Exercise Sheet 5

Exercise 5.1 - Likelihood

(a) Formulate the likelihood for each of the following problems.

(i) In a study of a root-infecting fungus in wheat, 250 seeds are planted. For technical reasons, it can only be observed that $x \leq 25$ seeds have germinated. Let θ be the probability that a seed will germinate. Give an expression for the likelihood of θ based on the information of the above experiment.

$n = 250$ seeds
X ... Number of germinated seeds, but: only $x \leq 25$ observed
θ ... Probability that a seed germinates
$\Rightarrow X \sim Bin(n = 250, \theta)$ (assuming independence)
We construct a new variable: $Y := (X \leq 25)$
Where $Y \sim Bin(1, \pi)$, with probability function $f_Y(y) = \pi^y (1-\pi)^{1-y}$.

$$\begin{aligned}
L(\theta; y) &= \pi^y (1-\pi)^{1-y}, \quad \text{for } y \in \{0, 1\} \\
&= P(X \leq 25)^y \left(P(X > 25)\right)^{1-y} \\
&\overset{y=1,\text{ since } x \leq 25 \text{ was observed}}{=} P(X \leq 25) \\
&= \sum_{x=0}^{25} \binom{250}{x} \theta^x (1-\theta)^{250-x}
\end{aligned}$$

(ii) Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\theta, 1)$. For various reasons, only the largest value of the sample is reported: $Y = \max(X_1, \ldots, X_n)$. Show that the density of Y is of the following form:

$$f_Y(y) = n \left(\Phi(y - \theta)\right)^{n-1} \phi(y - \theta),$$

where $\Phi(\cdot)$ is the distribution function and $\phi(\cdot)$ the density of the standard normal. What is the likelihood function $L(\theta; y)$? **Hint:** First, determine the distribution function of Y.

First, determining the distribution function of Y:

$$\begin{aligned}
F_Y(y) = P(Y \leq y) &= P(\max(X_1, \ldots, X_n) \leq y) \\
&= P(X_1 \leq y, \ldots, X_n \leq y) \\
&= \prod_{i=1}^{n} P(X_i \leq y) = \prod_{i=1}^{n} F_{X_i}(y) \\
&= \prod_{i=1}^{n} \Phi\left(\frac{y - \theta}{1}\right) = \left(\Phi(y - \theta)\right)^n
\end{aligned}$$

Remarks:

* If the maximum is to be smaller than a y, then each individual X_i must also be smaller than y, $i = 1, \ldots, n$.

* $X_i \sim N(\theta, 1) \Leftrightarrow X_i - \theta \sim \mathcal{N}(0,1), \forall i = 1, \ldots, n$

The density of Y results from the distribution function:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = n \cdot (\Phi(y-\theta))^{n-1} \cdot \phi(y-\theta)$$

Thus, we obtain the likelihood for theta for multiple samples of Y as:

$$L(\theta; y) = \prod_{i=1}^{n} n \cdot (\Phi(y_i - \theta))^{n-1} \cdot \phi(y_i - \theta)$$

(iii) Let X_1, X_2, X_3 be i.i.d. distributed with $X_i \sim \mathrm{C}(\theta, 1)$, where $\theta \in \mathbb{R}$ is the localisation parameter of the Cauchy distribution with density:

$$f(x;\theta) = \frac{1}{\pi} \frac{1}{1 + (x-\theta)^2} \quad , x \in \mathbb{R}$$

Determine the likelihood of θ.

X_1, X_2, X_3 i.i.d. $X_i \sim C(\theta, 1)$, where $\theta \in R$ is the location parameter of the Cauchy distribution

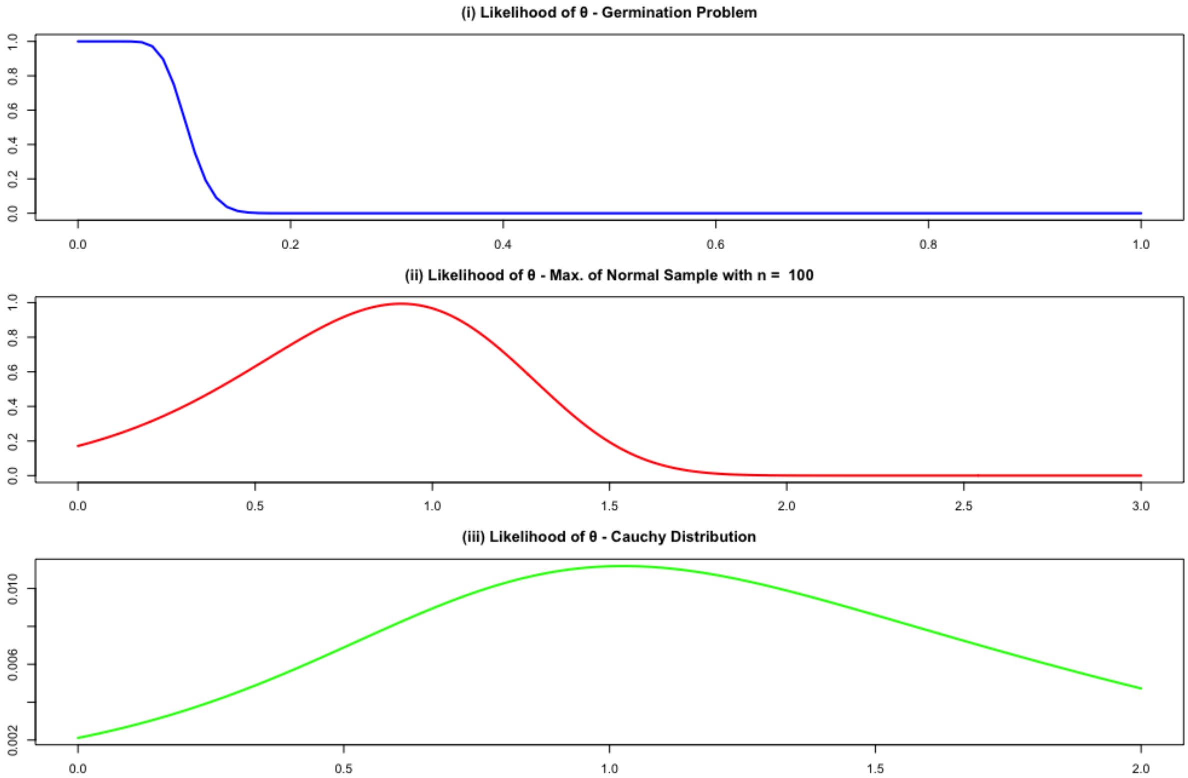
$$f_{X_i}(x_i; \theta) = \frac{1}{\pi} \cdot \frac{1}{1 + (x_i - \theta)^2}, \quad x_i \in \mathbb{R}$$

Likelihood of θ as usual:

$$\begin{aligned} L(\theta; x) &= \prod_{i=1}^{3} \frac{1}{\pi} \cdot \frac{1}{1 + (x_i - \theta)^2} \\ &= \frac{1}{\pi^3} \cdot \frac{1}{(1 + (x_1 - \theta)^2)(1 + (x_2 - \theta)^2)(1 + (x_3 - \theta)^2)} \end{aligned}$$

Remark: here, π is not a parameter, but the number (3.14).

(b) Use R to plot the 3 likelihoods obtained in part (a) for a random sample.



The first plot for part (i) illustrates that the likelihood for parameter values $\theta \in [0, 0.08]$ is positive, but essentially flat, allowing the inference that these values are equally likely given the data. For higher values, the likelihood declines sharply and reaches zero, due to the constraint $x \leq 25$. This sharp drop shows our information about θ is only based on the upper bound 25 of the sum of our sample, where for values far away from $0.1 = \frac{25}{250}$, there is no change in the likelihood.

In the second plot for part (ii), the likelihood derived from observing a single maximum in a sample of $n = 100$ has a pronounced peak near $\theta = 1$, which was the true parameter used to generate the data. Even one observed maximum provides meaningful information about θ. If we were to observe multiple maxima from independent samples of size $n = 100$, the likelihood would exhibit steeper curvature, indicating increased precision (or reduced variance) of the MLE. The likelihood is asymmetric, with a steeper drop-off for higher values of θ. This asymmetry is because the probability of observing an even higher maximum in the sample decreases more rapidly with increasing θ. Conversely, if we plotted the likelihood for θ based on the sample minimum, the relationship would invert due to the symmetry of the normal distribution.

The third plot for part (iii) shows the likelihood for the Cauchy distribution, with the true parameter set to $\theta = 1$ used to generate a random sample of size $n = 3$. The likelihood has a typical unimodal shape with a peak near the true parameter value. However, the Cauchy distribution is well-known for its heavy tails and undefined moments (e.g., no finite mean or variance), which complicates parameter estimation using maximum likelihood. The heavy tails make the likelihood more sensitive to extreme sample values, which can influence the MLE significantly in small samples. Despite these difficulties, the likelihood function itself proves useful in this plot.

Exercise 5.2 - Maximum Likelihood

Let $X = (X_1, \dots, X_n)^T$ be a sample of identically and independently distributed random variables X_i, $i = 1, \dots, n$, with density function

$$f(x;\theta) = \frac{1}{\theta^2} \cdot x \cdot e^{-\frac{x}{\theta}}$$

Determine:

(a) The likelihood and log-likelihood function of X.

We are looking at the gamma distribution with parameter k = 2.
Likelihood: $L(\theta, X) = \prod_{i=1}^{n} \frac{1}{\theta^2} \cdot x_i \cdot exp(-\frac{x_i}{\theta})$
Log-likelihood: $l(\theta, X) = \sum_{i=1}^{n} -log(\theta^2) + log(x_i) - \frac{x_i}{\theta}$

(b) The derivative of the log-likelihood function of X with respect to θ. This is also called the score function.

$\frac{\partial l(\theta,X)}{\partial \theta} = -\frac{2n}{\theta} + \frac{\sum_{i=1}^{n} x_i}{\theta^2} = \frac{2n}{\theta^2} - \frac{\sum_{i=1}^{n} 2x_i}{\theta^3} = s(\theta)$

(c) The Maximum Likelihood Estimator $\widehat{\theta}$ for θ and prove that it is indeed the maximum.

$$\begin{aligned}
s(\theta) = 0 \quad &\Leftrightarrow \quad 0 = \sum_{i=1}^{n} -\frac{2}{\theta} + \frac{x_i}{\theta^2} \\
&\Leftrightarrow \quad 0 = -\frac{2n}{\theta} + \frac{\sum_{i=1}^{n} x_i}{\theta^2} \\
&\Leftrightarrow \quad 2n\theta = \sum_{i=1}^{n} x_i \\
&\Rightarrow \quad \widehat{\theta} = \frac{\sum_{i=1}^{n} x_i}{2n} = \frac{\bar{x}}{2}
\end{aligned}$$

$\frac{\partial s(\theta,X)}{\partial \theta} = \frac{2n}{\theta^2} - \frac{\sum_{i=1}^{n} 2x_i}{\theta^3} \rightarrow$ Plugging in $\hat{\theta}$ for θ yields: $\frac{2n}{(\frac{\sum_{i=1}^{n} x_i}{2n})^2} - \frac{\sum_{i=1}^{n} 2x_i}{(\frac{\sum_{i=1}^{n} x_i}{2n})^3} = \frac{8n^3}{(\sum_{i=1}^{n} x_i)^2} - \frac{16n^3}{(\sum_{i=1}^{n} x_i)^2}$

Which is clearly negative, thus proving that the MLE is indeed the maximum.

(d) Calculate the expected value of the second derivative of the log-likelihood function (**Hint:** $\mathbb{E}(X_i) = 2\theta$). Generally speaking, what is the meaning of the second derivative? What does the expected value of the derivative of the score function mean in the context of the (log-)likelihood function?

$$\begin{aligned}
\mathbb{E}(\frac{\partial^2 l(\theta, X)}{\partial \theta^2}) &= \mathbb{E}(\frac{\partial(-\frac{2n}{\theta} + \frac{\sum_{i=1}^{n} x_i}{\theta^2})}{\partial \theta}) = \mathbb{E}(\frac{2n}{\theta^2} - \frac{2\sum_{i=1}^{n} x_i}{\theta^3}) \\
&= \mathbb{E}(\frac{2n}{\theta^2}) - \mathbb{E}(\frac{2\sum_{i=1}^{n} x_i}{\theta^3}) = \frac{2n}{\theta^2} - \frac{2\sum_{i=1}^{n} \mathbb{E}(x_i)}{\theta^3} \\
&= \frac{2n}{\theta^2} - \frac{2n2\theta}{\theta^3} = \frac{2n}{\theta^2} - \frac{4n}{\theta^2} \\
&= -\frac{2n}{\theta^2}
\end{aligned}$$

The second derivative denotes the curvature of a function at a given point. In context to the likelihood, this signifies the 'peakyness' of the likelihood function at a given θ, thus how much information about the true parameter we can expect to gain from our random variable or our data. This is also called the fisher information, which is denoted as the negative value of the above, so $\frac{2n}{\theta^2}$, since it intuitively makes more sense that information would be a positive value.

Exercise 5.3 - Multi-dimensional MLE

Let $(X_1, \dots, X_n)$ be a sample of i.i.d., Gamma distributed random variables, with parameters α, β. This means that their density, for $x > 0$, is:

$$f_X(x, \alpha, \beta) = \frac{x^{\alpha-1} e^{-\beta x} \beta^\alpha}{\Gamma(\alpha)},$$

where $\alpha, \beta > 0$, and $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$.

(a) Write down the likelihood and log-likelihood of the sample.

Given the density of the Gamma distribution, the likelihood will be:

$$\begin{aligned} L(\alpha, \beta; x_1, \dots, x_n) &= \prod_{i=1}^n \frac{x_i^{\alpha-1} e^{-\beta x_i} \beta^\alpha}{\Gamma(\alpha)} \\ &= \left(\prod_{i=1}^n x_i^{\alpha-1} e^{-\beta x_i} \right) \beta^{n\alpha} \Gamma(\alpha)^{-n}. \end{aligned}$$

The log-likelihood is:

$$\begin{aligned} \ell(\alpha, \beta; x_1, \dots, x_n) &= \log \left(L(\alpha, \beta; x_1, \dots, x_n) \right) \\ &= \sum_{i=1}^n \left((\alpha - 1) \log x_i - \beta x_i \right) + n\alpha \log \beta - n \log \left(\Gamma(\alpha) \right). \end{aligned}$$

(b) Derive the MLE for both parameters μ and σ^2. It is not necessary to show a closed form expression for the estimators.

*Hint: make use of the **digamma** function* $\psi(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$

The procedure to find the MLE is the same as in one dimension, with the only difference being that we need to calculate the derivatives for both variables and set them both to 0 simultaneously.

$$\mathbf{s}(\alpha, \beta) = \begin{pmatrix} \frac{\partial \ell}{\partial \alpha} \\ \frac{\partial \ell}{\partial \beta} \end{pmatrix} = \begin{pmatrix} n \log \beta + \sum \log x_i - n \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} \\ \frac{n\alpha}{\beta} - \sum x_i \end{pmatrix} \stackrel{!}{=} 0$$

The second equation is easily solved by:

$$\hat{\beta} = \frac{n\alpha}{\sum x_i} = \frac{\alpha}{\bar{x}}.$$

For the first equation we substitute what we obtain in the second and get:

$$\begin{aligned} n \left(\log(\alpha) - \log(\bar{x}) \right) + \sum log(x_i) - n\psi(\alpha) &= 0 \\ \left(\log(\alpha) - \log(\bar{x}) \right) + \overline{log(x)} - \psi(\alpha) &= 0, \end{aligned}$$

where $\overline{\log(x)} = \frac{1}{n} \sum \log(x_i)$. This equation does not have an easy closed form solution, so we stop here with the calculations.

(c) Show that if two estimators $\hat{\alpha}$ and $\hat{\beta}$ are such that $s(\hat{\alpha}, \hat{\beta}) = 0$, then they are maximum likelihood estimators.

Hint: the derivative of the digamma function $\psi'(\alpha)$ is always positive. Moreover $\alpha\psi'(\alpha) > 1$ for all values of α.

This is easily done in one dimension by showing that the log-likelihood function is concave, i.e. that its second derivative is negative. In 2 dimensions the concept is once again exactly the same: this time instead of showing that the second derivative is negative, we need to show that the *Hessian* matrix is *negative definite* in $\hat{\alpha}, \hat{\beta}$.

First let's calculate the Hessian matrix:

$$\mathbf{H}(\alpha, \beta) = \begin{pmatrix} \frac{\partial^2 \ell}{\partial \alpha^2} & \frac{\partial^2 \ell}{\partial \alpha \partial \beta} \\ \frac{\partial^2 \ell}{\partial \beta \partial \alpha} & \frac{\partial^2 \ell}{\partial \beta^2} \end{pmatrix} = \begin{pmatrix} -n\psi'(\alpha) & \frac{n}{\beta} \\ \frac{n}{\beta} & -\frac{n\alpha}{\beta^2} \end{pmatrix}.$$

Now, to see that H is negative definite it is sufficient to notice that it's diagonal elements are negative and it's off-diagonal elements are positive.

In fact we know from linear algebra that the *trace* of a 2x2 matrix is the sum of its eigenvalues, while its determinant is the product of the eigenvalues, so if the sum of two numbers is negative, and their product is positive, both of them need to be negative.

$$\begin{cases} tr(\mathbf{H}) = -n\left(\psi'(\alpha) + \frac{\alpha}{\beta^2}\right) < 0 \\ det(\mathbf{H}) = H_{\alpha,\alpha} \cdot H_{\beta,\beta} - H_{\alpha,\beta} \cdot H_{\beta,\alpha} = \frac{n^2\alpha\psi'(\alpha)}{\beta^2} - \frac{n^2}{\beta^2} = \frac{n^2}{\beta^2}(\alpha\psi'(\alpha) - 1) > 0 \end{cases}$$

So, using the hint, we conclude that the hessian matrix is negative definite not only for $\hat{\alpha}, \hat{\beta}$, but for any values of α, β.