Stk1100 Oblig 2

Soran Hussein Mohmmed / soranhm

April 2017

Oppgave 1

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

der k er konstant.

a)

for å finne k bruker jeg at tettheten skal være lik 1:

$$\begin{aligned} &\int_0^1 \int_0^x k(x-y)dydx = k\int_0^1 \Big[xy - \frac{1}{2}y^2\Big]_0^x dx \\ &= k\int_0^1 (x^2 - \frac{1}{2}x^2)dx = k\int_0^1 \frac{1}{2}x^2 dx = k\Big[\frac{1}{6}x^3\Big]_0^1 \\ &= \frac{k}{6} \end{aligned} \tag{1}$$

finner nå k ved å sette tettheten lik 1: $k/6 = 1 \rightarrow k = 6$.

b)

Skal nå bestemme P(2Y $\leq$ X):

$$\begin{aligned} P(2Y \leq X) &= P(Y \leq \frac{X}{2}) = 6\int_0^1 \int_0^{\frac{x}{2}} (x-y)dydx = 6\int_0^1 \Big[xy - \frac{1}{2}y^2\Big]_0^{\frac{x}{2}} dx \\ &= 6\int_0^1 \Big[\frac{x^2}{2} - \frac{x^2}{8}\Big]_0^{\frac{x}{2}} dx = 6\int_0^1 \frac{6}{16}x^2 dx = 6\Big[\frac{6}{16}\frac{1}{3}x^3\Big] = \frac{6*2}{16} * 1 \\ &= \frac{12}{16} = \frac{3}{4} \end{aligned} \tag{2}$$

c)

Marginale sannsynlighetstettheten til X: for $0 \leq x \leq 1$

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} d(x,y)dy = 6\int_0^x (x-y)dy = 6\Big[xy - \frac{1}{2}y^2\Big]_0^x \\ &= 6*(x^2 - \frac{1}{2}x^2) = 6*\frac{1}{2}x^2 = 3x^2 \end{aligned} \tag{3}$$

Dette gir da : (0 ellers)

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

d)

Marginale sannsynlighetstettheten til Y: for $0 \leq y \leq 1$

$$\begin{aligned} f_Y(x) &= \int_{-\infty}^{\infty} d(x,y)dx = 6\int_y^1 (x-y)dx = 6\Big[\frac{1}{2}x^2 - yx\Big]_y^1 \\ &= 6*\Big[(\frac{1}{2} - y) - (\frac{1}{2}y^2 - y^2)\Big] = 6\Big[\frac{1}{2} - y + \frac{1}{2}y\Big] \\ &= 3 - 6y + 2y^2 \end{aligned} \tag{4}$$

Dette gir da : (0 ellers)

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

e)

for at X og Y skal være uavhengige så må f(x,y) = f(x)f(y) for alle x og y. Vi har at $0 \leq y \leq x \leq 1$: $f_X(x)f_Y(y) = 3x^2 * (3 - 6y + 3y^2) = 9x^2 - 18x^2y + 9x^2y^2 \neq 6(x-y) = f(x,y)$ Dermed har vi at X og Y ikke er uavhengige.

Oppgave 2

$X_1, X_2, ..., X_n$ er uavhengig og identisk fordelt stokastiske varibler. $\mu = E(X_i)$ og $\sigma^2 = V(X_i)$. Gjennomsnittet:

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{5}$$

og den standardiserte gjennomsnittet:

$$Z_n = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \tag{6}$$

a)

Skal vise at $E(\bar{X}_n) = \mu, V(\bar{X}_n) = \frac{\sigma^2}{n}, E(Z_n) = 0$ og $V(Z_n) = 1$.

$$E(\bar{X}_n) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}E(\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu \tag{7}$$

$$V(\bar{X}_n) = V(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}V(\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n}V(X_i) = \frac{1}{n^2} * n * \sigma^2 = \frac{\sigma^2}{n} \tag{8}$$

$$\begin{aligned} E(Z_n) &= E(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}) = \frac{1}{\sigma/\sqrt{n}}E(\bar{X}_n - \mu) \\ &= \frac{1}{\sigma/\sqrt{n}}E(\bar{X}_n) - E(\mu) = \frac{1}{\sigma/\sqrt{n}}(\mu - \mu) = 0 \end{aligned} \tag{9}$$

$$\begin{aligned} V(Z_n) &= V(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}) = (\frac{1}{\sigma/\sqrt{n}})^2 V(\bar{X}_n - \mu) \\ &= \frac{1}{(\sigma/\sqrt{n})^2}V(\bar{X}_n) - V(\mu) = \frac{n}{\sigma^2}\frac{\sigma^2}{n} = 1 \end{aligned} \tag{10}$$

b)

Finner $\mu = E(X_i)$ og $\sigma^2 = V(X_i)$ for de tre fordelingene:
Uniforme fordelingen:

$$\begin{aligned} \mu = E(X_i) &= \int_{-1}^{1} x * \frac{1}{2}dx = \Big[\frac{1}{4}x^2\Big]_{-1}^{1} = \frac{1}{4} - \frac{1}{4} = 0 \\ E(X_i^2) &= \int_{-1}^{1} x^2 * \frac{1}{2}dx = \Big[\frac{1}{6}x^3\Big]_{-1}^{1} = \frac{1}{6} + \frac{1}{6} = \frac{1}{3} \\ \sigma^2 = V(X_i) &= E(X_i^2) - (E(X_i))^2 = \frac{1}{3} - 0^2 = \frac{1}{3} \end{aligned}$$

Gammafordelingen:

$$\begin{aligned} \mu = E(X_i) &= \alpha\beta = \frac{1}{2} * 1 = \frac{1}{2} \\ \sigma^2 = V(X_i) &= \alpha\beta^2 = \frac{1}{2} * 1^2 = \frac{1}{2} \end{aligned}$$

Bernoullifordelingen:

$$\begin{aligned} \mu = E(X_i) &= 0 * p(0) + 1 * p(1) = 0 + 0.75 = \frac{3}{4} \\ E(X_i^2) &= 0^2 * p(0) + 1^2 * p(1) = 0 + 0.75 = \frac{3}{4} \\ \sigma^2 = V(X_i) &= E(X_i^2) - (E(X_i))^2 = \frac{3}{4} - (\frac{3}{4})^2 = \frac{3}{16} \end{aligned}$$

c)

En normert historgram vi ligge nær sannsynlighetstettheten til den stokastiske variabelen. Når vi trekker fra uniforme og gamma-fordelingen, vil Z_n være kontinuerlig stokstisk variabel. Da vil histogrammet med observasjoner av Zn ligge nær sannsynlighetstettheten til Z_n. Når trekker fra Bernoulli fordelingen, vil Z_n være diskret stokastisk variabel. Denne vil også være det likt som det kontinuerlige tilfellet, da kan vi argumentere at histogramet av Z_n vil ligge nær punktsannsynligheten til Z_n.

d)

Bruker kommandoene som er gitt og lager et histogram med klasse bredde 0.25. Histogrammet er nesten symmetrisk om 0, pga den uniforme fordelingen er symmetrisk om sin forventningsverdi

e) og f)

Bruker tabell A.3 i læreboka og regner ut sannsynligheten for standardnormalfordelingen i intervallene som er gitt. Eks: [-2.5,-2.0) : $\Phi(-2.0)-\Phi(-2.5) = 0.0228-0.0062 = 0.0166$

Intervall	A.3	Z_5
$(-\infty, -2.5)$	0.0062	0.0038
$[-2.5, -2.0)$	0.0166	0.0147
$[-2.0, -1.5)$	0.0440	0.0438
$[-1.5, -1.0)$	0.0919	0.0944
$[-1.0, -0.5)$	0.1498	0.1549
$[-0.5, 0)$	0.1915	0.1833
$[0, 0.5)$	0.1915	0.1929
$[0.5, 1.0)$	0.1498	0.1438
$[1.0, 1.5)$	0.0919	0.0979
$[1.5, 2.0)$	0.0440	0.0483
$[2.0, 2.5)$	0.0166	0.0174
$[2.5, \infty)$	0.0062	0.0048

Har satt tallene fra matlab og fra tabellen i boka i tabellen ovenfor. Ser at den relative frekvensene av verdien Z_5 er ganske nærme standardnormalfordelingen fra boka.

g)

Setter nå n = 15 og n= 50 og gjør det samma som jeg gjorde i d) og f). Har satt tallene inn i tabellen over. ser at det blir nærmere og nærmere med høyere n verdi. Verdiene jeg får ut variere fra gang til gang.

```
n = 5; n2 = 15 ; n3 = 50;
my = 0; SD = sqrt(1/3);
```

```
X = unifrnd(−1,1,n,10000);     % uniform fordelt, n= 5
X2 = unifrnd(−1,1,n2,10000);   % uniform fordelt, n = 15
X3 = unifrnd(−1,1,n3,10000);   % uniform fordelt, n = 50

meanX = mean(X);               % Gjennomsnitt
meanX2 = mean(X2);             % Gjennomsnitt
meanX3 = mean(X3);             % Gjennomsnitt

Z = sqrt(n)*(meanX−my)/SD;

% OPPGAVE 2 e og f

int = [−Inf, −2.5, −2, −1.5, −1, −0.5, 0, 0.5, 1, 1.5, 2, 2.5, Inf];
ant = histc(Z, int);
relfrekv = ant(1:12)/10000;

% OPPGAVE 2 g

Z2 = sqrt(n2)*(meanX2−my)/SD;
Z3 = sqrt(n3)*(meanX3−my)/SD;
ant2 = histc(Z2, int);
ant3 = histc(Z3, int);
relfrekv2 = ant2(1:12)/10000;
relfrekv3 = ant3(1:12)/10000;

% plot
subplot(2,2,1); histogram(Z,−3:0.25:3); title('n= 5');
subplot(2,2,2); histogram(Z2,−3:0.25:3); title('n= 15');
subplot(2,2,3); histogram(Z3,−3:0.25:3); title('n= 50');
```

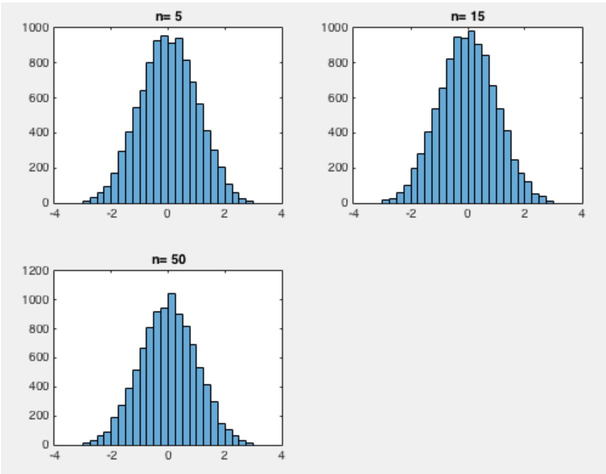

Figure 1: Grafene til uniforme fordelingen

Intervall	A.3	Z_{15}	Z_{50}
$(-\infty, -2.5)$	0.0062	0.0056	0.0064
$[-2.5, -2.0)$	0.0166	0.0182	0.0156
$[-2.0, -1.5)$	0.0440	0.0492	0.0455
$[-1.5, -1.0)$	0.0919	0.1005	0.0944
$[-1.0, -0.5)$	0.1498	0.1507	0.1466
$[-0.5, 0)$	0.1915	0.1801	0.1867
$[0, 0.5)$	0.1915	0.1866	0.1935
$[0.5, 1.0)$	0.1498	0.1498	0.1478
$[1.0, 1.5)$	0.0919	0.0918	0.0965
$[1.5, 2.0)$	0.0440	0.0440	0.0459
$[2.0, 2.5)$	0.0166	0.0165	0.0158
$[2.5, \infty)$	0.0062	0.0070	0.0053

h)

Ut fra tallene og figur ser jeg at denne bruker litt lengere tid (større n) for å bli lik standardnormalfordelingen.

```
n = 5; n2 = 15 ; n3 = 50;
my = 1/2; SD = sqrt(1/2);

X = gamrnd(0.5,1,n,10000);     % gamma fordelt, n = 5
X2 = gamrnd(0.5,1,n2,10000);   % gamma fordelt, n = 15
X3 = gamrnd(0.5,1,n3,10000);   % gamma fordelt, n = 50

meanX = mean(X);               % Gjennomsnitt
meanX2 = mean(X2);             % Gjennomsnitt
meanX3 = mean(X3);             % Gjennomsnitt

int = [-Inf, -2.5, -2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5, 2, 2.5, Inf];


Z = sqrt(n)*(meanX-my)/SD;
Z2 = sqrt(n2)*(meanX2-my)/SD;
Z3 = sqrt(n3)*(meanX3-my)/SD;

ant = histc(Z, int);
ant2 = histc(Z2, int);
ant3 = histc(Z3, int);
relfrekv = ant(1:12)/10000;
relfrekv2 = ant2(1:12)/10000;
relfrekv3 = ant3(1:12)/10000;

% plot
subplot(2,2,1); histogram(Z,-3:0.25:3); title('n= 5')
subplot(2,2,2); histogram(Z2,-3:0.25:3); title('n= 15')
subplot(2,2,3); histogram(Z3,-3:0.25:3); title('n= 50')
```

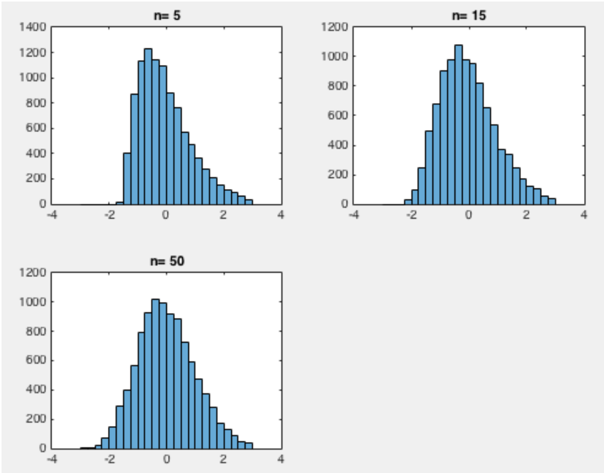

Figure 2: Grafene til Gammafordelingen

Intervall	A.3	Z_5	Z_{15}	Z_{50}
$(-\infty, -2.5)$	0.0062	0	0	0.0009
$[-2.5, -2.0)$	0.0166	0	0.0029	0.0089
$[-2.0, -1.5)$	0.0440	0.0017	0.0334	0.0431
$[-1.5, -1.0)$	0.0919	0.1272	0.1181	0.0967
$[-1.0, -0.5)$	0.1498	0.2363	0.1890	0.1717
$[-0.5, 0)$	0.1915	0.2243	0.2059	0.2019
$[0, 0.5)$	0.1915	0.1635	0.1779	0.1800
$[0.5, 1.0)$	0.1498	0.1034	0.1191	0.1319
$[1.0, 1.5)$	0.0919	0.0637	0.0709	0.0854
$[1.5, 2.0)$	0.0440	0.0365	0.0414	0.0449
$[2.0, 2.5)$	0.0166	0.0206	0.0225	0.0224
$[2.5, \infty)$	0.0062	0.0228	0.0189	0.0122

i)

Her tar det ekstra lengere tid (større n) for at vi skal nærme oss standardnormalfordelingen, desto høyere n vi bruker desto flere og bedre tall får vi.

```
n = 5; n2 = 15 ; n3 = 50;
my = 3/4; SD = sqrt(3/16);

X = binornd(1,0.75,n,10000);     % Bernoulli fordelt, n = 5
X2 = binornd(1,0.75,n2,10000);   % Bernoulli fordelt, n = 15
X3 = binornd(1,0.75,n3,10000);   % Bernoulli fordelt, n = 50

meanX = mean(X);               % Gjennomsnitt
meanX2 = mean(X2);               % Gjennomsnitt
meanX3 = mean(X3);               % Gjennomsnitt
```

```
int = [-Inf, -2.5, -2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5, 2, 2.5, Inf];

Z = sqrt(n)*(meanX-my)/SD;
Z2 = sqrt(n2)*(meanX2-my)/SD;
Z3 = sqrt(n3)*(meanX3-my)/SD;

ant = histc(Z, int);
ant2 = histc(Z2, int);
ant3 = histc(Z3, int);
relfrekv = ant(1:12)/10000;
relfrekv2 = ant2(1:12)/10000;
relfrekv3 = ant3(1:12)/10000;

% plot
subplot(2,2,1); histogram(Z,-3:0.25:3); title('n= 5');
subplot(2,2,2); histogram(Z2,-3:0.25:3); title('n= 15');
subplot(2,2,3); histogram(Z3,-3:0.25:3); title('n= 50');
```

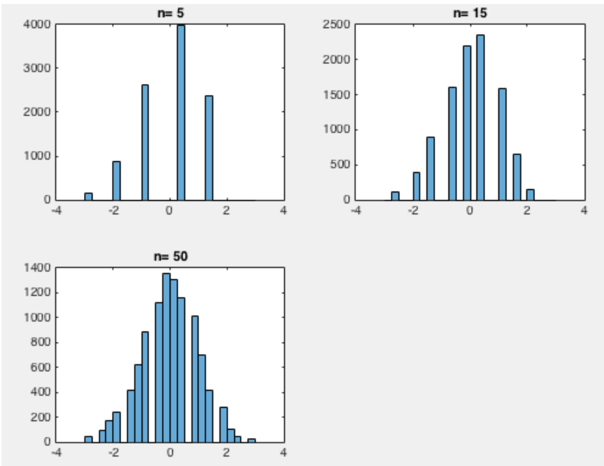

Figure 3: Grafene til Bernoulli-fordelingen

Intervall	A.3	Z_5	Z_{15}	Z_{50}
$(-\infty, -2.5)$	0.0062	0.0157	0.0171	0.0072
$[-2.5, -2.0)$	0.0166	0	0	0.0262
$[-2.0, -1.5)$	0.0440	0.0866	0.0392	0.0237
$[-1.5, -1.0)$	0.0919	0	0.0896	0.1037
$[-1.0, -0.5)$	0.1498	0.2626	0.1608	0.0889
$[-0.5, 0)$	0.1915	0	0.2196	0.2469
$[0, 0.5)$	0.1915	0.3977	0.2346	0.2470
$[0.5, 1.0)$	0.1498	0	0	0.1013
$[1.0, 1.5)$	0.0919	0.2374	0.1594	0.1108
$[1.5, 2.0)$	0.0440	0	0.0655	0.0276
$[2.0, 2.5)$	0.0166	0	0.0142	0.0143
$[2.5, \infty)$	0.0062	0	0	0.0024

j)

Ser at alle fordelingen går mot standardnormalfordelingen, men den uniformefordelingen går fortest, vi kan nesten se fra n = 5 at dette er bra fordeling. Dette kan skyldes at den er kontinuerlig og symmetrisk. Gammafordelingen startet med å gi en unøyaktig fordeling for n = 5 men desto høyere n vi valgte desto bedre ble tilnærmingen, mens bernoulli-fordelingen starter med veldig få punkter og vil ikke bli bra tilnærming før stor n, dette skyldes at det er en diskret fordeling. Oppsummering har vi at tilnærmingen går fortere for kontinuerlige og symmetriske fordelinger enn diskret og ikke symmetriske.