ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Quantum Information Theory
Solutions 1

HS 14
Dr. J.M. Renes

Exercise 1.1 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 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.*

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

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

The triangle inequality can be written as

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

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) *Show that definition (2) follows from (1).*

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

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

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

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

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

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

This proves that the two definitions (1) and (2) are equivalent.

c) *Suppose that P_X and Q_X represent the probability distributions of the outcomes of two dice. You are allowed to throw one of them 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 $\delta(P_X, Q_X)$?*

Your best strategy is to say it was the die more likely to outcome the result you obtained, ie. 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), than 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

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

by definition (2) of trace distance.

Exercise 1.2 Weak Law of Large Numbers

a) *Prove Markov's inequality*

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

This is done by multiplying the summands by a fraction $a/\varepsilon \geq 1$:

$$P[A \geq \varepsilon] = \sum_{a \geq \varepsilon} P_A(a) \leq \sum_{a \geq \varepsilon} \frac{a P_a(a)}{\varepsilon} \leq \sum_a \frac{a P_a(a)}{\varepsilon} = \frac{\langle A \rangle}{\varepsilon}.$$

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

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

where σ is the standard deviation of X.

c) *Use Chebyshev's inequality to prove the weak law of large numbers for i.i.d. X_i:* If we now substitute $X \to \frac{1}{n} \sum X_i$ the expectation value remains the same, whereas the variance scales with $\frac{1}{n}$. We get

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

and the weak law of large numbers follows with $n \to \infty$ for any fixed $\varepsilon > 0$.

Exercise 1.3 Jensen's inequality

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

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

One can prove this inequality by induction, starting from the definition of the convex function as the base of the induction. Then for the induction step we have, given a prob. distribution $\{p_1, ..., p_{n+1}\}$:

$$f(\sum_{k=1}^{n+1} p_k x_k) = f((1 - p_{n+1}) \sum_{k=1}^{n} \frac{p_k x_k}{1 - p_{n+1}} + p_{n+1} x_{n+1})$$

Now by the base of induction we have:

$$f((1-p_{n+1})\sum_{k=1}^{n}\frac{p_k x_k}{1-p_{n+1}}+p_{n+1}x_{n+1}) \leq (1-p_{n+1})f(\sum_{k=1}^{n}\frac{p_k x_k}{1-p_{n+1}})+p_{n+1}f(x_{n+1})$$

By the induction assumption that the Jensen's inequality holds for the case n probabilities in the distribution (here the distribution $\{\frac{p_k}{1-p_{n+1}}\}$), we have:

$$(1-p_{n+1})f(\sum_{k=1}^{n}\frac{p_k x_k}{1-p_{n+1}})+p_{n+1}f(x_{n+1}) \leq (1-p_{n+1})\sum_{k=1}^{n}\frac{p_k}{1-p_{n+1}}f(x_k)+p_{n+1}f(x_{n+1})$$

$$\rightarrow f(\sum_{k=1}^{n+1}p_k x_k) \leq (1-p_{n+1})\sum_{k=1}^{n}\frac{p_k}{1-p_{n+1}}f(x_k)+p_{n+1}f(x_{n+1}) = \sum_{k=1}^{n+1}p_k f(x_k)$$

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

You and your grandfather are trying to guess if it will rain tomorrow. All he knows is that it rains on 80% of the days. You know that and you also listen to the weather forecast and know that it is right 80% of the time and is always correct when it predicts rain.

Let us start by sorting the notation:

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

$$P_{R\hat{R}} = P_{R|\hat{R}}P_{\hat{R}}. \tag{4}$$

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

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

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 (4) you will compute the probabilities represented in Fig. 1 (note: this figure could be interpreted as a channel – check exercise sheet two for details).

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) *Both of you keep a record of your guesses and the actual weather for statistical analysis. After some time (*i.e. *enough that you can apply the weak law of large numbers), who will have guessed correctly more often?*

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.

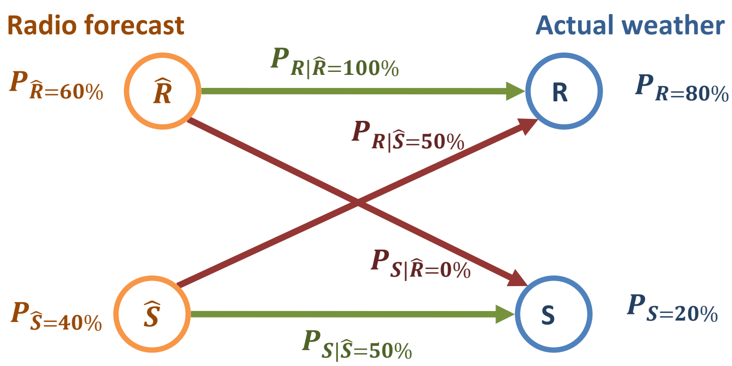

Figure 1: The radio forecast and the actual weather: marginal and conditional probabilities.