Freie Universität Berlin

Tutorials on Quantum Information Theory

Winter term 2020/21

Problem Sheet 7
Quantum Shannon Theory and Key Distribution

J. Eisert, J. Haferkamp, J. C. Magdalena De La Fuente

---

1. **On Shannon entropy...**

   To begin with let us first show some simple properties of entropies, in particular, of the mutual information.

   Recall the definition of the Shannon entropies for random variables X, Y which take values in $\mathcal{X}, \mathcal{Y}$ and are distributed according to probability distributions p, q over $\mathcal{X}$ and $\mathcal{Y}$, respectively.

   $$(1)\ H(X) = -\sum_{x\in\mathcal{X}} p(x)\log p(x)\ \text{(Shannon entropy)} \tag{1}$$

   $$(2)\ H(X|Y) = H(X,Y) - H(Y) = \sum_{x\in\mathcal{X}} p(x) H(Y|X=x)\ \text{(Conditional entropy)} \tag{2}$$

   $$(3)\ I(X:Y) = H(Y) - H(Y|X)\ \text{(Mutual information)} \tag{3}$$

   a) Show that $0 \leq H(X) \leq \log|\mathcal{X}|$, where the first equality holds *iff* there is an $x \in \mathcal{X}$ for which $p(x) = 1$ and the second inequality holds *iff* $p(x) = 1/|\mathcal{X}|$ for all x.

   Solution: *First inequality.* Since $p(x) \leq 1$ for all x, we have $\log(x) \leq 0$, which implies that with $p(x) \geq 0$ that $0 \leq H(X)$. We have equality if there exist $x \in \mathcal{X}$ with $p(x) = 1$, since implies $H(X) = 0$. The reverse direction follows from the fact that $H(X)$ is concave and the set of probability distributions is convex and, hence, it take its minimum at the extreme points, which have $p(x) = 1$ for one $x \in X$. *Second inequality.* The second inequality can be proven using Lagrange multipliers. In particular, if all $p_x := p(x) > 0$, we can compute the gradient $(\text{grad}H(X))_{p_x} = -\log(p_x) - 1$. Together with the restriction $\sum_x p_x = 1$, we obtain the equations $-\log(p_x) - 1 + \lambda = 0$. As log is an injective function, this can only be the case if all p_x are equal. Evaluating $H(X)$ at $p_x = 1/|X|$ yields the second inequality. It can easily be checked that the case with $p_x = 0$ for some multiple $x \in X$ does not yield a larger value simply by repeating the above argument on the non-vanishing p_x

   b) Show that the Shannon entropy is *subadditive*, i.e., that $H(X,Y) \leq H(X) + H(Y)$.
   Hint: Show that $H(X,Y) - H(X) - H(Y) \leq 0$ *using that* $\log_2 x \ln 2 = \ln x \leq x - 1$.

Solution: Using that $\sum_x p(x,y) = p(y)$ we have that

$$H(X,Y) - H(X) - H(Y) = -\sum_{x,y} p(x,y)(\log p(x,y) - \log p(x) - \log p(y)) \tag{4}$$

$$= \sum_{x,y} p(x,y) \log \frac{p(x)p(y)}{p(x,y)} \leq \frac{1}{\ln 2} \sum_{x,y} p(x,y) \left(\frac{p(x)p(y)}{p(x,y)} - 1 \right) \tag{5}$$

$$= \frac{1}{\ln 2} \sum_{x,y} (p(x)p(y) - p(x,y)) = \frac{1}{\ln 2}(1-1) = 0 \tag{6}$$

We also see that equality holds if and only if $p(x,y) = p(x)p(y)$ for all x, y, i.e. for independent random variables X and Y.

c) Show that $H(Y|X) \geq 0$ and hence $I(X:Y) \leq H(Y)$ with equality if and only if Y is a (deterministic) function of X.

Hint: Use Bayes' rule: $p(x,y) = p(y|x)p(x)$

Solution: We have

$$H(X,Y) = -\sum_{x,y} p(x,y) \log p(x)p(y|x) \tag{7}$$

$$= -\sum_{x} p(x) \log p(x) - \sum_{x,y} p(x,y) \log p(y|x) \tag{8}$$

$$= H(X) - \sum_{x,y} p(x,y) \log p(y|x) \tag{9}$$

The last term thus equals the conditional entropy and from $p(y|x) \leq 1$ we derive that the conditional entropy is positive. For $p(y|x) = 1$ we have equality, which is precisely the condition that y follows deterministically from x.

This translates directly to the claim for the mutual information: $I(X:Y) = H(Y) - H(Y|X) \leq H(Y)$.

d) Show that $H(Y|X) \leq H(Y)$ and hence that $I(X:Y) \geq 0$ with equality if and only if X and Y are independent random variables.

Solution: Using subadditivity, we have

$$H(Y|X) = H(X,Y) - H(X) \leq H(X) + H(Y) - H(X) = H(Y). \tag{10}$$

Equality holds if $H(X,Y) = H(X) + H(Y)$ which is the case if $p(x,y) = p(x)p(y)$ as we have seen when proving subadditivity.

Again, the statement for the mutual information is an immediate consequence.

2. **... and the von-Neumann entropy**

For any state $\rho \in \mathcal{D}(\mathcal{H})$ with $\dim\mathcal{H} = d$ the von-Neumann entropy is defined as $S(\rho) = -\operatorname{Tr}(\rho \log \rho)$.

a) Show that $0 \leq S(\rho)$ with equality if and only if ρ is pure. (One can also show the upper bound $S(\rho) \leq \log d$.)

Solution: Diagonalize ρ to obtain $S(\rho) = -\sum_x \lambda_x \log \lambda_x$. Using that λ_x form a probability distribution, the claim follows from the of the Shannon entropy.

b) Show that the von-Neumann entropy is *subadditive* in the sense that if two distinct systems A and B have a joint quantum state ρ^{AB} then $S(A,B) \leq S(A) + S(B)$.
Hint: You may use the inequality $S(\rho) \leq -\operatorname{Tr}[\rho \log \sigma]$ *for an arbitrary quantum state* σ*.*

Solution: Let $\rho = \rho_{AB}$, $\sigma = \rho_A \otimes \rho_B$

$$S(A,B) = S(\rho_{AB}) \leq S(\operatorname{Tr}_B[\rho_{AB}]) + S(\operatorname{Tr}_A[\rho_{AB}])$$

Now choose $\sigma = \rho_A \otimes \rho_B$. Then introducing the eigenvalue decomposition of $\rho_A = U\Lambda_A U^\dagger$ and $\rho_B = V\Lambda_B V^\dagger$, we write

$$\begin{aligned} S(\rho) &\leq -\operatorname{Tr}[\rho \log \rho_A \otimes \rho_B] \\ &= -\operatorname{Tr}[\rho(U\otimes V)\log(\Lambda_A \otimes \Lambda_B)(U\otimes V)^\dagger] \\ &= -\operatorname{Tr}[\rho(U\otimes V)(\log(\Lambda_A)\otimes \mathbb{1} + \mathbb{1}\otimes \log\Lambda_B)(U\otimes V)^\dagger] \\ &= -\operatorname{Tr}[\rho(\log(\rho_A)\otimes\mathbb{1})] - \operatorname{Tr}[\rho(\mathbb{1}\otimes\log\rho_B))] \\ &= S(\rho_A) + S(\rho_B). \end{aligned}$$

c) Suppose that $p = (p_i)_i$ is a probability vector and the states ρ_i are mutually orthogonal. Show that

$$S\left(\sum_i p_i \rho_i\right) = H(p) + \sum_i p_i S(\rho_i).$$

and use this result to infer that

$$S\left(\sum_i p_i \rho_i \otimes |i\rangle\langle i|\right) = H(p) + \sum_i p_i S(\rho_i),$$

where $\langle i|j\rangle = \delta_{ij}$ and the ρ_i are arbitrary quantum states.

Solution:

$$\begin{aligned} S\left(\sum_i p_i\rho_i\right) &= -\operatorname{Tr}\left[\sum_{i,j} p_i \lambda_i^j \Pi_i^j \log\left(\sum_{i,j} p_i \lambda_i^j \Pi_i^j\right)\right] && (11)\\ &= -\operatorname{Tr}\left[\sum_{i,j} p_i\lambda_i^j\Pi_i^j\left(\sum_{i,j}(\log p_i + \log\lambda_i^j)\Pi_i^j\right)\right] && (12)\\ &= -\left(\sum_{i,j,k,l} p_i\lambda_i^j(\log p_k + \log\lambda_k^l)\right)\operatorname{Tr}\left[\Pi_i^j\Pi_k^l\right] && (13)\\ &= -\left(\sum_{i,j} p_i\lambda_i^j(\log p_i + \log\lambda_i^j)\right) && (14)\\ &= H(p) + \sum_i p_i S(\rho_i) && (15) \end{aligned}$$

We use the equality in the subadditivity condition for $S(\rho\otimes\sigma) = S(\rho) + S(\sigma)$. This gives $S(\rho_i \otimes |i\rangle\langle i|) = S(\rho_i) + S(|i\rangle\langle i|) = S(\rho_i)$

d) Use the results from (b) and (c) to infer that the von-Neumann entropy S is concave.

Solution: To show: $S(\sum_i p_i\rho_i) \geq \sum_i p_i S(\rho_i)$.

$$\sum_i p_i S(\rho_i) = S\left(\sum_i p_i\rho_i \otimes |i\rangle\langle i|\right) - H(p)$$
$$\leq S\left(\sum_i p_i\rho_i\right) + S\left(\sum_i p_i\,|i\rangle\langle i|\right) - H(p) = S\left(\sum_i p_i\rho_i\right)$$

since $S(\sum_i p_i\,|i\rangle\langle i|) = H(p)$

3. **Classical capacities of quantum channels**
Although this exercise might look very long, it isn't. In the next paragraphs we just want to give you an overview on the formalism introduced in the lecture and needed for this exercise in a compressed fashion. No need to be intimidated ;)
In the lecture, we saw two alternative characterisations of the classical channel capacity of a quantum channel $\mathcal{E}$, which is given by the its Holevo-information $\chi(\mathcal{E})$. The task here is to establish the equivalence of these expressions.

To this end, recall the definition of the quantum mutual information of a bi-partite quantum system in a state ρ_{AB}

$$I(A:B)_{\rho_{AB}} \coloneqq S(\rho_A) + S(\rho_B) - S(\rho_{AB}). \tag{16}$$

The Holevo information of channel can be defined using the following scheme: Alices encodes the information of a classical random variable X taking values in $\mathcal{X}$ with probability distribution p_X into a quantum state using a set of states $\{\rho_x\}_{x\in\mathcal{X}}$. To keep track of the classical random variable but formulating everything quantum mechanically, we think of Alice encoding the result in another faithfully register N using orthogonal basis $\{|x\rangle\}_{x\in\mathcal{X}}$. From this notebook register N the classical information of X can be completely recovered. Altogether, Alice prepares the bi-partite state

$$\rho_{NA} = \sum_x p_X(x)\,|x\rangle\langle x|_N \otimes \rho_A^x. \tag{17}$$

Then, the state in system A is sent to Bob using the channel $\mathcal{E}$. Thus, we end up with a final state shared between Alice's notebook and Bob

$$\rho_{NB} = \sum_x p_X(x)\,|x\rangle\langle x|_N \otimes \mathcal{E}(\rho_A^x)_B. \tag{18}$$

We can now ask for the mutual information between the variable X encoded in N and Bob's output of the channel. Analogously to the classical result, maximizing the mutual information over all possible input variables X and encodings yields the capacity of the quantum channel to transmit classical informations, i.e.

$$\chi(\mathcal{E}) = \max_{(X,p_X,\{\rho^x\})} I(N,B)_{\rho_{NA}}. \tag{19}$$

a) Show that

$$\chi(\mathcal{E}) = \max_{(X,p_X,\{\rho^x\})} \left\{ S(\mathcal{E}(\sum_x p_X(x)\rho^x)) - \sum_x p_X(x) S(\mathcal{E}(\rho^x)) \right\}. \tag{20}$$

Solution: Let's simply call $\sigma_B^x \coloneqq \mathcal{E}(\rho_A^x)_B$. The marginal states of ρ_{NB} are

$$\rho_N = \mathrm{Tr}_B\, \rho_{NB} = \sum_x p_X(x)\, |x\rangle\langle x|\,, \tag{21}$$

$$\rho_B = \mathrm{Tr}_N\, \rho_{NB} = \sum_x p_X(x)\sigma_B^x. \tag{22}$$

In the calculation of Exercise 2.c) we have already seen that

$$S(\rho_{NB}) = H(X) + \sum_x p_X(x) S(\sigma_B^x). \tag{23}$$

Thus, the mutual information is

$$I(N:B)_{\rho_{NB}} = S(\rho_N) + S(\rho_B) - S(\rho_{NB}) \tag{24}$$

$$= H(X) + S\left(\sum_x p_X(x)\sigma_B^x\right) - H(X) - \sum_x p_X(x) S(\sigma_B^x), \tag{25}$$

from which the claim follows.

Remember that Shannon's noisy channel coding theorem states that the capacity of a noisy channel T is given by the maximum over all inputs of the mutual information:

$$C(T) = \max_{X, p_X} I(X:Y),$$

where Y is the random variable describing the output of the channel T with input X.

b) Determine the channel capacity of the binary symmetric channel defined by

$$\Pr(0|0) = \Pr(1|1) = 1 - p$$
$$\Pr(1|0) = \Pr(0|1) = p.$$

Hint: It may be useful to expand $H(Y|X)$ *as* $\sum_x p(x) H(Y|X = x)$.

Solution: We have

$$\begin{aligned} I(X:Y) &= H(Y) - H(Y|X) = H(Y) - \sum_x p_X(x) H(Y|X=x) \\ &= H(Y) - \sum_x p(x) H_2(p) = H(Y) - H_2(p) \\ &\leq 1 - H_2(p), \end{aligned}$$

where the last equation follows because Y is a binary random variable and $H_2(p) = -p\log(p) - (1-p)\log(1-p)$ is the binary entropy. Note that we can always ensure that Y is equally distributed by choosing X equally distributed. Hence, $C = 1 - H(p)$.

We now want to determine the channel capacity of the binary erasure channel as defined by

$$\Pr(0|0) = \Pr(1|1) = 1 - p$$
$$\Pr(e|0) = \Pr(e|1) = p.$$

c) First, use the expansion $H(Y) = H(Y,Z) = H(E) + H(Y|Z)$ to show that $H(Y) = H(p) + (1-p)H(\pi)$. Here, we let Z be the random variable distinguishing between the event $E = \{Y = e\}$ and $\neg E = \{Y \neq e\}$. We have that $\Pr(Z = E) = p$. Furthermore we call the propability defining the distribution of the input variable $\pi = \Pr(X = 1)$.

Hint: Use Eq. (2) *and* $\Pr(Y = y | Y \neq e) = \Pr(X = y)$.

Solution: First, we have that $\Pr(Z = E) = \Pr(Z = E|X = 0)\Pr(X = 0) + \Pr(Z = E|X = 1)Pr(X = 1) = p\pi + p(1-\pi) = p$. On the other hand

$$\begin{aligned} H(Y) &= H(Z) + \Pr(Z = \neg E)H(Y|Z = \neg E) + p(Z = E)H(Y|Z = E) \\ &= H_2(p) + pH(Y|Y = e) + (1-p)H(Y|Y \neq e) \\ &= H_2(p) + p \cdot 0 + (1-p)H_2(\pi), \end{aligned}$$

where we used that $H(Y|Y = e) = 0$ since Y can only take the value e with probability 1. On the other hand, $H(Y|Y = \neg e) = H(\pi)$ since if $Y \neq e$ then the case $Y = 1$ occurs with probability π whereas the case $Y = 0$ occurs with probability $1 - \pi$.

d) Use this result and proceed analogously to the binary symmetric channel to determine the channel capacity of the erasure channel.

Solution: Now, we need to maximize the obtaining expression as

$$\begin{aligned} C = \max_{p(x)} H(Y) - H(Y|X) &= \max_{p(x)} H_2(p) + (1-p)H_2(\pi) - H_2(p) \\ &= \max_{\pi}(1-p)H_2(\pi) = 1 - p \end{aligned}$$

where the maximum of H_2 of 1 is achieved for $\pi = 1/2$.

4. **Detecting Eve.** One key feature of the BB'84 protocol for quantum key distribution is that Alice and Bob are able to estimate how many bits were corrupted by the channel or Eve by comparing their results on a subset.

   In this excercise, we will prove this statement. More precisely, let Alice and Bob randomly select n of their $2n$ bits check for errors. We denote the number of errors in the test bits by e_T and the number of errors in the remaining, untested n bits by e_R. Then, for any $\delta > 0$

$$p \coloneqq \Pr\{e_T \leq \delta n \ \wedge \ e_R \geq (\delta + \epsilon)\} \leq \exp\left[-\mathcal{O}(n\epsilon^2)\right]. \tag{26}$$

   In other words, the probability that the number of errors in the unknown bits deviatiates by more than ϵ from the observed fraction δ in the test bits gets very small large n and ϵ.

   We denote the total number of errors that occur in the $2n$ bits by μn.

   a) Argue that

$$p \leq \binom{2n}{n}^{-1} \binom{\mu n}{\delta n} \binom{(2-\mu)n}{(1-\delta)n} \delta n. \tag{27}$$

Solution: Ok, we are given a bit strings of length $2n$. This string can be partitioned into two strings of equal size in $\binom{2n}{n}$ ways. The number of ways in which we end up with one substring containing exactly i of the μn corrupted bits is $\binom{\mu n}{i}\binom{2n-\mu n}{n-i}$. Therefore, the probability of getting up to δn corrupted bits is

$$p = \binom{2n}{n}^{-1} \sum_{i=1}^{\delta n} \binom{\mu n}{i} \binom{(2-\mu)n}{n-i} \leq \binom{2n}{n}^{-1} \binom{\mu n}{\delta n} \binom{(2-\mu)n}{(1-\delta)n} \delta n, \tag{28}$$

where we have used that $i/n \leq \delta = \frac{\mu}{2} - \frac{\epsilon}{2} \leq \frac{\mu}{2}$.

We will need a few identities to massage this term. To this end, let $H(p) = -p\log_2 p - (1-p)\log_2(1-p)$ be the binary entropy.

b) Show that

$$nH(p) + \mathcal{O}(\log_2 n) \leq \log_2 \binom{n}{pn} \leq nH(p) + \mathcal{O}(\log_2 n). \tag{29}$$

Hint: Recall Stirling's bound $\sqrt{2\pi}\sqrt{n}\, n^n \mathrm{e}^{-n} \leq n! \leq \mathrm{e}\sqrt{n}\, n^n \mathrm{e}^{-n}$.

Solution: Setting $q = 1-p$,

$$\binom{n}{np} = \frac{n!}{(np)!(nq)!} \leq \frac{\mathrm{e}\sqrt{n}\,\left(\frac{n}{\mathrm{e}}\right)^n}{2\pi n\sqrt{pq}\,\left(\frac{np}{\mathrm{e}}\right)^{np}\left(\frac{nq}{\mathrm{e}}\right)^{nq}} \tag{30}$$

$$= \frac{\mathrm{e}}{2\pi\sqrt{npq}} p^{-np} q^{-nq}. \tag{31}$$

Taking the logarithm yields

$$\log_2 \binom{n}{np} \leq -n[p\log_2 p + q\log_2 q] + \log_2 \frac{e}{2\pi} - \frac{1}{2}\log_2 npq = nH(p) + \mathcal{O}(\log_2 n). \tag{32}$$

The lower bound follows analogously and only differs in the constant offset.

Furthermore, one can derive the following simple bound for the binary entropy $H(x) \leq 1 - 2\left(x - \frac{1}{2}\right)^2$. (If you are curious, it is a good excercise to use Taylor's theorem including an estimate for the remainder to derive this bound.)

Solution: The first derivatives of H are

$$H'(x) = -\log_2(x) + \log_2(1-x) \tag{33}$$

$$H''(x) = -\frac{1}{x\ln(2)} - \frac{1}{(1-x)\ln(2)} \tag{34}$$

$$H'''(x) = \frac{1}{x^2\ln(2)} - \frac{1}{(1-x)^2\ln(2)}. \tag{35}$$

The maximum of $H(x)$ is at $x_{\max} = \frac{1}{2}$ with $H(x_{\max}) = 1$ and $H''(x_{\max}) = -\frac{4}{\ln(2)} \geq 4$. Thus, by Taylor's theorem

$$H(x) = 1 - \frac{2}{\ln 2}\left(x - \frac{1}{2}\right)^2 + R(x) \leq 1 - 2\left(x - \frac{1}{2}\right)^2 + R(x), \tag{36}$$

with $R(x) = \frac{1}{6}H'''(\xi)\left(x - \frac{1}{2}\right)^3$ for suitable $\xi \in (\frac{1}{2}, x)$ for $x \geq \frac{1}{2}$ or $\xi \in (x, \frac{1}{2})$ for $x \leq \frac{1}{2}$. The third derivative can be written as $H'''(\xi) = -2\frac{\xi - \frac{1}{2}}{\xi^2\ln(2)(\xi-1)^2}$. Thus, as x and ξ are always on the same side of $\frac{1}{2}$ we always end up with an overall minus sign. So since $R(x) \leq 0$ for all x, it can be dropped to arrive at the bound.

c) Plug everything together and show that $p \leq \exp\left[-\mathcal{O}(n\epsilon^2)\right]$.

Solution: The solution of (b) implies that up to log-factors

$$\binom{bn}{an} \approx 2^{bnH(a/b)}. \tag{37}$$

Thus, up to log-terms in n

$$\log_2 p \leq -2nH(1/2) + \mu n H(\delta/\mu) + (2-\mu)nH\left(\frac{1-\delta}{2-\mu}\right) \tag{38}$$

$$\leq -2n + \mu n\left(1 - 2\left(\frac{\delta}{\mu} - \frac{1}{2}\right)^2\right) + (2-\mu)n\left(1 - 2\left(\frac{1-\delta}{2-\mu} - \frac{1}{2}\right)^2\right) \tag{39}$$

$$= -2n + \mu n - \frac{1}{2}\mu n(\epsilon/\mu)^2 + 2n - \mu n - \frac{1}{2}n\epsilon^2/(2-\mu) \tag{40}$$

$$= -\frac{1}{2}n\left(\frac{1}{\mu} + \frac{1}{2-\mu}\right)\epsilon^2 \tag{41}$$

$$= -n\frac{1}{\mu(\mu-2)}\epsilon^2 \in -\mathcal{O}(n\epsilon^2). \tag{42}$$

Here we have used that from $\delta = \mu/2 - \epsilon/2$ it follows that

$$2\left(\frac{\delta}{\mu} - \frac{1}{2}\right)^2 = \frac{1}{2}\left(\frac{\epsilon}{\mu}\right)^2 \tag{43}$$

and

$$2\left(\frac{1-\delta}{2-\mu} - \frac{1}{2}\right)^2 = \frac{1}{2}\left(\frac{2-\mu+\epsilon}{2-\mu} - 1\right)^2 \tag{44}$$

$$= -\frac{1}{2}\frac{\epsilon^2}{(2-\mu)^2}. \tag{45}$$