Lösningar till tentamen i Matematisk Statistik, 5p

LGR00
6 juni, 2003 kl. 9.00 – 13.00

Kursansvarig: Eric Järpe
Maxpoäng: 30
Betygsgränser: 12p: G, 21p: VG
Hjälpmedel: Miniräknare samt tabell- och formelsamling som medföljer tentamenstexten.

1. Låt X vara en kontinuerlig stokastisk variabel sådan att
$P(X \leq x) = A(B + x + \sin x)$ där $x \in [-\pi, \pi]$ och A och B är konstanter.
   (a) Beräkna A och B. (2p)
   (b) Beräkna $E(X^2)$. (3p)

Lösning:

(a) $F(-\pi) = 0 \quad F(\pi) = 1 \quad \Rightarrow$

$$\Rightarrow \left\{ \begin{array}{rcl} A(B - \pi + \sin(-\pi)) & = & 0 \\ A(B + \pi + \sin \pi) & = & 1 \end{array} \right.$$

$$\Rightarrow \left\{ \begin{array}{rcl} AB - A\pi & = & 0 \\ AB + A\pi & = & 1 \end{array} \right.$$

$$\Rightarrow \left\{ \begin{array}{rcl} A & = & 1/2\pi \\ B & = & \pi \end{array} \right.$$

dvs $\underline{F(x) = \frac{1}{2\pi}(\pi + x + \sin x)}$.

(b) $f(x) = F'(x) = \frac{1}{2\pi}(1 + \cos x)$

$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_{-\pi}^{\pi} \frac{x^2}{2\pi}(1 + \cos x)\,dx = \frac{1}{2\pi}(\underbrace{\int_{-\pi}^{\pi} x^2 dx}_{I_1} + \underbrace{\int_{-\pi}^{\pi} x^2 \cos x\,dx}_{I_2})$

$I_1 = \int_{-\pi}^{\pi} x^2 dx = [\frac{x^3}{3}]_{-\pi}^{\pi} = \frac{\pi^3}{3} - (-\frac{\pi^3}{3}) = \frac{2\pi^3}{3}$

Eftersom $x^2 \cos x$ är integrerbar och jämn funktion är $\int_{-\pi}^{\pi} x^2 \cos x\,dx = 0$.

Om man inte känner till det får man partialintegrera:

$I_2 = \int_{-\pi}^{\pi} x^2 \cos x\,dx = \left\{ \begin{array}{ll} f = \cos x & g = x^2 \\ F = \sin x & g' = 2x \end{array} \right\} = [x^2 \sin x]_{-\pi}^{\pi} - 2\int_{-\pi}^{\pi} x \sin x\,dx =$

$= \left\{ \begin{array}{ll} f = \sin x & g = x \\ F = -\cos x & g' = 1 \end{array} \right\} = [x^2 \sin x]_{-\pi}^{\pi} - 2([x(-\cos x)]_{-\pi}^{\pi} - \int_{-\pi}^{\pi} 1 \cdot (-\cos x)\,dx) =$

$= [x^2 \sin x]_{-\pi}^{\pi} + 2([x \cos x]_{-\pi}^{\pi} - [\sin x]_{-\pi}^{\pi} = 0$

$\Rightarrow E(X^2) = \frac{1}{2\pi} \cdot \frac{2\pi^3}{3} = \underline{\frac{\pi^2}{3}}$.

$\square$

2. Till en utbildning söker dubbelt så många killar som tjejer. Oberoende av kön bär hälften glasögon.

   (a) Vad är sannolikheten att en sökande har glasögon eller är kille? (2p)

   (b) Antag att man slumpmässigt väljer 7 sökande. Vad är sannolikheten att 3 av dessa är tjejer och 4 är killar? (2p)

Lösning:

(a) Låt A vara händelsen att en sökande bär glasögon och B vara händelsen att denne är en kille. Då är
$P(\text{har glasögon eller är kille}) = P(A \cup B) = P(A) + P(B) - P(A \cap B) =$
$= P(A) + P(B) - P(A)P(B) = \frac{1}{2} + \frac{2}{3} - \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{6}(3 + 4 - 2) = \underline{5/6}$

(b) Låt X vara antalet killar bland de 7 sökande. Då är $X \in Bin(7, \frac{2}{3})$ och

$$P(X = 4) \;=\; \binom{7}{4}(\tfrac{2}{3})^4(1 - \tfrac{2}{3})^{7-4} \;=\; \frac{7 \cdot 6 \cdot 5}{3 \cdot 2 \cdot 1} \cdot \frac{16}{81} \cdot \frac{1}{27} \;=\; \underline{0.256}$$

□

3. Under perioden december-januari-februari snöar det på en viss plats 2.5 cm per dygn med standardavvikelse 2 cm per dygn. Samtidigt packas snön p.g.a. varmare väder med 1 cm per dygn med standardavvikelse 1/2 cm per dygn. Antag att väderleken är oberoende från dygn till dygn (dock ej normalfördelad). Vad är sannolikheten att snödjupet i slutet av februari är minst 1 meter? (3p)

Lösning:

Antal dygn $= 31 + 31 + 28 = 90$. Låt
$X_i =$ antal cm det snöar dygn i
$Y_i =$ antal cm som snön smälter dygn i
där $i = 1, 2, \ldots, 90$. Då är
$Z = \text{snödjupet} = \sum_{i=1}^{90} X_i \;-\; \sum_{i=1}^{90} Y_i$
$E(Z) = E(\sum_{i=1}^{90} X_i \;-\; \sum_{i=1}^{90} Y_i) = \sum_{i=1}^{90} E(X_i) \;-\; \sum_{i=1}^{90} E(Y_i) =$
$= 90 \cdot 2.5 - 90 \cdot 1 = 135.$
Eftersom $X_1, X_2, \ldots, X_{90}, Y_1, Y_2, \ldots, Y_{90}$ är oberoende är
$V(Z) = V(\sum_{i=1}^{90} X_i \;-\; \sum_{i=1}^{90} Y_i) = \sum_{i=1}^{90} V(X_i) \;+\; \sum_{i=1}^{90} V(Y_i) =$
$= 90(2.5)^2 + 90(0.5)^2 = 585$
Enligt *Centrala Gränsvärdessatsen* är snödjupet approximativt normalfördelat: $Z \in N(135, \sqrt{585})$ varmed
$\Rightarrow \; P(Z \geq 100) = 1 - \Phi(\frac{100-135}{\sqrt{585}}) = \Phi(\frac{35}{\sqrt{585}}) = \Phi(1.45) = \underline{0.9265}.$ □

4. Det sägs att ”en olycka kommer sällan ensam”. Det har Gunnar tagit fasta på och registrerar under en vecka ”oturs-händelser” som drabbar honom:

Veckodag	mån	tis	ons	tors	fre	lör	sön
Antal olyckor	10	1	3	4	8	2	7

Han vill bevisa att dessa olyckor tenderar att klumpa sig, dvs att de ej är jämnt fördelade över veckans sju dagar. Hjälp Gunnar med ett test på 1% signifikansnivå. (3p)

Lösning:
Antalet olyckor som Gunnar observerat är 35. Att dessa fördelar sig jämnt över veckans dagar innebär att man kan förvänta sig $\frac{35}{7} = 5$ olyckor varje dag. För att avgöra om Gunnars observerade frekvenser ska anses vara indikation på ”klumpning” görs ett chi-två-test: vi vill testa hypotesen $\begin{cases} H_1 : \mu_i = 5 \\ H_1 : \mu_i \neq 5 \end{cases}$ där $i = 1, 2, \ldots, 7$.

$$U = \sum_{i=1}^{7} \frac{(O_i - E_i)^2}{E_i} = \frac{5^2}{5} + \frac{(-4)^2}{5} + \frac{(-2)^2}{5} + \frac{(-1)^2}{5} + \frac{3^2}{5} + \frac{(-3)^2}{5} + \frac{2^2}{5} = \frac{68}{5} = 13.6$$

Men $\chi^2_{0.01}(6) = 16.81 \not< 13.6$ varmed H_0 *ej* kan förkastas på 1% signifikansnivå trots att olyckorna såg så ojämnt fördelade ut! Dvs <u>man kan ej skilja ojämnheten från slumpmässiga avvikelser på 1% signifikansnivå.</u> □

5. Inför en friluftsdag planeras vilka aktiviteter som ska vara valbara. Till att börja med måste man veta hur många som kommer vara med, speciellt om deltagarantalet blir fler än 100 elever, för att kunna genomföra vissa planerade aktiviteter. Från tidigare år har man följande statistik över deltagarantalet:

$$118 \quad 97 \quad 108 \quad 96 \quad 98 \quad 172 \quad 89 \quad 102$$

Antag att antalet deltagare är normalfördelat och

(a) Antag att deltagarantalet är oberoende för olika friluftsdagar. Kan man visa på 5% signifikansnivå att det förväntade deltagarantalet är fler än 100? (2p)

(b) Bilda ett 95% konfidensintervall för det förväntade deltagarantalet. (2p)

(c) Antag att man vet att standardavvikelsen är 21 deltagare per friluftsdag och att det förväntade deltagarantalet μ egentligen är 120. Hur stor styrka får ett test likt det ovan med 8 observationer att förkasta hypotesen att $\mu > 100$ på 1% signifikansnivå? (3p)

Lösning:

(a) Vi vill testa hypotesen $\begin{cases} \mu = 100 \\ \mu > 100 \end{cases}$.

$\bar{x} = \frac{880}{8} = 110$
$s^2 = \frac{1}{8-1}(\sum x_i^2 \ - 8(110)^2) = \frac{1}{7}(101\,726 - 96\,800) = 703.71$
$U = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} = \frac{110-100}{\sqrt{703.71/8}} = 1.066$
Men $\lambda_{0.05} = 1.64 \not< 1.066$ varmed H_0 ej kan förkastas dvs <u>man kan ej visa att deltagarantalet är större än 100 i genomsnitt.</u>

(b) Eftersom σ är okänd är ett 95% konf. intervall för μ
$(\bar{x} - t_{\alpha/2}(m-1)\frac{s}{\sqrt{m}}\ ,\ \bar{x} + t_{\alpha/2}(m-1)\frac{s}{\sqrt{m}})$.
I detta fall är $\bar{x} = 110$, $s = \sqrt{585}$, $m = 8$, $\alpha = 0.05$.
Därmed är $t_{0.025}(7) = 2.3646$ varmed intervallet är
$(110 - 2.3646 \cdot \sqrt{\frac{585}{8}}\ ,\ 110 + 2.3646 \cdot \sqrt{\frac{585}{8}})$ dvs <u>(89.7 , 130.3)</u>

(c) $\lambda_{0.01} = 2.326348$
Styrkan $= P(\text{förkasta } H_0 | H_1 : \mu = 120) = P(U > \lambda_{0.01} | \mu = 120) =$
$= P(\frac{\bar{X}-100}{21/\sqrt{8}} > \lambda_{0.01} | \mu = 120) = P(\bar{X} > \lambda_{0.01} \cdot 21/\sqrt{8} + 100 | \mu = 120) =$
$= P(\frac{\bar{X}-120}{21/\sqrt{8}} > \frac{\lambda_{0.01}\cdot 21/\sqrt{8}+100-120}{21/\sqrt{8}} = 1 - \Phi(2.326348 - \frac{20\sqrt{8}}{21}) =$
$= 1 - \Phi(-0.37) = \Phi(0.37) = \underline{0.6443}$

□

6. Ett prov består av 2 st 4-poängs-uppgifter och 3 st 6-poängs-uppgifter. Sannolikheten att en elev försöker sig på en 4-poängare är 0.7 och en 6-poängare 0.8. Om en elev försökt lösa en uppgift ges antingen full poäng (med sannolikhet 0.4) eller halv poäng (med sannolikhet 0.6).

   (a) Vad är sannolikheten att en elev får minst 20 poäng men högst 21 poäng totalt? Ange alla antaganden du gör. (3p)

   (b) Vad är den betingade sannolikheten att en elev har försökt lösa alla 5 uppgifterna givet att poängresultatet totalt är antingen 20 eller 21? (2p)

Lösning:

(a) Antag att uppgifternas poäng är oberoende av varandra. Antag dessutom för enkelhets skull att de två 4-poängarna kommer först och sedan de tre 6-poängarna, och låt oss beteckna t.ex. händelsen "full päng på första uppgiften, halv poäng på andra, full på tredje, full på fjärde och inget på femte" med $\{F, H, F, F, I\}$ (vilket alltså är en av de konfigurationer som ger totalt $4 + 2 + 6 + 6 + 0 = 18$ poäng). Eftersom detta ger en indelning i disjunkta händelser (en partition) har vi att
$P(\text{minst 20 och högst 21 poäng}) = P(\text{exakt 20}) + P(\text{exakt 21})$ där
$P(\text{exakt 20}) = \binom{2}{1}P(\{I, H, F, F, F\}) + \binom{3}{1}P(\{F, F, I, F, F\}) + \binom{3}{2}P(\{F, F, H, H, F\}) =$
$= 2 \cdot 0.3 \cdot 0.7 \cdot 0.8^3 0.6 \cdot 0.4^3 + 3 \cdot 0.7^2 0.2 \cdot 0.8^2 0.4^4 + 3 \cdot 0.7^2 0.8^3 0.4^3 0.6^2 = 0.0304$
$P(\text{exakt 21}) = \binom{2}{1}\binom{3}{2}P(\{H, F, H, F, F\}) = 2 \cdot 3 \cdot 0.7^2 0.8^3 0.6^2 0.4^3 = 0.0347$
varmed $P(\text{minst 20 och högst 21 poäng}) = 0.0304 + 0.0347 = \underline{0.0651}$

(b) $P(\text{försökt lösa alla} \,|\, \text{20 eller 21 poäng}) = \frac{P(\{\text{försökt lösa alla}\} \cap \{\text{20 eller 21 poäng}\})}{P(\text{20 eller 21 poäng})} =$

$$= \frac{\binom{3}{2}P(\{F, F, H, H, F\}) + \binom{2}{1}\binom{3}{2}P(\{H, F, H, F, F\})}{0.0651} = \frac{0.0173 + 0.0347}{0.0651} = \underline{0.7988}$$

□

7. Visa att för en exponentialfördelad variabel är väntevärdet större än medianen. (3p)

Lösning:
Antag att $X \in Exp(\lambda)$ där $\lambda > 0$.
Enligt formelsamlingen (om inte annat) vet vi att $E(X) = \frac{1}{\lambda}$.
Medianen är det tal m som satisfierar $P(X \leq m) = P(X > m)$ dvs $P(X \leq m) = \frac{1}{2}$.
Frekvensfunktion för X är $f(x) = \lambda e^{-\lambda x}$. Därmed definieras medianen av ekvationen
$\frac{1}{2} = P(X \leq m) = \int_0^m f(x)\, dx = \lambda \int_0^m e^{-\lambda x} dx = \lambda[-\frac{1}{\lambda}e^{-\lambda x}]_0^m = -e^{-\lambda m} - (-e^0) =$
$= 1 - e^{-\lambda m} \quad \Rightarrow \quad e^{-\lambda m} = 1 - \frac{1}{2} \quad \Rightarrow \quad \lambda m = -\ln\frac{1}{2} = \ln 2 \quad \Rightarrow \quad m = \frac{1}{\lambda}\ln 2.$
Men $1 > \ln 2$ (ty $2 < e$ och $\ln x$ växande så $\ln 2 < \ln e = 1$) varmed vi har att

$$E(X) = \frac{1}{\lambda} > \frac{1}{\lambda}\ln 2 = m$$

□