Exercise set 2

Christos Dimitrakakis: `christos.dimitrakakis@unine.ch`

March 4, 2025

Deadline: 5 March 2024

EXERCISE 1. Complete the meteorologists exercise for the Bayesian and Maximum-A-Posteriori models.

EXERCISE 2. Consider a Bernoulli random variable with unknown mean

$$x_t \mid \mu \sim \mathit{Bern}(\mu),$$

where the mean is sampled from a Beta distribution

$$\mu \sim \mathit{Beta}(\alpha, \beta).$$

The Bernoulli distribution satisfies

$$P_\mu(x_1, \ldots, x_T) = \prod_{t=1}^{T} P_\mu(x_t) = \mu^{s_T}(1-\mu)^{T-s_T},$$

where $s_T = \sum_{t=1}^{T} x_T$, while the Beta distribution has the form

$$\xi(\mu) \triangleq \frac{1}{B(\alpha, \beta)} \mu^{\alpha-1}(1-\mu)^{\beta-1},$$

where the factor $B(\alpha, \beta)$ is a normalisation constant independent of μ that ensures the density integrates to one.

1. Show that the maximum likelihood estimate

$$\arg\max_\mu P_\mu(x_1, \ldots, x_T)$$

   is $\mu = s_T/T$.

2. Show that the maximum a posteriori estimate

$$\arg\max_\mu P_\mu(x_1, \ldots, x_T)\xi(\mu)$$

   is $\mu = (\alpha + s_T - 1)/(\alpha + \beta + T - 2)$.

3. Show that the posterior expected value

$$\mathbb{E}_\xi(\mu|x_1, \ldots, x_t) = \int_{-\infty}^{\infty} \mu \, d\xi(\mu \mid x_1, \ldots, x_T)$$

   is $\mu = (\alpha + s_T)/(\alpha + \beta + T)$.

Exercise 2 solution. To make this simpler, we can take the logarithm of the likelihood

$$\ln P_\mu(x_1, \ldots, x_T) = s_T \ln \mu + (T - s_T) \ln(1 - \mu)$$

Since ln is a monotonic function, finding the maximum of the log-likelihood is equivalent to finding the maximum of the likelihood. We can do so by taking the derivative and setting it to zero:

$$
\begin{aligned}
0 &= \frac{d}{d\mu} \ln P_\mu(x_1, \ldots, x_T) \\
&= \frac{d}{d\mu} s_T \ln \mu + \frac{d}{d\mu}(T - s_T) \ln(1 - \mu) \\
&= s_T/\mu - (T - s_T)/(1 - \mu) \\
\mu &= s_T/T.
\end{aligned}
$$

For the *maximum a posteriori* estimate, we can use the same approach

$$
\begin{aligned}
0 &= \frac{d}{d\mu} \ln[P_\mu(x_1, \ldots, x_T)\xi(\mu)] \\
&= \frac{d}{d\mu} \ln P_\mu(x_1, \ldots, x_T) + \frac{d}{d\mu} \ln \xi(\mu) \\
&= \frac{s_T}{\mu} - \frac{T - s_T}{1 - \mu} + \frac{\alpha - 1}{\mu} - \frac{\beta - 1}{1 - \mu} \\
&= \frac{s_T + \alpha - 1}{\mu} - \frac{T - s_T + \beta - 1}{1 - \mu} \\
(s_T + \alpha - 1)(1 - \mu) &= (T - s_T + \beta - 1)\mu \\
(s_T + \alpha - 1) &= (T + \beta + \alpha - 2)\mu \\
\mu &= \frac{s_T + \alpha - 1}{T + \beta + \alpha - 2}
\end{aligned}
$$

Finally, we can look at the expected value of the distribution. Here we can use a shortcut. We know that the expected value of a Beta distribution with parameters α, β is $\alpha/(\alpha+\beta)$. Since we now have a Beta distribution with parameters $\alpha + s_T, \beta + T - s_T$, the expected value is $\alpha + s_T/(\alpha + \beta + T)$. We can prove that this is the correct expectation as follows:

$$1/B(\alpha, \beta) \int_0^1 \mu\mu^{\alpha-1}(1-\mu)^{\beta-1}d\mu = B(\alpha+1, \beta)/B(\alpha, \beta)$$

The Beta function has the property that $B(\alpha+1, \beta) = B(\alpha, \beta)\frac{\alpha}{\alpha+\beta}$. Replacing, we obtain the stated result. $\quad\square$

EXERCISE 3. James Randi claims that he is *psychic* and can *always predict a coin toss.* Let A denote the event that JR is psychic and let $P(A) = 2^{-16}$ be your prior belief that JR is a psychic. Let us now make a set of bets about Randi. In the following, make the following assumptions

ASSUMPTION 1. *All experiments are conducted with a fair coin, whose probability of coming heads at the k-th toss is $P(H_k) = 1/2$ and where H_k is independent of $H_{k-1}, \ldots, H_1$.*

ASSUMPTION 2. *If JR is a psychic, then he can perfectly predict the coin tosses. Thus, if B_k is the event that he predicts correctly, then $P(B_k \mid A, H_k) = P(B_k \mid A) = 1$.*

1. Let B_k denote the event that JR predicts the k-th coin toss correctly. What are the dependencies between A, H_k, B_k? (You can draw a graphical model to represent them). For simplicity, take $k = 1, 2$ only.

2. What is the marginal probability $P(B_1)$? What is the marginal probability $P(B_2)$?

3. Say that JR predicts the first coin toss correctly, i.e. B_1 holds. What is then the marginal probability $P(B_2 \mid B_1)$?

Hint: The only important events for the calculations are whether JR predicts correctly or not.
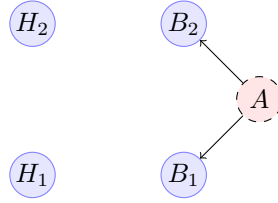
Solution. 1.

Figure 1: The Bayesian network for the prediction problem for a fair coin. The probability of correct predictions depends only on A.

2.

$$P(B_1) = P(B_1 \mid A)P(A) + P(B_1 \mid \neg A)P(\neg A) \approx 2^{-16} + 1/2 \quad (1)$$
$$P(B_2) = P(B_2 \mid A)P(A) + P(B_2 \mid \neg A)P(\neg A) \approx 2^{-16} + 1/2, \quad (2)$$

so $P(B_1) = P(B_2)$.

3. We need to calculate the following function

$$P(B_2 \mid B_1) = P(B_2 \mid A, B_1)P(A \mid B_1) + P(B_2 \mid \neg A, B_1)P(\neg A \mid B_1) \tag{3}$$

$$= P(B_2 \mid A)P(A \mid B_1) + P(B_2 \mid \neg A)P(\neg A \mid B_1) \tag{4}$$

$$= 1 \times P(A \mid B_1) + \frac{1}{2} \times [1 - P(A \mid B_1)]. \tag{5}$$

So we just need to evaluate:

$$P(A \mid B_1) = \frac{P(B_1 \mid A)P(A)}{P(B_1)} = \frac{1 \times 2^{-16}}{2^{-16} + 1/2} = \frac{1}{1 + 2^{15}} \approx 2^{-15}$$

We can now replace:

$$P(B_2 \mid B_1) = 2^{-15} + \frac{1}{2} \times [1 - 2^{-15}] \approx \frac{1}{2} + 2^{-15},$$

so a slight increase! □

EXERCISE 4. We continue the previous exercise, but with the following extra assumption.

ASSUMPTION 3. *Your utility for money is linear, i.e. $U(x) = x$ for any amount of money x.*

1. At the beginning of the experiment, JR bets you 100 CU that he can predict the *next four* coin tosses, i.e. he is willing to give you 100 CU if he doesn't predict all four tosses. How much are you willing to bet against that, i.e. how much are you willing to pay if he does predict them correctly?

2. You throw the coin 4 times, and JR guesses correctly all four times. Randi now bets you another 100 CU that he can predict the *next* four coin tosses. Up to how much would you bet now?

Proof. For this exercise, we simply need to calculate the expected utility. If JR can predict perfectly, then he has a chance of 1 of predicting. If not, then he has a chance of 2^{-4}. Marginalising over our prior:

$$P(B^4) = P(B^4|A)P(A) + P(B^4|\neg A)P(\neg A)$$
$$= 1 \times 2^{-16} + 2^{-4} \times (1 - 2^{-16}) < 2^{-16} + 2^{-4}$$

Since our utility is linear, we can now consider the two events and see what is the utility of taking the bet. If it is positive, then we agree. We can do this by calculating the point where the expected utility is greater than zero:

$$P(B^4)U(-x) + P(\neg B^4)U(100) \geq 0$$
$$x \leq 100 \times \frac{1 - 2^{-16} - 2^{-4}}{2^{-16} + 2^{-4}} \approx 1500$$

So you should bet up to 1500. We repeat the calculation for $P(B^{5:8}|B^4)$. Note that

$$P(A|B^4) = P(A)P(B^4|A)/P(B^4) \approx 2^{-16}/[2^{-16}+2^{-4}] = 1/[1+2^{12}] \approx 2^{-12}.$$

Plugging in:

$$P(B^{5:8}|B^4) = P(B^{5:8}|A)P(A|B^4) + P(B^{5:8}|\neg A)P(\neg A|B^4) = 1 \times 2^{-12} + 2^{-4} \times (1 - 2^{-12}$$

So, we still should be more than sure that JR is not psychic even after 4 straight tosses and bet up to approximately 1490.

Note: Of course, we are not being very careful here. We only considered the possibility that JR is a psychic or not. We haven't considered the possibility that JR may be performing a magic trick. This alternative hypothesis may have a much higher prior probability than JR being psychic.

$\square$

EXERCISE 5 (DO NOT HAND IN). Now assume that JR is a psychic, but that he can only predict some proportion $\theta \in [0, 1]$ of tosses (he could also be a *bad* psychic). We shall model our uncertainty about θ with a Beta distribution $\xi(\theta)$ with parameters $(2, 1)$, $\mathcal{Beta}(2, 1)$, which places higher probability on all values of θ closer to 1. So now our prior looks like this:

$$P(A) = 1 \tag{6}$$

$$P(B \mid A) = \int_0^1 P(B \mid \theta)\xi(\theta)d\theta = \int_0^1 \theta\xi(\theta)d\theta = \mathbb{E}_\xi\, \theta. \tag{7}$$

$$\tag{8}$$

So now we know that JR is psychic, but not how good a psychic he is.

1. What is the resulting posterior distribution $P(B_2 \mid B_1)$ of JR predicting the next toss correctly after he has predicted one correctly? Write this in terms of the resulting Beta posterior parameters, given that $P(A) = 2^{-16}$.

2. Write an expression for the marginal probability $P(B_{n+1} \mid B^n)$, where B^n is the event that he has predicted correctly n times.

Proof.

$$P(B_{n+1} \mid B^n) = P(B_{n+1} \mid A, B^n)P(A \mid B^n) + P(B_{n+1} \mid \neg A, B^n)P(\neg A \mid B^n)$$

$$P(B_{n+1} \mid A, B^n) = \frac{2 + n}{3 + n}$$

$$P(B_{n+1} \mid \neg A, B^n) = \frac{1}{2}$$

$$P(A \mid B^n) = \frac{P(B^n \mid A)P(A)}{P(B^n \mid A)P(A) + P(B^n \mid \neg A)P(\neg A)}$$

$$= \frac{1 \times 2^{-16}}{1 \times 2^{-16} + 2^{-n}(1 - 2^{16})} \approx 2^{n-16}, \quad \text{for } n < 16$$

$$P(B_{n+1} \mid B^n) \approx \frac{2 + n}{3 + n}2^{n-16} + \frac{1}{2}(1 - 2^{n-16}).$$

So as we can see, the rate approaches that of the Beta model exponentially fast, even though the Beta model itself only has a polynomial rate approaching to one. $\square$

EXERCISE 6. The following is an example of a *hierarchical model.* Instead of assuming that Randi is either a perfect psychic or a fraud, we can relax our assumption. Let A

denote the event that Randi is psychic, but that he can only predict some proportion $\theta \in [0, 1]$ of tosses (he could also be a *bad* psychic). We shall model our uncertainty about θ with a Beta distribution $\xi(\theta)$ with parameters $(2, 1)$, $\mathcal{B}eta(2, 1)$, which places higher probability on all values of θ closer to 1. So now our prior marginal distribution looks like this:

$$P(B) = P(B \mid A)P(A) + P(B \mid \neg A)P(\neg A) \tag{9}$$

$$P(B \mid A) = \int_0^1 P(B \mid \theta)\xi(\theta)d\theta = \int_0^1 \theta\xi(\theta)d\theta = \mathbb{E}_\xi \, \theta \tag{10}$$

$$P(B \mid \neg A) = \frac{1}{2}. \tag{11}$$

So now we think that JR is either a psychic or not, but if he is a psychic, we don't know how well he can predict. Still, we place a special emphasis on the case that he is not a psychic, where we'd expect him to predict correctly 50% of the time.

1. What is the resulting posterior distribution $P(B_2 \mid B_1)$ of JR predicting the next toss correctly after he has predicted one correctly? Write this in terms of the resulting $P(A \mid B_1)$ and the resulting Beta posterior parameters, given that $P(A) = 2^{-16}$.

2. Write an expression for the marginal probability $P(B_{n+1} \mid B^n)$.

3. Randi now offers you the opportunity to bet 100CU at every time step t that he would guess the next coin toss, with equal odds (i.e. you get 100CU if he doesn't guess correctly and lose 100CU if he does). If you accept the bet at time t, you throw the coin and Randi guesses, and you get/pay money depending on the outcome. Then he offers another bet at time $t + 1$. If at any time you decline the bet, the game stops. Define a strategy for playing this game.

Proof. Here I offer a high-level sketch. The main idea is to calculate a posterior distribution on A given the observations B. This can be written as follows:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

We can then replace the expressions for $P(B \mid A), P(B)$ and obtain the posterior distribution as usual. The rest of the exercise follows trivially. $\qquad \square$