Chapter 6
Strong Typicality

Department of Information Engineering
The Chinese University of Hong Kong

6.1 Strong AEP

Setup

- $\{X_k, k \geq 1\}$, X_k i.i.d. $\sim p(x)$.
- X denotes generic r.v. with entropy $H(X) < \infty$.
- $|\mathcal{X}| < \infty$

6.1 Strong AEP

Definition 6.1 The strongly typical set $T^n_{[X]\delta}$ with respect to $p(x)$ is the set of sequences $\mathbf{x} = (x_1, x_2, \cdots, x_n) \in \mathcal{X}^n$ such that $N(x; \mathbf{x}) = 0$ for $x \notin \mathcal{S}_X$, and

$$\sum_x \left| \frac{1}{n} N(x; \mathbf{x}) - p(x) \right| \leq \delta,$$

where $N(x; \mathbf{x})$ is the number of occurrences of x in the sequence $\mathbf{x}$, and δ is an arbitrarily small positive real number. The sequences in $T^n_{[X]\delta}$ are called strongly δ-typical sequences.

Theorem 6.2 (Strong AEP) There exists $\eta > 0$ such that $\eta \to 0$ as $\delta \to 0$, and the following hold:

1) If $\mathbf{x} \in T^n_{[X]\delta}$, then

$$2^{-n(H(X)+\eta)} \leq p(\mathbf{x}) \leq 2^{-n(H(X)-\eta)}.$$

2) For n sufficiently large,

$$\Pr\{\mathbf{X} \in T^n_{[X]\delta}\} > 1 - \delta.$$

3) For n sufficiently large,

$$(1-\delta)2^{n(H(X)-\eta)} \leq |T^n_{[X]\delta}| \leq 2^{n(H(X)+\eta)}.$$

Proof

1. If the relative frequency is about right, then everything else, including the empirical entropy, would be about right.

2. A consequence of WLLN.

3. Exactly the same as the proof of Part 3) of Theorem 5.3.

Theorem 6.3 For sufficiently large n, there exists $\varphi(\delta) > 0$ such that

$$\Pr\{\mathbf{X} \notin T^n_{[X]\delta}\} < 2^{-n\varphi(\delta)}.$$

Proof Chernoff bound.

6.2 Strong Typicality vs Weak Typicality

- Weak typicality: empirical entropy $\approx H(X)$
- Strong typicality: relative frequency $\sim p(x)$
- Strong typicality $\Rightarrow$ weak typicality (Proposition 6.5)
- Weak typicality $\not\Rightarrow$ strong typicality (see example in text)
- Both have AEP, but strong typicality has stronger conditional asymptotic properties (Theorem 6.10).
- Strong typicality works only for finite alphabet, i.e., $|\mathcal{X}| < \infty$.

Strong Typicality Implies Weak Typicality

Proposition 6.5 For any $\mathbf{x} \in \mathcal{X}^n$, if $\mathbf{x} \in T^n_{[X]\delta}$, then $\mathbf{x} \in W^n_{[X]\eta}$, where $\eta \to 0$ as $\delta \to 0$.

Proof By strong AEP and the definitions.

6.3 Joint Typicality

Setup

- $\{(X_k, Y_k), k \geq 1\}$, (X_k, Y_k) i.i.d. $\sim p(x, y)$.
- (X, Y) denotes pair of generic r.v. with entropy $H(X, Y) < \infty$.
- $|\mathcal{X}|, |\mathcal{Y}| < \infty$

Definition 6.6 The strongly jointly typical set $T^n_{[XY]\delta}$ with respect to $p(x, y)$ is the set of $(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ such that $N(x, y; \mathbf{x}, \mathbf{y}) = 0$ for $(x, y) \notin \mathcal{S}_{XY}$, and

$$\sum_x \sum_y \left| \frac{1}{n} N(x, y; \mathbf{x}, \mathbf{y}) - p(x, y) \right| \leq \delta,$$

where $N(x, y; \mathbf{x}, \mathbf{y})$ is the number of occurrences of (x, y) in the pair of sequences $(\mathbf{x}, \mathbf{y})$, and δ is an arbitrarily small positive real number. A pair of sequences $(\mathbf{x}, \mathbf{y})$ is called strongly jointly δ-typical if it is in $T^n_{[XY]\delta}$.

Theorem 6.7 (Consistency) If $(\mathbf{x}, \mathbf{y}) \in T^n_{[XY]\delta}$, then $\mathbf{x} \in T^n_{[X]\delta}$ and $\mathbf{y} \in T^n_{[Y]\delta}$.

Theorem 6.8 (Preservation) Let $Y = f(X)$. If

$$\mathbf{x} = (x_1, x_2, \cdots, x_n) \in T^n_{[X]\delta},$$

then

$$f(\mathbf{x}) = (y_1, y_2, \cdots, y_n) \in T^n_{[Y]\delta},$$

where $y_i = f(x_i)$ for $1 \leq i \leq n$.

Theorem 6.9 (Strong JAEP) Let

$$(\mathbf{X}, \mathbf{Y}) = ((X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)),$$

where (X_i, Y_i) are i.i.d. with generic pair of random variables (X, Y). Then there exists $\lambda > 0$ such that $\lambda \to 0$ as $\delta \to 0$, and the following hold:

1) If $(\mathbf{x}, \mathbf{y}) \in T^n_{[XY]\delta}$, then

$$2^{-n(H(X,Y)+\lambda)} \leq p(\mathbf{x}, \mathbf{y}) \leq 2^{-n(H(X,Y)-\lambda)}.$$

2) For n sufficiently large,

$$\Pr\{(\mathbf{X}, \mathbf{Y}) \in T^n_{[XY]\delta}\} > 1 - \delta.$$

3) For n sufficiently large,

$$(1-\delta)2^{n(H(X,Y)-\lambda)} \leq |T^n_{[XY]\delta}| \leq 2^{n(H(X,Y)+\lambda)}.$$

Stirling's Approximation

Lemma 6.11 (simplified) $\ln n! \sim n \ln n$.

Proof Write

$$\ln n! = \ln 1 + \ln 2 + \cdots + \ln n.$$

Since $\ln x$ is a monotonically increasing, we have

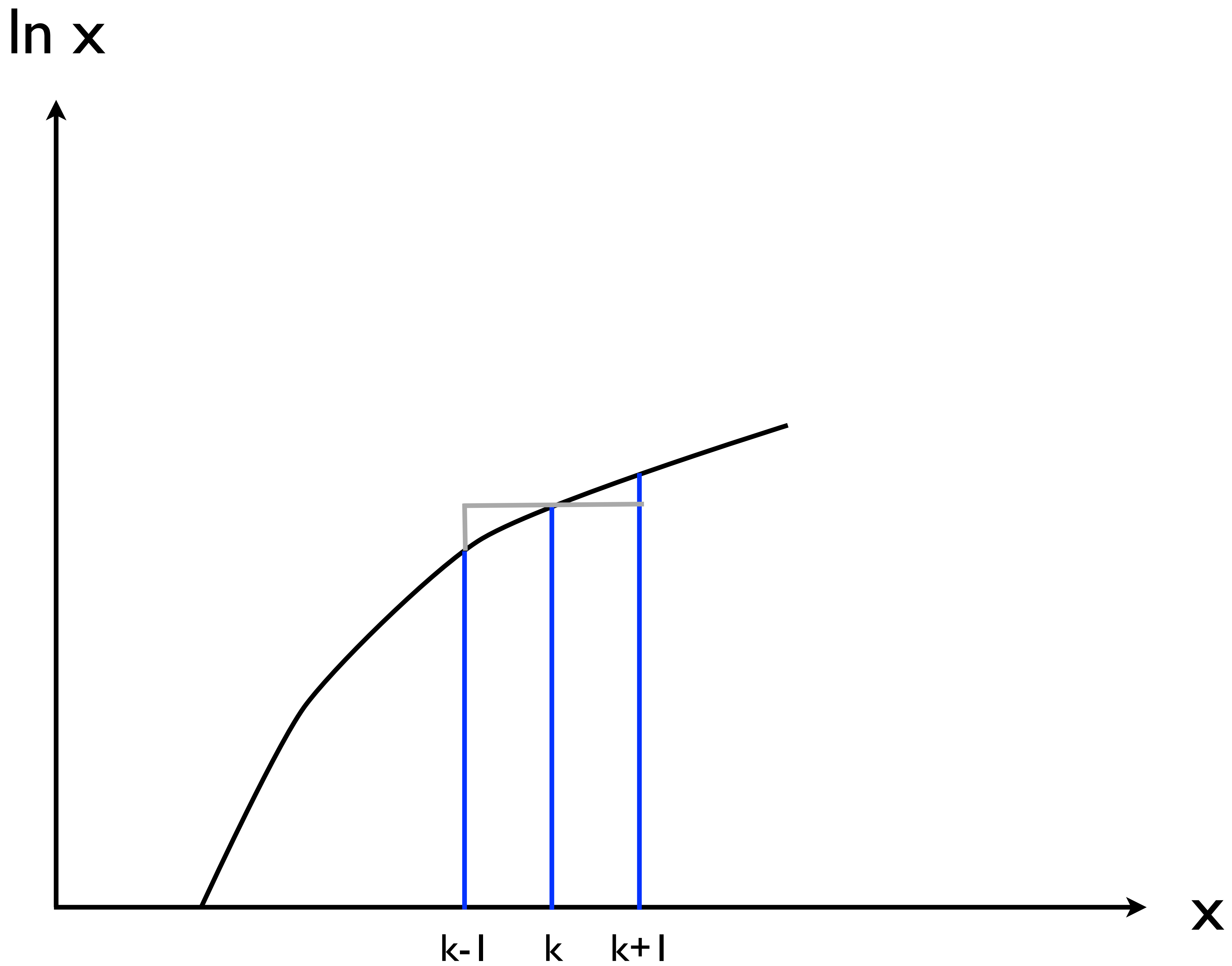
$$\int_{k-1}^{k} \ln x \; dx < \ln k < \int_{k}^{k+1} \ln x \; dx.$$

Summing over $1 \le k \le n$, we have

$$\int_{0}^{n} \ln x \; dx < \ln n! < \int_{1}^{n+1} \ln x \; dx,$$

or

$$n \ln n - n < \ln n! < (n+1)\ln(n+1) - n.$$



ln x
x
k-1
k
k+1

An Application

Lemma For large n,

$$\binom{n}{np, n(1-p)} \approx 2^{nH_2(\{p,1-p\})}$$

Proof

$$\begin{aligned}
\ln \binom{n}{np, n(1-p)} &\approx n \ln n - (np)\ln(np) - (n(1-p))\ln(n(1-p)) \\
&= n \ln n - np[\ln n + \ln p] - n(1-p)[\ln n + \ln(1-p)] \\
&= -n[p \ln p + (1-p)\ln(1-p)]
\end{aligned}$$

Changing to the base 2, we have

$$\log_2 \binom{n}{np, n(1-p)} \approx nH_2(\{p, 1-p\})$$

In general, for large n,

$$\binom{n}{np_1, np_2, \cdots, np_m} = \frac{n!}{\prod_i (np_i)!} \approx 2^{nH(\{p_i\})}$$

Theorem 6.10 (Conditional Strong AEP) For any $\mathbf{x} \in T^n_{[X]\delta}$, define

$$T^n_{[Y|X]\delta}(\mathbf{x}) = \{\mathbf{y} \in T^n_{[Y]\delta} : (\mathbf{x}, \mathbf{y}) \in T^n_{[XY]\delta}\}.$$

If $|T^n_{[Y|X]\delta}(\mathbf{x})| \geq 1$, then

$$2^{n(H(Y|X)-\nu)} \leq |T^n_{[Y|X]\delta}(\mathbf{x})| \leq 2^{n(H(Y|X)+\nu)},$$

where $\nu \to 0$ as $n \to \infty$ and $\delta \to 0$.

Remark Weak Typicality guarantees that the number of $\mathbf{y}$ that are jointly typical with a typical $\mathbf{x}$ is approximately equal to $2^{n(H(Y|X)}$ on the average. Strong typicality guarantees that this is so for each typical $\mathbf{x}$ as long as there exists at least one $\mathbf{y}$ that is jointly typical with $\mathbf{x}$.

Upper Bound in Theorem 6.10

- For any $\nu > 0$, consider

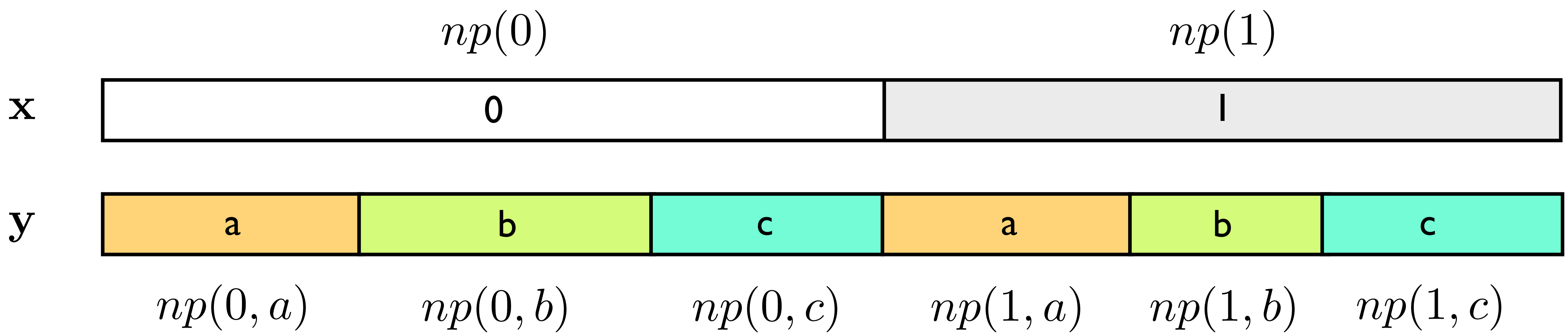
$$\begin{aligned}
2^{-n(H(X)-\nu/2)} &\overset{a)}{\geq} p(\mathbf{x}) \\
&= \sum_{\mathbf{y}\in\mathcal{Y}^n} p(\mathbf{x},\mathbf{y}) \\
&\geq \sum_{\mathbf{y}\in T^n_{[Y|X]\delta}(\mathbf{x})} p(\mathbf{x},\mathbf{y}) \\
&\overset{b)}{\geq} \sum_{\mathbf{y}\in T^n_{[Y|X]\delta}(\mathbf{x})} 2^{-n(H(X,Y)+\nu/2)} \\
&= |T^n_{[Y|X]\delta}(\mathbf{x})| 2^{-n(H(X,Y)+\nu/2)},
\end{aligned}$$

  a) and b) follows from strong joint AEP.

- Similar to the proof of the upper bound on $|T^n_{[X]\delta}|$ in Theorem 6.2 (SAEP).

Lower Bound in Theorem 6.10

$\mathcal{X} = \{0, 1\}, \mathcal{Y} = \{a, b, c\}$



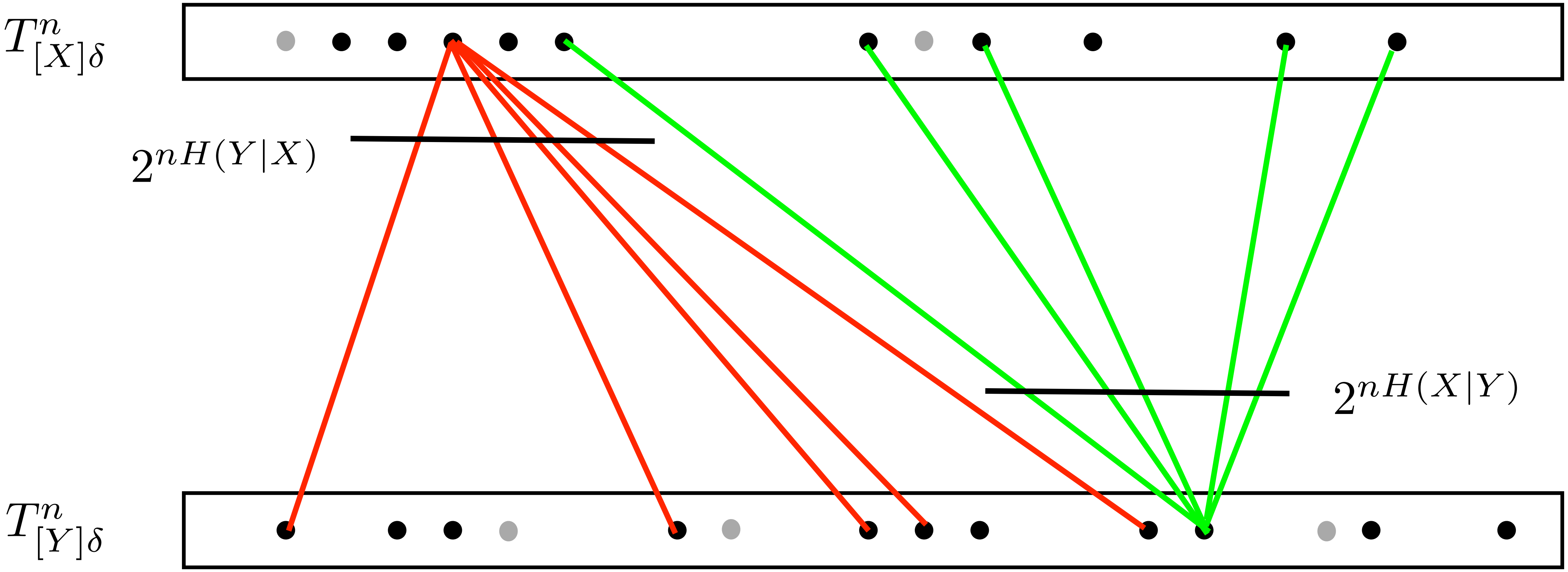
Rearrange the components of $\mathbf{y}$ corresponding to $x_k = 0$ and rearrange the components of $\mathbf{y}$ corresponding to $x_k = 1$. This preserves joint typicality.

$$
\begin{aligned}
\#arrangements &\approx \binom{np(0)}{np(0,a), np(0,b), np(0,c)} \binom{np(1)}{np(1,a), np(1,b), np(1,c)} \\
&\approx 2^{np(0)H(\{p(\cdot|0)\})} 2^{np(1)H(\{p(\cdot|1)\})} \\
&= 2^{n(p(0)H(Y|X=0)+p(1)H(Y|X=1))} \\
&= 2^{nH(Y|X)}
\end{aligned}
$$

Hence,

$$|T^n_{[Y|X]\delta}(\mathbf{x})| \geq 2^{n(H(Y|X)-\nu)}.$$

An Illustration of Conditional SAEP



Corollary 6.12 For a joint distribution $p(x, y)$ on $\mathcal{X} \times \mathcal{Y}$, let $S^n_{[X]\delta}$ be the set of all sequences $\mathbf{x} \in T^n_{[X]\delta}$ such that $T^n_{[Y|X]\delta}(\mathbf{x})$ is nonempty. Then

$$|S^n_{[X]\delta}| \geq (1 - \delta) 2^{n(H(X) - \psi)},$$

where $\psi \to 0$ as $n \to \infty$ and $\delta \to 0$.

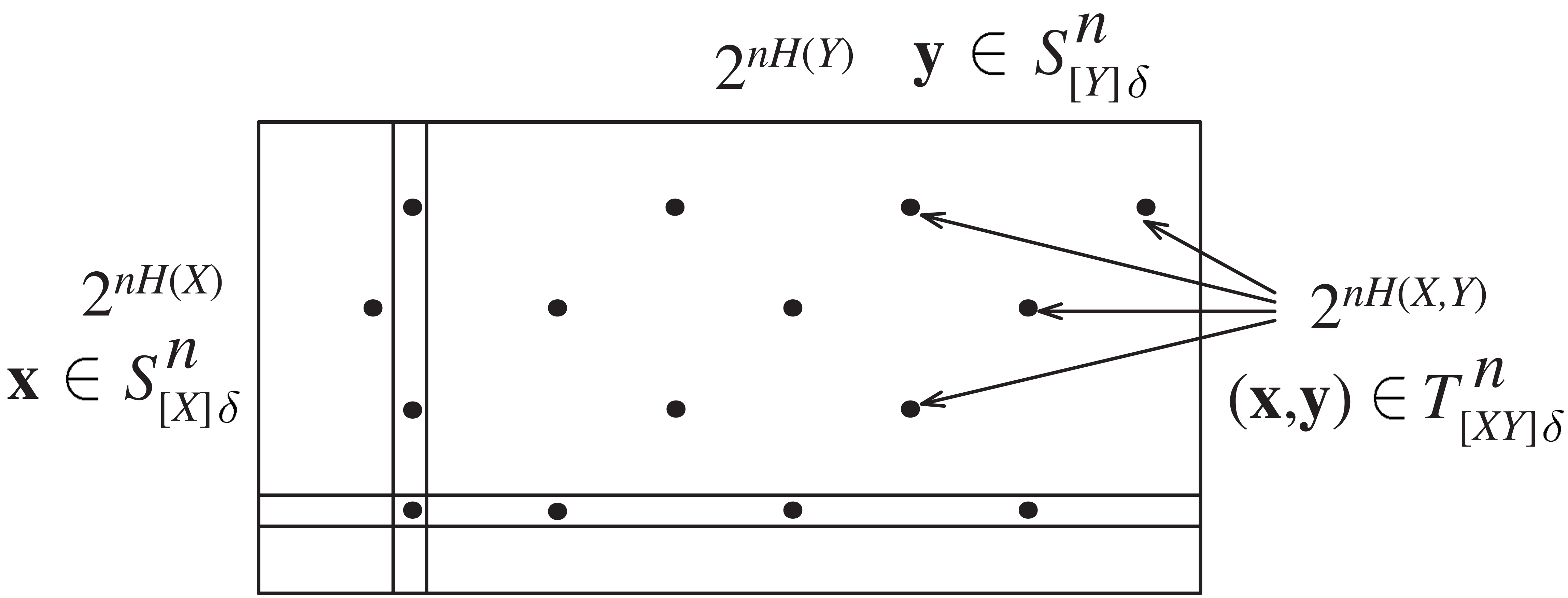
Proposition 6.13 With respect to a joint distribution $p(x, y)$ on $\mathcal{X} \times \mathcal{Y}$, for any $\delta > 0$,

$$\Pr\{\mathbf{X} \in S^n_{[X]\delta}\} > 1 - \delta$$

for n sufficiently large.

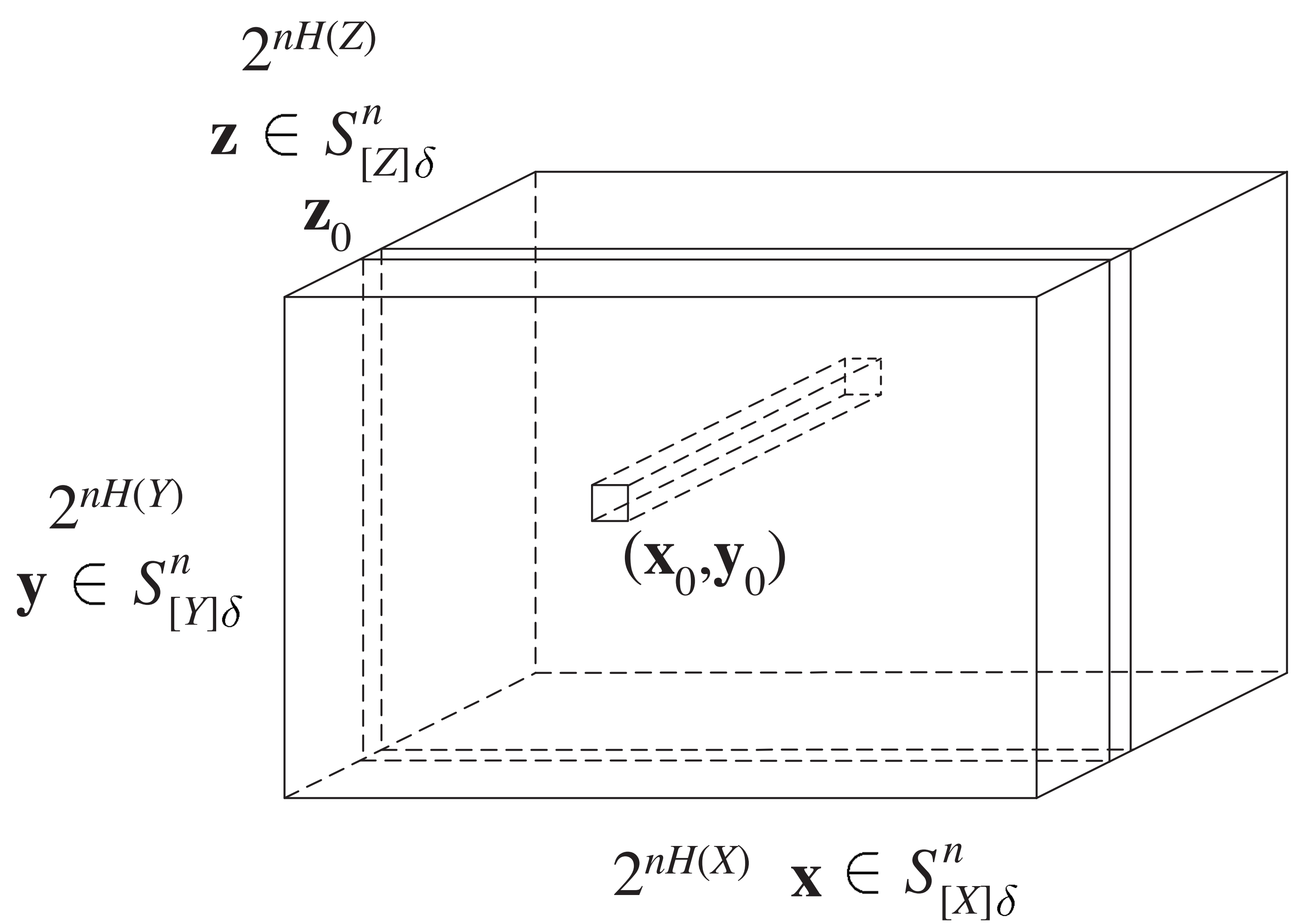
Strongly Joint Typicality Array

- Exhibits an "asymptotic quasi-uniform" structure.
- Two-Dimensional:



Strongly Joint Typicality Array

- Three-Dimensional:



Quasi-Uniform Array

- Provides a combinatorial interpretation of information inequalities.
- Related to many branches of information sciences: combinatorics, group theory (Ch. 16), Kolmogorov complexity, network coding, probability theory, matrix theory, quantum mechanics, ...
- Resources:
  - `http://temple.birs.ca/~09w5103/yeung_09w5103_talk.pdf` (slides)
  - `http://www.birs.ca/events/2009/5-day-workshops/09w5103/videos` (video)