Løysningsforslag Eksamen i SIF 5060 Statistikk
2.desember 1999

Oppgåve 1

a) Har at $X \sim \mathrm{N}(0, 2^2)$ og $Y \sim \mathrm{N}(4, 1.5^2)$ er uavhengige.

$P(X \leq 0.5) = P(\frac{X-0}{2} \leq \frac{0.5-0}{2}) = \Phi(0.25) = \underline{\underline{0.599}}$

Då X og Y er uavhengige er:

$P(X \geq 0.5 | Y \leq 0.5) = P(X \geq 0.5) = 1 - P(X \leq 0.5) = 1 - 0.599 = \underline{\underline{0.401}}$

$Z = X + Y$ er ein lineærkobinasjon av to uavhengige normalfordelte stokastiske variable, og er derfor sjølv normalfordelt med:

$\mathrm{E}(Z) = \mathrm{E}(X) + \mathrm{E}(Y) = 0 + 4 = \underline{\underline{4}}$

$\mathrm{Var}(Z) = \mathrm{Var}(X) + \mathrm{Var}(Y) = 2^2 + 1.5^2 = 6.25 = \underline{\underline{2.5^2}}$

$P(Z \leq 0.5) = P(\frac{Z-4}{2.5} \leq \frac{0.5-4}{2.5}) = \Phi(-1.4) = 1 - \Phi(1.4) = 1 - 0.919 = \underline{\underline{0.081}}$

Oppgåve 2

(Merk: I følgje oppgåveteksten skal konfidensintervallet *utleias*, ikkje berre setjas opp!)

Vi har at $X_1, \ldots, X_n$ er u.i.f. $\mathrm{N}(\mu_1, \sigma_0^2)$ og at $Y_1, \ldots, Y_m$ er u.i.f. $\mathrm{N}(\mu_2, \sigma_0^2)$, og også at alle X_i-ane er uavhengige av alle Y_j-ane, $i = 1, \ldots, n$, $j = 1, \ldots, m$. Forventningsverdiane μ_1 og μ_2 er ukjende, medan variansen σ_0^2 er felles og kjend.

Naturleg estimator: $\hat{\mu_1} - \hat{\mu_2} = \bar{X} - \bar{Y}$

Då estimatoren er ein lineærkombinasjon av uavhengige normalfordelte variable er han sjølv normalfordelt med:

$$\begin{aligned}
\mathrm{E}(\hat{\mu_1} - \hat{\mu_2}) &= \mathrm{E}(\bar{X}) - \mathrm{E}(\bar{Y}) = \mu_1 - \mu_2 \\
\mathrm{Var}(\hat{\mu_1} - \hat{\mu_2}) &= \mathrm{Var}(\bar{X}) + \mathrm{Var}(\bar{Y}) = \frac{\sigma_0^2}{n} + \frac{\sigma_0^2}{m}.
\end{aligned}$$

D.v.s:

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sigma_0 \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim \mathrm{N}(0,1)$$

som gjev:

$$\begin{aligned}
P\left(-z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sigma_0 \sqrt{\frac{1}{n} + \frac{1}{m}}} \leq z_{\frac{\alpha}{2}}\right) &= 1 - \alpha \\
P\left(\bar{X} - \bar{Y} - z_{\frac{\alpha}{2}} \sigma_0 \sqrt{\frac{1}{n} + \frac{1}{m}} \leq \mu_1 - \mu_2 \leq \bar{X} - \bar{Y} + z_{\frac{\alpha}{2}} \sigma_0 \sqrt{\frac{1}{n} + \frac{1}{m}}\right) &= 1 - \alpha
\end{aligned}$$

D.v.s. at vi får $(1 - \alpha)100\%$ konfidensintervall ved:

$$\left[\bar{X} - \bar{Y} - z_{\frac{\alpha}{2}} \sigma_0 \sqrt{\frac{1}{n} + \frac{1}{m}}, \bar{X} - \bar{Y} + z_{\frac{\alpha}{2}} \sigma_0 \sqrt{\frac{1}{n} + \frac{1}{m}}\right]$$

For å få numerisk svar set vi inn talverdiane: $\bar{x} = 28.80$, $\bar{y} = 26.07$, $\sigma_0 = 2$, $m = n = 10$ og $z_{0.025} = 1.96$. Får då eit 95%-konfidensintervall på:

$\underline{\underline{[0.977, 4.483]}}$

Oppgåve 3

a) $X \sim bin(n,p)$ fordi:

- Undersøker n uavhengige delar av DNA-strukturen.
- Finn for kvar del ut om denne delen av DNA-stukturen er samanfallande eller ikkje.
- Sannsynet for samanfallande er det same for alle delane ($P(\text{samanfell}) = p = 0.15$).

$$\begin{aligned}
P(X=2) &= \binom{5}{2} 0.15^2 (1-0.15)^{5-2} = \underline{\underline{0.138}} \\
P(X \geq 2) &= 1 - P(X<2) = 1 - P(X \leq 1) = 1 - 0.835 = \underline{\underline{0.165}} \\
P(X=2|X\geq 2) &= \frac{P(X=2 \cap X \geq 2)}{P(X \geq 2)} = \frac{P(X=2)}{P(X\geq 2)} \\
&= \frac{0.138}{0.165} = \underline{\underline{0.836}}
\end{aligned}$$

b)

$$\begin{aligned}
P(\text{Type-I-feil}) &= P(\text{forkaste } H_0|H_0) \\
&= P(X=5|p=0.15) = 0.15^5 = \underline{\underline{0.000076}} \\
P(\text{Type-II-feil}) &= P(\text{ikkje forkaste } H_0|H_1) \\
&= P(X<5|p=1) = \underline{\underline{0}}
\end{aligned}$$

Ser så på det generelle uttrykket for sannsynet for type-I-feil når vi har n forsøk. Finn derfrå kor stor n må vere for å oppnå ønska sannsyn for type-I-feil.

$$\begin{aligned}
P(\text{Type-I-feil}) = P(X=5|p=0.15) = 0.15^n &< 0.000001 \\
n \ln(0.15) &> \ln(0.000001) \\
n &> \frac{\ln(0.000001)}{\ln(0.15)} = 7.28
\end{aligned}$$

Minst $\underline{\underline{\text{8 delar}}}$ frå DNA-strukturen må undersøkast dersom sannsynet for type-I-feil skal vere mindre enn 0.000001.

Oppgåve 4

a)

$P(X>10)=1-P(X\leq 10)=1-F(10)=1-(1-e^{-\frac{\sqrt{10}}{2}})=\underline{\underline{0.206}}$

$$\begin{aligned}
P(X>20|X>10) &= \frac{P(X>20\cap X>10)}{P(X>10)}=\frac{P(X>20)}{P(X>10)}\\
&= \frac{1-F(20)}{1-F(10)}=\frac{e^{-\frac{\sqrt{20}}{2}}}{e^{-\frac{\sqrt{10}}{2}}}=\underline{\underline{0.519}}
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}(X) &= \int_{-\infty}^{\infty} xf(x)dx=\int_{0}^{\infty}\frac{1}{2\theta}\sqrt{x}e^{-\frac{\sqrt{x}}{\theta}}dx \stackrel{formel}{=} \frac{1}{2\theta}2\theta^{2\frac{1}{2}+2}\Gamma(2\cdot\frac{1}{2}+2)\\
&= \frac{1}{\theta}\theta^3\Gamma(3)=\underline{\underline{2\theta^2}}
\end{aligned}$$

b)

$U=\min(X_A,X_B)$, og X_A og X_B er uavhengige.

$$\begin{aligned}
F_U(u)=P(U\leq u) &= 1-P(U>u)=1-P(min(X_A,X_B)>u)\\
&= 1-P(X_A>u\cap X_B>u)\stackrel{uavh.}{=}1-P(X_A>u)P(X_B>u)\\
&= 1-(1-F_{X_A}(u))(1-F_{X_B}(u))=1-e^{-\frac{\sqrt{u}}{\theta_A}}e^{-\frac{\sqrt{u}}{\theta_B}}\\
&= 1-e^{-\sqrt{u}(\frac{1}{\theta_A}+\frac{1}{\theta_B})}
\end{aligned}$$

$$\begin{aligned}
f_U(u) &= F_U'(u)=\frac{1}{2}u^{-\frac{1}{2}}(\frac{1}{\theta_A}+\frac{1}{\theta_B})e^{-\sqrt{u}(\frac{1}{\theta_A}+\frac{1}{\theta_B})}\\
&= \underline{\underline{\frac{1}{2}(\frac{1}{\theta_A}+\frac{1}{\theta_B})\frac{1}{\sqrt{u}}e^{-\sqrt{u}(\frac{1}{\theta_A}+\frac{1}{\theta_B})}\ ,\ u>0}}
\end{aligned}$$

Gjenkjenner dette som same fordelinga som X og Y kjem frå, men med $\theta=(\frac{1}{\theta_A}+\frac{1}{\theta_B})^{-1}$.
Følgjer då frå a) at:
$\mathrm{E}(U)=2(\frac{1}{\theta_A}+\frac{1}{\theta_B})^{-2}=\underline{\underline{2(\frac{\theta_A\theta_B}{\theta_A+\theta_B})^2}}$

c)

Finn sannsynsmaksimeringsestimatoren (SME):

$$\begin{aligned}
L(\theta) &= f(x_1, \ldots, x_n; \theta) \\
&= \prod_{i=1}^{n} f(x_i; \theta) \\
&= \prod_{i=1}^{n} \frac{1}{2\theta} \frac{1}{\sqrt{x_i}} e^{-\frac{\sqrt{x_i}}{\Theta}} \\
&= (\frac{1}{2\theta})^n (\prod_{i=1}^{n} \frac{1}{\sqrt{x_i}}) e^{-\frac{1}{\theta}\sum_{i=1}^{n}\sqrt{x_i}}
\end{aligned}$$

$$l(\theta) = \ln L(\theta) = -n\ln(2\theta) - \frac{1}{2}\sum_{i=1}^{n} \ln x - \frac{1}{\theta}\sum_{i=1}^{n}\sqrt{x_i}$$

$$\begin{aligned}
\frac{\partial l(\theta)}{\partial \theta} = -\frac{2n}{2\theta} + \frac{1}{\theta}\sum_{i=1}^{n}\sqrt{x_i} &= 0 \\
n\theta &= \sum_{i=1}^{n}\sqrt{x_i}
\end{aligned}$$

D.v.s. SME blir: $\underline{\underline{\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}\sqrt{X_i}}}$

$$\mathrm{E}(\sqrt{X}) = \int_{-\infty}^{\infty} \sqrt{x} f(x) dx = \frac{1}{2\theta}\int_{0}^{\infty} e^{-\frac{\sqrt{x}}{\theta}} dx \stackrel{formel}{=} \frac{1}{2\theta} 2\theta^2 \Gamma(2) = \theta$$

D.v.s.

$$\mathrm{E}(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(\sqrt{X}) = \frac{1}{n}\sum_{i=1}^{n} \theta = \theta$$

Altså er estimatoren forventningsrett.

d)

Skal finne fordelinga til $Z_i = \frac{2\sqrt{X_i}}{\theta}$. Finn først X_i uttrykt ved Z_i: $X_i = (\frac{\theta Z_i}{2})^2 = h(Z_i)$. Får då at $h'(Z_i) = (\frac{\theta}{2})^2 2Z_i = \frac{\theta^2}{2} Z_i$. Vi finn nå tettleiken til Z_i ved transformasjonsformelen:

$$\begin{aligned} f_{Z_i}(z) &= f_{X_i}(h(z))|h'(z)| \\ &= \frac{1}{2\theta}\frac{1}{\sqrt{(\frac{\theta z}{2})^2}} e^{-\frac{1}{\theta}\sqrt{(\frac{\theta z}{2})^2}} \frac{\theta^2}{2} z \\ &= \frac{1}{\theta^2 z} e^{-\frac{z}{2}} \frac{\theta^2}{2} z \\ &= \frac{1}{2} e^{-\frac{z}{2}} \end{aligned}$$

som er tettleiken til ein χ^2_2-fordelt variabel.

$$\frac{2n\hat{\theta}}{\theta} = \frac{2}{\theta}\sum_{i=1}^{n}\sqrt{X_i} = \sum_{i=1}^{n}\frac{2\sqrt{X_i}}{\theta}$$

Der $\frac{2\sqrt{X_i}}{\theta}$ er χ^2-fordelt med 2 fridomsgrader. Har at summen av uavhengige χ^2-fordelte variable er χ^2-fordelt. Talet på fridomsgrader er lik summen av fridomsgradene til variablane. Dette gjev at:

$$\frac{2n\hat{\theta}}{\theta} \sim \chi^2_{2n}$$

som skulle visast.

e)

Tar utgangspunkt i at $\frac{2n\hat{\theta}}{\theta} \sim \chi^2_{2n}$ som gjev:

$$\begin{aligned} P\left(\chi^2_{1-\alpha/2,2n} \le \frac{2n\hat{\theta}}{\theta} \le \chi^2_{\alpha/2,2n}\right) &= 1-\alpha \\ P\left(\frac{1}{\chi^2_{1-\alpha/2,2n}} \le \frac{\theta}{2n\hat{\theta}} \le \frac{1}{\chi^2_{\alpha/2,2n}}\right) &= 1-\alpha \\ P\left(\frac{2n\hat{\theta}}{\chi^2_{1-\alpha/2,2n}} \le \theta \le \frac{2n\hat{\theta}}{\chi^2_{\alpha/2,2n}}\right) &= 1-\alpha \end{aligned}$$

D.v.s. eit $(1-\alpha)100\%$-konfidensintervall for θ blir:

$$\underline{\underline{\left[\frac{2n\hat{\theta}}{\chi^2_{\alpha/2,2n}}, \frac{2n\hat{\theta}}{\chi^2_{1-\alpha/2,2n}}\right]}}$$

For å få eit numerisk intervall set vi inn for: $\alpha = 0.05$, $\chi^2_{0.025,40} = 59.34$, $\chi^2_{0.975,40} = 24.43$ og $n\hat{\theta} = \sum_{i=1}^n \sqrt{x_i} = 29.902$:

$$\left[\frac{2 \cdot 29.902}{59.34}, \frac{2 \cdot 29.902}{24.43}\right] = \underline{\underline{[1.01, 2.45]}}$$

Oppgåve 5

I denne oppgåva er det fleire alternative løysningar, tre løysningar er gitt her:

Alternativ 1: Finn først momentgenerarande funksjon (MGF) til den aktuelle gammafordelinga:

$$\begin{aligned} M_Y(t) = \mathrm{E}(e^{tY}) &= \int_0^\infty e^{ty}\lambda^2 y e^{-\lambda y} dy \\ &= \int_0^\infty \lambda^2 y e^{-(\lambda - t)y} dy \\ &= \frac{\lambda^2}{(\lambda - t)^2} \int_0^\infty (\lambda - t)^2 y e^{-(\lambda - t)y} dy \\ &= (\frac{\lambda}{\lambda - t})^2 \end{aligned}$$

For $Y = X_1 + X_2$ har vi at:

$$M_Y(t) = M_{X_1+X_2}(t) = M_{X_1}(t) M_{X_2}(t)$$

Frå tabellen har vi at $M_{X_i}(t) = \frac{\lambda}{\lambda - t}$, d.v.s. $M_Y(t) = (\frac{\lambda}{\lambda - t})^2$.

Då $Y = X_1 + X_2$ har same MGF som gammafordalinga med parameter $\alpha = 2$ og $\beta = \frac{1}{\lambda}$ har vi vist at Y har denne gammafordelinga.

Alternativ 2:

Finn først fordelingsfunksjonen:

$$
\begin{aligned}
F_Y(y) &= P(Y \le y) = P(X_1 + X_2 \le y) \\
&= \int\int_{x_1+x_2 \le y} f(x_1, x_2) dx_1 dx_2 \\
&\overset{uavh.}{=} \int\int_{x_1+x_2 \le y} \lambda e^{-\lambda x_1} \lambda e^{-\lambda x_2} dx_1 dx_2 \\
&= \int_0^y \int_0^{y-x_2} \lambda e^{-\lambda x_1} \lambda e^{-\lambda x_2} dx_1 dx_2 \\
&= \int_0^y \lambda e^{-\lambda x_2}(1 - e^{-\lambda(y-x_2)}) dx_2 \\
&= \int_0^y (\lambda e^{-\lambda x_2} - \lambda e^{-\lambda y}) dx_2 \\
&= 1 - e^{-\lambda y} - y\lambda e^{-\lambda y}
\end{aligned}
$$

D.v.s. tettleiken er:

$$
\begin{aligned}
f(y) &= F_Y'(y) = \lambda e^{-\lambda y} - \lambda e^{-\lambda y} + y\lambda^2 e^{-\lambda y} \\
&= \underline{\underline{\lambda^2 y e^{-\lambda y}}},\ y > 0
\end{aligned}
$$

som skulle visast.

Alternativ 3:

$$
\begin{aligned}
F_Y(y) &= P(Y \le y) = P(\text{talet på hendingar i } [0, y] \ge 2) \\
&= P(Z \ge 2) \qquad \text{der } Z \sim P_0(\lambda y) \\
&= 1 - P(Z \le 1) \\
&= 1 - \frac{(\lambda y)^0}{0!} e^{-\lambda y} - \frac{(\lambda y)!}{1!} e^{-\lambda y} \\
&= 1 - e^{-\lambda y} - \lambda y e^{-\lambda y}
\end{aligned}
$$

D.v.s. tettleiken er:

$$
\begin{aligned}
f(y) &= F_Y'(y) = \lambda e^{-\lambda y} - \lambda e^{-\lambda y} + \lambda^2 y e^{-\lambda y} \\
&= \underline{\underline{\lambda^2 y e^{-\lambda y}}},\ y > 0
\end{aligned}
$$

som skulle visast.