Quantum Information Theory
Solutions 3.

HS 2015
Prof. R. Renner

Exercise 1. *Smooth min-entropy in the i.i.d. limit*

The smooth min-entropy of a random variable X over $\mathcal{X}$ is defined as

$$H^{\epsilon}_{\min}(X)_P = \max_{Q_X \in \mathcal{B}^{\epsilon}(P_X)} H_{\min}(X)_Q, \tag{1}$$

where the maximum is taken over all probability distributions Q_X that are ϵ-close to P_X. Furthermore, we define an i.i.d. random variable $\vec{X} = \{X_1, X_2, \dots, X_n\}$ on $\mathcal{X}^{\times n}$ with $P_{\vec{X}}(\vec{x}) = \prod_{i=1}^n P_X(x_i)$.

Use the weak law of large numbers to show that the smooth min-entropy converges to the Shannon entropy $H(X)$ in the i.i.d. limit:

$$\lim_{\epsilon\to 0} \lim_{n\to\infty} \frac{1}{n} H^{\epsilon}_{\min}(\vec{X})_{P_{\vec{X}}} = H(X)_{P_X}. \tag{2}$$

Solution. The weak law of large numbers as shown in exercise series 1 is given by:

$$\lim_{n\to\infty} P\left[\left(\frac{1}{n}\sum_i S_i - \mu\right)^2 \geq \eta\right] = 0 \quad \text{for any } \eta > 0,\ \mu = \mathbb{E}[S]. \tag{S.1}$$

Setting $S_i = h_P(x_i) := -\log P_X(x_i)$ we get $\mu = H(X)$ and thus

$$\lim_{n\to\infty} P\left[\left|\frac{1}{n}\sum_i h_P(x_i) - H(X)\right| < \nu\right] = 1 \quad \text{for any } \nu > 0. \tag{S.2}$$

Notice that to simplify the notation we dropped the subscript $P(X)$ for $H(X)$.
This knowledge allows us to restrict the set of vectors $\vec{x}$ to typical outcomes, namely we introduce a subset $\mathcal{G}_\nu$ of $\mathcal{X}^{\times n}$:

$$\mathcal{G}^{(n)}_\nu = \left\{\vec{x} \in \mathcal{X}^{\times n} : \left|\frac{1}{n}\sum_i h_P(x_i) - H(X)\right| < \nu\right\}. \tag{S.3}$$

The weak law of large numbers can now be restated simply as

$$\lim_{n\to\infty} P_{\vec{X}}[\mathcal{G}^{(n)}_\nu] = \lim_{n\to\infty} P_{\vec{X}}[\vec{x} \in \mathcal{G}^{(n)}_\nu] = 1. \tag{S.4}$$

Furthermore, let $(\mathcal{G}^{(n)}_\nu)^c$ denote the complement of $\mathcal{G}^{(n)}_\nu$ in $\mathcal{X}^{\times n}$. As a next step we choose

$$Q_{\vec{X}}(\vec{x}) = \begin{cases} P_{\vec{X}}(\vec{x})/P_{\vec{X}}[\mathcal{G}^{(n)}_\nu] & \text{if } \vec{x} \in \mathcal{G}^{(n)}_\nu \\ 0 & \text{if } \vec{x} \in (\mathcal{G}^{(n)}_\nu)^c \end{cases}. \tag{S.5}$$

This distribution has the property that for any fixed ν, the trace distance $\delta(P_{\vec{X}}, Q_{\vec{X}})$ vanishes in the limit of $n \to \infty$. This can be seen when we use the alternative definition of δ introduced in problem set 1 and when we take the probabilities over the set $(\mathcal{G}^{(n)}_\nu)^c$ of all events where $Q_{\vec{X}} < P_{\vec{X}}$:

$$\lim_{n\to\infty} \delta(P_{\vec{X}}, Q_{\vec{X}}) = \lim_{n\to\infty} P_{\vec{X}}[(\mathcal{G}^{(n)}_\nu)^c] - Q_{\vec{X}}[(\mathcal{G}^{(n)}_\nu)^c] = 0. \tag{S.6}$$

In particular, we can now evaluate the "smooth" min-entropy for any fixed $\epsilon > 0$ and $\nu > 0$:

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} H^{\epsilon}_{\min}(\vec{X}) &\geq \lim_{n\to\infty} \min_{\vec{x}\in\mathcal{X}^{\times n}} \frac{1}{n} h_Q(\vec{x}) \\
&= \lim_{n\to\infty} \min_{\vec{x}\in\mathcal{G}_\nu} \frac{1}{n} h_P(\vec{x}) + \lim_{n\to\infty} \frac{1}{n} \log P_{\vec{X}}[\mathcal{G}^{(n)}_\nu] \\
&= \lim_{n\to\infty} \min_{\vec{x}\in\mathcal{G}_\nu} \frac{1}{n} \sum_i h_P(x_i) \\
&\geq H(X) - \nu
\end{aligned} \tag{S.7}$$

The first inequality is a consequence of the fact that our $Q_{\vec{X}}$ is not necessarily optimal (as a matter of fact it could be shown that it actually is). It follows that this construction only gives us a lower bound on the i.i.d. limit once we set ν arbitrarily close to zero and let $\epsilon \to 0$. However, from the definition of Shannon and min-entropy follows directly that min-entropy can never exceed Shannon entropy,

$$H_{\min}(X)_P \leq H(X)_P, \tag{S.8}$$

since the information gain in the worst-case can never be higher than the average information gain. This concludes the proof.

Exercise 2. *An interpretation of the trace distance*

We have introduced the trace distance of two probability distributions in exercise sheet 1 and have shown that it is at least a reasonable distance measure in that it is positive and fulfils the triangle inequality. In this exercise we show an important property of this measure which is arguably the main reason why the trace distance is so frequently used, e.g. in security proofs of cryptographic protocols.

Consider two random variables X and X' on the same alphabet $\mathcal{X}$ distributed P_X and $P_{X'}$, respectively. The trace distance between them is $\delta(P_X, P_{X'}) =: \varepsilon$. The goal is to show that there exists a joint distribution of X and X', $\bar{P}_{XX'}$, which is compatible *with P_X and $P_{X'}$ and has the property that*

$$\bar{P}[X \neq X'] \leq \varepsilon. \tag{3}$$

Compatibility here means that the marginals of $\bar{P}_{XX'}$, $\bar{P}_X$ and $\bar{P}_{X'}$, coincide with P_X and $P_{X'}$, respectively.

(a) *Argue that for $\varepsilon \in \{0, 1\}$ the statement is (almost) trivially true.*

(b) *From now on we assume $0 < \varepsilon < 1$. For $x \in \mathcal{X}$ define*

$$P^{\min}_X(x) := \frac{\min\{P_X(x), P_{X'}(x)\}}{1-\varepsilon}\ , \quad P^{\text{diff}}_X(x) := \frac{P_X(x) - (1-\varepsilon)P^{\min}_X(x)}{\varepsilon} \quad \text{and} \tag{4}$$

$$P^{\text{diff}}_{X'}(x) := \frac{P_{X'}(x) - (1-\varepsilon)P^{\min}_X(x)}{\varepsilon}. \tag{5}$$

Check that $P^{\min}_X, P^{\text{diff}}_X$ and $P^{\text{diff}}_{X'}$ are valid probability distributions on $\mathcal{X}$.

(c) *Construct a possible joint distribution $\bar{P}$ as follows: throw a die with odds $\{1-\varepsilon, \varepsilon\}$. If the outcome corresponds to the probability $1-\varepsilon$ distribute XX' s.t. $X = X'$ and X distributed $P^{\min}_X$. If not, let X and X' be independently distributed P^{diff}_X and $P^{\text{diff}}_{X'}$, respectively. Check that $\bar{P}$ is compatible with P_X and $P_{X'}$ and that $\bar{P}[X \neq X'] \leq \varepsilon$.*

Why is this way of interpreting the trace distance so helpful? Think of X as an ideal system about which we can make precise statements (e.g. about its security w.r.t. attacks from adversaries) and about X'

as the real system. First the special case $\varepsilon = 0$: suppose we found in our theoretical analysis that the descriptions in terms of probability distributions of X and X' differ in trace distance by zero, $\varepsilon = 0$. Then, by the above, we know that there exists a joint distribution $\bar{P}$ describing both the ideal and the real system at the same time s.t. they never behave differently, i.e. always $X = X'$. In other words, with probability 1 the real system behaves ideally. In the general case, when the trace distance between P_X and $P_{X'}$ can be bounded by some ε, the statement that X' behaves like an ideal system still holds with probability $1 - \varepsilon$.

Solution.

(a) If $\varepsilon = 0$ this means that the probability distributions are equal. $\bar{P}$ should then just be

$$\bar{P}_{XX'}(x, x') = \delta_{xx'} P_X(x). \tag{S.9}$$

By definition $X = X'$ always in this case. In the case where $\varepsilon = 1$ the statement $\bar{P}[X \neq X'] \leq 1$ is trivially satisfied.

(b) We start with $P_X^{\min}$. There are two things we need to show: (i) $\forall x \in \mathcal{X} : P_X^{\min}(x) \geq 0$ and (ii) $\sum_{x\in\mathcal{X}} P_X^{\min}(x) = 1$. (i) is trivially satisfied by definition because both P_X and $P_{X'}$ are valid probability distributions. For (ii), define the (by now well-known) set $\mathcal{S} := \{x \in \mathcal{X} \mid P_X(x) \geq P_{X'}(x)\}$, showing up in the alternative definition of $\delta(\cdot, \cdot)$, and calculate

$$\begin{aligned}
\sum_{x\in\mathcal{X}} P_X^{\min}(x) &= \frac{1}{1-\varepsilon}\left(\sum_{x\in\mathcal{S}} P_{X'}(x) + \sum_{x\in\mathcal{S}^c} P_X(x)\right) \\
&= \frac{1}{1-\varepsilon}\left(\sum_{x\in\mathcal{S}}[P_{X'}(x) - P_X(x)] + \sum_{x\in\mathcal{S}} P_X(x) + \sum_{x\in\mathcal{S}^c} P_X(x)\right) \\
&= \frac{1}{1-\varepsilon}\left(-\delta(P_X, P_{X'}) + \sum_{x\in\mathcal{X}} P_X(x)\right) \\
&= \frac{1-\varepsilon}{1-\varepsilon} = 1.
\end{aligned} \tag{S.10}$$

Likewise we see that P_X^{diff} and $P_{X'}^{\text{diff}}$ are positive. Furthermore:

$$\sum_{x\in\mathcal{X}} P_X^{\text{diff}}(x) = \frac{1}{\varepsilon}\sum_{x\in\mathcal{X}}\big(P_X(x) - (1-\varepsilon)P_X^{\min}(x)\big) = \frac{1}{\varepsilon}\big(1 - (1-\varepsilon)\big) = 1, \tag{S.11}$$

where we used $\sum_{x\in\mathcal{X}} P_X^{\min}(x) = 1$ from above. The same calculation obviously works for $P_{X'}^{\text{diff}}$.

(c) The distribution $\bar{P}$ described in the exercise can be written as

$$\bar{P}_{XX'}(x, x') = (1-\varepsilon)\delta_{xx'} P_X^{\min}(x) + \varepsilon P_X^{\text{diff}}(x) P_{X'}^{\text{diff}}(x'). \tag{S.12}$$

We first check compatibility. By definition of $\bar{P}$ for $x \in \mathcal{X}$ we can write

$$\begin{aligned}
\bar{P}_X(x) = \sum_{x'\in\mathcal{X}} \bar{P}_{XX'}(x, x') &= (1-\varepsilon)P_X^{\min}(x) + \varepsilon P_X^{\text{diff}}(x) \\
&= (1-\varepsilon)P_X^{\min}(x) + \varepsilon\frac{P_X(x) - (1-\varepsilon)P_X^{\min}(x)}{\varepsilon} = P_X(x).
\end{aligned} \tag{S.13}$$

Again the same calculation works for $P_{X'}$, which concludes the argument that $\bar{P}$ is compatible with P_X and $P_{X'}$. Now, what is the probability that X and X' differ? By definition of $\bar{P}$ they could only differ if the 'coin flip' has the outcome associated to probability ε. Therefore we have almost trivially

$$\bar{P}[X \neq X'] \leq \varepsilon. \tag{S.14}$$

Exercise 3. ***Fano's inequality***

Given two random variables X and Y, how well can we predict X given Y? Fano's inequality bounds the probability of error in such a prediction in terms of the conditional entropy $H(X|Y)$. The goal of this exercise is to prove the inequality

$$P_{\text{error}} \geq \frac{H(X|Y) - 1}{\log |X|}. \tag{6}$$

(a) Representing the guess of X by the random variable $\widehat{X}$, which is some function, possibly random, of Y, show that $H(X|\widehat{X}) \geq H(X|Y)$.

(b) Consider the indicator random variable E which is 1 if $\widehat{X} \neq X$ and zero otherwise. Using the chain rule we can express the conditional entropy $H(E, X|\widehat{X})$ in two ways:

$$H(E, X|\widehat{X}) = H(E|X, \widehat{X}) + H(X|\widehat{X}) = H(X|E, \widehat{X}) + H(E|\widehat{X}). \tag{7}$$

Calculate each of these four expressions and complete the proof of the Fano inequality.

Hints: *For $H(E|\widehat{X})$ use the fact that conditioning reduces entropy: $H(E|\widehat{X}) \leq H(E)$. For $H(X|E, \widehat{X})$ consider the cases $E = 0, 1$ individually.*

Solution.

(a) By definition of the setting the random variable $\widehat{X}$ arises from Y through a channel (conditional probability distribution) $P_{\widehat{X}|Y}$. Thus we can use the data processing inequality. It leads directly to $H(X|\widehat{X}) \geq H(X|Y)$.
Notice that here another way of saying '$\widehat{X}$ arises from Y' is to state that X, Y, and $\widehat{X}$ form a Markov chain, $X \leftrightarrow Y \leftrightarrow \widehat{X}$.

(b) First, we have $H(E|X, \widehat{X}) = 0$ since E is determined from X and $\widehat{X}$. On the other hand, $H(E|\widehat{X}) \leq H(E) = H_{\text{bin}}(P_{\text{error}})$ since conditioning reduces entropy. Here, H_{bin} is the binary entropy. We then have

$$\begin{aligned} H(X|E, \widehat{X}) &= H(X|E = 0, \widehat{X})p(E = 0) + H(X|E = 1, \widehat{X})p(E = 1) \\ &= 0(1 - P_{\text{error}}) + H(X|E = 1, \widehat{X})P_{\text{error}} \leq P_{\text{error}} \log |X| \,. \end{aligned} \tag{S.15}$$

Putting this together we obtain

$$H(X|Y) \leq H(X|\widehat{X}) \leq H_{\text{bin}}(P_{\text{error}}) + P_{\text{error}} \log |X| \leq 1 + P_{\text{error}} \log |X|, \tag{S.16}$$

where the last inequality follows since $H_{\text{bin}}(x) \leq 1$. Rearranging terms gives the Fano inequality.