

Lösningar till tentamen: Matematisk statistik och signalbehandling (ESS011), 14.00-18.00 den 14/1-2009

Examinator: Serik Sagitov (Kursansvarig: Ottmar Cronie)

Tillåtna hjälpmedel: Tabell "Beta", utdelad formelsamling, valfri räknedosa(minnestömd)

Betygsgränser: 12p $\rightarrow$ Betyg 3, 18p $\rightarrow$ Betyg 4, 24p $\rightarrow$ Betyg 5

Maxpoäng: 30p

1 Antag att du kastar en rättvis sex-sidig tärning (alla sidor har lika sannolikhet). Betrakta nu händelsen $A = \{5, 6\}$, dvs händelsen att vi i ett visst slag får åtminstone 5. Ange i var och ett av fallen nedan en händelse B sådan att

- a) B är oberoende av A . **(1p)**
- b) $P(A|B) = \frac{1}{2}$. **(1p)**
- c) $P(A \cup B) = 1$ och $P(A \cap B) = \frac{1}{6}$. **(1p)**

Lösningar:

- a) $B = \{1, 2, 3, 4, 5, 6\} \Rightarrow A \cap B = A, P(A \cap B) = \frac{2}{6} \cdot 1 = P(A)P(B)$
- b) $B = \{3, 4, 5, 6\} \Rightarrow P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{2/6}{4/6} = \frac{1}{2}$
- c) $B = \{1, 2, 3, 4, 5\} \Rightarrow A \cap B = \{5\}, A \cup B = \{1, 2, 3, 4, 5, 6\} \Rightarrow P(A \cap B) = \frac{1}{6}, P(A \cup B) = 1$

2 Antag att den stokastiska variabeln X är normalfördelad med väntevärde 3 och varians 9. Låt nu $Y = \frac{1}{3}X - 1$.

- a) Vad har Y för typ av fördelning och vad är dess väntevärde och varians? **(1.5p)**
- b) Vad är sannolikheten att Y är åtminstone 1? **(1p)**
- c) Beräkna $\text{Cov}(X, Y)$. **(1.5p)**

Lösningar:

- a) $E[Y] = \frac{1}{3}E[X] - 1 = 0, \text{Var}(Y) = (1/3)^2 \text{Var}(X) = 1 \Rightarrow Y \sim N(0, 1)$
- b) Pga att $N(0, 1)$ -förd. är symmetrisk i 0 får vi $P(Y \geq 1) = P(Y \leq -1) = 0.1587$
- c) $\text{Cov}(X, Y) = \text{Cov}(X, \frac{1}{3}X - 1) = \frac{1}{3} \text{Cov}(X, X) = \frac{1}{3} \text{Var}(X) = 3$

3 Besvara följande små frågor:

- a) Vad är definitionen av ett typ I-fel respektive ett typ II-fel? **(1p)**
- b) Vad skall gälla för att en funktion $f(x)$ skall vara en täthetsfunktion för en kontinuerlig stokastisk variabel Y ? **(1p)**
- c) Låt $X \sim Exp(\pi)$. Vad är $P(X = 5)$? **(1p)**

Lösningar:

- a) Typ I-fel är slh att förkasta H_0 givet att H_0 är sann. Typ II-fel är slh att acceptera H_0 givet att mothypotesen är sann
- b) $f(x) \geq 0$, $\int_{-\infty}^{\infty} f(x)dx = 1$, $P(a \leq Y \leq b) = \int_a^b f(x)dx$
- c) $P(X = 5) = 0$ eftersom X är en kontinuerlig stokastisk variabel

4 Företaget "Däck som bara den AB" får varje dag in en låda med ventiler till de däck som tillverkas i fabriken. Varje låda består av 100 ventiler och från varje låda som inkommit plockar man slumpmässigt ut 10 stycken som testas för läckage. Om någon av de testade ventilerna läcker skickar man tillbaka hela lådan med ventiler till underleverantören. En viss dag får man in en låda som totalt innehåller 5 trasiga ventiler.

- a) Låt Z beteckna antalet trasiga ventiler som man stöter på. Vad har Z för sannolikhetsfunktion? **(1p)**
- b) Hur många av de undersökta ventilerna förväntas vara trasiga? **(1p)**
- c) Vad är sannolikheten att lådan inte skickas tillbaka? **(1p)**
- d) Vad är sannolikheten att man finner mer än en trasig ventil? **(1p)**

Lösningar:

- a) Z är hypergeometriskt fördelad med parameter $(N, n, m) = (100, 10, 5)$,
 $f_Z(k) = P(Z = k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}$
- b) $E[Z] = mn/N = 5$
- c) $P(Z = 0) = \frac{\binom{5}{0}\binom{100-5}{10-0}}{\binom{100}{10}} = 0.5838$
- d) $P(Z > 1) = 1 - P(Z \leq 1) = 1 - P(Z = 1) - P(Z = 0) = 1 - 0.3394 - 0.5838 = 0.0769$

5 Låt $X_1, \dots, X_n$ vara ett stickprov på $X \sim N(\mu, \sigma^2)$.

- a) Härled och skriv fullständigt ut log-likelihoodfunktionen (bara den icke-logaritmerade likelihood-funktionen ger inga poäng). **(1p)**
- b) Finn maximum-likelihoodskattaren för σ^2 . **(1.5p)**
- c) Är detta en väntevärdesriktig (unbiased) skattare för σ^2 (visa)? **(1.5p)**

Lösningar:

a)

$$l(\mu, \sigma) = \log \left(\prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i - \mu)^2 / 2\sigma^2} \right) = -n \log(\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

b)

$$\begin{aligned} \frac{\partial l(\mu, \sigma)}{\partial \sigma} &= -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - \mu)^2 = 0 \\ n\sigma^2 &= \sum_{i=1}^n (x_i - \mu)^2 = 0 \\ \implies \\ \hat{\sigma}^2(X_1, \dots, X_n) &= \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \end{aligned}$$

c)

$$E[\hat{\sigma}^2(X_1, \dots, X_n)] = \frac{1}{n} \sum_{i=1}^n E[(X_i - \mu)^2] = \frac{1}{n} \sum_{i=1}^n \underbrace{\text{Var}(X_i)}_{=\sigma^2} = \frac{1}{n} n\sigma^2 = \sigma^2$$

6 Koncentrationen av en aktiv ingrediens i ett flytande tvättmedel förmodas påverkas av typen av katalysator som används i tillverkningsprocessen. Man har gjort ett antal mätningar av koncentrationer med de två katalysatorer som finns tillgängliga (där enheten är gram aktiv ingrediens per 100kg tvättmedel) och fått följande data:

Katalysator 1	Katalysator2
$n_1 = 16$	$n_2 = 14$
$\bar{x}_1 = 6150$	$\bar{x}_2 = 5250$
$s_1 = 80$	$s_2 = 75$

- Konstruera ett tvåsidigt 99%-igt konfidenstervall för den förväntade koncentrationsdifferensen. **(1.5p)**
- Vad drar du för slutsats i a)? (motivera) **(0.5p)**
- Använd ett ensidigt hypotestest (signifikansnivå 0.01) för att se om katalysatorerna ger upphov till någon skillnad i förväntad koncentration. **(1.5p)**
- Ge och motivera de antaganden som du har gjort i a) och c)? **(0.5p)**

Lösningar:

a) Om varianserna antas vara lika:

$$\begin{aligned} \text{Konfidenstervall} &\text{ ges av } \bar{X}_1 - \bar{X}_2 \pm t_{(1-\alpha)/2}^{(n_1+n_2-2)} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \\ S_p^2 &= \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} \end{aligned}$$

och vi får $s_p = \sqrt{\frac{(16-1)80^2 + (14-1)75^2}{16+14-2}} = \sqrt{6040.179} = 77.719$ samt $t_{0.995}^{(28)} = 2.763$

Insättning ger intervallet $6150 - 5250 \pm 2.763 \cdot 77.719 \sqrt{\frac{1}{16} + \frac{1}{14}} = 900 \pm 79$,
dvs $\mu_1 - \mu_2 \in (821, 979)$

Om varianserna antas vara olika:

Konfidensintervallet ges av $\bar{X}_1 - \bar{X}_2 \pm t_{(1-\alpha)/2}^{(\gamma)} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$

$$\gamma = \left\lfloor \frac{(S_1^2/n_1 + S_2^2/n_2)^2}{(S_1^2/n_1)^2/(n_1-1) + (S_2^2/n_2)^2/(n_2-1)} \right\rfloor$$

$$\text{Vi får } \gamma = \left\lfloor \frac{(80^2/16 + 75^2/14)^2}{(80^2/16)^2/(16-1) + (75^2/14)^2/(14-1)} \right\rfloor = \lfloor 27.8481 \rfloor = 27$$

och därmed $t_{0.995}^{(27)} = 2.771$

Insättning ger intervallet $6150 - 5250 \pm 2.771 \sqrt{\frac{80^2}{16} + \frac{75^2}{14}} = 900 \pm 78$, dvs
 $\mu_1 - \mu_2 \in (822, 978)$

- b) Katalysator 1 är att föredra då den ger upphov till mer av den aktiva ingrediensen i slutprodukten. Hade 0 funnits med i intervallet hade det ej gått att göra något uttalande.
- c) Vi testar $H_0 : \mu_1 = \mu_2$ mot $H_1 : \mu_1 > \mu_2$.

Om varianserna antas vara lika:

$$\text{Teststatistikan (under } H_0) \text{ ges av } T = \frac{(\bar{X}_1 - \bar{X}_2) - \overbrace{(\mu_1 - \mu_2)}^{=0}}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

och den förkastas om $T > t_{0.99}^{(28)} = 2.467$. Eftersom den observerade teststatistikan $T_{obs} = 31.64 > 2.467$ kan vi förkasta H_0 på signifikansnivån 0.01.

Om varianserna antas vara olika:

$$\text{Teststatistikan (under } H_0) \text{ ges av } T = \frac{(\bar{X}_1 - \bar{X}_2) - \overbrace{(\mu_1 - \mu_2)}^{=0}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

och den förkastas om $T > t_{0.99}^{(27)} = 2.473$. Eftersom den observerade teststatistikan $T_{obs} = 31.78 > 2.473$ kan vi förkasta H_0 på signifikansnivån 0.01.

I fallet $H_0 : \mu_1 = \mu_2$ mot $H_1 : \mu_1 \neq \mu_2$ använder man de kritiska gränser som ges av kvantilerna i uppgift a).

- d) Förutom antagandet om huruvida de underliggande varianserna är lika eller ej så måste de två stickproven vara oberoende och normalfördelade.

7 En forskare hävdar att åtminstone 10% av alla hockeyhjälm har ett visst tillverkningsfel som skulle kunna ge upphov till skador hos bäraren. Då man kontrollerar 200 hjälm upptäcker man att 16 av dessa har det påtalade felet.

- a) Ta reda på, mha ett lämpligt hypotestest med signifikansnivån 0.01, om detta stödjer forskarens påstående. **(1.5p)**

- b) Finn testets p-värde. **(1.25p)**
- c) Vad skulle ett 99%-igt konfidensintervall baserat på informationen ovan ha för bredd? **(1.25p)**

Lösning:

- a) Vi testar $H_0 : p = 0.1$ mot $H_1 : p \neq 0.1$.
 Då antalet ($n = 200$) oberoende Bernoulli-försök som görs (ger oss en Binomialfördelning) är stort gäller det att X (antalet trasiga hjälmar) är approximativt normalfördelat enligt centrala gränsvärdessatsen.
 Teststatistikan ges av $T = \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \stackrel{app}{\sim} N(0, 1)$ (under H_0).
 Vi förkastar H_0 om $|T| > z_{1-\alpha/2}$
 $\hat{p} = 16/200 = 0.08$
 Eftersom den observerade teststatistikan $|T_{obs}| = |-0.943| = 0.943 < 2.576$ kan vi ej förkasta H_0 på signifikansnivån 0.01
- b) Låt $Z \sim N(0, 1)$. p -värdet $= P(Z < -0.943) + P(Z > 0.943) = 1 - P(0.943 < Z < 0.943) = 1 - 2(\Phi(0.943) - \Phi(0)) = 1 - 2(0.8264 - 0.5) = 0.3472$.
- c) $2z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} = 0.0988259$

8 Vikt (X) (pounds) och systoliskt blodtryck (Y) (blodtryck vid hjärtats sammandragning) hos 26 slumpmässigt utvalda (och oberoende) män i åldrarna 25-30 visas i tabellen nedan.

Patientnr	Vikt	Blodtryck	Patientnr	Vikt	Blodtryck
1	165	130	14	172	153
2	167	133	15	159	128
3	180	150	16	168	132
4	155	128	17	174	149
5	212	151	18	183	158
6	175	146	19	215	150
7	190	150	20	195	163
8	210	140	21	180	156
9	200	148	22	143	124
10	149	125	23	240	170
11	158	133	24	235	165
12	169	135	25	192	160
13	170	150	26	187	159

Antag nu att vikt och blodtryck är bivariat normalfördelade. Från värdena ovan får vi att

$$\sum_i x_i = 4743 \quad \sum_i y_i = 3786 \quad \sum_i x_i^2 = 880545 \quad \sum_i y_i^2 = 555802 \quad \sum_i x_i y_i = 697076$$

- a) Utför en linjär regression (där alla koefficienter samt regressionslinjens slutliga form anges) med blodtryck som svarsvariabel. **(2p)**
- b) Tolka koefficienterna (uttryckt i termer relaterade till problemet). **(0.5p)**
- c) Skatta korrelationskoefficienten. **(0.5p)**

- d) Testa hypotesen att korrelationen är 0 på signifikansnivån 0.05. **(1p)**

Lösningar:

- a) $\mu_i = \mathbb{E}[Y|X = x_i] = \beta_0 + \beta_1 x_i$ där X anger vikten och Y blodtrycket.
 $n = 26$.

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sum x_i^2 - n \bar{x}^2} = \frac{\sum x_i y_i - \frac{1}{n} \sum x_i \sum y_i}{\sum x_i^2 - \frac{1}{n} (\sum x_i)^2} = 0.4194$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 69.1048$$

$$\hat{Y}|x = \hat{\beta}_0 + \hat{\beta}_1 x = 69.1048 + 0.4194x$$

- b) Då man inte har någon vikt (låter konstigt) ger sambandet att man har 69.1048 i blodtryck och det ökar med 0.4194 enheter per pound.

- c)

$$R = \hat{\rho} = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = \frac{\sum x_i y_i - \frac{1}{n} \sum x_i \sum y_i}{\sqrt{\left(\sum x_i^2 - \frac{1}{n} (\sum x_i)^2\right) \left(\sum y_i^2 - \frac{1}{n} (\sum y_i)^2\right)}} = 0.7735$$

- d) Vi testar $H_0 : \rho = 0$ mot $H_1 : \rho \neq 0$.

Teststatistikan ges av $T = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}}$ som under H_0 är t-fördelad med $n - 2$ frihetsgrader.

Vi förkastar H_0 om $|T_{obs}| > t_{1-\alpha/2}^{(n-2)} = t_{0.975}^{(24)} = 2.064$.

Vi får $T_{obs} = \frac{0.7735\sqrt{24}}{\sqrt{1-0.7735^2}} = 5.97883$ och kan därmed förkasta H_0 .