ESS011: Matematisk statistik och signalbehandling – Tid: 14.00-18.00, Datum: 2009-05-27

Examinator: Ottmar Cronie (Serik Sagitov)
Jour: Ottmar Cronie, tel. 031–772 35 44
Hjälpmedel: Formelblad, Formelsamling "Beta", Räknedosa.
Betygsgränser: För betyg 3 krävs 17 poäng, för betyg 4 krävs 25.5 poäng, för betyg 5 krävs 34 poäng, max. 43 poäng.
Fullständig och välmotiverad lösning på uppgift krävs för att full poäng skall erhållas.

---

1. Ett lotteri innehåller 100 lotter varav 5 ger vinst. Först drar Ottmar en lott, därefter drar Bill en lott. Låt A vara händelsen att Ottmar drar en vinstlott och låt B vara händelsen att Bill drar en vinstlott.

   a) Ange $\mathbb{P}(A)$, $\mathbb{P}(B|A)$, $\mathbb{P}(B|A^c)$. **(1p)**
   b) Beräkna $\mathbb{P}(A \cap B)$ och $\mathbb{P}(A^c \cap B)$. **(1p)**
   c) Beräkna $\mathbb{P}(B)$. **(1p)**
   d) Är A och B oberoende händelser? Visa matematiskt. **(1p)**
   e) Beräkna $\mathbb{P}(A \cup B)$. **(1p)**

   Lösningar:

   a) $\mathbb{P}(A) = 5/100$ **(1/3p)**, $\mathbb{P}(B|A) = 4/99$ **(1/3p)**, $\mathbb{P}(B|A^c) = 5/99$ **(1/3p)**
   b) $\mathbb{P}(A \cap B) = P(B|A)P(A) = \frac{4 \cdot 5}{9900}$ **(1/2p)**, $\mathbb{P}(A^c \cap B) = P(B|A^c)P(A^c) = \frac{5 \cdot 95}{9900}$ **(1/2p)**
   c) Beräkna $\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B) = \frac{1}{20}$ **(1p)**
   d) $\mathbb{P}(A \cap B) = \frac{20}{9900} \neq \frac{5}{2000} = \frac{5}{100}\frac{1}{20} = \mathbb{P}(A)\mathbb{P}(B)$ **(1p)**
   e) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = 97/990$ **(1p)**

2. Man misstänker att en mottagare inte tar emot sända signaler på ett korrekt sätt. För att ta reda på om så är fallet skickar man 20 oberoende meddelanden där varje meddelande består av en sekvens av 3 oberoende kontrollsignaler. Ett meddelande godtages som korrekt mottaget om åtminstone 2 av kontrollsignalerna mottages korrekt. Antag nu att sannolikheten är 0.5 att en kontrollsignal mottages korrekt. Låt X beteckna antalet korrekt mottagna meddelanden.

   a) Vad har X för typ av fördelning? **(1p)**
   b) Bestäm parametrarna i fördelningen för X. **(2p)**
   c) Beräkna sannolikheten att åtminstone två meddelanden mottages korrekt. **(1p)**
   d) Man kan möjligen approximera fördelningen för X med en annan typ av fördelning. Vilken är fördelningen och vilket resultat motiverar approximationen? Motivera varför fördelningen för X uppfyller de krav som ställs för att approximationen skall gälla? Gör om fråga c) men använd dig denna gång av den approximativa fördelningen. **(2p)**

Lösningar:

a) X är Binomialfördelad **(1p)**

b) $X \sim Bin(n,p)$
$n = 20$ **(1/2p)**
Låt $Y \sim Bin(3, 0.5)$, $p = \mathbb{P}(Y \geq 2) = 1 - \mathbb{P}(Y \leq 1) = 1 - \mathbb{P}(Y = 1) - \mathbb{P}(Y = 0) = 1 - 3 \cdot 0.5^3 - 0.5^3 = 0.5$ **(3/2p)**

c) $\mathbb{P}(X \geq 2) = 1 - \mathbb{P}(X \leq 1) = 1 - \mathbb{P}(X = 1) - \mathbb{P}(X = 0) = 1 - 20 \cdot 0.5^{20} - 0.5^{20} = 0.99998$ **(1p)**

d) Enligt Centrala gränsvärdessatsen (CGS) gäller att $X \overset{app.}{\sim} N\left(np, np(1-p)\right) = N\left(10, 5\right)$ **(1/2p)**
Approximationen fungerar eftersom $np = 10 > 5$ (tumregeln) **(1/2p)**
Låt $Y \sim N\left(10, 5\right)$,
$\mathbb{P}(X \geq 2) \approx \mathbb{P}(Y \geq 1.5) = 1 - \mathbb{P}\left(\frac{Y - \mathbb{E}[Y]}{\sqrt{\mathrm{Var}(Y)}} \leq \frac{1.5 - \mathbb{E}[Y]}{\sqrt{\mathrm{Var}(Y)}}\right) = 1 - \Phi\left(-3.80132\right) > 1 - \Phi\left(-3.69\right) = 0.9999$ **(1p)**

3. Förklara följande begrepp inom punktskattning:

a) Stickprov. **(1p)**

b) Skattare och skattning. **(1p)**

c) Väntevärdesriktig (unbiased) skattare. **(1p)**

d) Vad är ett lämpligt kriterium för val mellan två väntevärdesriktiga skattare? **(1p)**

Lösningar:

a) Stokastiska variabler $X_1, \ldots, X_n$ som är oberoende och likafördelade **(1p)**

b) Skattare: En statistika $\hat{\theta} = \hat{\theta}\left(X_1, \ldots, X_n\right)$ (funktion av stickprovet) som används för att skatta en parameter θ **(1/2p)**
Skattning: Det värde skattaren får då den räknas ut för det observerade stickprovet, $\hat{\theta}\left(x_1, \ldots, x_n\right)$ **(1/2p)**

c) $\mathbb{E}[\hat{\theta}] = \theta$ **(1p)**

d) Man väljer den som har minst standardfel, dvs om $\mathbb{E}[\hat{\theta}_1] = \mathbb{E}[\hat{\theta}_2] = \theta$ och $\sqrt{\mathrm{Var}(\hat{\theta}_1)} < \sqrt{\mathrm{Var}(\hat{\theta}_2)}$ då väljer vi $\hat{\theta}_1$ **(1p)**

4. Låt $X_1, \ldots, X_{18}$ vara ett stickprov från $N(0,1)$-fördelningen och låt dessutom $Y_1, \ldots, Y_{36}$ vara ett stickprov från $N(1,4)$-fördelningen. Antag att båda stickproven är oberoende. I vanlig ordning betecknar $\bar{X}$ och $\bar{Y}$ medelvärdena av de respektive stickproven.

a) Varför är $\bar{X} - \bar{Y}$ normalfördelad? **(1p)**

b) Vad har $\bar{X} - \bar{Y}$ för exakt fördelning? **(1p)**

c) Vad är $\mathrm{Cov}\left(X_1, \bar{Y}\right)$? **(1p)**

d) Vad är $\mathbb{P}(\bar{X} - \bar{Y} \leq -1)$. **(1p)**

e) Betrakta nu ett stickprov $G_1, \ldots, G_n$ som kommer från en $Poi(\lambda)$-fördelning.
Finn momentgenererande funktionen för $G = \sum_{i=1}^{n} G_i$.
Vad för fördelning har G? **(3p)**

Lösningar:

a) Eftersom varje linjärkombination av oberoende normalfördelade stok.var. är normalfördelad får vi att $\bar{X} = \sum_{i=1}^{18} \frac{1}{18} X_i$ är normalfördelad pga att $X_1, \ldots, X_{18}$ är oberoende och normalfördelade (enl. definitionen av ett normalfördelat stickprov) och med samma resonemang får vi att $\bar{Y} = \sum_{i=1}^{36} \frac{1}{36} Y_i$ är normalfördelade pga att $Y_1, \ldots, Y_{36}$ är oberoende och normalfördelade. Detta medför att $\bar{X} - \bar{Y}$ i sin tur är en linjärkombination av oberoende normalfördelade stok.var. (de två stickproven är oberoende) och därför är normalfördelad. **(1p)**

b) $\mathbb{E}\left[\bar{X} - \bar{Y}\right] = \mathbb{E}[X_i] - \mathbb{E}[Y_i] = -1$ **(1/2p)**
$\mathrm{Var}\left(\bar{X} - \bar{Y}\right) \overset{ober.}{=} \mathrm{Var}(\bar{X}) + \mathrm{Var}(\bar{Y}) = \frac{\mathrm{Var}(X_i)}{18} + \frac{\mathrm{Var}(Y_i)}{36} = 3/18$ **(1/2p)**
$\bar{X} - \bar{Y} \sim N(-1, 3/18)$

c) $\mathrm{Cov}\left(X_1, \bar{Y}\right) \overset{ober.}{=} 0$ **(1p)**

d) Vad är $\mathbb{P}(\bar{X} - \bar{Y} \leq -1) = 1/2$ **(1p)**

e) Varje G_i har momentgenererande funktion $m_{G_i}(t) = \mathrm{e}^{\lambda(\mathrm{e}^t - 1)}$
$m_G(t) = \mathbb{E}\left[\mathrm{e}^{tG}\right] = \mathbb{E}\left[\mathrm{e}^{\sum_{i=1}^{n} tG_i}\right] \overset{ober.}{=} \prod_{i=1}^{n} \mathbb{E}\left[\mathrm{e}^{tG_i}\right] = \prod_{i=1}^{n} \mathrm{e}^{\lambda(\mathrm{e}^t - 1)} = \mathrm{e}^{n\lambda(\mathrm{e}^t - 1)}$ **(2p)**
G har samma momentgenererande funktion som en $Poi(n\lambda)$-fördelad stok.var. $\Leftrightarrow G \sim Poi(n\lambda)$ **(1p)**

5. Låt $X_1, \ldots, X_k$ vara ett stickprov från $Bin(n, p)$-fördelningen.

a) Härled $\hat{p}_{ml}$; maximum-likelihood-skattaren för p. **(2p)**

b) Antag att $n = 5$. Vad blir maximum-likelihood-skattningen då vi har följande observationer? **(1p)**

1	0	1	2	0
0	0	0	1	0

c) Är $\hat{p}_{ml}$ en väntevärdesriktig skattare för p? Vad har $\hat{p}_{ml}$ för standardfel? Visa. **(2p)**

Lösningar:

a) $L(p) = \prod_{i=1}^{k} f_{X_i}(x_i) = \prod_{i=1}^{k} \binom{n}{x_i} p^{x_i} (1-p)^{n-x_i}$ **(1/2p)**
$\Longrightarrow$
$l(p) := ln(L(p)) = \sum_{i=1}^{k} ln\left\{\binom{n}{x_i}\right\} + ln(p) \sum_{i=1}^{k} x_i + ln(1-p) \sum_{i=1}^{k} (n - x_i)$ **(1/2p)**
$\Longrightarrow$
$0 = \frac{dl(p)}{dp} = \frac{1}{p} \sum_{i=1}^{k} x_i - \frac{1}{1-p} \sum_{i=1}^{k} (n - x_i)$
$p \sum_{i=1}^{k} (n - x_i) = (1-p) \sum_{i=1}^{k} x_i$
$p \sum_{i=1}^{k} n = \sum_{i=1}^{k} x_i$
$\Longrightarrow$
$\hat{p}_{ml} = \frac{1}{nk} \sum_{i=1}^{k} X_i$ **(1p)**

b) Skattningen ges av $\hat{p} = \frac{1}{5 \cdot 10}(1+1+2+1) = 0.1$ **(1p)**

c) $\mathbb{E}[\hat{p}_{ml}] = \frac{1}{nk}\sum_{i=1}^{k} \underbrace{\mathbb{E}[X_i]}_{=np} = p$ **(1p)**

$$SE(\hat{p}_{ml}) = \sqrt{\text{Var}(\hat{p}_{ml})} = \sqrt{\left(\frac{1}{nk}\right)^2 \text{Var}\left(\sum_{i=1}^{k} X_i\right)} \overset{ober.}{=} \sqrt{\left(\frac{1}{nk}\right)^2 \sum_{i=1}^{k} \underbrace{\text{Var}(X_i)}_{=np(1-p)}} =$$

$$\sqrt{\frac{knp(1-p)}{n^2k^2}} = \sqrt{\frac{p(1-p)}{nk}}$$ **(1p)**

6. En skyddsfilm för LCD-skärmar är under framtagning. Syftet är att den skall reducera mängden skadlig strålning som skärmen avger (mäts i enheten *skstr*). Mätningar görs på 8 skärmar (prototypen som prövas är dyr att ta fram) dels före man sätter dit skyddsfilmen och dels efter skyddsfilmen har placerats på skärmarna (normalfördelning kan antagas gälla för alla linjärkombinationer av stickproven):

Dator	1	2	3	4	5	6	7	8
Utan film	5.6	6.3	4.9	5.8	5.5	5.7	6.1	5.4
Med film	5.2	5.8	4.3	5.2	4.6	4.7	5.8	5.5

a) Finn ett lämpligt dubbelsidigt 95%igt konfidensintervall för att göra ett utlåtande om den förväntade strålningsförändringen. **(2p)**

b) Kan man på basis av intervallet i a) dra slutsatsen att filmen minskar strålningen? Motivera. **(1p)**

c) Om normalfördelningsantagandet ej hade varit uppfyllt, vad hade man då varit tvungen att kräva för att konstruktionen av intervallet i a) (approximativt) skall gälla? **(1p)**

d) Genom att utföra ett signifikanstest som besvarar frågan om vi har fått förminskad strålning, motivera huruvida det erhållna p-värdet stödjer en förminskning, givet signifikansnivån 0.05. **(2p)**

Lösningar:

a)

Dator	1	2	3	4	5	6	7	8
Utan film, y_i	5.6	6.3	4.9	5.8	5.5	5.7	6.1	5.4
Med film, x_i	5.2	5.8	4.3	5.2	4.6	4.7	5.8	5.5
Differens, $d_i = x_i - y_i$	-0.4	-0.5	-0.6	-0.6	-0.9	-1	-0.3	0.1

$n = 8$
$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i = -0.525$
$s_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2 = 0.119286$
$s_d = 0.345378$
$\alpha = 0.05 \Rightarrow t_{1-\alpha/2}^{(n-1)} = t_{0.975}^{(7)} = 2.36462$
(1p)
$\left[\bar{d} - t_{1-\alpha/2}^{(n-1)}\frac{s_d}{\sqrt{n}}, \bar{d} + t_{1-\alpha/2}^{(n-1)}\frac{s_d}{\sqrt{n}}\right] = [-0.813742, -0.236258]$ **(1p)**

b) Eftersom alla värden i intervallet är negativa (det innehåller ej 0) konstaterar vi (med slh 0.95) att filmen minskar strålningen **(1p)**

c) Man hade varit tvungen att kräva ett stort stickprov **(1p)**

d) Vi testar $H_0 : \mu_D = \mu_X - \mu_Y = 0$ mot $H_1 : \mu_D = \mu_X - \mu_Y < 0$ **(1/2p)**

Teststatistika: $T = \frac{\bar{D}-\mu_D^0}{S_d/\sqrt{n}} = \frac{\bar{D}}{S_d/\sqrt{8}} \sim T_{8-1}$
Observerad teststatistika: $T_{obs} = \frac{\bar{d}}{s_d/\sqrt{8}} = \frac{-0.525}{0.345378/\sqrt{8}} = -4.29942$ **(1/2p)**
p-värdet: $X \sim T_7$, $P = \mathbb{P}(X < T_{obs}) = \mathbb{P}(X < -4.29942) = \mathbb{P}(X > 4.29942) < \mathbb{P}(X > 3.49948) \overset{Tabell}{=} 0.005$, så vi kan förkasta $H_0 : \mu_D = 0$ på signifikansnivån 0.05**(1p)**

7. Pastörisering av juice gör att juicens hållbarhet förlängs avsevärt. Man jämför två olika typer av pastöriseringsmaskiner och är intresserad av hur snabbt varje maskin förväntas kunna fylla ett juicepaket. Vi har fått följande information om våra (normalfördelade) stickprov.

Maskin 1	Maskin 2
$n_1 = 25$	$n_2 = 25$
$\bar{x}_1 = 115.5$	$\bar{x}_2 = 112.7$
$s_1^2 = 25.2$	$s_2^2 = 7.6$

a) Vad kan du statistiskt fastställa (signifikansnivå 0.2) om förhållandet mellan varianserna i de två stickproven? **(2p)**

b) Testa hypotesen att de två maskinerna i snitt är lika snabba (signifikansnivå 0.1). **(2p)**

c) Baserat på din utsaga i b), vilken typ av fel finns det risk att vi har begått? **(1p)**

Lösningar:

a) Vi testar $H_0 : \sigma_1^2 = \sigma_2^2$ mot $H_1 : \sigma_1^2 \neq \sigma_2^2$ **(1/2p)**
Teststatistika: $T = \frac{S_1^2}{S_2^2} \sim F(25-1, 25-1)$ **(1/2p)**
Förkasta $H_0 : \sigma_1^2 = \sigma_2^2$ på signifikansnivån $\alpha = 0.2$ om:
$f_{1-\alpha/2}^{(n_1-1,n_2-1)} = f_{0.9}^{(24,24)} = 1.70185 < T_{obs}$ eller $T_{obs} < 0.58760 = 1/f_{0.9}^{(24,24)} = 1/f_{1-\alpha/2}^{(n_2-1,n_1-1)} = f_{\alpha/2}^{(n_1-1,n_2-1)}$ **(1/2p)**
Observerad teststatistika: $T_{obs} = \frac{s_1^2}{s_2^2} = \frac{25.2}{7.6} = 3.31579$ **(1/4p)**
Vi kan förkasta $H_0 : \sigma_1^2 = \sigma_2^2$ på signifikansnivån $\alpha = 0.2$ **(1/4p)**

b) Vi testar $H_0 : \mu_1^2 = \mu_2^2$ mot $H_1 : \mu_1^2 \neq \mu_2^2$ **(1/2p)**
Vi betraktar varianserna som olika pga utfallet i uppgift a) och får teststatistikan:
$T = \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)_0}{\sqrt{S_1^2/n_1+S_2^2/n_2}} \sim T_\gamma$ där
$\gamma = \left\lfloor \frac{\left(s_1^2/n_1+s_2^2/n_2\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1}+\frac{(s_2^2/n_2)^2}{n_2-1}} \right\rfloor = \lfloor 37.2693 \rfloor = 37$ **(1/2p)**
Förkasta $H_0 : \mu_1^2 = \mu_2^2$ på signifikansnivån $\alpha = 0.1$ om:
$f_{1-\alpha/2}^{(\gamma)} = f_{0.95}^{(37)} \approx f_{0.95}^{(30)} = 1.69726 < |T_{obs}|$ **(1/2p)**
Observerad teststatistika:
$T_{obs} = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{s_1^2/n_1+s_2^2/n_2}} = \frac{115.5-112.7}{\sqrt{(25.2+7.6)/25}} = 2.44451$ **(1/4p)**
Vi kan förkasta $H_0 : \mu_1^2 = \mu_2^2$ på signifikansnivån $\alpha = 0.1$ **(1/4p)**

c) Typ I-fel = Förkasta $H_0 | H_0$ sann **(1p)**

8. För att undersöka hur temperaturen förändras ju längre norrut i Sverige man kommer, så mättes årsmedeltemperaturen år 2004 i 11 svenska orter. De återges i följande tabell tillsammans med orternas latituder.

Ort	Latitud	Medeltemperatur
Jokkmokk	66.6	-0.6
Umeå	63.5	4.0
Östersund	63.1	4.2
Gävle	60.4	5.8
Karlstad	59.2	7.0
Stockholm	59.3	7.6
Göteborg	57.8	7.7
Jönköping	57.4	6.0
Visby	57.6	7.6
Kalmar	56.7	7.5
Lund	55.7	8.5

$\sum_i x_i = 657.3$ $\quad\sum_i y_i = 65.3$
$\sum_i x_i^2 = 39389.5$ $\quad\sum_i y_i^2 = 455.95$
$\sum_i x_i y_i = 3820.4$

a) Sätt upp en linjär regressionsmodell. **(1p)**

b) Skatta samtliga parametrar i modellen. **(1p)**

c) Testa på signifikansnivå $\alpha = 0.05$ hurvida årsmedeltemperaturen 2004 sjönk ju längre norrut man kom.
Vilka antaganden krävs för testet? **(3p)**

d) Tolka koefficienterna (uttryckt i termer relaterade till problemet). **(1p)**

Lösningar:

a) $Y_i = \beta_0 + \beta_1 x_i + E_i$ **(1p)**

b) $b_1 = \frac{\sum x_i y_i - \frac{1}{n}\sum x_i \sum y_i}{\sum x_i^2 - \frac{1}{n}(\sum x_i)^2} = -0.7234$ **(1/2p)**
$b_0 = \bar{y} - b_1\bar{x} = 49.1636$ **(1/2p)**

c) Vi testar $H_0 : \beta_1 = 0$ mot $H_1 : \beta_1 < 0$ **(1/2p)**
Teststatistika: $T = \frac{B_1 - \beta_1^0}{S/\sqrt{S_{xx}}} = \frac{B_1}{S/\sqrt{\sum x_i^2 - \frac{1}{n}(\sum x_i)^2}} \sim T_{11-2}$ **(1/2p)**
Förkasta $H_0 : \beta_1 = 0$ på signifikansnivån $\alpha = 0.05$ om:
$T_{obs} < -t_{1-\alpha}^{(n-2)} = -t_{0.95}^{(9)} = -1.83311$ **(1/2p)**
Man finner att $s^2 = \frac{1}{n-2}\sum(y_i - b_0 - b_1 x_i)^2 = 1.0312 \Rightarrow s = 1.01548$ **(1/2p)**
Observerad teststatistika: $T_{obs} = \frac{b_1}{s/\sqrt{\sum x_i^2 - \frac{1}{n}(\sum x_i)^2}} = -7.5656$ **(1/4p)**
Vi kan förkasta H_0 på signifikansnivån $\alpha = 0.05$ **(1/4p)**
Vi kräver att $E_1, \ldots, E_n$ är oberoende och normalfördelade **(1/2p)**

d) Då latituden är 0 är temperaturen 49.16 grader **(1/2p)** och den avtar med 0.7234 grader per latitud **(1/2p)**