1. Let $X = \{X_1, \ldots, X_n\}$ be a random sample of $n$ i.i.d. Poisson random variables with parameter $\lambda$.

(a) (8pt) Find the maximum likelihood for $\lambda$ and its asymptotic sampling distribution.

Solution:
The log-likelihood can be written as:

$$\ell(X; \lambda) = -n\lambda + \left(\sum_{i=1}^{n} X_i \right) \log \lambda - \log \left(\prod_{i=1}^{n} X_i!\right)$$

so that

$$\dot{\ell}(X; \lambda) = -n + \frac{\sum_{i=1}^{n} X_i}{\lambda}$$

and

$$\ddot{\ell}(X; \lambda) = -\frac{\sum_{i=1}^{n} X_i}{\lambda^2} < 0$$

so that the MLE of $\lambda$ is

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}_n$$

By CLT,

$$\frac{\sqrt{n}(\bar{X}_n - \lambda)}{\sqrt{\lambda}} \xrightarrow{d} N(0, 1)$$

as $n \to \infty$. Therefore:

$$\hat{\lambda} \sim N(\lambda, \lambda/n)$$

(b) (8pt) Find the maximum likelihood estimator for the parameter $\mu = e^{-\lambda}$.

Solution:
By the invariance principle the MLE of $\mu$ is:

$$\hat{\mu} = e^{-\hat{\lambda}} = e^{-\bar{X}_n}$$

Suppose now that, rather than observing the actual values of the random variables $X_i$, we are just able to register whether they are null or positive. More precisely, only the events $X_i = 0$ or $X_i > 0$ for $i = 1, \ldots, n$ are observed.

(c) (8pt) Find the maximum likelihood for $\lambda$ for these new observations.

Solution:
Our sample now can be seen as $n$ realizations of a Bernoulli variable $Y$ with parameter $p = e^{\lambda}$, i.e. $P(Y = 0) = p$ and $P(Y = 1) = 1 - p$. Hence,

$$\ell(X; \lambda) = (n - \sum_{i=1}^{n} Y_i) \log p + \sum_{i=1}^{n} Y_i \log(1 - p)$$

By standard calculations we have that the MLE of $p$ is:

$$\hat{p} = (n - \sum_{i=1}^{n} Y_i)/n.$$
Therefore, by the invariance principle, the MLE of $\lambda$ is:

$$\hat{\lambda} = -\log \left( \frac{n - \sum_{i=1}^{n} Y_i}{n} \right)$$

that exists only for $n \neq \sum_{i=1}^{n} Y_i$, i.e. there is at least one null observation.

(d) (8pt) When does the maximum likelihood estimator not exist? Assuming that the true value of $\lambda$ is $\lambda_0$, compute the probability that the maximum likelihood estimator does not exist.

**Solution:**
The MLE exists for $n \neq \sum_{i=1}^{n} Y_i$. Therefore we have to calculate the probability:

$$\mathbb{P}_{\lambda_0} \left( n = \sum_{i=1}^{n} Y_i \right) = \prod_{i=1}^{n} \mathbb{P}_{\lambda_0}(Y_i = 1) = (1 - e^{-\lambda_0})^n$$

2. Let $X = \{X_1, \ldots, X_n\}$ be a random sample of $n$ i.i.d. random variables with densities:

$$f_X(x; \theta) = \begin{cases} \frac{\theta^x}{x!} e^{-\theta x} & \text{if } x > 0, \\ 0 & \text{otherwise} \end{cases}$$

with $\theta > 0$ is an unknown parameter. Moreover, consider another random sample $Y = \{Y_1, \ldots, Y_n\}$ of $n$ i.i.d. random variables with densities:

$$f_Y(y; \mu) = \begin{cases} \frac{\mu^y}{y!} e^{-\mu y} & \text{if } y > 0, \\ 0 & \text{otherwise} \end{cases}$$

with $\mu > 0$ is another unknown parameter. We further assume that the two sample are independent (i.e. $X_i \perp Y_j$, for all $i, j$).

(a) [10pt] Find the Generalized Likelihood Ratio Test (GLRT) statistic for testing:

$$\begin{cases} H_0: \theta = \mu, \\ H_1: \theta \neq \mu. \end{cases}$$

**Solution:**
Let us denote with:

$$V = \{X_1, \ldots, X_n, Y_1, \ldots, Y_n\}$$

the sample of size $2n$ obtained pooling together the samples $X$ and $Y$. The log–likelihood corresponding to $V$ is:

$$\text{lik}(V; \theta, \mu) = \text{lik}(X; \theta) \text{lik}(Y; \mu) = \frac{\theta^{3n} \mu^{3n}}{2^{2n}} e^{-\theta \sum_{i=1}^{n} X_i} e^{-\mu \sum_{i=1}^{n} Y_i} \prod_{i=1}^{n} X_i^2 Y_i^2$$

The GLRT can be written as:

$$\Lambda(V) = \frac{\sup_{0, \mu} \text{lik}(V; \hat{\theta}, \hat{\mu})}{\sup_{\theta, \mu} \text{lik}(X; \theta) \text{lik}(Y; \mu)} = \frac{\text{lik}(V; \hat{\theta}, \hat{\mu})}{\text{lik}(X; \hat{\theta}) \text{lik}(Y; \hat{\mu})}$$

where the hat denotes the MLE. Since

$$\partial_{\theta} \ell(X; \theta) = \frac{3n}{\theta} - \sum_{i=1}^{n} X_i$$

and

$$\partial_{\theta}^2 \ell(X; \theta) = -\frac{3n}{\theta^2} < 0$$

the MLE of $\theta$ is $\hat{\theta} = \frac{3n}{\sum_{i=1}^{n} X_i}$. Analogously, we have $\hat{\mu} = \frac{3n}{\sum_{i=1}^{n} Y_i}$ and $\hat{\theta}_0 = \frac{6n}{\sum_{i=1}^{n} X_i + \sum_{i=1}^{n} Y_i}$. Hence,

$$\Lambda(V) = \frac{\hat{\theta}^{6n} \exp(-\hat{\theta}_0 \sum_{i=1}^{n} (X_i + Y_i))}{\theta^{6n} \mu^{3n} \exp(-\theta \sum_{i=1}^{n} X_i - \mu \sum_{i=1}^{n} Y_i)} = \frac{\hat{\theta}^{6n}}{\hat{\theta}_0^{6n} \mu^{3n}}$$

2
Let us define now the following statistic:

\[ T := \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_i + \sum_{j=1}^{n} Y_j} \]

(b) [10pt] Show that the GLRT rejects \( H_0 \) if \( T(1 - T) < k \), for a suitable constant \( k \).

**Solution:**

The GLRT statistics reject for \( \Lambda(V) < c \), for a suitable constant \( c \). Then

\[
\Lambda(V) = \frac{\hat{\theta}^6}{\hat{\theta}^{3n} \hat{\mu}^{3n}} = \frac{\left( \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_i + \sum_{j=1}^{n} Y_j} \right)^{6n}}{\left( \frac{3n}{\sum_{i=1}^{n} X_i} \right)^{3n} \left( \frac{3n}{\sum_{i=1}^{n} Y_j} \right)^{3n}} = 2^{6n} \left( \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_i + \sum_{j=1}^{n} Y_j} \right)^{3n} \left( \frac{\sum_{i=1}^{n} Y_j}{\sum_{i=1}^{n} Y_j + \sum_{i=1}^{n} X_i} \right)^{3n}
\]

so that we reject for \( T(1 - T) < k \), with \( k = c^{1/3n}/4 \).

3. A company wants to monitor the efficiency of two employees in completing an assigned task. For this reason, the performances of two employees (denoted by \( A \) and \( B \)) were measured by recording the times needed to complete the assigned tasks. Hence, the following two samples have been collected:

\[ x_A = \{5.18, 13.43, 6.31, 3.18, 4.91, 11.07\}, \]
\[ x_B = \{5.50, 18.16, 8.14, 9.14, 14.24, 10.72\} \]

where the duration of each task is measured in hours.

(a) [10pt] Perform a test at 10% of significance for testing the hypothesis that employee \( A \) is faster than \( B \). Discuss critically the choice of the test used.

**Solution:**

Since we do not have any information on the distribution of the data, we can use the non-parametric Mann–Whitney for testing:

\[
H_0 : \quad F_A(x) = F_B(x), \quad \forall x \\
H_1 : \quad F_A(x) \geq F_B(x)
\]

We have that the sum of ranks are \( T_A = 30 \) and \( T_A = 48 \). The critical value for the one-tailed test is 31, so that \( T_A < 31 \), we can reject then \( H_0 \) at 10% of significance.

Suppose now that the time \( T \) needed by an employee for completing a task can be modeled by a continuous random variable with the following probability density function:

\[
f_T(t; \theta) = \begin{cases} 
\frac{1}{2\theta \sqrt{\pi}} e^{-\frac{t^2}{\theta}} & \text{if } t > 0, \\
0 & \text{otherwise}
\end{cases}
\]

with \( \theta > 0 \) an unknown parameter.

(b) [8pt] Given a sample \( T = \{T_1, \ldots, T_n\} \) of i.i.d random variables sampled from \( f_T(t; \theta) \), determine the maximum likelihood estimator of the probability \( \Pr(T > 7) \).

**Solution:**

\[
\Pr(T > 7) = \int_{7}^{\infty} \frac{1}{2\theta \sqrt{\pi}} e^{-\frac{t^2}{\theta}} dt = \int_{\sqrt{7}/\theta}^{\infty} e^{-y} dy = e^{-\sqrt{7}/\theta}
\]
Hence, by invariance principle, the MLE of \( P_\theta(T > 7) \) is \( e^{-\sqrt{T/\hat{\theta}}} \), where \( \hat{\theta} \) is the MLE of the parameter \( \theta \). By standard calculations or by noting that \( \sqrt{T} \sim \text{Exp}(\theta) \), we can derive that the MLE of \( \theta \) is:

\[
\hat{\theta} = \frac{\sum_{i=1}^{n} \sqrt{T_i}}{n}
\]

(3)

so that the MLE of \( P_\theta(T > 7) \) is \( P_{\hat{\theta}}(T > 7) \).

(c) Under the parametric model (1) for the random variable \( T \) and given the samples \( x_A, x_B \), estimate the probability that the time needed by an employee for completing a task is larger than 7 hours, under the further assumption that 55% of the employees are similar to employee \( A \) and 45% to employee \( B \).

Solution:
Using the samples \( x_A \) and \( x_B \), by (3) we find that following MLE estimates for the parameter \( \theta \):

\[
\hat{\theta}_A = 2.63, \quad \hat{\theta}_B = 3.26
\]

(4)

Therefore, by (2),(3) and (4), we have:

\[
0.55 \hat{\|}_{\hat{\theta}}^A(T > 7) + 0.45 \hat{\|}_{\hat{\theta}}^B(T > 7) = 0.42
\]

4. Let the independent random variables \( Y_1, Y_2, \ldots, Y_n \) be such that we have the following linear model:

\[
Y_i = \beta_0 + \beta_1 x_i + \beta_2 (x_i - 3.5)_+ + \epsilon_i
\]

for \( i = 1, \ldots, n \), where \( \epsilon_i \) are i.i.d. normal random variables such that \( \epsilon_i \sim N(0, \sigma^2) \) and with \((y)_+\) we denoted the positive part of the real number \( y \) (i.e. \((y)_+ := \max(0, y)\)). We collect the following sample of observations

\[
y = \{1, 2, 4, 5, 4, 3, 1\}
\]

\[
x = \{0, 1, 2, 3, 4, 5, 6\}
\]

(a) If we rewrite the linear model using the usual matrix formalism

\[
Y = X\beta + \epsilon
\]

write down the design matrix \( X \) of the linear model.

Solution:

\[
X = \begin{pmatrix}
1 & 0 & 0 \\
1 & 1 & 0 \\
1 & 2 & 0 \\
1 & 3 & 0 \\
1 & 4 & 0.5 \\
1 & 5 & 1.5 \\
1 & 6 & 2.5
\end{pmatrix}
\]

(b) Given that

\[
(X^\top X)^{-1} = \begin{pmatrix}
0.65 & -0.24 & 0.35 \\
-0.24 & 0.14 & -0.26 \\
0.35 & -0.26 & 0.65
\end{pmatrix}
\]

estimate the model coefficients and write down the fitted model.

Solution:
Since the LSE can be written as:

\[
\hat{\beta} = (X^\top X)^{-1} X^\top Y
\]

4
we have:
\[
\hat{\beta} = (1.27, 1.54, -3.27)^T
\]
and
\[
\hat{y} = 1.27 + 1.54x - 3.27(x - 3.5),
\]

(c) [8pt] Calculate the prediction of the fitted model at \(x = 4.5\). Assuming that the sum of squared residuals equals 7.8, calculate a 95% confidence interval for this prediction.

**Solution:**

The prediction is:
\[
\hat{y} = 1.27 + 1.54 \cdot 4.5 - 3.27(4.5 - 3.5) = 4.93
\]

The estimated covariance matrix of the fitted coefficient is:
\[
\Sigma_{\hat{\beta}, \hat{\beta}} = s^2(X^TX)^{-1}
\]

with \(s^2 = RSS/(7 - 3) = 7.8/4 = 1.95\). Then
\[
\text{Var}\hat{Y} = \text{Var}\hat{\beta}_0 + x^2\text{Var}\hat{\beta}_1 + (x - 3.5)^2\text{Var}\hat{\beta}_2 + 2x\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) + 2(x - 3.5)\text{Cov}(\hat{\beta}_0, \hat{\beta}_2) + 2x(x - 3.5)\text{Cov}(\hat{\beta}_1, \hat{\beta}_2)
\]
\[
= \Sigma_{1,1} + 4.5^2\Sigma_{2,2} + \Sigma_{3,3} + 9\Sigma_{1,2} + 2\Sigma_{1,3} + 9\Sigma_{2,3}
\]

Therefore a 95% CI for the prediction is:
\[
4.93 \pm t_{4.0.024} \sqrt{\text{Var}\hat{Y}} = 4.93 \pm 4.47
\]