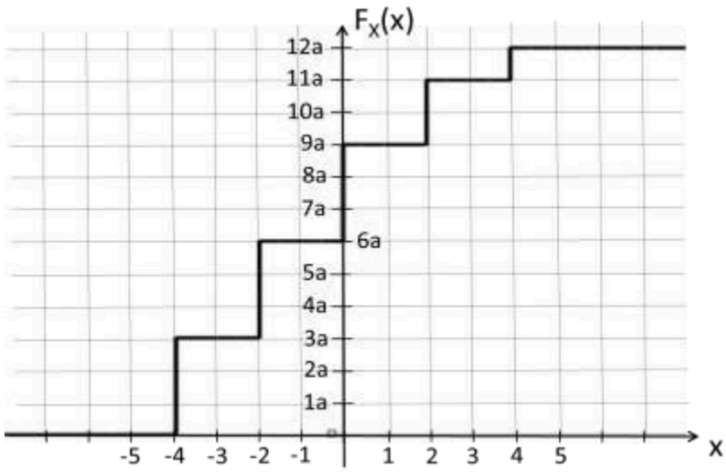
Forår2017Mat

```
clearvars
clc
close all
```

Opgave 1

En diskret stokastisk variabel X har følgende fordelingsfunktion (cdf) $F_X(x)$:



a) Bestem a, så $F_X(x)$ er en gyldig fordelingsfunktion.

```
clearvars
```

$$Fx(x) = 12a = 1 => a = \frac{1}{12}$$

b) Bestem og tegn tæthedsfunktionen (pmf) $f_X(x)$.

```
fx = [3/12 3/12 3/12 2/12 1/12];
x = [-4 -2 0 2 4];

figure(1);
bar(x,fx);

sum(fx)
```

```
ans = 1
```

c) Bestem middelværdien for X.

Ved forventningsværdien gælder følgende:

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

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

```
middel = -0.8333
```

d) Bestem variansen for X.

$$\boldsymbol{var}(\boldsymbol{x}) = \boldsymbol{E}\left|\boldsymbol{x}^2\right| - \boldsymbol{E}|\boldsymbol{x}|^2$$

```
var = sum(x.^2.*fx)-middel^2
```

```
var = 6.3056
```

Opgave 2

I juni måned (30 dage) regner det i gennemsnit 20% af dagene i den første halvdel af måneden og 30% af dagene i den sidste halvdel af måneden.

a) Hvor mange dage regner det i gennemsnit i juni måned?

hændelse a. Regner i starten af juni.

Hændelse b. Regner i slutnigen af juni.

```
Pr_R_Forst = 0.2;
Pr_R_Sidst = 0.3;
Pr_Forst = 0.5;
Pr_Sidst = 0.5;
Juni_Dage = 30;
total_prop = Pr_R_Forst*Pr_Forst + Pr_R_Sidst * Pr_Sidst
```

```
total_prop = 0.2500
```

```
gennemsnitRegn = 30*total_prop
```

```
gennemsnitRegn = 7.5000
```

b) Hvis vi oplever en dag med regn i juni måned, hvad er så sandsynligheden for at vi er i den sidste halvdel af måneden?

Bayes rule: $Pr(B|A) = \frac{Pr(A|B) \cdot Pr(B)}{Pr(A)}$

```
Pr_R_Sidst*Pr_Sidst/total_prop
```

```
ans = 0.6000
```

c) Hvad er sandsynligheden for at det regner **højst** 1 dag i den første halvdel af juni?

$$\boxed{Pr_n(k) = \frac{n!}{k!\,(n-k)!} p^k q^{n-k} = \binom{n}{k} p^k q^{n-k}}$$

```
Pr_IkkeRegn_forst = 1-Pr_R_Forst
```

```
Pr_IkkeRegn_forst = 0.8000
```

```
(15/1)*Pr_R_Forst^1*Pr_IkkeRegn_forst^14
```

```
ans = 0.1319
```

Opgave 3

En kontinuert stokastisk proces X(t) er givet ved:

$$X(t) = (-1)^n + W$$

hvor W er i.i.d. Gaussisk fordelte stokastiske variableW ~ $\mathcal{N}(0; 0{,}25)$, og n uafhængigt kan antage værdierne 0 og 1 med lige stor sandsynlighed.

a) Skitser 3 realisationer af processen X(t) i intervallet $0 \le t \le 5$. Angiv hvordan de 3 realisationer er opnået.

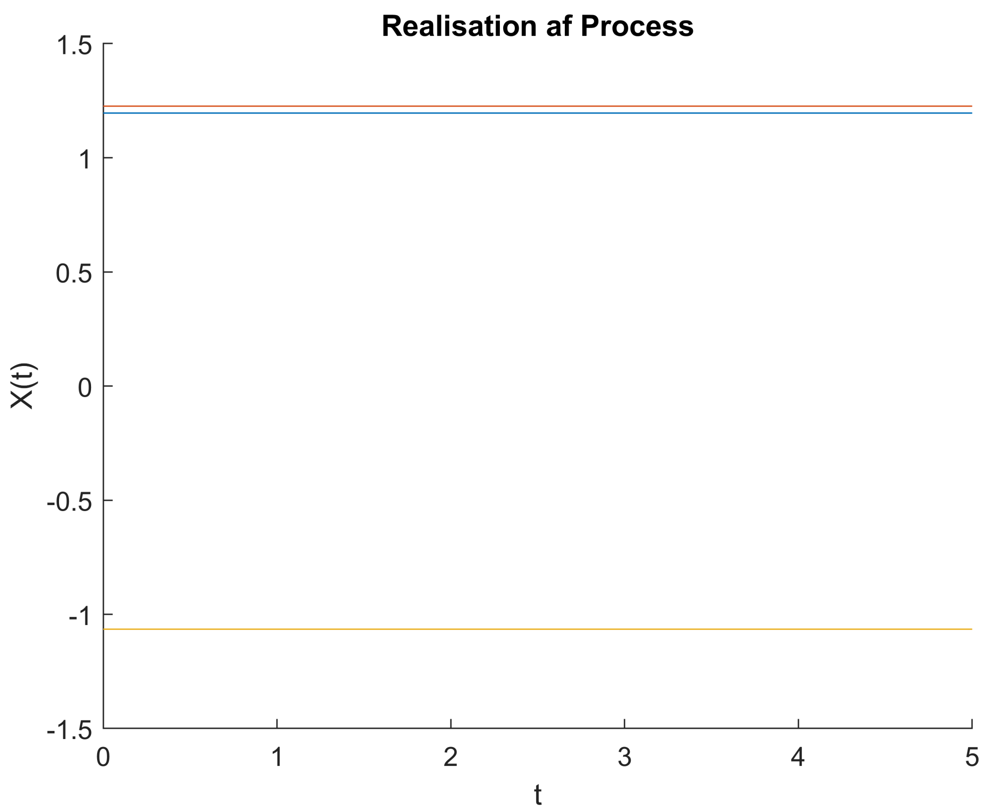
```
clearvars
close all
x = 0:1:5;
realisationer = 3;
% Først få værdierne til n
n = randi([0, 1], 1, 3)
```

```
n = 1×3
    0    0    1
```

```
% Ud fra W~N(0; 0,25) kan vi udlede:
mean = 0; variance = 0.25; sigma = sqrt(variance);
clear mean
% Få værdierne til w
w = randn(1, 3) * sigma;
figure(1)
hold on
xlabel("t"); ylabel("X(t)"); title("Realisation af Process");
```

```
for i = 1:realisationer
    y = (-1)^n(i) + w(i);
    plot(x, ones(1,length(x))*y);

end
```



b) Bestem middelværdien og variansen for én af realisationerne.

```
mean = mean(y)
```

```
mean = -1.0651
```

```
variane = var(y)
```

```
variane = 0
```

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

$E|X(t)| = E(-1^n) + E(w) = -1^0 * \Pr(0) + (-1) * \Pr(1) + E(w)$

```
middel = (-1)^0*(1/2)+(-1)^1*(1/2)+0
```

```
middel = 0
```

```
middel2 = ((-1)^0)^2*(1/2)+((-1)^1)^2*(1/2)+0
```

```
middel2 = 1
```

Varians er givet ved

$$var(x) = E|x^2| - E|x|^2 + Var(w)$$

```
variance = middel2-middel^2+0.25
```

```
variance = 1.2500
```

d) Angiv om processen er WSS (stationær i den brede forstand), og om den er ergodisk. Svarene skal begrundes.

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

Processen er ikke ergodisk da pmf ikke kan bestemmes udfra en realisation

Opgave 4

En kvalitetskontrol måler præcisionen af to forskellige typer gps'er. For begge typer blev målt afvigelsen mellem deres faktiske position ($d_{faktisk}$) og gps'ens angivelse (d_{gps}):

$$d_i = |d_{i,gps} - d_{i,faktisk}|$$

Det kan antages at afvigelserne er normalfordelte.
Der er testet 10 gps'er af type 1 og 12 gps'er af type 2.
For type 1 var middelafvigelsen $\widehat{\mu_1} = 5{,}21\ m$ med en estimeret varians $s_1^2 = 1{,}33\ m^2$.
For type 2 var middelafvigelsen $\widehat{\mu_2} = 4{,}18\ m$ med en estimeret varians $s_2^2 = 0{,}89\ m^2$.

a) Opstil en hypotese test for at bestemme om middelafvigelserne for de to typer gps'er er den samme.

Null hypothesis. Brugt som testen af ingen forskel.

$H_0 = \mu 1 = \mu 2$

Alternanive hypotese

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

b) Estimer forskellen i middelværdierne $\hat{\delta}$ for de to typer.

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(5.21-4.18)
```

```
middel = 1.0300
```

c) Estimer variansen $\hat{s}^2$ for forskellen mellem de to typer.

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

```
n1 = 10;
n2 = 12;
s1_2= 1.33;
s2_2 = 0.89;
standard2=1/(n1+n2-2)*((n1-1)*s1_2+(n2-1)*s2_2)
```

```
standard2 = 1.0880
```

d) Anvend en t-test til test af din hypotese. Kan NULL-hypotesen afvises med et signifikantniveau på 0,05? Svaret skal begrundes.

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

```
t = middel/(sqrt(standard2)*sqrt((1/n1)+(1/n2)))
```

```
t = 2.3062
```

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 = 0.0319
```

Da p er mindre end 0.05 så kan vi afvise hypotese 0.

e) Bestem 95% konfidens intervallet for forskellen i middelværdier δ.

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 = 2.0860
```

```
koefMinus = middel-t0*sqrt(standard2)*sqrt((1/n1)+(1/n2))
```

```
koefMinus = 0.0984
```

```
koefPositiv = middel+t0*sqrt(standard2)*sqrt((1/n1)+(1/n2))
```

```
koefPositiv = 1.9616
```