavvikelse $s = 0.2816$. Alltså är teststorheten $t = \frac{0.1708\sqrt{12}}{0.2816} = 2.101$. antalet frihetsgrader är 11. Eftersom t-kvantilen 0.025 för 11 frihetsgrader är $2.2 > 2.101$ kan vi inte ur dessa data avgöra om det finns någon systematisk skillnad i de två mäklarnas bedömningar.