

FORRÅR 2018 EKSAMEN

MADS STEINER KRISTENSEN - 201405230

AARHUS UNIVERSITY - AARHUS SCHOOL OF ENGINEERING

TABLE OF CONTENTS

FORRÅR 2018 EKSAMEN.....	1
1 STOKASTISKE VARIABLE.....	1
1.1 TÆTHEDSFUNKTION.....	1
1.2 GYLDIG TÆTHEDSFUNKTION.....	2
1.3 SKITSE AF TÆTHEDSFUNKTION.....	2
1.4 SANDSYNLIGHED.....	3
1.5 FORVENTNINGSVÆRDI & VARIANS.....	4
2 STOKASTISKE PROCESSER.....	4
2.1 REALISERING.....	4
2.2 MIDDELVÆRDI OG VARIANS.....	6
2.3 REALISERINGS MIDDELVÆRDI OG VARIANS.....	6
2.4 PROCESSEN.....	6
3 SANDSYNLIGHED.....	6
3.1 HÆNDELSE A.....	7
3.2 TOTALE SANDSYNLIGHED.....	7
3.3 BETINGET SANDSYNLIGHED.....	7
4 STATISTIK.....	8
4.1 HYPOTHESE.....	8
4.2 STATISTIK TEST.....	8
4.3 ESTIMERING AF VÆRDIER.....	8
4.4 T-TEST.....	9
4.5 KONFIDENSINTERVAL.....	9

```
clc
clear

addpath('[0] Library');
Color = load("colors.mat");
smp = load("library.mat");
```

1 STOKASTISKE VARIABLE

En kontinuert stokastisk variabel X har følgende fordelingsfunktion

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

1.1 TÆTHEDSFUNKTION

For at vise at tæthedsfunktionen er givet ved

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

Anvendes, at

$$f_X(x) = \frac{d}{dx} F_X(x)$$

```
syms x k
```

```
a = 0;
b = k * x - (2/3);
c = 1;
```

```
diff(a), diff(b), diff(c)
```

```
ans =
```

```
[]
```

```
ans = k
```

```
ans =
```

```
[]
```

Da [] betyder at værdien er 0 er tæthedsfunktionen $f_X(x)$ hermed eftervist at være korrekt.

1.2 GYLDIG TÆTHEDSFUNKTION

For at bestemme for hvilken værdi af k hvor $f_X(x)$ er en gyldig tæthedsfunktion anvendes, at

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

Da gælder i dette tilfælde, at

$$\int_a^b f_X(x) dx = 1$$

hvor $a = 2$ og $b = 5$.

$$\int_2^5 f_X(x) dx = 1 \Rightarrow \int_2^5 k dx = 1$$

Hvilket giver at for værdien af k lig med

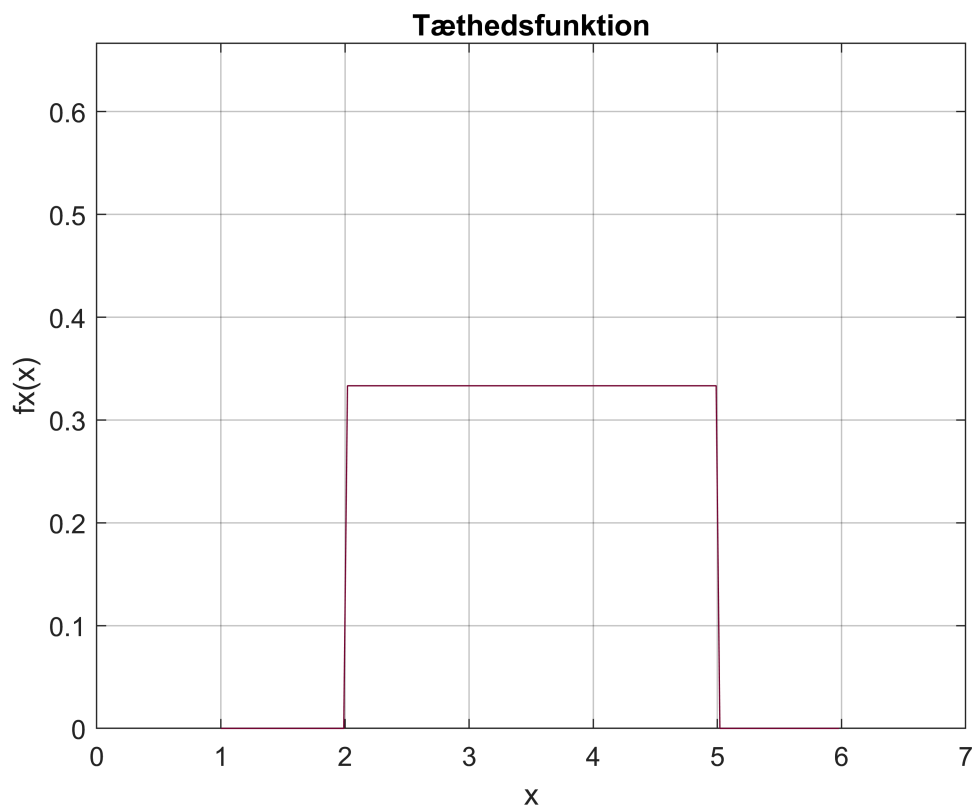
$$k = \frac{1}{3}$$

Opnåes en gyldig tæthedsfunktion.

1.3 SKITSE AF TÆTHEDSFUNKTION

En skitse af tæthedsfunktionen vises i matlab.

```
a = 2;  
b = 5;  
x = a-1:(b-a)/100:b+1;  
fx = unifpdf(x, a, b);  
  
figure(1)  
plot(x, fx, 'Color', Color.DeepCleret);  
axis([a-2, b+2, 0, 2/(b-a)]);  
grid on  
title('Tæthedsfunktion')  
ylabel('fx(x)')  
xlabel('x')
```



1.4 SANDSYNLIGHED

For at bestemme sandsynligheden $P(x \geq 3)$ antages, at $k = 3$ og der anvendes følgende

$$P(x \geq 3) = 1 - P(x < 3)$$

$$P(x \geq 3) = 1 - F_X(3)$$

```
k = 1/3;  
Fx = @(x) k*x-(2/3);  
  
Probability = 1 - Fx(3)
```

Probability = 0.6667

Dermed er sandsynligheden $P(x \geq 3)$ bestemt til at være $\frac{2}{3}$.

1.5 FORVENTNINGSVÆRDI & VARIANS

For at bestemme forventningsværdien anvendes, at

$$E[X] = \bar{X} = \mu_x = \int_{-\infty}^{\infty} x \cdot f_x(x) dx \Rightarrow E[X] = \int_2^5 x \cdot k$$

hvor værdien for k antages, at være $\frac{1}{3}$.

```
k = 1/3;  
fx = @(x) x * k;  
EstimationValue = sym(integral(fx, 2, 5))
```

EstimationValue =

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

Forventningsværdien er dermed bestemt til, at være $\frac{7}{2}$. Variansen bestemmes dernæst ved følgende,

$$\text{Var}(x) = \int_{-\infty}^{\infty} (x - \bar{x}) \cdot f_X(x) dx = E[X^2] - E[X]^2$$

```
fx = @(x) x.^2 * k;  
Variance = sym(integral(fx, 2, 5)) - EstimationValue^2
```

Variance =

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

Variansen er dermed bestemt til, at være $\frac{3}{4}$.

2 STOKASTISKE PROCESSER

En kontinuer stokastisk process er givet ved

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

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

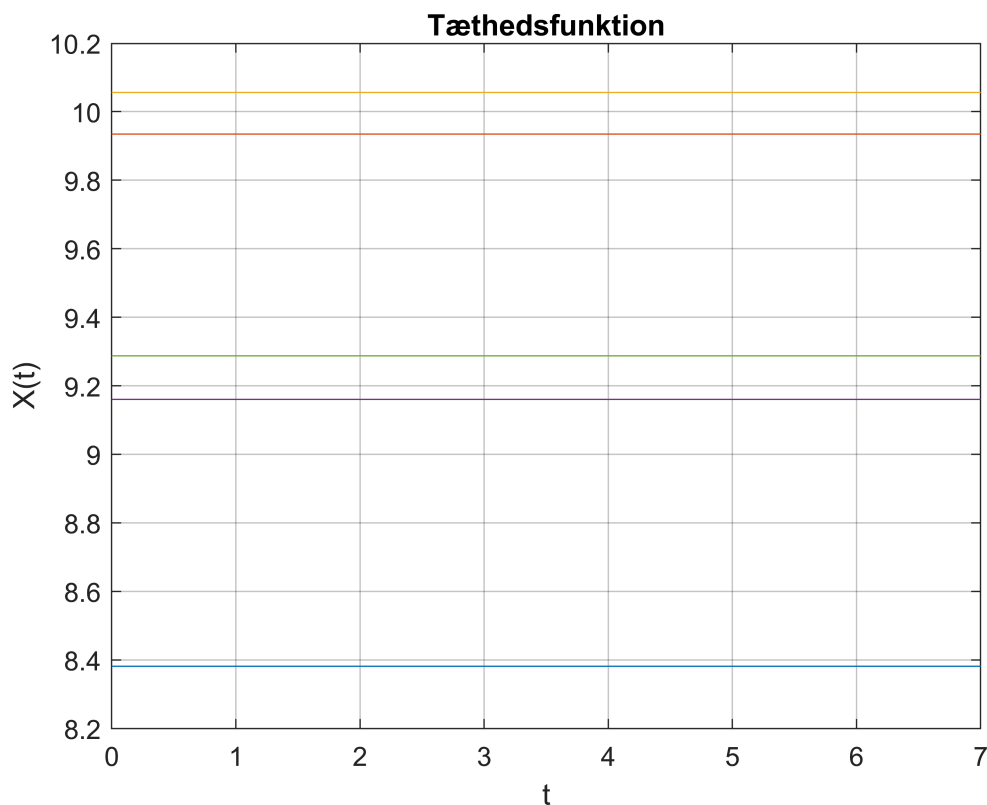
2.1 REALISERING

For at skitsere en realisering af den kontinuere process vides det, at

$$\sigma^2 = 1, \mu = 5$$

Dertil kan realiseringen laves ved brug af matlab.

```
t = 0:1:7;  
mu = 5;  
sigma = sqrt(1);  
  
Realisation = [],[],[],[],[];  
  
for n = 1:5  
    Realisation(n, :) = (ones(1, 8) * (sigma * randn + mu)) + 4;  
end  
  
figure(3)  
plot(t, Realisation(1, :));  
hold on  
plot(t, Realisation(2, :));  
plot(t, Realisation(3, :));  
plot(t, Realisation(4, :));  
plot(t, Realisation(5, :));  
grid on  
title('Tæthedsfunktion')  
ylabel('X(t)')  
xlabel('t')  
hold off
```



2.2 MIDDELVÆRDI OG VARIANS

For at bestemme middelværdien for ensemble anvendes, at

$$E[X(t)] = E[w + 4] = E[w] + E[4]$$

```
Ensemble.MeanValue = mu + 4;
```

Variansen kan bestemmes som

$$\text{Var}(X(t)) = \text{Var}[w + 4] = \text{Var}[w] + \text{Var}[4]$$

```
Ensemble.Variance = sigma^2 + 0;
```

```
Ensemble
```

```
Ensemble = struct with fields:  
  MeanValue: 9  
  Variance: 1
```

Dermed er middelværdien bestemt til at være 9 og variansen bestemt til at være 1.

2.3 REALISERINGS MIDDELVÆRDI OG VARIANS

```
Real.Mean = mean(Realisation(1, :));  
Real.Variance = round(var(Realisation(1, :)), 10);
```

```
Real
```

```
Real = struct with fields:  
  Mean: 8.3813  
  Variance: 0
```

Middelværdi og variansen for en udvalgt realisering er bestemt til at være værdierne vist ovenfor.

2.4 PROCESSEN

Da ensemble middelværdi og variansen er uafhængig af tiden og samtidig konstant er processen WSS. Processen er ikke ergodisk da middelværdien og variansen af en realisering ikke siger noget om en middelværdi og varians for hele processen.

3 SANDSYNLIGHED

A: GRAVID FØDTE EN PIGE I 2012

B: GRAVID FØDTE EN DRENG

C: GRAVID FØDTE ET BARN OVER 4000 G.

20.2% af alle nyfødte drenge vejede i 2012 over 4000g. 12.8% af nyfødte piger vejede i 2012 over 4000g.

$$P(C|B) = 0.202$$

$$P(C|A) = 0.128$$

3.1 HÆNDELSE A

For at bestemme hvad sandsynligheden er for hændelse A, vides det, at der blev født 29.785 drenge og 28.131 piger i 2012. Da anvendes den relative frekvens som er bestemt ved følgende udtryk.

$$P(A) = \frac{N_A}{N} = \frac{28.131}{28.131 + 29.785}$$

```
Boys = 29785;
Girls = 28131;
```

```
A = Girls / (Girls + Boys)
```

```
A = 0.4857
```

Sandsynligheden for hændelse **A** er dermed bestemt til at være 48.57%.

3.2 TOTALE SANDSYNLIGHED

Den totale sandsynlighed for hændelse **C** kan bestemmes ved, at anvende følgende.

$$P(A) = P(A \cap B) + P(A \cap \bar{B})$$

$$P(A) = P(A|B) \cdot P(B) + P(A|\bar{B}) \cdot P(\bar{B})$$

Der antages i dette tilfælde for opgaven, at $P(B) = 1 - P(A)$ da den modsatte hændelse at at få en pige vil være at få en dreng.

```
B = 1 - A;
CGivenB = 0.202;
CGivenA = 0.128;
```

```
C = CGivenA * A + CGivenB * B
```

```
C = 0.1661
```

Den totale sandsynlighed for hændelse **C** er dermed bestemt til at være 16.61%.

3.3 BETINGET SANDSYNDLIGHED

For at bestemme hvad sandsynligheden er for at den fødende fik en pige, hvis det oplyses, at hendes barn vejede over 4000g ved fødslen anvendes, at

$$P(A|C) = \frac{P(C|A) \cdot P(A)}{P(C)}$$

```
AGivenC = (CGivenA * A) / C
```

```
AGivenC = 0.3744
```

Dermed er sandsynligheden for, at den fødende fik en pige givet at barnet vejde over 4000g ved fødslen, bestemt til at være 37.44%.

4 STATISTIK

Der måles 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 $\hat{\mu}_1 = 1.68m$, med en estimeret varians på $\sigma_1^2 = 0.10$.

Middelværdien for mænd i klassen er $\hat{\mu}_1 = 1.78m$, med en estimeret varians på $\sigma_1^2 = 0.20$.

```
Women.Sigma = 0.1;
Women.Mu = 1.68;
Women.Amount = 19;

Men.Sigma = 0.20;
Men.Mu = 1.78;
Men.Amount = 35;
```

4.1 HYPOTESE

Der opstilles en hypotese test for at bestemme om middelværdien af mænd og kvinder i klassen er den samme.

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

4.2 STATISTIK TEST

Da varianserne er ukendte og data er normal-fordelt anvendes en t-test på forskellen af middelværdierne.

4.3 ESTIMERING AF VÆRDIER

Estimeringen af forskellen i middelværdierne kan bestemmes som følgende.

$$\hat{\delta} = |\bar{x}_1 - \bar{x}_2|$$

```
Delta = abs(Women.Mu - Men.Mu)
```

```
Delta = 0.1000
```

Forskellen i middelværdierne er bestemt til at være 0.1. Dernæst til at bestemme forskellen i standard afvigelsen og dertil anvendes, at

$$s^2 = \frac{1}{n_1 + n_2 - 2} [(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2]$$

```
PooledVariance = sqrt((1 / (Women.Amount + Men.Amount - 2)) ...
* ((Women.Amount - 1)*Women.Sigma + (Men.Amount - 1)*Men.Sigma))
```

```
PooledVariance = 0.4067
```


Forskellen i standard afvigelsen er dermed bestemt til at være 0.4067.

4.4 T-TEST

Der anvendes nu en t-test til test af hypotesen. For at undersøge om NULL hypotesen kan afvises med et signifikansniveau på 0.05 anvendes, at

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad p = 2 \cdot (1 - t_{\text{cdf}}(|t|, n_1 + n_2 - 2))$$

```
t = (Delta - 0) / (PooledVariance * sqrt(1/Women.Amount + 1/Men.Amount));
```

```
p = 2 * (1 - tcdf(abs(t), Women.Amount + Men.Amount - 2))
```

```
p = 0.3921
```

NULL hypotesen kan ikke afvises da $p > 0.05$.

4.5 KONFIDENSINTERVAL

For at finde 95% konfidens intervallet for δ anvendes, at

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

```
TZero = tinv(0.975, Women.Amount + Men.Amount - 2)
```

```
TZero = 2.0066
```

```
Step = TZero * PooledVariance * sqrt(1/Women.Amount + 1/Men.Amount);
```

```
Interval.Min = Delta - Step;
```

```
Interval.Max = Delta + Step;
```

```
Interval
```

```
Interval = struct with fields:
```

```
Min: -0.1325
```

```
Max: 0.3325
```

Konfidensintervallet for δ er dermed bestemt til at være [-0.1325; 0.3325].