Exercise 1. *Trace distance*

The trace distance (or L_1-distance) between two probability distributions P_X and Q_X over a discrete alphabet $\mathcal{X}$ is defined as

$$\delta(P_X, Q_X) = \frac{1}{2} \sum_{x \in \mathcal{X}} |P_X(x) - Q_X(x)| \,. \tag{1}$$

The trace distance may also be written as

$$\delta(P_X, Q_X) = \max_{\mathcal{S} \subseteq \mathcal{X}} |P_X[\mathcal{S}] - Q_X[\mathcal{S}]| \,, \tag{2}$$

where we maximise over all events $S \subseteq \mathcal{X}$ and the probability of an event $\mathcal{S}$ is given by $P_X[\mathcal{S}] = \sum_{x \in \mathcal{S}} P_X(x)$.

(a) Show that $\delta(\cdot, \cdot)$ is a good measure of distance by proving that $0 \leq \delta(P_X, Q_X) \leq 1$ and the triangle inequality $\delta(P_X, R_X) \leq \delta(P_X, Q_X) + \delta(Q_X, R_X)$ for arbitrary probability distributions P_X, Q_X and R_X.

(b) Show that definitions (2) and (1) are equivalent.

(c) Let us now find an operational meaning for the trace distance. Suppose that P_X and Q_X represent the probability distributions of the outcomes of two dice that look identical. You are allowed to throw one of them only once and then have to guess which die that was. What is your best strategy? What is the probability that you guess correctly and how can you relate that to the trace distance $\delta(P_X, Q_X)$?

Solution.

(a) The lower bound follows from the fact that each element of the sum (1) is non-negative. We get the upper bound from

$$(P_X, Q_X) = \frac{1}{2} \sum_{x \in \mathcal{X}} |P_X(x) - Q_X(x)| \leq \frac{1}{2} \sum_{x \in \mathcal{X}} P_X(x) + Q_X(x) = 1. \tag{S.1}$$

The triangle inequality can be written as

$$\frac{1}{2} \sum_{x \in \mathcal{X}} |P_X(x) - R_X(x)| \leq \frac{1}{2} \sum_{x \in \mathcal{X}} |P_X(x) - Q_X(x)| + |Q_X(x) - R_X(x)| \,. \tag{S.2}$$

If the inequality is true for every $x \in \mathcal{X}$, it is also true for the above sum. It is thus sufficient to prove that $|P_X(x) - R_X(x)| \leq |P_X(x) - Q_X(x)| + |Q_X(x) - R_X(x)|$ for all $x \in \mathcal{X}$. We know that $|\alpha + \beta| \leq |\alpha| + |\beta|$ for $\alpha, \beta \in \mathbb{R}$. Hence the inequality follows with $\alpha = P_X(x) - Q_X(x)$ and $\beta = Q_X(x) - R_X(x)$.

(b) To maximise $|P_X[\mathcal{S}] - Q_X[\mathcal{S}]| = \left|\sum_{x \in \mathcal{S}} P_X(x) - Q_X(x)\right|$ in (2), we choose

$$S = \{x \in \mathcal{X} \;:\; P_X(x) \geq Q_X(x)\} \,. \tag{S.3}$$

Let $\bar{\mathcal{S}}$ be its complement, such that $\mathcal{S} \cup \bar{\mathcal{S}} = \mathcal{X}, \mathcal{S} \cap \bar{\mathcal{S}} = \emptyset$. We may now write

$$0 = \sum_{x \in \mathcal{X}} P_X(x) - Q_X(x) = \sum_{x \in S} |P_X(x) - Q_X(x)| - \sum_{x \in \bar{\mathcal{S}}} |P_X(x) - Q_X(x)| \,. \tag{S.4}$$

The terms $P_X(x) - Q_X(x)$ are positive in the first sum on the right-hand side and negative in the second sum. We can thus take the modulus after the sum in the first term and write

$$\begin{aligned}\left|\sum_{x\in\mathcal{S}} P_X(x) - Q_X(x)\right| &= \sum_{x\in\mathcal{S}} |P_X(x) - Q_X(x)| = \sum_{x\in\bar{\mathcal{S}}} |P_X(x) - Q_X(x)| \\ &= \frac{1}{2}\sum_{x\in\mathcal{X}} |P_X(x) - Q_X(x)| .\end{aligned} \tag{S.5}$$

The last equality is obtained because $\sum_{x\in\mathcal{X}} = \sum_{x\in\mathcal{S}} + \sum_{x\in\bar{\mathcal{S}}}$. Altogether this proves that the two definitions (1) and (2) are equivalent.

(c) Your best strategy is to say it was the die more likely to outcome the result you obtained, i.e. if you define the event $\mathcal{S} = \{x \in \mathcal{X} : P_X(x) \geq Q_X(x)\}$ (the results that are more likely with die P), then you better say that you threw die P if you get an outcome $x \in \mathcal{S}$ and Q if $x \in \bar{\mathcal{S}}$.

The probability that your guess is right is

$$P_{\checkmark} = \frac{1}{2}P_X(\mathcal{S}) + \frac{1}{2}Q_X(\bar{\mathcal{S}}) = \frac{1}{2}\left(P_X(\mathcal{S}) + 1 - Q_X(\mathcal{S})\right) = \frac{1}{2}\left(1 + \delta(P_X, Q_X)\right), \tag{S.6}$$

by definition (2) of trace distance.

Exercise 2. ***Weak law of large numbers***

Let A be a positive random variable with expectation value $\langle A\rangle = \sum_a a\ P_A(a)$. Let $P[A \geq \varepsilon]$ denote the probability of an event $\{A \geq \varepsilon\}$ for some $\varepsilon > 0$.

(a) Prove Markov's inequality

$$P[A \geq \varepsilon] \leq \frac{\langle A\rangle}{\varepsilon}. \tag{3}$$

(b) Use Markov's inequality to prove Chebyshev's inequality, i.e.,

$$P\left[(X-\mu)^2 \geq \epsilon\right] \leq \frac{\sigma^2}{\epsilon}, \tag{4}$$

where σ denotes the standard deviation of X.

(c) Use Chebyshev's inequality to prove the weak law of large numbers for i.i.d. X_i:

$$\lim_{n\to\infty} P\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)^2 \geq \varepsilon\right] = 0 \quad \text{for any } \varepsilon > 0, \mu = \langle X_i\rangle . \tag{5}$$

Solution.

(a) This is done by multiplying the summands by a fraction a/ε which is large than or equal to 1 for $a \geq \varepsilon$:

$$P[A \geq \varepsilon] = \sum_{a\geq\varepsilon} P_A(a) \leq \sum_{a\geq\varepsilon} \frac{aP_a(a)}{\varepsilon} \leq \sum_{a} \frac{aP_a(a)}{\varepsilon} = \frac{\langle A\rangle}{\varepsilon}. \tag{S.7}$$

(b) Note that we can substitute $A = (X - \mu)^2$ into Markov's inequality to get Chebyshev's inequality

$$P[(X - \mu)^2 \geq \varepsilon] \leq \frac{\sigma^2}{\varepsilon}, \tag{S.8}$$

where σ is the standard deviation of X.

(c) If we now substitute $X = \frac{1}{n} \sum X_i$ the expectation value remains the same, whereas the variance scales with $\frac{1}{n}$ because the X_i are independent and all have the same distribution (i.i.d.). We get

$$P\left[\left(\frac{1}{n}\sum_i X_i - \mu\right)^2 \geq \varepsilon\right] \leq \frac{{\sigma_i}^2}{n\varepsilon}, \tag{S.9}$$

where σ_i is the standard deviation of X_i. The weak law of large numbers now follows with $n \to \infty$ for any fixed $\varepsilon > 0$.

Exercise 3. ***Conditional probabilities I: Mrs. Smith's children***

(a) You are strolling in a park when a woman, Mrs. Smith, is approaching you with a covered twin buggy. She tells you that she has fraternal twins and in the course of the conversation you learn that one of the twins is a girl named Jane. What is the probability that the other twin is a girl, too?

(b) We now change the situation slightly. After having a short conversation with Mrs. Smith you know that she has fraternal twins. You ask her 'are they both boys?' and she answers 'no'. What is the probability that she has two girls?

(c) Another version of this story goes as follows: during the conversation with Mrs. Smith you ask 'could you please tell me the sex of one of your twins?' and she answers 'one of them is a girl'. What is now the probability that the other is a girl?

(d) Explain the difference between the three situations in words and in terms of conditional probabilities. What are the hidden assumptions?

Solution.

(a) For fraternal twins the probabilities are

$$P(\text{two boys}) = P(\text{two girls}) = \frac{1}{2}P(\text{one boy, one girl}) = \frac{1}{4} \tag{S.10}$$

and the sexes of the twins are independent. Considering the table

Jane	other child	possible?
boy	boy	no
boy	girl	no
girl	boy	yes
girl	girl	yes

and taking Mrs. Smith's answer into account we see that only 2 of initially 4 equally likely events are still possible. This gives probability $\frac{1}{2}$ that the other twin is also a girl.

(b) Consider the following table:

firstborn	secondborn	Mrs. Smith's answer	sex of other child
boy	boy	'yes'	irrelevant
boy	girl	'no'	boy
girl	boy	'no'	boy
girl	girl	'no'	girl

The three relevant cases (the cases for which Mrs. Smith answers with 'no') are now equally likely, hence occur with probability $\frac{1}{3}$ each. Thus she has two girls only with probability $\frac{1}{3}$. Note that the distinction into firstborn and secondborn is arbitrary and artificial – we only need something to distinguish the twins.

(c) We go over all possibilities the mother has to mention the sex of one of her twins:

firstborn	secondborn	Mrs. Smith mentions	you learn	other child
boy	boy	firstborn	≥ 1 boy	boy
boy	boy	secondborn	≥ 1 boy	boy
boy	girl	firstborn	≥ 1 boy	girl
boy	girl	secondborn	≥ 1 girl	boy
girl	boy	firstborn	≥ 1 girl	boy
girl	boy	secondborn	≥ 1 boy	girl
girl	girl	firstborn	≥ 1 girl	girl
girl	girl	secondborn	≥ 1 girl	girl

Assuming that Mrs. Smith mentions the sex of each child with equal probability all 8 cases are equally probable. However, only 4 of them are possible in our case (the ones in which you learn '≥ 1 girl'). In half of these cases the other child is a girl, hence this happens with probability $\frac{1}{2}$.

(d) In (a) you are given the name of one child to distinguish it from the other. A priori there are 4 equally likely events, but after excluding the 2 events in which Jane is a boy you are left with 2 only. In terms of conditional probabilities this is

$$P(\text{two girls} \mid \text{Jane is a girl}) = \frac{P(\text{Jane is a girl} \mid \text{two girls})P(\text{two girls})}{P(\text{Jane is a girl})} = \frac{1 \cdot \frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}. \tag{S.11}$$

The situation in (b) is pretty clear. Before asking you have 4 possibilities with equal probability, $\frac{1}{4}$ each. After getting the answer that not both are boys one is excluded and you are left with 3 equally likely possibilities. In terms of conditional probabilities we can again use Bayes' law to obtain

$$P(\text{two girls} \mid \text{not two boys}) = \frac{P(\text{not two boys} \mid \text{two girls})P(\text{two girls})}{P(\text{not two boys})} = \frac{1 \cdot \frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}. \tag{S.12}$$

Hence the probability for Mrs. Smith to have two girls is $\frac{1}{3}$.

The situation in (c) is again a bit different. Here we went over all possibilities the mother has to mention the sex of one of her twins. To do so we again artificially distinguished the children as firstborn and secondborn. Doing the same maths as above we obtain

$$P(\text{two girls} \mid \text{you learn} \geq 1 \text{ girl}) = \frac{P(\text{you learn} \geq 1 \text{ girl} \mid \text{two girls})P(\text{two girls})}{P(\text{you learn} \geq 1 \text{ girl})} = \frac{1 \cdot \frac{2}{8}}{\frac{4}{8}} = \frac{1}{2}. \tag{S.13}$$

Note that we implicitly assume that Mrs. Smith would have approached you also if she had two boys, in this case telling you that one of the twins is a boy. In addition, it is assumed that if she had a boy and a girl she would have told you that she has at least one boy/girl with probability $\frac{1}{2}$ each. These hidden assumptions are important when we use the last table to deduce the conditional probabilities in (S.13). If they do not hold the eight cases are not equally likely.

Exercise 4. ***Conditional probabilities II: how knowing more does not always help***

Suppose you are visiting your grandfather in his hut in Scotland. You had offered him a radio for Christmas three years ago, but he is not so fond of such modern technologies and has not used it since. You decide to initiate a game to prove to him that technology is helpful: every evening you alone listen to the weather forecast on the radio and then both you and your grandfather try to guess if it will rain next morning. Having lived there since birth, your grandfather knows that it rains on 80% of the days. You had reached the same conclusion on previous summer holidays. You also know that the weather forecast is right 80% of the time and is always correct when it predicts rain.

(a) What is the optimal strategy for your grandfather? And for you?

(b) Both of you keep a record of your guesses and the actual weather for statistical analysis. After some months who will have guessed correctly more often?

(c) Can you think of an argument to convince him that listening to the forecast is useful?

Solution. Let us start by sorting the notation:

P_R - probability that it rains;
$P_{\hat{R}}$ - probability that the radio predicts rain;
P_S - probability that it is sunny (no rain);
$P_{\hat{S}}$ - probability that the radio predicts sunshine;
$P_{R|\hat{R}}$- probability that it rains *when* radio predicts rain;
$P_{R\hat{R}}$ - probability that it rains *and* radio predicted rain.

Notice that $P_{R|\hat{R}}$ is a conditioned probability while $P_{R\hat{R}}$ is a joint probability,

$$P_{R\hat{R}} = P_{R|\hat{R}} P_{\hat{R}}. \tag{S.14}$$

(a) You were given the probabilities $P_R = 80\%, \quad P_{R\hat{R}} + P_{S\hat{S}} = 80\%, \quad P_{R|\hat{R}} = 100\%.$

The best thing your grandfather can do is to say it will rain every morning – this way he will win 80% of the time. As for you, if you use (S.14) you will compute the probabilities represented in Fig. 1.

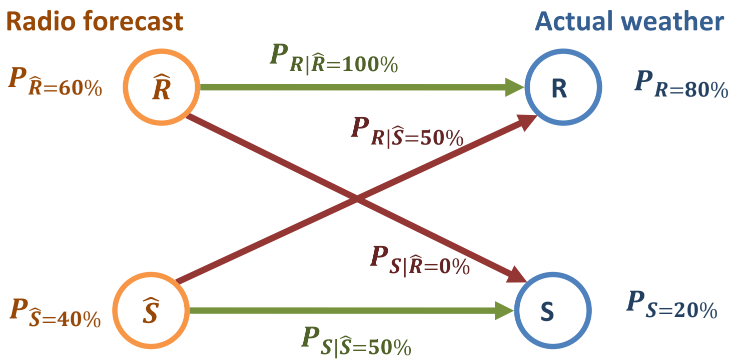

Figure 1: The radio forecast and the actual weather: marginal and conditional probabilities.

Note: this figure could be interpreted as a channel – check exercise sheet 2 for more details on channels.

When the forecast is rain you should believe it. When the report predicts sun it fails with 50% chance, so any strategy in this case is equally good (or bad). You may for instance say it will always rain or follow the forecast.

(b) “After some months” means “after so many days that you can apply the law of large numbers”. Both you and your grandfather will be correct on approximately 80% of the days – this is easy to see since one of your optimal strategies is to copy your grandfather and say it will always rain.

(c) In exercise sheet 2.