EQ2341 – Tutorial 3

Recall:

- Parametrized model $f_{\boldsymbol{X}|S}(\boldsymbol{x} \mid s) = \widehat{f}_j(\boldsymbol{x} \mid \boldsymbol{\theta}_j)$
- Log-likelihood function $\mathcal{L}(\boldsymbol{\theta} \mid \mathcal{D}_j) = \sum_{n=1}^{N_j} \ln \widehat{f}_j(\boldsymbol{x}_{j,n} \mid \boldsymbol{\theta})$
- ML parameter estimate $\widehat{\boldsymbol{\theta}}_{\text{ML}}(\mathcal{D}_j) = \operatorname{argmax}_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta} \mid \mathcal{D}_j)$
- MAP parameter estimate $\widehat{\boldsymbol{\theta}}_{\text{MAP}}(\mathcal{D}_j) = \operatorname{argmax}_{\boldsymbol{\Theta}} \widehat{f}_{\boldsymbol{\Theta}|\mathcal{D}}(\boldsymbol{\theta} \mid \mathcal{D}_j)$

4.1 (MLE for the Exponential Distribution) A common problem in engineering is to assess the reliability of a system, given statistical data on how long the different components tend to last before they fail. It is then important to determine when and how often each component can be expected to break down. Let the data

$$\mathcal{D} = (x_1, ..., x_N)$$

be a set of measured life-lengths for different examples of a certain component. This data is by nature nonnegative and continuous-valued. A very simple model of such life-length data is the *exponential distribution*

$$f_X(x|\mu) = \frac{1}{\mu}\exp(-\frac{x}{\mu})$$

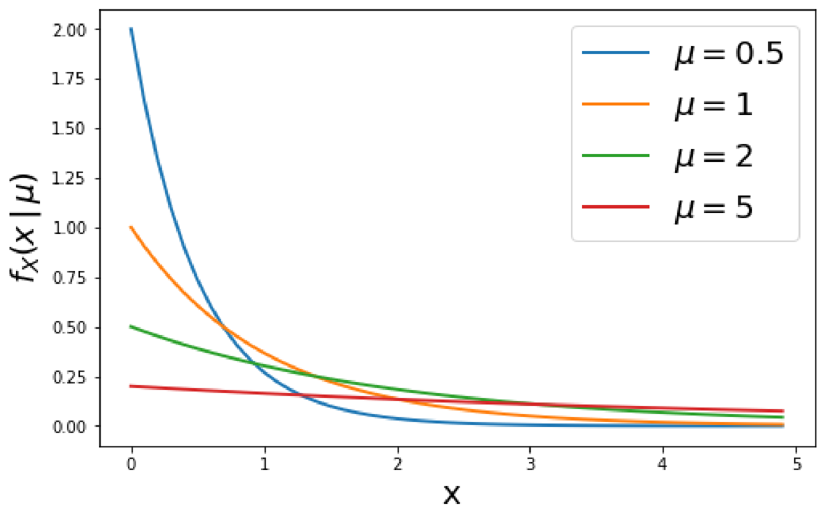
for $x \geq 0$, where $\mu > 0$ is a life-length parameter. The exponential distribution has the unique property that it is memoryless. This means that new and old components have the same probability of failing with in the same time span, and there is no good way to guess when a part will fail based on its age. This is not always a good representation of reality, but can be a reasonable first guess for light bulbs and other components that break suddenly and unpredictably.

4.1.a Assume that all samples are independent. Write down the probability of the dataset $\mathcal{D}$ for a given parameter value μ.

4.1.b Find the maximum-likelihood parameter estimate $\widehat{\mu}_{\mathrm{ML}}$ as a function of the data.

4.1.c A certain type of light bulb has an exponentially distributed lifetime. Your other light bulbs of this kind have lasted 35050, 15450, and 15800 hours. What is the maximum-likelihood parameter estimate for this data?

Bonus Prove that the exponential distribution is memoryless. Hint: think about the probability that a light bulb lasts beyond a certain time t.



Note that this is only a **model** of the life-length! Ideally, the notation should have been $\widehat{f}_X(x_i|\mu)$ to distinguish it from the 'true' life-length distribution.

Sol. 4.1.a Since the samples are *independent and identically distributed*, their probability of dataset is

$$P(\mathcal{D}|\mu) = \prod_{i=1}^{N} \widehat{f}_X(x_i|\mu)$$
$$= \prod_{i=1}^{N} \frac{1}{\mu} e^{-\frac{x_i}{\mu}}$$

Sol. 4.1.b The log-likelihood is given by

$$\begin{aligned}
\mathcal{L}(\mu \,|\, \mathcal{D}) &= \sum_{i=1}^{N} \ln \widehat{f}_X(x_i|\mu) \\
&= \sum_{i=1}^{N} \ln \frac{1}{\mu}\, e^{-\frac{x_i}{\mu}} \\
&= \sum_{i=1}^{N} \ln \frac{1}{\mu} + \sum_{i=1}^{N} \ln e^{-\frac{x_i}{\mu}} \\
&= -N \ln \mu - \frac{1}{\mu} \sum_{i=1}^{N} x_i.
\end{aligned}$$

The *maximum likelihood* estimate of μ is the value of μ that maximizes the likelihood $\mathcal{L}$. For this, we set the first derivative of $\mathcal{L}$ with respect to μ to zero.

$$\begin{aligned}
& \frac{\partial \mathcal{L}}{\partial \mu} = 0 \\
\Longrightarrow & \left(-\frac{N}{\mu} + \frac{1}{\mu^2} \sum_{i=1}^{N} x_i \right)_{\mu=\mu_{\mathrm{ML}}} = 0 \\
\Longrightarrow & \mu_{\mathrm{ML}} = \frac{\sum_i^N x_i}{N}.
\end{aligned}$$

That is, the ML life-length is equal to the sample mean.

Sol. 4.1.c $\mu_{\mathrm{ML}} = \frac{35050+15450+15800}{3} = 22100.$

Sol. Bonus The probability that a light bulb survives *beyond* time t is

$$\begin{aligned}
P(X > t) &= 1 - \int_0^t \frac{1}{\mu} e^{-\frac{x}{\mu}}\, dx \\
&= 1 - (1 - e^{-\frac{t}{\mu}}) \\
&= e^{-\frac{t}{\mu}}
\end{aligned}$$

Given that a bulb has survived byond time t, the probability that it will also survive beyond $t + \Delta t$ is

$$\begin{aligned} P(X > t + \Delta t | X > t) &= \frac{P(X > t + \Delta t)}{P(X > t)} \\ &= \frac{e^{-\frac{t+\Delta t}{\mu}}}{e^{-\frac{t}{\mu}}} \\ &= e^{-\frac{\Delta t}{\mu}}, \end{aligned}$$

that is, independent of the time elapsed so far.

4.3 (Doping Testing) In this problem we will look a bit deeper at the doping test scenario in example 4.1 on page 62. We will also look at what happens when the model is wrong. Read the example first, then answer the following questions:

4.3.a Assume that the a-priori probability that a tested individual is under the influence of a certain forbidden substance is $p_{cheat} = 0.9$. Also assume that the individual tests give the right indication 90% of the time, so that $p_{k|clean} = 0.1$ and $p_{k|cheat} = 0.9$, and that $k = 2$ tests are made. What is the probability that a tested athlete is influenced by the substance if both tests are positive? What if only one test is positive? What about the case when both test come up clean?

4.3.b In practice, many tests are correlated, and tend to make mistakes on the same examples. Drug tests to indicate heroin use, for instance, typically also react to other, innocent substances, such as poppy seeds commonly used in baking bread. To get an impression of how such corre- lations can affect the conclusions, we consider a second model, in which the two tests are totally correlated, so that the second test always gives the same result as the first. The first test is still assumed to be accurate 90% of the time.

What is the probability in this new model that an athlete is influenced by the substance if both tests are positive? What if both tests are clean? (Note that we cannot get only one positive result in this case, since the tests never disagree.)

4.3.c Compare your answer to the previous two problems. What would be the practical consequences of using the wrong model, acting as if the tests are independent as in the first case, when in reality they are not? How can we notice or avoid these issues in practice?

Sol 4.3.a **If the athlete cheated**, the conditional probability of a test turning up positive is

$$P(x_k \,|\, p_{\text{cheat}}) = (1 - p_{k\,|\,\text{cheat}})^{1-x_k} (p_{k\,|\,\text{cheat}})^{x_k},$$

that is,

$$\begin{aligned} P(x_k = 1 \,|\, p_{\text{cheat}}) &= p_{k\,|\,\text{cheat}}, \\ P(x_k = 0 \,|\, p_{\text{cheat}}) &= 1 - p_{k\,|\,\text{cheat}} \end{aligned}$$

where k is the index of the test. Similarly, **if the athlete is clean**,

$$P(x_k \,|\, p_{\text{clean}}) = (1 - p_{k\,|\,\text{clean}})^{1-x_k} (p_{k\,|\,\text{clean}})^{x_k}$$

We can calculate the marginal probability for the test outcomes, for example,

$$\begin{aligned} P([1,1]) &= P([1,1] \,|\, p_{\text{cheat}}) \times p_{\text{cheat}} + P([1,1] \,|\, p_{\text{clean}}) \times p_{\text{clean}} \\ &= 0.9 \times 0.9 \times 0.9 + 0.1 \times 0.1 \times 0.1 \\ &= 0.730 \end{aligned}$$

For $x_1 = 1$, $x_2 = 1$, if the tests are independent,

$$\begin{aligned} P(p_{\text{cheat}} \,|\, [1,1]) &= \frac{P([1,1] \,|\, p_{\text{cheat}}) \times p_{\text{cheat}}}{P([1,1] \,|\, p_{\text{cheat}}) \times p_{\text{cheat}} + P([1,1] \,|\, p_{\text{clean}}) \times p_{\text{clean}}} \\ &= \frac{0.9 \times 0.9 \times 0.9}{0.9 \times 0.9 \times 0.9 + 0.1 \times 0.1 \times 0.1} \\ &> 99\%. \end{aligned}$$

Similarly, for $x_1 = 1$, $x_2 = 0$,

$$\begin{aligned} P(p_{\text{cheat}} \,|\, [1,0]) &= \frac{P([1,0] \,|\, p_{\text{cheat}}) \times p_{\text{cheat}}}{P([1,0] \,|\, p_{\text{cheat}}) \times p_{\text{cheat}} + P([1,0] \,|\, p_{\text{clean}}) \times p_{\text{clean}}} \\ &= \frac{0.9 \times 0.1 \times 0.9}{0.9 \times 0.1 \times 0.9 + 0.1 \times 0.9 \times 0.1} \\ &\approx 90\%. \end{aligned}$$

Finally, for $x_1 = 0$, $x_2 = 0$,

$$\begin{aligned} P(p_{\text{cheat}} \,|\, [0,0]) &= \frac{P([0,0] \,|\, p_{\text{cheat}}) \times p_{\text{cheat}}}{P([0,0] \,|\, p_{\text{cheat}}) \times p_{\text{cheat}} + P([0,0] \,|\, p_{\text{clean}}) \times p_{\text{clean}}} \\ &= \frac{0.1 \times 0.1 \times 0.9}{0.1 \times 0.1 \times 0.9 + 0.9 \times 0.9 \times 0.1} \\ &= 1\%. \end{aligned}$$

Sol 4.3.b Since the second test is correlated,

$$\begin{aligned}
P([1,1]) &= P([1,1] \mid p_{\text{cheat}}) \times p_{\text{cheat}} + P([1,1] \mid p_{\text{clean}}) \times p_{\text{clean}} \\
&= 0.9 \times 0.9 + 0.1 \times 0.1 \\
&= 0.82 \\
P([0,0]) &= P([0,0] \mid p_{\text{cheat}}) \times p_{\text{cheat}} + P([0,0] \mid p_{\text{clean}}) \times p_{\text{clean}} \\
&= 0.1 \times 0.9 + 0.9 \times 0.1 \\
&= 0.18
\end{aligned}$$

Using the two values above, we get the new accuracy of the tests

$$P(p_{\text{cheat}} \mid [1,1]) = \frac{0.9 \times 0.9}{0.82} \approx 99\%.$$

$$P(p_{\text{cheat}} \mid [0,0]) = \frac{0.1 \times 0.9}{0.18} \approx 50\%.$$

Sol 4.3.c Discuss.