Information Theory I

Fall Semester 2017

Prof. Dr. A. Lapidoth

Signal and Information Processing Laboratory
Institut für Signal- und Informationsverarbeitung

Strong Typicality and Covering Lemma

http://www.isi.ee.ethz.ch/teaching/courses/it1.html

Let P be a probability mass function on a finite set $\mathcal{X}$. For a fixed $\epsilon \in (0,1)$, the set of strongly typical x-sequences of length n is defined as

$$\mathcal{T}_\epsilon^{(n)}(P) \triangleq \left\{ x^n \in \mathcal{X}^n \;\middle|\; \forall \xi \in \mathcal{X} \colon \left|\tfrac{1}{n} N(\xi|x^n) - P(\xi)\right| \leq \epsilon P(\xi) \right\}, \tag{1}$$

where $N(\xi|x^n)$ denotes the number of occurences of the symbol ξ in the sequence x^n. Using $\delta \triangleq \epsilon H(P)$, the set $\mathcal{T}_\epsilon^{(n)}(P)$ has the following properties:

1) For all $x^n \in \mathcal{T}_\epsilon^{(n)}(P)$, $2^{-n(H(P)+\delta)} \leq \prod_{i=1}^{n} P(x_i) \leq 2^{-n(H(P)-\delta)}$.
2) $\left|\mathcal{T}_\epsilon^{(n)}(P)\right| \leq 2^{n(H(P)+\delta)}$.
3) For sufficiently large n, $\left|\mathcal{T}_\epsilon^{(n)}(P)\right| \geq (1-\epsilon)\, 2^{n(H(P)-\delta)}$.
4) For X^n IID $\sim P$, $\lim_{n\to\infty} \Pr\left[X^n \in \mathcal{T}_\epsilon^{(n)}(P)\right] = 1$.

Let P_{XY} be a joint probability mass function on finite sets $\mathcal{X}$ and $\mathcal{Y}$. For a fixed $\epsilon \in (0,1)$, the set of jointly typical sequences of length n is defined as

$$\begin{aligned}\mathcal{T}_\epsilon^{(n)}(P_{XY}) \triangleq \bigl\{ (x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n \;\bigm|\; \\ \forall (\xi,\gamma) \in \mathcal{X} \times \mathcal{Y} \colon \left|\tfrac{1}{n} N((\xi,\gamma)|x^n,y^n) - P_{XY}(\xi,\gamma)\right| \leq \epsilon P_{XY}(\xi,\gamma) \bigr\}.\end{aligned} \tag{2}$$

Using $\delta_{XY} \triangleq \epsilon H(P_{XY})$ and denoting the marginal probability mass functions of P_{XY} by P_X and P_Y, the set $\mathcal{T}_\epsilon^{(n)}(P_{XY})$ has the following properties:

5) For all $(x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}(P_{XY})$, $x^n \in \mathcal{T}_\epsilon^{(n)}(P_X)$ and $y^n \in \mathcal{T}_\epsilon^{(n)}(P_Y)$.
6) For all $(x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}(P_{XY})$, $2^{-n(H(P_{XY})+\delta_{XY})} \leq \prod_{i=1}^{n} P_{XY}(x_i, y_i) \leq 2^{-n(H(P_{XY})-\delta_{XY})}$.
7) $\left|\mathcal{T}_\epsilon^{(n)}(P_{XY})\right| \leq 2^{n(H(P_{XY})+\delta_{XY})}$.
8) For sufficiently large n, $\left|\mathcal{T}_\epsilon^{(n)}(P_{XY})\right| \geq (1-\epsilon)\, 2^{n(H(P_{XY})-\delta_{XY})}$.
9) For (X^n, Y^n) IID $\sim P_{XY}$, $\lim_{n\to\infty} \Pr\left[(X^n, Y^n) \in \mathcal{T}_\epsilon^{(n)}(P_{XY})\right] = 1$.

Let $I(X;Y)$ be with respect to P_{XY}. For a fixed ϵ' satisfying $0 < \epsilon' < \epsilon$ and sufficiently large n, i.e., for $n \geq n_0$, where n_0 depends only on P_{XY}, ϵ, and ϵ', the following holds:

10) For all $x^n \in \mathcal{T}_{\epsilon'}^{(n)}(P_X)$ and for Y^n IID $\sim P_Y$, $\Pr\left[(x^n, Y^n) \in \mathcal{T}_\epsilon^{(n)}(P_{XY})\right] \geq 2^{-n(I(X;Y)+4\delta_{XY})}$.

Lemma 1. *For all* $x^n \in \mathcal{T}_\epsilon^{(n)}(P)$, $2^{-n(H(P)+\delta)} \leq \prod_{i=1}^{n} P(x_i) \leq 2^{-n(H(P)-\delta)}$.

Proof. Fix $x^n \in \mathcal{T}_\epsilon^{(n)}(P)$. Observe that by (1), $N(\xi|x^n) = 0$ for all $\xi \in \mathcal{X}$ with $P(\xi) = 0$, which implies $\prod_{i=1}^{n} P(x_i) > 0$. Defining $\mathcal{X}' \triangleq \operatorname{supp}(P)$, we have

$$\begin{aligned}
-\frac{1}{n}\log\prod_{i=1}^{n} P(x_i) &= -\frac{1}{n}\sum_{i=1}^{n}\log P(x_i) \\
&= -\frac{1}{n}\sum_{i=1}^{n}\sum_{\xi\in\mathcal{X}'} \mathrm{I}\{x_i = \xi\}\log P(\xi) \\
&= -\frac{1}{n}\sum_{\xi\in\mathcal{X}'}\sum_{i=1}^{n} \mathrm{I}\{x_i = \xi\}\log P(\xi) \\
&= -\frac{1}{n}\sum_{\xi\in\mathcal{X}'} N(\xi|x^n)\log P(\xi) \\
&= \sum_{\xi\in\mathcal{X}'} \frac{1}{n} N(\xi|x^n)\log\frac{1}{P(\xi)} \\
&\overset{\text{(i)}}{\leq} \sum_{\xi\in\mathcal{X}'} (1+\epsilon)P(\xi)\log\frac{1}{P(\xi)} \\
&= (1+\epsilon)\, H(P) \\
&= H(P) + \delta,
\end{aligned}$$

where (i) holds because $N(\xi|x^n) \leq n(1+\epsilon)P(\xi)$ by (1) and because $\log\frac{1}{P(\xi)} \geq 0$ for all $\xi \in \mathcal{X}'$. This implies $\prod_{i=1}^{n} P(x_i) \geq 2^{-n(H(P)+\delta)}$. Similarly, using $N(\xi|x^n) \geq n(1-\epsilon)P(\xi)$ in (i), we obtain

$$-\frac{1}{n}\log\prod_{i=1}^{n} P(x_i) \geq H(P) - \delta$$

and $\prod_{i=1}^{n} P(x_i) \leq 2^{-n(H(P)-\delta)}$. ∎

Lemma 2. $\left|\mathcal{T}_\epsilon^{(n)}(P)\right| \leq 2^{n(H(P)+\delta)}$.

Proof. We have

$$\begin{aligned}
1 &= \sum_{x^n\in\mathcal{X}^n}\prod_{i=1}^{n} P(x_i) \\
&\geq \sum_{x^n\in\mathcal{T}_\epsilon^{(n)}(P)}\prod_{i=1}^{n} P(x_i) \\
&\overset{\text{(i)}}{\geq} \sum_{x^n\in\mathcal{T}_\epsilon^{(n)}(P)} 2^{-n(H(P)+\delta)} \\
&= \left|\mathcal{T}_\epsilon^{(n)}(P)\right| \cdot 2^{-n(H(P)+\delta)},
\end{aligned}$$

where (i) follows from Lemma 1. Multiplication by $2^{n(H(P)+\delta)}$ leads to the desired inequality. ∎

Lemma 3. *For sufficiently large n, $\left|\mathcal{T}_\epsilon^{(n)}(P)\right| \geq (1-\epsilon)\, 2^{n(H(P)-\delta)}$.*

Proof. Let X^n be IID $\sim P$. For sufficiently large n,

$$\begin{aligned} 1-\epsilon &\overset{\text{(i)}}{\leq} \Pr\bigl[X^n \in \mathcal{T}_\epsilon^{(n)}(P)\bigr] \\ &= \sum_{x^n \in \mathcal{T}_\epsilon^{(n)}(P)} \Pr[X^n = x^n] \\ &= \sum_{x^n \in \mathcal{T}_\epsilon^{(n)}(P)} \prod_{i=1}^{n} P(x_i) \\ &\overset{\text{(ii)}}{\leq} \sum_{x^n \in \mathcal{T}_\epsilon^{(n)}(P)} 2^{-n(H(P)-\delta)} \\ &= \left|\mathcal{T}_\epsilon^{(n)}(P)\right| \cdot 2^{-n(H(P)-\delta)}, \end{aligned}$$

where (i) follows from Lemma 4 for sufficiently large n; and (ii) follows from Lemma 1. Multiplication by $2^{n(H(P)-\delta)}$ leads to the desired inequality. ∎

Lemma 4. *For X^n IID $\sim P$, $\lim_{n\to\infty} \Pr\bigl[X^n \in \mathcal{T}_\epsilon^{(n)}(P)\bigr] = 1$.*

Proof. This follows from the weak law of large numbers and the union bound. For convenience, we provide a proof with explicit constants. Define

$$C \triangleq \min_{x \in \operatorname{supp}(P)} P(x) > 0. \tag{3}$$

We show below that for all $\xi \in \mathcal{X}$,

$$\Pr\left[\left|\frac{1}{n} N(\xi|X^n) - P(\xi)\right| > \epsilon P(\xi)\right] \leq \frac{1}{n\epsilon^2 C^2}. \tag{4}$$

Consequently,

$$\begin{aligned} \Pr\bigl[X^n \notin \mathcal{T}_\epsilon^{(n)}(P)\bigr] &= \Pr\left[\exists \xi \in \mathcal{X} \colon \left|\frac{1}{n} N(\xi|X^n) - P(\xi)\right| > \epsilon P(\xi)\right] \\ &\overset{\text{(i)}}{\leq} \sum_{\xi \in \mathcal{X}} \Pr\left[\left|\frac{1}{n} N(\xi|X^n) - P(\xi)\right| > \epsilon P(\xi)\right] \\ &\overset{\text{(ii)}}{\leq} \sum_{\xi \in \mathcal{X}} \frac{1}{n\epsilon^2 C^2} \\ &= \frac{|\mathcal{X}|}{n\epsilon^2 C^2}, \end{aligned} \tag{5}$$

where (i) follows from the union bound; and (ii) follows from (4). Since the RHS of (5) tends to zero as n tends to infinity, $\Pr\bigl[X^n \notin \mathcal{T}_\epsilon^{(n)}(P)\bigr]$ must also tend to zero as n tends to infinity, which proves the claim.

It remains to show (4). For all $\xi \in \mathcal{X}$ with $P(\xi) = 0$, the LHS of (4) is zero, so (4) holds in these cases. Now, fix $\xi \in \mathcal{X}$ with $P(\xi) > 0$. Introduce indicator random variables $Z_1, \ldots, Z_n$ as follows:

$$Z_i \triangleq \begin{cases} 1 & \text{if } X_i = \xi, \\ 0 & \text{otherwise.} \end{cases}$$

Define $Z \triangleq \sum_{i=1}^{n} Z_i$, so $Z = N(\xi|X^n)$. By the linearity of the expectation,

$$\mu_Z \triangleq \mathsf{E}[Z] = \sum_{i=1}^{n} \mathsf{E}[Z_i] = \sum_{i=1}^{n} P(\xi) = nP(\xi).$$

Now,

$$\begin{aligned}
\Pr\left[\left|\frac{1}{n}N(\xi|X^n) - P(\xi)\right| > \epsilon P(\xi)\right] &= \Pr[|Z - nP(\xi)| > n\epsilon P(\xi)] \\
&= \Pr[|Z - \mu_Z| > n\epsilon P(\xi)] \\
&\overset{\text{(i)}}{\leq} \frac{\mathsf{Var}(Z)}{n^2\epsilon^2 P(\xi)^2} \\
&\overset{\text{(ii)}}{\leq} \frac{1}{n\epsilon^2 P(\xi)^2} \\
&\overset{\text{(iii)}}{\leq} \frac{1}{n\epsilon^2 C^2},
\end{aligned}$$

where (i) follows from Chebyshev's inequality (see Exercise 1, Problem 4); (ii) holds since $Z_1, \ldots, Z_n$ are independent, so

$$\mathsf{Var}(Z) = \mathsf{Var}\left(\sum_{i=1}^{n} Z_i\right) = \sum_{i=1}^{n} \mathsf{Var}(Z_i) = \sum_{i=1}^{n} P(\xi)(1 - P(\xi)) \leq \sum_{i=1}^{n} 1 = n;$$

and (iii) follows from (3). ■

Lemma 5. *For all* $(x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}(P_{XY})$, $x^n \in \mathcal{T}_\epsilon^{(n)}(P_X)$ *and* $y^n \in \mathcal{T}_\epsilon^{(n)}(P_Y)$.

Proof. Fix $(x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}(P_{XY})$. For all $\xi \in \mathcal{X}$,

$$\begin{aligned}
N(\xi|x^n) &= \sum_{\gamma \in \mathcal{Y}} N((\xi, \gamma)|x^n, y^n) \\
&\overset{\text{(i)}}{\leq} \sum_{\gamma \in \mathcal{Y}} n(1+\epsilon) P_{XY}(\xi, \gamma) \\
&= n(1+\epsilon) P_X(\xi),
\end{aligned}$$

where (i) follows from (2). Similarly, $N(\xi|x^n) \geq n(1-\epsilon)P_X(\xi)$ for all $\xi \in \mathcal{X}$. Together, this implies $\left|\frac{1}{n}N(\xi|x^n) - P_X(\xi)\right| \leq \epsilon P_X(\xi)$ for all $\xi \in \mathcal{X}$, so $x^n \in \mathcal{T}_\epsilon^{(n)}(P_X)$. Analogously, $y^n \in \mathcal{T}_\epsilon^{(n)}(P_Y)$. ■

Lemma 6. *For all* $(x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}(P_{XY})$, $2^{-n(H(P_{XY})+\delta_{XY})} \leq \prod_{i=1}^{n} P_{XY}(x_i, y_i) \leq 2^{-n(H(P_{XY})-\delta_{XY})}$.

Lemma 7. $\left|\mathcal{T}_\epsilon^{(n)}(P_{XY})\right| \leq 2^{n(H(P_{XY})+\delta_{XY})}$.

Lemma 8. *For sufficiently large* n, $\left|\mathcal{T}_\epsilon^{(n)}(P_{XY})\right| \geq (1-\epsilon)\, 2^{n(H(P_{XY})-\delta_{XY})}$.

Lemma 9. *For* (X^n, Y^n) *IID* $\sim P_{XY}$, $\lim_{n\to\infty} \Pr\left[(X^n, Y^n) \in \mathcal{T}_\epsilon^{(n)}(P_{XY})\right] = 1$.

Proof of Lemmas 6–9. They follow from applying Lemmas 1–4 to the set $\mathcal{X}' \triangleq (\mathcal{X} \times \mathcal{Y})$. ■

Lemma 10. *For a fixed ϵ' satisfying $0 < \epsilon' < \epsilon$ and sufficiently large n, the following holds: For all $x^n \in \mathcal{T}_{\epsilon'}^{(n)}(P_X)$ and for Y^n IID $\sim P_Y$, $\Pr\bigl[(x^n, Y^n) \in \mathcal{T}_{\epsilon}^{(n)}(P_{XY})\bigr] \geq 2^{-n(I(X;Y)+4\delta_{XY})}$.*

Proof. We have

$$\begin{aligned}
\Pr\bigl[(x^n, Y^n) \in \mathcal{T}_{\epsilon}^{(n)}(P_{XY})\bigr] &= \sum_{y^n:\,(x^n,y^n)\in\mathcal{T}_{\epsilon}^{(n)}(P_{XY})} \Pr[Y^n = y^n] \\
&= \sum_{y^n:\,(x^n,y^n)\in\mathcal{T}_{\epsilon}^{(n)}(P_{XY})} \prod_{i=1}^{n} P_Y(y_i) \\
&\overset{\text{(i)}}{\geq} \sum_{y^n:\,(x^n,y^n)\in\mathcal{T}_{\epsilon}^{(n)}(P_{XY})} 2^{-n(H(P_Y)+\epsilon H(P_Y))} \\
&\overset{\text{(ii)}}{\geq} \sum_{y^n:\,(x^n,y^n)\in\mathcal{T}_{\epsilon}^{(n)}(P_{XY})} 2^{-n(H(P_Y)+\delta_{XY})} \\
&= \bigl|\{\, y^n \in \mathcal{Y}^n : (x^n, y^n) \in \mathcal{T}_{\epsilon}^{(n)}(P_{XY}) \,\}\bigr| \cdot 2^{-n(H(P_Y)+\delta_{XY}))} \\
&\overset{\text{(iii)}}{\geq} 2^{n(H(P_{XY})-H(P_X)-3\delta_{XY})} \cdot 2^{-n(H(P_Y)+\delta_{XY}))} \\
&= 2^{-n(H(P_X)+H(P_Y)-H(P_{XY})+4\delta_{XY})} \\
&= 2^{-n(I(X;Y)+4\delta_{XY})},
\end{aligned}$$

where (i) follows from Lemma 1 because $y^n \in \mathcal{T}_{\epsilon}^{(n)}(P_Y)$ by Lemma 5; (ii) holds because $H(P_Y) \leq H(P_{XY})$; and (iii) follows from Lemma 12 below. ■

Lemma 11. *For a fixed ϵ' satisfying $0 < \epsilon' < \epsilon$ and sufficiently large n, the following holds: For all $x^n \in \mathcal{T}_{\epsilon'}^{(n)}(P_X)$ and for $Y^n \sim \prod_{i=1}^{n} P_{Y|X}(y_i|x_i)$, $\lim_{n\to\infty} \Pr\bigl[(x^n, Y^n) \in \mathcal{T}_{\epsilon}^{(n)}(P_{XY})\bigr] = 1$.*

Proof. Define

$$C \triangleq \min_{(x,y)\in\operatorname{supp}(P_{XY})} P_{XY}(x, y) > 0. \tag{6}$$

We show below that for all $x^n \in \mathcal{T}_{\epsilon'}^{(n)}(P_X)$ and for all $(\xi, \gamma) \in \mathcal{X} \times \mathcal{Y}$,

$$\Pr\left[\left|\frac{1}{n}N((\xi,\gamma)|x^n, Y^n) - P_{XY}(\xi,\gamma)\right| > \epsilon P_{XY}(\xi,\gamma)\right] \leq \frac{1}{n(\epsilon-\epsilon')^2 C^2}. \tag{7}$$

Consequently, for all $x^n \in \mathcal{T}_{\epsilon'}^{(n)}(P_X)$,

$$\begin{aligned}
\Pr\bigl[(x^n, Y^n) \notin \mathcal{T}_{\epsilon}^{(n)}(P_{XY})\bigr] &= \Pr\left[\exists(\xi,\gamma) \in \mathcal{X}\times\mathcal{Y}\colon \left|\frac{1}{n}N((\xi,\gamma)|x^n, Y^n) - P_{XY}(\xi,\gamma)\right| > \epsilon P_{XY}(\xi,\gamma)\right] \\
&\overset{\text{(i)}}{\leq} \sum_{(\xi,\gamma)\in\mathcal{X}\times\mathcal{Y}} \Pr\left[\left|\frac{1}{n}N((\xi,\gamma)|x^n, Y^n) - P_{XY}(\xi,\gamma)\right| > \epsilon P_{XY}(\xi,\gamma)\right] \\
&\overset{\text{(ii)}}{\leq} \sum_{(\xi,\gamma)\in\mathcal{X}\times\mathcal{Y}} \frac{1}{n(\epsilon-\epsilon')^2 C^2} \\
&= \frac{|\mathcal{X}|\cdot|\mathcal{Y}|}{n(\epsilon-\epsilon')^2 C^2},
\end{aligned} \tag{8}$$

where (i) follows from the union bound; and (ii) follows from (7). Since the RHS of (8) tends to zero as n tends to infinity, $\Pr\bigl[(x^n, Y^n) \notin \mathcal{T}_{\epsilon}^{(n)}(P_{XY})\bigr]$ must also tend to zero as n tends to infinity, which proves the claim.

It remains to show (7). Fix $x^n \in \mathcal{T}_{\epsilon'}^{(n)}(P_X)$, $\xi \in \mathcal{X}$, and $\gamma \in \mathcal{Y}$. We first treat the case $P_{XY}(\xi,\gamma)=0$.

- If $P_X(\xi)=0$, then $N(\xi|x^n)=0$, so $\Pr[N((\xi,\gamma)|x^n,Y^n)=0]=1$;
- otherwise, $P_X(\xi)>0$ and $P_{Y|X}(\gamma|\xi)=0$, so $\Pr[N((\xi,\gamma)|x^n,Y^n)=0]=1$, too.

Since $\Pr[N((\xi,\gamma)|x^n,Y^n)=0]=1$, the LHS of (7) is zero, so (7) holds in this case. Assume from now on that $P_{XY}(\xi,\gamma)>0$, which implies $P_X(\xi)>0$. Define the set $\mathcal{A} \triangleq \{i \in \{1,\dots,n\} \mid x_i=\xi\}$. Define $k \triangleq |\mathcal{A}|$. Observe that $k=N(\xi|x^n)$. By Lemma 5, $x^n \in \mathcal{T}_{\epsilon'}^{(n)}(P_X)$, so

$$n(1-\epsilon')P_X(\xi) \le k \le n(1+\epsilon')P_X(\xi). \tag{9}$$

In particular, $k>0$ because $P_X(\xi)>0$ and $0<\epsilon'<\epsilon<1$. Label the elements of $\mathcal{A}$ by $\alpha_1,\dots,\alpha_k$. Introduce indicator random variables $Z_1,\dots,Z_k$ as follows:

$$Z_i \triangleq \begin{cases} 1 & \text{if } Y_{\alpha_i}=\gamma, \\ 0 & \text{otherwise.} \end{cases}$$

Define $Z \triangleq \sum_{i=1}^k Z_i$, so $Z=N((\xi,\gamma)|x^n,Y^n)$. By the linearity of the expectation,

$$\mu_Z \triangleq \mathsf{E}[Z] = \sum_{i=1}^k \mathsf{E}[Z_i] = \sum_{i=1}^k P_{Y|X}(\gamma|\xi) = kP_{Y|X}(\gamma|\xi).$$

Observe that

$$\begin{aligned} \mu_Z - nP_{XY}(\xi,\gamma) &= kP_{Y|X}(\gamma|\xi) - nP_{XY}(\xi,\gamma) \\ &\overset{\text{(i)}}{\le} n(1+\epsilon')P_X(\xi)P_{Y|X}(\gamma|\xi) - nP_{XY}(\xi,\gamma) \\ &= n(1+\epsilon')P_{XY}(\xi,\gamma) - nP_{XY}(\xi,\gamma) \\ &= n\epsilon' P_{XY}(\xi,\gamma), \end{aligned}$$

where (i) follows from (9). Similarly, we obtain $\mu_Z - nP_{XY}(\xi,\gamma) \ge -n\epsilon' P_{XY}(\xi,\gamma)$. Together, these inequalities imply

$$|\mu_Z - nP_{XY}(\xi,\gamma)| \le n\epsilon' P_{XY}(\xi,\gamma). \tag{10}$$

Observe that if $|Z-\mu_Z| \le n(\epsilon-\epsilon')P_{XY}(\xi,\gamma)$, then

$$\begin{aligned} |Z - nP_{XY}(\xi,\gamma)| &= |Z-\mu_Z+\mu_Z-nP_{XY}(\xi,\gamma)| \\ &\overset{\text{(i)}}{\le} |Z-\mu_Z| + |\mu_Z - nP_{XY}(\xi,\gamma)| \\ &\overset{\text{(ii)}}{\le} n(\epsilon-\epsilon')P_{XY}(\xi,\gamma) + n\epsilon' P_{XY}(\xi,\gamma) \\ &= n\epsilon P_{XY}(\xi,\gamma), \end{aligned}$$

where (i) follows from the triangle inequality; and (ii) follows from the above condition and from (10). Now,

$$\begin{aligned} \Pr\left[\left|\frac{1}{n}N((\xi,\gamma)|x^n,Y^n) - P_{XY}(\xi,\gamma)\right| > \epsilon P_{XY}(\xi,\gamma)\right] &= \Pr[|Z-nP_{XY}(\xi,\gamma)| > n\epsilon P_{XY}(\xi,\gamma)] \\ &\overset{\text{(i)}}{\le} \Pr[|Z-\mu_Z| > n(\epsilon-\epsilon')P_{XY}(\xi,\gamma)] \\ &\overset{\text{(ii)}}{\le} \frac{\mathsf{Var}(Z)}{n^2(\epsilon-\epsilon')^2 P_{XY}(\xi,\gamma)^2} \end{aligned}$$

$$\overset{\text{(iii)}}{\leq} \frac{1}{n(\epsilon-\epsilon')^2 P_{XY}(\xi,\gamma)^2}$$
$$\overset{\text{(iv)}}{\leq} \frac{1}{n(\epsilon-\epsilon')^2 C^2},$$

where (i) follows from the above observation; (ii) follows from Chebyshev's inequality; (iii) holds because $Z_1, \ldots, Z_k$ are independent, so

$$\mathsf{Var}(Z) = \mathsf{Var}\left(\sum_{i=1}^{k} Z_i\right) = \sum_{i=1}^{k} \mathsf{Var}(Z_i) = \sum_{i=1}^{k} P_{Y|X}(\gamma|\xi)(1 - P_{Y|X}(\gamma|\xi)) \leq \sum_{i=1}^{k} 1 = k \leq n;$$

and (iv) follows from (6). ■

Lemma 12. *For a fixed ϵ' satisfying $0 < \epsilon' < \epsilon$ and sufficiently large n, the following holds: For all $x^n \in \mathcal{T}_{\epsilon'}^{(n)}(P_X)$, $\left|\left\{ y^n \in \mathcal{Y}^n : (x^n, y^n) \in \mathcal{T}_{\epsilon}^{(n)}(P_{XY}) \right\}\right| \geq 2^{n(H(P_{XY})-H(P_X)-3\delta_{XY})}$.*

Proof. Note that $\delta_{XY} = 0$ is only possible if $H(P_{XY}) = 0$; in that case, the claim of the lemma can be easily verified. Assume $\delta_{XY} > 0$ from now on. Let n be sufficiently large. Fix $x^n \in \mathcal{T}_{\epsilon'}^{(n)}(P_X)$. Let $Y^n \sim \prod_{i=1}^{n} P_{Y|X}(y_i|x_i)$. Then,

$$\begin{aligned}
2^{-n\delta_{XY}} &\overset{\text{(i)}}{\leq} \frac{1}{2} \\
&\overset{\text{(ii)}}{\leq} \Pr\left[(x^n, Y^n) \in \mathcal{T}_{\epsilon}^{(n)}(P_{XY})\right] \\
&= \sum_{y^n :\, (x^n,y^n) \in \mathcal{T}_{\epsilon}^{(n)}(P_{XY})} \Pr[Y^n = y^n] \\
&= \sum_{y^n :\, (x^n,y^n) \in \mathcal{T}_{\epsilon}^{(n)}(P_{XY})} \prod_{i=1}^{n} P_{Y|X}(y_i|x_i) \\
&= \sum_{y^n :\, (x^n,y^n) \in \mathcal{T}_{\epsilon}^{(n)}(P_{XY})} \prod_{i=1}^{n} \frac{P_{XY}(x_i, y_i)}{P_X(x_i)} \\
&= \sum_{y^n :\, (x^n,y^n) \in \mathcal{T}_{\epsilon}^{(n)}(P_{XY})} \frac{\prod_{i=1}^{n} P_{XY}(x_i, y_i)}{\prod_{i=1}^{n} P_X(x_i)} \\
&\overset{\text{(iii)}}{\leq} \sum_{y^n :\, (x^n,y^n) \in \mathcal{T}_{\epsilon}^{(n)}(P_{XY})} \frac{2^{-n(H(P_{XY})-\delta_{XY})}}{2^{-n(H(P_X)+\epsilon H(P_X))}} \\
&\overset{\text{(iv)}}{\leq} \sum_{y^n :\, (x^n,y^n) \in \mathcal{T}_{\epsilon}^{(n)}(P_{XY})} \frac{2^{-n(H(P_{XY})-\delta_{XY})}}{2^{-n(H(P_X)+\delta_{XY})}} \\
&= \left|\left\{ y^n \in \mathcal{Y}^n : (x^n, y^n) \in \mathcal{T}_{\epsilon}^{(n)}(P_{XY}) \right\}\right| \cdot 2^{-n(H(P_{XY})-H(P_X)-2\delta_{XY})},
\end{aligned}$$

where (i) holds for sufficiently large n because $\delta_{XY} > 0$; (ii) follows from Lemma 11 for sufficiently large n; (iii) follows from Lemma 6 because $(x^n, y^n) \in \mathcal{T}_{\epsilon}^{(n)}(P_{XY})$ and from Lemma 1 because $x^n \in \mathcal{T}_{\epsilon}^{(n)}(P_X)$ by Lemma 5; and (iv) holds because $H(P_X) \leq H(P_{XY})$. Multiplication by $2^{n(H(P_{XY})-H(P_X)-2\delta_{XY})}$ leads to the desired inequality. ■