Oppgåve 1 Pinnekjøtt

a) La X vere Poissonfordelt med punktsannsyn $P(X = x) = \frac{\mu^x}{x!}e^{-\mu}$, $x = 0, 1, \ldots$, og anta parameter $\mu = 3$.

Rekn ut sannsynet for at X er mindre eller lik 3.

Rekn ut $P(X \leq 3 | X > 0)$.

b) Ei null-trunkert Poissonfordeling med parameter μ er karakterisert med punktsannsyn $f(y) = P(X = y|X > 0)$, der X er Poissonfordelt med parameter μ.

Vis at punktsannsynet til Y er

$$f(y) = \frac{\mu^y}{y!}\frac{e^{-\mu}}{1 - e^{-\mu}}, \qquad y = 1, 2, \ldots$$

Finn forventinga $E(Y)$ og rekn ut for $\mu = 3$.

Karin skal ha gjestar til pinnekjøttmiddag i jula. Ho antar at antalet kjøttstykker som kvar person et er null-trunkert Poissonfordelt med parameter μ og uavhengig for ulike personar. Karin bruker data frå julebordet i pensjonistlaget til å estimere parameteren μ.

c) Under julebordet observerer ho antal kjøttstykker; $y_1, \ldots, y_{21}$, som dei 21 gjestane et.

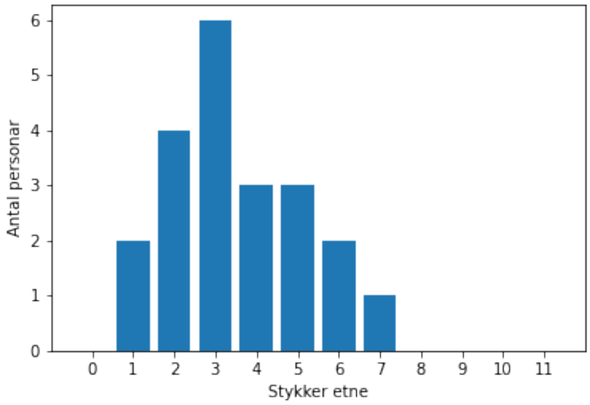
Vis at log-rimelighetsfunksjonen (log likelihood function) for μ vert

$$l(\mu) = \ln(\mu) \sum_{i=1}^{21} y_i - \sum_{i=1}^{21} \ln y_i! - 21\mu - 21 \ln(1 - e^{-\mu})$$

Bruk første-orden Taylorutvikling: $\ln(1 - e^{-\mu}) \approx \ln(1 - e^{-3}) + \frac{e^{-3}}{1-e^{-3}}(\mu - 3)$ til å finne ein approksimasjon til sannsynsmaksimeringsestimatoren $\hat{\mu}$.

Gjestane et tilsaman 74 stykjer. Rekn ut estimatet frå oppgjeve tal.

Figur 1 viser histogrammet av antal stykker julebordsdeltakarane et. Bruk punktsannsynet for null-trunkert Poissonfordeling til å diskutere om modellen stemmer høveleg med histogrammet si oppsummering av data for 21 personar. (Du kan bruke tabell for Poissonfordeling som del av utrekninga, med parameter nærmast den du fann ved sannsynlighetsestimering.)



Figur 1: Datainnsamling av etne stykker i forbindelse med julebord. Det er 21 julebordsdeltakarar.

Oppgåve 2 Lyspærer

a) La ein stokastisk variabel X vere eksponensialfordelt slik at sannsynlighetstettleiken er $f(x) = \frac{1}{\beta}\exp(-x/\beta)$, $x > 0$. Vi set her $\beta = 1$.

Rekn ut $P(X < 3)$.

Finn teoretisk median m til X, altså verdien av konstanten m slik at $P(X \leq m) = 0.5$.

b) Eit toalett har tre lyspærer. Man antar at levetidene til lyspærene, X_1, X_2 og X_3, er uavhengige og eksponensialfordelte med parameter $\beta = 1$ år. Det vil vere lys i rommet så lenge minst ei av lyspærene virker. Vi definerer Y som tida til toalettet vert heilt mørkt.

Korleis kan hendinga $Y \leq y$ uttrykkjast som ei hending som involverer X_1, X_2 og X_3?

Vis at $P(Y \leq y) = (1 - e^{-y})^3$, $y > 0$.

Rekn ut sannsynet til Y.

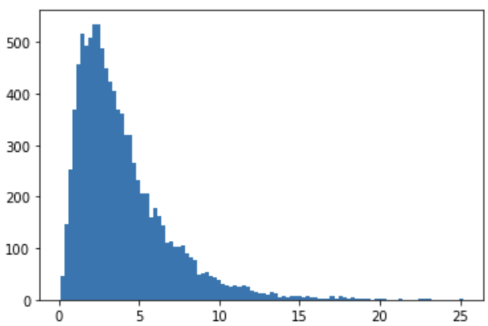
c) Anta no at dei tre lyspærene er av ulik type. Dei har eksponensialfordelte uavhengige levetider X_1 med parameter $\beta = 1$, X_2 med parameter $\beta = 2$ og X_3 med parameter $\beta = 3$.

Ein kan bruke Monte Carlo simulering til å forstå dei statistiske eigenskapane til denne situasjonen. Forklar korleis dette kan gjerast ved hjelp av Pythonkoden i Figur 2. Kva syner resultata og plott til høgre i figuren?

```python
import numpy as np

N = 10000
cnt1 = 0
cnt2 = 0
cnt3 = 0
y = np.zeros(N)
ex10 = 0
i = 0
while(i<N):
    x1 = np.random.exponential(1)
    x2 = np.random.exponential(2)
    x3 = np.random.exponential(3)
    if(x1 > x2 and x1 > x3):
        cnt1 = cnt1 + 1
    if(x2 > x1 and x2 > x3):
        cnt2 = cnt2 + 1
    if(x3 > x1 and x3 > x2):
        cnt3 = cnt3 + 1
    y[i] = np.max([x1,x2,x3])
    if(y[i] > 10):
        ex10 = ex10 + 1
    i = i + 1

print(cnt1/N)
print(cnt2/N)
print(cnt3/N)
print(np.mean(y))
print(np.median(y))
print(ex10/N)

import matplotlib.pyplot as plt
plt.hist(y,100)
```

Resultat (print):

```
0.1278
0.3393
0.5329
3.9291
3.2104
0.0405
```



Figur 2: Monte Carlo simulering av levetider for tre lyspærer. Pythonkode (venstre) og resultat av koden (høgre).

Oppgåve 3 Skigåing

a) X er normalfordelt med forventning $\mu = 10$ og standardavvik $\sigma = 0.5$.

Rekn ut $P(X < 9)$.

Kva er c gjeve at $P(X > c) = 0.99$?

Johannes er ein dedikert skiløpar. Ein konkurrent seier at han bruker 10 minutt opp ein bakke, under gjennomsnittleg dagsform og føreforhald. Johannes tenkjer at han er raskare enn konkurrenten sin i denne bakken. For å undersøkje dette, går han på ski opp bakken ti ulike dagar, under ulike føreforhald og avhengig av dagsform. La x_i = tid på dag $i = 1, \ldots, 10$. Tidene hans vert 10.1, 9.3, 9.75, 9.9, 10.15, 9.55, 9.8, 9.95, 9.45, 9.6. Det er oppgjeve at $\sum_{i=1}^{10} x_i = 97.55$ og at $\sum_{i=1}^{10} x_i^2 = 952.3$

b) Anta at tida det tek å gå opp bakken på ski er normalfordelt med forventning μ og ukjend varians σ^2. Vi antar vidare at tidene er uavhengige.

Formuler Johannes sin situasjon som ein hypotesetest.

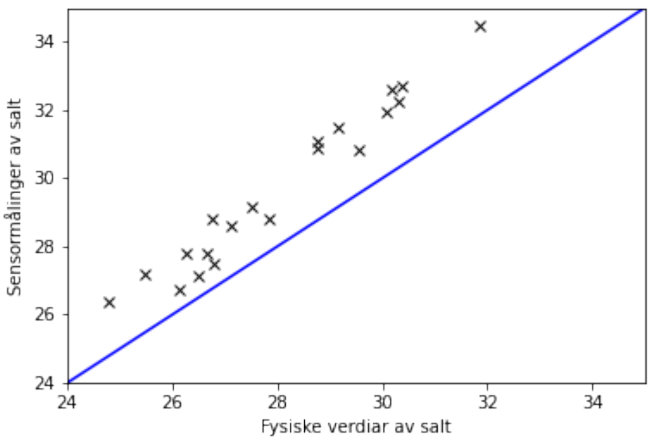
Gjennomfør hypotesetesten med signifikansnivå 0.05.

Det er vanskeleg å undersøkje om data faktisk er normalfordelt her. Gjennomfør i staden ein fortegnstest for hypotesen, med signifikansnivå 0.05.

Oppgåve 4 Saltmålinger

Ein fysisk modell vert brukt for å forstå variabilitet i saltinnhald i Trondheimsfjorden. I tillegg kan ein måla saltinnhaldet med ein sensor på utvalde stader. Slike målinger vert mellom anna brukt til å kalibrere den fysiske modellen. Vi lar x_i og y_i vere resultat av fysisk modell og sensormåling. Her indikererer $i = 1, \ldots, n$ stader i fjorden der ein på gjeve tidspunkt har tilgjengeleg begge variablane.

Ein foreslår regresjonsmodellen $Y_i = \alpha + \beta x_i + \epsilon_i$, $i = 1, \ldots, n$, der det vert antatt at $\epsilon_i \sim N(0, \sigma^2)$ og uavhengige. Frå $n = 20$ målinger (Figur 3) har vi gjennomsnitt $\bar{x} = 28.0$ og $\bar{y} = 29.6$, og $\sum_{i=1}^{n}(x_i - \bar{x})^2 = 64.6$ og $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = 77.2$.



Figur 3: Plott av fysiske modellberekninger (førsteaksen) og sensormålinger (andreaksen). Den heiltrukne blå linja syner $y = x$.

a) Rekn ut estimat for β og α.

Frå residuala estimerer ein σ^2 med $s^2 = 0.48^2$. Rekn ut eit 90 % konfidensintervall for β.

b) Ein stad gjev den fysiske modellen $x_0 = 30$. Finn predikert sensormåling der.

For kva resultat i den fysiske modellen vert usikkerheita i predikert sensormåling minst?

Uten å rekna, kva kan du seie om vi i staden skulle predikert ei sensormåling for $x_0 = 20$?