Lösningar till Tentamen i Matematisk Statistik, 7.5 hp

Distanskurs
15 januari, 2011 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 $X \in Exp(\ln \lambda)$ där $\lambda > 1$. Beräkna
   (a) medianen av X om $\lambda = 2$. (3p)
   (b) λ om $P(\frac{1}{2} < X < 1) = \frac{1}{6}$. (3p)

Lösning:

(a) $\lambda = 2 \;\Rightarrow\; f_X(x) = \ln 2\, e^{-x \ln 2}$
$\Rightarrow\; 0.5 = \int_0^m f_X(x)\,dx = \int_0^m \ln 2\, e^{-x\ln 2} dx = \ln 2\,[-\frac{1}{\ln 2} e^{-x \ln 2}]_0^m = 1 - e^{-m \ln 2} =$
$= 1 - 2^{-m} \;\Rightarrow\; 2^{-m} = 0.5 \;\Rightarrow\; -m = \frac{\ln \frac{1}{2}}{\ln 2} = -1 \;\Rightarrow\; m = 1$

(b) $\frac{1}{6} = P(\frac{1}{2} < X < 1) = P(X < 1) - P(X < \frac{1}{2})$. Eftersom $P(X \le x) = \int_0^x \ln \lambda\, e^{-y \ln \lambda} dy = 1 - e^{-x \ln \lambda} = 1 - \lambda^{-x}$ så är $\frac{1}{6} = 1 - \lambda^{-1} - (1 - \lambda^{-1/2}) = \lambda^{-1/2} - \lambda^{-1}$. Multiplicerar vi med λ får vi $\frac{1}{6}\lambda - \lambda^{1/2} + 1 = 0$ och substituerar vi med $\alpha = \lambda^2 \geq 0$ får vi andragradaren $\alpha^2 - 6\alpha + 6 = 0$ med lösningarna $\alpha = 3 \pm \sqrt{3}$. Dessa svarar mot parametervärdena $\lambda_1 = 6(2 + \sqrt{3})$ och $\lambda_2 = 6(2 - \sqrt{3})$ varav endast λ_1 satisfierar ekvationen $\lambda^{-1/2} - \lambda^{-1} = \frac{1}{6}$. Därmed är $\lambda = 6(2 + \sqrt{3})$ den entydiga lösningen till $P(\frac{1}{2} < X < 1) = \frac{1}{6}$. □

2. I ett hus installeras ett inbrottslarm som ger falsklarm (dvs larmet går givet att det inte är inbrott) med sannolikhet 0.02. Efter ett års användning har det utlösts 3% av tiden och varit inbrott 8% av tiden.
   (a) Vad är den betingade sannolikheten att det blir inbrott under ett år då larmet inte går? (3p)
   (b) Om händelserna "larm" och "inbrott" antas oberoende, vad är sannolikheten att det under ett år blir inbrott eller larm? (2p)

Lösning:

(a) Låt I = "inbrott" och L = "larm".
Då är $P(L|I^C) = 0.02 \quad P(L) = 0.03 \quad P(I) = 0.08$ och
$P(I|L^C) = 1 - P(I^C|L^C) = 1 - \frac{P(I^C \cap L^C)}{P(L^C)} = 1 - \frac{P(L^C|I^C)P(I^C)}{1-P(L)} =$
$= 1 - \frac{(1-P(L|I^C))(1-P(I))}{1-P(L)} = 1 - \frac{(1-0.02)(1-0.08)}{1-0.03} = 1 - 0.92948 = 0.07051$

(b) $P(I \cup L) = P(I) + P(L) - P(I \cap L) \overset{I \perp L}{=} 0.08 + 0.03 - 0.08 \cdot 0.03 = 0.1076$. □

3. Ett parkeringsbolag har 100 parkeringshus runt om i Sverige. En viss dag finns det X_1 bilar i hus 1, X_2 bilar i hus 2, osv. Antag att $X_1, X_2, \ldots, X_{100}$ är oberoende av varandra och att $X_k \in Poi(50)$ där $k = 1, 2, \ldots, 100$. Vad är sannolikheten

   (a) att det sammanlagt finns högst 2 bilar i hus 1, 2 och 3 tillsammans? (3p)
   (Tips: Om $Y \perp Z$, $Y \in Poi(\lambda_Y)$ och $Z \in Poi(\lambda_Z)$ så är $X+Y \in Poi(\lambda_Y + \lambda_Z)$.)

   (b) approximativt att det i genomsnitt beräknat på de 100 parkeringshusen finns mellan 49 och 51 bilar per parkeringshus? (3p)

Lösning:

(a) $X_1, X_2, \ldots, X_{100}$ oberoende och $X_k \in Poi(50)$ för alla $k = 1, 2, \ldots, 100$.
Låt $Z = X_1 + X_2 + X_3$. Då är $Z \in Poi(150)$ och $P(Z \leq 2) = P(Z = 0) + P(Z = 1) + P(Z = 2) = \frac{150^0}{0!}e^{-150} + \frac{150^1}{1!}e^{-150} + \frac{150^2}{2!}e^{-150} = (1 + 150 + \frac{150^2}{2})e^{-150} = 8.180327 \cdot 10^{-62}$

(b) $\bar{X} = \frac{1}{100}\sum_{k=1}^{100} X_k \overset{appr.}{\in} N(\mu, \sigma^2)$ där
$\mu = E(\frac{1}{100}\sum_{k=1}^{100} X_k) = \frac{1}{100} \cdot 100 \cdot 50 = 50$
$\sigma^2 = V(\frac{1}{100}\sum_{k=1}^{100} X_k) = \frac{1}{100^2} \cdot 100 \cdot 50 = 0.5$
varmed
$P(49 < \bar{X} < 51) = P(\bar{X} < 51) - P(\bar{X} < 49) = \Phi(\frac{51-50}{\sqrt{0.5}}) - \Phi(\frac{49-50}{\sqrt{0.5}}) =$
$= \Phi(\sqrt{2}) - (1 - \Phi(\sqrt{2})) = 2\Phi(1.4142) - 1 = 0.8414.$ □

4. I en lärobok läser Kalle att *"... under hösten flyttar ladusvalan söderut och under oktober månad kan man se hur antalet observationer halveras från vecka till vecka..."*. Kalle betvivlar detta och under oktober observerar han

Vecka 41	*Vecka 42*	*Vecka 43*	*Vecka 44*
64	39	12	4

Kan du på 1% signifikansnivå bevisa att det citerade påståendet inte gäller hos Kalle? (3p)

Lösning: Hypotes $\begin{cases} H_0 : E(N_i) = \frac{1}{2}E(N_{i+1}), \text{ för alla } i = 41, 42, 43 \\ H_1 : E(N_i) \neq \frac{1}{2}E(N_{i+1}), \text{ för något } i = 41, 42, 43 \end{cases}$
där N_i är antalet observationer vecka i. Enligt nollhypotesen ska det förväntade antalet observationer fördela sig enligt

Vecka 41	*Vecka 42*	*Vecka 43*	*Vecka 44*
$8x$	$4x$	$2x$	x

varvid vi totalt får $15x$. Det totala antalet observationer är $n = 64 + 39 + 12 + 4 = 119$. Därmed är det förväntade antalet observationer vecka 44 under nollhypotesen $x = 119/15 = 7.93$ och under de andra veckorna $2x, 4x, 8x$ varmed vi får

i	1	2	3	4
O_i	64	39	12	4
E_i	63.47	31.73	15.87	7.93

$U = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i} = 0.004 + 1.666 + 0.944 + 1.948 = 4.5621 < 11.3449 = \chi^2_{0.01}(4-1).$
Nej, man kan inte förkasta H_0, dvs man kan inte bevisa på 1% sign.nivå att antalet observationer inte halveras från vecka till vecka hos Kalle. □

5. Pelle tränar för att kvalificera sig till DM i längdhopp. Under dagen har han hoppat

   5.48 m, 5.73 m, 5.69 m, 5.58 m, 5.71 m

   Hoppen kan antas vara normalfördelade.

   (a) Kan man på 5% signifikansnivå visa att Pelle hoppar längre än 5.5 m? (3p)

   Pelles konkurrent Arne hoppar (tillika normalfördelat)

   5.37 m, 5.40 m, 5.81 m, 5.58 m

   (b) Bilda ett 95% konfidensintervall för den förväntade differensen mellan Pelles och Arnes hopp. (3p)

Lösning:

(a) Hypotes $\begin{cases} H_0 : \mu = 5.5 \\ H_1 : \mu > 5.5 \end{cases}$

$\bar{x} = 5.638$

$\sum x_i^2 = 158.9799 \;\Rightarrow\; s^2 = \frac{1}{5-1}(\sum x_i^2 - 5\bar{x}^2) = 0.01117$

$U = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} = \frac{5.638-5.5}{\sqrt{0.01117/5}} = 2.9197 > 2.1318 = t_{0.05}(4)$

Ja, man kan ej förkasta H_0, dvs det är bevisat på 5% sign.nivå att Pelle hoppar längre än 5.5 m.

(b) $\bar{y} = 5.54 \quad \bar{\Delta} = \bar{x} - \bar{y} = 0.098$

$\sum y_i^2 = 122.8894 \;\Rightarrow\; s_Y^2 = \frac{1}{4-1}(\sum y_i^2 - 4\bar{y}^2) = 0.041$

$\Rightarrow\; s_\Delta^2 = \frac{(5-1)s_X^2 + (4-1)s_Y^2}{5+4-2} = \frac{0.16768}{7} = 0.02395$

$t_{0.025}(7) = 2.3646 \;\Rightarrow\; t_{0.025}(7) \cdot \sqrt{0.02395} \cdot \sqrt{\frac{1}{5} + \frac{1}{4}} = 0.2455$

Därmed är konfidensintervallet

$$(\underbrace{0.098 - 0.2455}_{\text{avrunda nedåt}}, \underbrace{0.098 + 0.2455}_{\text{avrunda uppåt}})$$

$$(\quad -0.147 \quad , \quad 0.344 \quad)$$

□

6. Antag att X och Y är oberoende och rektangulärfördelade på $(a, 1)$ där $0 < a < 1$. Bestäm konstanten C så att $S^* = C - \frac{\sqrt{3}}{2}\max(X, Y)$ blir en väntevärdesriktig skattning av standardavvikelsen av X. (4p)

Lösning:

$S^* = C - \frac{\sqrt{3}}{2}\max(X, Y)$ och X, Y oberoende och fördelade $R(a, 1)$. Låt $Z = \max(X, Y)$.

Då är $F_Z(z) = P(\max(X, Y) \le z) = P(X \le z, Y \le z) \overset{X \perp Y}{=} P(X \le z)P(Y \le z) =$

$= \begin{cases} 1 & z \ge a \\ (\frac{z-a}{1-a})^2 & a < z < 1 \\ 0 & z \le a \end{cases} \;\Rightarrow\; f_Z(z) = \frac{d}{dz}F_Z(z) = \frac{2}{1-a} \cdot \frac{z-a}{1-a} = \frac{2(z-a)}{(1-a)^2}$ då $a \le z \le 1 \;\Rightarrow$

$\Rightarrow\; E(Z) = \int_{\mathbb{R}} z\, f_Z(z)\, dz = \int_a^1 z \cdot \frac{2(z-a)}{(1-a)^2}\, dz = \frac{2}{(1-a)^2}\int_a^1 (z^2 - az)\, dz = \frac{2}{(1-a)^2}[\frac{z^3}{3} - \frac{az^2}{2}]_a^1 =$
$\frac{2}{(1-a)^2}(\frac{1}{3} - \frac{a}{2} - \frac{a^3}{3} + \frac{a^3}{2}) = \frac{1}{3(1-a)^2}(2 - 3a + a^3) = \frac{(1-a)(2-a-a^2)}{3(1-a)^2} = \frac{2-a-a^2}{3(1-a)} = \frac{(1-a)(2+a)}{3(1-a)} = \frac{2+a}{3}$.

Därmed är $E(S^*) = C - \frac{\sqrt{3}}{2}E(\max(X, Y)) = C - \frac{1}{\sqrt{12}}(2+a)$ och eftersom (enligt formelsamlingen) $V(X) = \frac{(1-a)^2}{12}$ så standardavvikelsen är $\frac{1-a}{\sqrt{12}}$ där $0 < a < 1$, ska C väljas så att $C - \frac{1}{\sqrt{12}}(2+a) = \frac{1-a}{\sqrt{12}}$, dvs $C = \frac{1-a+2+a}{\sqrt{12}} = \frac{3}{\sqrt{12}} = \frac{\sqrt{3}}{2}$.

□