

Kontinuasjonseksamen i ISTx100y, august 2021

Oppgave 1 (15%, tallsvar)

A: tilfeldig valgt batteri er defekt $P(A) = 0.05$

a)

Vi trekker tilfeldig $n = 15$ batterier. Finn sannsynligheten for at akkurat $x = 3$ av dem er defekte.

De $n = 15$ tilfeldig valgte batteriene kan anses som en binomisk forsøksrekke der sannsynligheten for 'suksess' (defekt batteri) er $p = 0.05$ i hvert forsøk.

Dersom vi lar X telle antall suksesser i 15 forsøk, finner vi punktsannsynlighet i binomisk fordeling som

$$P(X = 3) = \binom{15}{3} 0.05^3 (1 - 0.05)^{12} = 0.03073298 \approx 0.03$$

Vi skal gi svaret avrundet til 2 desimaler.

Slik regner vi det i Python.

```
In [1]: from scipy import stats  
        round(stats.binom.pmf(3,15,0.05),2)
```

```
Out[1]: 0.03
```

b)

La B betegne hendelsen at hurtigtesten viser "defekt". Vi antar at $P(B|A) = 0.80$ og at $P(B'|A') = 0.85$.

Vi skal finne sannsynligheten for at hurtigtesten viser "ikke defekt" gitt at batteriet faktisk er defekt, det vil si $P(B'|A)$. Dette finner vi ved komplementregelen;

$$P(B'|A) = 1 - P(B|A) = 1 - 0.80 = 0.20$$

Det trenger vi ikke å regne i Python.

c)

Vi skal finne $P(B)$, og denne kan vi finne ved setningen om total sannsynlighet;

$$P(B|A)P(A) + P(B|A')P(A') = 0.80 \cdot 0.05 + (1 - 0.85) \cdot (1 - 0.05) = 0.1825 \approx 0.18$$

```
In [2]: 0.80*0.05 + (1-0.85)*(1-0.05)
```

```
Out[2]: 0.18250000000000002
```

Oppgave 2 (15 %, tallsvar)

La X være en stokastisk variabel med forventning $E(X) = 3$ og varians $\text{Var}(X) = 2$.

a)

Forventningsverdien til $2X$, finner vi ved å bruke at $E(aX) = aE(X)$.

$$E(2X) = 2 \cdot E(X) = 2 \cdot 3 = 6$$

b)

La Y være en stokastisk variabel, uavhengig av X , som har forventningsverdi $E(Y) = 3$ og varians $\text{Var}(Y) = 2$.

$$\text{Vi finner } E(2X + 3Y) = 2 \cdot E(X) + 3 \cdot E(Y) = 2 \cdot 3 + 3 \cdot 3 = 15$$

c)

$$\text{Vi finner } \text{Var}(2X + 3Y) = 2^2 \cdot \text{Var}(X) + 3^2 \cdot \text{Var}(Y) = 4 \cdot 2 + 9 \cdot 2 = 26$$

```
In [3]: print(2*3+3*3)
        print(2**2*2 + 3**2*2)
```

```
15
26
```

Oppgave 3 (25%)

Anders har nettopp kjøpt seg en brukt el-sykkel. Vi skal studere to typer feil som kan opptre på denne sykkelmodellen; feil med ladekabelen og feil med de elektroniske girene. Basert på kjennskap til lignende bilmodeller antar Anders at feil på ladekabelen kan modelleres som en poissonprosess med rate $\lambda_L = 0.5$ / år, og at feil med girene kan modelleres som en poissonprosess med rate $\lambda_G = 2$ / år. De to prosessene kan antas å være uavhengige.

a)

i. Hva er sannsynligheten for at ladekabelen må repareres minst én gang i løpet av ett år?

Vi ser på en poissonfordelt stokastiske variabel X_L med rate $\lambda_L = 0.5$ /år. Vi finner

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{(\lambda_L \cdot 1)^0}{0!} e^{-\lambda_L \cdot 1} = 1 - e^{-\lambda_L} = 1 - e^{-0.5} = 0.3934693 \approx 0.393$$

ii. Hva er sannsynligheten for at det går mer enn to år før det inntreffer en feil på girene?

Vi ser på en Poissonprosess med rate $\lambda_G = 2$ /år. Tiden T frem til første hendelse inntreffer er eksponensialfordelt, og dermed finner vi

$$P(T > 2) = 1 - P(T \leq 2) = 1 - (1 - e^{-\lambda_G \cdot 2}) = e^{-2 \cdot 2} = 0.01831564 \approx 0.018$$

Alternativt kan man bruke poissonfordelingen og regne ut sannsynligheten for ingen feil for to år. Da vil vi se på en rate på $2 \cdot 2 = 4$ pr 2 år

$$P(X_{L,2} = 0) = \frac{(\lambda_G \cdot 2)^0}{0!} e^{-\lambda_G \cdot 2} = e^{-2 \cdot 2} = 0.01831564 \approx 0.018$$

```
In [4]: import numpy as np
        np.exp(-4)
```

```
Out[4]: 0.01831563888873418
```

b)

Det koster 3000 kr for å reparere ladekabelen og 1000 kr for å reparere girene hver gang de feiler.

i. Hvor mye må Lars forvente å betale for reparasjoner av disse i løpet av to år?

Vi bruker samme formel som i Oppgave 2, og vi må vite at forventningsverdien i en poissonfordeling med rate λ i tidsperioden t er lik λt .

La $X_{L,2}$ være en poissonfordelt stokastisk variabel som teller antall feil på ladekabelen over en periode på to år, og la $X_{G,2}$ være en poissonfordelt stokastisk variabel som teller antall feil med girene over en periode på to år. Vi kan da regne ut

$$E(3000 \cdot X_{L,2} + 1000 \cdot X_{G,2}) = 3000 \cdot \lambda_L \cdot 2 + 1000 \lambda_G \cdot 2 = 7000$$

```
In [5]: 3000*0.5*2+1000*2*2
```

```
Out[5]: 7000.0
```

ii. Og hva blir standardavviket i disse reparasjonskostnadene?

Vi setter først opp ett uttrykk for variansen. Da bruker vi samme formel som i oppgave 2 med korrelasjon lik 0. Her må vi vite at variansen i en poissonprosess med rate λ i tidsperioden t er lik λt . Det er en viktig egenskap ved poissonfordelingen av variansen er lik forventningsverdien.

$$\text{Var}(3000 \cdot X_{L,2} + 1000 \cdot X_{G,2}) = 3000^2 \cdot \text{Var}(X_{L,2}) + 1000^2 \cdot \text{Var}(X_{G,2}) = 3000^2 \cdot \lambda_L \cdot 2 + 1000$$

Deretter kan vi regne ut standardavviket ved $\sqrt{3000^2 \cdot 0.5 \cdot 2 + 1000^2 \cdot 2 \cdot 2} = 3605.551 \approx 3605.6$

```
In [6]: np.sqrt(3000**2*0.5*2+1000**2*2*2)
```

```
Out[6]: 3605.5512754639894
```

c)

Lars bestemmer seg for at han vil skrote sykkelens dersom kostnadene for reparasjon av ladekabel og gir overskrider 10 000 kr de neste to årene. Bruk normaltilnærming for å regne ut sannsynligheten for at Lars skroter sykkelens etter to år.

Vi lar Y være en stokastisk variabel som modellerer reparasjonskostnader for ladekabel og gir over en periode på to år. Vi skal anta at Y er tilnærmet normalfordelt, og fra oppgave b) vet vi at $E(Y) = 7000$ og $SD(Y) = 3605.6$.

Vi skal regne ut

$$P(Y > 10000) = 1 - P(Y \leq 10000) = 1 - P\left(Z \leq \frac{10000 - 7000}{3605.6}\right) \approx 1 - \Phi(0.83) = 1 - 0.7967$$

Her har vi brukt tabellene til å slå opp $\Phi(0.83)$.

Det er også mulig å bruke Python direkte og da blir det slik:

```
In [7]: (10000-7000)/3605.6
```

```
Out[7]: 0.8320390503660972
```

```
In [8]: 1-stats.norm.cdf(10000, loc=7000, scale=3605.6)
```

```
Out[8]: 0.20269345144183482
```

Oppgave 4 (30%)

En elektronisk enhet, A, kan sende ut et binært signal (0 eller 1). En annen elektronisk enhet, B, skal motta signalet.

Dessverre er ikke overføringen feilfri. Dersom enhet A sender ut signalet a ($a = 0$, eller $a = 1$), vil det signalet som enhet B mottar kunne antas å være en normalfordelt størrelse Y med forventningsverdi a og varians $\sigma^2 = 0.36$.

a)

i. Anta at enhet A har sendt ut signalet $a = 1$. Hva blir sannsynligheten for at enhet B observer et signal som er mindre enn 0.6?

Dersom enhet A sender ut signalet $a = 1$ så er $Y \sim N(1, \sqrt{0.36})$ og

$$P(Y \leq 0.6) = P\left(Z \leq \frac{0.6-1}{\sqrt{0.36}}\right) \approx \Phi(-0.67) = 0.2514$$

```
In [9]: stats.norm.cdf(0.6,1,np.sqrt(0.36))
print((0.6-1)/np.sqrt(0.36))
stats.norm.cdf(-0.67,0,1)
-0.6666666666666667
```

```
Out[9]: 0.25142889509531013
```

ii. Anta at enhet A har sendt ut signalet $a = 0$. Hva blir sannsynligheten for at enhet B observer et signal i intervallet $[-0.6, 0.6]$?

Dersom enhet A sender ut signalet $a = 0$ så er $Y \sim N(0, \sqrt{0.36})$ og

$$P(-0.6 \leq Y \leq 0.6) = P\left(\frac{0.6-0}{\sqrt{0.36}} Z \leq \frac{-0.6-0}{\sqrt{0.36}}\right) = \Phi(1) - \Phi(-1) = 0.8413 - 0.1587 = 0.6826$$

Her kan man evt bemerke at vi er ute etter sannsynligheten for at en normalfordelt variabel tar en verdi pluss/minus ett standardavvik fra forventningsverdien, og det er et kjent resultat at dette skjer med sannsynlighet 0.68.

```
In [10]: stats.norm.cdf(0.6,0,np.sqrt(0.36))-stats.norm.cdf(-0.6,0,np.sqrt(0.36))
```

```
Out[10]: 0.6826894921370859
```

b)

Enhet A inneholder en sensor som monitorer fukt i et lager. Dersom luftfuktigheten er lav sender enheten med jevne mellomrom ut signalet $a = 0$.

Dersom fuktigheten i lageret overskrider en terskelverdi sender enhet A ut signalet $a = 1$. Basert på de tre nyeste (uavhengige) observasjonene gjort med enhet B, skal vi vurdere om vi skal gjennomføre planlagt kontroll ved lageret, eller om vi vil utsette kontrollen. Vi ønsker dermed å teste

$$H_0 : a = 1 \text{ mot } H_1 : a = 0$$

ved signifikansnivå $\alpha = 0.05$. Dersom observasjonene gir grunnlag for å forkaste H_0 kan vi utsette kontrollen.

i.

Skriv ned formelen for en egnet testobservator, og forklar hvilke uttrykk som inngår i formelen.

Vi skal se på målingene Y_1, Y_2, Y_3 der $Y_i \sim N(a, \sqrt{0.36})$ for $i = 1, 2, 3$ og der Y -ene er uavhengige stokastiske variabler med kjent varians. Som testobservator kan vi bruke

$$Z = \frac{\bar{Y} - 1}{\sqrt{0.36/3}} = \frac{\bar{Y} - 1}{\sqrt{0.12}}$$

Vi ser altså på gjennomsnittet av målingene ($\bar{Y}$) og trekker fra forventningsverdien til gjennomsnittet under H_0 ($a = 1$), og deler på standardavviket til gjennomsnittet ($SD(\bar{Y}) = \sqrt{0.36/3} = \sqrt{0.12}$).

Alternativt kan også $\bar{Y}$ brukes og når nullhypotesen er riktig er den normalfordelt med forventningsverdi 1 og standardavvik $SD(\bar{Y}) = \sqrt{0.36/3} = \sqrt{0.12}$, eller bare summen av de tre målingene.

ii. Hva er testobservatorens fordeling dersom H_0 er sann? Skissér testens forkastningsområde.

Dersom nullhypotesen er sann så er testobservatoren Z standard normalfordelt. Dersom vi observerer en lav verdi av Z vil dette peke mot alternativhypotesen (venstresidig test). Ved signifikansnivå $\alpha = 0.05$ forkaster vi H_0 dersom vi observerer $z \leq -z_{0.05} = -1.645$

```
In [11]: import matplotlib.pyplot as plt

ZunderH0=stats.norm(loc=0,scale=1)

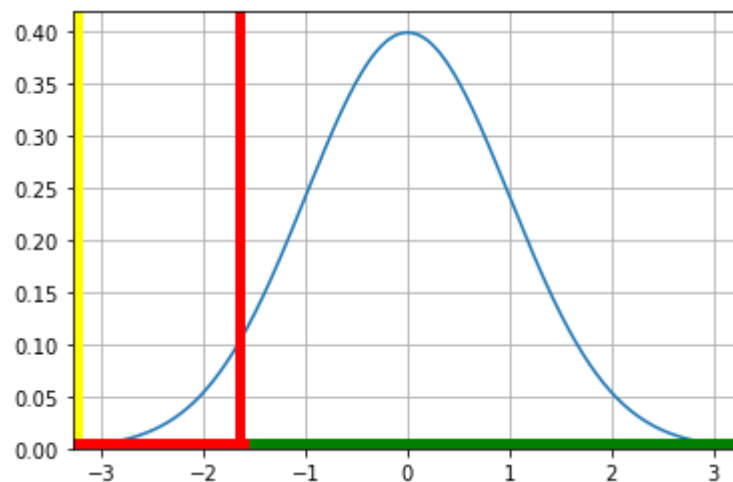
X=ZunderH0
xmin, xmax = X.interval(0.999)
if xmin > 0 and xmin < 0.1: xmin = 0
xverdier = np.linspace(xmin, xmax, 1001)

fig, ax = plt.subplots()
ax.plot(xverdier, X.pdf(xverdier))
ax.set_xlim(xmin, xmax)
ax.set_ylim(bottom=0)
ax.grid()

plt.axvline(xmin, 0, 1, lw=10, label='forkastningsgrense',c="yellow")

plt.axvline(-1.645, 0, 1, lw=5, label='forkastningsgrense',c="red")
plt.axhline(0, (-1.645-xmin)/(xmax-xmin), 1, lw=10, label='ikke fo
rkast innen området',c="green")
plt.axhline(0, 0, (-1.645-xmin)/(xmax-xmin), lw=10, label='forkast
innen området',c="red")
```

Out[11]: <matplotlib.lines.Line2D at 0x7fdd0ccceca0>



iii. Hva blir teststyrken?

Teststyrker en sannsynligheten for å forkaste H_0 dersom H_1 er sann. Vi må dermed se på fordelingen til testobservatoren Z dersom H_1 er sann.

Når H_1 er sann så er $\bar{Y} \sim N(0, 0.12)$, slik at $Z = \frac{\bar{Y}}{\sqrt{0.12}} - \frac{1}{\sqrt{0.12}} \sim N(-1/\sqrt{0.12}, 1)$.

Da er $P(Z \leq -1.645) = P\left(\frac{Z - (-1/\sqrt{0.12})}{1} \leq \frac{-1.645 - (-1/\sqrt{0.12})}{1}\right) \approx \Phi(1.24) = 0.8925$

```
In [12]: X=ZunderH0
xmin, xmax = X.interval(0.999)
if xmin > 0 and xmin < 0.1: xmin = 0
xverdier = np.linspace(xmin, xmax, 1001)

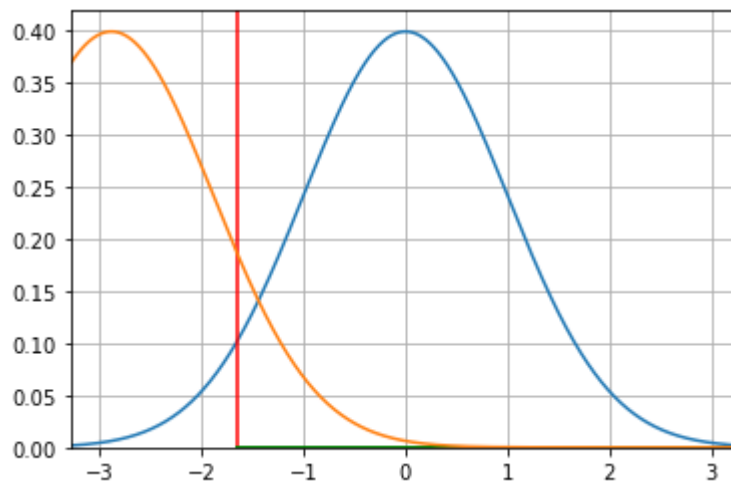
fig, ax = plt.subplots()
ax.plot(xverdier, X.pdf(xverdier))
ax.set_xlim(xmin, xmax)
ax.set_ylim(bottom=0)
ax.grid()

plt.axvline(-1.645, 0, 1, label='forkastningsgrense',c="red")
plt.axhline(0, (-1.645-xmin)/(xmax-xmin), 1, lw=2,label='ikke for
kast innen området',c="green")

ZunderH1=stats.norm(loc=-1/np.sqrt(0.12),scale=1)

XX=ZunderH1
ax.plot(xverdier, XX.pdf(xverdier))
print(XX.cdf(-1.645))
```

0.8928358404822117



iv. Vi har gjort følgende tre observasjoner med enhet B:

$$y_1 = -0.31, y_2 = 1.08, y_3 = 0.85.$$

Hva blir utfallet av hypotesetesten?

Vi observerer $\bar{y} = 0.54$ og $z = -1.327906$. Siden z ikke ligger innenfor forkastningsområdet har vi ikke grunnlag for å forkaste nullhypotesen. Vi bør dermed ikke avlyse den planlagte kontrollen ved lageret.

```
In [13]: Yvec=np.array([-0.31,1.08,0.85])
print(Yvec)
Yvec.mean()
z=(Yvec.mean()-1)/np.sqrt(0.12)
print(z)
```

[-0.31 1.08 0.85]
-1.3279056191361391

Oppgave 5 (15%)

Et datasett består av 25 observasjoner av en responsvariabel Y og en kovariat x . Et kryssplott av observasjonene er vist.

En lineær regresjonsmodell $Y = \beta_0 + \beta_1 x + \varepsilon$ har blitt tilpasset i Python. Deler av utskriften (`modell.summary()`) er vist. Bruk utskriften til å svare på oppgave a, b og c.

a)

Den estimerte regresjonslinjen kan leses fra kolonnen `coef` og $\hat{\beta}_0$ er gitt i raden `Intercept` og $\hat{\beta}_1$ i raden som heter `x`.

Den estimerte regresjonslinja er $\hat{y} = 2.5897 + 0.5104x$.

b)

Hvilken hypotesetest utføres i utskriften i linjen som starter med "x" ?

Hypotesetesten som utføres er $H_0 : \beta_1 = 0$ mot $H_1 : \beta_1 \neq 0$.

c)

Regn ut et 90% konfidensintervall for stigningstallet β_1 .

Konfidensintervall for β_1 finner vi ved

$$[\hat{\beta}_1 - t_{\alpha/2, n-2} \cdot \text{SE}(\hat{\beta}_1), \hat{\beta}_1 + t_{\alpha/2, n-2} \cdot \text{SE}(\hat{\beta}_1)]$$

I dette tilfellet er $n - 2 = 23$ og $\alpha/2 = 0.05$ slik at $t_{\alpha/2, n-2} = 1.714$. Fra utskriften vet vi at $\hat{\beta}_1 = 0.5104$ og $\text{SE}(\hat{\beta}_1) = 0.1340$.

Dermed kan vi regne ut konfidensintervallet $[0.2807, 0.7401]$.

```
In [14]: beta1hat=0.5104
         sebeta1hat=0.1340
         nm2=23
         alpha=0.1
         alphahalve=alpha/2
         talphahalve=stats.t.ppf(1-alphahalve,nm2) # sannsynlighet alpha/2
         til høyre har 1-alpha/2 til venstre
         print(talphahalve)

         add=talphahalve*sebeta1hat
         print(add)
         nedre=beta1hat-add
         ovre=beta1hat+add
         print(ovre)
         print(nedre)

1.7138715277470473
0.22965878471810436
0.7400587847181044
0.2807412152818956
```

In []: