STK1100 Oblig 2

Daniel Heinesen, daniehei

27. april 2017

Oppgave 1)

a)

Vi har en simultan sannsynlighetsfordelig

$$f(x,y) = \begin{cases} k(x-y) & 0 \leq y \leq x \leq 1 \\ 0 & \text{ellers} \end{cases} \tag{1}$$

for å finne k kan vi bruke at fordeligen må være normalisert. Dette betyr at

$$I = \int_0^1 \int_0^x k(x-y)dydx = 1 \tag{2}$$

Vi kan regne dette ut:

$$I = \int_0^1 \int_0^x k(x-y)dydx = k\int_0^1 \left[xy - \frac{1}{2}y^2\right]_0^x dx$$
$$= k\int_0^1 \frac{1}{2}x^2 dx = k\frac{1}{6}x^3\Big|_0^1 = \frac{k}{6} = 1$$

Dette gir oss at

$$\underline{\underline{k = 6}} \tag{3}$$

b)

Vi ønsker å finne $P(2Y \leq X)$. Dette er sannsynligheten at om vi velger en tilfeldig x og en tilfeldig y, så er $2Y \leq X \Leftrightarrow Y \leq X/2$. Vi finner sannsynligheten ved

$$P(2Y \leq X) = \int_0^1 \int_0^{x/2} k(x-y)dydx = k\int_0^1 \left(xy - \frac{1}{2}y^2\right)\Big|_0^{x/2} dx = k\int_0^{x/2} \frac{1}{2}x^2 - \frac{1}{8}x^2 dx$$
$$= k\frac{1}{8}x^3\Big|_0^1 = \frac{k}{8}$$

Setter vi inn verdien for k det endelig svaret:

$$\underline{\underline{P(2Y \leq X) = \frac{3}{4}}} \tag{4}$$

c)

Vi ønsker så å finne den marginale sannsynlighetsfordeligen for X. Vi ser først på hvor $0 \leq y \leq x \leq 1$. For å få den marginale sannsynlighetsfordeligen må vi integrere sannsynlighetsfordeligen over alle mulige verdier av y, som er $0 \leq y \leq x$, vi får da

$$f_X(x) = \int_0^x k(x-y)dy = k\left(xy - \frac{1}{2}y^2\right)\Big|_0^x = \frac{k}{2}x^2 = \underline{3x^2} \tag{5}$$

For vi ser nå på de øvrige funksjonene. Her vil $f(x,y) = 0$, som fører til at også $f_X(x) = 0$. Så vi ender opp med at

$$f_X(x) = \begin{cases} 3x^2 & 0 \leq x \leq 1 \\ 0 & \text{ellers} \end{cases} \tag{6}$$

d)

Vi ønsker så å finne den marginale sannsynlighetsfordeligen for Y. Vi starter også her med å se på tilfellene hvor $0 \leq y \leq x \leq 1$. Vi ønsker å integrere $f(x,y)$ over alle de mulige verdiene av x, hvilket er $y \leq x \leq 1$:

$$f_Y(x) = \int_y^1 k(x-y)dy = k\left(\frac{1}{2}x^2 - yx\right)\Bigg|_y^1 = k\left(\frac{1}{2}y^2 - y + \frac{1}{2}\right) = \underline{3y^2 - 6y + 3} \tag{7}$$

Akkurat som i deloppgaven over vil $f(x,y) = 0$ for de øvrige tilfellen, noe som gjør at $f_Y(x) = 0$ her. Så vi ender opp med:

$$f_X(x) = \begin{cases} 3y^2 - 6y + 3 & 0 \leq y \leq 1 \\ 0 & \text{ellers} \end{cases} \tag{8}$$

e)

For å sjekke om X og Y er uavhenginge kan vi gange samme de marginale sannsynlighetsfordelingene og se om vi før den orginale sannsynlighetsfordelingen (1):

$$f_X(x) \cdot f_Y(y) = 3x^2 \cdot (3y^2 - 6y + 3) = 9x^2y^2 - 6x^2y + 9x^2 \neq f(x,y) \tag{9}$$

Siden $f_X(x) \cdot f_Y(y) \neq f(x,y)$ er X og Y **ikke** uavhenginge.

Oppgave 2)

Koden finnes som opp2b.py. Denne koden vil lage alle plottene som er brukt i denne oppgaven.(Alle plotten vil lages og vises når man kjører programmet, så vær forberedt på den når den som retter dette kjører programmet)

a)

Vi har flere identiske stokatiske variabler X_i som følger:

$$E(X_i) = \mu, \qquad V(X_i) = \sigma^2 \tag{10}$$

Vi innfører så at

$$\overline{X}_i = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{11}$$

Vi har også det standariserte gjennomsnittet

$$Z_n = \frac{\overline{X}_i - \mu}{\sigma/\sqrt{n}} = \sqrt{n}\frac{\overline{X}_i - \mu}{\sigma} \tag{12}$$

Vi kommer til å bruke at for forventingsverdien til flere uavhengige stokastiske variabler er

$$E\left(a + \sum_{i=1}^{n} b_i X_i\right) = a + \sum_{i=1}^{n} b_i E(X_i) \tag{13}$$

Og for variansen er gjelder

$$V\left(a + \sum_{i=1}^{n} b_i X_i\right) = \sum_{i=1}^{n} b_i^2 V(X_i) \tag{14}$$

Vi kan fra dette se at for de uavhengige stokastiske variablene X_i gjelder:

$$E(\overline{X}_i) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}n\mu = \underline{\underline{\mu}} \tag{15}$$

og

$$V(\overline{X}_i) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^2}n\sigma^2 = \underline{\underline{\frac{\sigma^2}{n}}} \tag{16}$$

Og for det standariserte gjennomsnittet gjelder

$$E(Z_n) = E\left(\sqrt{n}\frac{\overline{X}_i - \mu}{\sigma}\right) = \sqrt{n}\left(\frac{E(\overline{X}_i)}{\sigma} - \frac{\mu}{\sigma}\right) = \sqrt{n}\left(\frac{\mu}{\sigma} - \frac{\mu}{\sigma}\right) = \underline{\underline{0}} \tag{17}$$

og

$$V(Z_n) = V\left(\sqrt{n}\frac{\overline{X}_i - \mu}{\sigma}\right) = n\frac{V(\overline{X}_i)}{\sigma^2} = n\frac{\sigma^2/n}{\sigma^2} = \underline{\underline{1}} \tag{18}$$

b)

Vi skal her bestemme forventingsverdien og variansen til de 3 følgende fordelingene:

Uniform Fordeling:

Vi har en uniform fordeling:

$$f(x) = \begin{cases} \frac{1}{2} & -1 \leq x \leq 1 \\ 0 & ellers \end{cases} \tag{19}$$

Vi kan regne ut fordelingen:

$$\mu = E(X_i) = \int_{-1}^{1} x f(x) dx = \int_{-1}^{1} x\frac{1}{2} dx = \underline{\underline{0}} \tag{20}$$

Siden integranden er en 'odd' funksjon. For å regne ut variansen regner vi først ut:

$$E(X_i^2) = \int_{-1}^{1} x^2 f(x) dx = \int_{-1}^{1} \frac{1}{2}x^2 dx = \frac{1}{6}x^3\Big|_{-1}^{1} = \frac{1}{3} \tag{21}$$

Vi har da at variansen er

$$\sigma^2 = V(X_i) = E(X_i^2) - E(X_i)^2 = \underline{\underline{\frac{1}{3}}} \tag{22}$$

Gammafordeling:

Vi har gammafordelingen:

$$f(x) = \begin{cases} \frac{1}{\sqrt{\pi x}}e^{-x} & x > 0 \\ 0 & ellers \end{cases} \tag{23}$$

Dette er en gammafordeling med $\alpha = 1/2$ og $\beta = 1$. Vi vet hva definisjonene på forventingsverdien og variansen er for gammafordelinger, så vi kan fort finne at for denne fordelingen er:

$$\mu = \alpha\beta = \underline{\underline{\frac{1}{2}}} \tag{24}$$

og

$$\sigma^2 = \alpha\beta^2 = \underline{\underline{\frac{1}{2}}} \tag{25}$$

Bernoullifordeling:

Vi har en Bernoullifordeling med punktfordelingen $p(x) = P(X = x)$, med en $p = 0.75$, som gir at $p(1) = 0.75$ og $p(0) = 0.25$. Vi vet også de generelle formelene for forventingsverdien og variansen til Bernoullifordelinger. Vi finner så at

$$\mu = p = \underline{\underline{0.75}} \tag{26}$$

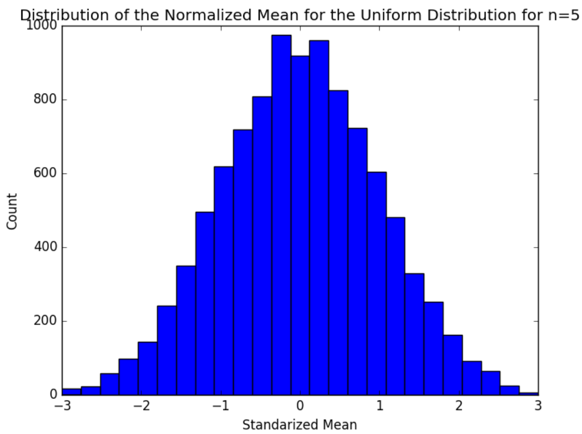
og

$$\sigma^2 = p(1-p) = \frac{1}{4} \cdot \frac{3}{4} = \underline{\underline{\frac{3}{16}}} \tag{27}$$

c)

Vi forventer å se at det normerte histogrammet av observasjonene nærmer seg sannsynlighetstettheten til Z_n blir større og større. Teller vi opp antall observasjoner med en viss verdi/innen for et vist intervall vil forvente at jo høyere sannsynlighet for å finne denne observasjonen, jo flere tilfeller vil vi observere, og vis versa. Med nok antall observasjoner vil observasjonen alltid legge seg etter fordlingen til Z_n. Siden fordelingen til Z_n er normalisert, så må vi også normere histogrammet for at det skal likne på fordelingen til Z_n.

d)



Figur 1: Histogram for en uniform fordeling med $n = 5$

Vi ser over et histogram med de standariserte gjennomsnittene for en uniform fordeling. Siden vi regnet ut standardiserte gjennomsnitt av observasjoner av fordelingene, så vil vi etter sentralgrenseteoremet forvente at Z_n og dermed histogrammet skal gå mot en normalfordeling. Vi kan se at histogrammet over er veldig nært en normalfordeling. Rett rundt $\mu = 0$ er det noen ekstra topper, men ellers er fordelingen ganske normalfordelt.

e)

Verdiene under finnes i utskrifen til python-programmet.

Tabell for sannsynlighetene til en standardnomalfordelt variable:

Interval	Sannsynlighet
$-\infty$ til -2.5	0.0062
-2.5 til -2.0	0.0165
-2.0 til -1.5	0.0441
-1.5 til -1.0	0.0919
-1.0 til -0.5	0.1499
-0.5 til 0.0	0.1915
0.0 til 0.5	0.1915
0.5 til 1.0	0.1499
1.0 til 1.5	0.0919
1.5 til 2.0	0.0441
2.0 til 2.5	0.0165
2.5 til ∞	0.0062

Tabell 1: Sannsynlighetene til en standardnomalfordelt variable

f)

På samme måte som vi fant verdiene for en standardnomalfordelt variable, kan vi finne verdiene for vår fordelinge av det standardiserte gjennomsnittet av den uniforme fordelingen:

Interval	Sannsynlighet
$-\infty$ til -2.5	0.0044
-2.5 til -2.0	0.0171
-2.0 til -1.5	0.0457
-1.5 til -1.0	0.0926
-1.0 til -0.5	0.1486
-0.5 til 0.0	0.1905
0.0 til 0.5	0.1931
0.5 til 1.0	0.1494
1.0 til 1.5	0.0901
1.5 til 2.0	0.0469
2.0 til 2.5	0.0175
2.5 til ∞	0.0041

Tabell 2: Sannsynlighetene til en unifrom fordeling med $n = 5$

Vi kan se at disse verdiene er veldig nære de vi fant i deloppgaven over 1. Vi kan finne absoluttveriden til forskjellen mellom fordelingen:

Interval	Forskjell
$-\infty$ til -2.5	0.0018
-2.5 til -2.0	0.0056
-2.0 til -1.5	0.00164
-1.5 til -1.0	0.0008
-1.0 til -0.5	0.0012
-0.5 til 0.0	0.0010
0.0 til 0.5	0.0016
0.5 til 1.0	0.0005
1.0 til 1.5	0.0015
1.5 til 2.0	0.0028
2.0 til 2.5	0.0010
2.5 til ∞	0.0021

Tabell 3: Forskjellen til en unifrom fordeling med $n = 5$

Vi ser her at forskjellen mellom standardnormalfordelingen og vår fordeling er svært liten, som vi forventet ut i fra histogrammet i 1.

Jeg kommer til å ha sannsynligheten og forskjellen fra normalfordelingen i de samme tabellene i resten av oppgaven.