KUNGL TEKNISKA HÖGSKOLAN

Matematiska institutionen
avd matematisk statistik

TENTAMEN I 5B1501 SANNOLIKHETSTEORI OCH STATISTIK I
TORSDAGEN DEN 26 AUGUSTI 1999 KL 8.00 – 13.00.

Examinator för M: Timo Koski, tel. 790 8466
Examinator för övriga linjer: Jan Grandell, tel. 790 7136

Tillåtna hjälpmedel: Formel- och tabellsamling i matematisk statistik. L. Råde - B. Westergren: Mathematics Handbook for Science and Engineering. Räknare.

Införda beteckningar skall förklaras och definieras. Resonemang och uträkningar skall vara så utförliga 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 25 poäng.

Resultatet anslås senast torsdagen den 16 september 1999 på Matematisk statistiks anslagstavla i entréplanet, Lindstedtsvägen 25, rakt fram innanför porten.

Ingen kontrollskrivning får tillgodoräknas vid denna tentamen.

Eventuella klagomål på bedömningen av tentamen skall lämnas till Elevexpeditionen senast sju veckor efter den dag då tentamen ägt rum.

Lycka till!

\- -

Uppgift 1

Waloddi Weibull, som var professor vid KTH, gjorde viktiga insatser bl.a. vid studiet av utmattningsfenomen av maskinelement. Efter honom är den s.k. Weibullfördelningen uppkallad; en stokastisk variabel X säges vara $\mathrm{W}(\lambda, \beta)$-fördelad, där λ och $\beta > 0$, om

$$F_X(x) = \begin{cases} 1 - e^{-(\lambda x)^\beta}, & \text{för } x \geq 0, \\ 0, & \text{annars.} \end{cases}$$

Ett system består av två komponenter med livslängder som är $\mathrm{W}(\lambda, \beta)$-fördelade. Livslängderna är vidare oberoende stokastiska variabler. Systemet fungerar så länge som båda komponenterna fungerar.

Bestäm fördelningsfunktionen för livslängden hos systemet. (10 p)

Uppgift 2

I en ergonomisk undersökning av ett experimentellt tangentbord för arbetsstationer har man rekryterat 31 mycket rutinerade operatörer och för var och en av dem provat sig fram till den bekvämaste höjden x_i, $i = 1, \ldots, 31$, (mätt i cm från golvnivån) för armstödet. Den genomsnittliga observerade höjden $\frac{1}{31}(x_1 + \ldots + x_{31})$ blev lika med 80.0 cm.
Vi modellerar de 31 observationerna som oberoende stickprov från en normalfördelning $N(m, 2.0)$. Testa med ett tvåsidigt test med risknivån 5 % hypotesen

$$H_0 : m = 81.5 \quad \text{(cm)}$$

mot

$$H_1 : m \neq 81.5 \quad \text{(cm)}.$$

Det skall klart framgå om hypotesen H_0 förkastas eller ej. (10 p)

Uppgift 3

Nederbördsmängden i mm under ett dygn kan betraktas som en stokastisk variabel X. Det gäller att

$$X = YZ$$

där Y och Z är oberoende stokastiska variabler med följande egenskaper

$$P(Y = 0) = 1 - p,\ P(Y = 1) = p,\ 0 \leq p \leq 1 \text{ och } P(Z > x) = e^{-x/m},\ x > 0,\ m > 0.$$

Detta betyder att sannolikheten att det regnar är p, och om det regnar är nederbördsmängden en Exp(m)-fördelad stokastisk variabel.

(a) Beräkna $E(X)$ och $V(X)$ som funktioner av p och m. (5 p)

(b) Antag att nederbördsmängderna under olika dygn är oberoende likafördelade stokastiska variabler som alla har samma fördelning som X ovan. Vi har ur nederbördsdata erhållit värdena $E(X) = 1.5$ och $V(X) = 6.75$ (svarande mot $m = 3$ och $p = 1/2$ i a-uppgiften. Detta kan vara en bra kontroll för Dig som räknat a-uppgiften). Beräkna med lämplig och välmotiverad approximation sannolikheten att den totala nederbördsmängden under ett år (=365 dagar) överstiger 500 mm. (5 p)

Uppgift 4

I vissa problem i partikelfysik möter vi följande sannolikhetstäthet

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

där $-1 \leq \theta \leq 1$ är en okänd parameter. Låt $x_1, \ldots, x_n$ vara utfall av oberoende stokastiska variabler $X_1, \ldots, X_n$, respektive, som alla är fördelade enligt denna fördelning.

(a) Härled *minsta-kvadrat-skattningen* (MK-skattningen) av parametern θ. (5 p)

(b) Utred om MK-skattningen är väntevärdesriktig. (5 p)

Uppgift 5

En grupp av matematiklärare vid en större teknisk högskola har utvecklat ett nytt diagnostiskt prov i matematik, som de s.k. 'nollorna' (dvs. de nyintagna studenterna) skall ta. Man tror att det diagnostiska provet är av sådan svårighetsgrad att 40% av alla 'nollor' kommer att få underkänt i provet, 40% kommer att få betyget 3 och resten kommer att få 4 eller 5.
Man gjorde ett slumpmässigt urval av 200 'nollor', som skrev det diagnostiska testet. Det visade sig att 60 'nollor' fick betyget 3 och 45 'nollor' fick 4 eller 5 och de övriga blev underkända.
Testa med χ^2-metoden på den approximativa risknivån 5 % om den tänkta betygsfördelningen verkar vara förenlig med ovanstående resultat. (10 p)

Uppgift 6

Datachefen vid en tärningsfabrik meddelar nöjt att han nu "säkrat" den datoriserade maskinen för automatisk målning av tärningarna inför 2000-talet. Detta må vara sant, men maskinen blev felprogrammerad på så sätt att den nu på varje tärningssida målar något av talen $1, \ldots, 6$ helt slumpmässigt, dvs den målar något av talen med samma sannolikhet, och oberoende av vad den målat på tärningens övriga sidor. Datachefen tar nu en tärning från den felprogramerade maskinen.

(a) Vad är sannolikheten för att han får en sexa i ett kast med tärningen. (3 p)

(b) Vad är sannolikheten för att han får två sexor i två kast med tärningen. (7 p)

KUNGL TEKNISKA HÖGSKOLAN
Matematiska institutionen
avd matematisk statistik

LÖSNINGAR TILL
TENTAMEN I 5B1501 SANNOLIKHETSTEORI OCH STATISTIK I
TORSDAGEN DEN 26 AUGUSTI 1999 KL 8.00 – 13.00.

Uppgift 1

Låt X_1 och X_2 vara livslängderna hos komponenterna och låt Y vara livslängden hos systemet. Då gäller, för $y \geq 0$

$$F_Y(y) = 1 - P(Y > y) = 1 - P(X_1 > y, X_2 > y) = 1 - P(X_k > y)^2 = 1 - \big(1 - F_X(y)\big)^2$$

$$= 1 - \big(1 - (1 - e^{-(\lambda y)^\beta})\big)^2 = 1 - \big(e^{-(\lambda y)^\beta}\big)^2 = \underline{1 - e^{-2(\lambda y)^\beta}},$$

dvs Y är W$(2^{1/\beta}\lambda, \beta)$-fördelad.

Uppgift 2

Eftersom standardavvikelsen $\sigma = 2.0$ är känd, bildar vi ett tvåsidigt konfidensintervall med konfidensgrad 95 % enligt λ -metoden. Detta intervall är

$$\frac{1}{31}(x_1 + \ldots + x_{31}) \pm \lambda_{0.025} \cdot \frac{\sigma}{\sqrt{31}} =$$

$$= 80.0 \pm 1.96 \cdot \frac{2.0}{\sqrt{31}} = 80.0 \pm 0.7 = [79.3, 80.7].$$

Eftersom $m = 81.5$ inte ligger inom detta intervall kommer H_0 att <u>förkastas</u>.

Uppgift 3

(a) $E(X) = E(YZ) =$ (p g a oberoendet) $= E(Y)E(Z)$. Vi ser att $E(Z) = m$ eftersom Z är Exp(m) och $E(Y) = (1-p)\cdot 0 + p \cdot 1 = p$ vilket ger $E(X) = \underline{mp}$. Vidare är $E(X^2) = E(Y^2 Z^2) =$ (p g a oberoendet) $= E(Y^2)E(Z^2)$. Eftersom $V(Z) = E(Z^2) - (E(Z))^2$ och $Y^2 = Y$ (som ger $E(Y^2) = E(Y)$) får vi därför $E(X^2) = p \cdot (m^2 + m^2) = 2m^2 p$ som ger $V(X) = E(X^2) - (E(X))^2 = 2m^2 p - (mp)^2 = \underline{m^2 p(2-p)}$.

(b) Låt $X_j =$ nederbörden dag nr j, $j = 1, 2, \ldots, 365$ och $S_{365} = X_1 + X_2 + \cdots + X_{365}$. Enligt centrala gränsvärdessatsen är S_{365} approximativt

$$N\big(E(S_{365}), D(S_{365})\big)\text{-fördelad} = N\big(365 \cdot 1.5, \sqrt{365 \cdot 6.75}\,\big)\text{-fördelad}$$

och således får vi den sökta sannolikheten

$$P(S_{365} > 500) = P\left(\frac{S_{365} - 365 \cdot 1.5}{\sqrt{365 \cdot 6.75}} > \frac{500 - 365 \cdot 1.5}{\sqrt{365 \cdot 6.75}}\right)$$

$$\approx 1 - \Phi\left(\frac{500 - 365 \cdot 1.5}{\sqrt{365 \cdot 6.75}}\right) \approx 1 - \Phi(-0.956) = \Phi(0.956) \approx \underline{0.83}.$$

Uppgift 4

(a) MK-skattningen fås som det varde på θ som minimerar

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

Vi behöver ta fram värdet på $E(X_i)$ som funktion av θ. Här gäller

$$E(X_i) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx = \frac{1}{2}\int_{-1}^{1} x \cdot (1 + \theta x)\, dx$$

och

$$\frac{1}{2}\left[\int_{-1}^{1} x \cdot (1 + \theta x)\, dx\right] = \frac{1}{2}\left[\int_{-1}^{1} x dx + \theta \int_{-1}^{1} x^2 dx\right]$$

$$= \frac{1}{2}\left[\left[\frac{x^2}{2}\right]_{-1}^{1} + \theta\left[\frac{x^3}{3}\right]_{-1}^{1}\right] = \frac{1}{2}\left[\left(\frac{1}{2} - \frac{1}{2}\right) + \theta\left(\frac{1}{3} - \left(-\frac{1}{3}\right)\right)\right] = \frac{1}{2} \cdot \theta \frac{2}{3} = \frac{\theta}{3}.$$

Således har vi fått

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

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

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

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

(b) Väntevärdesriktighet avgörs genom att kontrollera om väntevärdet $E(\theta^*)$ är lika med θ eller inte. Vi betraktar θ^* som funktion av de stokastiska variablerna $X_1, \ldots, X_n$ dvs.

$$\theta^* = \frac{3}{n}\sum_{i=1}^{n} X_i.$$

Från ovan erhåller vi

$$E(\theta^*) = \frac{3}{n}\sum_{i=1}^{n} E(X_i) = \frac{3}{n}\sum_{i=1}^{n}\frac{\theta}{3} = \theta.$$

Detta visar att MK-skattningen är väntevärdesriktig.

Uppgift 5

Vi vill testa H_0 formulerad av matematiklärarna, dvs.

$$H_0 : P(\text{underkänt}) = 0.40, P(\text{betyg 3}) = 0.40, P(\text{betyg 4 eller 5}) = 0.20.$$

χ^2-metoden ger teststorheten

$$Q = \frac{(95 - 200 \cdot 0.4)^2}{200 \cdot 0.4} + \frac{(60 - 200 \cdot 0.4)^2}{200 \cdot 0.4} + \frac{(45 - 200 \cdot 0.2)^2}{200 \cdot 0.2} = 8.4375$$

Detta värde på teststorheten jämförs med en tröskel eller en 0.05-kvantil för χ^2-fördelningen med $3 - 1 = 2$ frihetsgrader, vilket ger

$$\chi^2_{0.05}(2) = 5.99.$$

Eftersom $Q = 8.4375 > \chi^2_{0.05}(2) = 5.99$, kommer vi att förkasta H_0.

Uppgift 6

Uppgiften går säkert att lösa på många olika sätt. Vi ska ge två lösningar av olika karaktär. På en tentamen ska dock *aldrig* alternativa lösningar ges.

"Betingningslösning":

Låt X vara antalet sexor på på tärningen. X är Bin$(6, 1/6)$-fördelad, vilket innebär att $E(X) = 6 \cdot \frac{1}{6} = 1$ och att $V(X) = 6 \cdot \frac{1}{6}\frac{5}{6} = 5/6$. För senare bruk noterar vi att $E(X^2) = V(X) + E(X)^2 = \frac{5}{6} + 1 = \frac{11}{6}$.

(a)

$$P(\text{sexa}) = \sum_{k=1}^{6} P(\text{sexa} \mid X = k)P(X = k) = \sum_{k=1}^{6} \frac{k}{6} P(X = k) = \frac{E(X)}{6} = \underline{\frac{1}{6}}.$$

(b)

$$P(\text{två sexor}) = \sum_{k=1}^{6} P(\text{två sexor} \mid X = k)P(X = k) = \sum_{k=1}^{6} \frac{k^2}{36} P(X = k) = \frac{E(X^2)}{36} = \frac{11}{6} \cdot \frac{1}{36} = \underline{\frac{11}{216}}.$$

"Resonemangslösning":

(a) Oavsett vilken sida av tärningen som kommer upp så är sannolikheten för en sexa 1/6.

(b) Vi skiljer på fallen då samma sida kommer upp i de bägge kasten, vilket sker med sannolikhet 1/6, och då olika sidor kommer upp. I det sista fallet är resultaten oberoende. Detta ger

$$P(\text{två sexor})$$

$$= P(\text{två sexor} \mid \text{samma sida}) \cdot P(\text{samma sida}) + P(\text{två sexor} \mid \text{olika sidor}) \cdot P(\text{olika sidor})$$

$$= \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{36} \cdot \frac{5}{6} = \left(1 + \frac{5}{6}\right)\frac{1}{36} = \frac{11}{6} \cdot \frac{1}{36} = \frac{11}{216}.$$