Opgave 1: Stokastiske Variable

En kontinuert stokastisk variabel X har følgende fordelingsfunktion (cdf):

$$F_X(x) = \begin{cases} 0, & 2 \geq x \\ k \cdot x - \frac{2}{3}, & 2 < x \leq 5 \\ 1, & 5 < x \end{cases}$$

1) Vis at tæthedssfunktionen (pdf) er givet ved:

$$f_X(x) = \begin{cases} 0, & 2 \geq x \\ k, & 2 < x \leq 5 \\ 0, & 5 < x \end{cases}$$

From cdf to pdf: $f_X(x) = \frac{dF_X(x)}{dx}$

```
syms x k
Fx(x) = piecewise(2 >= x, 0, 2 < x <= 5, k * x -2/3, 5 < x, 1)
```

Fx(x) =

$$\begin{cases} 0 & \text{if } x \leq 2 \\ k\,x - \frac{2}{3} & \text{if } x \in (2, 5] \\ 1 & \text{if } 5 < x \end{cases}$$

```
fx(x) = diff(Fx(x))
```

fx(x) =

$$\begin{cases} 0 & \text{if } x < 2 \\ k & \text{if } x \in (2, 5) \\ 0 & \text{if } 5 < x \end{cases}$$

2) For hvilken værdi af k er $f_X(x)$ en gyldig tæthedsfunktion? Begrund svaret.

For at fx(x) er en gyldig tæthedsfunktion gælder følgende:

$$\int_{-\infty}^{\infty} fx(x)dx = \ 1$$

Samt at alle værdierne af pdf er større eller lig 0

```
teathed = int(fx(x), 'x', -Inf, Inf)
```

teathed = $3\,k$

```
k = solve(teathed==1, k)
```

k =

$$\frac{1}{3}$$

```

```

3) Skitsér tæthedsfunktionen og angiv navnet på fordelingsfunktionen.

```
clearvars
syms x
k = 1/3
```

k = 0.3333

```
Fx(x) = piecewise(2 >= x, 0, 2 < x <= 5, k * x -2/3, 5 < x, 1);
fx(x) = diff(Fx(x));

fplot(fx(x), [-4, 13],'-r','LineWidth',2)

title('Tæthedsfunktion')
xlabel('x')
ylabel('y')
```

Fordelingsfunktionen er uniform fordelt

4) Brug $F_X(x)$ til at beregne sandsynligheden $\Pr(x \geq 3)$. Antag at $k = \frac{1}{3}$.

```
Pr=1-Fx(3)
```

Pr =

$$\frac{2}{3}$$

5) Bestem forventningsværdien og variansen af X udfra $f_X(x)$. Angiv desuden hvilken formler, der bruges til at bestemme værdierne. Antag at $k = \frac{1}{3}$.

Forventningsværdien givet ved

$$E\ |X| = \int x \cdot f(x)\,\mathrm{d}x$$

```
Ex=int(x*fx(x),'x',-Inf,Inf)
```

Ex =

$$\frac{7}{2}$$

Variance er givet ved

$$Var(x) = E\ |x^2| - E\ |X|^2$$

```
var = int(x^2*fx(x),'x',-Inf,Inf)-(Ex^2)
```

var =

$$\frac{3}{4}$$

Opgave 2: Stokastiske Processer

En kontinuer stokastisk process er givet ved:

$$X(t) = w + 4$$

Hvor w er normalfordelt efter $w \sim N(5,1)$.

1) Skitsér fem realisationer af processen $X(t)$ mellem $t \in [0; 7]$. Brug en Gauss-generator, det kan evt. være matlabs indbyggede generator, *randn()*. Angiv desuden hvordan de fem realisationer er fremkommet.

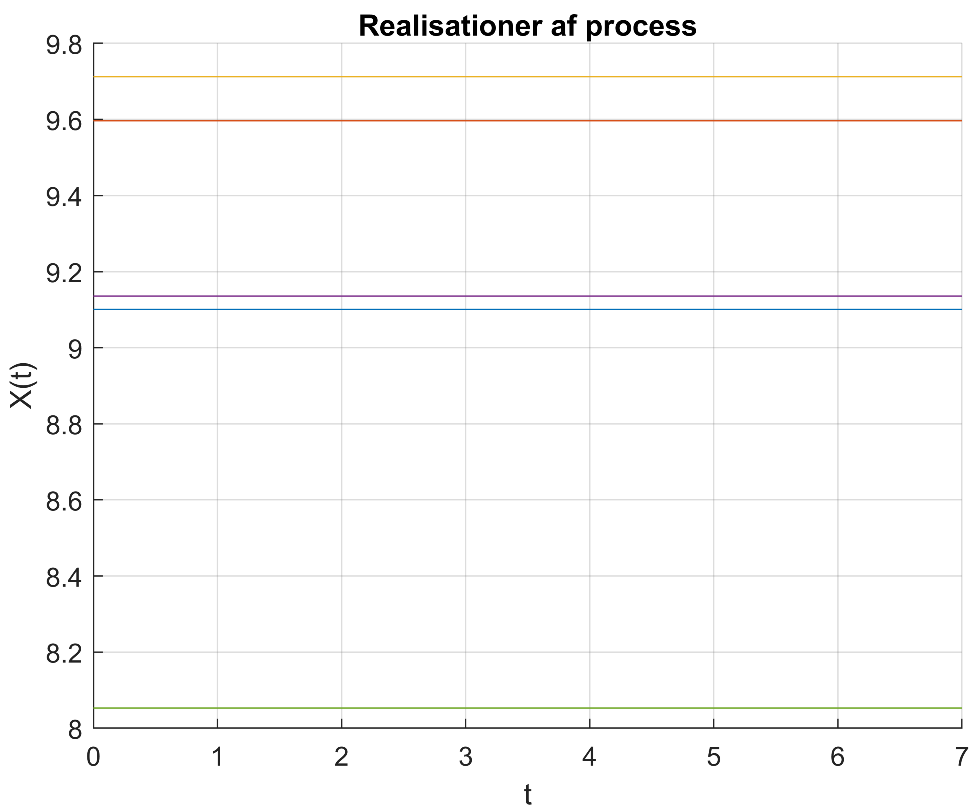
Normal fordelt - Middelværdi = 5, varians = 1

```
clearvars
close all
% Antal realisationer
```

```
N = 5;
t_start = 0;
t_end = 7;
interval = t_end-t_start;
mean = 5; variance = 1; sigma = sqrt(variance);
w = normrnd(mean, sigma, [1 N+1])
```

```
w = 1×6
    5.1008    5.5964    5.7122    5.1356    4.0531    4.1531
```

```
t = t_start:1:t_end;
figure(1)
hold on
grid
xlabel("t"); ylabel("X(t)"); title("Realisationer af process");
for i=1:N
    y = w(i)+4;
    fplot(y,[0 7])
end
```



2) Bestem ensemble middelværdien og ensemble variansen for processen $X(t)$.

Vi kender middelværdien for normalfordelingen W.

```
ensembleMean = mean+4
```

```
ensembleMean = 9
```

Vi kender variansen for norfordelingen W

$$Var(x(t)) = Var(w) + Var(k)$$

```
ensembleVar = variance + (4^2-4^2)
```

```
ensembleVar = 1
```

3) Udvælg én af de fem realisationer, og bestem middelværdien og variansen for denne realisation.

```
firstsRealisation = w(1)+4
```

```
firstsRealisation = 9.1008
```

```
Ex = firstsRealisation
```

```
Ex = 9.1008
```

```
% Varians = E[X^2] - E[X]^2
Ex2 = firstsRealisation^2
```

```
Ex2 = 82.8248
```

```
var = Ex2-Ex^2
```

```
var = 0
```

4) Angiv om processen $X(t)$ er WSS (stationær i den brede forstand) og om den er ergodisk. Begrund dine svar.

Da processen er uafhængig af tiden, og at middelværdien og variansen ikke ændrer sig, så kan vi konstanter at processen er stationær.

Processen er ikke ergodisk, da en realisation ikke siger noge om middelværdien, og variansen af hele prossen.

Opgave 3: Sandsynlighedsregning

Hændelse A er, at en gravid fødte en pige i 2012.

Hændelse B er, at hun fødte en dreng.

Hændelse C er, at hun fødte et barn, der vejer over 4000g.

20,2% af alle nyfødte drenge vejede i 2012 over 4000g. 12,8% af nyfødte piger vejede i 2012 over 4000g.

1) Hvis der blev født 29.785 drenge og 28.131 piger i 2012, hvad er sandsynligheden for hændelse A?

```
piger = 28131
```

```
piger = 28131
```

```
drenge = 29785
```

```
drenge = 29785
```

```
total = piger + drenge
```

```
total = 57916
```

```
Pr_a=piger/(total)
```

```
Pr_a = 0.4857
```

2) Hvad er den totale sandsynlighed for hændelse C?

Total sandsynlighed er givet ved

$$\begin{aligned}\Pr(A) &= \Pr(A \cap B) + \Pr(A \cap \bar{B}) \\ &= \Pr(A|B) \cdot \Pr(B) + \Pr(A|\bar{B}) \cdot \Pr(\bar{B})\end{aligned}$$

```
Pr_b= 1-Pr_a
```

```
Pr_b = 0.5143
```

```
Pr_C_A = 0.128
```

```
Pr_C_A = 0.1280
```

```
Pr_C_B = 0.202
```

```
Pr_C_B = 0.2020
```

```
Pr_C = ((Pr_C_A*Pr_a)+(Pr_C_B*Pr_b))
```

```
Pr_C = 0.1661
```

3) Hvad var sandsynligheden for at den fødende fik en pige, hvis det oplyses, at hendes barn vejede over 4000g ved fødslen?

Bayes Rule:

$$\boxed{Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)} = \frac{Pr(B|A) \cdot Pr(A)}{Pr(B)}}$$

```
(Pr_C_A*Pr_a)/Pr_C
```

```
ans = 0.3744
```

Opgave 4: Statistik

Vi måler højden på studerende i en klasse, der består af 19 kvinder og 35 mænd. Højderne antages at være normalfordelt. Middelværdien for kvinder i klassen er $\widehat{\mu_1} = 1{,}68m$, med en estimeret varians på ${s_1}^2 = 0{,}10$. Middelværdien for mænd i klassen er $\widehat{\mu_2} = 1{,}78m$, med en estimeret varians på ${s_2}^2 = 0{,}20$.

1) Opstil et hypotese test, for at bestemme om middelværdien af mænd og kvinder i klassen er den samme.

Null hypothesis. Brugt som testen af ingen forskel.

$H_0 = \mu 1 = \mu 2$

Alternanive hypotese

$H_1 = \mu 1 \neq \mu 2$

2) Angiv hvilken statistisk test du vil udføre for at teste hypotesen.

Begrund dit svar.

Da variansen er ukendt, og dataen er normalt fordelt, så bruges der en t-test på forskellen af middelværdierne.

3) Estimér forskellen i middelværdierne δ og standard afvigelsen σ for forskellen.

Estimator

$$\hat{\delta} = \bar{x}_1 - \bar{x}_2 \sim N\left(\mu_1 - \mu_2, \sigma_1^2/n_1 + \sigma_2^2/n_2\right)$$

```
middel= abs(1.68-1.78)
```

```
middel = 0.1000
```

Standard deviation er givet ved:

$$\sigma = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1) * s_2^2}{n_1 + n_2 - 2}}$$

```
n1 = 19;
n2 = 35;
s1_2= 0.1;
s2_2 = 0.2;
standard=sqrt(((n1-1)*s1_2+(n2-1)*s2_2)/(n1+n2-2))
```

```
standard = 0.4067
```

4) Anvend en t-test til hypotese test af din hypotese. Kan NULL hypotesen afvises med et signifikansniveau på 0,05? Begrund dit svar.

$$t = \frac{\hat{\delta}}{s * \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$$

```
t = middel/(standard*sqrt((1/n1)+(1/n2)))
```

```
t = 0.8629
```

P-value

$$2 \cdot \left(1 - t_{cdf}(|t|, n_1 + n_2 - 2)\right)$$

```
pval = 2*(1-tcdf(abs(t),(n1+n2)-2))
```

```
pval = 0.3921
```

Da pval > 0.05 så fejler vi med at afvise null hypotesen.

5) Opstil og find 95% konfidens intervallet for δ. Angiv hvilken formel, der er brugt.

Lower bound:

$$\delta_- = (\bar{x}_1 - \bar{x}_2) - t_0 \cdot s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Upper bound:

$$\delta_+ = (\bar{x}_1 - \bar{x}_2) + t_0 \cdot s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

```
t0=tinv(1-0.05/2,(n1+n2-2))
```

```
t0 = 2.0066
```

```
koefMinus = middel-t0*standard*sqrt((1/n1)+(1/n2))
```

```
koefMinus = -0.1325
```

```
koefPositiv = middel+t0*standard*sqrt((1/n1)+(1/n2))
```

```
koefPositiv = 0.3325
```