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:

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

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 $1 - \varepsilon$.

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.