Gabriel Buginga

Machine Learning
CPS 863
Lista 4

October 28, 2019

Date Performed:	October 28, 2019
Institution:	PESC/COPPE/UFRJ
Instructors:	Edmundo de Souza e Silva
	Rosa M. M. Leão
	Daniel Sadoc Menasché

Problem 1

$\mathcal{D} = \{x_1, x_2, \ldots, x_N\}$, with $n_H = 4$, $N = 5$ and $P(heads) = p$.

1,2: Then we have:

$$\begin{aligned}\mathcal{L}(p|\mathcal{D}) &= p(\mathcal{D}/p) \\ &= p^{n_H} \cdot (1-p)^{N-n_H} \\ &= p^4 \cdot (1-p)\end{aligned}$$

3: Aiming for $log\, p(\mathcal{D}|p) = (N - N_H)log(1-p) + N_H log(p)$:

$$\begin{aligned}\frac{\partial log\, p(\mathcal{D}|p)}{\partial p} = 0 \Rightarrow & (N - N_H)\frac{-1}{1-p} + N_H\frac{1}{p} = 0 \\ & \frac{N_H}{p} = \frac{N - N_H}{1-p} \\ & p = \frac{N_H}{N} \\ & p = \frac{4}{5} = 0.8\end{aligned}$$

4: The previous derivation already have the formula:

$$\begin{aligned}\hat{p}^{mle} &= \underset{p}{\text{argmax}} p(\mathcal{D}|p) \\ &= \underset{p}{\text{argmax}} log\, p(\mathcal{D}|p) = (N - N_H)log(1-p) + N_H log(p) \\ &= \frac{N_H}{N}\end{aligned}$$

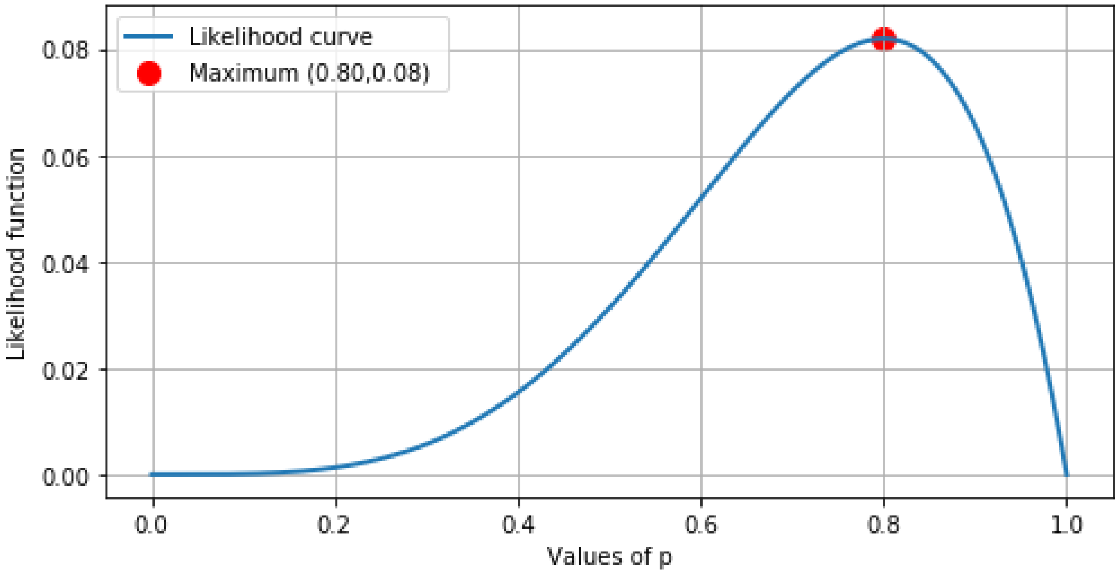

Figure 1: Likelihood function $\mathcal{L}(p|\mathcal{D})$.

Problem 2

1:$\mathcal{D} = \{x_1, x_2, \dots, x_N\}$ for a exponential distribution with mean $\frac{1}{\mu}$. Then we directly have for the likelihood:

$$\begin{aligned}\mathcal{L}(\mu|\mathcal{D}) &= p(\mathcal{D}/\mu) \\ &= \prod_{i=1}^{n} \mu \exp\left(-\mu x_i\right) \\ &= \mu^n \exp\left(-\mu \sum_{i=1}^{n} x_i\right)\end{aligned}$$

2,3:For the MLE, and using the dataset *data-l4-p-1.txt*:

$$\begin{aligned}\frac{\partial log\, p(\mathcal{D}|\mu)}{\partial \mu} = 0 \Rightarrow & \frac{\partial\left(n\, log\mu - \mu \sum_{i=1}^{n} x_i\right)}{\partial \mu} = 0 \\ & \frac{n}{\mu} - \sum_{i=1}^{n} x_i = 0 \\ & \mu = \frac{n}{\sum_{i=1}^{n} x_i}\end{aligned}$$

$$\begin{aligned}\hat{\mu}^{mle} &= \underset{\mu}{\operatorname{argmax}} p(\mathcal{D}|\mu) \\ &= \underset{p}{\operatorname{argmax}} log\, p(\mathcal{D}|p) \\ &= \frac{n}{\sum_{i=1}^{n} x_i} \\ &= 0.0969515\end{aligned}$$

Problem 3

1:$\mathcal{D} = \{x_1, x_2, \ldots, x_N\}$ from $\mathcal{N}\left(\mu, \sigma^2\right)$. Now $\theta = \left(\mu, \sigma^2\right)$, with two parameters to optimize. The likelihood:

$$\begin{aligned}
\mathcal{L}(\left(\mu, \sigma^2\right)|\mathcal{D}) &= p(\mathcal{D}/\left(\mu, \sigma^2\right)) \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right)
\end{aligned}$$

2:For the MLE, first we optimize then we plug it, obtaining the answer from [1]:

$$\begin{aligned}
\frac{\partial log\, p\left(\mathcal{D}/\left(\mu,\sigma^2\right)\right)}{\partial \mu} = 0 \Rightarrow & \frac{\partial\left(-\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2\right)}{\partial\mu} = 0 \\
& -\frac{\sum_{i=1}^{n}\left(x_i-\mu\right)}{\sigma^2} = 0 \\
& \hat{\mu}^{mle} = \frac{1}{n}\sum_{i=1}^{n} x_i \\
\Rightarrow & \frac{\partial\left(-\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2\right)}{\partial\sigma^2} = 0 \\
& -\frac{n}{2\sigma^2} + \frac{2\sum_{i=1}^{n}\left(x_i-\mu\right)^2}{\left(2\sigma^2\right)^2} = 0 \\
& \hat{\sigma^2}^{mle} = \frac{\sum_{i=1}^{n}\left(x_i-\hat{\mu}^{mle}\right)^2}{n}
\end{aligned}$$

Problem 4

1,2:For $\mathcal{D}_1 = data-l4-p-2a.txt$ and $\mathcal{D}_2 = data-l4-p-2b.txt$, we can rightly calculate the MLE using Problem's 3 results and Lista 3 for the uniform case:

$$\begin{aligned}
D_1 \Rightarrow \hat{\mu}^{mle} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\
\hat{\sigma^2}^{mle} &= \frac{\sum_{i=1}^{n}\left(x_i-\hat{\mu}^{mle}\right)^2}{n} \\
\hat{\mu}^{mle} &= 12.9857 \quad \hat{\sigma^2}^{mle} = 27.9379 \\
log_{10}\,\mathcal{L}(\left(\hat{\mu}^{mle}, \hat{\sigma^2}^{mle}\right)|\mathcal{D}_1) &= -308.3931 \\
D_1 \Rightarrow \hat{a}^{mle} &= min\{x_1, x_2, \cdots, x_n\} \\
\hat{b}^{mle} &= max\{x_1, x_2, \cdots, x_n\} \\
\hat{a}^{mle} &= 1.3182 \quad \hat{b}^{mle} = 29.6183 \\
log\,\mathcal{L}(\left(\hat{a}^{mle}, \hat{b}^{mle}\right)|\mathcal{D}_1) &= -334.2867
\end{aligned}$$

$$D_2 \Rightarrow \hat{\mu}^{mle} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\hat{\sigma^2}^{mle} = \frac{\sum_{i=1}^{n}\left(x_i - \hat{\mu}^{mle}\right)^2}{n}$$

$$\hat{\mu}^{mle} = 10.7887 \quad \hat{\sigma^2}^{mle} = 32.7648$$

$$log\, \mathcal{L}(\left(\hat{\mu}^{mle}, \hat{\sigma^2}^{mle}\right)|\mathcal{D}_2) = -316.3616$$

$$D_2 \Rightarrow \hat{a}^{mle} = min\{x_1, x_2, \cdots, x_n\}$$

$$\hat{b}^{mle} = max\{x_1, x_2, \cdots, x_n\}$$

$$\hat{a}^{mle} = 1.0552 \quad \hat{b}^{mle} = 19.5786$$

$$log\, \mathcal{L}(\left(\hat{a}^{mle}, \hat{b}^{mle}\right)|\mathcal{D}_2) = -291.9034$$

3: Using the likelihood test for each dataset:

$$\frac{\mathcal{L}(\left(\hat{\mu}^{mle}, \hat{\sigma^2}^{mle}\right)|\mathcal{D}_1)}{\mathcal{L}(\left(\hat{a}^{mle}, \hat{b}^{mle}\right)|\mathcal{D}_1)} = 1.759827 \cdot 10^{11}$$

$$\frac{\mathcal{L}(\left(\hat{\mu}^{mle}, \hat{\sigma^2}^{mle}\right)|\mathcal{D}_2)}{\mathcal{L}(\left(\hat{a}^{mle}, \hat{b}^{mle}\right)|\mathcal{D}_2)} = 2.387483 \cdot 10^{-11}$$

Then for D_1 we can infer that it came from a normal distribution, conversely for D_2 it came from the uniform distribution.

Problem 5

1: We can readily generate the following figure:

2: For $\Theta_0 = \{\theta_0\}$:

$$\begin{aligned}\max_{\theta\in\Theta_0}\{\mathcal{L}(\theta|\mathcal{D})\} &= p(\mathcal{D}/\theta_0) \\ &= \prod_{i=1}^{n}\theta_0 \exp\left(-\theta_0 x_i\right) \\ &= \theta_0^n \exp\left(-\theta_0\sum_{i=1}^{n} x_i\right) \\ &= \theta_0^n \exp\left(-\theta_0 \cdot n \cdot a\right)\end{aligned}$$

3:For $\Theta_1 = [\theta_0, \infty)$, referring to figure 2 as we have the peak at $1/a$, the denominator changes its value depending the this maximum lies in relation to θ_0:

$$\begin{aligned}if \quad \theta_0 \leqslant 1/a \Rightarrow \max_{\theta\in\Theta_1}\{\mathcal{L}(\theta|\mathcal{D})\} &= p\left(\mathcal{D}/\frac{1}{a}\right) \\ &= \frac{1}{a^n}\exp\left(-n\right) \\ if \quad \theta_0 > 1/a \Rightarrow \max_{\theta\in\Theta_1}\{\mathcal{L}(\theta|\mathcal{D})\} &= p\left(\mathcal{D}/\theta_0\right) \\ &= \theta_0^n \exp\left(-\theta_0 \cdot n \cdot a\right)\end{aligned}$$

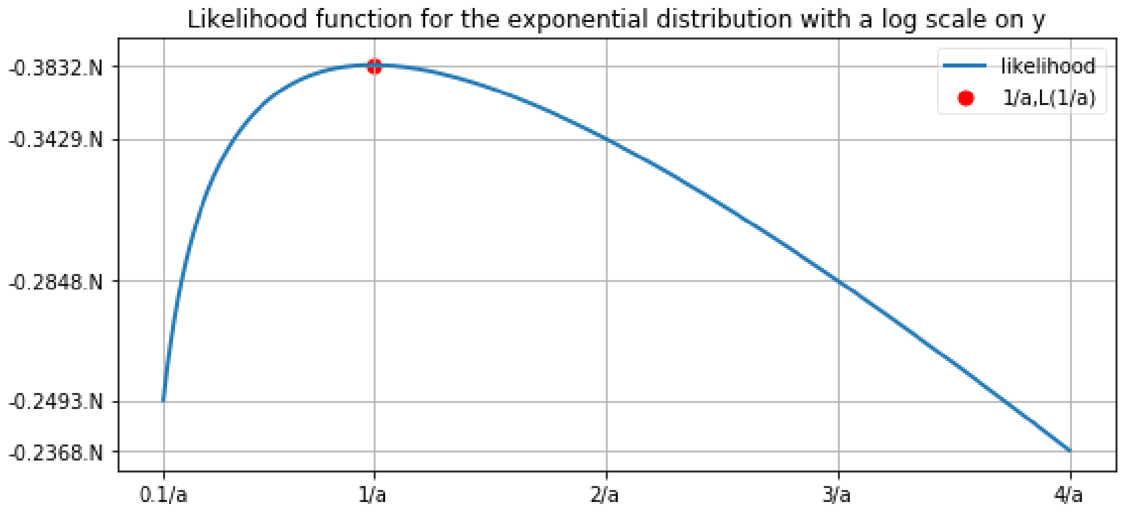

Figure 2: Likelihood function $\mathcal{L}(\mu|\mathcal{D})$.

4:Using what we already calculated, and knowing that a has all the information for calculating the likelihood ratio test, it's our sufficient statistic:

$$
\begin{aligned}
if \quad \theta_0 \leqslant 1/a \Rightarrow \Lambda(\mathbf{x}) = \frac{\max_{\theta\in\Theta_0}\{\mathcal{L}(\theta|\mathcal{D})\}}{\max_{\theta\in\Theta_1}\{\mathcal{L}(\theta|\mathcal{D})\}} &= \frac{\theta_0^n \exp\left(-\theta_0 \cdot n \cdot a\right)}{\frac{1}{a^n}\exp\left(-n\right)} \\
&= (a\theta_0)^n \exp\left(n(1-a\theta_0)\right) \\
if \quad \theta_0 > 1/a \Rightarrow \Lambda(\mathbf{x}) = \frac{\max_{\theta\in\Theta_0}\{\mathcal{L}(\theta|\mathcal{D})\}}{\max_{\theta\in\Theta_1}\{\mathcal{L}(\theta|\mathcal{D})\}} &= \frac{\theta_0^n \exp\left(-\theta_0 \cdot n \cdot a\right)}{\theta_0^n \exp\left(-\theta_0 \cdot n \cdot a\right)} \\
&= 1
\end{aligned}
$$

Problem 6

We have $E[X], E[Y], \mathrm{Var}(X), \mathrm{Var}(Y), \mathrm{Cov}(X,Y)$, $f(X) = aX + b$ and want to minimize $E[D^2] = E[(Y - f(X))^2]$:

$$
\begin{aligned}
E[(Y-f(X))^2] &= E[(Y-aX-b)^2] \\
&= E[Y^2 + a^2X^2 + b^2 - 2bY - 2aXY + 2abX] \\
\frac{\partial E[(Y-f(X))^2]}{\partial a} = 0 \Rightarrow &- 2E[XY] + 2bE[X] + 2aE[X^2] = 0 \\
a &= \frac{E[XY] - bE[X]}{E[X^2]} \\
a &= \frac{\mathrm{Cov}(X,Y) + E[X](E[Y]-b)}{\mathrm{Var}(X) + E[X]^2} \\
\frac{\partial E[(Y-f(X))^2]}{\partial b} = 0 \Rightarrow &- 2E[Y] + 2aE[X] + 2b = 0 \\
b &= E[Y] - aE[X]
\end{aligned}
$$

References

[1] K. P. Murphy, *Machine learning: a probabilistic perspective.* MIT press, 2012.