MVE090 Matematisk statistik Z, 7.5 hp

Tentamen 18 oktober 2010 em

Tillåtna hjälpmedel är räknedosa utan lagrad information om kursen, Beta samt kursens formel- och tabellsamling.

Examinator är Tommy Norberg, ankn 3528 eller 0730 79 42 09.

Övningsledare är Anna Rudvik, ankn 5338 eller 0730 57 96 26.

Jour är Anna.

Maximalt antal tentamenspoäng är 30, av dessa krävs normalt 12 för godkänt betyg och 18 resp 24 för 4:a och 5:a. Lösningar till tentamensproblemen går att ladda ner från kurshemsidan.

Svar och lösningar skall motiveras om ej annat sägs i uppgiften.

Uppgifter

1. I ett försök observarades utfallet A. Man vet att en av de två hypoteserna H^+ och H^- måste ha orsakat det observerade utfallet, samt att det inträffar i ca 10% av fallen om det är H^+ och i ca 30% av fallen om det är H^- som gäller. Man drog slutsatsen att H^- var den mest troliga orsaken, samt att sannolikheten för H^- då A har observerats är ca

$$\frac{0.30}{0.10 + 0.30} = 0.75$$

   (a) Vilket antagande om förekomsten av de två hypoteserna H^+ och H^- har man då gjort? (2 p)

   (b) Antag nu att det var ett blodprov från en (i en stor population) godtyckligt vald individ som undersöktes och att man vet att proportionen H^- är ca 10%. Är H^- även nu den mest troliga orsaken? (2 p)

2. Låt frekvensen f vara binomialfördelad med parametrar $n = 1, 2, \ldots$ och $p \in (0, 1)$. Visa att $E[f] = np$ och $\mathrm{Var}[f] = np(1-p)$. För vilket värde på p är variansen störst? (4 p)

3. Hur stor är sannoliheten att du, om du skulle singla en enkrona 100 gånger, får mellan 40 och 60 klave? Svaret skall ej motiveras. (2 p)

4. De två stokastiska variablerna X, Y har tätheten

$$f(x, y) = cx^{\alpha-1} \quad \text{för} \quad 0 < x \leq y < 1$$

   där parametern är $\alpha > 0$ och c är en konstant, som naturligtvis beror av α.

   (a) Beräkna c. (2 p)

   (b) Antag att $\alpha = 1$. Beräkna $P(X \leq 0.4, 0.2 \leq Y \leq 0.6)$. (2 p)

5. Under 3 på varandra följande år har man observerat $x_1 = 2$, $x_2 = 5$, samt $x_3 = 3$ olyckor med dödlig utgång utefter en motortrafikled.

   (a) Skatta olycksintensiteten λ. (1 p)

   (b) Ange en vettig modell för observationerna. (1 p)

   (c) Är din skattning väntevärdesriktig m.a.p denna modell? (2 p)

   Deluppgifterna (a) och (b) behöver ej motiveras.

6. Det är viktigt för bibehållandet av hög kvalitet i en produkt att variansen σ^2 av en viss variabel ej är för stor. Som största acceptabla värde har man satt 2.00. Varje dag väljer man slumpmässigt ut $n = 15$ exemplar ur produktionen, mäter variabeln ifråga och beräknar stickprovsvariansen s^2. Hur stor får denna vara för att man med minst 90% konfidens ska kunna hävda att $\sigma^2 < 2.00$? Antag normalfördelade observationer. (4 p)

7. (J.f.r uppgift 6.) En dag mätte man upp $\bar{x} = 3.74$ och $s^2 = 1.072$ i $n = 15$ oberoende observationer. Beräkna ett 95% konfidensintervall för μ. Du kan anta att observationerna är normalfördelade. (4 p)

8. (J.f.r uppgift 6.) Av kvalitetsskäl är det även viktigt att variabeln som mäts ej är för stor. Under två på varandra följande dagar mätte man upp $\bar{x}_1 = 3.74$ resp $\bar{x}_2 = 4.29$. Stickprovsvarianserna var $s_1^2 = 1.072$ och $s_2^2 = 1.201$ (obs $n_1 = n_2 = 15$). Det ser ju ut som att väntevärdet är större dag 2, d.v.s att $\mu_2 > \mu_1$, och att man därför kanske behöver korrigera produktionen. Detta är emellertid kostsamt, eftersom det innebär minst en dags produktionsbortfall, så man undrar i vilken grad det går att verifiera statistiskt att $\mu_2 > \mu_1$? Antag normalfördelade observationer. (4 p)

Lycka till med lösandet av uppgifterna!

Svar eller lösningar till Mat stat Z den 18/10-2010

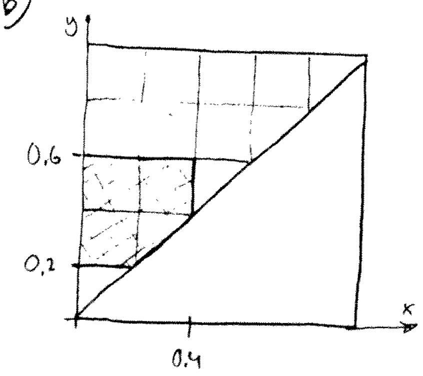
1. Bayes formel säger att $P(H^-|A) = P(H^-)P(A|H^-)/P(A)$, där $P(A) = P(H^+)P(A|H^+) + P(H^-)P(A|H^-)$. Om $P(H^+) = P(H^-) = 0.5$ följer att $P(H^-|A) = 0.3/(0.3 + 0.4) = 0.75$ så svaret på fråga (a) är att H^+ och H^- förekommer ungefär lika ofta i populationen. Om $P(H^-) = 0.1$, så följer $P(H^-|A) = 0.1 * 0.3/(0.1 * 0.3 + 0.9 * 0.1) = 0.03/0.12 = 0.25$ och $P(H^+|A) = 1 - 0.25 = 0.75$. Nu är det H^+ som är troligast, så svaret på fråga (b) är nej.

2. Vi använder att $f = \sum_{i=1}^{n} X_i$, där $X_i \sim \text{Ber}(p)$ och $X_1, \ldots, X_n$ är oberoende. Notera nu att $E[X_i] = p$ (ty $P(X_i = 1) = p$ och $P(X_i = 0) = 1 - p$) och $\text{Var}[X_i] = p(1-p)$ (ty $\text{Var}[X_i] = E[X_i^2] - E[X_i]^2 = E[X_i] - E[X_i]^2 = p - p^2 = p(1-p)$). Vi får nu, eftersom väntevärdet är en linjär operator, att $E[f] = \sum_i E[X_i] = np$ och, tack vare oberoendet, att $\text{Var}[f] = \sum_i \text{Var}[X_i] = np(1-p)$. Ett annat sätt att visa detta är via mgf. Se läroboken, sats 3.5.1 och övn 43 på sida 90. Variansen är störst då $p = 0.5$. För att se detta, derivera $\text{Var}[f] = np(1-p)$ m.a.p p, och sätt det erhållna uttrycket till 0. Då fås ekvationen $n - 2np = 0$, som implicerar $p = 0.5$. Att $\text{Var}[f]$ har ett lokalt och även globalt maximum då $p = 0.5$ inses av att andraderivatan är $-2n < 0$.

3. Ca 0.95.

4. (a) c fås ur ekvationen $1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy$. Högerledet $= \int_0^1 \int_0^y cx^{\alpha-1}\,dx\,dy = \int_0^1 \frac{cy^\alpha}{\alpha}\,dy = \frac{c}{\alpha(\alpha+1)}$, så $c = \alpha(\alpha+1)$. (b) Då $\alpha = 1$ är tätheten likformig (unform) på utfallsrummet $0 \le x \le y \le 1$, som har arean 0.5, varför $c = 2$. Detta värde behövs dock inte, eftersom ett enkelt geometriskt argument (se figur nedan) direkt ger att den sökta sannolikheten är $3.5/12.5 = 0.28$.

4b)



$P = \frac{3,5}{12,5} = 0{,}28$

5. (a) $\widehat{\lambda} = \frac{2+5+3}{3} = \frac{10}{3} \approx 3.33$ olyckor per år. Olycksintensiteten är ju förväntat antal olyckor under ett år. (b) En lämplig modell är $X \sim \text{Poi}(\lambda)$, där X representerar antalet olyckor under ett godtyckligt år. (c) Skattningen i (a) är väntevärdesriktig, eftersom $E[X] = \lambda$ och medelvärdet väntevärdesriktigt skattar väntevärdet.

6. Händelsen $\frac{(n-1)s^2}{\sigma^2} \ge \chi^2_{0.9}$ inträffar med 90 % sannolikhet, så vi får att påståendet $\sigma^2 \le \frac{(n-1)s^2}{\chi^2_{0.9}}$ har 90% konfidens. Om $\frac{(n-1)s^2}{\chi^2_{0.9}} \le 2.00$ kan vi således med minst 90% konfidens hävda att $\sigma^2 \le 2.00$. Notera nu att $\frac{(n-1)s^2}{\chi^2_{0.9}} \le 2.00$ om, och endast om, $s^2 \le \frac{2.00 \cdot \chi^2_{0.9}}{n-1} \approx \frac{2.00 \cdot 7.790}{14} \approx 1.113$.

7. $\mu = \bar{x} \pm t_{0.025}(n-1)s/\sqrt{n} = 3.74 \pm 2.145 \cdot \sqrt{\frac{1.072}{15}} = 3.74 \pm 0.57 = (3.17, 4.31)$. Ingen ytterligare motivering behövs.

8. Kvoten $s_1^2/s_2^2 = 1.072/1.201 = 0.893$ är, under $H_0 : \sigma_1 = \sigma_2$, $F(14,14)$-fördelad, och normala (med sannolikheten 80%) värden ligger i intervallet $(0.494, 2.022)$. Det går alltså inte ens på 20%-nivån att förkasta $H_0 : \sigma_1 = \sigma_2$, så vi utgår i den fortsatta analysen att variansen är densamma och lika med σ^2 i alla 30 mätningarna. Vi ska nu testa $H_0 : \mu_2 = \mu_1$ mot alternativet $H_1 : \mu_2 > \mu_1$. Vi börjar med att väga samman de två varianskattningarna $s_1^2 = 1.072$ och $s_2^2 = 1.201$ till en med $n_1 + n_2 - 2 = 28$ frihetsgrader: $s^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} = \frac{14\cdot 1.072+14\cdot 1.201}{14+14} = 1.137 = 1.07^2$. Teststatistikan är $t = \frac{\bar{x}_2-\bar{x}_1}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{4.29-3.74}{1.07\sqrt{\frac{1}{15}+\frac{1}{15}}} = 1.408$. I t-kvantiltabell läser vi av $t_{0.10}(28) = 1.313$ och $t_{0.05}(28) = 1.701$. Vi konstaterar att vi kan förkasta $H_0 : \mu_2 = \mu_1$, till förmån för $H_1 : \mu_2 > \mu_1$, på 10%-nivån, men ej på 5%-nivån. Konfidensen i påståendet $\mu_2 > \mu_1$ är alltså mellan 90% och 95%.