Opgave 1: Stokastiske Variable

En diskret stokastisk variabel X har følgende tæthedsfunktion (pmf):

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

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

```
clearvars
```

For at det er en gyldig tæthedsfunktion er følgende givet:

$$\sum_x fx(x) = 1$$

```
syms k
x = [-3 0 2 4 7 10 12];
k = solve(k*length(x)== 1 , k)
```

k =

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

2) Skitsér tæthedsfunktionen.

```
fx = [k k k k k k k];
plot(x,fx)

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

3) Bestem forventningsværdien og variansen af X udfra $f_X(x)$. Angiv hvilke formler, der bruges til at finde værdierne. Antag at $k = \frac{1}{7}$.

Ved forventningsværien gælder følgende:

$$\boldsymbol{E}|\boldsymbol{x}| = \sum_{x} x * fx(x)$$

```
Ex= sum(x.*fx)
```

Ex =

$$\frac{32}{7}$$

Variance er givet ved

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

```
Ex2= sum(x.^2.*fx)
```

Ex2 = 46

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

var =

$$\frac{1230}{49}$$

4) Beregn sandsynlighederne $\Pr(x \geq 2)$ og $\Pr(x > 2)$. Antag at $k = \frac{1}{7}$.

```
Pr1 = sum(fx(x>=2))
```

Pr1 =

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

```
Pr2 = sum(fx(x>2))
```

Pr2 =

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

5) Bestem fordelingsfunktionen (cdf) for X. Angiv desuden mellemregninger. Antag at $k = \frac{1}{7}$.

$$Fx(x) = \int_{nedre\ værdi}^{x} x\, dx + tidl.\ led$$

```
syms x
Fx(x) = piecewise( x < -3, 0, ...
              -3 <= x < 0, 1/7, ...
              0 <= x < 2, 2/7, ...
              2 <= x < 4, 3/7, ...
              4 <= x < 7, 4/7, ...
              7 <= x < 10, 5/7, ...
              10 <= x < 12, 6/7, ...
              12 <= x, 7/7)
```

Fx(x) =

$$\begin{cases} 0 & \text{if } x < -3 \\ \frac{1}{7} & \text{if } x \in [-3, 0) \\ \frac{2}{7} & \text{if } x \in [0, 2) \\ \frac{3}{7} & \text{if } x \in [2, 4) \\ \frac{4}{7} & \text{if } x \in [4, 7) \\ \frac{5}{7} & \text{if } x \in [7, 10) \\ \frac{6}{7} & \text{if } x \in [10, 12) \\ 1 & \text{if } 12 \le x \end{cases}$$

Opgave 2: Stokastiske Processer

En kontinuer stokastisk process er givet ved:

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

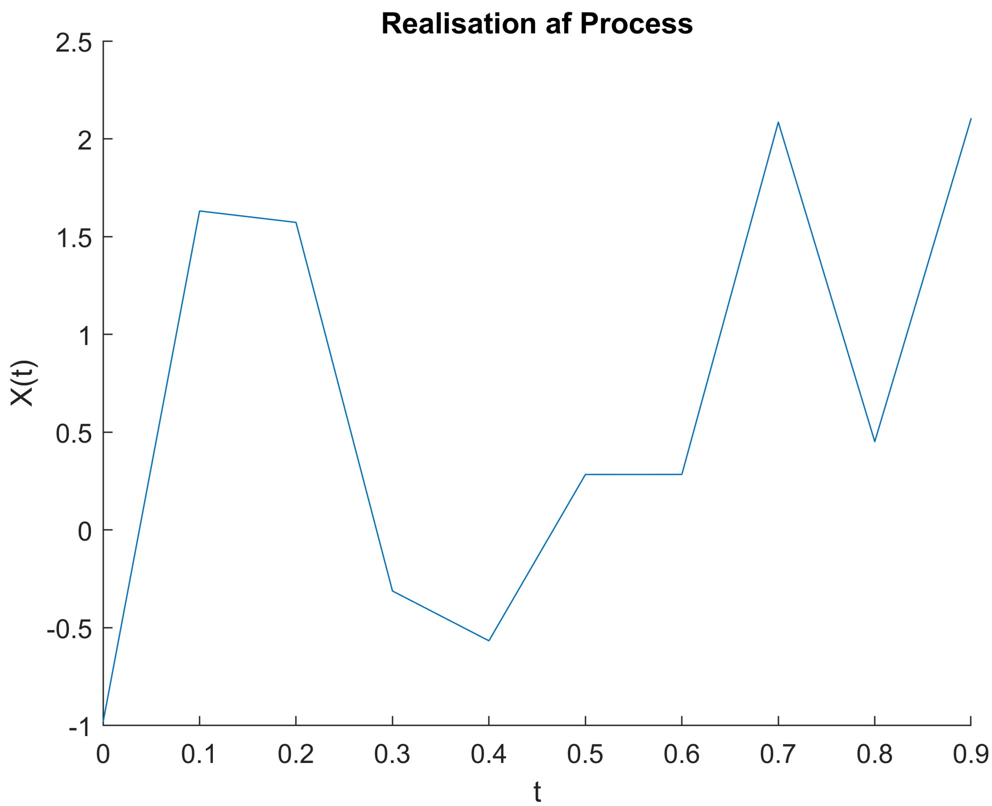
Hvor $w(t)$ er i.i.d. (uafhængig og ens fordelt) og normalfordelt efter $w(t) \sim N(t,1)$.

1) Skitsér én realisation af processen $X(t)$, hvor den er samplet til tiderne: $t = [0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]$. Brug en Gauss-generator, f.eks. *randn*() i matlab. Angiv desuden hvordan realisationen er fremkommet.

```
close all
t = 0:0.1:0.9;
mean = t; variance = 1; sigma = sqrt(variance);
w = normrnd(mean, sigma, [1 length(t)])
```

```
w = 1×10
   -0.9819    1.6311    1.5727   -0.3121   -0.5667    0.2836    0.2839    2.0853 ···
```

```
figure(1)
hold on;
xlabel("t"); ylabel("X(t)"); title("Realisation af Process");
plot(t, w)
```



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

Ensemble middelværdi ved w(t)~N(t,1) er t pga normalfordelingsfunktionen,

0 ved den givne stokastiske process

```
EnsembleMean = t + 0
```

```
EnsembleMean = 1×10
         0    0.1000    0.2000    0.3000    0.4000    0.5000    0.6000    0.7000 ⋯
```

ensemble variansen ved w(t)~N(t,1) er 1 pga normalfordelingsfunktionen

og +0 ved den givne stokastiske process

```
EnsembleVar = 1 + 0
```

```
EnsembleVar = 1
```

3) Hvad forventer du den tidslige middelværdi af en vilkårlig realisation af $X(t)$ i et tidsinterval $t = [0; 100]$ vil blive, begrund dit svar?

```
t = 0:1:100;
mean = t; variance = 1; sigma = sqrt(variance);
w = normrnd(mean, sigma, [1 length(t)])
```

```
w = 1×101
    1.2372    3.5197    1.6205    2.5247    5.6757    4.2616    4.1845    7.0867 ···
```

```
clear mean
mean(w)
```

```
ans = 50.1860
```

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

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

Da processen ikke er stationær, så kan den heller ikke være ergodisk.

5) Opstil ligningen til bestemmelse af Autocorrelationen $R_{X(t_1)X(t_2)}(t_1 = 1, t_2 = 2)$ og udregn værdien.

$$R_{XX}(t_1, t_2) = E[X(t_1)X(t_2)^*]$$

```
t1 = 1; t2=2;
E_T1 = 1 + 0;
E_T2= 2 + 0;
E_T1 * E_T2
```

```
ans = 2
```

Opgave 3: Sandsynlighedsregning

Et studie viser at hvis et barn på 14 er flyttet mere end én gang på et år, vil barnet med en sansynlighed 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.

1) Hvad er den totale sandsynlighed for at et af børnene i studiet begik alvorlig kriminalitet indenfor de næste 10 år?

Hændelse A : Barn er flyttet mere end en gang

Hændelse B : Begår kriminalitet

```
format
Pr_B_A = 0.06;
Pr_B_NOT_A = 0.03;
Pr_a = 0.31;
Pr_not_a = 1-Pr_a;
```

Total sandsynlighed er givet ved

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

```
Pr_b = Pr_B_A*Pr_a+Pr_B_NOT_A*Pr_not_a
```

```
Pr_b = 0.0393
```

2) Hvis et barn fra studiet har begået alvorlig kriminalitet indenfor de 10 år studiet rakte sig over, hvad er sandsynligheden for at barnet tilhørte gruppen, der havde flyttet mere end én gang?

Bayes Rule:

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

```
SandsynlighedTilhorteGruppe = (Pr_B_A*Pr_a)/Pr_b
```

```
SandsynlighedTilhorteGruppe = 0.4733
```

Opgave 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

```
fri = [50 43 11 35 7 62 70 67 25 1 ];
fange = [7 2 1 3 15 6 14 1 5 9];
```

1) Opstil en NULL og Alternativ hypotese, for at bestemme om middelværdien af de to grupper 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) Bør testen, der udføres, være parret eller uparret? Begrund dit svar.

Testen være udføres uparrat, da dataens subject (fangenskab eller fri) er to forskellige ting.

3) Estimér middelværdierne for begge grupper.

```
friMean = mean(fri)
```

```
friMean = 37.1000
```

```
fangeMean = mean(fange)
```

```
fangeMean = 6.3000
```

4) Estimér varianserne for begge grupper, samt den samlede varians (pooled variance).

Sample variance for 1: $s_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(x_{1i} - \bar{x}_1)^2$

Sample variance for 2: $s_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(x_{2i} - \bar{x}_2)^2$

```
S2_fange = (1/(length(fange)-1)) * sum((fange-fangeMean).^2)
```

```
S2_fange = 25.5667
```

```
S2_fri = (1/(length(fri)-1)) * sum((fri-friMean).^2)
```

```
S2_fri = 648.7667
```

Pooled variance

$$s^2 = \frac{1}{n_1 + n_2 - 2}\left((n_1 - 1)s_1^2 + (n_2 - 1)s_2^2\right)$$

```
PooledVar = ...
    (((length(fange)-1) * S2_fange) +...
    ((length(fri)-1) * S2_fri)) /...
    (length(fange)+length(fri)-2)
```

```
PooledVar = 337.1667
```

5) Anvend en uparret 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 = abs(friMean-fangeMean)/(sqrt(PooledVar)*sqrt((1/length(fri))+(1/length(fange))))
```

```
t = 3.7507
```

P-value

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

```
pval = 2*(1-tcdf(abs(t),(length(fri)+length(fange))-2))
```

```
pval = 0.0015
```

Da pval er mindre end 0.05 så afvises hypotesen.

6) Opstil og find 95% konfidens intervallet for forskellen i middelværdierne. 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:

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

```
n1 = length(fri)
```

```
n1 = 10
```

```
n2 = length(fange)
```

```
n2 = 10
```

```
middel = abs(friMean-fangeMean)
```

```
middel = 30.8000
```

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

```
t0 = 2.1009
```

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

```
koefMinus = 13.5477
```

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

```
koefPositiv = 48.0523
```