Obligatorisk innlevering 2 i STAT110/622

Basert på kapittel 1-6 i boken. Les informasjon “Obligatoriske innleveringer” på MittUiB->Hjem.

1. **Defekte objekter** Vi antar at det er 3 defekte objekt i en mengde på 21 objekter. Et utvalg å 7 blir plukket ut tilfeldig uten tilbakelegg. La X betegne antall defekte objekter i utvalget.

   (a) Finn sannsynligheten for at utvalget inneholder akkurat defekte 2 objekter.

   Solution: Dette er en hypergeometrisk fordeling:

   $$P(X = 2) = \frac{\binom{3}{2}\binom{21-3}{7-2}}{\binom{21}{7}} = 0.22$$

   (b) Finn forventning og varians til X.

   Solution:

   $$E(X) = 7 \cdot \frac{3}{21} = 1$$
   $$V(X) = \left(\frac{21-7}{21-1}\right) \cdot 7 \cdot \frac{3}{21}\left(1 - \frac{3}{21}\right) = 0.6$$

2. **Stokastiske algoritmer**

   En programmerer kan velge mellom to algoritmer (A og B) for å løse et problem som vi kaller “Q”. Algoritme A er deterministisk, dvs. den er garantert å finne et svar, og bruker alltid samme tid (20 timer) for å løse Q. Algoritme B er stokastisk, dvs. at den gjør bruk av tilfeldige tall internt, men vi skal ikke “kikke inn” i selve algoritmen. Algoritme B klarer ikke alltid å finne et svar, men til gjengjeld kjører den raskt. Den har alltid en kjøretid på 0.5 timer (på problem Q), uavhengig av om den klarer å finne et svar eller ikke, og den gir beskjed når den feiler. Derfor må B kjøres flere ganger, helt til den finner svaret. Vi definerer $p = P(\text{B finner svaret})$ for én enkelt kjøring, og at ulike kjøringer av B er uavhenige (med tanke på om de finner svaret eller ikke).

   (a) Dersom $p = 0.05$, hva er forventet total kjøretid før algoritme B finner et svar? Bør programmereren velge algoritme A eller B for problemet Q? (Hint: kikk på geometrisk fordeling)

   Solution: Vi lar X være antall kjøringer av B som må til før den finner svaret. Vi har at X er geometrisk fordelt, og med suksessannsynlighet $p = 0.05$. Da er i følge slides fra kapittel 3.3 forventningen til X:

   $$E(X) = \frac{1-p}{p} + 1 = \frac{1}{p} = 20$$

   Husk at det er to varianter av den geometriske fordelingen: en der du teller med suksessen (og det gjør vi), og en der du ikke gjør det. Forventet total kjøretid før B finner svaret blir $0.5 \cdot E(X) = 10$ timer, og programmereren bør derfor velge B, siden A tar 20 timer.

3. **Om hjort**

   En økolog ønsker å undersøke hjortebestanden i to forskjellige områder, Område 1 og Område 2, på Vestlandet. Han antar at antall hjort, X og Y, i Område 1 og 2 er Poisson fordelt.

   (a) Økologen regner med at forventet antall hjort er henholdsvis $\lambda_1 = 3$ i område 1 og $\lambda_2 = 5$ i område 2. Finn $P(X = 2)$ og $P(X \geq 3)$, og finn et tilnærmet uttrykk for $P(X = Y)$. Presiser den forutsetningen du må bruke for å regne ut det siste svaret.

Solution: $P(X = 2) = 0.224$ og $P(X \geq 3) = 1 - P(X \leq 2) = 0.577.$

$$P(X = Y) = \sum_{i=0}^{\infty} P(X = i)P(Y = i) = 0.05 \cdot 0.007 + 0.149 \cdot 0.034 + \cdots = 0.113$$

Vi har her antatt at X og Y er uavhengige, slik at vi kan multiplisere $P(X = i)P(Y = i)$.

(b) Dersom en hjort befinner seg i et slikt område, regner økologen med en sannsynlighet på 0.2 for at den blir observert. Dersom det i virkeligheten er 5 hjort i området, hva er sannsynligheten for at tre av dem blir observert? Hva er forventet antall hjort observert?

Solution: La Z være antall observerte hjort. Da er Z binomisk fordelt med $n = 5$ og $p = 0.2$. $P(Z = 3) = \binom{5}{3} 0.2(1 - 0.2) = 0.051$ og $E(Z) = np = 5 \cdot 0.2 = 1$

4. **Investeringer** To investeringer X og Y gir avkastning som følger (forventning og varians):

$$E(X) = 0.6 \text{ og } E(Y) = 0.5; \qquad V(X) = 1, \quad V(Y) = 2.$$

Korrelasjonen mellom X og Y er $\rho(X, Y) = 0.5$.

(a) Finn forventning og varians for de "kombinerte" investeringene

$$U = X + Y, \quad W = 2Y.$$

Finn til slutt korrelasjonen mellom U og W. Hint: Finn først $V(U + W)$.

Solution:

$$E(U) = E(X) + E(Y) = 0.6 + 0.5 = 1.1, \quad E(V) = 2E(Y) = 2 \cdot 0.5 = 1.$$

Vi får bruk for $\text{Cov}(X, Y) = 0.5 \cdot \sqrt{1} \cdot \sqrt{2} = 0.71$

$$V(U) = V(X + Y) = V(X) + V(Y) + 2\text{Cov}(X, Y) = 1 + 2 + 2 \cdot 0.71 = 4.42$$

$$V(W) = V(2Y) = 4 \cdot V(Y) = 8.$$

$$V(U + W) = V(X + 3Y) = V(X) + 9 \cdot V(Y) + 6\text{Cov}(X, Y) = 1 + 18 + 2 \cdot 30.71 = 23.26$$

$$\text{Cov}(U, W) = \frac{1}{2}(V(U + W) - V(U) - V(W)) = \frac{23.26 - 4.42 - 8}{2} = 5.42.$$

$$\text{Corr}(U, W) = \frac{\text{Cov}(U, W)}{\text{SD}(U) \cdot \text{SD}(W)} = \frac{5.42}{\sqrt{4.42} \cdot \sqrt{8}} = 0.91.$$

5. **Sandkorn**

Diameteren av sandkorn i et spesielt sandtak, målt i mm, kan betraktes som en tilfeldig variabel X med sannsynlighetstetthet

$$f(x) = \begin{cases} k(x - x^4) & \text{if } 0 \le x \le 1 \\ 0 & \text{ellers} \end{cases}$$

(a) Bestem konstanten k slik at dette blir en gyldig sannsynlighetstetthet.

> **Solution:** Vi har $\int_0^1 (x - x^4)dx = \frac{1}{2} - \frac{1}{5} = 0.3$, slik at vi må velge $k = 1/0.3 = 3.33$ for at det totale arealet under f skal bli 1.

(b) Finn forventningen $E(X)$ og variansen $V(X)$

> **Solution:** Vi har $E(X) = \int_0^1 xk(x - x^4)dx = 0.555$ og $E(X^2) = \int_0^1 x^2 k(x - x^4)dx = 0.357$ slik at $V(X) = E(X^2) - E(X)^2 = 0.357 - 0.555^2 = 0.049$.

(c) Finn forventningen $E(4X + 3)$ og variansen $V(4X + 3)$

> **Solution:**
> $$E(4X + 3) = 4E(X) + 3 = 5.22.$$
> $$V(4X + 3) = V(4X) = 4^2 V(X) = 4^2 0.049 = 0.784.$$

6. **Simultanfordeling**

Vi betrakter to tilfeldige variable X og Y med simultan sannsynlighetsfordeling $p(x, y) = P(X = x \cap Y = y)$ gitt ved denne tabellen:

$x \backslash y$	0	1	2
0	0.05	0.25	0.1
1	0.1	0.05	0.1
2	0.25	0	0.1

(a) Finn de marginale sannsynlighetsfordelingene for X og Y. Hva er $P(X = 2|Y = 2)$? Er X og Y uavhengige tilfeldige variable? Begrunn svaret.

> **Solution:** Benytter formler $p_X(x) = \sum_y p(x, y)$ og $p_Y(y) = \sum_x p(x, y)$ for marginalsannsynligheter, dvs. summerer rekkevis og kolonnevis i tabellen:
>
> | $x \backslash y$ | 0 | 1 | 2 | $p_X(x)$ |
> |---|---|---|---|---|
> | 0 | 0.05 | 0.25 | 0.1 | 0.4 |
> | 1 | 0.1 | 0.05 | 0.1 | 0.25 |
> | 2 | 0.25 | 0 | 0.1 | 0.35 |
> | $p_Y(y)$ | 0.4 | 0.3 | 0.3 | |
>
> Betinget sannsynlighet: $P(X = 2|Y = 2) = \frac{p(2,2)}{p_Y(2)} = \frac{0.1}{0.3} = 0.333$. Nei, X og Y er ikke uavhengige fordi $P(X = 2|Y = 2) \neq P(X = 2) = p_X(2) = 0.35$.

(b) Finn korrelasjonen mellom X og Y når du får vite at forventningene er $E(X) = 0.95$ og $E(Y) = 0.9$, og variansene $V(X) = 0.747$ og $V(Y) = 0.69$.

Solution: Finner først

$$E(XY) = \sum_{x=0}^{2}\sum_{y=0}^{2} xy \cdot p(x,y) = 0 \cdot 0 \cdot 0.05 + \cdots + 2 \cdot 2 \cdot 0.1 = 0.65$$

og kovarians $cov(X,Y) = E(XY) - E(X)E(Y) = -0.205$, slik at korrelasjonen blir

$$\rho = \frac{cov(X,Y)}{\sqrt{V(X)V(Y)}} \approx -0.286$$

(c) Finn betinget forventning $E(Y|X=2)$ og betinget varians $V(Y|X=2)$.

Solution: Regner først ut betinget fordeling $p_{Y|X}(y|x=2) = p(2,y)/p_X(2)$:

y	0	1	2
$p_{Y\|X}(y\|x=2)$	2.5/3.5	0/3.5	1/3.5

som gir

$$E(Y|X=2) = 0 \cdot 2.5/3.5 + 1 \cdot 0/3.5 + 2 \cdot 1/3.5 = 0.571$$

og varians

$$V(Y|X=2) = (0-0.571)^2 \cdot 2.5/3.5 + (1-0.571)^2 \cdot 0/3.5 + (2-0.571)^2 \cdot 1/3.5 = 0.816$$

7. **Ventetider**

Ventetiden T (i timer) mellom hver telefonsamtale en person mottar er tilfeldig med tetthetsfunksjon $f(t) = 1.8e^{-1.8t}$ for $t \geq 0$.

(a) Finn forventet tid mellom oppringninger og finn tilhørende varians.

Solution:

$$E(T) = \int_0^\infty 1.8te^{-1.8t} = \frac{1}{1.8} \quad E(T^2) = \int_0^\infty t^2 1.8e^{-1.8t} = \frac{2}{3.24}, \quad \text{dvs}$$

$$\text{Var}(T) = E(T^2) - (E(T))^2 = \frac{2}{3.24} - \Big(\frac{1}{1.8}\Big)^2 = \frac{1}{3.24}.$$

(b) For to uavhengige ventetider T_1 og T_2, finn sannsynligheten for at både T_1 og T_2 er større enn 1.

Solution:

$$P(T_1 \geq 1, T_2 \geq 1) = P(T_1 \geq 1)P(T_2 \geq 1) = \Big(\int_1^\infty 1.8e^{-1.8t}dt\Big)^2 = (e^{-1.8})^2 = e^{-3.6} \approx 0.027.$$

8. **Bohrs atommodell**

Figur 1 viser Bohrs modell for hydrogenatomet (les figurteksten). Vi lar (X,Y) betegne posisjonen til elektronet i forhold til atomkjernen. I kvantemekanikk kan man aldri bestemme (X,Y) eksakt, men må i stedet bruke en sannsynlighetsfordeling. Vi antar at $X \sim N(0,1)$ og $Y \sim N(0,1)$, dvs. normalfordelte med forventning 0 og varians 1, og at X og Y er uavhengige av hverandre.

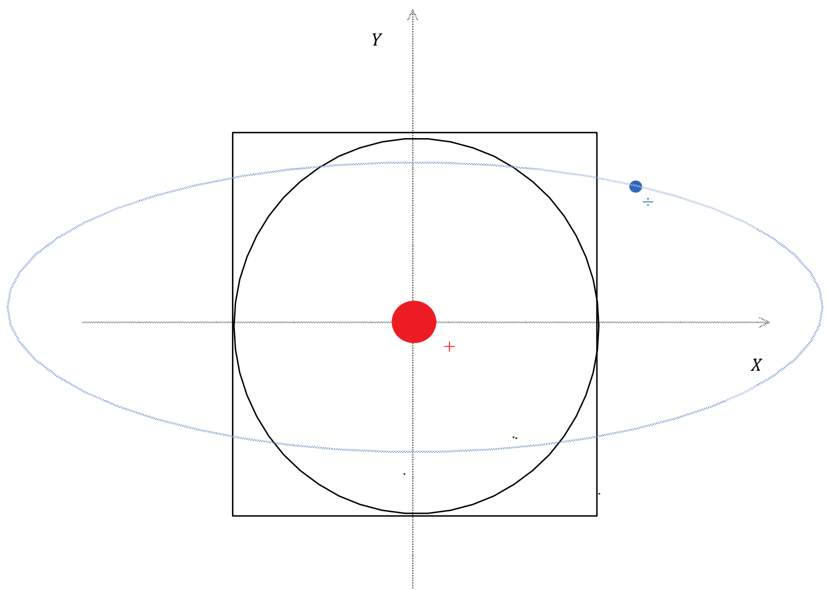

Figure 1: Bohrs atommodell med elektron (blått) og atomkjerne (rød). Den sorte sirkelen har radius $r = 2.15$.

(a) Hva er sannsynligheten for $-2.15 \leq X \leq 2.15$? Hva er sannsynligheten for at elektronet befinner seg inne i det sorte kvadratet. Hint: vi har antatt at X og Y er uavhengige tilfeldige variable.

Solution: Siden X er standard normalfordelt kan vi direkte bruke tabell A3 og finner

$$P(-2.15 \leq X \leq 2.15) = 1 - 2P(X \leq -2.15) = 1 - 2 \cdot 0.016 = 0.968.$$

Det at elektronet befinner seg i kvadratet er det samme som at $-2.15 \leq X \leq 2.15$ og $-2.15 \leq Y \leq 2.15$. Pga. uavhengighet har vi

$$P(\text{Elektron er inne i kvadratet}) = 0.968^2 = 0.937.$$

(b) Hva er sannsynligheten for at elektronet befinner seg inne i den sorte sirkelen? Hjelp: avstanden fra elektronet til kjernen er $R = \sqrt{X^2 + Y^2}$, og det kan vises at den tilfeldige variabelen $U = R^2$ har en kji-kvadratfordeling med $\nu = 2$ frihetsgrader som også er en eksponensialfordeling (men hvilken?).

Solution: Det at elektronet befinner seg i den sorte sirkelen er det samme som at $R \leq 2.15$, dvs. $U \leq 2.15^2 = 4.62$. Som hintet sier skal vi sette inn $\nu = 2$ i uttrykket tettheten i kji-kvadratfordelingen, og vi finner da at U har tetthet $f(u) = \frac{1}{2}\exp(-u/2)$ som er tettheten i en eksponentialfordeling. Dette gir $P(U \leq 4.62) = 1 - \exp(-4.62/2) = 0.90$.

Alternativ løsning som gjør bruk av tabell A6: Det at elektronet befinner seg i den sorte sirkelen er det samme som at $R \leq 2.15$, dvs $R^2 \leq 2.15^2 = 4.62$. Fra tabell A6 finner vi at $P(R^2 \leq 2.15) \approx 1 - 0.1 = 0.9$. Her har vi brukt cellen $\nu = 2$ og $\alpha = 0.1$ i tabellen (den eksakte verdien i tabellen er 4.605.)