Løsningsforslag til eksamen i

STK1100 8. juni 2006

Oppgave 1

a) Vi har at $Z = (Y - \mu)/\sigma$ er standardnormalfordelt. For $x > 0$ er derfor den kumulative fordelingen til X gitt ved

$$\begin{aligned}
F(x) &= P(X \leq x) = P(\log X \leq \log x) = P(Y \leq \log x) \\
&= P\left(\frac{Y-\mu}{\sigma} \leq \frac{\log x - \mu}{\sigma}\right) = P\left(Z \leq \frac{\log x - \mu}{\sigma}\right) \\
&= \Phi\left(\frac{\log x - \mu}{\sigma}\right)
\end{aligned}$$

b) For $\mu = 1.3$ og $\sigma^2 = 0.8$ finner vi at

$$P(X > 10) = 1 - F(10) = 1 - \Phi\left(\frac{\log 10 - 1.3}{\sqrt{0.8}}\right) = 1 - \Phi(1.12) = 1 - 0.869 = 0.131$$

Ca. 13 % av norske menn i tjueårene (som ikke er avholdende) drikker minst 10 liter ren alkohol i løpet av ett år.

c) Vi har at $\Phi'(z)$ er tettheten i standardnormalfordelingen, dvs.

$$\Phi'(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

For $x > 0$ er dermed sannsynlighetstettheten til X gitt ved

$$f(x) = F'(x) = \Phi'\left(\frac{\log x - \mu}{\sigma}\right)\frac{1}{\sigma x} = \frac{1}{\sqrt{2\pi}\,\sigma x} e^{-(\log x - \mu)^2/(2\sigma^2)}$$

d) Medianen $x_{0.50}$ er gitt ved at $F(x_{0.50}) = 0.50$. Siden $\Phi(0) = 0.50$, gir punkt a at $x_{0.50}$ er gitt ved ligningen $(\log x_{0.50} - \mu)/\sigma = 0$. Det gir $\log x_{0.50} = \mu$, slik at $x_{0.50} = e^{\mu}$.

e) Vi har at $Y = \log X \sim N(\mu, \sigma^2)$, og at $M_Y(t) = E(e^{tY}) = e^{\mu t + \sigma^2 t^2/2}$. Vi får dermed at

$$E(X) = E(e^Y) = M_Y(1) = e^{\mu + \sigma^2/2}$$

På samme måte finner vi at

$$E(X^2) = E(e^{2Y}) = M_Y(2) = e^{2\mu + 4\sigma^2/2},$$

slik at

$$\mathrm{Var}(X) = E(X^2) - (EX)^2 = e^{2\mu+4\sigma^2/2} - e^{2\mu+2\sigma^2/2} = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right)$$

f) For $\mu = 1.3$ og $\sigma^2 = 0.8$ finner vi at

$$\begin{aligned} x_{0.50} &= e^{1.3} = 3.7 \\ E(X) &= e^{1.3+0.8/2} = 5.5 \end{aligned}$$

At forventningsverdien er 5.5, betyr at det gjennomsnittlige forbruket av alkohol blant ikke-avholdende norske menn i tjueårene er 5.5 liter ren alkohol per år. At medianen er 3.7, betyr at halvparten av ikke-avholdende norske menn i tjueårene drikker mindre enn 3.7 liter ren alkohol per år, og halvparten drikker mer enn det. Siden gjennomsnittsverdien blir trukket opp av de som drikker mye, er medianen best egnet til å beskrive alkoholforbruket for en "typisk" ikke-avholdende norsk mann i tjueårene.

Oppgave 2

a) Vi bestemmer k slik at $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx\,dy = 1$. Nå har vi at

$$\begin{aligned} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx\,dy &= \int_0^{\infty}\int_x^{\infty} k\,e^{-(x+y)}\,dy\,dx \\ &= k\int_0^{\infty} e^{-x}\left(\int_x^{\infty} e^{-y}\,dy\right)dx = k\int_0^{\infty} e^{-x}\,e^{-x}\,dx \\ &= k\int_0^{\infty} e^{-2x}\,dx = \frac{k}{2} \end{aligned}$$

Det gir $k = 2$.

b) Vi finner at

$$\begin{aligned} P(2X \le Y) &= \int_0^{\infty}\int_{2x}^{\infty} 2\,e^{-(x+y)}\,dy\,dx = 2\int_0^{\infty} e^{-x}\left(\int_{2x}^{\infty} e^{-y}\,dy\right)dx \\ &= 2\int_0^{\infty} e^{-x}\,e^{-2x}\,dx = 2\int_0^{\infty} e^{-3x}\,dx = \frac{2}{3} \end{aligned}$$

c) For $x > 0$ blir den marginale fordelingen til X:

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy = \int_x^{\infty} 2\,e^{-(x+y)}\,dy$$

$$= \quad 2\,e^{-x} \int_x^\infty e^{-y}\,dy = 2\,e^{-x}\,e^{-x} = 2e^{-2x}$$

Oppgave 3

a) Sannsynligheten er 1/2 for at barnet vil arve a fra mor. Vi skriver kort $P(a \text{ fra mor}) = 1/2$. Tilsvarende har vi at

$$P(A \text{ fra mor}) = 1/2, \quad P(a \text{ fra far}) = 1/2 \quad \text{og} \quad P(A \text{ fra far}) = 1/2$$

Siden arven fra de to foreldrene er uavhengig, får vi at

$$P(\text{barnet får genotypen } aa) = P(a \text{ fra mor og } a \text{ fra far})$$
$$= P(a \text{ fra mor}) \cdot P(a \text{ fra far}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

På tilsvarende måte får vi at

$$P(\text{barnet får genotypen } Aa) = P(A \text{ fra mor og } a \text{ fra far}) + P(a \text{ fra mor og } A \text{ fra far})$$
$$= P(A \text{ fra mor}) \cdot P(a \text{ fra far}) + P(a \text{ fra mor}) \cdot P(A \text{ fra far})$$
$$= \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$$

b) Sannsynligheten er 0.05 for at kvinnen har genotypen Aa. Vi skriver kort $P(\text{kvinnen er } Aa) = 0.05$. Tilsvarende gir opplysningene i oppgaven at

$$P(\text{kvinnen er } AA) = 0.95,\ P(\text{mannen er } Aa) = 0.05 \text{ og } P(\text{ mannen er } AA) = 0.95$$

Siden genotypene for kvinnen og mannen er uavhengige får vi at

$$P(\text{begge er bærere av sykdomsgenet}) = P(\text{kvinnen er } Aa \text{ og mannen er } Aa)$$
$$= P(\text{kvinnen er } Aa) \cdot P(\text{mannen er } Aa)$$
$$= 0.05 \cdot 0.05 = 0.0025$$

På tilsvarende måte får vi at

$$P(\text{nøyaktig én er bærer av sykdomsgenet})$$
$$= P(\text{kvinnen er } Aa \text{ og mannen er } AA) + P(\text{kvinnen er } AA \text{ og mannen er } Aa)$$
$$= P(\text{kvinnen er } Aa) \cdot P(\text{mannen er } AA) + P(\text{kvinnen er } AA) \cdot P(\text{mannen er } Aa)$$
$$= 0.05 \cdot 0.95 + 0.95 \cdot 0.05 = 0.095$$

Nedenfor får vi også bruk for at

$$P(\text{ingen er bærere av sykdomsgenet}) = P(\text{kvinnen er } AA \text{ og mannen er } AA)$$

$$= P(\text{kvinnen er } AA) \cdot P(\text{mannen er } AA)$$

$$= 0.95 \cdot 0.95 = 0.9025$$

c) Setningen om total sannsynlighet gir at

$$P(\text{barn får cystisk fibrose}) = P(\text{barn får genotypen } aa)$$

$$= P(\text{begge foreldrene er bærere}) \cdot P(\text{barn får genotypen } aa \,|\, \text{ begge bærere})$$

$$+P(\text{én av foreldrene er bærer}) \cdot P(\text{barn får genotypen } aa \,|\, \text{ én er bærer})$$

$$+P(\text{ingen av foreldrene er bærer}) \cdot P(\text{barn får genotypen } aa \,|\, \text{ingen er bærer})$$

$$= 0.0025 \cdot \frac{1}{4} + 0.095 \cdot 0 + 0.9025 \cdot 0 = 0.000625$$

Sannsynligheten er ca. 0.06% for at barnet vil lide av cystisk fibrose.

Hvis nøyaktig én av foreldrene er bærer av sykdomsgenet (har genotypen Aa), finner vi på lignende måte som i punkt a at sannsynligheten er 1/2 for at barnet blir en frisk bærer. Setningen om total sannsynlighet gir dermed at

$$P(\text{barn blir frisk bærer av sykdomsgenet}) = P(\text{barn får genotypen } Aa)$$

$$= P(\text{begge foreldrene er bærere}) \cdot P(\text{barn får genotypen } Aa \,|\, \text{ begge bærere})$$

$$+P(\text{én av foreldrene er bærer}) \cdot P(\text{barn får genotypen } Aa \,|\, \text{ én er bærer})$$

$$+P(\text{ingen av foreldrene er bærer}) \cdot P(\text{barn får genotypen } Aa \,|\, \text{ingen er bærer})$$

$$= 0.0025 \cdot \frac{1}{2} + 0.095 \cdot \frac{1}{2} + 0.9025 \cdot 0 = 0.04875$$

Sannsynligheten er ca. 4.9% for at barnet blir en frisk bærer av sykdomsgenet.

d) Setningen om total sannsynlighet gir at

$$P(\text{tre friske barn}) = P(\text{ingen av barna har genotypen } aa)$$

$$= P(\text{begge er bærere}) \cdot P(\text{ingen av barna har genotypen } aa \,|\, \text{ begge bærere})$$

$$+P(\text{én er bærer}) \cdot P(\text{ingen av barna har genotypen } aa \,|\, \text{ én er bærer})$$

$$+P(\text{ingen er bærer}) \cdot P(\text{ingen av barna har genotypen } aa \,|\, \text{ingen er bærer})$$

$$= 0.0025 \cdot \left(1 - \frac{1}{4}\right)^3 + 0.095 \cdot 1 + 0.9025 \cdot 1 = 0.9986$$

Av definisjonen av betinget sannsynlighet og produktsetningen (eller direkte av Bayes' setning) finner vi dermed at

$$
\begin{aligned}
&P(\text{ingen av foreldrene er bærere} \,|\, \text{tre friske barn}) \\
&\qquad = \frac{P(\text{ingen er bærere}) \cdot P(\text{tre friske barn} \,|\, \text{ingen er bærere})}{P(\text{tre friske barn})} \\
&\qquad = \frac{0.9025 \cdot 1}{0.9986} = 0.904 \qquad ,
\end{aligned}
$$

Gitt at ekteparet har fått tre friske barn, er sannsynligheten 90.4% for at hverken kvinnen eller mannen er bærer av sykdomsgenet.