ETH Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Quantum Information Theory
Tips 1

HS 14

Exercise 1.1 Trace distance

Let us try to recap on probability distributions with a fast example. Imagine you have a die that you know to be biased. The probabilities of obtaining the different outcomes are

$$P(\text{“1”}) = 50\%, \quad P(\text{“2”}) = 10\%, \quad P(\text{“3”}) = 20\%, \quad P(\text{“4”}) = 5\%, \quad P(\text{“5”}) = 10\%, \quad P(\text{“6”}) = 5\%,$$

We call the set of outcomes $\mathcal{X} = \{\text{“1”}, \text{“2”}, \text{“3”}, \text{“4”}, \text{“5”}, \text{“6”}\}$ the *alphabet* of the probability distribution, and in this note we refer to its elements “1”, “2”, etc. as *outcomes*. Any subset of the alphabet is an *event* – for instance the event “obtaining an even number” is represented by the set $\mathcal{S} = \{\text{“2”}, \text{“4”}, \text{“6”}\}$, and the probability of an event is obviously just the sum of the probabilities of all elements of that event, $P[\mathcal{S}] = \sum_{x\in\mathcal{S}} P(x) = 10\% + 5\% + 5\%$.
To simplify the notation we will assume the alphabet to be known and represent the probability distributions like this:

$$P = \begin{pmatrix} 0.5 \\ 0.10 \\ 0.20 \\ 0.05 \\ 0.10 \\ 0.05 \end{pmatrix}.$$

Needless to say the probabilities of all outcomes are non-negative and sum up to one.
Now we will introduce the trace distance. A distance between two entities is supposed to quantify how far apart they are in some sense. The most immediate example is the usual distance between two cities, but we can think of something a bit more imaginative. Imagine you have two bowls with some beans inside each. We could define a distance as the number of beans we would have to move from one bowl to the other in order to equilibrate them, which happens to be half of the difference between the number of beans in the bowls.
The trace distance generalises this concept to probability distributions: given two probability distributions P_X and Q_X (the subscript refers to the random variable X — check pages 4–5 of the script for details) how much weight would we have do move between them to end up with two identical distributions? Like in the case of the beans, the answer is half of the difference between the probabilities of all outcomes $P_X(x)$ and $Q_X(x)$,

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

and this defines the *trace distance*. In a quick example, if we have two D4 dice that follow the probability distributions

$$P_X = \begin{pmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{pmatrix}, \qquad Q_X = \begin{pmatrix} 0.7 \\ 0.1 \\ 0.1 \\ 0.1 \end{pmatrix},$$

the trace distance between them would be $\delta(P_X, Q_X) = \frac{1}{2}(|0.25 - 0.7| + 3 \cdot |0.25 - 0.1|) = 0.45$.
As we will see, the trace distance has a very neat operational meaning and will be used a lot in this course, especially its generalisation to the quantum case (we will get there in a few weeks).
In part a) you have to prove that the trace distance is a proper distance: that it is non-negative is trivial to show. Upper bound also follows straightforwardly. The triangle inequality (meaning it takes longer to get

from one place to another if you have to stop somewhere else first) follows from the triangle inequality for the absolute values. All is fairly direct if you use definition (1).
In part *b*) you are introduced to another formulation of the trace distance,

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

and have to prove that it is equivalent to (1). To do so, you should first stop to think what is the event $\mathcal{S}$ that maximises $|P_X[\mathcal{S}] - Q_X[\mathcal{S}]|$. This is easy to see if you expand $P_X[\mathcal{S}]$ according to the definition above. Once you know what $\mathcal{S}$ is you only need to play with the sums until you reach (1). Two very trivial, very useful little things you will need are $\mathcal{S} \cup \bar{\mathcal{S}} = \mathcal{X}$ and $\sum_{x\in\mathcal{X}} P_X(x) = 1$. The script (pages 8–9) also helps.
Finally, in part *c*) you will find a physical interpretation of the trace distance: a way of relating it to the probability of correctly distinguishing two states (like our two dice) after a single measurement in one of them. Again, this exercise is easy once you figure what your best strategy and probability of winning are. Definition (2) will be useful. In the end you should get

$$P_{\checkmark} = \frac{1}{2}\left(1 + \delta(P_X, Q_X)\right),$$

which tells us that the more *distant* (according to the trace distance) two distributions are, the easier it is to distinguish them. No surprises here!

Exercise 1.2 Weak Law of Large Numbers

We all know the law of large numbers from secondary school: if you repeat an experiment many many times in the exact same conditions, the average outcome will be the expectation value of the experiment. In this exercise you will have to prove it. This is marginally more exciting than proving $0 < 1$ in Calculus I.
Recap: the expectation value of a random variable is given by $\langle X\rangle = \sum_{x\in\mathcal{X}} x P_X(x)$ and the standard deviation is $\sigma_X^2 = \langle (X - \langle X\rangle)^2\rangle$. Mathematically, "repeating the same experiment many times in the exact same conditions" is expressed by *independent, identical distributed* random variables. In a nutshell, that means that $P_{X_1 X_2\cdots X_N} = P_{X_1}^N$ — details on page 8 of the script.
You will start by proving Markov's inequality in part *a*). Later you will use it to prove the weak law of large numbers. So first you want to prove that the probability of the event $\mathcal{S} = A \geq \varepsilon$ is smaller than $\langle A\rangle/\varepsilon$. Expand the definition of probability of an event and see how you can to the expectation value. It is just a very simple clever trick, really. In part *b*) you will prove the Chebyshev inequality, an easy consequence of the Markov inequality. In part *c*) you need to prove the weak law of large numbers, by applying the Chebyshev inequality.
From this law it follows that:

$$\lim_{n\to\infty} P(|X - \mu| \geq \varepsilon) = 0$$

Hence, the probability that the average of n iid RVs X_i approaches $\langle X_i\rangle$ arbitrarily close as $n \to \infty$, approaches 1.

Exercise 1.3 Jensen's inequality

For f a convex function, and the probability distribution $\{p_1, .., p_n\}$ we have:

$$f(\sum p_k x_k) \leq \sum p_k f(x_k)$$

One can prove this inequality by induction, starting from the definition of the convex function as the base of the induction.
Another way of expressing this inequality, for the random variable X and the convex function ϕ is:

$$\phi[E(X)] \leq E[\phi(X)]$$

This can easily be shown as a consequence of the previous inequality.

Exercise 1.4 Conditional probabilities: how knowing more does not always help

Think of the probabilities given by the previous knowledge, and once given by the forecast. Consider conditional probabilities. You can think of the set up from the information theoretical view point.