Refinement of Two Fundamental Tools in Information Theory

Raymond W. Yeung
Institute of Network Coding
The Chinese University of Hong Kong

Joint work with Siu Wai Ho and Sergio Verdu

Discontinuity of Shannon's Information Measures

- Shannon's information measures: $H(X)$, $H(X|Y)$, $I(X;Y)$ and $I(X;Y|Z)$.
- They are described as continuous functions [Shannon 1948] [Csiszár & Körner 1981] [Cover & Thomas 1991] [McEliece 2002] [Yeung 2002].
- All Shannon's information measures are indeed discontinuous everywhere when random variables take values from countably infinite alphabets [Ho & Yeung 2005].
- e.g., X can be any positive integer.

Discontinuity of Entropy

- Let $P_X = \{1, 0, 0, \ldots\}$ and

$$P_{X_n} = \left\{1 - \frac{1}{\sqrt{\log n}}, \frac{1}{n\sqrt{\log n}}, \frac{1}{n\sqrt{\log n}}, \ldots, 0, 0, \ldots\right\}.$$

- As $n \to \infty$, we have

$$\sum_i \left|P_X(i) - P_{X_n}(i)\right| = \frac{2}{\sqrt{\log n}} \to 0.$$

- However,

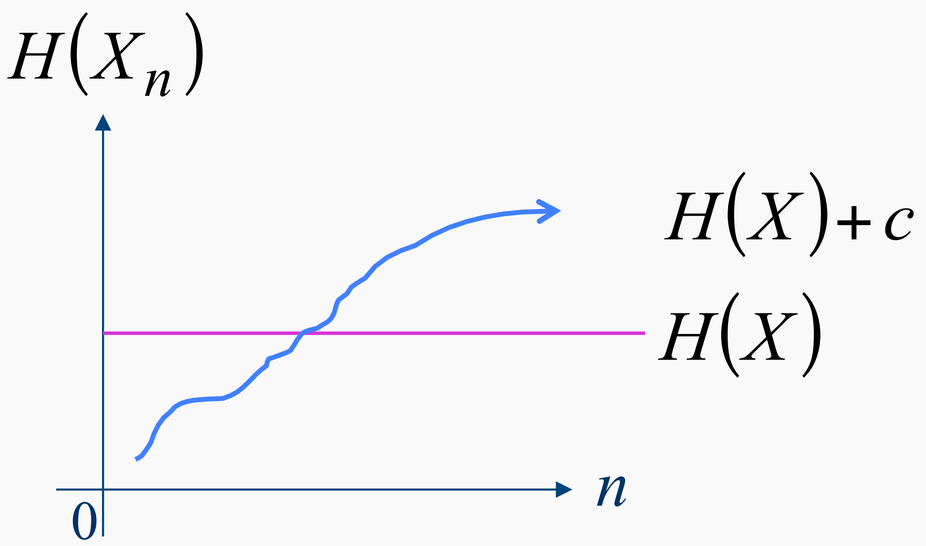
$$\lim_{n \to \infty} H(X_n) = \infty.$$

Discontinuity of Entropy

- Theorem 1: For any $c \geq 0$ and any X taking values from a countably infinite alphabet with $H(X) < \infty$,

$$\exists P_{X_n} \quad \text{s.t.} \quad V\left(P_X, P_{X_n}\right) = \sum_i \left|P_X(i) - P_{X_n}(i)\right| \to 0$$

$$\text{but} \quad H\left(X_n\right) \to H\left(X\right) + c$$

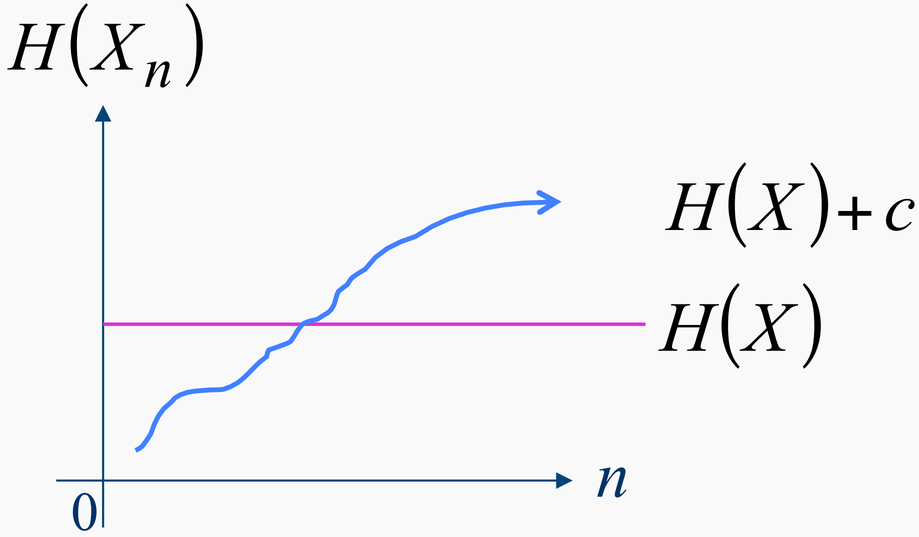


Discontinuity of Entropy

- Theorem 2: For any $c \geq 0$ and any X taking values from countably infinite alphabet with $H(X) < \infty$,

$$\exists P_{X_n} \quad \text{s.t.} \quad D\left(P_X \| P_{X_n}\right) = \sum_i P_X(i) \log \frac{P_X(i)}{P_{X_n}(i)} \to 0$$

$$\text{but} \quad H\left(X_n\right) \to H\left(X\right) + c$$

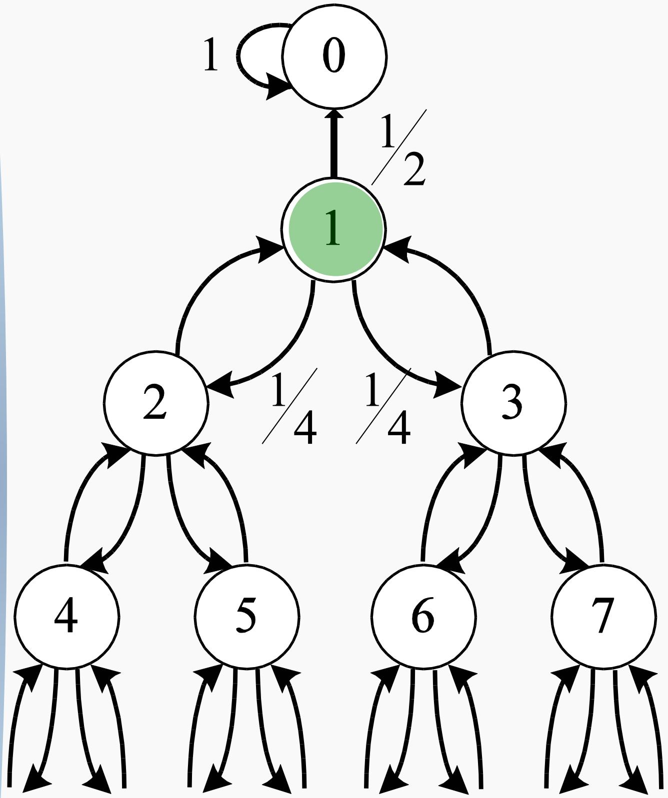


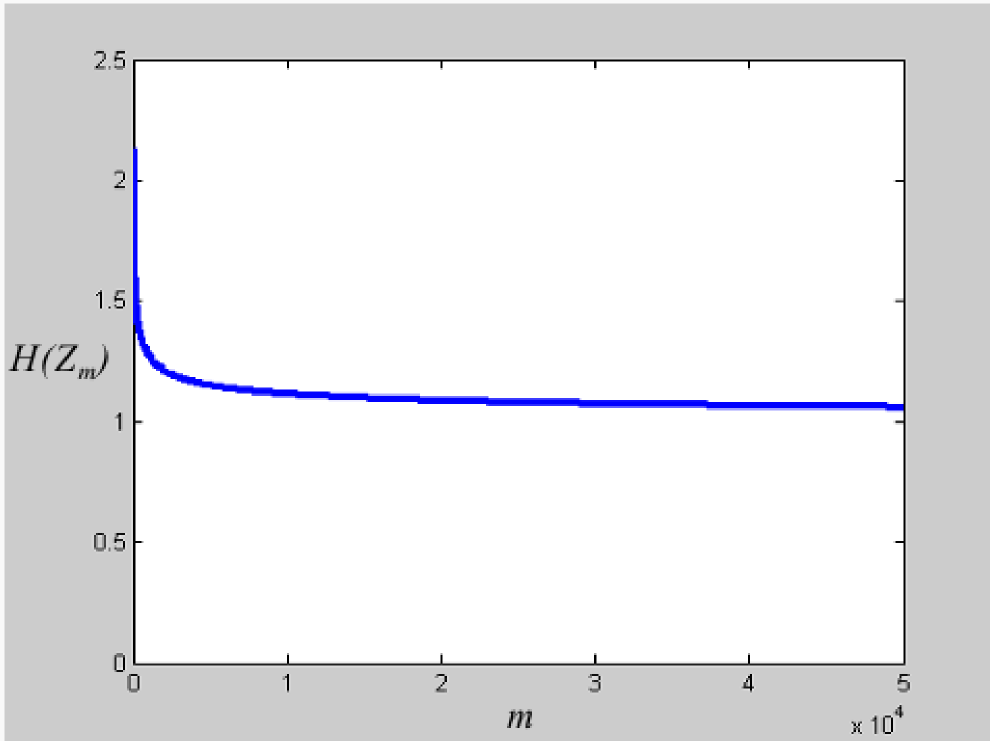
Pinsker's inequality

$$D(p \parallel q) \geq \frac{1}{2\ln 2} V^2(p,q)$$

- By Pinsker's inequality, convergence w.r.t. $D(\cdot \parallel \cdot)$ implies convergence w.r.t. $V(\cdot,\cdot)$.
- Therefore, Theorem 2 implies Theorem 1.

Discontinuity of Entropy



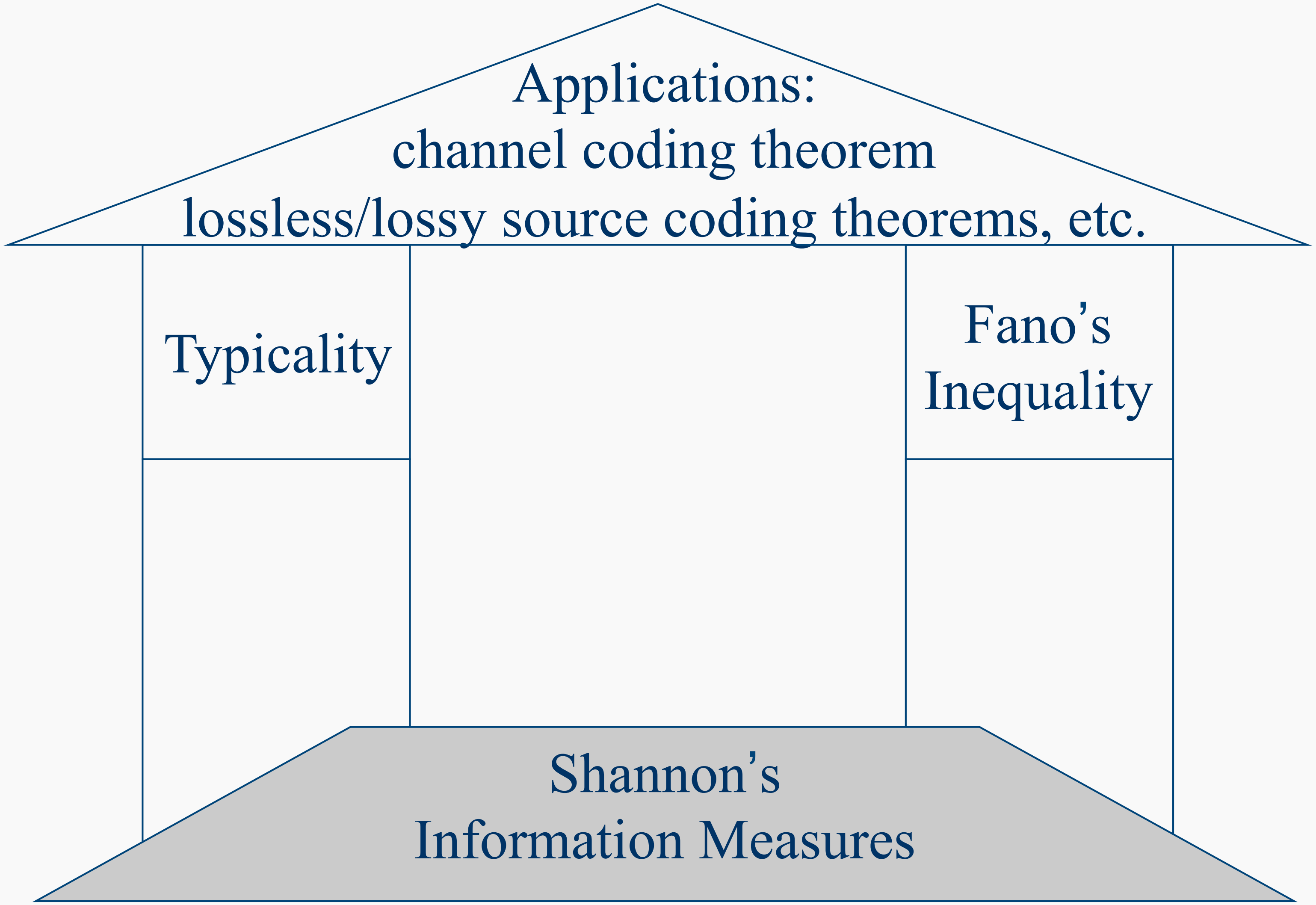


Discontinuity of Shannon's Information Measures

- Theorem 3: For any X, Y and Z taking values from countably infinite alphabet with $I(X;Y|Z) < \infty$,

$$\exists P_{X_n Y_n Z_n} \quad \text{s.t.} \quad \lim_{n\to\infty} D\left(P_{XYZ} \,\|\, P_{X_n Y_n Z_n}\right) = 0$$
$$\text{but} \quad \lim_{n\to\infty} I\left(X_n; Y_n \,|\, Z_n\right) = \infty.$$

Discontinuity of Shannon's Information Measures

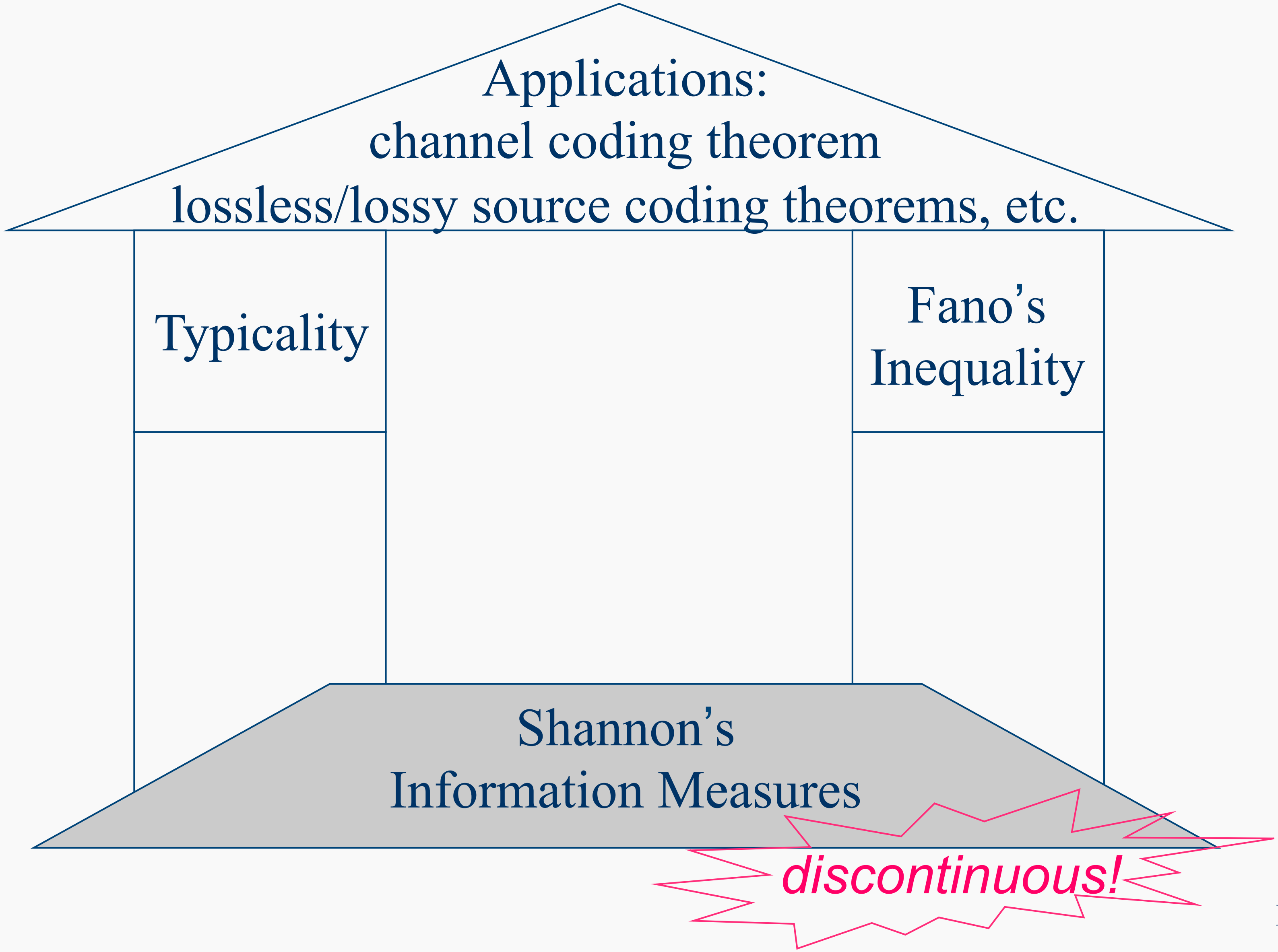


To Find the Capacity of a Communication Channel

Capacity ≥ C_1 — Typicality

Capacity ≤ C_2 — Fano's Inequality

On Countably Infinite Alphabet



Typicality

- Weak typicality was first introduced by Shannon [1948] to establish the source coding theorem.
- Strong typicality was first used by Wolfowitz [1964] and then by Berger [1978]. It was further developed into the method of types by Csiszár and Körner [1981].
- Strong typicality possesses stronger properties compared with weak typicality.
- It can be used only for random variables with finite alphabet.

Notations

- Consider an i.i.d. source $\{X_k, k \geq 1\}$, where X_k taking values from a countable alphabet $\mathcal{X}$.
- Let $P_X = P_{X_k}$ for all k.
- Assume $H(P_X) < \infty$.
- Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$
- For a sequence $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$,
- $N(x; \mathbf{x})$ is the *number of occurrences* of x in $\mathbf{x}$
- $q(x; \mathbf{x}) = n^{-1}N(x; \mathbf{x})$ and
- $Q_X = \{q(x; \mathbf{x})\}$ is the *empirical distribution* of $\mathbf{x}$
- e.g., $\mathbf{x} = (1, 3, 2, 1, 1)$.

$$N(1; \mathbf{x}) = 3,\ N(2; \mathbf{x}) = N(3; \mathbf{x}) = 1$$

$$Q_X = \{3/5, 1/5, 1/5\}.$$

Weak Typicality

- Definition (Weak typicality): For any $\varepsilon > 0$, the weakly typical set $W^n_{[X]\varepsilon}$ with respect to P_X is the set of sequences $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ such that

$$\left| -\frac{1}{n} \log P_X(\mathbf{x}) - H(P_X) \right| \leq \varepsilon$$

Weak Typicality

- **Definition 1 (Weak typicality):** For any $\varepsilon > 0$, the weakly typical set $W^n_{[X]\varepsilon}$ with respect to P_X is the set of sequences $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ such that

$$|D(Q_X \| P_X) + H(Q_X) - H(P_X)| \leq \varepsilon$$

- Note that

$$H(Q_X) = -\sum_x Q_X(x) \log Q_X(x)$$

while

$$\text{Empirical entropy} = -\sum_x Q_X(x) \log P_X(x)$$

Asymptotic Equipartition Property

- Theorem 4 (Weak AEP): For any $\varepsilon > 0$:
  - 1) If $\mathbf{x} \in W^n_{[X]\varepsilon}$, then

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

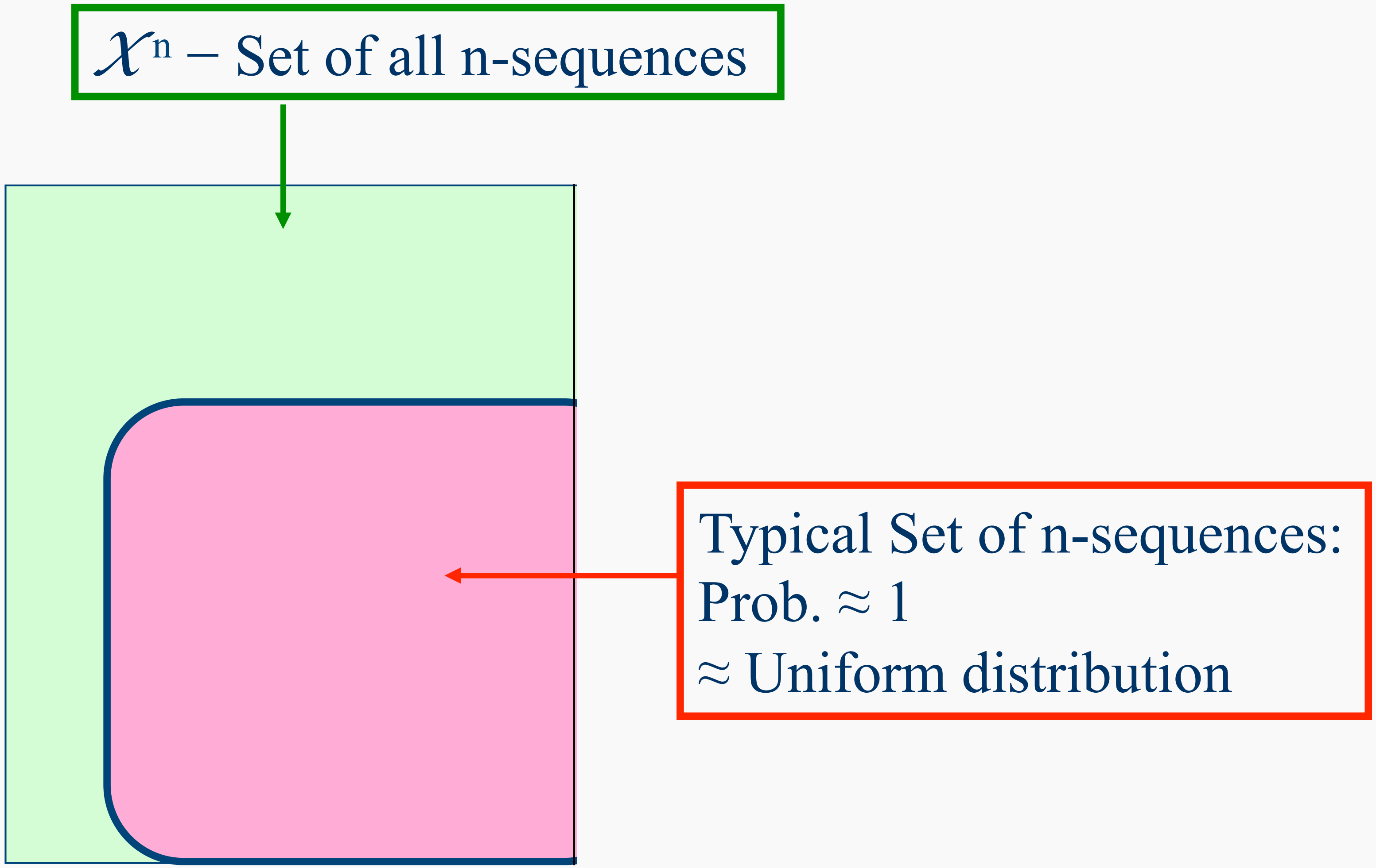
  - 2) For sufficiently large n,

$$\Pr\left\{\mathbf{X} \in W^n_{[X]\varepsilon}\right\} > 1 - \varepsilon$$

  - 3) For sufficiently large n,

$$(1-\varepsilon)2^{n(H(X)-\varepsilon)} \leq \left|W^n_{[X]\varepsilon}\right| \leq 2^{n(H(X)+\varepsilon)}$$

Illustration of AEP



Strong Typicality

- Strong typicality has been defined in slightly different forms in the literature.
- Definition 2 (Strong typicality): For $|\mathcal{X}| < \infty$ and any $\delta > 0$, the strongly typical set $T^n_{[X]\delta}$ with respect to P_X is the set of sequences $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ such that

$$V(P_X, Q_X) = \sum_x |P_X(x) - q(x;\mathbf{x})| \le \delta$$

the variational distance between the empirical distribution of the sequence $\mathbf{x}$ and P_X is small.

Asymptotic Equipartition Property

- Theorem 5 (Strong AEP): For a finite alphabet $\mathcal{X}$ and any $\delta > 0$:
  - 1) If $\mathbf{x} \in T^n_{[X]\delta}$, then

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

  - 2) For sufficiently large n,

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

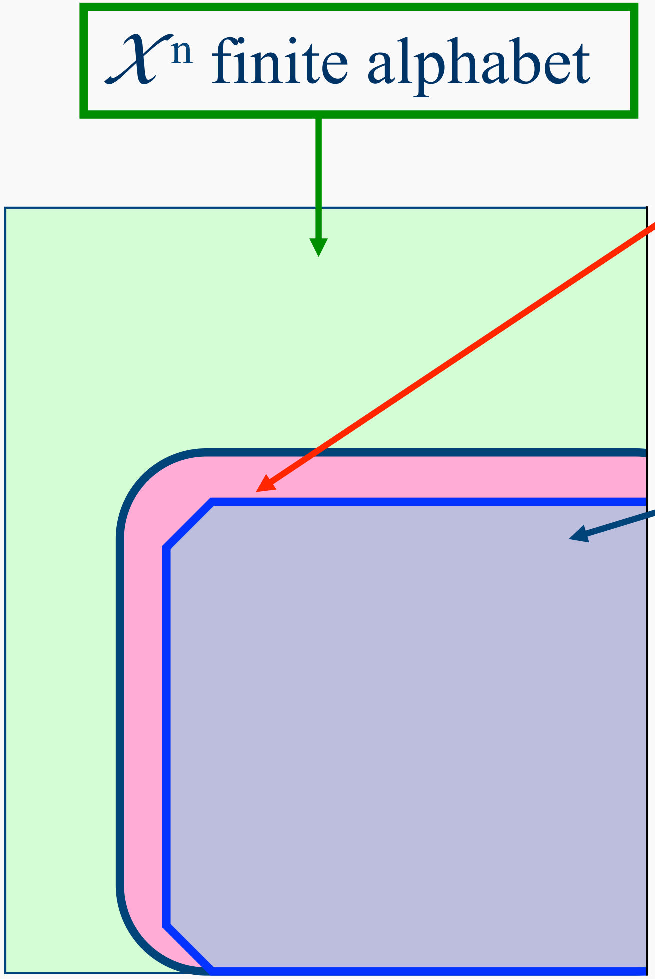
  - 3) For sufficiently large n,

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

Breakdown of Strong AEP

- If strong typicality is extended (in the natural way) to countably infinite alphabets, strong AEP no longer holds
- Specifically, Property 2 holds but Properties 1 and 3 do not hold.

Typicality



Weak Typicality:

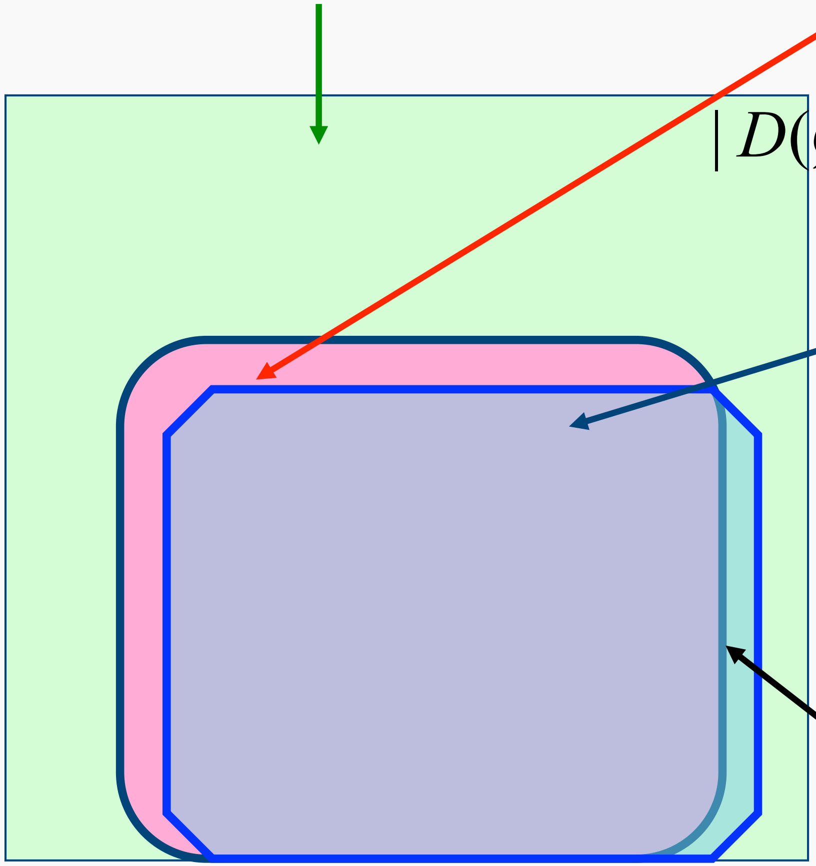
$$|D(Q_X \| P_X) + H(Q_X) - H(P_X)| \leq \varepsilon$$

Strong Typicality:

$$V(P_X, Q_X) \leq \delta$$

Unified Typicality

$\mathcal{X}^n$ countably infinite alphabet



Weak Typicality:

$$|D(Q_X \| P_X) + H(Q_X) - H(P_X)| \leq \varepsilon$$

Strong Typicality:

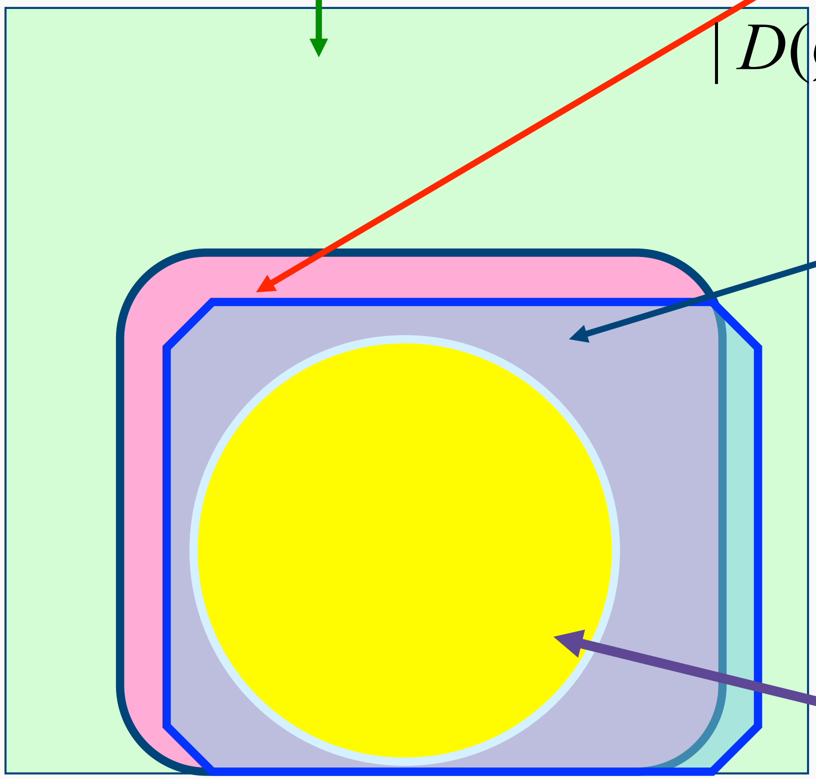
$$V(P_X, Q_X) \leq \delta$$

$\exists \mathbf{x}$ s.t. $D(Q_X \| P_X)$ is small

but $|H(Q_X) - H(P_X)|$ is large

Unified Typicality

$\mathcal{X}^n$ countably infinite alphabet



Weak Typicality:

$$|D(Q_X \| P_X) + H(Q_X) - H(P_X)| \le \varepsilon$$

Strong Typicality:

$$V(P_X, Q_X) \le \delta$$

Unified Typicality:

$$D(Q_X \| P_X) + |H(Q_X) - H(P_X)| \le \eta.$$

Unified Typicality

- Definition 3 (Unified typicality): For any $\eta > 0$, the unified typical set $U^n_{[X]\eta}$ with respect to P_X is the set of sequences $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ such that

$$D(Q_X \| P_X) + |H(Q_X) - H(P_X)| \le \eta$$

- Weak Typicality: $|D(Q_X \| P_X) + H(Q_X) - H(P_X)| \le \varepsilon$

  Strong Typicality: $V(P_X, Q_X) \le \delta$

- Each typicality corresponds to a “distance measure”
- Entropy is continuous w.r.t. the distance measure induced by unified typicality

Asymptotic Equipartition Property

- **Theorem 6 (Unified AEP)**: For any > 0:
  - 1) If $\mathbf{x} \in U^n_{[X]\eta}$, then

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

  - 2) For sufficiently large n,

$$\Pr\left\{\mathbf{X} \in U^n_{[X]\eta}\right\} > 1-\eta$$

  - 3) For sufficiently large n,

$$(1-\eta)2^{n(H(X)-\mu)} \leq \left|U^n_{[X]\eta}\right| \leq 2^{n(H(X)+\mu)}$$

Unified Typicality

- Theorem 7: For any $\mathbf{x} \in \mathcal{X}^n$,

  if $\mathbf{x} \in U^n_{[X]\eta}$, then $\mathbf{x} \in W^n_{[X]\varepsilon}$ and $\mathbf{x} \in T^n_{[X]\delta}$,

  where $\varepsilon = \eta$ and $\delta = \sqrt{\eta \cdot 2 \ln 2}$.

Unified Jointly Typicality

- Consider a bivariate information source $\{(X_k, Y_k), k \geq 1\}$ where (X_k, Y_k) are i.i.d. with generic distribution P_{XY} .
- We use (X, Y) to denote the pair of generic random variables.
- Let $(\boldsymbol{X}, \boldsymbol{Y}) = ((X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n))$.
- For the pair of sequence $(\mathbf{x}, \mathbf{y})$, the empirical distribution is $Q_{XY} = \{q(x,y; \mathbf{x},\mathbf{y})\}$ where $q(x,y; \mathbf{x},\mathbf{y}) = n^{-1}N(x,y; \mathbf{x},\mathbf{y})$.

Unified Jointly Typicality

- Definition 4 (Unified jointly typicality): For any $\eta > 0$, the unified typical set $U^n_{[XY]\eta}$ with respect to P_{XY} is the set of sequences $(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ such that

$$D(Q_{XY} \| P_{XY}) + |H(Q_{XY}) - H(P_{XY})|$$
$$+ |H(Q_X) - H(P_X)| + |H(Q_Y) - H(P_Y)| \leq \eta.$$

- This definition cannot be simplified.

Conditional AEP

- **Definition 5**: For any $\mathbf{x} \in U^n_{[X]\eta}$, the conditional typical set of Y is defined as

$$U^n_{[Y|X]\eta}(\mathbf{x}) = \left\{ \mathbf{y} \in U^n_{[Y]\eta} : (\mathbf{x}, \mathbf{y}) \in U^n_{[XY]\eta} \right\}$$

- **Theorem 8**: For $\mathbf{x} \in U^n_{[X]\eta}$, if

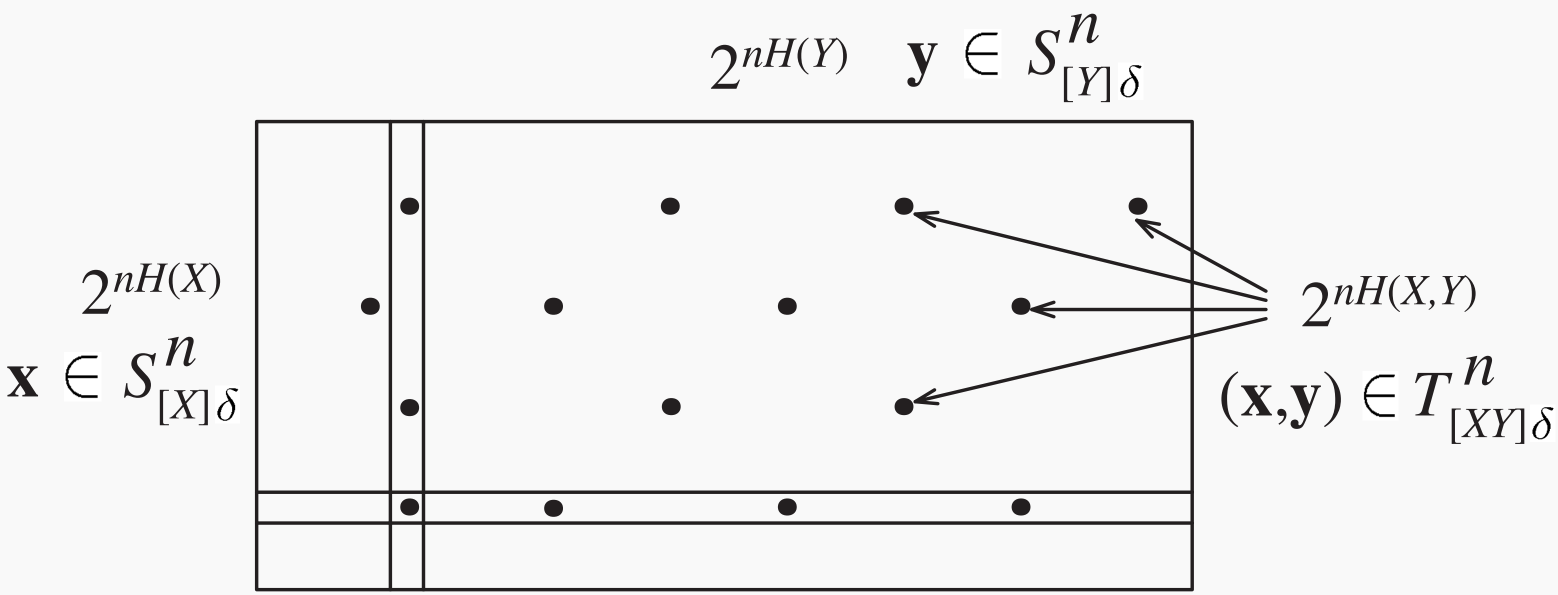
$$\left| U^n_{[Y|X]\eta}(\mathbf{x}) \right| \geq 1,$$

then

$$2^{n(H(Y|X)-\nu)} \leq \left| U^n_{[Y|X]\eta}(\mathbf{x}) \right| \leq 2^{n(H(Y|X)+\nu)}$$

where $\nu \to 0$ as $\eta \to 0$ and then $n \to \infty$

Illustration of Conditional AEP



Applications

- Rate-distortion theory
  - A version of rate-distortion theorem was proved by strong typicality [Cover & Thomas 1991][Yeung 2008]
  - It can be easily generalized to countably infinite alphabet
- Multi-source network coding
  - The achievable information rate region in multisource network coding problem was proved by strong typicality [Yeung 2008]
  - It can be easily generalized to countably infinite alphabet

Fano's Inequality

- Fano's inequality: For discrete random variables X and Y taking values on the same alphabet $\mathcal{X} = \{1, 2, \ldots\}$, let

$$\varepsilon = \mathbf{P}[X \neq Y] = 1 - \sum_{w \in \mathcal{X}} P_{XY}(w, w)$$

- Then

$$\boxed{H(X \mid Y) \leq \varepsilon \log(|\mathcal{X}| - 1) + h(\varepsilon),}$$

where

$$h(x) = x \log \frac{1}{x} + (1 - x) \log \frac{1}{1 - x}$$

for $0 < x < 1$ and $h(0) = h(1) = 0$.

Motivation 1

$$H(X|Y) \le \varepsilon \log(|\mathcal{X}|-1) + h(\varepsilon)$$

- This upper bound on $H(X|Y)$ is not tight.
- For fixed ε and $|\mathcal{X}|$, can always find X such that

$$H(X|Y) \le H(X) < \varepsilon \log(|\mathcal{X}|-1) + h(\varepsilon)$$

- Then we can ask, for fixed P_X and ε, what is

$$\max_{P_{Y|X}: P[X \ne Y] = \varepsilon} H(X|Y) < \varepsilon \log(|\mathcal{X}|-1) + h(\varepsilon)$$

Motivation 2

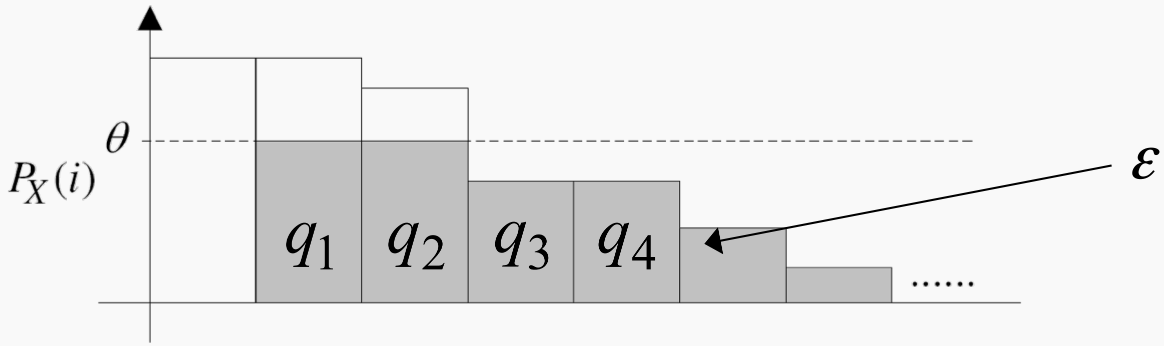
- If $\mathcal{X}$ is countably infinite, Fano's inequality no longer gives an upper bound on $H(X|Y)$.
- It is possible that $H(X|Y) \not\rightarrow 0$ as $\varepsilon \rightarrow 0$ which can be explained by the discontinuity of entropy.
- $P_{X_n} = \left\{1 - \frac{1}{\sqrt{\log n}}, \frac{1}{n\sqrt{\log n}}, \ldots, \frac{1}{n\sqrt{\log n}}\right\}$ and $P_{Y_n} = \{1, 0, 0, \ldots\}$
- Then $H(X_n|Y_n) = H(X_n) \rightarrow \infty$ but $\varepsilon_n = \frac{1}{\sqrt{\log n}} \rightarrow 0$
- Under what conditions $\varepsilon \rightarrow 0 \Rightarrow H(X|Y) \rightarrow 0$ for countably infinite alphabets?

Tight Upper Bound on $H(X|Y)$

- Theorem 9: Suppose $\varepsilon = \mathbf{P}[X \neq Y] \leq 1 - P_X(1)$, then

$$H(X \mid Y) \leq \varepsilon\, H(\mathcal{Q}(P_X, \varepsilon)) + h(\varepsilon)$$

where the right side is the tight bound dependent on ε and P_X. (This is the simplest of the 3 cases.)

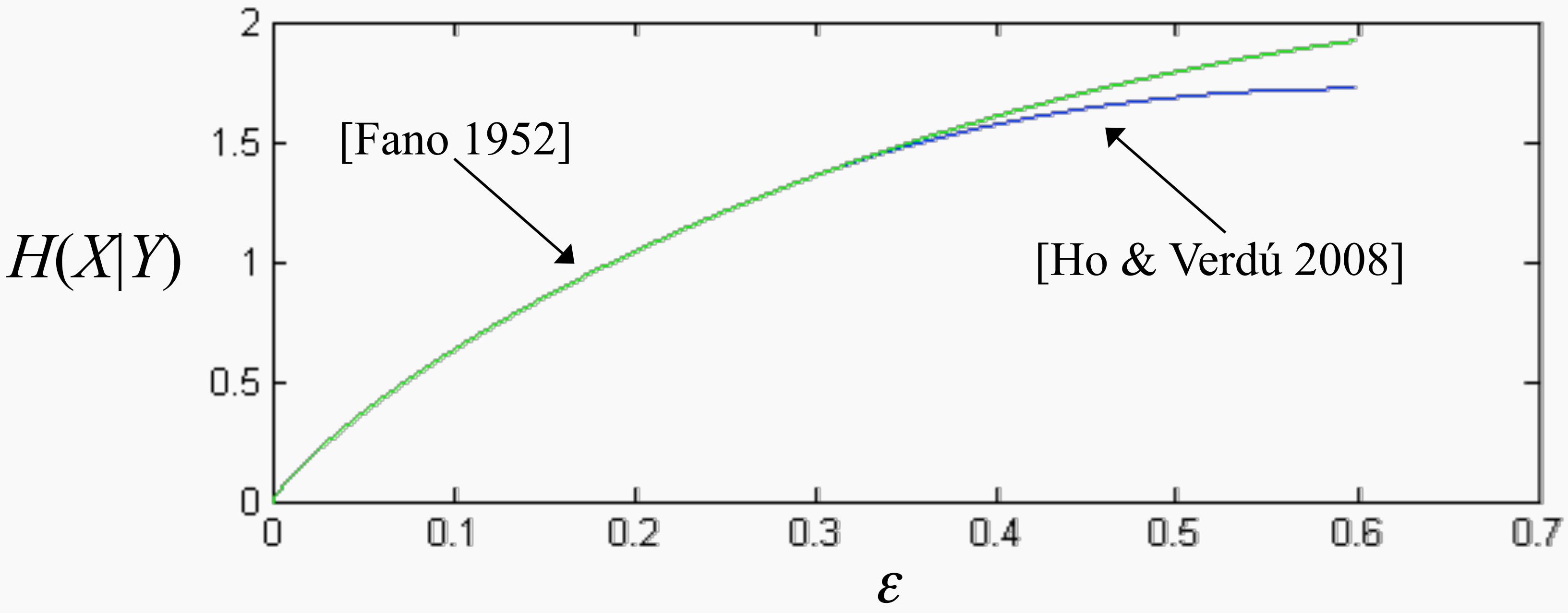


$$\mathcal{Q}(P_X, \varepsilon) = \{\varepsilon^{-1} q_1, \varepsilon^{-1} q_2, \varepsilon^{-1} q_3, \ldots\}$$

- Let $\Phi_X(\varepsilon) = \varepsilon\, H(\mathcal{Q}(P_X, \varepsilon)) + h(\varepsilon)$

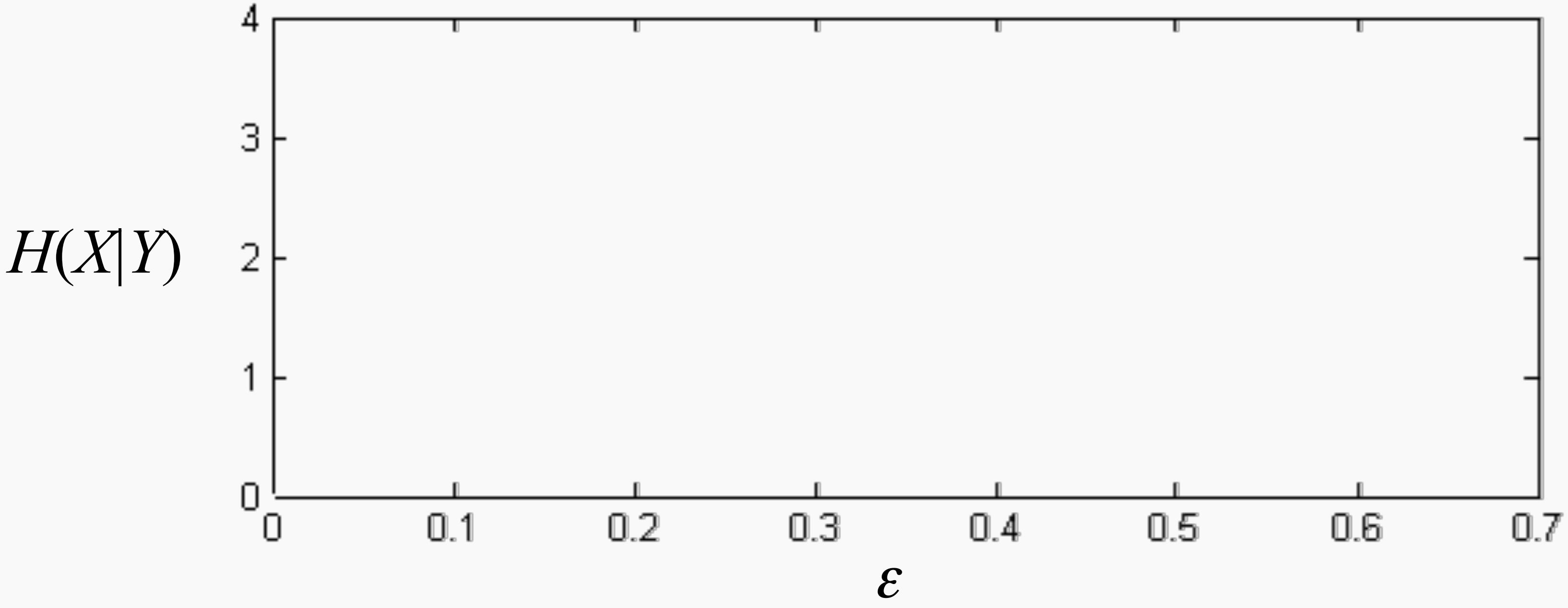
Generalizing Fano's Inequality

- Fano's inequality [Fano 1952] gives an upper bound on the conditional entropy $H(X|Y)$ in terms of the error probability $\varepsilon = \Pr\{X \neq Y\}$.
- e.g. $P_X = [0.4, 0.4, 0.1, 0.1]$



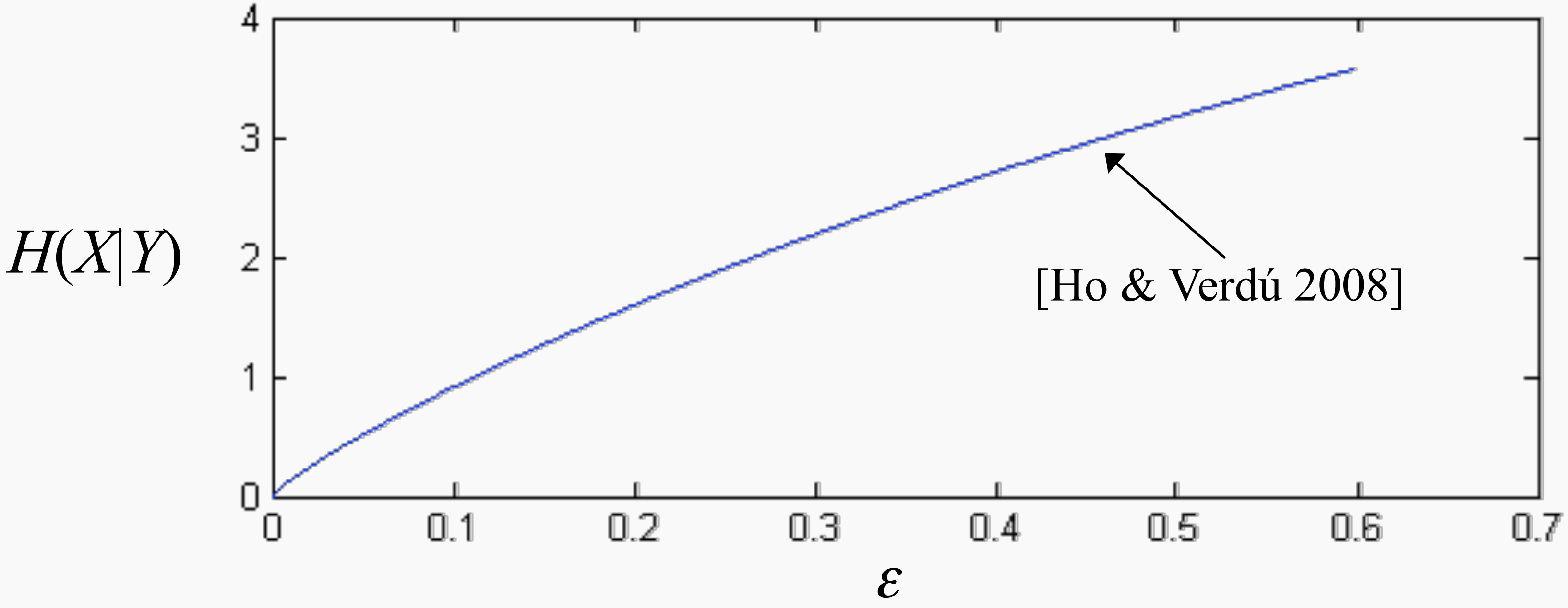
Generalizing Fano's Inequality

- e.g., X is a Poisson random variable with mean equal to 10.
- Fano's inequality no longer gives an upper bound on $H(X|Y)$.

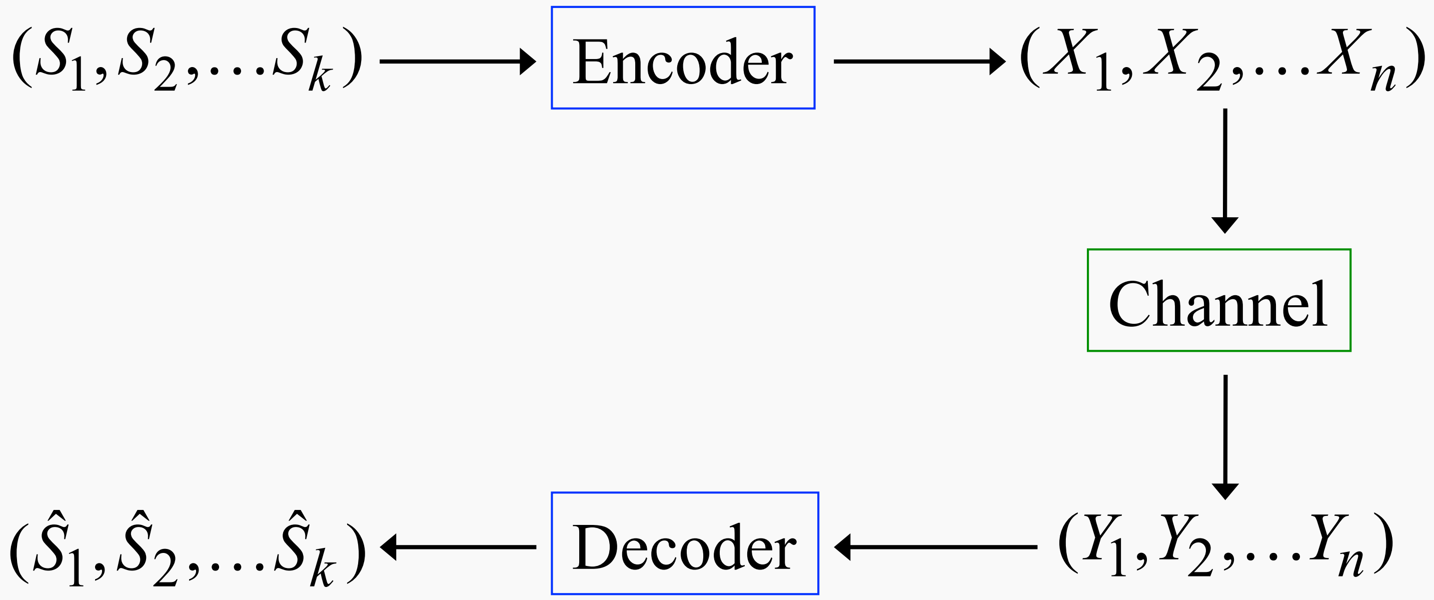


Generalizing Fano's Inequality

- e.g. X is a Poisson random variable with mean equal to 10.
- Fano's inequality no longer gives an upper bound on $H(X|Y)$.



Joint Source-Channel Coding



k-to-*n* joint source-channel code

Error Probabilities

- The average symbol error probability is defined as

$$\lambda_k = \frac{1}{k} \sum_{i=1}^{k} \mathbf{P}[S_i \neq \hat{S}_i]$$

- The block error probability is defined as

$$\mu_k = \mathbf{P}[(S_1, S_2, \ldots S_k) \neq (\hat{S}_1, \hat{S}_2, \ldots \hat{S}_k)]$$

Symbol Error Rate

- Theorem 10: For any discrete memoryless source and general channel, the rate of a *k*-to-*n* joint source-channel code with symbol error probability λ_k satisfies

$$\frac{k}{n} \le \frac{\sup_{X^n} n^{-1} I(X^n;Y^n)}{k^{-1} H(S^k) - \Phi_{S^*}(\lambda_k)}$$

where S^* is constructed from $\{S_1, S_2, \ldots, S_k\}$ according to

$$P_{S^*}(1) = \min_j P_{S_j}(1),$$

$$P_{S^*}(a) = \min_j \sum_{i=1}^{a} P_{S_j}(i) - \sum_{i=1}^{a-1} P_{S^*}(i) \quad a \ge 2.$$

Block Error Rate

- Theorem 11: For any general discrete source and general channel, the block error probability μ_k of a k-to-n joint source-channel code is lower bounded by

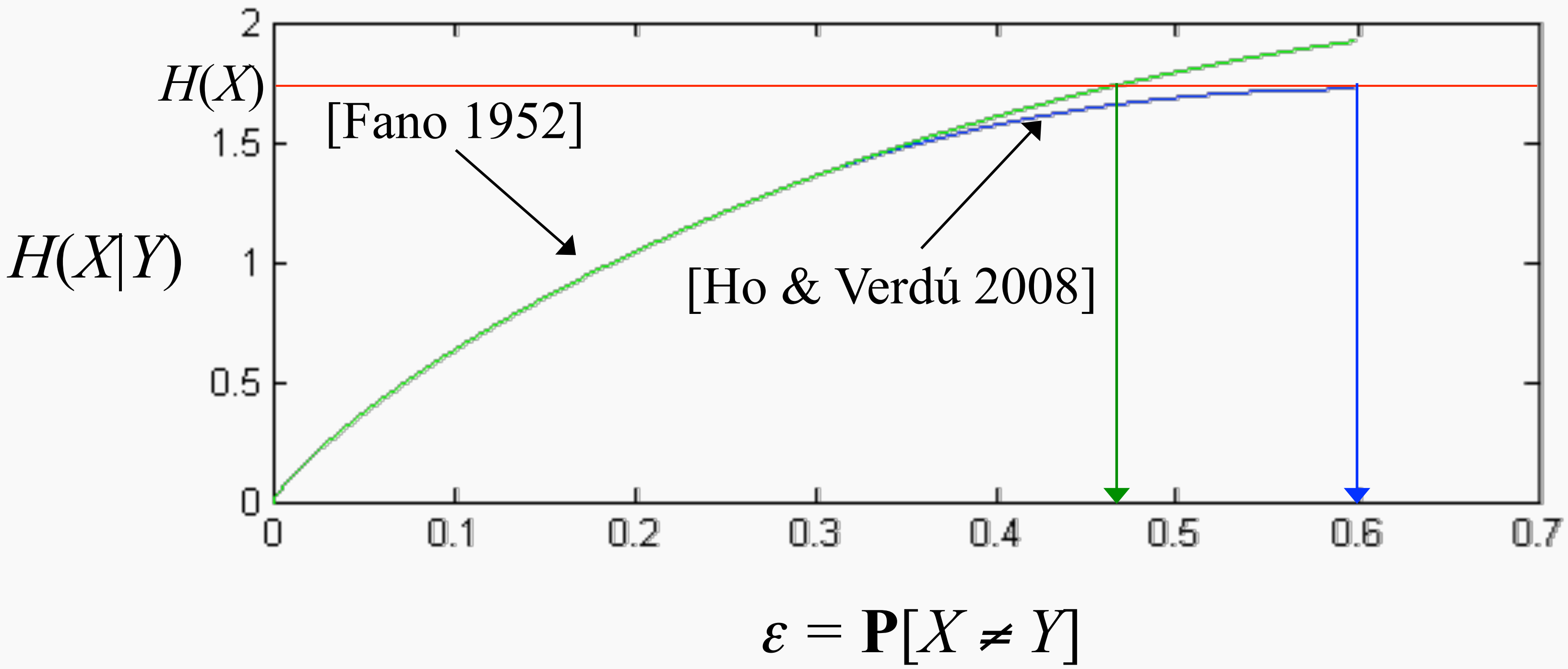
$$\Phi_{S^k}^{-1}\left(H(S^k) - \sup_{X^n} I(X^n;Y^n) \right) \leq \mu_k$$

Information Theoretic Security

- Weak secrecy $\lim_{n \to \infty} n^{-1} I(X^n; Y^n) = 0$ has been considered in [Csiszár & Körner 78, Broadcast channel] and some seminal papers.
- [Wyner 75, Wiretap channel I] only stated that "a large value of the equivocation implies a large value of P_{ew}", where the equivocation refers to $n^{-1} H(X^n \mid Y^k)$ and P_{ew} means μ_n.
- It is important to clarify what exactly weak secrecy implies.

Weak Secrecy

- E.g., $P_X = (0.4, 0.4, 0.1, 0.1)$.



Weak Secrecy

- Theorem 12: For any discrete stationary memoryless source (i.i.d. source) with distribution P_X, if

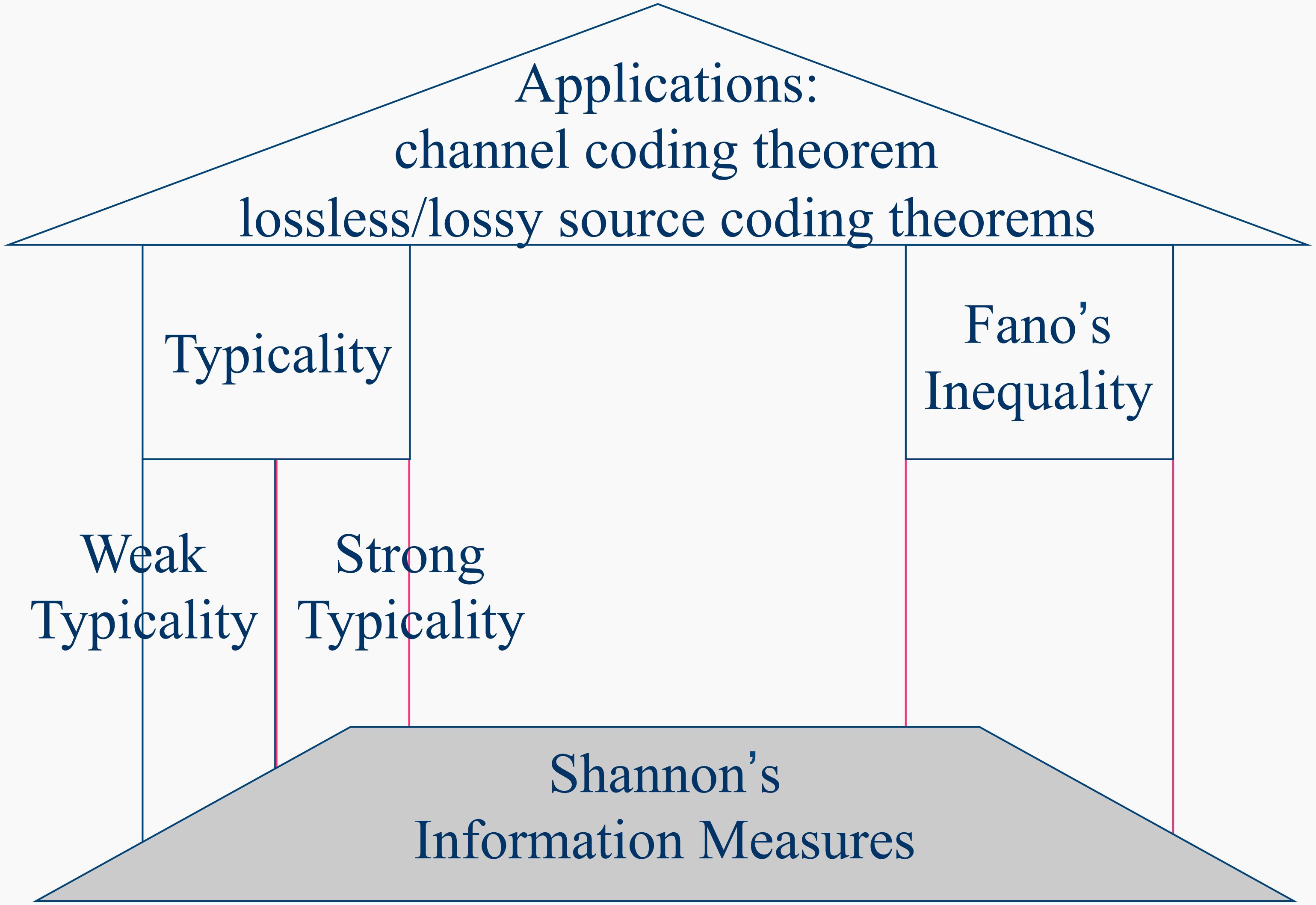
$$\lim_{n\to\infty} n^{-1} I(X^n;Y^n) = 0,$$

- Then

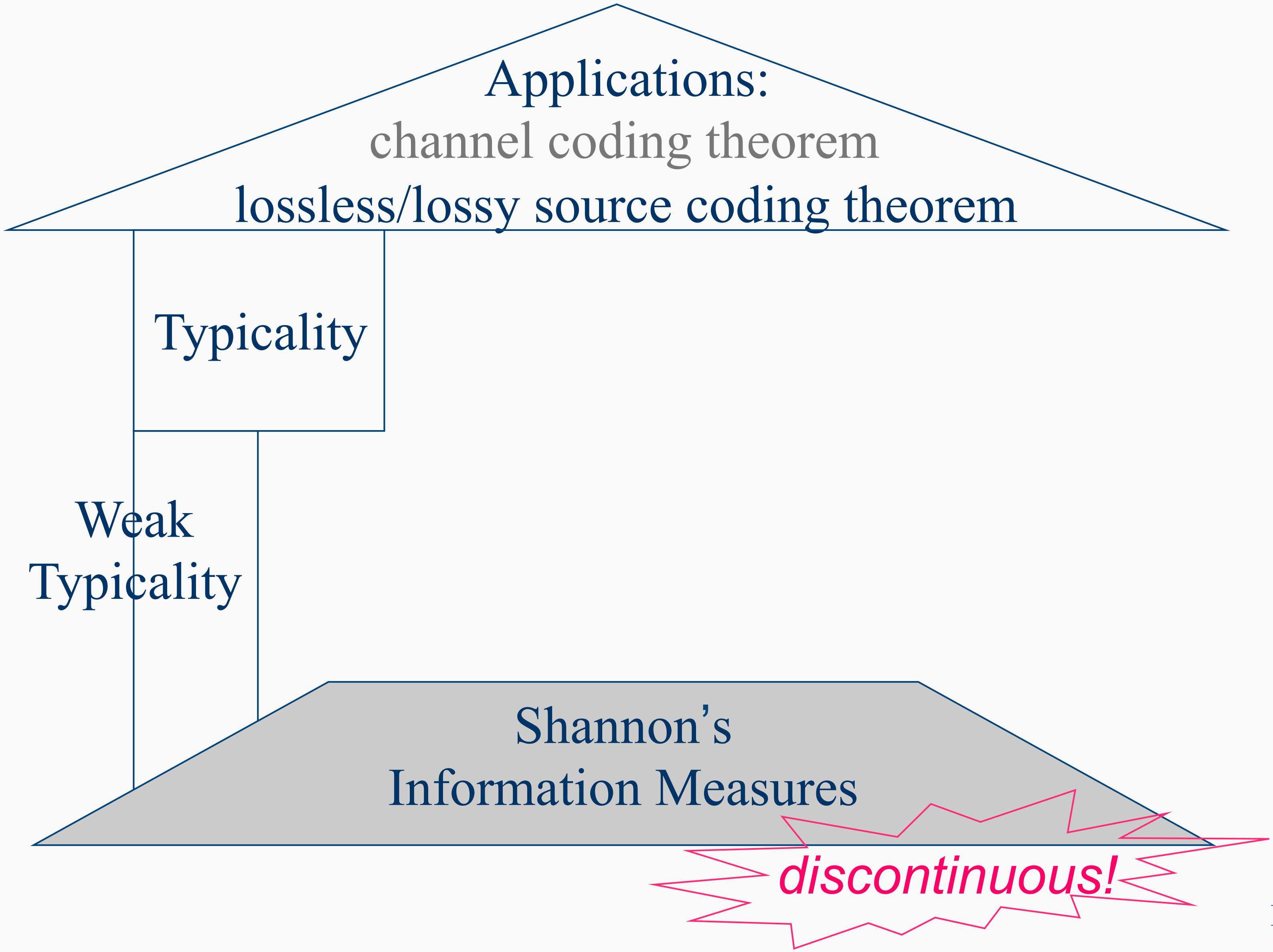
$$\lim_{n\to\infty} \lambda_n = \lambda_{\max} \quad \text{and} \quad \lim_{n\to\infty} \mu_n = 1.$$

- Remark:
  - Weak Secrecy together with the stationary source assumption is insufficient to show the maximum error probability.
  - The proof is based on the tight upper bound on $H(X|Y)$ in terms of error probability.

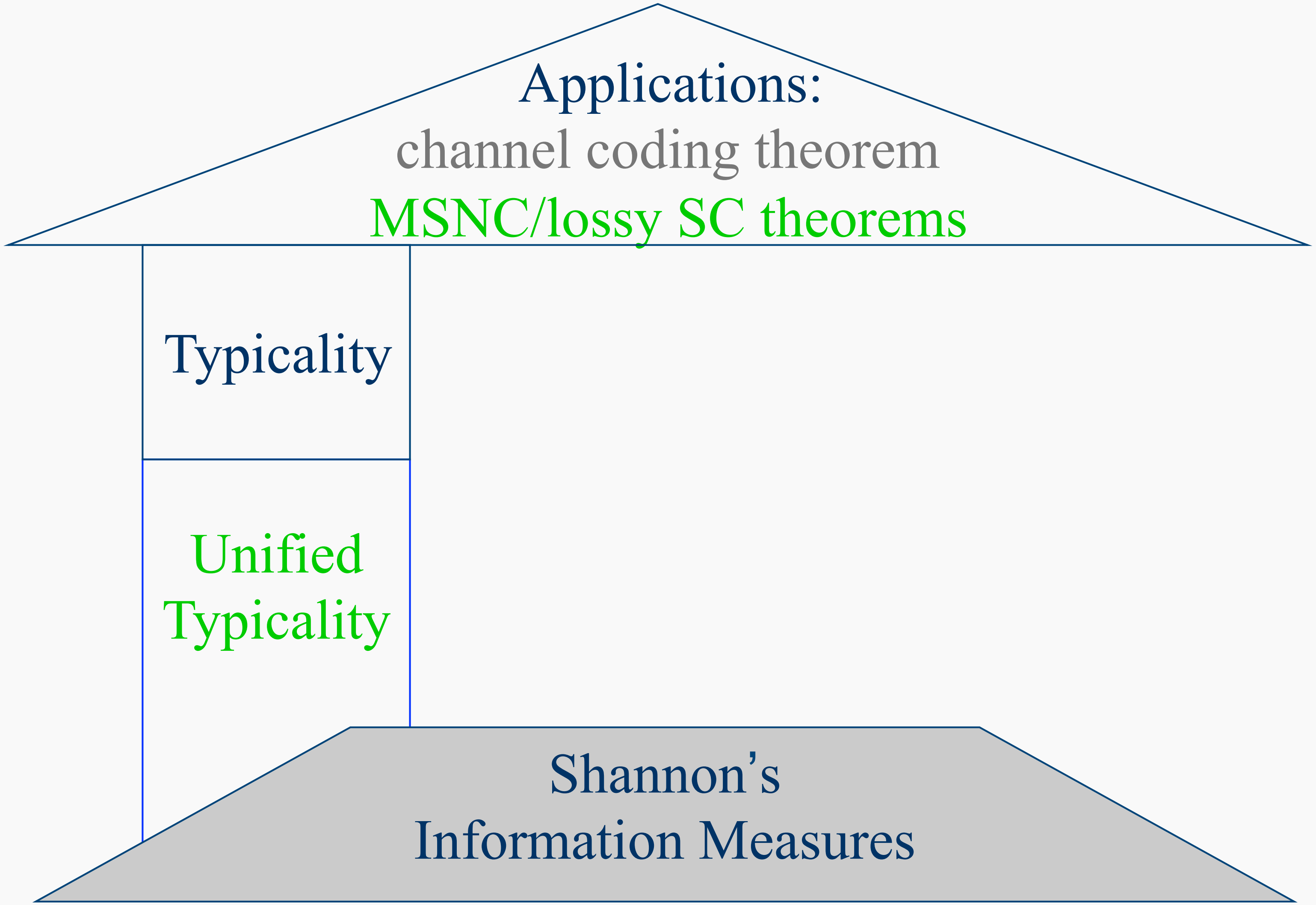
Summary


Applications:
channel coding theorem
lossless/lossy source coding theorems
Typicality
Fano's Inequality
Weak Typicality
Strong Typicality
Shannon's Information Measures

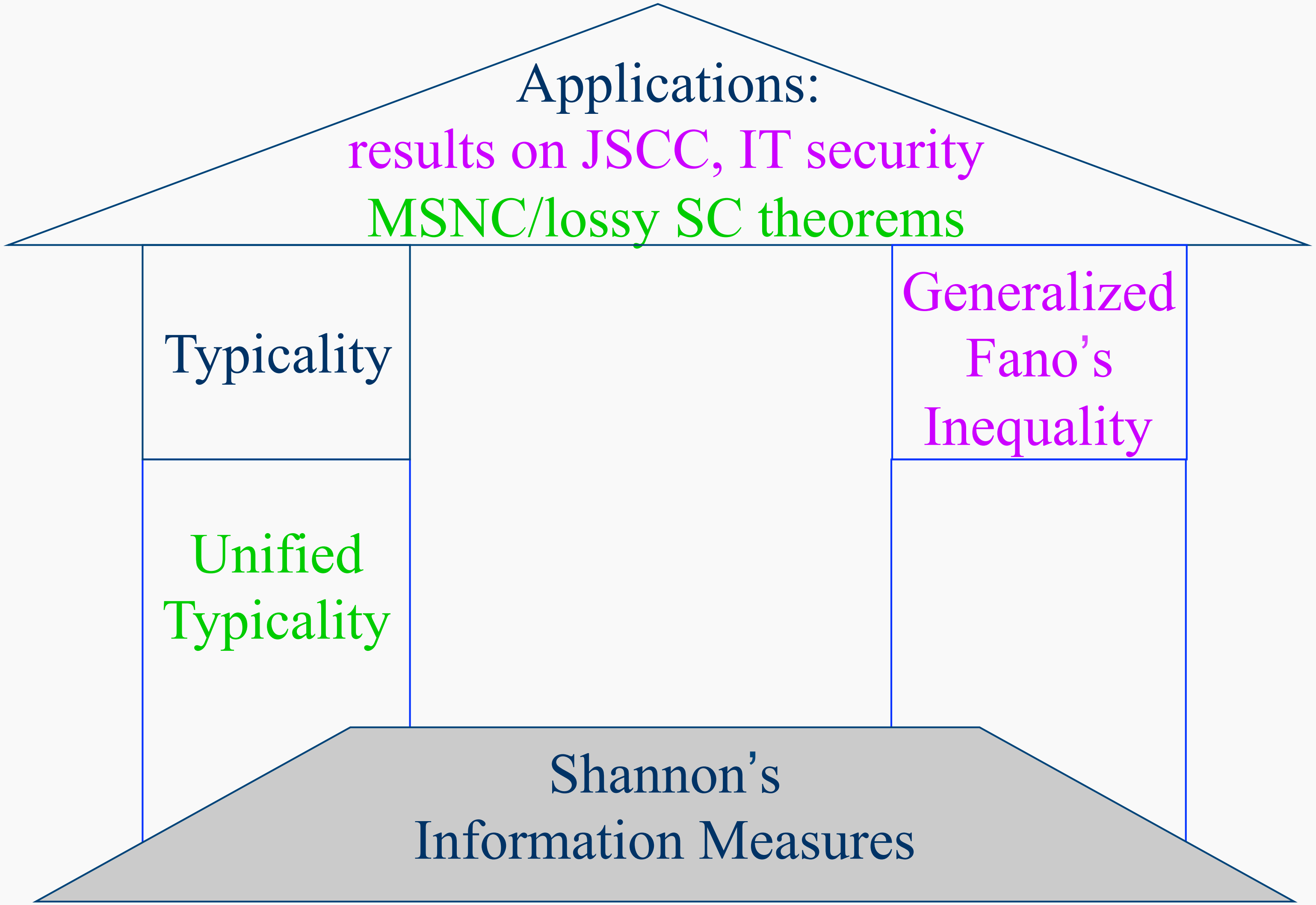
On Countably Infinite Alphabet



Unified Typicality



Generalized Fano's Inequality



Perhaps...

A lot of fundamental research in information theory are still waiting for us to investigate.

References

- S.-W. Ho and R. W. Yeung, “On the Discontinuity of the Shannon Information Measures,” *IEEE Trans. Inform. Theory*, vol. 55, no. 12, pp. 5362–5374, Dec. 2009.
- S.-W. Ho and R. W. Yeung, “On Information Divergence Measures and a Unified Typicality,” *IEEE Trans. Inform. Theory*, vol. 56, no. 12, pp. 5893–5905, Dec. 2010.
- S.-W. Ho and S. Verdú, “On the Interplay between Conditional Entropy and Error Probability,” *IEEE Trans. Inform. Theory*, vol. 56, no. 12, pp. 5930–5942, Dec. 2010.
- S.-W. Ho, “On the Interplay between Shannon's Information Measures and Reliability Criteria,” in *Proc. 2009 IEEE Int. Symposium Inform. Theory (ISIT 2009)*, Seoul, Korea, June 28-July 3, 2009.
- S.-W. Ho, “Bounds on the Rates of Joint Source-Channel Codes for General Sources and Channels,” in *Proc. 2009 IEEE Inform. Theory Workshop (ITW 2009)*, Taormina, Italy, Oct. 11-16, 2009.

Q & A

Why Countably Infinite Alphabet?

- An important mathematical theory can provide some insights which cannot be obtained from other means.
- Problems involve random variables taking values from countably infinite alphabets.
- Finite alphabet is the special case.
- Benefits: tighter bounds, faster convergent rates, etc.
- In source coding, the alphabet size can be very large, infinite or unknown.

Discontinuity of Entropy

- Entropy is a measure of uncertainty.
- *We can be more and more sure that a particular event will happen as time goes, but at the same time, the uncertainty of the whole picture keeps on increasing*.
- If one found the above statement counter-intuitive, he/she may have the concept that entropy is continuous rooted in his/her mind.
- The limiting probability distribution may not fully characterize the asymptotic behavior of a Markov chain.

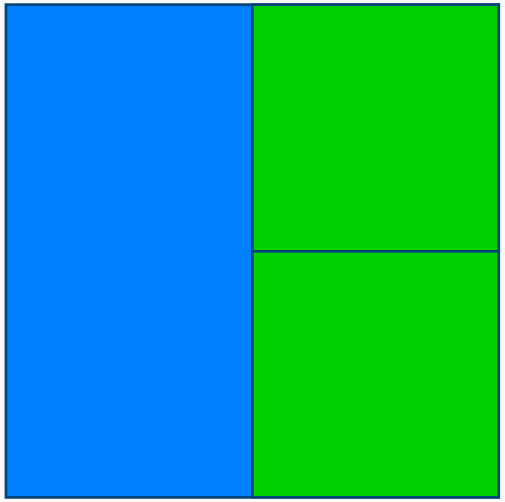
Discontinuity of Entropy

Suppose a child hides in a shopping mall where the floor plan is shown in the next slide.

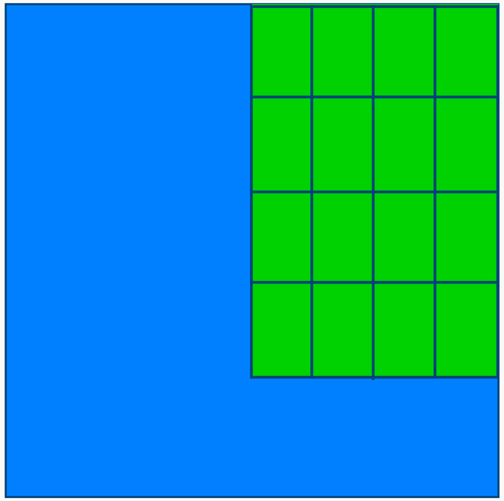
In each case, the chance for him to hide in a room is directly proportional to the size of the room.

We are only interested in which room the child locates in but not his exact position inside a room.

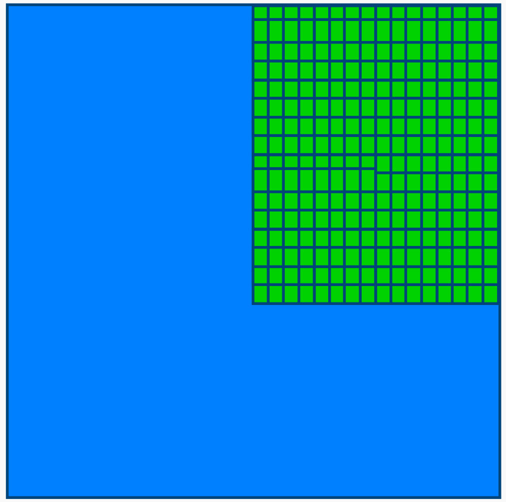
Which case do you expect is the easiest to locate the child?



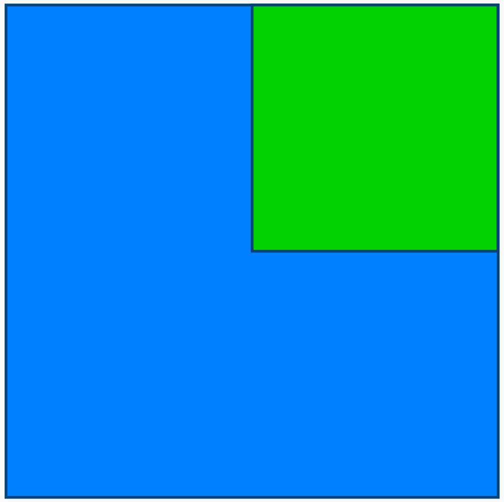
Case A
1 blue room +
2 green rooms



Case B
1 blue room +
16 green rooms



Case C
1 blue room +
256 green rooms



Case D
1 blue room +
4096 green rooms

	Case A	Case B	Case C	Case D
The chance in the blue room	0.5	0.622	0.698	0.742
The chance in a green room	0.25	0.0326	0.00118	0.000063

Discontinuity of Entropy

From Case A to Case D, the difficulty is increasing. By the Shannon entropy, the uncertainty is increasing although the probability of the child being in the blue room is also increasing.

We can continue to construct this example and make the chance in the blue room approaching to 1!

The critical assumption is that the number of rooms can be unbounded. So we have seen that
"There is a very sure event" and "large uncertainty of the whole picture" can exist at the same time.

Imagine there is a city where everyone has a normal life everyday with probability 0.99.
With probability 0.01, however, any kind of accident that beyond our imagination can happen.
Would you feel a big uncertainty about your life if you were living in that city?

$$\lim_{n\to\infty} n^{-1} I(X^n; Y^k) = 0$$

- Weak secrecy is insufficient to show the maximum error probability.
- Example 1: Let W, V and X_i be binary random variables.
- Suppose W and V are independent and uniform.
- Let

$$X_i = \begin{cases} W & V = 0 \\ \text{independent and uniform} & V = 1 \end{cases}$$

$$\tilde{\lambda}_{\max} = 1 - \max_x P_X(x) = 0.5$$

$$\tilde{\mu}_{\max} = \lim_{n\to\infty}\left(1 - \max_{x^n} P_{X^n}(x^n)\right) = \lim_{n\to\infty}\left(1 - \frac{1}{2}\cdot\frac{1}{2}\right) = \frac{3}{4}$$

Example 1

- Let

Y_1	Y_2	Y_3	Y_4	$\cdots$
$\parallel$	$\parallel$	$\parallel$	$\parallel$	
X_1	X_4	X_9	X_{16}	
X_2	X_3	X_8	X_{15}	
X_5	X_6	X_7	X_{14}	
X_{10}	X_{11}	X_{12}	X_{13}	

$$0 \le \lim_{n\to\infty} n^{-1} I(X^n; Y^k)$$
$$\le \lim_{n\to\infty} n^{-1}\sqrt{n} = 0$$

Choose $\hat{x}^n = \begin{cases} (0,0,\ldots,0) & \text{if } Y_i = 0 \quad \forall i \\ (1,1,\ldots,1) & \text{if } Y_i = 1 \quad \forall i. \end{cases}$

- Then

$$\lim_{n\to\infty} \mu_n = \mathbf{P}[V=1] = \frac{1}{2} < \tilde{\mu}_{\max} = \frac{3}{4}$$

$$\lim_{n\to\infty} \lambda_n = \mathbf{P}[V=1] \cdot \frac{1}{2} = \frac{1}{4} < \tilde{\lambda}_{\max} = \frac{1}{2}$$

Joint Unified Typicality

Can

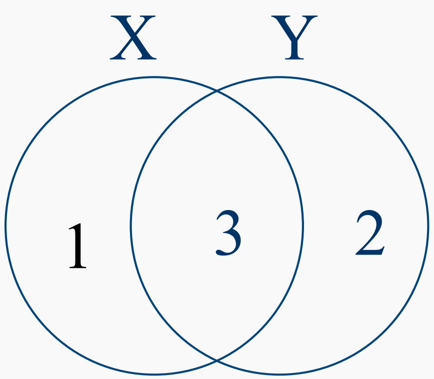
$$D(Q_{XY} \| P_{XY}) + |H(Q_{XY}) - H(P_{XY})| + |H(Q_X) - H(P_X)| + |H(Q_Y) - H(P_Y)| \le \eta.$$

be changed to

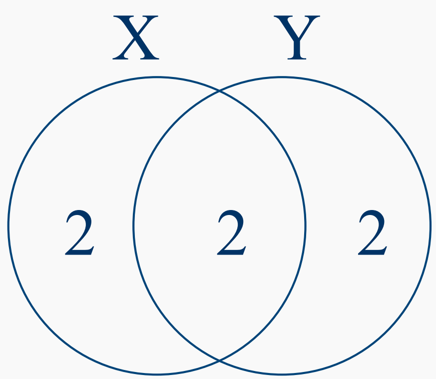
$$D(Q_{XY} \| P_{XY}) + |H(Q_{XY}) - H(P_{XY})| + |H(Q_X) - H(P_X)| \le \eta.$$

?

Ans:



$Q = \{q(xy)\}$



$P = \{p(xy)\}$

$D(Q||P) << 1$

Joint Unified Typicality

Can

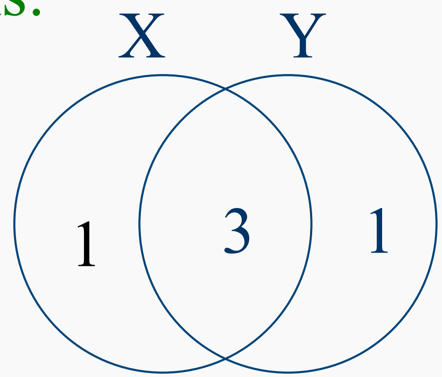
$$D(Q_{XY} \| P_{XY}) + |H(Q_{XY}) - H(P_{XY})| + |H(Q_X) - H(P_X)| + |H(Q_Y) - H(P_Y)| \le \eta.$$

be changed to

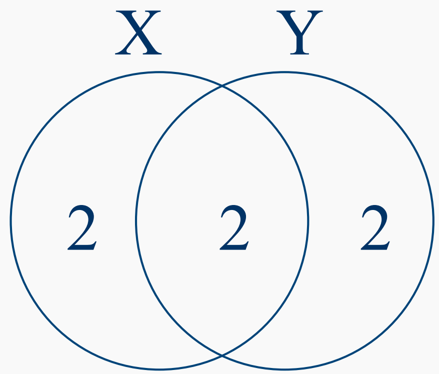
$$D(Q_{XY} \| P_{XY}) + |H(Q_X) - H(P_X)| + |H(Q_Y) - H(P_Y)| \le \eta.$$

?

Ans:



$Q = \{q(xy)\}$



$P = \{p(xy)\}$

$D(Q||P) << 1$

Asymptotic Equipartition Property

- Theorem 5 (Consistency): For any $(\mathbf{x},\mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$,

  if $(\mathbf{x},\mathbf{y}) \in U^n_{[XY]\eta}$, then $\mathbf{x} \in U^n_{[X]\eta}$ and $\mathbf{y} \in U^n_{[Y]\eta}$.

- Theorem 6 (Unified JAEP): For any $\eta > 0$:
  - 1) If $(\mathbf{x}, \mathbf{y}) \in U^n_{[XY]\eta}$, then
    $$2^{-n(H(XY)+\eta)} \leq p(\mathbf{x},\mathbf{y}) \leq 2^{-n(H(XY)-\eta)}$$
  - 2) For sufficiently large n,
    $$\Pr\{(\mathbf{X},\mathbf{Y}) \in U^n_{[XY]\eta}\} > 1-\eta$$
  - 3) For sufficiently large n,
    $$(1-\eta)2^{n(H(XY)-\eta)} \leq \left|U^n_{[XY]\eta}\right| \leq 2^{n(H(XY)+\eta)}$$