Sheet 1

1. We consider the following sequence

4	1	1	1	1	1	2	2	1	2
1	1	1	2	1	3	4	1	1	1
2	1	1	4	1	1	1	1	1	1

   (a) Give a 95% confidence interval of the mean. In order to get full marks, you need to justify the sampling distribution you use.

   (b) We suspect that the sample is i.i.d. from the geometric distribution defined by the probability mass function

$$\mathbb{P}[X_i = k \mid p] = p(1-p)^{k-1}, \quad k \in \mathbb{N}.$$

      i. Show that the cumulative distribution function is

$$\mathbb{P}[X_i \leq x \mid p] = 1 - (1-p)^{\lfloor x \rfloor}, \quad x \geq 0$$

      where $\lfloor x \rfloor$ designates the integral part of x.

      ii. Perform a χ^2 goodness of fit test to check if we may assume that the sample follows the geometric distribution with parameter $p = 0.5$.

2. (a) We observe n independent realisations of a geometric distribution with unknown parameter p. Derive the maximum likelihood estimator of p.

   (b) Now we assume that the first m observations are right-censored. Derive the corresponding maximum likelihood estimator.

   (c) We consider the Poisson distribution defined by the probability mass function

$$\mathbb{P}[X_i = k \mid \lambda] = \frac{\lambda^k \exp(-\lambda)}{k!}, \quad k \in \mathbb{N}_0.$$

      i. Derive the maximum likelihood estimator of λ.

      ii. Is this estimator biased?

      iii. Compare its variance to the Cramér–Rao bound. Conclude.

3. (a) What is the Delta method?

   (b) Derive the maximum likelihood estimator of the mean of the exponential distribution.

   (c) Use the Delta method to derive the asymptotic distribution of the rate parameter of an i.i.d. sample from an exponential distribution.

4. In this problem, we consider an i.i.d. sample from a mixture of two uniform distributions. With probability a, the sample is uniformly distributed over $[-1, 1]$ and with probability $1 - a$ it is uniformly distributed over $[0, 1]$.

   (a) Give the density function.

   (b) Calculate the cumulative distribution function.

   (c) Use the Fisher–Neyman theorem to prove that the number of samples that belong to $[-1, 0)$ is a sufficient statistic for a.

   (d) Find the maximum likelihood estimator of a.

   (e) Calculate the expectation of this distribution and deduce an other estimator of a.

5. The table below provides data on height (in cm), stride length (in cm), and gender (encoded as an indicator for males) collected from various individuals.

Height (cm)	176	168	170	167	166	167	170
Gender (0:female, 1:male)	1	0	0	0	1	0	0
Stride (cm)	75	61	50	58	74	61	55

   Use the method of least squares to fit a linear model to predict the stride length based on height and gender. For this exercise, construct the model without including an intercept coefficient. Describe your findings on how each predictor influences stride length.

Solution Sheet 1

1. (a) In this case the dataset is sufficiently large ($n = 30$), thus we may use the confidence interval based on the normal distribution

$$\left[\bar{x} - \frac{1.96S}{\sqrt{30}},\ \bar{x} + \frac{1.96S}{\sqrt{30}}\right] = [1.19, 1.88]$$

(b) i. First of all, we observe that since the support of the probability mass function is $\mathbb{N}$, we have $\mathbb{P}[X_i \leq x \mid p] = \mathbb{P}[X_i \leq \lfloor x \rfloor \mid p]$. Then, it suffices to use the formula for the partial sum of a geometric series

$$F(x) = \mathbb{P}[X_i \leq \lfloor x \rfloor \mid p] = p \sum_{k=0}^{\lfloor x \rfloor - 1} (1-p)^k = p \cdot \frac{1-(1-p)^{\lfloor x \rfloor}}{p} = 1 - (1-p)^{\lfloor x \rfloor}.$$

ii. For simplicity we partition the range into three bins

$$\{1\}, \quad \{2\}, \quad [3, \infty).$$

The numbers of entries per bin are respectively $o_1 = 21$, $o_2 = 5$ and $o_3 = 4$ (the last bin contains only 4 entries, although we can't do better unless taking only 2 bins). The expected frequencies of the bins are

$$e_1 = 0.5, \quad e_2 = 0.25, \quad e_3 = 0.25$$

We compare the expected number of entries per bin to the observations

$$\sum_{i=1}^{3} \frac{(o_i - 30 \cdot e_i)^2}{30 \cdot e_i \cdot (1-e_i)} = 8.09.$$

For the χ^2 with 2 degrees of freedom, this would yield a p-value of 0.0175 which is below the threshold of 5%. According to this test we would reject the assumption that the data follow the geometric distribution with parameter 0.5.

2. (a) The maximum likelihood estimator is a solution of

$$\begin{aligned}\hat{p} &= \operatorname*{argmax}_{p \in [0,1]} \prod_{i=1}^{n} p(1-p)^{X_i - 1} = \operatorname*{argmax}_{p \in [0,1]} n \log(p) + \sum_{i=1}^{n} (X_i - 1) \log(1-p) \\ &= \operatorname*{argmax}_{p \in [0,1]} \underbrace{n \log(p) + n(\bar{X} - 1) \log(1-p)}_{\ell(p)}.\end{aligned}$$

In particular $\hat{p}$ must cancel the first order derivative of the log-likelihood ℓ

$$0 = \frac{n}{\hat{p}} + \frac{n(1-\bar{X})}{(1-\hat{p})} \Leftrightarrow \hat{p} = \frac{1}{\bar{X}}.$$

Finally we need to check the sign of the second order derivative of the log likelihood

$$\frac{\partial^2 \ell(p)}{\partial p^2} = -\frac{n}{p^2} - \frac{n(1-\bar{X})}{(1-p)^2}.$$

This quantity is strictly negative since $n \geq 1$ and $\bar{X} \geq 1$ (because the range of the geometric distribution is $\mathbb{N}$).

(b) This time, the likelihood must incorporate the censoring

$$\begin{aligned}\hat{p} &= \underset{p\in[0,1]}{\operatorname{argmax}} \left(\prod_{i=1}^{m}(1-p)^{X_i}\right)\cdot\left(\prod_{i=m+1}^{n} p(1-p)^{X_i-1}\right)\\ &= \underset{p\in[0,1]}{\operatorname{argmax}}\ p^{n-m}(1-p)^{n(\bar{X}-1)+m}\\ &= \underset{p\in[0,1]}{\operatorname{argmax}}\ \underbrace{(n-m)\log(p)+(n(\bar{X}-1)+m)\log(1-p)}_{\ell(p)}.\end{aligned}$$

As before, it must cancel the first order derivative of the log likelihood

$$0=\frac{n-m}{\hat{p}}-\frac{n(\bar{X}-1)+m}{1-\hat{p}} \Leftrightarrow \hat{p}=\frac{n-m}{n\bar{X}}.$$

Again we should check the sign of the second order derivative.

(c) i. As usual

$$\begin{aligned}\hat{\lambda} &= \underset{\lambda\in\mathbb{R}_+}{\operatorname{argmax}} \prod_{i=1}^{n}\frac{\lambda^{X_i}\exp(-\lambda)}{k!} = \underset{\lambda\in\mathbb{R}_+}{\operatorname{argmax}} \prod_{i=1}^{n}\lambda^{X_i}\exp(-\lambda)\\ &= \underset{\lambda\in\mathbb{R}_+}{\operatorname{argmax}}\ \lambda^{n\bar{X}}\exp(-n\lambda) = \underset{\lambda\in\mathbb{R}_+}{\operatorname{argmax}}\ \underbrace{n\bar{X}\log(\lambda)-n\lambda}_{\ell(\lambda)}.\end{aligned}$$

The estimator must cancel the first order derivative of the log likelihood

$$0=\frac{n\bar{X}}{\hat{\lambda}}-n \Leftrightarrow \hat{\lambda}=\bar{X}.$$

We may either compute the second order derivative or notice that $\ell(\lambda)$ is the sum of a concave function and of a linear function, thus it is concave.

ii. It is immediate if we remember that the expectation of a Poisson random variable is equal to λ

$$\mathbb{E}[\hat{\lambda}]=\mathbb{E}[\bar{X}]=\mathbb{E}[X_i]=\lambda.$$

iii. First of all, it is straightforward to observe that

$$\mathbb{V}[\hat{\lambda}]=\frac{\lambda}{n}$$

if we use the properties of the variance and we remember that the variance of a Poisson random variable is equal to λ. Now we need to calculate the Fisher information

$$I(\lambda)=n\cdot\mathbb{E}\left[\left(\frac{X_i}{\lambda}-1\right)^2\right]=\frac{n}{\lambda^2}\cdot\mathbb{E}\left[(X_i-\lambda)^2\right]=\frac{n}{\lambda^2}\cdot\mathbb{V}[X_i]=\frac{n}{\lambda}.$$

In particular, the estimator $\hat{\lambda}$ is an unbiased estimator that matches the Cramér–Rao bound, thus it is a minimum variance unbiased estimator.

3. (a) Lecture notes.

   (b) Standard question : $\hat{\mu} = \bar{X}$

   (c) First of all we use the central limit theorem to observe that

$$\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} \mathcal{N}(0, \mu^2),$$

   where we denote by μ the mean. Besides, the invariance of the maximum likelihood estimator ensures the $1/\bar{X}$ is the maximum likelihood estimator of λ. Then, it suffices to apply the delta method with the function $\lambda = g(\mu) = 1/\mu$ to obtain

$$\sqrt{n}\left(\frac{1}{\bar{X}} - \lambda\right) \xrightarrow{d} \mathcal{N}(0, \lambda^2).$$

4. (a) *(This is a less standard question)*
   We use the total probability formula

$$f(x|a) = a \cdot f_1(x) + (1-a) \cdot f_2(x) = a \cdot \frac{\mathbf{1}_{[-1,1]}(x)}{2} + (1-a) \cdot \mathbf{1}_{[0,1]}(x)$$
$$\frac{a\mathbf{1}_{[-1,0]}(x)}{2} + (1 - a/2) \cdot \mathbf{1}_{[0,1]}(a).$$

   (b) If $x \leq 0$

$$F(x|a) = \int_{-1}^{x} f(t|a)\ dt = \frac{a}{2} \cdot \int_{-1}^{x} dt = \frac{a(x+1)}{2}$$

   if $x \geq 0$

$$F(x|a) = \frac{a}{2} \cdot \int_{-1}^{0} dt + (1 - a/2) \cdot \int_{0}^{x} dt = x(1 - a/2) + a/2.$$

   (c) It is immediate as the likelihood of a series of observations $X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n$ is

$$f(\mathbf{x}|a) = \prod_{i=1}^{n} f(x_i|a) = \left(\frac{a}{2}\right)^m (1 - a/2)^{n-m},$$

   where we denote by m the number of samples in $[-1, 0)$. (the function $h(\mathbf{x})$ is equal to 1 if we refer to the notation of the lecture notes).

   (d)

$$\begin{aligned}\hat{a} &= \underset{a \in [0,1]}{\operatorname{argmax}}\ f(\mathbf{X}|a) = \left(\frac{a}{2}\right)^m (1 - a/2)^{n-m} \\ &= \underset{a \in [0,1]}{\operatorname{argmax}}\ m(\log(a) - \log(2)) + (n-m)\log(1 - a/2).\end{aligned}$$

   In particular, it cancel the first order derivative

$$0 = \frac{m}{\hat{a}} - \frac{1}{2} \cdot \frac{n-m}{1 - \hat{a}/2} \Leftrightarrow \hat{a} = \frac{2m}{n}.$$

(e) The expectation for a single sample is

$$\mathbb{E}[X_i] = \int_{-1}^{1} x \cdot f(x|a)dx = \frac{a}{2}\int_{-1}^{0} x\ dx + (1 - a/2)\int_{0}^{1} x\ dx = \frac{1-a}{2}$$

in particular we deduce an estimator by the method of moments by setting

$$\bar{X} = \frac{1-\hat{a}}{2} \Leftrightarrow \hat{a} = 1 - 2\bar{X}.$$

5. For stability, it is preferable to express the lengths into meters in this case. A linear model would be

$$\mathbb{E}[S] = \begin{pmatrix} H & G \end{pmatrix} \cdot \beta$$

for S is the stride length, G is the gender indicator and H is the height. The regression vector β should be estimated with the method of least squares. In particular it should be a solution of the normal equation

$$\hat{\beta} = (X^\intercal X)^{-1} X^\intercal y$$

where

$$X = \begin{pmatrix} 1.76 & 1 \\ 1.68 & 0 \\ 1.70 & 0 \\ 1.67 & 0 \\ 1.66 & 1 \\ 1.67 & 0 \\ 1.70 & 0 \end{pmatrix}, \qquad y = \begin{pmatrix} 0.75 \\ 0.61 \\ 0.50 \\ 0.58 \\ 0.74 \\ 0.61 \\ 0.55 \end{pmatrix}.$$

The solution is

$$\hat{\beta} = \begin{pmatrix} 0.34 \\ 0.17 \end{pmatrix}.$$

According to this, an increase of the height by $1cm$ is expected to lead to increase of the stride by $0.34cm$ and a males are expected to have a stride that is $17cm$ longer than women.

Sheet 2

1. In this problem, we consider the power law with the density

$$f(x \mid \alpha) = C(\alpha) \cdot x^{-\alpha}, \quad x \geq 1, \; \alpha > 2.$$

   (a) Calculate the normalisation constant $C(\alpha)$.

   (b) Calculate the expectation of a random variable following this distribution and deduce an estimator of α.

   (c) Find the maximum likelihood estimator of α.

   (d) Calculate the cumulative distribution function and deduce the median.

2. In this problem we consider the Laplace distribution with the density

$$f(x \mid \lambda) = \frac{\lambda}{2} \cdot \exp\left(-\lambda |x|\right).$$

   (a) Derive the maximum likelihood $\hat{\lambda}$ estimator of λ.

   (b) Show that $1/\hat{\lambda}$ it is a minimum variance unbiased estimator of $1/\lambda$.

3. We want to weigh two objects A and B of respective mass θ_A and θ_B. k weighings of the object of mass A yield values

$$y_1, \cdots, y_k,$$

   then $n-k$ weighings of both objects together yield values

$$y_{k+1}, \cdots, y_n.$$

   Derive the least squares estimates of (θ_A, θ_B) assuming that the weighings are unbiased, independent, and with the same variance.

4. We consider the following table

1	5	17	3	4	6	4	9	36	4
21	2	12	5	4	13	23	2	14	6
1	23	5	10	5	10	13	19	4	2

   (a) Can we assume that the entries are normally distributed?

   (b) Give a 95% confidence interval for the population mean. Explain your choice of sampling distribution.

   (c) Check the independence of the sequence with the Wald–Wolfowitz run test.

5. (a) Recall the definition of the empirical cumulative distribution function.

   (b) State and demonstrate the Glivenko–Cantelli theorem.