Assignment 4

MATH 667 Quantum Information Theory

Mark Girard

24 March 2016

1 Problem 1

Problem 1. Let $\mathcal{E} : \mathrm{H}_n \to \mathrm{H}_m$ be a linear map from the space of $n \times n$ Hermitian matrices to the space of $m \times m$ Hermitian matrices. The dual map $\mathcal{E}^* : \mathrm{H}_m \to \mathrm{H}_n$ is the linear map defined by the relation

$$\mathrm{Tr}[\rho \mathcal{E}^*(\sigma)] = \mathrm{Tr}[\mathcal{E}(\rho)\sigma] \quad \forall \rho \in \mathrm{H}_n \text{ and } \sigma \in \mathrm{H}_m .$$

(a) Show that if $\mathcal{E}$ is a CP map then $\mathcal{E}^*$ is also a CP map.

(b) Show that if $\mathcal{E}$ is TP then $\mathcal{E}^*$ is unital (i.e. show that $\mathcal{E}^*(I_m) = I_n$).

(c) Let $\rho, \sigma \in \mathrm{H}_{n,+,1}$ be two density matrices. Show that $\sigma \prec \rho$ (i.e. the vector of eigenvalues of σ is majorized by that of ρ) if and only if there exists a unital CPTP map $\mathcal{E} : \mathrm{H}_n \to \mathrm{H}_n$ such that $\sigma = \mathcal{E}(\rho)$.

Solution. In the following, we use $\mathrm{id}_{\mathrm{H}_\ell}$ to denote the identity mapping on H_ℓ for some integer ℓ. Recall that a linear map $\Lambda : \mathrm{H}_n \to \mathrm{H}_m$ is completely positive if and only if the map $\Lambda \otimes \mathrm{id}_{\mathrm{H}_\ell}$ is positivity preserving for some integer $\ell \geq n$.

(a) *Proof.* Suppose that $\mathcal{E}$ is a completely positive map, let $\ell = \max\{m, n\}$ and let $P \in \mathrm{H}_m \otimes \mathrm{H}_\ell$ be a positive operator. We need to show that $(\mathcal{E}^* \otimes \mathrm{id}_{\mathrm{H}_\ell})(P) \in \mathrm{H}_n \otimes \mathrm{H}_\ell$ is a positive operator. Recall that an operator $A \in \mathrm{H}_n \otimes \mathrm{H}_\ell$ is positive if and only if it holds that $\mathrm{Tr}[AQ] \geq 0$ for all positive operators Q. Let $Q \in \mathrm{H}_n \otimes \mathrm{H}_\ell$ be a positive, then

$$\mathrm{Tr}[Q(\mathcal{E}^* \otimes \mathrm{id}_{\mathrm{H}_\ell})(P)] = \mathrm{Tr}[(\mathcal{E} \otimes \mathrm{id}_{\mathrm{H}_\ell})(Q)P] \geq 0$$

where $(\mathcal{E} \otimes \mathrm{id}_{\mathrm{H}_\ell})(Q)$ is positive by the complete positivity of $\mathcal{E}$. Hence $\mathcal{E}^*$ is completely positive. $\square$

(b) *Proof.* Suppose that $\mathcal{E}$ is trace preserving. Note that the identity matrix I_n is the only operator $A \in \mathrm{H}_n$ with the property that $\langle \psi | A | \psi \rangle = 1$ holds for all unit vectors $|\psi\rangle \in \mathbb{C}^n$. Let $|\psi\rangle \in \mathbb{C}^n$ be an arbitrary unit vector. Then

$$\langle \psi | \mathcal{E}^*(I_m) | \psi \rangle = \mathrm{Tr}[|\psi\rangle\langle\psi| \mathcal{E}^*(I_m)] = \mathrm{Tr}[\mathcal{E}(|\psi\rangle\langle\psi|) I_m] = \mathrm{Tr}[\mathcal{E}(|\psi\rangle\langle\psi|)] = \mathrm{Tr}[|\psi\rangle\langle\psi|] = 1$$

since $\mathcal{E}$ is trace-preserving. It follows that $\mathcal{E}^*(I_m) = I_n$, as desired. $\square$

(c) *Proof.* Let $\rho, \sigma \in \mathrm{H}_{n,+,1}$ be two density matrices and let $\vec{\lambda}$ and $\vec{\mu}$ be the vectors of the eigenvalues of ρ and σ respectively. There exist orthonormal bases $\{|u_i\rangle\}$ and $\{|v_i\rangle\}$ of $\mathbb{C}^n$ such that

$$\rho = \sum_i \lambda_i |u_i\rangle\langle u_i| \quad \text{and} \quad \sigma = \sum_i \mu_i |v_i\rangle\langle v_i|.$$

Furthermore, define unitary operators $U = \sum_i |u_i\rangle\langle i|$ and $V = \sum_i |v_i\rangle\langle i|$ such that $\rho = U \operatorname{diag}(\vec{\lambda}) U^*$ and $\sigma = V \operatorname{diag}(\vec{\mu}) V^*$.

We first suppose that $\sigma \prec \rho$, that is $\vec{\mu} \prec \vec{\lambda}$. By assumption, there exists a doubly stochastic matrix D such that $\vec{\mu} = D\vec{\lambda}$. Since D is doubly stochastic, it can be written as a convex combination of permutation matrices

$$D = \sum_{\pi} p_\pi P_\pi,$$

where the sum is taken over all permutations π of n elements, P_π is the permutation operator corresponding to π defined by

$$P_\pi = \sum_{i=1}^{n} |\pi(i)\rangle\langle i|,$$

and each $p_\pi \geq 0$ is a probability such that $\sum_\pi p_\pi = 1$. Note that $P_\pi^* = P_\pi^{-1}$ and thus $P_\pi^* P_\pi = P_\pi P_\pi^* = I_n$ holds for every permutation π. Therefore

$$\sum_\pi p_\pi P_\pi^* P_\pi = \sum_\pi p_\pi P_\pi P_\pi^* = \sum_\pi p_\pi I_n = I_n.$$

Consider the collection of operators $\{K_\pi\}$ where $K_\pi = \sqrt{p_\pi} V P_\pi^* U^*$ for each π. We see that

$$\sum_\pi K_\pi^* K_\pi = \sum_\pi p_\pi U P_\pi V^* V P_\pi^* U^* = U\Big(\sum_\pi p_\pi P_\pi P_\pi^*\Big)U^* = U I_n U^* = I_n$$

and thus the collection $\{K_\pi\}$ is a Kraus representation for the channel $\mathcal{E} : \mathrm{H}_n \to \mathrm{H}_n$ defined by

$$\mathcal{E}(A) = \sum_\pi K_\pi A K_\pi^* = \sum_\pi p_\pi V P_\pi^* U^* A U P_\pi V$$

for all $A \in \mathrm{H}_n$. Note that $\mathcal{E}$ is unital. Indeed, the dual map $\mathcal{E}^*$ can be given by $\mathcal{E}^*(A) = \sum_\pi K_\pi^* A K_\pi$ for all operators A and

$$\sum_\pi K_\pi K_\pi^* = \sum_\pi p_\pi V P_\pi^* U^* U P_\pi V^* = V\Big(\sum_\pi p_\pi P_\pi^* P_\pi\Big)V^* = V I_n V^* = I_n.$$

Hence $\mathcal{E}^*$ is also a CPTP map with Kraus operators $\{K_\pi^*\}$. Since $\mathcal{E}^*$ is trace preserving and $(\mathcal{E}^*)^* = \mathcal{E}$, it follows from part (b) that $\mathcal{E}$ is unital.

Finally, note that $P_\pi^* \operatorname{diag}(\vec{\lambda}) P_\pi = \operatorname{diag}(P_\pi \vec{\lambda})$ holds for every π, since

$$P_\pi^* \operatorname{diag}(\vec{\lambda}) P_\pi = \sum_{i,j,k} |i\rangle\langle \pi(i)| (\lambda_j |j\rangle\langle j|) |\pi(k)\rangle\langle k| = \sum_i \lambda_i |\pi(i)\rangle\langle \pi(i)| = \sum_i \lambda_{\pi^{-1}(i)} |i\rangle\langle i| = \operatorname{diag}(P_\pi \vec{\lambda})$$

and $P_\pi \vec{\lambda} = \sum_i \lambda_i |\pi(i)\rangle = \sum_i \lambda_{\pi^{-1}(i)} |i\rangle$. Now

$$\begin{aligned}
\mathcal{E}(\rho) = \sum_\pi p_\pi V P_\pi^* U^* \rho U P_\pi V^* &= \sum_\pi p_\pi V P_\pi^* \operatorname{diag}(\vec{\lambda}) P_\pi V^* \\
&= \sum_\pi p_\pi V \operatorname{diag}(P_\pi \vec{\lambda}) V^* \\
&= V \operatorname{diag}\Big(\sum_\pi p_\pi P_\pi \vec{\lambda}\Big) V^* = V \operatorname{diag}(\vec{\mu}) V^* = \sigma,
\end{aligned}$$

and thus $\mathcal{E}(\rho) = \sigma$ for the unital CPTP map $\mathcal{E}$.

Now suppose that there exists a unital CPTP map $\mathcal{E}$ such that $\sigma = \mathcal{E}(\rho)$. Define an $n \times n$ matrix D whose elements are given by

$$D_{ij} = \mathrm{Tr}[|v_i\rangle\langle v_i| \mathcal{E}(|u_j\rangle\langle u_j|)].$$

Note that $\sum_i |u_i\rangle\langle u_i| = \sum_i |v_i\rangle\langle v_i| = I_n$, since both $\{|u_j\rangle\}$ and $\{|v_i\rangle\}$ are orthonormal bases. All of the columns and rows of D each sum to 1, since

$$\sum_i D_{ij} = \mathrm{Tr}[I_n \mathcal{E}(|u_j\rangle\langle u_j|)] = \mathrm{Tr}[\mathcal{E}(|u_j\rangle\langle u_j|)]] = 1$$

holds for all j by the fact that $\mathcal{E}$ is trace-preserving and

$$\sum_j D_{ij} = \mathrm{Tr}[|v_i\rangle\langle v_i|\mathcal{E}(I_n)] = \mathrm{Tr}[|v_i\rangle\langle v_i|] = 1$$

holds for all i by the fact that $\mathcal{E}$ is unital. Furthermore, each D_{ij} is nonnegative, since $\mathcal{E}(|u_j\rangle\langle u_j|)$ is a positive operator by positivity of $\mathcal{E}$, and thus

$$D_{ij} = \langle v_i|\mathcal{E}(|u_j\rangle\langle u_j|)|v_i\rangle \geq 0$$

for all i and j. It follows that D is a doubly stochastic matrix.

We now show that $\vec{\mu} = D\vec{\lambda}$. The i^{th} entry of $D\vec{\lambda}$ is

$$\begin{aligned}(D\vec{\lambda})_i &= \sum_j \mathrm{Tr}[|v_i\rangle\langle v_i|\mathcal{E}(|u_j\rangle\langle u_j|)]\lambda_j \\ &= \mathrm{Tr}\Big[|v_i\rangle\langle v_i|\mathcal{E}\Big(\underbrace{\sum_j \lambda_j |u_j\rangle\langle u_j|}_{\rho}\Big)\Big] = \mathrm{Tr}[|v_i\rangle\langle v_i|\sigma] = \langle v_i|\sigma|v_i\rangle = \mu_i,\end{aligned}$$

and thus $\vec{\mu} = D\vec{\lambda}$ for a doubly stochastic matrix D. It follows that $\vec{\mu} \prec \vec{\lambda}$ and thus $\sigma \prec \rho$, as desired. $\square$

2 Problem 2

Problem 2. Find necessary and sufficient conditions for which the following equality holds:

$$S(\rho^{\mathsf{AB}}) = |S(\rho^{\mathsf{A}}) - S(\rho^{\mathsf{B}})|. \tag{2.1}$$

Give an example.

Solution. The necessary and sufficient conditions for equality come from the following observation together with Proposition 1. The problem can be broken down into two sub-statements:

(i) $S(\rho^{\mathsf{AB}}) = S(\rho^{\mathsf{B}}) - S(\rho^{\mathsf{A}})$ if and only if $\rho^{\mathsf{AR}} = \rho^{\mathsf{A}} \otimes \rho^{\mathsf{R}}$ holds for all purifications ρ^{ABR} of ρ^{AB}, and

(ii) $S(\rho^{\mathsf{AB}}) = S(\rho^{\mathsf{A}}) - S(\rho^{\mathsf{B}})$ if and only if $\rho^{\mathsf{BR}} = \rho^{\mathsf{B}} \otimes \rho^{\mathsf{R}}$ holds for all purifications ρ^{ABR} of ρ^{AB}.

The following proposition proves only part (i), but flipping the A and B yields part (ii). Note that $S(\rho^{\mathsf{AB}})$ must be nonnegative. Hence equality in (2.1) holds if and only if the condition in either (i) or (ii) holds.

Proposition 1. *Let ρ^{AB} be a bipartite state. Then the equality $S(\rho^{\mathsf{AB}}) = S(\rho^{\mathsf{B}}) - S(\rho^{\mathsf{A}})$ holds if and only if it holds that $\rho^{\mathsf{AR}} = \rho^{\mathsf{A}} \otimes \rho^{\mathsf{R}}$ for all possible purifications ρ^{ABR} of ρ^{AB}.*

Proof. Let ρ^{ABR} be a purification of ρ^{AB} such that $\rho^{\mathsf{ABR}} = |\psi\rangle\langle\psi|^{\mathsf{ABR}}$ for a pure state vector $|\psi\rangle^{\mathsf{ABR}}$ and

$$\rho^{\mathsf{AB}} = \mathrm{Tr}_{\mathsf{R}}\, \rho^{\mathsf{ABR}}.$$

Suppose that $\rho^{\mathsf{AR}} = \rho^{\mathsf{A}} \otimes \rho^{\mathsf{R}}$. Recall that the von Neumann entropy is sub-additive, $S(\rho^{\mathsf{AR}}) \leq S(\rho^{\mathsf{A}}) + S(\rho^{\mathsf{R}})$ with equality if and only if $\rho^{\mathsf{AR}} = \rho^{\mathsf{A}} \otimes \rho^{\mathsf{R}}$ (i.e., systems A and R are uncorrelated). It follows that

$$S(\rho^{\mathsf{AR}}) = S(\rho^{\mathsf{A}}) + S(\rho^{\mathsf{R}}) \tag{2.2}$$

from the assumption. Since the state $\rho^{\mathsf{ABR}} = |\psi\rangle\langle\psi|^{\mathsf{ABR}}$ is pure, it holds that

$$S(\rho^{\mathsf{R}}) = S(\rho^{\mathsf{AB}}) \qquad \text{and} \qquad S(\rho^{\mathsf{B}}) = S(\rho^{\mathsf{AR}}). \tag{2.3}$$

Putting together equations (2.2) and (2.3) yields the equality $S(\rho^{\mathsf{AB}}) = S(\rho^{\mathsf{B}}) - S(\rho^{\mathsf{A}})$.

For the converse, the exact same argument works in reverse. That is, if we suppose that $S(\rho^{\mathsf{AB}}) = S(\rho^{\mathsf{B}}) - S(\rho^{\mathsf{A}})$, we can make the same replacements as above to yield the equality $S(\rho^{\mathsf{AR}}) = S(\rho^{\mathsf{A}}) + S(\rho^{\mathsf{R}})$ for any possible purification ρ^{ABR}. This equality occurs if and only if $\rho^{\mathsf{AR}} = \rho^{\mathsf{A}} \otimes \rho^{\mathsf{R}}$, as desired. □

Example. The "trivial" examples of states ρ^{AB} that satisfy this equality in (2.1) are those for which either

- ρ^{AB} is pure, in which case $S(\rho^{\mathsf{AB}}) = 0$ and $S(\rho^{\mathsf{A}}) = S(\rho^{\mathsf{B}})$, or
- $\rho^{\mathsf{AB}} = \rho^{\mathsf{A}} \otimes \rho^{\mathsf{B}}$ and at least one of ρ^{A} or ρ^{B} is pure, in which case $S(\rho^{\mathsf{AB}}) = S(\rho^{\mathsf{A}})$ if ρ^{B} is pure or $S(\rho^{\mathsf{AB}}) = S(\rho^{\mathsf{B}})$ if ρ^{A} is pure.

We can construct a non-trivial example of a state ρ^{AB} that satisfies the equality in (2.1) in the following. Consider the pure state vectors $|u\rangle$ and $|v\rangle$ in $\mathbb{C}^2 \otimes \mathbb{C}^4$ defined by

$$|u\rangle^{\mathsf{AB}} = \frac{1}{\sqrt{2}}(|00\rangle^{\mathsf{AB}} + |11\rangle^{\mathsf{AB}}) \qquad \text{and} \qquad |v\rangle^{\mathsf{AB}} = \frac{1}{\sqrt{2}}(|02\rangle^{\mathsf{AB}} + |13\rangle^{\mathsf{AB}}).$$

Note that $\langle u|v\rangle = 0$. Define the following mixed state (which is certainly not pure)

$$\rho^{\mathsf{AB}} = \frac{1}{2}|u\rangle\langle u| + \frac{1}{2}|v\rangle\langle v|$$

and note that $S(\rho^{\mathsf{AB}}) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} = \log 2 = 1$. The reduced density operators on systems A and B are

$$\rho^{\mathsf{A}} = \frac{1}{2}\big(|0\rangle\langle 0| + |1\rangle\langle 1|\big) \qquad \text{and} \qquad \rho^{\mathsf{B}} = \frac{1}{4}\big(|0\rangle\langle 0| + |1\rangle\langle 1| + |2\rangle\langle 2| + |3\rangle\langle 3|\big),$$

both of which are clearly not pure. The entropies of these reduced density operators are $S(\rho^{\mathsf{A}}) = \log 2 = 1$ and $S(\rho^{\mathsf{B}}) = \log 4 = 2$. We see that

$$|S(\rho^{\mathsf{A}}) - S(\rho^{\mathsf{B}})| = |1 - 2| = 1 = S(\rho^{\mathsf{AB}}),$$

as desired.

3 Problem 3

Problem 3. Prove the strong sub-additivity of the Shannon entropy: For three random variables X, Y, Z,

$$H(X,Y,Z) + H(Y) \leq H(X,Y) + H(Y,Z)$$

where H is the Shannon entropy.

Solution. *Proof.* Define the mutual information of two variables X and Y as

$$I(X\colon Y) = H(X) + H(Y) - H(X,Y).$$

From the sub-additivity of the Shannon entropy, $H(X,Y) \leq H(X) + H(Y)$, it is clear that $I(X\colon Y) \geq 0$. Define the conditional mutual information as

$$I(X\colon Z \,|\, Y) = \sum_y p(y) I(X|_{Y=y}\colon Z|_{Y=y}).$$

Since both $p(y)$ and $I(X|_{Y=y}\colon Z|_{Y=y})$ are nonnegative for every y, it follows that $I(X\colon Z \,|\, Y) \geq 0$. Recall that $p(x|Y=y) = \frac{p(x,y)}{p(y)}$. By definition of the mutual information, we have

$$\begin{aligned}
p(y)I(X|_{Y=y}\colon Z|_{Y=y}) &= p(y)H(X|_{Y=y}) + p(y)H(Z|_{Y=y}) - p(y)H(X,Z|_{Y=y}) \\
&= -\sum_x p(x,y)\log\frac{p(x,y)}{p(y)} - \sum_z p(z,y)\log\frac{p(z,y)}{p(y)} + \sum_{x,z} p(x,y,z)\log\frac{p(x,y,z)}{p(y)} \\
&= -\sum_x p(x,y)\log p(x,y) - H(Y) - \sum_z p(z,y)\log p(z,y) - H(Y) \\
&\qquad + \sum_{x,z} p(x,y,z)\log p(x,y,z) + H(Y) \\
&= -\sum_x p(x,y)\log p(x,y) - \sum_z p(z,y)\log p(z,y) + \sum_{x,z} p(x,y,z)\log p(x,y,z) - H(Y).
\end{aligned}$$

By taking the sum over all $y's$ in the definition of $I(X\colon Z \,|\, Y)$, we have

$$\begin{aligned}
I(X\colon Z \,|\, Y) &= \sum_y p(y) I(X|_{Y=y}\colon Z|_{Y=y}) \\
&= H(X,Y) + H(Y,Z) - H(X,Y,Z) - H(Y).
\end{aligned}$$

Since $I(X\colon Z \,|\, Y) \geq 0$, from this it follows that

$$H(X,Y) + H(Y,Z) - H(X,Y,Z) - H(Y) \geq 0$$

and thus $H(X,Y,Z) + H(Y) \leq H(X,Y) + H(Y,Z)$ as desired. □

4 Problem 4

Problem 4. Consider an i.i.d. source characterized by a random variable X with alphabet $x \in \{1, \ldots, d\} = \mathcal{X}$ corresponding to probability $p(x) > 0$. Consider a sequence of size n, denoted as $x^n = (x_1, \ldots, x_n) \in \mathcal{X}^n$. The empirical distribution of the sequence x^n is defined by

$$q_{x^n}(x) = \frac{1}{n} N(x|x^n)$$

where $N(x|x^n)$ is the number of times the symbol $x \in \{1, \ldots, d\}$ appears in x^n. For $\delta > 0$ denote

$$T(n, \delta) = \left\{ x^n \in \mathcal{X}^n \,\middle|\, |q_{x^n}(x) - p(x)| < \delta \;\forall x \in \mathcal{X} \right\}.$$

(a) Show that for any $\epsilon, \delta > 0$ and sufficiently large n

$$\Pr(T(n, \delta)) \geq 1 - \epsilon.$$

(b) Show that for any $\epsilon, \delta > 0$ and sufficiently large n

$$(1 - \epsilon) 2^{n(H(X) - c\delta)} \leq |T(n, \delta)| \leq 2^{n(H(X) + c\delta)}$$

for some positive constant c.

(c) Show that if $x^n \in T(n, \delta)$ then

$$2^{-n(H(X)+c\delta)} \leq p(x^n) \leq 2^{-n(H(X)-c\delta)} \tag{4.1}$$

for some positive constant c, and $p(x^n) = p(x_1)p(x_2)\cdots p(x_n)$.

Solution. The idea for this solution is due to Wilde[1] and Yeung[2].

(a) Let $\epsilon, \delta > 0$. For each $x \in \mathcal{X}$, consider the i.i.d. indicator random variables $I_1(x), \ldots, I_k(x)$ obtained by sampling x_i and then setting

$$I_i(x) = \begin{cases} 1 \text{ if } x_i = x \\ 0 \text{ if } x_i \neq x \end{cases}.$$

We can write $N(x|x^n)$ as

$$N(x|x^n) = \sum_{i=1}^{n} I_i(x).$$

Since $\Pr(I_i(x) = 1) = p(x)$, we have the expected values $E[I_i(x)] = p(x)$ for all $x \in \mathcal{X}$. By the weak law of large numbers, for every $a \in \mathcal{X}$, there is a sufficiently large n_a such that

$$\Pr\left(\left\{\left|q_{x^n}(a) - p(a)\right| \geq \delta\right\}\right) = \Pr\left(\left\{\left|\frac{1}{n}\sum_{i=1}^{n} I_i(a) - p(a)\right| \geq \delta\right\}\right) < \frac{\epsilon}{d} \tag{4.2}$$

holds for all $n > n_a$, where $d = |\mathcal{X}|$. Let $n_0 = \max\{n_a \,|\, a \in \mathcal{X}\}$. Then for all $n > n_0$, we have

$$\begin{aligned} \Pr\left(\left\{\left|q_{x^n}(a) - p(a)\right| \geq \delta \text{ for some } a \in \mathcal{X}\right\}\right) &= \Pr\left(\left\{\left|\frac{1}{n}\sum_{i=1}^{n} I_i(a) - p(a)\right| \geq \delta \text{ for some } a \in \mathcal{X}\right\}\right) \\ &= \Pr\left(\bigcup_{a \in \mathcal{X}} \left\{\left|\frac{1}{n}\sum_{i=1}^{n} I_i(a) - p(a)\right| \geq \delta\right\}\right) \end{aligned}$$

[1]See Section 14.7.1 - 14.7.3 in *Classical to Quantum Shannon Theory* by Mark Wilde.
[2]See Chapter 6 in *Information Theory and Network Coding* by Raymond Yeung.

$$
\begin{aligned}
&\leq \sum_{a\in\mathcal{X}} \Pr\left(\left\{\left|\frac{1}{n}\sum_{i=1}^{n} I_i(a) - p(a)\right| \geq \delta\right\}\right) \\
&< \sum_{a\in\mathcal{X}} \frac{\epsilon}{d} = \epsilon,
\end{aligned}
$$

where the first inequality follows from the union bound[3] and the last inequality follows from (4.2). The probability of the complement of the above event is thus

$$
\Pr\left(\left\{\left|q_{x^n}(a) - p(a)\right| < \delta \text{ for all } a \in \mathcal{X}\right\}\right) = 1 - \epsilon
$$

for all $n > n_0$, as desired.

(b) We will prove part (b) from part (c). Hence, assuming part (c) is true, we can use the upper bound in (4.1) (i.e. $p(x^n) \leq 2^{-n(H(X)-c\delta)}$) to find that

$$
\Pr\big(T(n,\delta)\big) = \sum_{x^n \in T(n,\delta)} p(x^n) \leq |T(n,\delta)| 2^{-n(H(X)-c\delta)}
$$

for some positive constant c. For n sufficiently large, we have $1-\epsilon \leq \Pr\big(T(n,\delta)\big)$, and it follows from part (a) that

$$
1-\epsilon \leq \Pr\big(T(n,\delta)\big) \leq |T(n,\delta)| 2^{-n(H(X)-c\delta)}.
$$

Thus $(1-\epsilon)2^{n(H(X)-c\delta)} \leq |T(n,\delta)|$, which yields the desired lower bound. For the upper bound, we use the lower bound in (4.1) (i.e. $\leq 2^{-n(H(X)+c\delta)} \leq p(x^n)$). Now

$$
|T(n,\delta)| 2^{-n(H(X)+c\delta)} \leq \sum_{x^n \in T(n,\delta)} p(x^n) = \Pr\big(T(n,\delta)\big) \leq 1,
$$

and rearranging yields the desired upper bound $|T(n,\delta)| \leq 2^{n(H(X)+c\delta)}$. (Note that this upper bound holds for all n.)

(c) If $p(a) \neq 0$ for every $a \in \mathcal{X}$, for any sequence x^n we can write

$$
p(x^n) = \prod_{a\in\mathcal{X}} p(a)^{N(a|x^n)}.
$$

If $p(a) = 0$ for any $a \in \mathcal{X}$, then we must consider the subset $\mathcal{X}^+$ of $\mathcal{X}$ containing all a for which $p(a) > 0$. For n large enough, it must hold that

$$
p(x^n) = \prod_{a\in\mathcal{X}^+} p(a)^{N(a|x^n)}. \tag{4.3}
$$

for any $x^n \in T(n,\delta)$. That is, no symbols $a \in \mathcal{X}$ with probability $p(a) = 0$ can appear in typical sequences x^n if n is sufficiently large. Taking the logarithm of (4.3) yields

$$
\begin{aligned}
\log p(x^n) &= \sum_{a\in\mathcal{X}^+} N(a|x^n) \log p(a) \\
&= \sum_{a\in\mathcal{X}^+} \big(N(a|x^n) + np(a) - np(a)\big) \log p(a) \\
&= n \sum_{a\in\mathcal{X}^+} p(a)\log p(a) + n \sum_{a\in\mathcal{X}^+} \left(\frac{1}{n} N(a|x^n) - p(a)\right) \log p(a)
\end{aligned}
$$

[3]The union bound refers to the fact that $\Pr\left(\bigcup_i A_i\right) \leq \sum_i \Pr(A_i)$ for any set of events A_i.

$$= -n\left(H(X) + \sum_{a\in\mathcal{X}^+}\left(\frac{1}{n}N(a|x^n) - p(a)\right)\big(-\log p(a)\big)\right).$$

Since $x^n \in T(n,\delta)$, we know

$$\left|\frac{1}{n}N(a|x^n) - p(a)\right| \leq \delta$$

which implies

$$\begin{aligned}
\left|\sum_{a\in\mathcal{X}^+}\left(\frac{1}{n}N(a|x^n) - p(a)\right)\big(-\log p(a)\big)\right| &\leq \sum_{a\in\mathcal{X}^+}\left|\left(\frac{1}{n}N(a|x^n) - p(a)\right)\right|\big(-\log p(a)\big)\\
&\leq \delta\sum_{a\in\mathcal{X}^+}(-\log p(a))\\
&= \delta c
\end{aligned}$$

where we take c to be the positive constant

$$c = -\sum_{a\in\mathcal{X}^+}\log p(a).$$

From this we see that

$$-\delta c \leq \sum_{a\in\mathcal{X}^+}\left(\frac{1}{n}N(a|x^n) - p(a)\right)\big(-\log p(a)\big) \leq \delta c$$

and thus

$$-n(H(X)+\delta c) \leq \underbrace{-n\left(H(X) + \sum_{a\in\mathcal{X}^+}\left(\frac{1}{n}N(a|x^n) - p(a)\right)\log p(a)\right)}_{\log p(x^n)} \leq -n(H(X)-\delta c),$$

which yields

$$2^{-n(H(X)+\delta c)} \leq p(x^n) \leq 2^{-n(H(X)-\delta c)},$$

as desired.