Innlevering 4

Oppg. 1 a)

i) $G = g(S,T) = \dfrac{2S}{T^2}$

$$E(G) = E(g(S,T)) \approx g(\mu_S, \mu_T) = \frac{2\mu_S}{\mu_T^2} = \frac{2 \cdot 241{,}3\,m}{(7{,}02\,s)^2}$$
$$= \underline{\underline{9{,}79}}$$

ii) $Var(G) = Var(g(S,T))$

$$\approx \left(\frac{\partial g}{\partial S}(\mu_S, \mu_T)\right)^2 Var(S) + \left(\frac{\partial g}{\partial T}(\mu_S, \mu_T)\right)^2 Var(T)$$

$$\frac{\partial g}{\partial S} = \frac{2}{T^2}, \quad \frac{\partial g}{\partial T} = 2S \cdot \frac{\partial g}{\partial T}(T^{-2}) = -4S \cdot \frac{1}{T^3}$$

$$Var(G) \approx \left(\frac{2}{\mu_T^2}\right)^2 \cdot \sigma_S^2 + \left(-\frac{4S}{\mu_T^3}\right)^2 \sigma_T^2 = \underline{\underline{\frac{4}{\mu_T^4}\sigma_S^2 + \frac{16}{\mu_T^6}\sigma_T^2 \mu_S^2}}$$

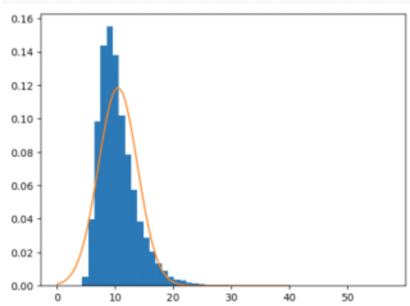
iii) $\sigma_S = 2m, \quad \sigma_t = 1s$

$$\sigma_g^2 = \frac{4}{7{,}02^4} \cdot 2^2 + \frac{16}{7{,}02^6} \cdot 1^2 \cdot 241{,}3^2 = 7{,}791$$

$$\Rightarrow \underline{\underline{\sigma_g = 2{,}791}}$$

b)

```
import numpy as np
import matplotlib.pyplot as plt
mu_S = 241.3
mu_T = 7.02
sigma_S = 2
sigma_T = 1
N = 10000
def Gaussian(x,mu,sigma):
    return 1/np.sqrt(2*np.pi*sigma**2)*np.exp(-0.5*((x-mu)/sigma)**2)
S_simulations = np.random.normal(mu_S,sigma_S,N)
T_simulations = np.random.normal(mu_T,sigma_T,N)
G_simulations = 2*S_simulations / T_simulations**2
plt.hist(G_simulations,50,density = True)
X = np.linspace(0,40,100)
plt.plot(X,Gaussian(X,np.mean(G_simulations),np.std(G_simulations)))
plt.show()
```



Jeg tror ikke G er normalfordelt siden histogrammet ikke samsvarer med grafen til normalfordelingen.

2a) $P\left(\frac{X+Y}{2} > 10,2\right) = P(\tilde{X} > 10,2)$

$\tilde{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) = N(10, 0,02)$

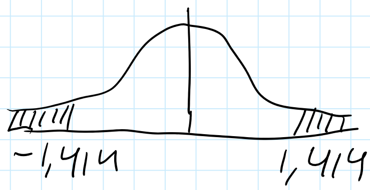
$\frac{10,2-10}{\sqrt{0,02}} = 1,414$

$\Phi(-1,414) = 0,0793$

$P(|X-Y| > 0,4) = P(|Z| > 0,4)$

$Z \sim N(0, 0,2^2 \cdot 2) = N(0, 0.08)$

$\frac{0,4-0}{\sqrt{0,08}} = 1,414$



$2\Phi(-1,414) = 0,1586$

b) $\bar{X} \sim N\left(\mu_A, \frac{\sigma^2}{5}\right)$ $E(\hat{\mu}_A) = E(\bar{X}) = \mu_A$

$\bar{Y} \sim N\left(\mu_B, \frac{\sigma^2}{5}\right)$ $E(\hat{\mu}_B) = E(\bar{Y}) = \mu_B$

$Var(\hat{\mu}_A) = \frac{\sigma^2}{5}, \quad Var(\hat{\mu}_B) = \frac{\sigma^2}{5}$

$U = X+Y, \quad V = X-Y$

$U \sim N(\mu_A + \mu_B, \sigma^2) \Rightarrow \bar{U} \sim N\left(\mu_A + \mu_B, \frac{\sigma^2}{5}\right)$

$V \sim N(\mu_A - \mu_B, \sigma^2) \Rightarrow \bar{V} \sim N\left(\mu_A - \mu_B, \frac{\sigma^2}{5}\right)$

$E(\tilde{\mu}_A) = E\left(\frac{1}{2}\bar{U} + \frac{1}{2}\bar{V}\right) = \frac{1}{2}\left(E(\bar{U}) + E(\bar{V})\right) = \mu_A$

$E(\tilde{\mu}_B) = E\left(\frac{1}{2}\bar{U} - \frac{1}{2}\bar{V}\right) = \frac{1}{2}\left(E(\bar{U}) - E(\bar{V})\right) = \mu_B$

$Var(\tilde{\mu}_A) = Var\left(\frac{1}{2}\bar{U}\right) + Var\left(\frac{1}{2}\bar{V}\right) = \frac{1}{4}\left(Var(\bar{U}) + Var(\bar{V})\right) = \frac{1}{4}\frac{2\sigma^2}{5} = \frac{\sigma^2}{10}$

$Var(\tilde{\mu}_B) = Var\left(\frac{1}{2}\bar{U}\right) + Var\left(\frac{1}{2}\bar{V}\right) = \frac{1}{4}\left(Var(\bar{U}) + Var(\bar{V})\right) = \frac{1}{4}\frac{2\sigma^2}{5} = \frac{\sigma^2}{10}$

Alternativ 2 er best fordi $\tilde{\mu} = \mu_A$, $\tilde{\mu}_B = \mu_B$
og fordi $Var(\tilde{\mu}_A) < Var(\hat{\mu}_A)$, $Var(\tilde{\mu}_B) < Var(\hat{\mu}_B)$

3a) $L(\lambda) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{(x_i)!} e^{-\lambda}$

$\ln L = \sum_{i=1}^{n} \frac{x_i}{\lambda} - n$

$0 = \frac{\sum_{i=1}^{n} x_i}{\hat{\lambda}} - n \Rightarrow \hat{\lambda} = \sum_{i=1}^{n} x_i, \quad Y = \sum_{i=1}^{n} x_i$

$Y \sim Poi(n\lambda)$

$$0 = \frac{\sum\limits_{i=1}^{n}}{\hat{\lambda}} - n \Rightarrow \lambda = \sum_{i=1} x_i \; , \quad y = \sum_{i=1} x_i$$

$$Y \sim Poi(n\lambda)$$

$$E(\hat{\lambda}) = E\left(\frac{Y}{n}\right) = \frac{1}{n} E(Y) = \lambda \leftarrow \text{Forventningsrett}$$

$$E(\hat{\lambda}) = Var\left(\frac{Y}{n}\right) = \frac{1}{n^2} Var(Y) = \frac{n\lambda}{n^2} = \frac{\lambda}{n}$$

b) $$E(\tilde{\lambda}) = \alpha E(\bar{X}) + \beta E(\bar{Y}) = \alpha\lambda + \beta\frac{\lambda}{2} = \lambda$$

$$\Rightarrow \alpha + \frac{\beta}{2} = 1$$

$$Var(\bar{X}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \cdot n \, Var(X_i) = \frac{\lambda}{n}$$

$$Var(\bar{Y}) = Var\left(\frac{1}{m}\sum_{i=1}^{m} Y_i\right) = \frac{1}{m^2} \cdot m \, Var(Y_i) = \frac{\lambda}{2m}$$

$$Var(\tilde{\lambda}) = \alpha^2 Var(\bar{X}) + \beta^2 Var(\bar{Y}) = \alpha^2 \frac{\lambda}{n} + \beta^2 \frac{\lambda}{2m}$$

$$\alpha = 1 - \frac{\beta}{2} \Rightarrow Var(\tilde{\lambda}) = \left(1 - \frac{\beta}{2}\right)^2 \frac{\lambda}{n} + \beta^2 \frac{\lambda}{2m}$$

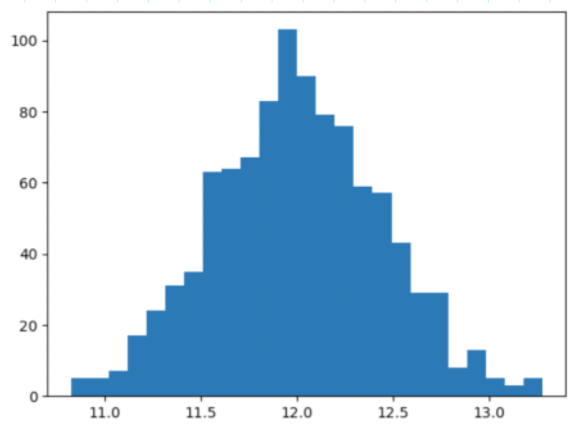
$$\frac{\partial Var(\tilde{\lambda})}{\partial \beta} = \left(\frac{\beta}{2} - 1\right)\frac{\lambda}{n} + \frac{\beta\lambda}{m} = 0$$

$$\beta\left(\frac{\lambda}{2n} + \frac{\lambda}{m}\right) = \frac{\lambda}{n} \Rightarrow \beta = \frac{1}{n\left(\frac{1}{2n} + \frac{1}{m}\right)} = \frac{2m}{m+2n}$$

$$\alpha + \frac{m}{m+2n} = 1 \Rightarrow \underline{\underline{\alpha = \frac{2n}{m+2n}}}$$

c)

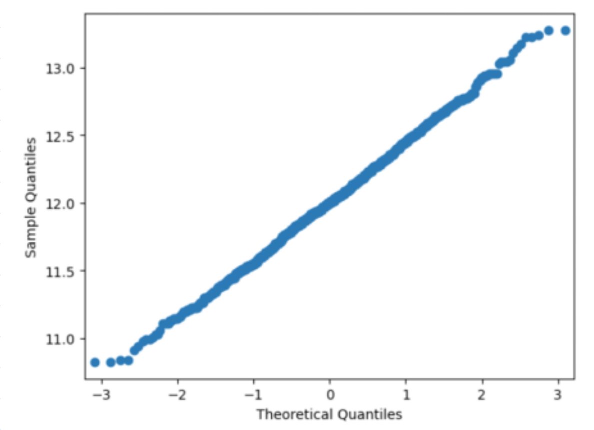
```python
import numpy as np
import matplotlib.pyplot as plt
import statsmodels.api as sm
n = 42
m = 35
l = 12
N = 1000
L = np.zeros(N)
for r in range(1000):
    alpha = 2*n/(m+2*n)
    beta = 2*m/(m+2*n)
    X = np.random.poisson(l,n)
    Y = np.random.poisson(l/2,m)
    X_bar = np.mean(X)
    Y_bar = np.mean(Y)
    l_tilde = alpha*X_bar + beta*Y_bar
    L[r] = l_tilde

plt.hist(L,bins=25,density = False)
sm.qqplot(L)
plt.show()
```



Basert på dette virker $\tilde{\lambda}$ ganske normalfordelt.

Hvis man tar gjennomsnittet $n = 42$, $m = 35$ så er

$\bar{X}$ og $\bar{Y}$ ca normalfordelt $\Rightarrow$ $\tilde{\lambda}$ tilnærmet normalfordelt



Her ser vi en tydelig normalfordeling.

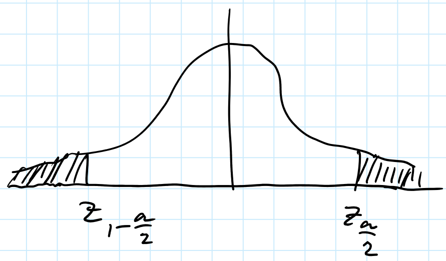
Teorien bak er sentralgrenseteoremet: Vi tar gjennomsnittet / $\frac{\text{summen}}{n}$ av mange uavhengige stokastiske variabler.

4a) μ er gjennomsnittlig/forventet koffeininnhold i 1 dL av Cola

σ er slik at ca. 68 % av 1 dl Cola vil inneholde mellom $\mu - \sigma$ til $\mu + \sigma$ koffein

b) $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$\Rightarrow Z \sim \dfrac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$



$P(|Z| < z_{\frac{\alpha}{2}}) = 1 - \alpha$

$\Rightarrow P(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}) = P\left(-z_{\frac{\alpha}{2}} < \dfrac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} < z_{\frac{\alpha}{2}}\right)$

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

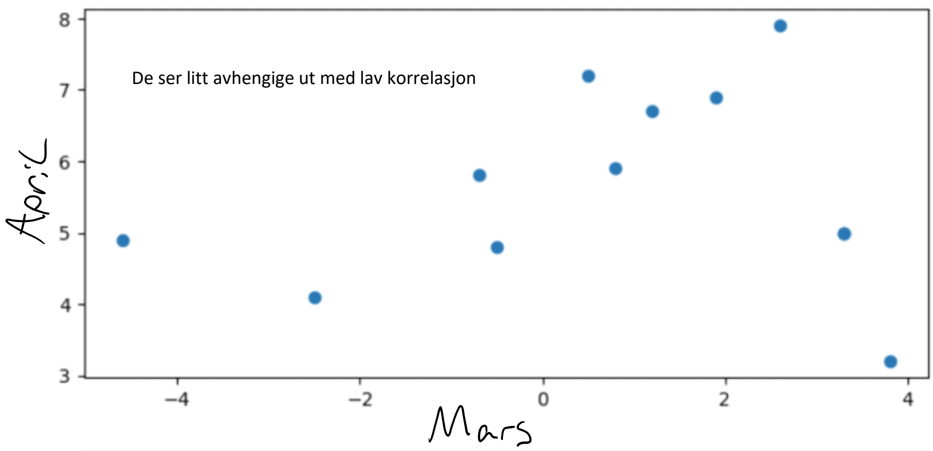
øvre og nedre grense endres fordi $\bar{X}$ endres med hvert forsøk.

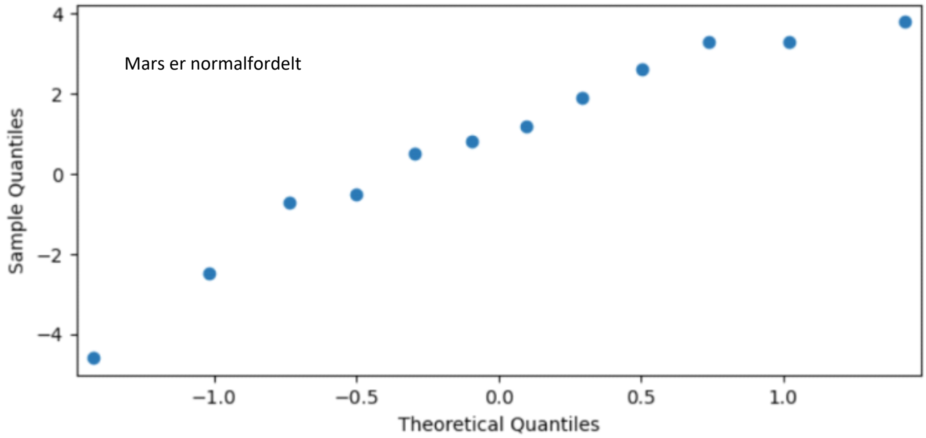
$\alpha = 0{,}05 \Rightarrow \mu : \left[8{,}2 - 1{,}96 \cdot \frac{0{,}19}{\sqrt{12}},\ 8{,}2 + 1{,}96 \cdot \frac{0{,}19}{\sqrt{12}}\right]$

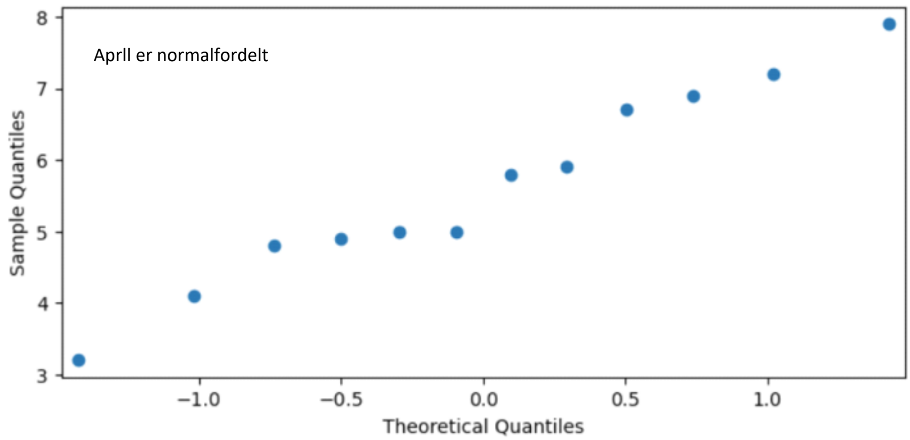
$\mu : [8{,}09,\ 8{,}31]$

c) $z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \frac{0{,}1}{2} \Rightarrow \left(z_{\frac{\alpha}{2}} \frac{\sigma}{0{,}05}\right)^2 < n \Rightarrow n = 55{,}47 \approx 56$

5 a) Man kan legge X og Y i qq plot for å sjekke om de er normalfordelte, dvs på en linje. Man kan plotte X og Y i et scatterplot for å se om det er trender (en rett linje f.eks).








```python
import numpy as np
import matplotlib.pyplot as plt
import statsmodels.api as sm
x = np.array([-2.5, 0.5, 3.3, 2.6, -0.7, -4.6, 3.3, 0.8, 1.9, -0.5, 1.2, 3.8])
y = np.array([4.1, 7.2, 5.0, 7.9, 5.8, 4.9, 5.0, 5.9, 6.9, 4.8, 6.7, 3.2])
figs, axs = plt.subplots(3, 1, figsize=(8, 12))
axs[0].scatter(x, y)
sm.qqplot(x, ax=axs[1])
sm.qqplot(y, ax=axs[2])

plt.show()
```

$$s^2 = \frac{n}{n-1}\sigma^2 \Rightarrow \sqrt{\frac{n}{n-1}}\,\sigma = s$$

$$\mu_m : \left[\bar{X} - t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}},\ \bar{X} + t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}} \right]$$

$$\mu_m : \left[0{,}75833 - 3{,}106 \cdot \frac{2{,}41815}{\sqrt{11}},\ 0{,}75833 + 3{,}106 \cdot \frac{2{,}41815}{\sqrt{11}} \right]$$

$$\mu_m : [-1{,}506,\ 3{,}023]$$