Gabriel Buginga

Machine Learning
CPS 863
Lista 3

October 16, 2019

Date Performed:	October 16, 2019
Institution:	PESC/COPPE/UFRJ
Instructors:	Edmundo de Souza e Silva
	Rosa M. M. Leão
	Daniel Sadoc Menasché

Problem 1

1: We have $\mathcal{D} = \{3, 3, 4, 2, 3\}$, and we want its likelihood for the $h = U(1, 5)$, considering i.i.d:

$$\begin{aligned} p(\mathcal{D}|h) &= \left(\frac{1}{5}\right)^5 \\ &= 0.00032. \end{aligned}$$

2: Directly from above $logp(\mathcal{D}|h) = -8.04718$.

Problem 2

1:$h = U(1, a)$ and $\mathcal{D} = \{3, 3, 4, 2, 3\}$, we want $p(\mathcal{D}|h)$. For that we need to be careful about if some data point is in fact covered by h, so we need $\mathbb{I}(x_i \in h)$ which is 1 if what is inside is true, and 0 otherwise:

$$p(\mathcal{D}|h) = \mathbb{I}(2 \in h)\left(\frac{1}{a-1}\right) \cdot \mathbb{I}(3 \in h)\left(\frac{1}{a-1}\right)^3 \cdot \mathbb{I}(4 \in h)\left(\frac{1}{a-1}\right).$$

Problem 3

1:We want $p(\mathcal{D}|h)$, $\mathcal{D} = \{TTHTHHTHHH\}$ with $p = 0.2$ or $p = 0.6$:

$$
\begin{aligned}
p(\mathcal{D}|h) &= (1-p)^4 \cdot (p)^6 \\
[p = 0.2] &= 2.62144 \cdot 10^{-5} \\
[p = 0.6] &= 1.19439 \cdot 10^{-3}
\end{aligned}
$$

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

$$
\begin{aligned}
\frac{\partial log\, p(\mathcal{D}|h)}{\partial p} = 0 \Rightarrow (N - N_H)\frac{-1}{1-p} + N_H \frac{1}{p} &= 0 \\
\frac{N_H}{p} &= \frac{N - N_H}{1-p} \\
p &= \frac{N_H}{N} \\
p &= \frac{6}{10} = 0.6
\end{aligned}
$$

Problem 4

1:For $\mathcal{D} = \{3, 3, 4, 2, 3\}$, $p(\mathcal{D}|dado) = \left(\frac{1}{6}\right)^5 = 0.0001286$ and $p(\mathcal{D}|U(1,5) = \left(\frac{1}{5}\right)^5 = 0.00032$.

2:We have the following posteriors:

$$
\begin{aligned}
p(dado|\mathcal{D}) &= \frac{p(\mathcal{D}|dado) \cdot p(dado)}{p(\mathcal{D}|dado) \cdot p(dado) + p(\mathcal{D}|U(1,5)) \cdot p(U(1,5))} \\
&= \frac{\left(\frac{1}{5}\right)^5 \cdot 0.2}{\left(\frac{1}{6}\right)^5 \cdot 0.8 + \left(\frac{1}{5}\right)^5 \cdot 0.2} \\
&= 0.61649 \\
p(U(1,5)|\mathcal{D}) &= \frac{p(\mathcal{D}|U(1,5)) \cdot p(U(1,5))}{p(\mathcal{D}|dado) \cdot p(dado) + p(\mathcal{D}|U(1,5)) \cdot p(U(1,5))} \\
&= \frac{\left(\frac{1}{6}\right)^5 \cdot 0.8}{\left(\frac{1}{6}\right)^5 \cdot 0.8 + \left(\frac{1}{5}\right)^5 \cdot 0.2} \\
&= 0.38350
\end{aligned}
$$

3,4: The probability for seeing $\tilde{x} = 5$ and $\tilde{x} = 6$, we weight across hypothesis and posteriors[1]

$$
\begin{aligned}
p(\tilde{x} = 5|\mathcal{D}) &= p(\tilde{x} = 5|dado) \cdot p(dado|\mathcal{D}) + p(\tilde{x} = 5|U(1,5)) \cdot p(U(1,5)|\mathcal{D}) \\
&= \frac{1}{6} \cdot 0.61649 + \frac{1}{5} \cdot 0.38350 \\
&= 0.17945 \\
p(\tilde{x} = 6|\mathcal{D}) &= p(\tilde{x} = 6|dado) \cdot p(dado|\mathcal{D}) + p(\tilde{x} = 6|U(1,5)) \cdot p(U(1,5)|\mathcal{D}) \\
&= \frac{1}{6} \cdot 0.61649 + 0 \cdot 0.38350 \\
&= 0.10274
\end{aligned}
$$

5,6: We're doing MLE and MAP for the parameter p, using the uniform prior for :

$$
\begin{aligned}
\hat{p}^{MLE} = \underset{p}{\operatorname{argmax}}\, p(\mathcal{D}|p) &= \underset{p}{\operatorname{argmax}}\, p(\mathcal{D}|dado) \cdot p(dado|p) + p(\mathcal{D}|U(1,5)) \cdot p(U(1,5)|p) \\
&= \underset{p}{\operatorname{argmax}} \left(\frac{1}{6}\right)^5 \cdot p + \left(\frac{1}{5}\right)^5 \cdot (1-p) \\
&= 0 \\
\hat{p}^{MAP} = \underset{p}{\operatorname{argmax}}\, p(\mathcal{D}|p) \cdot p(p) &= \underset{p}{\operatorname{argmax}}\, p(\mathcal{D}|dado) \cdot p(dado|p) \cdot p(p) + p(\mathcal{D}|U(1,5)) \cdot p(U(1,5)|p) \cdot p(p) \\
&= \underset{p}{\operatorname{argmax}} \left(\frac{1}{6}\right)^5 \cdot p \cdot \frac{1}{p} + \left(\frac{1}{5}\right)^5 \cdot (1-p) \cdot \frac{1}{p} \\
&= 1
\end{aligned}
$$

7: All posteriors change because they're clearly a function of p, so all the results of **2,3,4** change. However, **1** don't because we marginalize directly within the hypothesis, and **5,6** notoriously don't change by definition - as we used the uniform prior.

Problem 5

1: $\mathcal{D} = \{3, 3, 4, 2, 3\}$ and now they're i.i.d and we need to construct a product:

$$
\begin{aligned}
p(\mathcal{D}|p) = \prod_{x_i \in \mathcal{D}} (p(x_i|U(1,5)) \cdot p(U(1,5)) + p(x_i|dado) \cdot p(dado)) &= \left(\frac{1}{5} \cdot (1-p) + \frac{1}{6} \cdot p\right)^5 \\
&= \left(\frac{1}{5} \cdot (1-0.7) + \frac{1}{6} \cdot 0.7\right)^5 \\
&= 0.000172
\end{aligned}
$$

2,3: As $p = 0.7$, it has a fixed value, a delta function at 0.7 so even seeing data it doesn't change our beliefs about it:

$$
\begin{aligned}
p(\tilde{x} = 5|\mathcal{D}) &= p(\tilde{x} = 5|dado) \cdot p(dado|\mathcal{D}) + p(\tilde{x} = 5|U(1,5)) \cdot p(U(1,5)|\mathcal{D}) \\
&= \frac{1}{6} \cdot p + \frac{1}{5} \cdot (1-p) \\
&= 0.17666 \\
p(\tilde{x} = 6|\mathcal{D}) &= p(\tilde{x} = 6|dado) \cdot p(dado|\mathcal{D}) + p(\tilde{x} = 6|U(1,5)) \cdot p(U(1,5)|\mathcal{D}) \\
&= \frac{1}{6} \cdot p + 0 \cdot (1-p) \\
&= 0.11666
\end{aligned}
$$

4: From *dado*: $\frac{(\frac{1}{6} \cdot 0.7)^5}{0.00172} = 0.12559$; from $U(1,5)$: $\frac{(\frac{1}{5} \cdot 0.3)^5}{0.00172} = 0.00451$

5:Remembering from **1**:

$$
\begin{aligned}
\hat{p}^{MLE} = \underset{p}{\operatorname{argmax}}\, p(\mathcal{D}|p) &= \left(\frac{1}{5} \cdot (1-p) + \frac{1}{6} \cdot p\right)^5 \\
\hat{p}^{MLE} = \underset{p}{\operatorname{argmax}}\, log\, p(\mathcal{D}|p) &= 5 \cdot log\left(\frac{1-p}{5} + \frac{p}{6}\right) \\
&= 0
\end{aligned}
$$

6:For this new dataset:

$$\hat{p}^{MLE} = \underset{p}{\text{argmax}}\, p(\mathcal{D}|p) = \left(\frac{1}{5}\cdot(1-p) + \frac{1}{6}\cdot p\right)^{19}\cdot\left(\frac{1}{6}\cdot p\right)$$
$$\hat{p}^{MLE} = \underset{p}{\text{argmax}}\, log\, p(\mathcal{D}|p) = 19\cdot log\left(\frac{1-p}{5}+\frac{p}{6}\right) + log\frac{p}{6}$$
$$= 0.30372$$

Problem 6

1: Our likelihood:

$$\hat{a}^{MLE} = \prod_{x_i\in\mathcal{D}} \frac{\mathbb{I}(x\in[-a,a])}{2a} = \left(\frac{\mathbb{I}(x\in[-a,a])}{2a}\right)^n$$
$$= 0\; if\; a < \max\{X_1,\ldots,X_n\}\; and\; \left(\frac{1}{2a}\right)^n,\; otherwise.$$

2: Rigorously we have the relation below, the problem that we don't have the prior to calculate it:

$$p(V=v|\mathcal{D}) = \int p(v|a)p(a|\mathcal{D})d\boldsymbol{a}$$
$$= \int p(v|a)\left(\frac{p(\mathcal{D}|a)p(a)}{\int p(a|\mathcal{D})p(a)da}\right)d\boldsymbol{a}$$

Problem 7

1: $\mathcal{D}$ = TTTTHHHHTH. For a uniform prior on the coin chosen, our $\hat{p}^{MLE} = 0.5$ which can induce us to choose the hypothesis of the fair coin.

2:Our estimate stays exactly the same because $\hat{p}^{MLE} = 0.5$ even with the new priors.

3:$p(\mathcal{D}|h_{biased}) = p^5\cdot(1-p)^5$ and $p(\mathcal{D}|h_{biased}) = (\frac{1}{2})^{10}$.

4:For that:

$$p(h_{biased}|\mathcal{D}) = \frac{p(\mathcal{D}|h_{biased})p(h_{biased})}{p(\mathcal{D}|h_{biased})p(h_{biased}) + p(\mathcal{D}|h_{fair})p(h_{fair})}$$
$$= \frac{p^5(1-p)^5 0.7}{p^5(1-p)^5 0.7 + 0.5^{10}0.3}$$
$$p(h_{fair}|\mathcal{D}) = \frac{p(\mathcal{D}|h_{fair})p(h_{fair})}{p(\mathcal{D}|h_{biased})p(h_{biased}) + p(\mathcal{D}|h_{fair})p(h_{fair})}$$
$$= \frac{0.5^{10}0.3}{p^5(1-p)^5 0.7 + 0.5^{10}0.3}$$

References

[1] A. N. Shiryaev, "Probability-1," 2016.