Homework 3

2021-31645-T1 Machine Learning
Emmanouil Palaiologos

December 8, 2021

Bayesian Inference

MacKay, D. J., Mac Kay, D. J. (2003). Information theory, inference and learning algorithms. Cambridge university press.

Example 2.7 & Exercise 2.8

Taking advantage of **Bayesian Theorem** we can express the **posterior** probability of f_H given n_H as follows:

$$P(f_H|n_H, N) = \frac{P(n_H|f_H, N) \cdot P(f_H)}{P(n_H, N)} \tag{1}$$

Now we are able to break-down the problem :

- For the **likelihood**, we use the Naive-Bayes *assumption* for **independent** data:

$$P(n_H|f_H, N) = \prod_{i=1}^{N} f_H^{n_i} \cdot (1 - f_H)^{(1-n_i)} = f_H^{n_H}(1 - f_H)^{(N-n_H)} \tag{2}$$

- Since we have no further information, our **prior** *belief* about f_H can be described by a **uniform** (*"flat"*) distribution $P(f_H) = 1$ as well.

- Calculating the normalizing factor (the **'evidence'**) $P(n_H, N)$, we observe that:

$$
\begin{aligned}
P(n_H, N) &= \int P(n_H, f_H, N)\, df_H && (3)\\
&= \int \underbrace{P(n_H|f_H, N)}_{likelihood}\, \underbrace{P(f_H)}_{prior}\, df_H && (4)\\
&= \int_0^1 f_H^{n_H}(1-f_H)^{(N-n_H)}\cdot 1\,\cdot df_H && (5)\\
&= \int_0^1 f_H^{n_H+1-1}(1-f_H)^{(N-n_H+1-1)}\, df_H && (6)\\
&= B(n_H+1, N-n_H+1) && (7)
\end{aligned}
$$

,where B is the **Beta** function and therefore:

$$
P(n_H, N) = \frac{n_H!\,(N-n_H)!}{(N+1)!} \tag{8}
$$

Inserting the results of eq.(8) and eq.(2) into eq.(1), we get:

$$
\boxed{P(f_H|n_H, N) = f_H^{n_H}(1-f_H)^{(N-n_H)}\cdot\frac{(N+1)!}{n_H!\,(N-n_H)!}} \tag{9}
$$

In order to perform the prediction for the $N+1$th outcome, we marginalize so that

$$
\begin{aligned}
P(heads|n_H, N) &= \int P(heads, f_H|n_H)\, df_H && (10)\\
&= \int P(heads|f_H)\, \underbrace{P(f_H|n_H)}_{posterior}\, df_H && (11)\\
&= \int f_H \cdot f_H^{n_H}(1-f_H)^{(N-n_H)}\cdot\frac{(N+1)!}{n_H!\,(N-n_H)!}\, df_H && (12)\\
&= \frac{(n_H+1)!\,\cancel{(N-n_H)!}}{(N+2)!}\cdot\frac{(N+1)!}{n_H!\,\cancel{(N-n_H)!}} && (13)\\
&= \frac{(n_H+1)\cancel{n_H!}}{(N+2)\cancel{(N+1)!}}\frac{\cancel{(N+1)!}}{\cancel{n_H!}} && (14)
\end{aligned}
$$

$$
\implies \boxed{P(heads|n_H, N) = \frac{n_H+1}{N+2}} \tag{15}
$$

The famous *Laplace's rule of succession* (eq.15) gives us the undermentioned results for different values of N, n_H:

1. $P(heads|n_H = 0, N = 3) = \frac{1}{5}$

2. $P(heads|n_H = 2, N = 3) = \frac{3}{5}$
3. $P(heads|n_H = 3, N = 10) = \frac{1}{3}$
4. $P(heads|n_H = 29, N = 300) = \frac{30}{302}$

Moreover, we should pay attention back to (eq.9) for the *posterior* probability, that is actually described by *Beta* distribution :

$$P(f_H|n_H, N) = \frac{1}{\frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}} \cdot f_H^{\alpha-1}(1 - f_H)^{\beta-1} \tag{16}$$

,where $\alpha = n_H + 1$ and $\beta = N - n_H + 1$.Through this scope we can rewrite the previous 4 cases as : $(\alpha = 1, \beta = 4), (\alpha = 3, \beta = 2), (\alpha = 4, \beta = 8) and (\alpha = 30, \beta = 272)$. In graph (1) the posterior distribution is plotted for these specific values. It is quite obvious that the curve tends to sharpen as N is rising. The more data we have, the more we *"believe"* the probabilities we calculate and therefore the prior slowly becomes less significant, enabling the whole distribution to be placed around the most probable value.

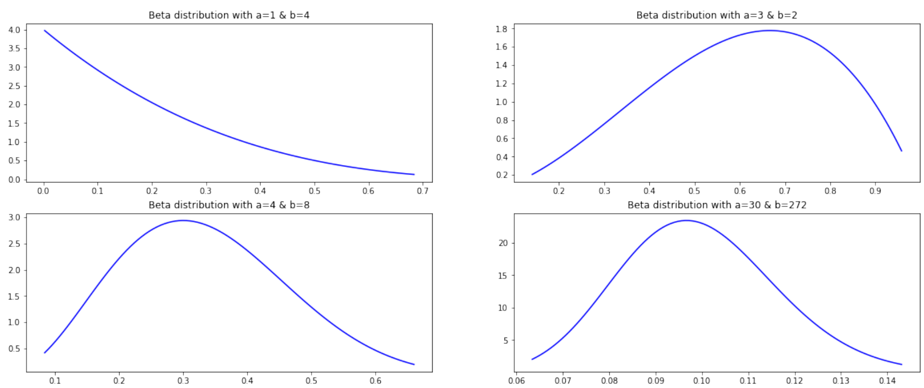
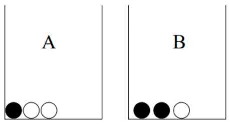


Figure 1: Posterior probability as beta distribution for different values of α and β.

Example 2.10

Since we have no extra information for a biased decision between two urns, it follows that :

$$P(A) = P(B) = \frac{1}{2}$$



The probability that the urn A is selected under the constraint that a black ball has been extracted, can be found using *Bayes Theorem*:

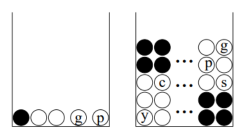
$$
\begin{aligned}
P(A|b = black) &= \frac{P(A, b = black)}{P(b = black)} && (17)\\
&= \frac{P(b = black|A)\, P(A)}{\sum_i P(b = black|URN_i)\, P(URN_i)} && (18)\\
&= \frac{P(b = black|A)\, P(A)}{P(b = black|A)P(A) + P(b = black|B)P(B)} && (19)\\
&= \frac{\frac{1}{3}\cdot\frac{1}{2}}{\frac{1}{3}\cdot\frac{1}{2} + \frac{2}{3}\cdot\frac{1}{2}} = \frac{1}{3} && (20)
\end{aligned}
$$

Example 2.11

The same idea applies here:

$$P(A) = P(B) = \frac{1}{2}$$

The result will be independent of the detailed contents of the urns (and all other possible outcomes) :



$$
\begin{aligned}
P(A|b = black) &= \frac{P(A, b = black)}{P(b = black)} && (21)\\
&= \frac{P(b = black|A)\, P(A)}{\sum_i P(b = black|URN_i)\, P(URN_i)} && (22)\\
&= \frac{P(b = black|A)\, P(A)}{P(b = black|A)P(A) + P(b = black|B)P(B)} && (23)\\
&= \frac{\frac{1}{5}\cdot\frac{1}{2}}{\frac{1}{5}\cdot\frac{1}{2} + \frac{2}{5}\cdot\frac{1}{2}} = \frac{1}{3} && (24)
\end{aligned}
$$

(as expected!)

Exercise 3.5

(α) $s = aba$ **and** $F = 3$

Taking under consideration Mackey's notes on the basic inference problem with the unknown parameter p_a, we write for the **posterior** probability:

$$P(p_a|s, F, \mathcal{H}_1) = \frac{p_a^{F_a}(1-p_a)^{F_b}(F_a + F_b + 1)!}{F_a!F_b!} \tag{25}$$

Since $F = 3$ and the given sequence is $s = aba$, then $F_a = 2$ and $F_b = 1$. We rewrite (25):

$$P(p_a|s = aba, F = 3, \mathcal{H}_1) = \frac{p_a^2(1-p_a)4!}{2!1!} \tag{26}$$
$$= 12\, p_a^2(1-p_a) \tag{27}$$

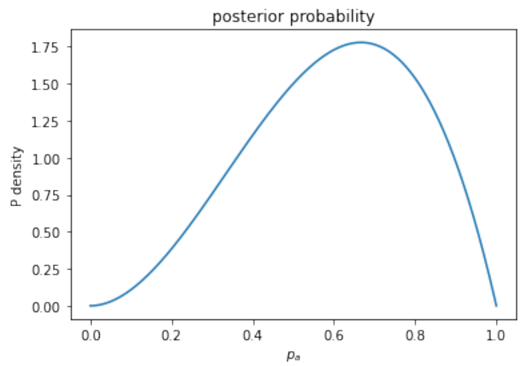


Figure 2: Graph of the posterior probability, with s=aba.

The value of p_a that maximizes $P(p_a|s = aba, F = 3, \mathcal{H}_1)$ is simply given by :

$$\hat{p_a} = \arg\max_{p_a \in [0,1]} 12\, p_a^2(1-p_a) \tag{28}$$

That is

$$\frac{d(12\, p_a^2(1-p_a))}{dp_a} = 0 \tag{29}$$

$$\implies 36p_a^2 - 24p_a = 0 \tag{30}$$

$$\implies \hat{p_a} = \arg\max_{p_a \in [0,1]} 12\, p_a^2(1-p_a) \tag{31}$$

$$\implies \boxed{\hat{p_a} = \frac{2}{3}} \tag{32}$$

As for the **mean value** of p_a under the given distribution ($Beta$), we can generalize :

$$\mathbb{E}[p_a] = \frac{\int_0^1 p_a \cdot P(p_a|s,F,\mathcal{H}_1)\, dp_a}{\underbrace{P(s|F,\mathcal{H}_1)}_{normalizer}} \tag{33}$$

$$= \frac{(F_a+F_b+1)!}{F_a!F_b!} \int_0^1 p_a^{F_a+1}(1-p_a)^{F_b}\, dp_a \tag{34}$$

$$= \frac{(F_a+F_b+1)!}{F_a!F_b!} \frac{\Gamma(F_a+2)\Gamma(F_b+1)}{\Gamma(F_a+F_b+3)} \tag{35}$$

$$= \frac{(F_a+F_b+1)!}{F_a!\cancel{F_b}!} \frac{(F_a+1)!\cancel{F_b}!}{(F_a+F_b+2)!} \tag{36}$$

$$= \frac{\cancel{(F_a+F_b+1)!}}{\cancel{F_a}!(F_a+F_b+2)} \frac{(F_a+1)\cancel{F_a}!}{\cancel{(F_a+F_b+1)!}} \tag{37}$$

$$\implies \boxed{\mathbb{E}[p_a] = \frac{(F_a+1)}{(F_a+F_b+2)}} \tag{38}$$

which is basically $\propto \frac{\alpha}{\alpha+\beta}$. If we set $F_a = 2$, $F_b = 1$, we get:

$$< p_a >= \mathbb{E}[p_a] = \frac{3}{5} \tag{39}$$

(β) $s = bbb$ **and** $F = 3$

Here we are faced with $F_a = 0$ and $F_b = 3$. Following similar reasoning, we can find :

- The *Maximum A Posteriori* estimate

$$\hat{p_a} = \arg\max_{p_a \in [0,1]} P(p_a|s=bbb, F=3, \mathcal{H}_1) \tag{40}$$

$$= \arg\max_{p_a \in [0,1]} 4(1-p_a)^3 \tag{41}$$

$$\implies \hat{p_a} = 0 \tag{42}$$

  which might seem extreme but since we have $F_a = 0$ it is only reasonable.

- The *mean value*, using (38)

$$\mathbb{E}[p_a] = \frac{(0+1)}{(0+3+2)} = \frac{1}{5} \tag{43}$$

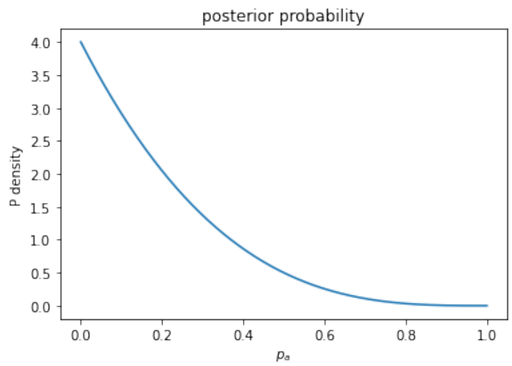


Figure 3: Graph of the posterior probability, with s=bbb.

Exercise 3.10

By hypothesis, we have evidence that **the first room belongs to a girl** and **at least one room belongs to a boy**

(α) <u>**'Sequence' of children= (g,g,b)**</u>

Out of all possible $2^3 = 8$ outcomes, we have to rule out all cases that are **not** conditioned to $P(b, g)$, where $P(b)$ and $P(g)$ are the probabilities of having *"at least one b"* and *"g as first"*, respectively. That means that we eliminate the cases (g, g, g) and $(b, b, b), (b, b, g), (b, g, b), (b, g, g)$from our sample set, so that we are left with 3 possible cases : $(g, g, b), (g, b, g)$ and (g, b, b). Out of those, only $n(ggb \cap b, g) = 1$ is complying with the desired outcome. Hence:

$$P(ggb|b, g) = \frac{P(ggb \cap b, g)}{P(b, g)} = \frac{1}{3} \tag{44}$$

which is not affected directly by the prior probability (i.e. the equal probabilities assumption $\frac{1}{8}$.

Of course we can acquire an identical result by applying directly the Bayes theorem, while noticing that $P(ggb \cap b, g) = P(b, g|ggb)P(ggb) = 1 \cdot \frac{1}{8}$ and $P(b, g) = P(b|g)P(g) = \frac{3}{4} \cdot \frac{4}{8} = \frac{3}{8}$. Thus:

$$(44) \implies \boxed{P(ggb|b, g) = \frac{\frac{1}{8}}{\frac{3}{8}} = \frac{1}{3}}$$

(β) <u>**'Sequence' of children= (g,b,b)**</u>

Similarly, $P(gbb|b, g) = \frac{1}{3}$

Exercise 3.12

The problem itself is formulated in a tricky way. Behind the puzzle of *"What is now the chance of drawing a white counter?"* is lying the real question: ***"What is now the probability of the remaining counter to be white? "***. This slight reformulation allows us to state that if an initial **hypothesis** $\mathcal{H}_0$ supposes *the original content of the bag was a white ball*, then our question changes to **'what is the probability of this hypothesis, given the data (D) ?'** (that is the **evidence** of drawing a white ball).

According to *Bayes Theorem* :

$$P(\mathcal{H}_0|D) = \frac{P(D|\mathcal{H}_0)\,P(\mathcal{H}_0)}{P(D)} \tag{45}$$

In detail:

- $P(D|\mathcal{H}_0) = 1$ as a certain event that actually happened
- $P(\mathcal{H}_0) = \frac{1}{2}$ as a prior possibility, with the assumption of equal chance for $\mathcal{H}_0$ and $\overline{\mathcal{H}_0} = 1 - \mathcal{H}_0$
- We can expand $P(D)$ using **sum rule** :

$$\begin{aligned} P(D) &= \sum_i P(D \cap \mathcal{H}_i) && (46)\\ &= P(D|\mathcal{H}_0)P(\mathcal{H}_0) + P(D|\overline{\mathcal{H}_0})P(\overline{\mathcal{H}_0}) && (47)\\ &= 1 \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{4} && (48) \end{aligned}$$

Finally, eq.45 $\implies$

$$P(\mathcal{H}_0|D) = \frac{1 \cdot 1/2}{3/4} = \frac{2}{3} \tag{49}$$

After drawing the first ball, the posterior has been updated so that the chances are increasing from 50% to $\approx$ 67%!

Exercise 3.14

This example is more straightforward :

$$\begin{aligned} P(B = heads|A = heads) &= \frac{P(A = heads \cap B = heads)}{P(A = heads)} && (50)\\ &= \frac{N_{(\text{'hh'})}}{N_{(\text{'hh,ht,th'})}} && (51)\\ &= \frac{1/4}{3/4} = \frac{1}{3} && (52) \end{aligned}$$

Where we considered initially the outcomes $\{\ hh, ht, th, tt\ \}$ (h :heads, t :tails).