16. **Probability of conditionally typical sequences.** In Chapter 7, we calculated the probability that two independently drawn sequences $X^n$ and $Y^n$ are weakly jointly typical. To prove the rate distortion theorem, however, we need to calculate this probability when one of the sequences is fixed and the other is random.

The techniques of weak typicality allow us only to calculate the average set size of the conditionally typical set. Using the ideas of strong typicality on the other hand provides us with stronger bounds which work for all typical $x^n$ sequences. We will outline the proof that $\Pr\{x^n , Y^n \in A^{(n)}_\epsilon \} \approx 2^{-nI(X;Y)}$ for all typical $x^n$. This approach was introduced by Berger[1] and is fully developed in the book by Csiszár and Körner[4].

Let $(X_i, Y_i)$ be drawn i.i.d. $\sim p(x,y)$. Let the marginals of $X$ and $Y$ be $p(x)$ and $p(y)$ respectively.

(a) Let $A^{(n)}_\epsilon$ be the strongly typical set for $X$. Show that

$$|A^{(n)}_\epsilon| \geq 2^{nH(X)} \quad (10.167)$$

*Hint: Theorem 11.1.1 and 11.1.3.*

(b) The **joint type** of a pair of sequences $(x^n, y^n)$ is the proportion of times $(x_i, y_i) = (a, b)$ in the pair of sequences, i.e.,

$$p_{x^n,y^n}(a, b) = \frac{1}{n} N(a, b|x^n, y^n) = \frac{1}{n} \sum_{i=1}^{n} I(x_i = a, y_i = b). \quad (10.168)$$

The **conditional type** of a sequence $y^n$ given $x^n$ is a stochastic matrix that gives the proportion of times a particular element of $Y$ occurred with each element of $X$ in the pair of sequences. Specifically, the conditional type $V^{(n)}_{y^n|x^n}(b|a)$ is defined as

$$V^{(n)}_{y^n|x^n}(b|a) = \frac{N(a, b|x^n, y^n)}{N(a|x^n)}. \quad (10.169)$$

Show that the number of conditional types is bounded by $(n + 1)^{||X||||Y||}$.

(c) The set of sequences $y^n \in \mathcal{Y}^n$ with conditional type $V$ with respect to a sequence $x^n$ is called the conditional type class $T_V(x^n)$. Show that

$$\frac{1}{(n + 1)^{||X||||Y||}} 2^{nH(Y|X)} \leq |T_V(x^n)| \leq 2^{nH(Y|X)}. \quad (10.170)$$

(d) The sequence $y^n \in \mathcal{Y}^n$ is said to be $\epsilon$-**strongly conditionally typical** with the sequence $x^n$ with respect to the conditional distribution $V(\cdot | \cdot)$ if the conditional type is close to $V$. The conditional type should satisfy the following two conditions:

i. For all $(a, b) \in X \times Y$ with $V(b|a) > 0$,

$$\frac{1}{n} \left| N(a, b|x^n, y^n) - V(b|a)N(a|x^n) \right| \leq \frac{\epsilon}{||Y|| + 1}. \quad (10.171)$$
ii. $N(a, b|x^n, y^n) = 0$ for all $(a, b)$ such that $V(b|a) = 0$.

The set of such sequences is called the conditionally typical set and is denoted $A^e_{\epsilon}(Y|x^n)$. Show that the number of sequences $y^n$ that are conditionally typical with a given $x^n \in X^n$ is bounded by

$$\frac{1}{(n+1)^{X||Y}}2^{n(H(Y|X)-\epsilon_1)} \leq |A^e_{\epsilon}(Y|x^n)| \leq (n + 1)^{|X||Y|}2^{n(H(Y|X)+\epsilon_1)}, \quad (10.172)$$

where $\epsilon_1 \to 0$ as $\epsilon \to 0$.

(e) For a pair of random variables $(X, Y)$ with joint distribution $p(x, y)$, the $\epsilon$-strongly typical set $A^s_{\epsilon}(n)$ is the set of sequences $(x^n, y^n) \in X^n \times Y^n$ satisfying

i. $N(a, b|x^n, y^n) = 0$ for all $(a, b) \in X \times Y$ with $p(a, b) = 0$.

The set of $\epsilon$-strongly jointly typical sequences is called the $\epsilon$-strongly jointly typical set and is denoted $A^s_{\epsilon}(n)(X, Y)$.

Let $(X, Y)$ be drawn i.i.d. $\sim p(x, y)$. For any $x^n$ such that there exists at least one pair $(x^n, y^n) \in A^s_{\epsilon}(n)(X, Y)$, the set of sequences $y^n$ such that $(x^n, y^n) \in A^s_{\epsilon}(n)$ satisfies

$$\frac{1}{(n+1)^{X||Y}}2^{n(H(Y|X)-\delta(\epsilon))} \leq |\{y^n : (x^n, y^n) \in A^s_{\epsilon}(n)\}| \leq (n + 1)^{|X||Y|}2^{n(H(Y|X)+\delta(\epsilon))}, \quad (10.174)$$

where $\delta(\epsilon) \to 0$ as $\epsilon \to 0$. In particular, we can write

$$2^{n(H(Y|X)-\epsilon_2)} \leq |\{y^n : (x^n, y^n) \in A^s_{\epsilon}(n)\}| \leq 2^{n(H(Y|X)+\epsilon_2)}, \quad (10.175)$$

where we can make $\epsilon_2$ arbitrarily small with an appropriate choice of $\epsilon$ and $n$.

(f) Let $Y_1, Y_2, \ldots, Y_n$ be drawn i.i.d. $\sim \prod p(y_i)$. For $x^n \in A^s_{\epsilon}(n)$, the probability that $(x^n, Y^n) \in A^s_{\epsilon}(n)$ is bounded by

$$2^{-n(I(X;Y)+\epsilon_3)} \leq \Pr((x^n, Y^n) \in A^s_{\epsilon}(n)) \leq 2^{-n(I(X;Y)-\epsilon_3)}, \quad (10.176)$$

where $\epsilon_3$ goes to 0 as $\epsilon \to 0$ and $n \to \infty$. 
Solution:

Probability of conditionally typical sequences.

(a) The set of strongly typical sequences is the set of sequence whose type is close the distribution $p$. We have two conditions - that the proportion of any symbol $a$ in the sequence is close to $p(a)$ and that no symbol with $p(a) = 0$ occurs in the sequence. The second condition may seem a technical one, but is essential in the proof of the strong equipartition theorem below.

By the strong law of large numbers, for a sequence drawn i.i.d. $\sim p(x)$, the asymptotic proportion of any letter $a$ is close to $p(a)$ with high probability. So for appropriately large $n$, the proportion of every letter is within $\epsilon$ of $p(a)$ with probability close to 1, i.e., the strongly typical set has a probability close to 1. We will show that

$$2^{n(H(p) - \epsilon')} \leq |A_{\epsilon}^{s(n)}| \leq 2^{n(H(p) + \epsilon')},$$  

(10.177)

where $\epsilon'$ goes to 0 as $\epsilon \to 0$ and $n \to \infty$.

For sequences in the strongly typical set,

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

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

(10.178)

and since $\left|\frac{1}{n} N(a|x^n) - p(a)\right| < \epsilon$ if $p(a) > 0$, and $\left|\frac{1}{n} N(a|x^n) - p(a)\right| = 0$ if $p(a) = 0$, we have

$$\left|-H(p) - \frac{1}{n} \log p(x^n)\right| < \epsilon_1.$$  

(10.179)

where $\epsilon_1 = \epsilon \sum_{a : p(a) > 0} \log \frac{1}{p(a)}$. It follows that $\epsilon_1 \to 0$ as $\epsilon \to 0$.

Recall the definition of weakly typical sequences in Chapter 3. A sequence was defined as $\epsilon_1$-weakly typical if $| - \log p(x^n) - H(p) | \leq \epsilon_1$. Hence a sequence that is $\epsilon$-strongly typical is also $\epsilon_1$-weakly typical. Hence the strongly typical set is a subset of the corresponding weakly typical set, i.e., $A^{s(n)}_\epsilon \subset A^{(n)}_{\epsilon_1}$.

Similarly, by the continuity of the entropy function, it follows that for all types in the typical set, the entropy of the type is close to $H(p)$. Specifically, for all $x^n \in A^{s(n)}_\epsilon$, $|p_{x^n}(a) - p(a)| < \epsilon$ and hence by Lemma 10.0.5, we have

$$|H(p_{x^n}) - H(p)| < \epsilon_2,$$  

(10.180)

where $\epsilon_2 = -|\mathcal{X}| \epsilon \log e \to 0$ as $\epsilon \to 0$.

There are only a polynomial number of types altogether and hence there are only a polynomial number of types in the strongly typical set. The type class of any type $q \in A^{s(n)}_\epsilon$, by Theorem 12.1.3, has a size bounded by

$$\frac{1}{(n + 1)|\mathcal{X}|} 2^{nH(q)} \leq |T(q)| \leq 2^{nH(q)}.$$  

(10.181)
By the previous part of this theorem, for \( q \in A^*_{\epsilon(n)} \), \( |H(q) - H(p)| \leq \epsilon_2 \), and hence

\[
\frac{1}{(n + 1)^{|X|}} 2^{n(H(p) - \epsilon_2)} \leq |T(q)| \leq 2^{n(H(p) + \epsilon_2)}. \tag{10.182}
\]

Since the number of elements in the strongly typical set is the sum of the sizes of the type classes in the strongly typical set, and there are only a polynomial number of them, we have

\[
\frac{1}{(n + 1)^{|X|}} 2^{n(H(p) - \epsilon_2)} \leq |A^*_{\epsilon(n)}| \leq (n + 1)^{|X|} 2^{n(H(p) + \epsilon_2)}, \tag{10.183}
\]

i.e., \( \frac{1}{n} \log |A^*_{\epsilon(n)}| - H(p) | \leq \epsilon' \), where \( \epsilon' = \epsilon_2 + \frac{|X|}{n} \log(n + 1) \) which goes to 0 as \( \epsilon \to 0 \) and \( n \to \infty \).

It is instructive to compare the proofs of the strong AEP with the AEP for weakly typical sequences. The results are similar, but there is one important difference. The lower bound on size of the strongly typical set does not depend on the probability of the set—instead, the bound is derived directly in terms of the size of type classes. This enables the lower bound in the strong AEP to be extended to conditionally typical sequences and sets; the weak AEP cannot be extended similarly. We will consider the extensions of the AEP to conditional distributions in the next part.
(b) The concept of types for single sequences can be extended to pairs of sequences for which we can define the concept of the joint type and the conditional type.

**Definition:** The joint type of a pair of sequences \((x^n, y^n)\) is the proportion of times a pair of symbols \((a, b)\) occurs jointly the the pair of sequences, i.e.,

\[
p_{x^n,y^n}(a,b) = \frac{1}{n}N(a,b|x^n,y^n).
\] (10.184)

**Definition:** The conditional type of a sequence \(y^n\) given \(x^n\) is a stochastic matrix that gives the proportion of times a particular element of \(Y\) occurred with each element of \(X\) in the pair of sequences. Specifically, the conditional type \(V_{y^n|x^n}(b|a)\) is defined as

\[
V_{y^n|x^n}(b|a) = \frac{N(a,b|x^n,y^n)}{N(a|x^n)},
\] (10.185)

The set of sequences \(y^n \in Y^n\) with conditional type \(V\) with respect to a sequence \(x^n\) is called the conditional type class \(T_V(x^n)\).

**Lemma 10.0.2** The number of conditional types for sequences of length \(n\) from the alphabet \(X\) and \(Y\) is bounded by \((n+1)^{|X||Y|}\).

**Proof:** By Theorem 12.1.1, the number of ways of choosing a row of the matrix \(V(\cdot|a)\) is bounded by \(n+1)^{|Y|}\) and there are \(|X|\) different choices of rows. So the total number of different conditional types is bounded by \((n+1)^{|X||Y|}\). □

(c) Since \(V_{y^n|x^n}\) is a stochastic matrix, we can multiply it with \(p_{x^n}\) to find the joint type of \((x^n, y^n)\). We will denote the conditional entropy of \(Y\) given \(X\) for this joint distribution as \(H(V_{y^n|x^n}|p_{x^n})\).

**Lemma 10.0.3** For \(x^n \in X^n\), let \(T_V(x^n)\) denote the set of sequences \(y^n \in Y^n\) with conditional type \(V\) with respect to \(x^n\). Then

\[
\frac{1}{(n+1)^{|X||Y|}}2^{nH(V|p_{x^n})} \leq |T_V(x^n)| \leq 2^{nH(V|p_{x^n})}.
\] (10.186)

**Proof:** This is a direct consequence of the corresponding lemma about the size of unconditional type classes. We can consider the subsequences of the pair corresponding each element of \(X\). For any particular element \(a \in X\), the number of conditionally typical sequences depends only the conditional type \(V(\cdot|a)\), and hence the number of conditionally typical sequences is bounded by

\[
\prod_{a \in X} \frac{1}{(N(a|x^n)+1)|Y|}2^{N(a|x^n)H(V|p_{x^n})} \leq |T_V(x^n)| \leq \prod_{a \in X} 2^{N(a|x^n)H(V|p_{x^n})}
\] (10.187)

which proves the lemma. □

The above two lemmas generalize the corresponding lemmas for unconditional types. We can use these to extend the strong AEP to conditionally typical sets.
(d) We begin with the definition of strongly conditionally typical sequences.

**Definition:** The sequence \( y^n \in \mathcal{Y}^n \) is said to be \( \epsilon \)-strongly conditionally typical with the sequence \( x^n \) with respect to the conditional distribution \( V(\cdot|\cdot) \) if the conditional type is close to \( V \). The conditional type should satisfy the following two conditions:

i. For all \( (a, b) \in \mathcal{X} \times \mathcal{Y} \) with \( V(b|a) > 0 \),
\[
\frac{1}{n} |N(a, b|x^n, y^n) - V(b|a)N(a|x^n)| \leq \epsilon. \tag{10.188}
\]

ii. \( N(a, b|x^n, y^n) = 0 \) for all \( (a, b) \) such that \( V(b|a) = 0 \).

The set of such sequences is called the conditionally typical set and is denoted \( A^{(n)}_{\epsilon}(\mathcal{Y}|x^n) \).

Essentially, a sequence \( y^n \) is conditionally typical with \( x^n \) if the subsequence of \( y^n \) corresponding to the occurrences of a particular symbol \( a \) in \( x^n \) is typical with respect to the conditional distribution \( V(\cdot|a) \). Since the number of such conditionally typical sequences is just the product of the number of subsequences conditionally typically corresponding to each choice of \( a \in \mathcal{X} \), we can now extend the strong AEP to derive a bound on the size of the conditionally typical set.

**Lemma 10.0.4** The number of sequences \( y^n \) that are conditionally typical with a given \( x^n \in \mathcal{X}^n \) is bounded by
\[
\frac{1}{(n + 1)|\mathcal{X}||\mathcal{Y}|} 2^{n(H(V|p_{x^n}) - \epsilon_4)} \leq |A^{(n)}_{\epsilon}(\mathcal{Y}|x^n)| \leq (n + 1)|\mathcal{X}||\mathcal{Y}| 2^{n(H(V|p_{x^n}) + \epsilon_4)}, \tag{10.189}
\]
where \( \epsilon_4 = -|\mathcal{X}||\mathcal{Y}|\epsilon \log \epsilon \to 0 \) as \( \epsilon \to 0 \).

**Proof:** Just as in the proof of the strong AEP (Theorem 12.2.1), we will derive the bounds using purely combinatorial arguments. The size of the conditional type class is bounded in Lemma 10.0.3 in terms of the entropy of the conditional type. By Lemma 10.0.5 and the definition of the conditionally typical set, we have

\[
|H(p_{y^n|x^n}|p_{x^n}) - H(V|p_{x^n})| \leq -|\mathcal{X}||\mathcal{Y}|\epsilon \log \epsilon \tag{10.190}
\]

Combining this with the bound on the number of conditional types (Lemma 10.0.2) we have the theorem. \( \square \)

(e) We now extend the definition of strongly typical sequences to pairs of sequences. The joint type of a pair of sequences is the proportion of occurrences of a pair of symbols together in the pair. A pair of sequences \( (x^n, y^n) \) is called jointly strongly typical with respect to a distribution \( p(x, y) \) if the joint type is close to \( p(x, y) \).

**Definition:** For a pair of random variables \( (X, Y) \) with joint distribution \( p(x, y) \), the \( \epsilon \)-strongly typical set \( A^{(n)}_{\epsilon}(n) \) is the set of sequences \( (x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n \) satisfying
i. \[ \frac{1}{n} N(a,b|x^n,y^n) - p(a,b) < \epsilon \] (10.191)

for every pair \((a,b) \in \mathcal{X} \times \mathcal{Y}\) with \(p(a,b) > 0\).

ii. \(N(a,b|x^n,y^n) = 0\) for all \((a,b) \in \mathcal{X} \times \mathcal{Y}\) with \(p(a,b) = 0\).

The set of \(\epsilon\)-strongly jointly typical sequences is called the \(\epsilon\)-strongly jointly typical set and is denoted \(A_{\epsilon}^{(n)}(X,Y)\).

**Theorem 10.0.1 (Joint AEP.)** Let \((X^n,Y^n)\) be sequences of length \(n\) drawn i.i.d. according to \(p(x^n,y^n) = \prod_{i=1}^{n} p(x_i,y_i)\). Then

\[ P(A_{\epsilon}^{(n)}) \to 1, \quad \text{as } n \to \infty. \] (10.192)

**Proof:** Follows directly from the weak law of large numbers. \(\square\)

From the definition, it is clear that strongly jointly typical sequences are also individually typical, i.e., for \(x^n\) such that \((x^n,y^n) \in A_{\epsilon}^{(n)}(X,Y)\),

\[ |p_{x^n}(a) - p(a)| \leq \sum_{b \in \mathcal{Y}} |p_{x^n,y^n}(a,b) - p(a,b)| \leq \epsilon|\mathcal{Y}|, \quad \text{for all } a \in \mathcal{X}. \] (10.193)

Hence \(x^n \in A_{\epsilon}^{(n)}(\mathcal{Y})\). This in turn implies that the pair is also conditionally typical for the conditional distribution \(p(y|x)\), i.e., for \((x^n,y^n) \in A_{\epsilon}^{(n)}(X,Y)\),

\[ |p_{x^n,y^n}(a,b) - p(b|a)p_{x^n}(a)| < \epsilon(|\mathcal{Y}| + 1) < \epsilon|\mathcal{X}||\mathcal{Y}|. \] (10.194)

Since conditional entropy is also a continuous function of the distribution, the conditional entropy of the type of a jointly strongly typical sequence, \(p_{x^n,y^n}\), is close to conditional entropy for \(p(x,y)\). Hence we can also extend Lemma 10.0.3 for elements of the typical set as follows:

**Theorem 10.0.2** (Size of conditionally typical set)

Let \((X,Y)\) be drawn i.i.d. \(\sim p(x,y)\). For any \(x^n\) such that there exists at least one pair \((x^n,y^n) \in A_{\epsilon}^{(n)}(X,Y)\), the set of sequences \(y^n\) such that \((x^n,y^n) \in A_{\epsilon}^{(n)}(X,Y)\) satisfies

\[ \frac{1}{(n+1)|\mathcal{X}||\mathcal{Y}|} \leq \frac{e^{n(H(X|Y) - \delta(\epsilon))}}{2^{e^{n(H(X|Y)+\delta(\epsilon))}}} \leq \frac{e^{n(H(Y|X) - \delta(\epsilon))}}{2^{e^{n(H(Y|X)+\delta(\epsilon))}}}. \]

where \(\delta(\epsilon) \to 0\) as \(\epsilon \to 0\). In particular, we can write

\[ 2^{e^{n(H(Y|X) - \epsilon_3)}} \leq |\{y^n : (x^n,y^n) \in A_{\epsilon}^{(n)}(X,Y)\}| \leq 2^{e^{n(H(Y|X)+\epsilon_3)}}, \] (10.196)

where we can make \(\epsilon_3\) arbitrarily small with an appropriate choice of \(\epsilon\) and \(n\).
Proof: The theorem follows from Theorem 10.0.2 and the continuity of conditional entropy as a function of the joint distribution. Now the set of sequences that are jointly typical with a given $x^n$ are also $\epsilon \times \mathcal{X} \times |\mathcal{Y}|$-strongly conditionally typical, and hence from the upper bound of Theorem 10.0.2, we have

$$|\{y^n : (x^n, y^n) \in A_{\epsilon}^{n(n)}\}| \leq (n + 1)|\mathcal{X}||\mathcal{Y}|2^{n(H(p(b|a)p_x^n) + \delta(\epsilon))},$$

(10.198)

where $\delta(\epsilon) \to 0$ as $\epsilon \to 0$. Now since

$$H(Y|X) = -\sum_x p(x) \sum_p(y|x) \log p(y|x)$$

(10.199)

is a linear function of the distribution $p(x)$, we have

$$|H(p(b|a)p_x^n) - H(Y|X)| \leq \epsilon|\mathcal{Y}| \max_{a \in \mathcal{X}} H(Y|X = a) \leq \epsilon|\mathcal{Y}| \log |\mathcal{Y}|,$$

(10.200)

which gives us the upper bound of the theorem.

For the lower bound, assume that $(x^n, y^n) \in A_{\epsilon}^{n(n)}(X,Y)$. Then since the joint type of a pair of sequences is determined by the type of $x^n$ and the conditional type of $y^n$ given $x^n$, all sequences $y^n$ with this conditional type will also be in $A_{\epsilon}^{n(n)}(X,Y)$. Hence the number of sequences $|\{y^n : (x^n, y^n) \in A_{\epsilon}^{n(n)}\}|$ is at least as much as the number of sequences of this conditional type, which by the lower bound of Lemma 10.0.4, and the continuity of conditional entropy as a function of the joint distribution (Lemma 10.0.5 and (10.200)), we have

$$|\{y^n : (x^n, y^n) \in A_{\epsilon}^{n(n)}\}| \geq (n + 1)|\mathcal{X}||\mathcal{Y}|2^{n(H(p(b|a)p_x^n) - \delta(\epsilon))},$$

(10.201)

where $\delta(\epsilon) \to 0$ as $\epsilon \to 0$. This gives us the theorem with

$$\epsilon_5 = \frac{|\mathcal{X}||\mathcal{Y}|}{n} \log(n + 1) + \epsilon|\mathcal{Y}| \log |\mathcal{Y}| - \epsilon|\mathcal{X}|^2|\mathcal{Y}|^2 \log \epsilon|\mathcal{X}||\mathcal{Y}|.$$  

(10.202)

To use this result, we have to assume that there is at least one $y^n$ such that $(x^n, y^n) \in A_{\epsilon}^{n(n)}(X,Y)$. From the definitions of the strongly typical sets, it is clear that if $|p_x^n(a) - p(a)| < \epsilon$, there exists at least one conditional distribution $\hat{p}(b|a)$ such that $|\hat{p}(b|a)p_x^n(a) - p(a, b)| < \epsilon$ and hence for large enough $n$, we have at least one conditional type such that $|p_x^n, y^n(a, b) - p(a, b)| \leq \epsilon$ and hence if $x^n$ is $\epsilon$-strongly typical, then there exists a conditional type such the joint type is jointly typical. For such an $x^n$ sequence, we can always find a $y^n$ such that $(x^n, y^n)$ is jointly typical.

(f) Notice that for the results of Theorems 10.0.2, we have used purely combinatorial arguments to bound the size of the conditionally type class and the conditionally typical set. These theorems illustrate the power of the method of types. We will now use the last theorem to bound the probability that a randomly chosen $Y^n$ will be conditionally typical with a given $x^n \in A_{\epsilon}^{n(n)}$. 
Theorem 10.0.3 Let $Y_1, Y_2, \ldots, Y_n$ be drawn i.i.d. $\sim \prod p(y)$. For $x^n \in A^*(n)$, the probability that $(x^n, Y^n) \in A^*(n)$ is bounded by

$$2^{-n(I(X;Y)+\epsilon)} \leq Pr((x^n, Y^n) \in A^*(n)) \leq 2^{-n(I(X;Y)-\epsilon)} \quad (10.203)$$

where $\epsilon = \epsilon_5 - \epsilon |\mathcal{X}||\mathcal{Y}| \log \epsilon |\mathcal{X}|$ which goes to 0 as $\epsilon \to 0$ and $n \to \infty$.

Proof: If $Y^n \in A^*(n)$, then $p(Y^n) \geq 2^{-n(H(Y))}$, and hence

$$P((x^n, Y^n) \in A^*(n)) = \sum_{y^n: (x^n, y^n) \in A^*(n)} p(y^n) \quad (10.204)$$

$$\leq \sum_{y^n: (x^n, y^n) \in A^*(n)} 2^{-n(H(Y)-\epsilon_6)} \quad (10.205)$$

$$= |A^*(n)(Y|x^n)|2^{-n(H(Y)-\epsilon)} \quad (10.206)$$

$$\leq 2^{n(H(Y|x)+\epsilon_5)}2^{-n(H(Y)-\epsilon_6)} \quad (10.207)$$

$$= 2^{-n(I(X;Y)-\epsilon)} \quad (10.208)$$

where $\epsilon_6 = -\epsilon |\mathcal{X}||\mathcal{Y}| \log \epsilon |\mathcal{X}|$ since $|p_{y^n} - p| \leq \epsilon |\mathcal{X}|$ if $(x^n, y^n) \in A^*(n)$.

Also

$$P((x^n, Y^n) \in A^*(n)) = \sum_{y^n: (x^n, y^n) \in A^*(n)} p(y^n) \quad (10.209)$$

$$\geq \sum_{y^n: (x^n, y^n) \in A^*(n)} 2^{-n(H(Y)+\epsilon_6)} \quad (10.210)$$

$$= |A^*(n)(Y|x^n)|2^{-n(H(Y)+\epsilon_6)} \quad (10.211)$$

$$\geq 2^{n(H(Y|x)-\epsilon_5)}2^{-n(H(Y)+\epsilon_6)} \quad (10.212)$$

$$= 2^{-n(I(X;Y)+\epsilon)} \quad (10.213)$$

Hence

$$2^{-n(I(X;Y)+\epsilon)} \leq Pr((x^n, Y^n) \in A^*(n)) \leq 2^{-n(I(X;Y)-\epsilon)}. \quad (10.214)$$

\[ \square \]

The main result of this problem is the last theorem, which gives upper and lower bounds on the probability that a randomly chosen sequence $y^n$ will be jointly typical with a given $x^n$. This was used in the proof of the rate distortion theorem.

To end this solution, we will prove a theorem on the continuity of entropy:

Lemma 10.0.5 If $|p(x) - q(x)| \leq \epsilon$ for all $x$, then $|H(p) - H(q)| \leq -\epsilon |\mathcal{X}| \log \epsilon$.

Proof: We will use some simple properties of the function

$$f(x) = -x \ln x \quad \text{for} \ 0 \leq x \leq \frac{1}{e}. \quad (10.215)$$
Since $f'(x) = -1 - \ln x > 0$ and $f''(x) = -\frac{1}{x}$, $f(x)$ is an increasing concave function. Now consider

$$g(x) = f(x + \epsilon) - f(x) = x \ln x - (x + \epsilon) \ln(x + \epsilon). \quad (10.216)$$

Then again by differentiation, it is clear that $g'(x) < 0$ so the function is strictly decreasing. Hence $g(x) < g(0) = -\epsilon \ln \epsilon$ for all $x$.

For any $a \in \mathcal{X}$, assume $p(a) > q(a)$, and hence we have

$$p(a) - q(a) \leq \epsilon \quad (10.217)$$

Hence by the fact that $f$ is an increasing function, we have

$$| - p(a) \ln p(a) + q(a) \ln q(a) | = -p(a) \ln p(a) + q(a) \ln q(a) \quad (10.218)$$

$$\leq - (q(a) + \epsilon) \ln(q(a) + \epsilon) + q(a) \ln q(a)$$

$$\leq -\epsilon \ln \epsilon. \quad (10.219)$$

Summing this over all $a \in \mathcal{X}$, we have the lemma. □