

Tentamen: Matematisk statistik och signalbehandling (ESS011), 2008-08-26

Tid och plats: 8.30-12.30, V-huset

Betygsgränser: 16p $\rightarrow$ Betyg 3, 24p $\rightarrow$ Betyg 4, 32p $\rightarrow$ Betyg 5

Kortfattade lösningar:

1 Lösningar:

- a) $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) = 0.5 \cdot 0.5 = 0.15$ **(2/3p)**
 $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = 0.3 + 0.5 - 0.15 = 0.65$ **(2/3p)**
 $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A) = 0.3$ **(2/3p)**
- b) $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) = 0.6 \cdot 0.5 = 0.3$
 $\Rightarrow$
 $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = 0.3 + 0.5 - 0.3 = 0.5$ **(1p)**
 $\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = 0.3/0.3 = 1$ **(1p)**
- c) Nej, ty summan av sannolikheter för disjunkta händelser får aldrig överstiga 1 **(1p)**.

2 Lösningar:

- a) X är geometriskt fördelad med parameter p **(1p)**
- b)

$$0.75 \leq \mathbb{P}(X \leq 2) = \sum_{k=1}^2 p(1-p)^{k-1} = p + p(1-p) = 2p - p^2 \quad \textbf{(2p)}$$

$$p^2 - 2p + 0.75 \leq 0$$

$$p \leq 1 \pm \sqrt{1 - 0.75} = 1 \pm 0.5$$

Svar: $p \leq 0.5$ **(2p)**

3 Lösningar:

- a) Låt $N(t)$ vara den stokastiska variabel som anger antalet storkar vid tiden t . Det gäller att $N(t) \sim \text{Poi}(t\lambda)$ så att $N(5) \sim \text{Poi}(5\lambda) \stackrel{\lambda=0.5}{=} \text{Poi}(2.5)$ **(2p)**.

$$\mathbb{P}(n_1 \leq N(t) \leq n_2) = \sum_{k=n_1}^{n_2} \mathbb{P}(N(t) = k) = \sum_{k=n_1}^{n_2} \frac{(t\lambda)^k e^{-t\lambda}}{k!}$$

Alltså

$$\begin{aligned} \mathbb{P}(2 \leq N(5) \leq 4) &= \sum_{k=2}^4 \mathbb{P}(N(5) = k) = \sum_{k=2}^4 \frac{2.5^k e^{-2.5}}{k!} \\ &= 0.2565 + 0.2138 + 0.1336 = 0.6039 \end{aligned}$$

(2p)

b) Pga glömske-egenskapen hos exponentialfördelningen får vi att

$$\begin{aligned}\mathbb{P}(T_1 \leq s+t | T_1 > s) &= 1 - \mathbb{P}(T_1 > s+t | T_1 > s) = 1 - \mathbb{P}(T_1 > t) \\ &= \mathbb{P}(T_1 \leq t) = 1 - e^{-\lambda t}\end{aligned}$$

(1p)

4 Lösningar: ML-skattaren:

$$L(\theta) = \prod_{i=1}^n f_{X_i}(x_i; \theta, k) = \prod_{i=1}^n x_i^{k-1} \frac{e^{-x_i/\theta}}{\theta^k \Gamma(k)} \quad (1p)$$

$\Rightarrow$

$$l(\theta) := \log(L(\theta)) = \sum_{i=1}^n \log(x_i^{k-1}) - \sum_{i=1}^n \frac{x_i}{\theta} - nk \log(\theta) - n \log(\Gamma(k)) \quad (1p)$$

$\Rightarrow$

$$0 = \frac{dl(\theta)}{d\theta} = \frac{1}{\theta^2} \sum_{i=1}^n x_i - \frac{nk}{\theta} \quad (1p)$$

$\Rightarrow$

$$\hat{\theta} = \frac{1}{nk} \sum_{i=1}^n x_i \quad (1p)$$

(4p)

Väntevärdesriktigheten:

$$\mathbb{E}[\hat{\theta}] = \mathbb{E}\left[\frac{1}{nk} \sum_{i=1}^n X_i\right] = \frac{1}{nk} \sum_{i=1}^n \mathbb{E}[X_i] = \frac{1}{nk} \sum_{i=1}^n k\theta = \frac{1}{nk} nk\theta = \theta \quad (1p)$$

5 Lösningar:

a) $X_1, \dots, X_n$ oberoende och likafördelade $N(\mu, \sigma^2)$. Då gäller för $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim N(\mu, \sigma^2/n)$ att

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

vilket medför att

$$1-\alpha = \mathbb{P}\left(-z_{1-\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{1-\alpha/2}\right) = \mathbb{P}\left(\bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\right)$$

(där $z_{1-\alpha/2}$ är kvantilen i fråga i $N(0, 1)$ -fördelningen) så att det $(1 - \alpha)100\%$ iga konfidensintervallet ges av

$$\left(\bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\right)$$

(2.5p)

b)

$$\left(\bar{X} - t_{n-1, 1-\alpha/2} \frac{S}{\sqrt{n}}, \bar{X} + t_{n-1, 1-\alpha/2} \frac{S}{\sqrt{n}} \right)$$

där $t_{n-1, 1-\alpha/2}$ är kvantilen i fråga i T_{n-1} -fördelningen. **(1p)**

c) $\bar{x} = 9.881$, $s = 1.075$ och $t_{19, 0.975} = 2.093$

$$\left(9.881 - 2.093 \frac{1.075}{\sqrt{20}}, 9.881 + 2.093 \frac{1.075}{\sqrt{20}} \right) = (9.769, 9.994)$$

(1.5p)

6 Lösningar:

a) Med $n = 350$, $\hat{p} = 66/350$ och $z_{1-\alpha/2} = 1.96$ får vi ett 95%igt konfidensintervall enligt

$$\left(\hat{p} - z_{1-\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n}, \hat{p} + z_{1-\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n} \right) = (0.1475, 0.2296)$$

(2p)

b) $H_0 : p = 0.2 (= p_0)$, $H_1 : p \neq 0.2 (= p_0)$

$$\frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{66/350 - 0.2}{\sqrt{0.2(1-0.2)/350}} = -0.5345$$

Eftersom $-z_{1-\alpha/2} = -1.96 < -0.5345 < 1.96 = z_{1-\alpha/2}$ kan vi ej förkasta H_0 . **(3p)**

7 Lösningar:

a) Testa hypotesen $H_0 : \mu_1 = \mu_2$ mot $H_1 : \mu_1 \neq \mu_2$ på signifikansnivån 0.05.

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(32 - 1)12.96^2 + (34 - 1)11.85^2}{32 + 34 - 2} = 153.76$$

$$T = \frac{(\bar{X}_1 - \bar{X}_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim T_{32+34-2} = T_{64}$$

$$T_{obs} = \frac{48.9 - 48.4}{\sqrt{153.76} \sqrt{\frac{1}{32} + \frac{1}{34}}} = 0.164$$

$$-t_{64, 0.975} = -1.998 < T_{obs} < 1.998 = t_{64, 0.975}$$

Vi kan ej förkasta H_0 på signifikansnivån 0.05.

p -värdet ges av

$$p = \mathbb{P}(T_{64} \leq -0.164) + \mathbb{P}(T_{64} \geq 0.164) = 2 \cdot 0.435 = 0.87$$

Då antalet frihetsgrader är stort kan vi approximera T_{64} med $N(0, 1)$

$$\begin{aligned} p &= \mathbb{P}(Z \leq -0.164) + \mathbb{P}(Z \geq 0.164) = 1 - \mathbb{P}(-0.164 \leq Z \leq 0.164) \\ &= 1 - 2(\Phi(0.164) - \Phi(0)) = 1 - 2(0.565 - 0.5) = 0.87 \end{aligned}$$

(3p)

- b) Konfidensintervallet ges av $\bar{x}_1 - \bar{x}_2 \pm t_{64,0.975} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 0.5 \pm 6.1$.
Eftersom $0 \in (-5.6, 6.6)$ kan vi inte, baserat på konfidensintervallet, säga att det finns någon skillnad. **(1.5p)**
- c) Man måste antaga att de två populationerna har samma varians. Normalfördelningsantagande behövs ej då stickprovsstorlekarna är tillräckligt stora. **(0.5p)**

8 Lösningar:

$\mu_i = \mathbb{E}[Y|X = x_i] = \beta_0 + \beta_1 x_i$ där X anger avståndet och Y halten. $n = 20$.

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sum x_i^2 - n \bar{x}^2} = \frac{\sum x_i y_i - \frac{1}{n} \sum x_i \sum y_i}{\sum x_i^2 - \frac{1}{n} (\sum x_i)^2} = -0.243$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{1}{20} 129.2 - (-0.243) \frac{1}{20} 210 = 9.012$$

$$\hat{Y}|x = \hat{\beta}_0 + \hat{\beta}_1 x = 9.012 - 0.243x$$

(2.5p) Determinationskoefficienten:

$$R = \hat{\rho}^2 = \left(\frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} \right)^2 = \left(\frac{\sum x_i y_i - \frac{1}{n} \sum x_i \sum y_i}{\sqrt{\left(\sum x_i^2 - \frac{1}{n} (\sum x_i)^2 \right) \left(\sum y_i^2 - \frac{1}{n} (\sum y_i)^2 \right)}} \right)^2 = 0.35$$

(1.5p) Halten kemikalie per kilo vatten avtar med 0.243 gram per km. Direkt vid fabriken är halten kemikalie i bäcken 9.012 gram per kilo vatten. Ungefär 99% av variationen i y kan förklaras av x . **(1p)**