TMA4240/4245
Statistikk
11. august 2012

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Eksamen - løsningsforslag

Oppgave 1

Vi har $N \sim N(\mu, \sigma^2)$, $\mu = 85$ og $P(X > 88) = 0.1$

a)

$$P(X > 88) = P\left(\frac{X-\mu}{\sigma} > \frac{88-\mu}{\sigma}\right) = P\left(Z > \frac{88-85}{\sigma}\right) = P(Z > 3/\sigma) = 0.1 \quad (1.1)$$

$$\Leftrightarrow 1 - P(Z \leq 3/\sigma) = 0.1 \quad (1.2)$$

$$\Rightarrow 3/\sigma = 1.28 \Leftrightarrow \sigma = 3/1.28 = \underline{\underline{2.344}} \quad (1.3)$$

$$P(X > 90) = P\left(\frac{X-\mu}{\sigma} > \frac{90-\mu}{\sigma}\right) = P\left(Z > \frac{90-85}{2.34}\right) = 1 - P(Z \leq 2.14) \quad (1.4)$$

$$= 1 - 0.9838 = \underline{\underline{0.0162}} \quad (1.5)$$

b) Vi har $n = 6$ uavhengige forsk (spydkast), kast over 88 meter = suksess, P(suksess) = $0.1 = p$, mao. en binomisk fordeling.

$$P(X \geq 3) = 1 - P(X < 3) = 1 - (P(X=2) + P(X=1) + P(X=0)) \quad (1.6)$$

$$P(X=2) = \binom{n}{x} p^x (1-p)^{n-x} = \binom{6}{2} 0.1^2 0.9^4 = 0.0984 \quad (1.7)$$

$$P(X=1) = \binom{6}{1} 0.1^1 0.9^5 = 0.3543 \quad (1.8)$$

$$P(X=0) = \binom{6}{0} 0.1^0 0.9^6 = 0.5314 \quad (1.9)$$

$$P(X \geq 3) = 1 - (0.0984 + 0.3543 + 0.5314) = \underline{\underline{0.0159}} \quad (1.10)$$

Lengste kast over 88 meter = minst en suksess.

$$P(X \geq 1) = 1 - P(X=0) = 1 - 0.5314 = \underline{\underline{0.4686}} \quad (1.11)$$

c) Gjennomsnitt:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{5} = \sum_{i=1}^{5} x_i = \frac{1}{5}(86.8 + 84.4 + 88.3 + 90.6 + 85.4) = \underline{\underline{87.1}}\text{m} \quad (1.12)$$

Empirisk standardavvik:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sqrt{\frac{1}{4}\sum_{i=1}^{5}(x_i - 87.1)^2} = \Bigg(\frac{1}{4}\big[(86.8-87.1)^2 + (84.4-87.1)^2 \tag{1.13}$$

$$+(88.3-87.1)^2 + (90.6-87.1)^2 + (85.4-87.1)^2\big]\Bigg)^{\frac{1}{2}} = \underline{\underline{2.447}}\text{m} \tag{1.14}$$

d) Vi får følgende hypotese:

$$H_0 : \mu = 85\text{m} \qquad \text{mot} \qquad H_1 : \mu > 85\text{m} \tag{1.15}$$

Teststatistikk:

$$t^* = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{87.1 - 85}{2.447/\sqrt{5}} = 1.92 \tag{1.16}$$

Kritisk verdi: $t_{\alpha,n-1} = t_{0.05,4} = 2.132$. Siden $t^* < t_{0.05,4}$ kan vi ikke forkaste H_0 ved signifikansnivå 0.05. Alternativt, p-verdi: $0.050 < P(T > t^*) < 0.075$, siden p-verdi> 0.05 kan vi ikke forkaste H_0 ved signifikansnivå 0.05.

e) Et konfidensintervall er et intervallestimat av en parameter. Et prediksjonsintervall er et intervallestimat av en ny observasjon. Et prediksjonsintervall er gitt ved

$$\left[\bar{x} - t_{\alpha/2,n-1}s\sqrt{1+\frac{1}{n}}, \bar{x} + t_{\alpha/2,n-1}s\sqrt{1+\frac{1}{n}}\right] = [87.1 \pm 2.776 \cdot 2.447\sqrt{6/5}] \tag{1.17}$$

Så prediksjonsintervallet til det sjette spydkastet er $[79.66, 94.54]$.

Oppgave 2 Normalfordeling: Dybdemålingene fra Rybekken ser normalfordelt ut fordi 1. og 3. kvantil er tilnærmet like langt fra medianen og det er få ekstremobservasjoner. Dette vil gi en nærmest symmetrisk fordeling som kjennetegner (bl.a.) normalfordelingen. Dataene fra Sylsjødammen har en mer skjev fordeling da 3. kvantil ligger betraktelig lengre fra medianen enn 1. kvantil. Den har i tillegg flere og mer ekstreme observasjoner. Derfor er disse trolig ikke normalfordelt.

Forventet snødybde: Vil være høyere ved Sylsjødammen enn ved Rybekken pga noen ekstremt høye verdier, men den store variasjonen i dataene vil trolig gjøre at forskjellen mellom stedene ikke er statistisk signifikant, mao forventet snødybde er like.

Varians: ser ut til å være betraktelig større ved Sylsjødammen enn ved Rybekken pga lengre avstand mellom medianen og 3. kvantil og de store ekstremobservasjonene.

Oppgave 3 Vi finner først den kumulative sannsynlighetsfordelingen:

$$F_X(x) = P(X \leq x) = \int_0^x f_X(u)du = \int_0^x \frac{1}{\beta}e^{-\frac{u}{\beta}}du = \left[-e^{-\frac{u}{\beta}}\right]_0^x = 1 - e^{-\frac{x}{\beta}} \tag{3.1}$$

a)

$$P(X>2) = 1 - P(X \leq 2) = 1 - (1 - e^{-\frac{2}{4}}) = e^{-\frac{1}{2}} = \underline{\underline{0.607}} \tag{3.2}$$

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^{\infty} \frac{x}{\beta} e^{-\frac{x}{\beta}} dx \tag{3.3}$$

Setter inn $u = \frac{x}{\beta}$ og $du = \frac{1}{\beta}dx \Leftrightarrow \beta du = dx$,

$$\beta \int_0^{\infty} u e^{-u} du = \beta \left[u(-e^{-u})|_0^{\infty} - \int_0^{\infty} (-e^{-u}) du \right] = \beta \left[0 + \int_0^{\infty} e^{-u} du \right] \tag{3.4}$$

$$= \beta \left[-e^{-u} \right]_0^{\infty} = \beta(0+1) = \beta = \underline{\underline{4}} \text{ mg} \tag{3.5}$$

$$P(X<4|X>2) = \frac{P(X<4 \cap X>2)}{P(X>2)} = \frac{P(2<X<4)}{P(X>2)} = \frac{\left[-e^{-x/4} \right]_2^4}{e^{-\frac{1}{2}}} \tag{3.6}$$

$$= \frac{-e^{-\frac{4}{4}} + e^{-\frac{2}{4}}}{e^{-\frac{2}{4}}} = 1 - e^{-\frac{1}{2}} = \underline{\underline{0.393}} \tag{3.7}$$

$$\mathrm{E}[X|X>2] = \frac{\int_2^{\infty} \frac{x}{\beta} e^{-\frac{x}{\beta}} dx}{P(X>2)} = \frac{\beta \left[u(-e^{-u})|_{2/\beta}^{\infty} - \int_{2/\beta}^{\infty} (-e^{-u}) du \right]}{P(X>2)} \tag{3.8}$$

$$= \frac{\beta \left(\frac{2}{\beta} e^{-\frac{2}{\beta}} + [-e^{-u}]_{2/\beta}^{\infty} \right)}{P(X>2)} = \frac{\beta \left(\frac{2}{\beta} e^{-\frac{2}{\beta}} + e^{-\frac{2}{\beta}} \right)}{e^{-\frac{2}{\beta}}} \tag{3.9}$$

$$= \beta \left(\frac{2}{\beta} + 1 \right) = 2 + \beta = \underline{\underline{6}} \text{ mg} \tag{3.10}$$

b) La $x = x_1 \ldots, x_n$.

$$L(x|\beta) = \prod_{i=1}^{n} f_X(x_i) = \prod_{i=1}^{n} \frac{1}{\beta} e^{-\frac{x_i}{\beta}} = \left(\frac{1}{\beta} \right)^n e^{-\sum_{i=1}^{n} \frac{x_i}{\beta}} \tag{3.11}$$

$$l(x|\beta) = \ln L(x|\beta) = -n \ln \beta - \frac{1}{\beta} \sum_{i=1}^{n} x_i \tag{3.12}$$

$$\frac{\partial l(x|\beta)}{\partial \beta} = -\frac{n}{\beta} + \frac{1}{\beta^2} \sum_{i=1}^{n} x_i = 0 \Leftrightarrow \frac{1}{\beta^2} \sum_{i=1}^{n} x_i = \frac{n}{\beta} \tag{3.13}$$

$$\Leftrightarrow \beta = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x} \tag{3.14}$$

Siden $\frac{\partial l(x|\beta)}{\partial \beta} > 0$ for $\beta < \bar{x}$ og $\frac{\partial l(x|\beta)}{\partial \beta} < 0$ for $\beta > \bar{x}$ er $\bar{x}$ et globalt maksimum og $\hat{\beta}_{\mathrm{SME}} = \bar{x}$.

$$\mathrm{E}[\hat{\beta}] = \mathrm{E}[\bar{x}] = \mathrm{E}\left[\frac{1}{n} \sum_{i=1}^{n} x_i \right] = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[x_i] = \frac{1}{n} \sum_{i=1}^{n} \beta = \underline{\underline{\beta}} \tag{3.15}$$

Finner $\mathrm{E}[x^2]$:

$$\mathrm{E}[x^2] = \int_0^\infty x^2 f_X(x)dx = \int_0^\infty \frac{x^2}{\beta}e^{-\frac{x}{\beta}}dx = \frac{\beta}{\beta}\int_0^\infty \frac{x^2}{\beta}e^{-\frac{x}{\beta}}dx = \beta\int_0^\infty \frac{x^2}{\beta^2}e^{-\frac{x}{\beta}}dx \tag{3.16}$$

$$= \beta^2\int_0^\infty u^2 e^{-u}du = \beta^2\left[u^2(-e^{-u})|_0^\infty - 2\int_0^\infty u(-e^{-u})du\right] \tag{3.17}$$

$$= 2\beta^2\left[\int_0^\infty ue^{-u}du\right] = 2\beta^2\left[u(-e^{-u})|_0^\infty - \int_0^\infty(-e^{-u})du\right] = 2\beta^2 \tag{3.18}$$

Dermed er $\mathrm{Var}[x] = \mathrm{E}[x^2] - (\mathrm{E}[x])^2 = 2\beta^2 - \beta^2 = \beta^2$. Man kan også gjenkjenne $f_X(x) = \frac{1}{\beta}e^{-\frac{x}{\beta}}$ som eksponentialfordelingen med forventing β og varians β^2.

$$\mathrm{Var}[\hat{\beta}] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^n x_i\right] = \left(\frac{1}{n}\right)^2\sum_{i=1}^n \mathrm{Var}[x_i] = \frac{1}{n^2}\sum_{i=1}^n \beta^2 = \underline{\underline{\frac{\beta^2}{n}}} \tag{3.19}$$

c) Siden X_i'ene er uavhengige og identisk fordelt, sier sentralgrenseteoremet at $\bar{X}$ er normalfordelt med forventning β og varians β^2/n. Dette kan brukes til å tilnærme konfidensintervallet til β da

$$P\left(-z_{\alpha/2} < \frac{\bar{X}-\beta}{\bar{X}/\sqrt{n}} < z_{\alpha/2}\right) = 1-\alpha \tag{3.20}$$

$$\underline{\underline{P\left(\bar{X} - z_{\alpha/2}\frac{\bar{X}}{\sqrt{n}} < \beta < \bar{X} + z_{\alpha/2}\frac{\bar{X}}{\sqrt{n}}\right) = 1-\alpha}} \tag{3.21}$$

Med $\alpha = 0.05$ og $\bar{X} = 5.2$ mg får vi:

$$5.2 \pm 1.96\frac{5.2}{\sqrt{50}} = \underline{\underline{(3.76, 6.64)}} \tag{3.22}$$

Alternativt, kan man regne konfidensintervallet eksakt:

$$P\left(-z_{\alpha/2} < \frac{\bar{X}-\beta}{\beta/\sqrt{n}} < z_{\alpha/2}\right) = 1-\alpha \tag{3.23}$$

$$\underline{\underline{P\left(\frac{\bar{X}}{1+z_{\alpha/2}/\sqrt{n}} < \beta < \frac{\bar{X}}{1-z_{\alpha/2}/\sqrt{n}}\right) = 1-\alpha}} \tag{3.24}$$

Med $\alpha = 0.05$ og $\bar{X} = 5.2$ mg får vi:

$$\frac{5.2}{1 \mp 1.96/\sqrt{50}} = \underline{\underline{(4.07, 7.19)}} \tag{3.25}$$

d) Vi har $Y = \frac{2X}{\beta}$ og $f_X(x) = \frac{1}{\beta}e^{-\frac{x}{\beta}}$. Vi setter

$$y = \frac{2x}{\beta} \Leftrightarrow x = \frac{y\beta}{2} = w(y) \tag{3.26}$$

og får $w'(y) = \frac{\beta}{2}$. Vi finner tetthetsfunksjonen til Y:

$$g_Y(y) = f_X(w(y))|w'(y)| = \frac{1}{\beta}e^{-\frac{y\beta}{2\beta}}\frac{\beta}{2} = \frac{1}{2}e^{-\frac{y}{2}} \qquad y > 0 \tag{3.27}$$

Videre er en kjikvadratfordeling definert som

$$h_Y(y|\nu) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} y^{\nu/2-1} e^{-y/2} \tag{3.28}$$

og vi ser at $g_Y(y) = h_Y(y|\nu)$ når $\nu = 2$, mao Y er en kjikvadratfordeling med 2 frihetsgrader.

La $\tilde{Y} = \sum_{i=1}^{n} Y_i$. Dette er en sum av uavhengige kjikvadratfordelte variabler med 2 frihetsgrader, derfor er $\tilde{Y}$ kjikvadrafordelt med $\nu = \sum_{i=1}^{n} 2 = 2n$ frihetsgrader. Vi kan skrive

$$\tilde{Y} = \sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} 2X_i/\beta = 2n\frac{1}{n}\sum_{i=1}^{n} X_i/\beta = 2n\bar{X}/\beta \tag{3.29}$$

Et konfidensintervall for en kjikvadrafordelt variabel er

$$P\left(\chi^2_{1-\alpha/2,\nu} < \tilde{Y} < \chi^2_{\alpha/2,\nu}\right) = 1-\alpha \tag{3.30}$$

$$P\left(\chi^2_{1-\alpha/2,\nu} < \frac{2n\bar{X}}{\beta} < \chi^2_{\alpha/2,\nu}\right) = 1-\alpha \tag{3.31}$$

$$P\left(\frac{2n\bar{X}}{\chi^2_{\alpha/2,\nu}} < \beta < \frac{2n\bar{X}}{\chi^2_{1-\alpha/2,\nu}}\right) = 1-\alpha \tag{3.32}$$

$$P\left(\frac{2\cdot 50\cdot 5.2}{129.6} < \beta < \frac{2\cdot 50\cdot 5.2}{74.2}\right) = 0.95 \tag{3.33}$$

Så det eksakte 95% konfidensintervallet for β er $(4.01, 7.01)$.