Problem 1 Lamb chops

a) Let X be Poisson distributed with probability mass function $P(X = x) = \frac{\mu^x}{x!}e^{-\mu}$, $x = 0, 1, \ldots$, and set parameter $\mu = 3$.

Compute the probability of X being equal or less than 3.

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

b) A zero-truncated Poisson distribution with parameter μ is characterized by probability mass function $f(y) = P(X = y|X > 0)$, where X is Poisson distributed with parameter μ.

Show that the probability mass function of Y is

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

Find the expected value $E(Y)$ and compute it for $\mu = 3$.

Karin will have guests for lamb chops this Christmas. She assumes that the number of lamb chops pieces each person eats is zero-truncated Poisson distributed with parameter μ and independent for different persons. Karin uses data from the Christmas party at work to estimate the parameter μ.

c) At this work Christmas party she observes the number of pieces; $y_1, \ldots, y_{21}$, each of the 21 participants eats.

Show that the log likelihood function for μ is

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

Use first-order Taylor expansion: $\ln(1 - e^{-\mu}) \approx \ln(1 - e^{-3}) + \frac{e^{-3}}{1-e^{-3}}(\mu - 3)$ to find an approximation to the maximum likelihood estimator $\hat{\mu}$.

The guests eat a total of 74 pieces of lamb chops. Compute the estimate from the available data.

Figure 1 shows the histogram of the number of pieces eaten by the participants at the Christmas party. Use the probability mass function of the zero-truncated Poisson distribution to discuss if the model fits well with the histogram summarizing the data from the 21 participants. (Use the Poisson table with value closest to the maximum likelihood estimates.)

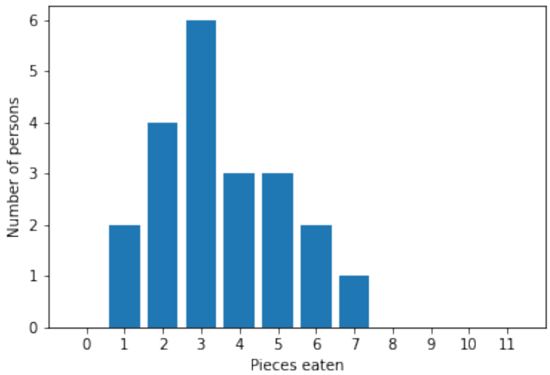

Figure 1: Data of eaten lamb chops pieces at the work Christmas party. There are 21 participants.

Problem 2 Light bulbs

a) Let the stochastic variable X be exponential distributed such that the probability density is $f(x) = \frac{1}{\beta}\exp(-x/\beta)$, $x > 0$. We set $\beta = 1$ here.

Compute $P(X < 3)$.

Find the theoretical median m of X, that is the value of the constant m such that $P(X \leq m) = 0.5$.

b) A toilet has three light bulbs. We assume that the life time of the bulbs, X_1, X_2 and X_3, are independent and exponential distributed with parameter $\beta = 1$ year. There will be light in the room as long as one of the bulbs lives. We define Y as the time when the toilet gets dark.

Show how the event $Y \leq y$ can be expressed as an event involving X_1, X_2 and X_3.

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

Find the probability density function of Y.

c) Assume next that the three light bulbs are of different types. They are exponential distributed and independent, but with different parameters; X_1 has parameter $\beta = 1$, X_2 has parameter $\beta = 2$ and X_3 has parameter $\beta = 3$.

One can use Monte Carlo simulation to understand the statistical properties of this situation. Explain how this is done in the Python code in Figure 2. What do the results and plot to the right in this figure show?

```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)
```

Results (print):

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

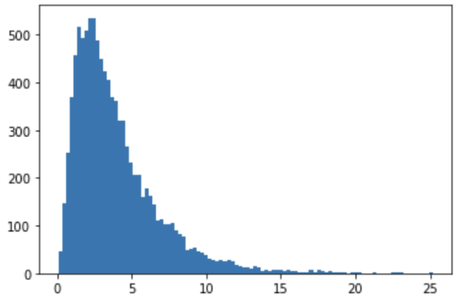

Figure 2: Monte Carlo simulation of life times for three light bulbs. Python code (left) and results of the code (right).

Problem 3 Skiing

a) X is normal distributed with expected value $\mu = 10$ and standard deviation $\sigma = 0.5$.

Compute $P(X < 9)$.

Find c such that $P(X > c) = 0.99$.

Johannes is a dedicated skier. A competitor of Johannes says that he uses 10 minutes up a particular hill, under average conditions. Johannes thinks he is faster than his competitor up this hill.To check this, he skis up the hill ten different days, under various conditions. Let $x_i =$ be the ski time on day $i = 1, \ldots, 10$. The times are 10.1, 9.3, 9.75, 9.9, 10.15, 9.55, 9.8, 9.95, 9.45, 9.6. From this data, we have $\sum_{i=1}^{10} x_i = 97.55$ and $\sum_{i=1}^{10} x_i^2 = 952.3$

b) Assume that the time to ski up the hill is normal distributed with expected value μ and unknown variance σ^2. We further assume that different times are independent.

Formulate the situation of Johannes as a hypothesis test.

Perform the hypothesis test at significance level 0.05.

It is difficult to check if the data are normal distributed in this case. Test the hypothesis using a sign test instead, with significance level 0.05.

Problem 4 Salt data

A physical model is used to understand the salt variability in the Trondheim fjord. One can also gauge the salt level by a sensor at selected locations. This is for instance used to calibrate the physical model. Let x_i and y_i be the result of the physical model and sensor measurements, respectively. Here, $i = 1, \ldots, n$ indicates locations in the fjord where one has both data available at some points in time.

One suggests a regression model $Y_i = \alpha + \beta x_i + \epsilon_i$, $i = 1, \ldots, n$, where $\epsilon_i \sim N(0, \sigma^2)$ and independent. From $n = 20$ measurements (Figure 3) we get means $\bar{x} = 28.0$ and $\bar{y} = 29.6$, and $\sum_{i=1}^{n}(x_i - \bar{x})^2 = 64.6$ and $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = 77.2$.

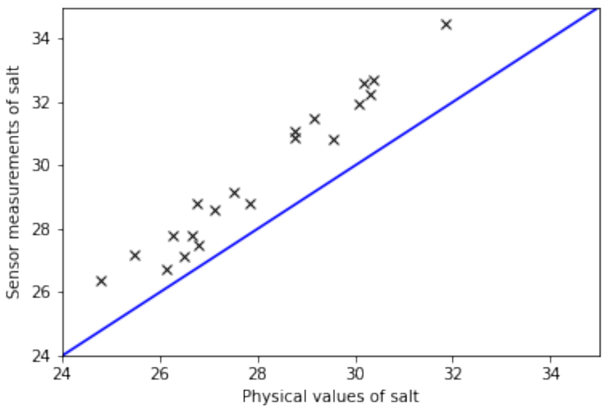

Figure 3: Plot of physical model (first axis) and sensor measurements (second axis).Solid blue line shows $y = x$.

a) Calculate estimates for β and α.

From residuals one estimates σ^2 by $s^2 = 0.48^2$. Find a 90 % confidence interval for β.

b) At a certain place and time, the physical model gives $x_0 = 30$. Find the associated predicted sensor measurement.

For which results in the physical model is uncertainty in predicted sensor measurement at its smallest?

Without doing any calculations, what can you say about doing a sensor prediction when $x_0 = 20$?