UiO : Universitetet i Oslo

Skriftlig hjemmeeksamen i STK1100

Vår 2020

Emnekode: STK1100

Kandidatnummer: 15220

Antall sider: 21

Innleveringsdato: 09.06.2020

Oppgave 1

a. Vi definerer A og $P1$ som

A – Personen har antistoffer i blodet

$P1$ – Testen $T1$ gir et positivt resultat

Sannsynlighet for at en person har antistoffer i blodet er

$$P(A) = 0.01 \tag{1.a.I}$$

Sannsynligheten for at testen $T1$ gir et positivt resultat, gitt at man faktisk har antistoffer i blodet

$$P(P1|A) = 0.955 \tag{1.a.II}$$

Sannsynligheten for at testen $T1$ gir et negativt resultat, gitt at man ikke har antistoffer i blodet

$$P(P1'|A') = 0.98 \tag{1.a.III}$$

Vi skal finne sannsynligheten for at en tilfeldig valgt testperson får et positivt resultat av test $T1$, altså $P(P1)$. Vi finner først sannsynligheten for at en person ikke har antistoffer i blodet

$$P(A') = 1 - P(A) \tag{1.a.IV}$$
$$= 1 - 0.01 \tag{1.a.V}$$
$$= 0.99 \tag{1.a.VI}$$

Videre finner vi sannsynligheten for at testen $T1$ gir et positivt resultat, gitt at man ikke har antistoffer i blodet

$$P(P1|A') = 1 - P(P1'|A') \tag{1.a.VII}$$
$$= 1 - 0.98 \tag{1.a.VIII}$$
$$= 0.02 \tag{1.a.IX}$$

Vi kan nå finne sannsynligheten $P(P1)$ ved å bruke setningen om total sannsynlighet

$$P(P1) = P(P1|A) \cdot P(A) + P(P1|A') \cdot P(A') \qquad (1.a.X)$$

$$= 0.955 \cdot 0.01 + 0.02 \cdot 0.99 \qquad (1.a.XI)$$

$$= 0.02935 \approx 2.9\% \qquad (1.a.XII)$$

Sannsynligheten for at en tilfeldig valgt testperson får et positivt resultat av test $T1$ er

$$P(P1) = 0.02935 \approx 2.9\%$$

b. Vi skal nå beregne sannsynlighet for at en tilfeldig valgt testperson har antistoffer i blodet, gitt en positiv test $T1$, altså $P(A|P1)$. Videre skal vi kommentere resultatet. Vi starter med å finne $P(A|P1)$ ved å bruke Bayes setning

$$P(A|P1) = \frac{P(P1|A) \cdot P(A)}{P(P1)} \qquad (1.b.I)$$

Vi setter inn sannsynlighetene fra a, og får

$$P(A|P1) = \frac{0.955 \cdot 0.01}{0.02935} \qquad (1.b.II)$$

$$= 0.32538 \approx 32.5\% \qquad (1.b.III)$$

Pga. høyere sensitivitet og lavere spesifisitet vil denne utregninger gir betydelige mindre verdier av sannsynligheten for at en tilfeldig valgt testperson har antistoffer i blodet enn det testen $T2$ ville ha gjord.

Sannsynlighet for at en tilfeldig valgt testperson har antistoffer i blodet, gitt en positiv test $T1$ er

$$P(A|P1) = 0.32538 \approx 32.5\%$$

c. Vi skal nå beregne sannsynlighet for at en tilfeldig valgt testperson har antistoffer i blodet, gitt en positiv test $T2$, altså $P(A|P2)$. Videre skal vi sammenligne resultatene med resultatene i b og kommentere de. Vi starter med å definere $P2$

$P2$ – Testen $T2$ gir et positivt resultat

Sannsynligheten for at testen $T2$ gir et positivt resultat, gitt at man faktisk har antistoffer i blodet

$$P(P2|A) = 0.942 \tag{1.c.I}$$

Sannsynligheten for at testen $T2$ gir et negativt resultat, gitt at man ikke har antistoffer i blodet

$$P(P2'|A') = 0.998 \tag{1.c.II}$$

Videre finner vi sannsynligheten for at testen $T2$ gir et positivt resultat, gitt at man ikke har antistoffer i blodet

$$P(P2|A') = 1 - P(P2'|A') \tag{1.c.III}$$
$$= 1 - 0.998 \tag{1.c.IV}$$
$$= 0.002 \tag{1.c.V}$$

Vi finner nå sannsynligheten for at en tilfeldig valgt testperson får et positivt resultat av test $T2$, altså $P(P2)$ ved å bruke setningen om total sannsynlighet og resultatene fra a

$$P(P2) = P(P2|A) \cdot P(A) + P(P2|A') \cdot P(A') \tag{1.c.VI}$$
$$= 0.942 \cdot 0.01 + 0.002 \cdot 0.99 \tag{1.c.VII}$$
$$= 0.0114 \approx 1.1\% \tag{1.c.VIII}$$

Videre beregner vi sannsynligheten for at en tilfeldig valgt testperson har antistoffer i blodet, gitt en positiv test $T2$ ved å bruke Bayes setning

$$P(A|P2) = \frac{P(P2|A) \cdot P(A)}{P(P2)} \tag{1.c.IX}$$

Vi setter inn sannsynlighetene fra a, og får

$$P(A|P2) = \frac{0.942 \cdot 0.01}{0.0114} \tag{1.c.X}$$
$$= 0.82631 \approx 82.6\% \tag{1.c.XI}$$

Som vi også nevnte i siste deloppgave, så har denne sannsynligheten blitt betydelig større enn det vi har beregnet for testen $T1$. Dette skyldes høy spesifisitet og lav sensitivitet. Her ser vi at testen $T2$ er mer nøyaktige enn testen $T1$, samtidig som det er viktig å understreke at i slike tilfeller som denne er det mer verdig å fokusere (velge) på testene som kanskje gir mer falsk-positiv resultat, enn de testene som gir mer falsk-negativ resultat. Siden da blir det lettere å forhindre videre spredning av viruset.

Sannsynlighet for at en tilfeldig valgt testperson har antistoffer i blodet, gitt en positiv test $T2$ er

$$P(A|P2) = 0.82631 \approx 82.6\%$$

d. Vi skal nå finne sannsynligheten for å ikke ha antistoffer i blodet, gitt at man har testet positivt. Vi skal finne denne sannsynligheten både for testen $T1$ og $T2$, og senere kommentere. Vi starter med å finne sannsynligheten for å ikke ha antistoffer i blodet, gitt at man har testet positivt for testen $T1$, altså $P(A'|P1)$. Vi bruker Bayes setning og resultatene fra a og b og får

$$P(A'|P1) = \frac{P(P1|A') \cdot P(A')}{P(P1)} \tag{1.d.I}$$
$$\tag{1.d.II}$$
$$= \frac{0.02 \cdot 0.99}{0.029} \tag{1.d.III}$$

$$= 0.68275 \approx 68{,}3\%$$

Videre finner vi sannsynligheten for å ikke ha antistoffer i blodet, gitt at man har testet positivt for testen $T2$, altså $P(A'|P2)$. Vi bruker Bayes setning og resultatene fra a og c, og får

$$P(A'|P2) = \frac{P(P2|A') \cdot P(A')}{P(P2)} \quad (1.d.IV)$$

$$= \frac{0.002 \cdot 0.99}{0.0114} \quad (1.d.V)$$

$$= 0.17368 \approx 17{,}4\% \quad (1.d.VI)$$

Sannsynligheten for å ikke ha antistoffer i blodet, gitt at man har testet positivt for testen $T1$ er

$$P(A'|P1) = 0.68275 \approx 68{,}3\%$$

Og sannsynligheten for å ikke ha antistoffer i blodet, gitt at man har testet positivt for testen $T2$ er

$$P(A'|P2) = 0.17368 \approx 17{,}4\%$$

e. Vi skal nå beregne på hvor høy måtte spesifisiteten ha vært for test $T2$ for at sannsynligheten i punkt d skulle bli mindre enn 5%. Spesifisitet er sannsynligheten for at testen gir et negativt resultat, gitt at man ikke har antistoffer i blodet, altså $P(P2'|A')$. Vi bruker resultatene fra (1.c.III), at

$$P(P2|A') = 1 - P(P2'|A') \quad (1.e.I)$$

Videre setter vi opp ligningen med $P(P2|A')$ som den ukjente . Vi bruker (1.d.IV) og setter at $P(A'|P2)$ skal bli mindre enn 5% og får

$$P(A'|P2) = \frac{P(P2|A') \cdot P(A')}{P(P2)} \quad (1.e.II)$$

$$0.05 > \frac{P(P2|A') \cdot P(A')}{P(P2)} \quad (1.e.III)$$

$$0.05 > \frac{(1 - P(P2'|A')) \cdot P(A')}{P(P2)} \tag{1.e.IV}$$

Vi løser ulikheten med hensyn på $P(P2|A')$

$$0.05 \cdot \frac{P(P2)}{P(A')} > 1 - P(P2'|A') \tag{1.e.V}$$

$$P(P2'|A') > 1 - 0.05 \cdot \frac{P(P2)}{P(A')} \tag{1.e.VI}$$

Vi bruker resultatene fra a og c, og får

$$\begin{aligned} P(P2'|A') &> 1 - 0.05 \cdot \frac{P(P2)}{P(A')} && \text{(1.e.VII)} \\ && & \text{(1.e.VIII)} \\ &> 1 - 0.05 \cdot \frac{0.0114}{0.99} && \text{(1.e.IX)} \\ &> 0.99942 \end{aligned}$$

For at sannsynlighet i punkt d skal bli mindre enn 5% må spesifisiteten $P(P2'|A')$ bli

$$P(P2'|A') > 0.99942 \approx 99.942\%$$

f. Vi skal nå finne sannsynligheten for at minst 1 av 10 tilfeldig valgte personer som testet negativt på testen $T1$ hadde likevel antistoffer i blodet. Vi finner først sannsynligheten for at personen har antistoffer i blodet, gitt at testen er negativt. Vi bruker Bayes setning og får

$$P(A|P1') = \frac{P(P1'|A)P(A)}{P(P1')} \tag{1.f.I}$$

Vi bruker resultatene fra a og finner de ukjente

$$P(P1'|A) = 1 - P(P1|A) = 0.045 \tag{1.f.II}$$

Og

$$P(P1') = 1 - P(P1) = 0.97065 \tag{1.f.III}$$

Vi setter dem inn i ligning (1.f.I) og får

$$P(A|P1') = \frac{0.045 \cdot 0.01}{0.97065} = 0.00046$$

Videre regner vi ut sannsynligheten for at personen ikke har antistoffer i blodet, gitt at testen er negativt

$$P(A'|P1') = 1 - 0.00046 = 0.99587 \qquad \text{(1.f.VI)}$$

Vi finner nå sannsynlighet for at ingen av de 10 personer har antistoffer i blodet gitt at testen er negativt

$$P(ingen) = \left(P(A'|P1')\right)^{10} = 0.95946 \qquad \text{(1.f.VII)}$$

Nå finner vi sannsynlighet for at minst 1 av 10 har antistoffer i blodet, gitt at testen er negativt

$$P(minst\ 1) = 1 - P(ingen) = 0.04054 \qquad \text{(1.f.VIII)}$$

Sannsynligheten for at minst 1 av 10 tilfeldig valgte personer som testet negativt på testen T1 hadde likevel antistoffer i blodet er

$$P(minst\ 1) = 0.04054$$

Oppgave 2

a. Vi har en simultan sannsynlighetstetthet

$$f(x,y) = \begin{cases} kx(x+y) & \text{når } 0 \le x \le 2 \text{ og } 0 \le y \le 2 \\ 0 & \text{ellers} \end{cases} \qquad \text{(2.a.I)}$$

Der k er en konstant. Vi skal vise at $k = \frac{3}{28}$. Dersom vi integrerer den simultane sannsynligheten $f_{XY}(x,y)$ over alle mulige kombinasjoner av x og y, får vi 1

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dxdy = \int_0^2\int_0^2 kx(x+y)dydx = 1 \qquad \text{(2.a.II)}$$

Vi bruker dette til å finne k. Først regner vi ut integralet

$$\int_0^2 kx(x+y)\,dy \qquad \text{(2.a.III)}$$

Vi setter konstanten k og x utenfor

$$kx\int_0^2 (x+y)\,dy \qquad \text{(2.a.IV)}$$

Videre bruker vi sumregelen for integraler

$$kx\left(\int_0^2 x\,dy + \int_0^2 y\,dy\right) \qquad \text{(2.a.V)}$$

Vi regner ut de bestemte integralene

$$kx\left([xy]_0^2 + \left[\frac{y^2}{2}\right]_0^2\right) \qquad \text{(2.a.VI)}$$

$$kx(2x+2) \qquad \text{(2.a.VII)}$$

Da får vi at

$$\int_0^2 kx(x+y)\,dy = kx(2x+2) \qquad \text{(2.a.VIII)}$$

Setter vi det inn i integralet (2.a.II), og får

$$\int_0^2\int_0^2 kx(x+y)\,dydx = \int_0^2 kx(2x+2)\,dx \qquad \text{(2.a.IX)}$$

Vi setter konstanten k utenfor

$$k\int_0^2 x(2x+2)\,dx \qquad \text{(2.a.X)}$$

Vi bruker delvis integrasjon der $u = (2x+2)$ og $v' = x$

$$k\left[x^2(x+1) - \int x^2dx\right]_0^2 \qquad \text{(2.a.XI)}$$

Siden integralet $\int x^2dx = \frac{x^3}{3}$, får vi at

$$k\left[x^2(x+1)-\frac{x^3}{3}\right]_0^2 \tag{2.a.XII}$$

$$k\left(12-\frac{8}{3}\right) \tag{2.a.XIII}$$

$$k\left(\frac{36}{3}-\frac{8}{3}\right) \tag{2.a.XIV}$$

$$\frac{28}{3}k \tag{2.a.XV}$$

Vi setter det i ligningen (2.a.II) og får

$$\int_0^2\int_0^2 kx(x+y)\;dydx = 1 \tag{2.a.XVI}$$

$$\frac{28}{3}k = 1 \tag{2.a.XVII}$$

$$k = \frac{3}{28} \tag{2.a.XVIII}$$

Dermed har vi vist at konstanten $k = \frac{3}{28}$.

b. Vi skal nå beregne sannsynligheten for at Y er større eller lik X, dvs. $P(Y \geq X)$. I denne tilfelle vil kan vi sette opp integralet på denne måten

$$P(Y \geq X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\;dxdy \tag{2.b.I}$$

$$= \int_0^2\int_0^y f_{XY}(x,y)\;dxdy \tag{2.b.II}$$

Der $f_{XY}(x,y) = kx(x+y)$, se (2.a.I)

$$P(Y \geq X) = \int_0^2\int_0^y kx(x+y)\;dxdy \tag{2.b.III}$$

Vi starter med å regne ut integralet $\int_0^y kx(x+y)\;dx$

$$\int_0^y kx(x+y)\;dx \tag{2.b.VI}$$

$$\int_0^y k(x^2 + xy)\,dx \tag{2.b.V}$$

Vi setter konstanten k utenfor

$$k\int_0^y (x^2 + xy)\,dx \tag{2.b.VI}$$

Videre bruker vi sumregelen for integraler

$$k\left(\int_0^y x^2\,dx + \int_0^y xy\,dx\right) \tag{2.b.VII}$$

Vi regner ut de bestemte integralene

$$k\left(\left[\frac{1}{3}x^3\right]_0^y + \left[\frac{1}{2}x^2y\right]_0^y\right) \tag{2.b.VIII}$$

$$k\left(\frac{y^3}{3} + \frac{y^3}{2}\right) \tag{2.b.IX}$$

Da får vi at

$$\int_0^y kx(x+y)\,dx = k\left(\frac{y^3}{3} + \frac{y^3}{2}\right) \tag{2.b.X}$$

Setter vi det inn i integralet (2.b.III), og får

$$P(Y \geq X) = \int_0^2 k\left(\frac{y^3}{3} + \frac{y^3}{2}\right)\,dy \tag{2.b.XI}$$

Vi setter konstanten k utenfor

$$P(Y \geq X) = k\int_0^2 \left(\frac{y^3}{3} + \frac{y^3}{2}\right)\,dy \tag{2.b.XII}$$

Og vi regner ut den bestemte integralet ved å bruke sumregelen for integraler

$$P(Y \geq X) = k\left(\int_0^2 \frac{y^3}{3}\,dy + \int_0^2 \frac{y^3}{2}\,dy\right) \tag{2.b.XIII}$$

$$P(Y \geq X) = k\left(\frac{1}{3}\left[\frac{y^4}{4}\right]_0^2 + \frac{1}{2}\left[\frac{y^4}{4}\right]_0^2\right) \tag{2.b.XIV}$$

$$P(Y \geq X) = k\left(2 + \frac{4}{3}\right) \tag{2.b.XV}$$

$$P(Y \geq X) = \frac{10}{3}k \tag{2.b.XVI}$$

Setter vi verdien for k som vi fant i a, så får vi

$$P(Y \geq X) = \frac{10}{3} \cdot \frac{3}{28} = \frac{5}{14} \approx 35.7\% \tag{2.b.XVI}$$

Sannsynligheten for at Y er større eller lik X, dvs. $P(Y \geq X)$ er lik

$$P(Y \geq X) = \frac{5}{14} \approx 35.7\%.$$

c. Vi skal nå finne den marginale sannsynlighetstettheten for X og Y, og bestemme om X og Y er uavhengige. Vi starter med å finne den marginale sannsynlighetstettheten for X. Vi setter opp integralet for $0 \leq x \leq 2$

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy = \int_{0}^{2} f_{XY}(x, y)\, dy \tag{2.c.I}$$

Der $f_{XY}(x, y) = kx(x + y)$, se (2.a.I)

$$f_X(x) = \int_{0}^{2} kx(x + y)\, dy \tag{2.c.II}$$

Vi setter konstanten k og x utenfor

$$f_X(x) = kx \int_{0}^{2} (x + y)\, dy \tag{2.c.III}$$

Videre bruker vi sumregelen for integraler

$$f_X(x) = kx\left(\int_{0}^{2} x\, dy + \int_{0}^{2} y\, dy\right) \tag{2.c.IV}$$

$$f_X(x) = kx\left([xy]_0^2 + \left[\frac{y^2}{2}\right]_0^2\right) \tag{2.c.V}$$

$$f_X(x) = kx(2x + 2) \tag{2.c.VI}$$

$$f_X(x) = 2kx(x+1) \tag{2.c.VII}$$

Videre finner vi den marginale sannsynlighetstettheten for Y. Vi setter opp integralet for $0 \leq y \leq 2$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx = \int_0^2 f_{XY}(x,y)\,dx \tag{2.c.I}$$

Der $f_{XY}(x,y) = kx(x+y)$, se (2.a.I)

$$f_Y(y) = \int_0^2 kx(x+y)\,dx \tag{2.c.II}$$

$$f_Y(y) = \int_0^2 k(x^2+xy)\,dx \tag{2.c.II}$$

Vi setter konstanten k utenfor

$$f_Y(y) = k\int_0^2 (x^2+xy)\,dx \tag{2.c.III}$$

Videre bruker vi sumregelen for integraler

$$f_Y(y) = k\left(\int_0^2 x^2\,dx + \int_0^2 xy\,dx\right) \tag{2.c.IV}$$

$$f_Y(y) = k\left(\left[\frac{x^3}{3}\right]_0^2 + y\left[\frac{x^2}{2}\right]_0^2\right) \tag{2.c.V}$$

$$f_Y(y) = k\left(\frac{8}{3} + 2y\right) \tag{2.c.VI}$$

Videre sjekker vi om X og Y er uavhengige ved å multiplisere de marginale sannsynlighetstetthetene sammen

$$f_X(x) \cdot f_Y(y) = 2kx(x+1) \cdot k\left(\frac{8}{3} + 2y\right) \tag{2.c.VII}$$

$$= 2k^2x(x+1)\left(\frac{8}{3} + 2y\right) \tag{2.c.VIII}$$

$$\neq\ kx(x+y) \tag{2.c.IX}$$

$$\neq\ f_{XY}(x,y) \tag{2.c.X}$$

Vi ser at $f_X(x) \cdot f_Y(y) \neq f_{XY}(x,y)$ derfor er de ikke uavhengige.

Den marginale sannsynligheten for X er

$$f_X(x) = 2kx(x+1) \qquad \text{for } 0 \leq x \leq 2$$

og for Y er

$$f_Y(y) = k\left(\frac{8}{3} + 2y\right) \qquad \text{for } 0 \leq y \leq 2$$

X og Y er ikke uavhengige siden $f_X(x) \cdot f_Y(y) \neq f_{XY}(x,y)$.

d. Vi skal nå finne sannsynlighetstettheten til $U = X + Y$. Vi starter med å gjøre et variabelskifte der $X = V$, og vi bruker dette til å finne Y

$$U = X + Y \rightarrow Y = U - V \tag{2.d.I}$$

For å finne marginaltettheten til U, finner vi først simultantettheten til U og V. Simultantettheten til U og V er definert som

$$f_{UV}(u,v) = f_{XY}(x(u,v), y(u,v))|J| \tag{2.d.II}$$

Der Jacobideterminanten $|J|$ er definert som

$$|J| = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} \tag{2.d.III}$$

Vi starter med å regne ut Jacobideterminanten $|J|$

$$|J| = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} \tag{2.d.IV}$$

$$= \begin{vmatrix} \frac{\partial(V)}{\partial u} & \frac{\partial(V)}{\partial v} \\ \frac{\partial(U-V)}{\partial u} & \frac{\partial(U-V)}{\partial v} \end{vmatrix} \tag{2.d.V}$$

$$= \begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix} \tag{2.d.VI}$$

$$= 0 \cdot (-1) - 1 \cdot 1 \tag{2.d.VII}$$

$$= -1 \tag{2.d.VIII}$$

Dermed blir simultantettheten til U og V

$$f_{UV}(u,v) = f_{XY}(x(u,v), y(u,v))|J| \qquad (2.d.IX)$$
$$= f_{XY}(v, u-v)(-1) \qquad (2.d.X)$$

Og siden $f_{XY}(v,u-v) = kv\big(v + (u-v)\big) = kvu$, så får vi at

$$f_{UV}(u,v) = kvu(-1) \qquad (2.d.XI)$$
$$= -kvu \qquad (2.d.XII)$$

Videre finner vi ut marginaltettheten til U for $0 \le u \le 2$ ved å integrere ut V fra simultantettheten

$$f_U(u) = \int_{-\infty}^{\infty} f_{UV}(u,v)\ dv \qquad (2.d.XIII)$$
$$= \int_0^2 -kvu\ dv \qquad (2.d.XIV)$$

Vi regner ut den bestemte integralet

$$f_U(u) = \int_0^2 -kvu\ dv \qquad (2.d.XV)$$

Vi setter konstantene k og u utenfor

$$f_U(u) = -ku\int_0^2 v\ dv \qquad (2.d.XVI)$$

Og får

$$f_U(u) = -ku\int_0^2 v\ dv \qquad (2.d.XVII)$$
$$= -ku\left[\frac{v^2}{2}\right]_0^2 \qquad (2.d.XVIII)$$
$$= -ku(2-0) \qquad (2.d.XIX)$$
$$= -2ku \qquad (2.d.XX)$$

Bruker vi at $k = \frac{3}{28}$, se a. Og får

$$f_U(u) = -2ku \qquad (2.d.XXI)$$

$$= -2u \cdot \frac{3}{28} \qquad (2.d.XXII)$$

$$= -\frac{3u}{14} \qquad (2.d.XXIII)$$

Sannsynlighetstettheten til $U = X + Y$ er

$$f_U(u) = -\frac{3u}{14} \qquad \text{for } 0 \leq u \leq 2.$$

Oppgave 3

a. Vi starter med å vise at $\eta = e^{\mu}$ er medianen i den log-normale fordelingen. Medianen i den log-normale fordelingen kan bli funnet ved å sette den kumulative fordelingen til X lik 0.5, altså

$$F(\eta) = 0.5$$

Vi innfører en ny stokastisk variabel Z, som defineres ved

$$Z = \frac{\mathrm{Y} - \mu}{\sigma}$$

Z er standard normaltfordelt. Vi finner nå den kumulative fordeling til X

$$\begin{aligned} F(\eta) &= P(X \leq \eta) \\ &= P(e^{Y} \leq \eta) \\ &= P(Y \leq ln\eta) \\ &= P\left(\frac{\mathrm{Y} - \mu}{\sigma} \leq \frac{ln\eta - \mu}{\sigma}\right) \\ &= P\left(Z \leq \frac{ln\eta - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{ln\,\eta - \mu}{\sigma}\right) \end{aligned}$$

For at

$$\Phi(0) = 0.5$$

Så må vi sette $\frac{ln\,\eta - \mu}{\sigma} = 0$, og vi får

$$\frac{ln\,\eta - \mu}{\sigma} = 0$$

$$ln\,\eta = \mu$$

$$\eta = e^{\mu}$$

Vi har dermed vist at $\eta = e^{\mu}$ er medianen i den log-normale fordelingen

Videre skal vi vise at forventningen $E(X) = E(e^{Y}) = \eta e^{\frac{\sigma^2}{2}}$. Siden vi har at $Y \sim N(\mu, \sigma^2)$, så blir den momentgeneredende funksjonen til Y

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

For $t = 1$ for vi at

$$\begin{aligned} M_Y(1) &= E(X) \\ &= E(e^{Y}) \\ &= e^{\mu + \frac{\sigma^2}{2}} \\ &= e^{\mu} e^{\frac{\sigma^2}{2}} \end{aligned}$$

Bruker vi at $\eta = e^{\mu}$, så får vi

$$E(X) = E(e^{Y}) = e^{\mu} e^{\frac{\sigma^2}{2}} = \eta e^{\frac{\sigma^2}{2}}$$

Dermed har vi vist at $E(X) = E(e^{Y}) = \eta e^{\frac{\sigma^2}{2}}$

b. Vi skal vise Observatoren $\bar{Y}$ er en forventningsrett estimator for μ. For at dette skal være sant må

$$E(\bar{Y}) = \mu$$

Vi vett at $X_1, \dots, X_n$ er uavhengige og log-normal fordelte, derfor

$$E(Y_i) = \mu$$

Da vett at $\bar{Y} = \hat{\mu} = \frac{1}{n}(Y_1 + \cdots + Y_n)$, og siden $Y_i = \ln(X_i)$ så får vi at

$$\bar{Y} = \hat{\mu} = \frac{1}{n}(Y_1 + \cdots + Y_n) = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\sum_{i=1}^{n} \ln(X_i)$$

Vi får derfor at

$$E(\bar{Y}) = E(\hat{\mu})$$

Altså

$$E(\bar{Y}) = E(\hat{\mu}) = E\left(\frac{1}{2}(Y_1 + \cdots + Y_N)\right) = \frac{1}{n}(n\mu) = \mu$$

Dermed har vi vist at $\hat{\mu}$ er en forvetningsrett estimator for μ.

c. η^* kunne våre en estimator for η dersom fordelingen skulle våre symmetrisk. Vi har en symmetrisk fordeling når gjennomsnitt $E(X)$ eksisterer og er lik medianen η. Gjennomsnitt $E(X)$ er definert som

$$E(X) = e^{(\mu + \frac{\sigma^2}{2})}$$

Som ikke er lik $\eta = e^{\mu}$

$$e^{(\mu + \frac{\sigma^2}{2})} \neq e^{\mu}$$

Pga. symmetri kan ikke η^* kan ikke være en estimator for η

Oppgave 4

a. Vi bruker samme framgangsmåte som i oppgave 3 a. og finner først den kumulative fordelingen til X

$$\begin{aligned} F(x) &= P(X \leq x) \\ &= P(e^Y \leq x) \end{aligned}$$

$$= P(Y \leq lnx)$$

$$= P\left(\frac{Y - \mu}{\sigma} \leq \frac{lnx - \mu}{\sigma}\right)$$

$$= P\left(Z \leq \frac{lnx - \mu}{\sigma}\right)$$

$$= \Phi\left(\frac{ln\, x - \mu}{\sigma}\right)$$

Videre for $\mu = 0.7$ og $\sigma = 1.2$ finner vi at

$$P(X > 10) = 1 - F(10)$$
$$= 1 - \Phi\left(\frac{ln\, x - \mu}{\sigma}\right)$$
$$= 1 - \Phi\left(\frac{ln\, 10 - 0.7}{1.2}\right)$$
$$= 1 - \Phi(1.34)$$
$$= 1 - 0.90914$$
$$= 0.09086$$

Ca. 9 % av norske kvinner (som ikke er avholdende) drikker minst 10 liter ren alkohol i løpet av ett år.

b. Medianen til en log-normal fordeling er definert som (se oppg.3 a.)

$$\eta = e^{\mu}$$

Vi setter at $\mu = 0.7$ og får

$$\eta = e^{0.7} = 2.014$$

Forventet årlig alkohol forbruk er definer som (se oppg.3 a.)

$$E(X) = E(e^{Y}) = e^{\mu} e^{\frac{\sigma^2}{2}}$$

Vi setter at $\mu = 0.7$ og $\sigma = 1.2$, og får

$$E(X) = E(e^{Y}) = e^{0.7} e^{\frac{(1.2)^2}{2}} = 4.137$$

Medianen egner seg bedre til å beskrive alkoholforbruket for voksne norske kvinner enn forventningsverdi siden den tar mindre hensyn på de kvinnene som drikker ekstrem mye (og som derfor øker kraftig gjennomsnittet), og heller tar for seg den mengden på alkohol som flertallet av kvinner drikker.

Medianen til det årlige alkoholforbruket er 2.014 liter og forventet årlig alkohol forbruk for (ikke-avholdende) voksne norske kvinner er 4.137 liter. Medianen egner seg best til å beskrive alkoholforbruket for voksne norske kvinner.

c. Vi skriver følgende koden i Python

```
import numpy as np
import scipy.stats as stats

x = [1, 3.4, 5, 14.4, 11.5, 8.2, 0.6, 2.7, 26.8, 3.0,
     1.3, 20.2, 4, 14, 3.3, 1.8, 1.7, 4.6, 7.4, 7.1,
     5.2, 23.6, 1.6, 1.1, 15.5, 3, 1.9, 4.2, 27.4, 1.5]

m = np.median(x) #estimat for medianen
n = np.mean(x) #estimat for gjennomsnitt

l=np.size(x) #størrelse
s=np.std(x) #sigma

h = stats.norm.interval(0.95, loc=n, scale=s/l) #intervallet for gjennomsnitt
f = stats.norm.interval(0.95, loc=m, scale=s/l) #intervallet for median

print("Estimat for median er %g"%(m))
print("Estimat for gjennomsnitt er %g"%(n))
print("Tilnærmet 95% konfidensintervall for medianen er")
print(f)
print(" tilnærmet 95% konfidensintervall for gjennomsnitt")
print(h)
```

Kjøreeksempel gir

```
Estimat for median er 4.1
Estimat for gjennomsnitt er 7.56667
Tilnærmet 95% konfidensintervall for medianen er
(3.588621914484165, 4.611378085515835)
 tilnærmet 95% konfidensintervall for gjennomsnitt
(7.0552885811508315, 8.078044752182501)
```

d. Vi skriver videre på koden fra forrige deloppgave, resten av koden er

```python
B = 10000
meanvec = []
medianvec = []
for i in range(B):
    xstar = stats.norm.rvs(n, s, size=l)
    meanvec.append(np.mean(xstar))
    medianvec.append(np.median(xstar))
print("Standardfeilen til estimat for gjennomsnitt er %g" %(np.std(meanvec)))
print("Standardfeilen til estimat for median er %g" %(np.std(medianvec)))
```

Kjøreeksempel gir

```
Standardfeilen til estimat for gjennomsnitt er 1.42898
Standardfeilen til estimat for median er 1.7454
```

e. Sammenlignet med resultatene fra oppgave b. ser vi at alkoholforbruket er høyere hos de kvinnelige studentene sammenlignet med norske kvinner. Vi ser også at forskjellen mellom gjennomsnitt og median er mye større, og som nevnt tidligere skyldes at noen individer drikker mer alkohol enn andre.