

FORRÅR 2016 REEKSAMEN

MADS STEINER KRISTENSEN - 201405230

AARHUS UNIVERSITY - AARHUS SCHOOL OF ENGINEERING

TABLE OF CONTENTS

FORRÅR 2016 REEKSAMEN.....	1
1 STOKASTISKE VARIABLE.....	1
1.1 GYLDIG TÆTHEDSFUNKTION.....	1
1.2 SKITSERING AF TÆTHEDSFUNKTION.....	2
1.3 FORVENTNINGSVÆRDI OG VARIANS.....	3
1.4 SANDSYNLIGHEDER.....	4
1.5 FORDELINGSFUNKTION.....	4
2 STOKASTISKE PROCESSER.....	5
2.1 REALISERING.....	5
2.2 ENSEMBLE.....	6
2.3 TEMPORAL MEAN.....	6
2.4 PROCESSEN.....	7
2.5 AUTOKORRELATION.....	7
3 SANDSYNLIGHED.....	7
3.1 TOTALE SANDSYNLIGHED.....	8
3.2 BETINGET SANDSYNLIGHED.....	8
4 STATISTIK.....	8
4.1 HYPOTESER.....	9
4.2 PARRET VS. UPARRET.....	9
4.3 ESTIMERING AF MIDDELVÆRDI.....	9
4.4 ESTIMERING AF VARIANS.....	9
4.5 UPARRET T-TEST.....	10
4.6 KONFIDENSINTERVAL.....	10

```
clc
clear

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

1 STOKASTISKE VARIABLE

En diskret stokastisk variabel X har følgende tæthedsfunktion.

x	-3	0	2	4	7	10	12
$f_X(x)$	k	k	k	k	k	k	k

1.1 GYLDIG TÆTHEDSFUNKTION

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

$$\sum_{i=1}^n f_x(x_i) = \sum_{i=1}^n P(X = x_i) = 1$$

```
syms k
```

```
fx = k + k + k + k + k + k + k == 1;
```

```
solve(fx, k)
```

```
ans =
```

```
 $\frac{1}{7}$ 
```

For værdien af k lig med $\frac{1}{7}$ er $f_X(x)$ en gyldig tæthedsfunktion.

1.2 SKITSERING AF TÆTHEDSFUNKTION

Følgende er en skitsering af tæthedsfunktionen lavet i matlab under antagelse af at k har værdien $\frac{1}{7}$.

```
k = 1/7;
```

```
X = [-3, 0, 2, 4, 7, 10, 12];
```

```
P = [k, k, k, k, k, k, k];
```

```
figure(1)
```

```
stem(X, P, 'o', 'Color', Color.DeepCleret);
```

```
grid on
```

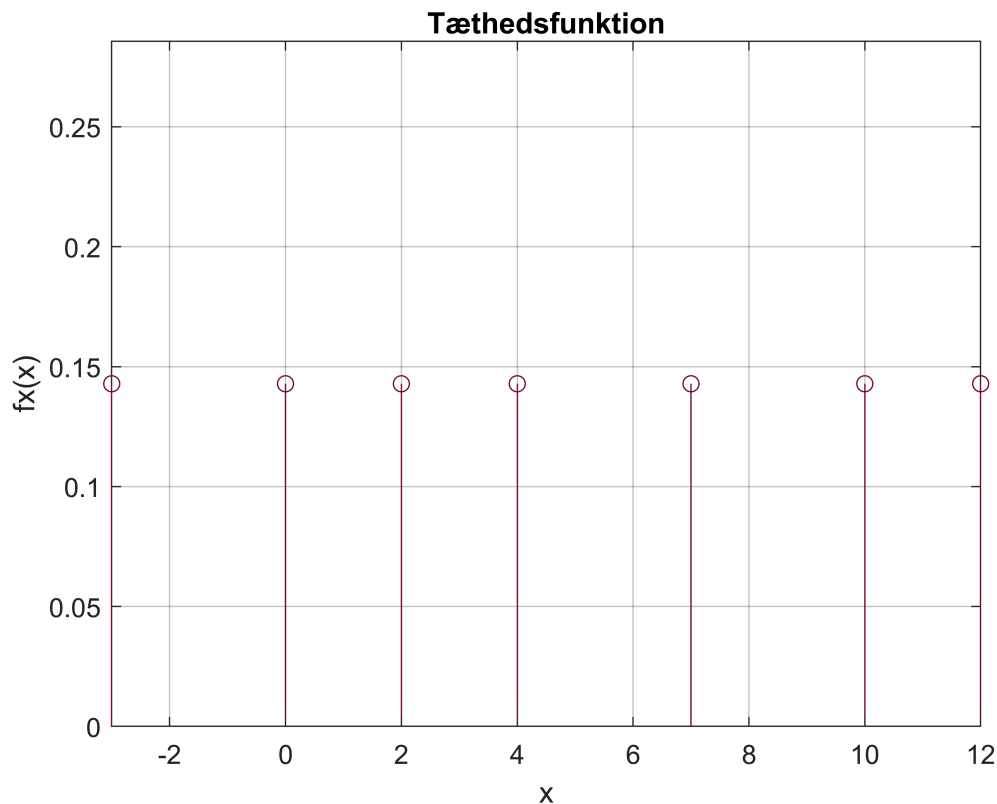
```
title('Tæthedsfunktion')
```

```
ylabel('fx(x)')
```

```
xlabel('x')
```

```
xlim([-3 12])
```

```
ylim([0 k*2])
```



1.3 FORVENTNINGSVÆRDI OG VARIANS

For at bestemme forventningsværdien og variansen ud fra $F_X(x)$ anvendes at der for forventningsværdien gælder, at

$$\bar{X} = E[X] = \sum_{i=1}^n x_i \cdot f_X(x_i)$$

```
EstimationX = 0;

for n = 1:length(P)
    EstimationX = EstimationX + (X(n) * P(n));
end

EstimationX
```

EstimationX = 4.5714

For variansen gælder der, at

$$\text{Var}(X) = \sigma_x^2 = E[X^2] - E[X]^2$$

```
EstimationXSquared = 0;

for n = 1:length(P)
    EstimationXSquared = EstimationXSquared + (X(n)^2 * P(n));
end
```

$$\text{VarianceX} = \text{EstimationXSquared} - \text{EstimationX}^2$$

$$\text{VarianceX} = 25.1020$$

Variansen er dermed bestemt til at være 25.1.

1.4 SANDSYNLIGHEDER

For at bestemme $P(x \geq 2)$ samt $P(x > 2)$ anvendes, at arealet for en tæthedsfunktion er lig med 1, hvormed

$$\sum_{i=1}^n f_X(x_i) = \sum_{i=1}^n P(X = x_i) = 1$$

Samtidig antages, at $k = \frac{1}{7}$.

$$f_x = @ (x) \ k;$$

$$\text{ProbabilityEqualOrAbove} = \text{sym}(f_x(2) + f_x(4) + f_x(7) + f_x(10) + f_x(12))$$

$$\text{ProbabilityEqualOrAbove} =$$

$$\frac{5}{7}$$

Dermed er $P(x \geq 2)$ bestemt til at være $\frac{5}{7}$.

$$f_x = @ (x) \ k;$$

$$\text{ProbabilityAbove} = \text{sym}(f_x(4) + f_x(7) + f_x(10) + f_x(12))$$

$$\text{ProbabilityAbove} =$$

$$\frac{4}{7}$$

Dermed er $P(x > 2)$ bestemt til at være $\frac{4}{7}$.

1.5 FORDELINGSFUNKTION

For at bestemme fordelingsfunktionen for X anvendes, at

$$F_x(x) = \sum_{x_i \leq x}^n f_x(x_i) = P(X \leq x)$$

Hvor det for eksempler af mellemregninger giver, at

$$F_x(-3) = \sum_{x_i \leq -3}^n f_x(-3) = f_x(-3) = \frac{1}{7}$$

$$F_x(0) = \sum_{x_i \leq 0}^n f_x(0) = f_x(-3) + f_x(0) = \frac{2}{7}$$

$$F_x(2) = \sum_{x_i \leq 2}^n f_x(2) = f_x(-3) + f_x(0) + f_x(2) = \frac{3}{7} \dots$$

Hvor k antages at være $\frac{1}{7}$, hvilket dermed giver, at

$$F_X(x) = \left\{ \begin{array}{ll} 0 & x < -3 \\ k & -3 \leq x < 0 \\ 2k & 0 \leq x < 2 \\ 3k & 2 \leq x < 4 \\ 4k & 4 \leq x < 7 \\ 5k & 7 \leq x < 10 \\ 6k & 10 \leq x < 12 \\ 7k & x \geq 12 \end{array} \right\}$$

2 STOKASTISKE PROCESSER

En kontinuer stokastisk process er givet ved

$$X(t) = w(t)$$

Hvor $w(t)$ er i.i.d (uafhængig og ens fordelt) og normalfordelt efter $w(t) \sim N(t, 1)$.

2.1 REALISERING

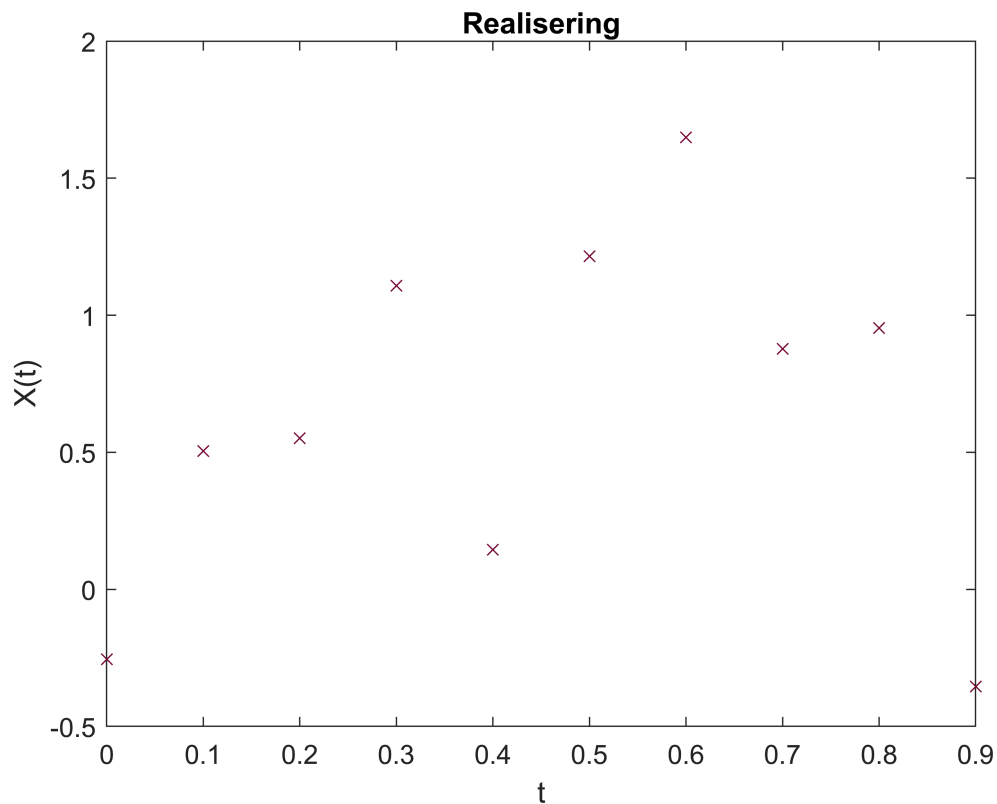
Realiseringen laves ved brug af matlab.

```
t = 0:0.1:0.9;

mu = t;
sigma = sqrt(1);

Realisation = randn(1, length(t)) + t;

figure(2)
plot(t, Realisation(1, :), 'x', 'Color', Color.DeepCleret);
title('Realisering')
ylabel('X(t)')
xlabel('t')
hold off
```



2.2 ENSEMBLE

For at bestemme middelværdien for ensemble anvendes, at

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

```
Ensemble.MeanValue = t;
```

Variansen kan bestemmes som

$$\text{Var}(X(t)) = \text{Var}[w(t)] = 1$$

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

```
Ensemble
```

```
Ensemble = struct with fields:
```

```
MeanValue: [0 0.1000 0.2000 0.3000 0.4000 0.5000 0.6000 0.7000 0.8000 0.9000]
```

```
Variance: 1
```

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

2.3 TEMPORAL MEAN

Den tidslige middelværdi kan bestemmes som

$$\mu_{X(t)}(t) = E[X(t)] = \int_{-\infty}^{\infty} x(t) f_{X(t)}(x(t)) dx(t)$$

Hvor den forventede tidslige middelværdi i en intervallet kan bestemmes som gennemsnittet i intervallet bestem som følgende.

$$\frac{1}{n} \cdot \int_0^{100} E[X(t)] dt$$

```
t = 0:1:100;
```

```
TemporalMeanEstimate = sum(t) / length(t)
```

```
TemporalMeanEstimate = 50
```

2.4 PROCESSEN

Da middelværiden er afhængig af tiden er processen ikke WSS. Da processen ikke er WSS kan processen heller ikke være ergodisk.

2.5 AUTOKORRELATION

Ligningen til at bestemme autokorrelationen er som følgende

$$R_{x(t_1)x(t_2)}(t_1 = 1, t_2 = 2) = \int_{x(t_1)} \int_{x(t_2)} x(t_1)x(t_2) \cdot f_{X(t_1),X(t_2)}(x(t_1), x(t_2)) dx(t_1)dx(t_2)$$

Da $x(t_1)$ og $x(t_2)$ er uafhængige gælder, at

$$R_{x(t_1)x(t_2)}(t_1 = 1, t_2 = 2) = E[x(t_1)] \cdot E[x(t_2)]$$

```
Rx = @(x, y) t(x + 1) * t(y + 1);
```

```
AutoCorrelationValue = Rx(1, 2)
```

```
AutoCorrelationValue = 2
```

Værdien for autokorrelation er dermed bestemt til at være 2.

3 SANDSYNLIGHED

Et studie viser at hvis et barn på 14 er flyttet mere end én gang på et år, vil barnet med en sandsynlighed på 0,06 begå alvorlig kriminalitet indenfor de næste 10 år. For børn, der flyttede én eller færre gange på et år, var sandsynligheden 0,03. 31% af børnene i studiet tilhørte gruppen, der var flyttet mere end en gang.

A: FLYTTET MERE END EN GANG

B: BEGÅR KRIMINALITET

$$P(B|A) = 0.06$$

$$P(B|\bar{A}) = 0.03$$

$$P(A) = 0.31$$

3.1 TOTALE SANDSYNLIGHED

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

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

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

```

BGivenA = 0.06;
BGivenNotA = 0.03;
A = 0.31;

```

```

NotA = 1 - A;

```

```

B = BGivenA * A + BGivenNotA * NotA

```

```

B = 0.0393

```

Sandsynligheden for hændelse B er dermed bestemt til at være 0.0393 altså 3.93%.

3.2 BETINGET SANDSYNLIGHED

For at bestemme sandsynligheden for om et barn tilhørte gruppen der havde flyttet mere end en gang, hvis det havde begået kriminalitet anvendes, at

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

```

AGivenB = (BGivenA * A) / B

```

```

AGivenB = 0.4733

```

Sandsynligheden er dermed bestemt til at være 0.4733 altså 47.33%.

4 STATISTIK

I et studie af tandhvalers forventede levetid, registrerede man dødsalderen på individuelle tandhvaler. Der blev i studiet registreret 10 hvaler, der var døde i fangenskab, og 10 hvaler, der var døde i det fri.

Død i det fri (alder i år)	Død i Fangenskab (alder i år)
50	7
43	2
11	1
35	3
7	15
62	6
70	14
67	1
25	5
1	9

4.1 HYPOTESER

Der opstilles en hypotese test for at bestemme om middelværdien af de to grupper er den samme.

$$H_0 : \mu_1 = \mu_2$$

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

4.2 PARRET VS. UPARRET

Da data ikke opfylder samme betingelser, altså eventuelle forhold vil det være optimalt at anvende en uparret test. Dette er vurderet på baggrund af dyr i det fri og i fangenskab som ikke har samme vilkår.

4.3 ESTIMERING AF MIDDELVÆRDI

Den estimerede middelværdi for de to sample set kan bestemmes som følgende.

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i$$

```
Free.Data = [50, 43, 11, 35, 7, 62, 70, 67, 25, 1];
Captivity.Data = [7, 2, 1, 3, 15, 6, 14, 1, 5, 9];
```

```
Free.EstimatedMean = mean(Free.Data)
```

```
Free = struct with fields:
    Data: [50 43 11 35 7 62 70 67 25 1]
    EstimatedMean: 37.1000
```

```
Captivity.EstimatedMean = mean(Captivity.Data)
```

```
Captivity = struct with fields:
    Data: [7 2 1 3 15 6 14 1 5 9]
    EstimatedMean: 6.3000
```

De to estimerede middelværdier er dermed bestemt til at være 37.1 samt 6.3.

4.4 ESTIMERING AF VARIANS

Den estimerede varians kan bestemmes ud fra de to varianse fra de to sample set givet ved

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{\mu})^2$$

```
Free.Variance = (1/(length(Free.Data) - 1)) * ...  
  sum((Free.Data - Free.EstimatedMean).^2);  
  
Captivity.Variance = (1/(length(Captivity.Data) - 1)) * ...  
  sum((Captivity.Data - Captivity.EstimatedMean).^2);
```

Den estimerede varians bestemmes nu som følgende.

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

```
Lead = (1 / (length(Free.Data) + length(Captivity.Data) - 2));  
  
Back = (length(Free.Data) - 1) * Free.Variance + ...  
  (length(Captivity.Data) - 1) * Captivity.Variance;  
  
PooledVariance = Lead * Back
```

```
PooledVariance = 337.1667
```

Dermed er den estimerede varians bestemt til at være 337.1667.

4.5 UPARRET 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 = (Free.EstimatedMean - Captivity.EstimatedMean) / (sqrt(PooledVariance) * ...  
  sqrt(1/length(Free.Data) + 1/length(Captivity.Data)));  
  
p = 2 * (1 - tcdf(abs(t), length(Free.Data) + length(Captivity.Data) - 2))
```

```
p = 0.0015
```

NULL hypotesen kan afvises da $p < 0.05$.

4.6 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, length(Free.Data) + length(Captivity.Data) - 2);  
Step = TZero * sqrt(PooledVariance) * sqrt(1/length(Free.Data) + 1/length(Captivity.Data));  
Interval.Min = (Free.EstimatedMean - Captivity.EstimatedMean) - Step;  
Interval.Max = (Free.EstimatedMean - Captivity.EstimatedMean) + Step;  
  
Interval
```

```
Interval = struct with fields:  
  Min: 13.5477  
  Max: 48.0523
```

Konfidensintervallet for δ er dermed bestemt til at være [13.5477; 48.0523].