Lösningar till Tentamen i Matematisk Statistik, 7.5p

Distanskurs
28 maj, 2010 kl. 9.00–13.00

Maxpoäng: 30p. **Betygsgränser**: 12p: betyg G, 21p: betyg VG. **Hjälpmedel**: Miniräknare samt formelsamling som medföljer tentamenstexten. **Kursansvarig**: Eric Järpe (035-16 76 53, 0702-822 844).

1. Låt $\xi \in N(2, 1)$. Beräkna a sådant att $P(2\xi - 1 \geq a) = 0.99$. (2p)

   Lösning: $0.99 = P(2\xi - 1 \geq a) = 1 - P(2\xi - 1 \leq a) = 1 - P(\xi \leq \frac{1}{2}(a+1)) = = 1 - \Phi\left(\frac{\frac{1}{2}(a+1)-2}{1}\right) \Rightarrow \Phi(\frac{1}{2}(a+1) - 2) = 0.01 \Rightarrow \Phi(-\frac{1}{2}(a+1) + 2) = 0.99 \Rightarrow$ $\Rightarrow -\frac{1}{2}(a+1) + 2 = 2.326348 \Rightarrow a = -2 \cdot (2.326348 - 2) - 1 = -1.652696.$ □

2. Antalet uppgifter vid ett centralprov i matematik är 10. Antag att alla uppgifter löses med sannolikhet 0.6 oberoende av varandra och att 100 elever skriver provet. Vad är chansen att någon har alla rätt? (2p)

   Lösning: $P(\text{en elev har alla rätt}) = 0.6^{10} = 0.006046618$. Därmed är
   $P(\text{någon av de 100 elverna har alla rätt}) = 1-P(\text{ingen av de 100 eleverna har alla rätt})$
   $= 1-P(\text{elev 1 har inte alla rätt})P(\text{elev 2 har inte alla rätt}) \cdots P(\text{elev 100 har inte alla rätt})$
   $= 1-(1-P(\text{elev 1 har alla rätt}))(1-P(\text{elev 2 har alla rätt})) \cdots (1-P(\text{elev 100 har alla rätt}))$
   $= 1 - (1 - 0.6^{10})(1 - 0.6^{10}) \cdots (1 - 0.6^{10}) = 1 - (1 - 0.006046618)^{100} = 0.4547426.$ □

3. Antag att en elektrisk krets har två seriekopplade komponenter med livslängd X_1 resp. X_2 år som är oberoende och exponentialfördelade med parametrar $\lambda_1 = 0.08$ resp. $\lambda_2 = 0.05$. Vad är sanolikheten att kretsen fungerar efter 10 år? (3p)

   Lösning:
   $P(\text{Kretsen fungerar efter 10 år}) = P(X_1 > 10, X_2 > 10) = (1 - P(X_1 \leq 10)) \cdot (1 - P(X_2 \leq 10)) = (1 - (1 - e^{-0.08 \cdot 10}))(1 - (1 - e^{-0.05 \cdot 10})) = e^{-0.8-0.5} = 0.2725318.$ □

4. För att planera personaltätheten observerar matvarukedjan MINI hur många som passerar genom entrén mellan kl. 11 och 13 och finner under en vecka

måndag	tisdag	onsdag	torsdag	fredag	lördag	söndag
115	87	73	92	91	101	89

(a) Bilda ett 95% konfidensintervall för det förväntade antalet kunder mellan kl. 11 och 13. (3p)

(b) Gör ett hypotestest på 1% signifikansnivå av om det förväntade antalet kunder mellan kl. 11 och 13 är fler än 85. (2p)

Lösning:

(a) $\bar{x} = 92.57143$, $s^2 = 167.2857$ och $t_{0.05/2}(7-1) = 2.4469$ varmed konfidensintervallet fås av att beräkna $92.57143 \pm 2.4469\sqrt{\frac{167.2857}{7}}$ och avrunda undre gränsen nedåt och övre gränsen uppåt, dvs vi får $(80.609, 104.534)$.

(b) $\begin{cases} H_0 : \mu = 85 \\ H_1 : \mu > 85 \end{cases}$ $\quad \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{92.571428 - 85}{\sqrt{167.28573/7}} = 1.5488$ och $t_{0.01}(6) = 3.1427$.
Eftersom $1.5488 \not> 3.1427$ kan man ej förkasta H_0, dvs man kan ej visa att det förväntade antalet kunder mellan 11 och 13 är fler än 85 på 1% signifikansnivå.

$\square$

5. Ett pappersbruk tillverkar kollegieblock som är exakt 1 cm tjocka.

(a) Vad är approximativt sannolikheten att ett block innehåller minst 100 ark, om varje ark i blocket har en tjocklek som är fördelad $R(0.06, 0.13)$ mm oberoende av varandra? (3p)

Antag nu att man räknar antalet papper i 8 block vilket ger

Block	1	2	3	4	5	6	7	8
Antal ark	108	79	118	103	110	81	90	100

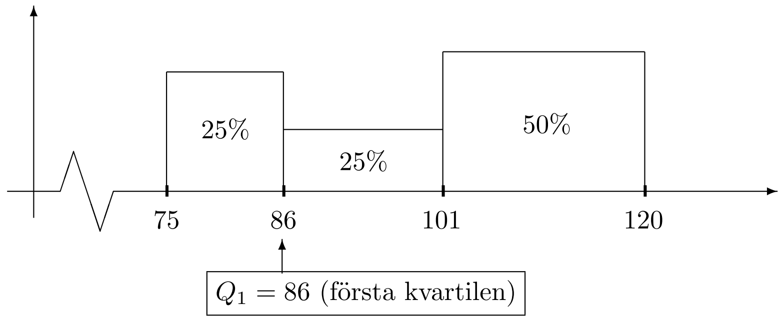
(b) Bilda ett histogram för stickprovet med klassgränserna $[75, 86)$, $[86, 101)$, $[101, 120)$ och beräkna första kvartilen. (3p)

(c) Någon påstår att antalet sidor i ett block är normalfördelat. Gör ett test, med utgångspunkt från histogrammet i uppgift (b), av om man kan förkasta detta påstående på 5% signifikansnivå. (3p)

Lösning:

(a) Om 100 ark blir mindre än en 1 cm tjock bunt så får det plats fler ark i blocket. Därmed är $P(\text{Ett block har minst 100 ark}) = P(\text{100-arksbunten är} \leq 1 \text{ cm}) = P(\sum_{i=1}^{100} X_i \leq 10)$ där X_i är tjockleken hos ark i. Vi har att $E(X_i) = \frac{0.06+0.13}{2} = 0.095$ och $V(X_i) = \frac{(0.06-0.13)^2}{12} = 0.000408333$. Därmed är $E(\sum_{i=1}^{100} X_i) = 100 \cdot 0.095 = 9.5$ och $V(\sum_{i=1}^{100} X_i) = 100 \cdot 0.000408333 = 0.0408333$. Enligt CGS är därmed $P(\sum_{i=1}^{100} X_i \leq 10) \approx \Phi(\frac{10-9.5}{\sqrt{0.0408333}}) = \Phi(2.47) = 0.9932$.

(b)



(c) Låt X beteckna antalet ark i ett block. Är då X normalfördelat?
Denna fråga beskrivs av hypotesen $\begin{cases} H_0 : X \text{ är normalfördelad} \\ H_1 : X \text{ är ej normalfördelad} \end{cases}$
som testas m.h.a. ett χ^2-test. Vi får:
$\bar{x} = 98.625$ och $\sum_i x_i^2 = 79199 \Rightarrow s^2 = \frac{1}{8-1}(79199 - 8 \cdot 98.625^2) = 197.69.$
Därmed får vi intervallsannolikheterna (under normalfördelningsantagande):

$[75, 86)$: $P(75 < X < 86) = \Phi(\frac{86-98.625}{\sqrt{197.69}}) - \Phi(\frac{75-98.625}{\sqrt{197.69}}) =$
$= (1 - \Phi(0.90)) - (1 - \Phi(1.68)) = 0.1376.$

$[86, 101)$: $P(86 < X < 101) = \Phi(\frac{101-98.625}{\sqrt{197.69}}) - \Phi(\frac{86-98.625}{\sqrt{197.69}}) =$
$= \Phi(0.17) - (1 - \Phi(0.90)) = 0.2484.$

$[101, 120)$: $P(101 < X < 120) = \Phi(\frac{120-98.625}{\sqrt{197.69}}) - \Phi(\frac{101-98.625}{\sqrt{197.69}}) =$
$= \Phi(1.52) - \Phi(0.17)) = 0.3682.$

Med hjälp av dessa sannolikheter kan vi beräkna det förväntade antalet (av det totalt 8 observationerna) vi hade fått under normalfördelningsantagandet:
$E_1 = 8 \cdot 0.1376 = 1.1008$, $E_2 = 8 \cdot 0.3834$, $E_3 = 8 \cdot 0.3682 = 2.9456$.
Detta ger värde på χ^2-statistikan:
$\sum_{i=1}^{3} \frac{(O_i - E_i)^2}{E_i} = \frac{(2-1.1008)^2}{1.1008} + \frac{(2-3.0672)^2}{3.0672} + \frac{(4-2.9456)^2}{2.9456} = 0.7345 + 0.3713 + 0.3774 = 1.4832$. Men $\chi^2_{0.05}(3-1) = 5.9915$ så man kan ej förkasta H_0 : på 5% signifikansivå. $\square$

6. Låt funktionen $f(x) = Cxe^{-\theta x^2}$ vara definierad för alla $x > 0$, där θ är ett positivt reellt tal. Beräkna

(a) C så att $f(x)$ är en täthetsfunktion för en positiv stokastisk variabel X. (2p)

(b) medianen för X. (3p)

(c) Antag att $X_1, X_2, \ldots, X_{2n}$ är ett stickprov på X. Visa att

$$\pi^{-n} \prod_{i=1}^{2n} \frac{1}{X_i}$$

är en väntevärdesriktigt skattning av θ^n. (4p)

Lösning:

(a) $\displaystyle\int_0^\infty xe^{-\theta x^2}dx = \left[-\frac{1}{2\theta}e^{-\theta x^2}\right]_0^\infty = -\frac{1}{2\theta}(0-1) = \frac{1}{2\theta} \quad \Rightarrow \quad C = 2\theta.$

(b) $\frac{1}{2} = \int_0^m 2\theta x e^{-\theta x^2}dx = 2\theta[-\frac{1}{2\theta}e^{-\theta x^2}]_0^m = -(e^{-\theta m^2} - 1) \quad \Rightarrow \quad e^{-\theta m^2} = \frac{1}{2} \quad \Rightarrow$
$\Rightarrow \ \theta m^2 = \ln 2 \ \Rightarrow \ m = \sqrt{\frac{1}{\theta}\ln 2}.$

(c) Ska visa att $E(\pi^{-n}\prod_i^{2n}\frac{1}{X_i}) = \theta^n$.
Eftersom stickprovet $X_1, X_2, \ldots, X_{2n}$ består av oberoende variabler har vi att $E(\prod_i^{2n}\frac{1}{X_1}) = \prod_i^{2n}E(\frac{1}{X_i})$. Vidare är $E(\frac{1}{X_i}) = \int_0^\infty \frac{1}{x}\cdot 2\theta x e^{-\theta x^2}dx = 2\theta\int_0^\infty e^{-\theta x^2}dx$. Uttryck av typen e^{-x^2} låter sig dock inte integreras så lätt, men en lösning är att utnyttja något vi känner till ganska väl; eftersom integrering av täthetsfunktionen av en standard normalfördelad variabel renderar 1, dvs $\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx = 1$, så får vi $\int_0^\infty e^{-x^2/2}dx = \sqrt{\frac{\pi}{2}}$. Därmed är $\sqrt{\frac{\pi}{2}} = \int_0^\infty e^{-x^2/2}dx \left\{ \begin{array}{c} \sqrt{\theta}\,y = \frac{x}{\sqrt{2}} \\ \sqrt{\theta}\,dy = \frac{1}{\sqrt{2}}dx \end{array} \right\} = \int_0^\infty e^{-\theta y^2}\sqrt{2\theta}\,dy \ \Rightarrow \ \int_0^\infty e^{-\theta y^2}dy = \frac{1}{2}\sqrt{\frac{\pi}{\theta}}$. Alltså har vi slutligen att

$$\begin{aligned}
E\left(\pi^{-n}\prod_{i=1}^{2n}\frac{1}{X_i}\right) &= \pi^{-n}\prod_{i=1}^{2n}E(\tfrac{1}{X_i}) \\
&= \pi^{-n}\prod_{i=1}^{2n}2\theta\frac{1}{2}\sqrt{\frac{\pi}{\theta}} \\
&= \pi^{-n}\theta^{2n}(\pi^{1/2})^{2n}(\theta^{-1/2})^{2n} \\
&= \theta^n
\end{aligned}$$

$\square$