

Løsningsforslag kontinuasjonseksamen

TMA4240/45

August 2024

Oppgave 1

- $E(X + Y) = 2.00$
- $\text{Var}(2X + Y) = 2^2\text{Var}(X) + \text{Var}(y) + 2 \cdot 2\text{Cov}(X, Y) = 8 + 1 + 4 = 13.00$
- $P(X < 3) = P(Z \leq \frac{1}{\sqrt{2}}) = \Phi(0.71) = 0.7611 \approx 0.76$
- $P(Y > 0) = 0.50$

Oppgave 2

- Gjennomsnitt 6.0
- Median 6.0
- Standardavvik $2.768875 \approx 2.77$

Oppgave 3

- $E(X) = 2$
- $\text{SD}(Y) = 1$
- $\text{Cor}(X, Y) = 0.8$

Oppgave 4 Streikeuttak

(i) Siden 110 av 250 er organisert i Unio, er sannsynligheten for at en tilfeldig valgt ansatt er organisert i Unio gitt ved $110/250 = 0.44$.

(ii) La S betegne hendelsen ‘tatt ut i streik’, la U betegne ‘medlem i Unio’, la A betegne ‘medlem i Akademikerne’ og la O' betegne ‘uorganisert’. Sannsynligheten for at en tilfeldig valgt ansatt blir tatt ut i streik kan da regnes ut ved

$$P(S) = P(S|U)P(U) + P(S|A)P(A) + P(S|O')P(O'),$$

der $P(S|O') = 0$ siden uorganiserte ikke streiker. Dermed blir

$$P(S) = P(S|U)P(U) + P(S|A)P(A) = 0.1 \cdot \frac{110}{250} + 0.5 \cdot \frac{80}{250} = 0.204.$$

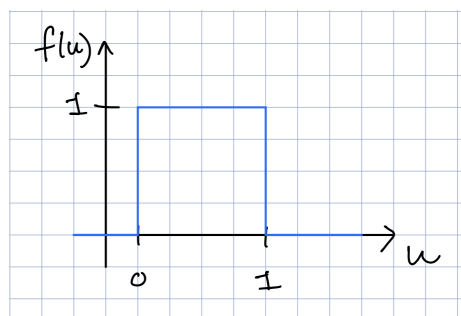
(iii) Gruppen av streikere for Unio består av $110 \cdot 0.1 = 11$ AA-ansatte, $50 \cdot 0.5 = 25$ BB-ansatte og $60 \cdot 0.3 = 18$ CC-ansatte, totalt er da 54 personer tatt ut i streik. Sannsynligheten for at en tilfeldig valgt streiker er ansatt ved CC-fakultetet er dermed $P(CC) = 18/54 = 0.333$, der hendelsen CC betegner ‘streiker fra CC-fakultetet’. Men, dersom vi bare ser på streikende som ikke er ansatt ved AA-fakultet, har vi totalt 43 streikende, hvorav 18 er fra CC-fakultet. Den betingede sannsynligheten for at en streikende - som vi vet ikke er ansatt ved AA-fakultet - er fra CC-fakultet er da $P(CC|AA') = 18/43 = 0.419$.

(iv) La V betegne ‘streikevakt mandag morgen’. I følge oppgaven må V og CC være uavhengige hendelser siden sannsynligheten for streikevakt (etter at en person er tatt ut i streik) ikke påvirkes av fakultetstilhørighet. Dermed er

$$P(CC|V) = P(CC) = 18/54 = 0.333$$

Oppgave 5 Stokastisk simulering

(i) U er standard uniformfordelt og fordelingsfunksjonen er skissert i figuren under. Videre er $E(U) = \frac{(0+1)}{2} = 0.5$ og $\text{Var}(U) = \frac{(1-0)^2}{12} = 1/12 \approx 0.083$.



(ii) Vi kan lese ut fra koden at X tar heltallsverdiene 3, 4, 5 og 6. Den kumulative

fordelingen kan vi lese rett fra koden;

$$P(X \leq x) = 0 \text{ for } x < 3$$

$$P(X \leq 3) = 0.167$$

$$P(X \leq 4) = 0.528$$

$$P(X \leq 5) = 0.956$$

$$P(X \leq x) = 1 \text{ for } x \geq 6$$

Punktsannsynlighetene for $x \in \{3, 4, 5, 6\}$ blir

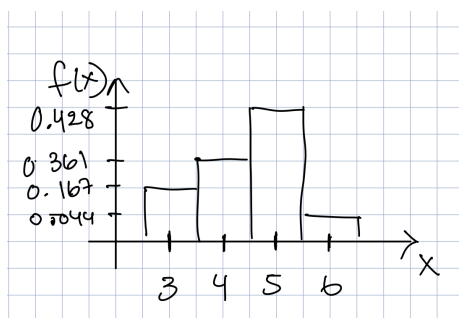
$$P(X = 3) = 0.167$$

$$P(X = 4) = 0.528 - 0.167 = 0.361$$

$$P(X = 5) = 0.956 - 0.528 = 0.428$$

$$P(X = 6) = 1 - 0.956 = 0.044$$

En grov skisse av punktsannsynlighetene:



(iii) Følgende kode kan brukes for å tilnærme $E(X^2)$ med stokastisk simulering

```
import numpy as np
# antar simX allerede definert
n = 1000
x = simX(n)
ex2 = np.sum(x**2)/n
```

Vi har satt $n = 1000$ slik at forventningsverdien kan approksimeres ved å regne ut gjennomsnittet av kvadrerte simulerte trekninger fra X . Lavere n kan også gi gode nok estimat, men jo høyere n jo bedre blir tilnærmingen.

Oppgave 6 Eksponensialfordelte stokastiske variabler

a)

Kumulativ fordeling finner en ved

$$F(x) = \int_0^x f(u)du = \int_0^x \frac{1}{\mu} e^{-u/\mu} du = [-e^{-u/\mu}]_0^x = -e^{-x/\mu} - 1 = 1 - e^{-x/\mu}$$

Sannsynlighet for at X er mindre enn 2000 er når vi antar at $\mu = 1000$:

$$P(X \leq 2000) = F(2000) = 1 - e^{-2} = 0.865 \quad (1)$$

Medianen m er slik at $P(X \leq m) = 0.5$, dvs:

$$0.5 = F(m) = 1 - e^{-m/1000} \quad (2)$$

Det gir:

$$\begin{aligned} e^{-m/1000} &= 0.5 \\ -\frac{m}{1000} &= -\log(2) \\ m &= \log(2)/1000 = 693.15 \end{aligned}$$

b)

(i) En stokastisk variabel Y som er kjikvadratfordelt med $\nu = 2$ vil ha sannsynlighetstetthet

$$f(y) = \frac{1}{2^{2/2}\Gamma(2/2)} y^{2/2-1} e^{-y/2} = \frac{1}{2} e^{-y/2}.$$

Vi definerer en stokastisk variabel $Z = 2X/\mu$ som er en funksjon av X . Da er

$$\begin{aligned} F_Z(z) &= P(Z \leq z) = P(2X/\mu \leq z) = P(X \leq z\mu/2) \\ &= 1 - e^{-\frac{z\mu}{2\mu}} = 1 - e^{-z/2}. \end{aligned}$$

Videre er

$$f_Z(z) = \frac{dF(z)}{dz} = \frac{1}{2} e^{-z/2}$$

Dette kjenner vi igjen som sannsynlighetstettheten til en stokastisk variabel som er kjikvadratfordelt med $\nu = 2$.

Vi kan også vise at Z er kjikvadratfordelt ved å se på de momentgenererende funksjonene. En stokastisk variabel Y som er kjikvadratfordelt med $\nu = 2$ vil ha MGF

$$M_Y(t) = \left(\frac{1}{1-2t} \right)^{2/2} = \frac{1}{1-2t}$$

Den eksponensialfordelte variabelen X har MGF $M_X(t) = \frac{1}{1-\mu t}$ og siden $M_{aX}(t) = M_X(at)$ for en konstant a så har vi at

$$M_Z(t) = M_{2X/\mu}(t) = M_X(2t/\mu) = \frac{1}{1-2t}$$

som vi kjenner igjen om MGF-en til en kjikvadratfordelt stokastisk variabel med $\nu = 2$.

(ii) Den stokastiske variabelen $2 \sum_{i=1}^n X_i/\mu = \sum_{i=1}^n 2X_i/\mu$ er en sum av n uavhengige kjikvadratfordelte stokastiske variabler med 2 frihetsgrader. Det betyr at $2 \sum_{i=1}^n X_i/\mu$ er χ -kvadratfordelt med $2n$ frihetsgrader.

c)

For å utlede et konfidensintervall for μ tar vi utgangspunkt i at vi kjenner fordelingen til $2 \sum_{i=1}^n X_i/\mu$. Da er:

$$P(\chi_{1-\alpha/2, 2n}^2 < \frac{2 \sum_{i=1}^n X_i}{\mu} < \chi_{\alpha/2, 2n}^2) = P(\frac{2 \sum_{i=1}^n X_i}{\chi_{\alpha/2, 2n}^2} < \mu < \frac{2 \sum_{i=1}^n X_i}{\chi_{1-\alpha/2, 2n}^2}) = 1 - \alpha$$

I vårt tilfelle er $n = 5$ og da har $2 \sum_{i=1}^n X_i/\mu$ en χ -kvadrat fordeling med $\nu = 10$ frihetsgrader. Vi setter så $\alpha = 0.05$ og finner fra tabellen at $\chi_{0.975, 10} = 3.247$ og $\chi_{0.025, 10} = 20.483$. Videre er $2 \sum_{i=1}^5 x_i = 16986$. Vårt 95% KI for μ blir da

$$(829.27, 5231.29).$$

d)

Rimelighetsfunksjonen for parameteren μ_A er gitt ved:

$$\begin{aligned} L(\mu_A; x_1, \dots, x_m, y_1, \dots, y_n) &= \prod_i f_X(x_i) \prod_j f_Y(y_j) \\ &= \prod_{i=1}^m \frac{1}{\mu_A} e^{-x_i/\mu_A} \prod_{j=1}^n \frac{1}{\mu_B} e^{-y_j/\mu_B} = \prod_{i=1}^m \frac{1}{\mu_A} e^{-x_i/\mu_A} \prod_{j=1}^n \frac{2}{\mu_A} e^{-2y_j/\mu_A} \end{aligned}$$

Log-rimelighetsfunksjonen er:

$$\begin{aligned} l(\mu_A; x_1, \dots, x_m, y_1, \dots, y_n) &= \sum_{i=1}^m \left[\log\left(\frac{1}{\mu_A}\right) - \frac{x_i}{\mu_A} \right] + \sum_{j=1}^n \left[\log\left(\frac{2}{\mu_A}\right) - \frac{2y_j}{\mu_A} \right] \\ &= -m \log(\mu_A) - \frac{1}{\mu_A} \sum_{i=1}^m x_i - n \log(\mu_A) + n \log(2) - \frac{2}{\mu_A} \sum_{j=1}^n y_j \\ &= -(m+n) \log(\mu_A) - \frac{1}{\mu_A} \left(\sum_{i=1}^m x_i + 2 \sum_{j=1}^n y_j \right) + n \log(2) \end{aligned}$$

For å finne SME må vi derivere $l(\cdot)$ med hensyn på μ_A :

$$\frac{d l(\mu_A; x_1, \dots, x_m, y_1, \dots, y_m)}{d\mu_A} = -\frac{m+n}{\mu_A} + \frac{1}{\mu_A^2} \left(\sum_i x_i + 2 \sum_j y_j \right)$$

Når vi setter det lik 0 får vi at:

$$-\frac{m+n}{\mu_A} + \frac{1}{\mu_A^2} \left(\sum_i x_i + 2 \sum_j y_j \right) = 0$$

$$\mu_A = \frac{\sum_i x_i + 2 \sum_j y_j}{m+n}$$

Innsatt stokastiske variabler $X_1, \dots, X_m, Y_1, \dots, Y_n$ finner vi SME

$$\hat{\mu}_A = \frac{\sum_{i=1}^m X_i + 2 \sum_{j=1}^n Y_j}{m+n} = \frac{m\bar{X} + 2n\bar{Y}}{m+n}$$

e)

Den stokastiske variabelen Z er binomisk fordelt med parametere n og $p = \frac{\mu_A}{\mu_A + 2\mu_B}$. Under $H_0 : \mu_A = 2\mu_B$ er $p = 0.5$. Under H_0 er dermed fordelingen til Z kjent (ingen ukjente parameterverdier) og kan derfor brukes for å konstruere en test. Under den alternative hypotesen er $p > 0.5$.

For å ta en beslutning vi kan bruke p -verdi. Den er definert som sannsynlighet for å observere en lik eller mer ekstrem verdi enn det som er observert, under H_0 . I vårt tilfelle blir p -verdien:

$$P(Z > z_{\text{obs}} | H_0) = P(Z \geq 7 | p = 0.5, n = 10) =$$

$$1 - P(Z \leq 6 | p = 0.5, n = 10) = 1 - 0.828 = 0.172 > 0.05.$$

Nullhypotesen forkastes ikke.

Alternativt kan man bruke kritisk verdi. Vi forkaster H_0 dersom $Z > k$, der k er en kritisk verdi. Vi må først finne et størst mulig heltall k slik at $P(Z > k) \leq \alpha$, og her er $\alpha = 0.05$. Dette kan skrives om til $1 - P(Z \leq k) \leq \alpha$ og $P(Z \leq k) \geq 1 - \alpha = 0.95$. Fra tabell finner vi at kritisk verdi k må være $k = 8$, som gir $P(Z > k | H_0) = 1 - 0.989 = 0.011$. Siden observasjonen vår er lavere enn den kritiske verdien kan vi ikke forkaste H_0 .