From Non-Information-Spectra to Information-Spectra

Te Sun Han
韓太舜

電子情報通信学会総合大会

March 22, 2018

- Sources

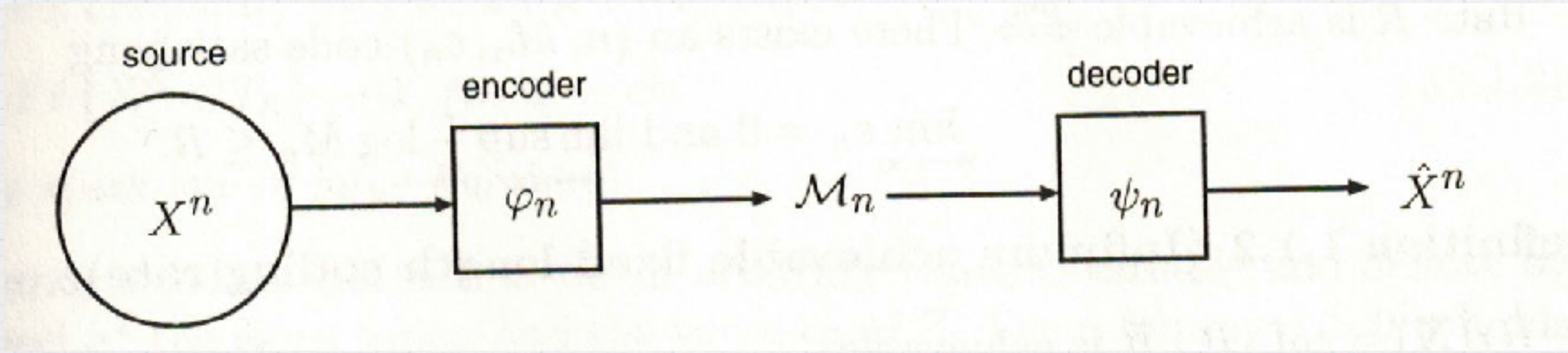
i.i.d. sources,
Markov sources
ergodic sources
non-ergodic sources
non-stationary sources,
general sources

$$X^n = (X_1, X_2, \cdots, X_n), \quad n = 1, 2, \cdots$$

$$\mathbf{X} = \{X^n\}_{n=1}^{\infty}$$

I. Source Coding

- Source coding system



$$R \simeq \frac{1}{n} \log M_n \to R_{\min} \quad (M_n = |\mathcal{M}_n|)$$

$$\varepsilon_n \to 0 \quad (\text{error probability})$$

Minimum achievable rate

$$\overset{=3}{R_{\min}} = H(X)$$

for memoryless sources

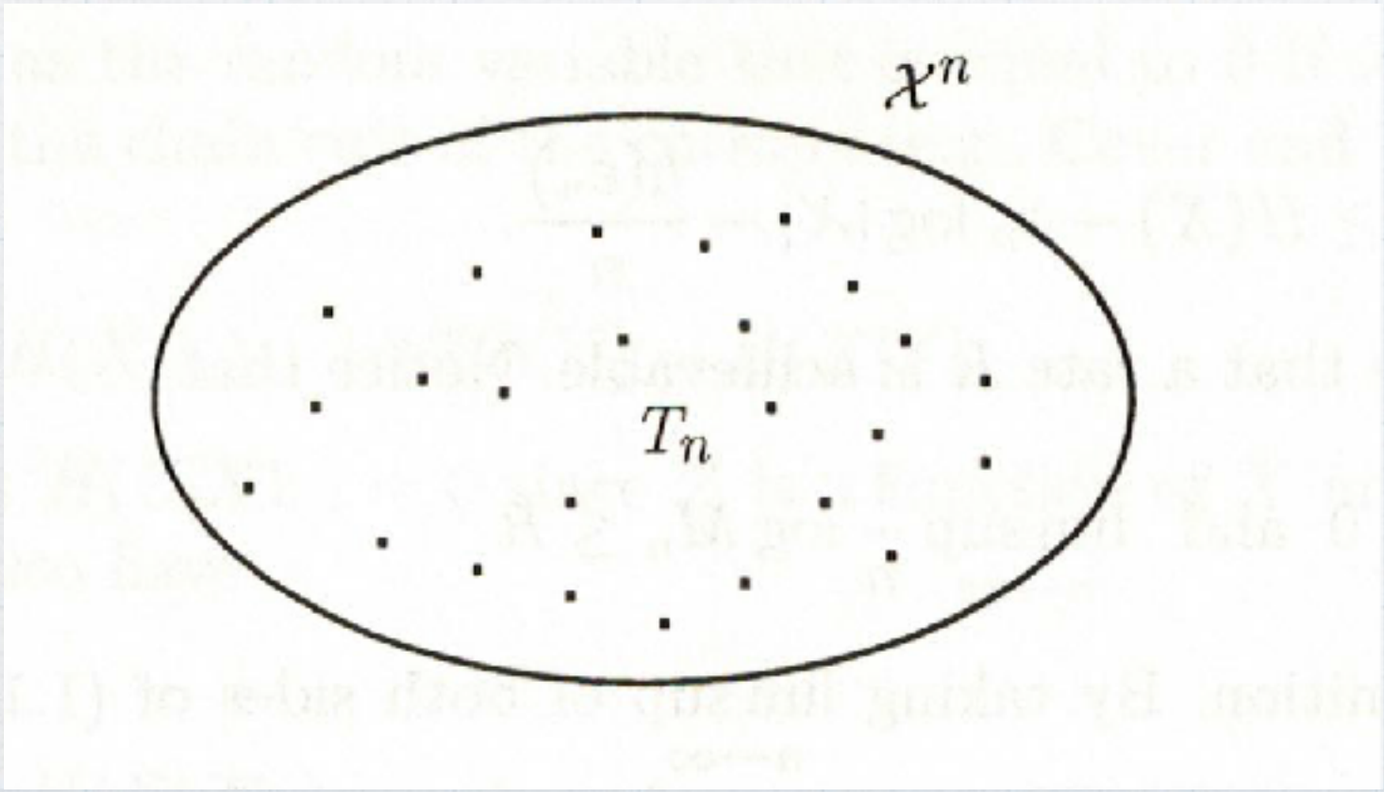
- Typical sequences (for i.i.d. source)

$$T_n = \left\{ \mathbf{x} \in \mathcal{X}^n : \left| \frac{1}{n} \log \frac{1}{P_{X^n}(\mathbf{x})} - H(X) \right| < \gamma \right\},$$

$$p(\mathbf{x}) \simeq 2^{-nH(X)}, \quad \mathbf{x} \in T_n \qquad (H(X) \text{:entropy})$$

$$\Pr\{X^n \in T_n\} \to 1, \quad n \to \infty \qquad \text{(law of large numbers)}$$

cf. Chebyshev



分散が有限に限る

1) Direct part:

Typical sequences だけを送れば良い

=3

$$M_n = |T_n| \simeq 2^{nH(X)}$$

1) Converse : Berger

$$
\begin{aligned}
\log M_n &\overset{=3}{\geq} H(\varphi_n(X^n)) \geq H(\psi_n(\varphi_n(X^n))) \\
&= H(\hat{X}^n) \geq I(X^n; \hat{X}^n) \\
&\equiv H(X^n) - H(X^n|\hat{X}^n) \\
&= nH(X) - H(X^n|\hat{X}^n),
\end{aligned}
$$

$$\overset{=3}{\hat{X}^n} = \psi_n(\varphi_n(X^n))$$

ここで, Fano inequality を使う:

$$
\begin{aligned}
H(X^n|\hat{X}^n) &\overset{=3}{\leq} \varepsilon_n \log(|\mathcal{X}|^n - 1) + h(\varepsilon_n) \\
&\leq n\varepsilon_n \log|\mathcal{X}| + h(\varepsilon_n),
\end{aligned}
$$

where $\varepsilon_n = \Pr\{X^n \neq \hat{X}^n\}$ denotes the error probability.

- $\overset{=3}{|\mathcal{X}|} < \infty,\ \varepsilon_n \to 0$ でないと使えない

1)' Converse: Csiszar-Korner

$$\begin{aligned} S_n &= \{\mathbf{x} \in \mathcal{X}^n | \mathbf{x} = \psi_n(\varphi_n(\mathbf{x}))\} \\ \varepsilon_n &= \Pr\{X^n \neq S_n\}, \\ \delta_n &= \Pr\{X^n \notin T_n\} \end{aligned}$$

Then, since $\Pr\{X^n \in T_n \cap S_n\} \geq 1 - (\varepsilon_n + \delta_n)$

$$\begin{aligned} M_n 2^{-nH(X)} &\geq \Pr\{X^n \in T_n \cap S_n\} \\ &\geq 1 - (\varepsilon_n + \delta_n) \end{aligned}$$

Hence

$$\frac{1}{n} \log M_n \geq H(X) + \frac{1}{n} \log[1 - (\varepsilon_n + \delta_n)]$$

- $|\mathcal{X}| < \infty,\ \varepsilon_n \to 0$ でなくとも良い

- How to generalize the argument to the case of $|\overset{=3}{\mathcal{X}}| = \infty$ or $\varepsilon_n \not\to 0$

- Typical sequences in the case of $|\mathcal{X}| = \infty$

(Poisson分布，幾何分布など)

$$T_n = \left\{ \mathbf{x} \in \mathcal{X}^n : \left| \frac{1}{n} \log \frac{1}{P_{X^n}(\mathbf{x})} - H(X) \right| < \gamma \right\},$$

$$p(\mathbf{x}) \simeq 2^{-nH(X)}, \quad \mathbf{x} \in T_n$$

$$\Pr\{X^n \in T_n\} \to 1, \quad n \to \infty$$

分散が無限でも良い

- Khintchin's weak law of large numbers
- Individual ergodic theorem

- What is the characteristics of the above arguments:

- Let us consider the probability distribution of

$$\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)}$$ (self-entropy density)

Example 1

1) i.i.d. source

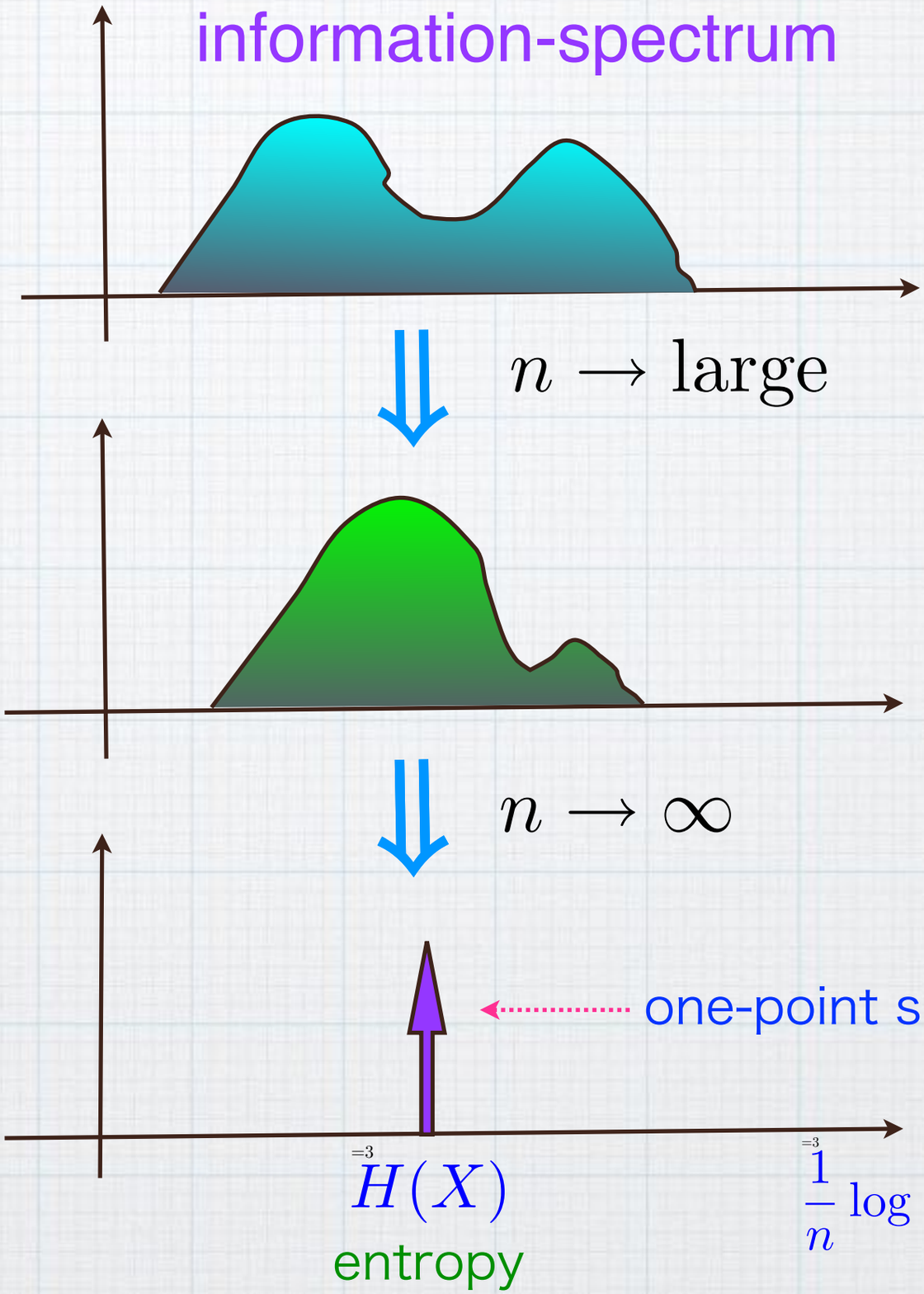
2)Markov source

3)ergordic source

⇓

AEP

(Asymptotic Equi-Partition)



- How about the case of non-ergodic cases

Example 2

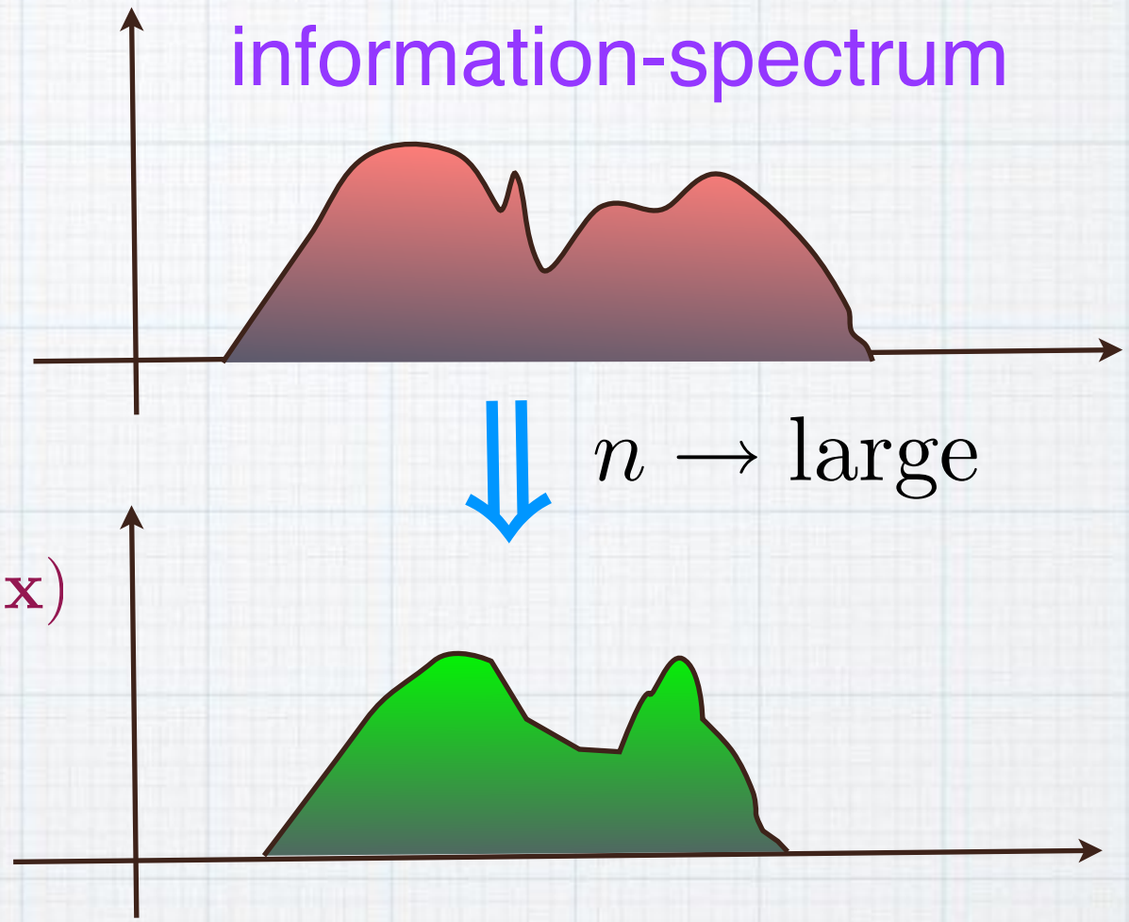
4)mixed source

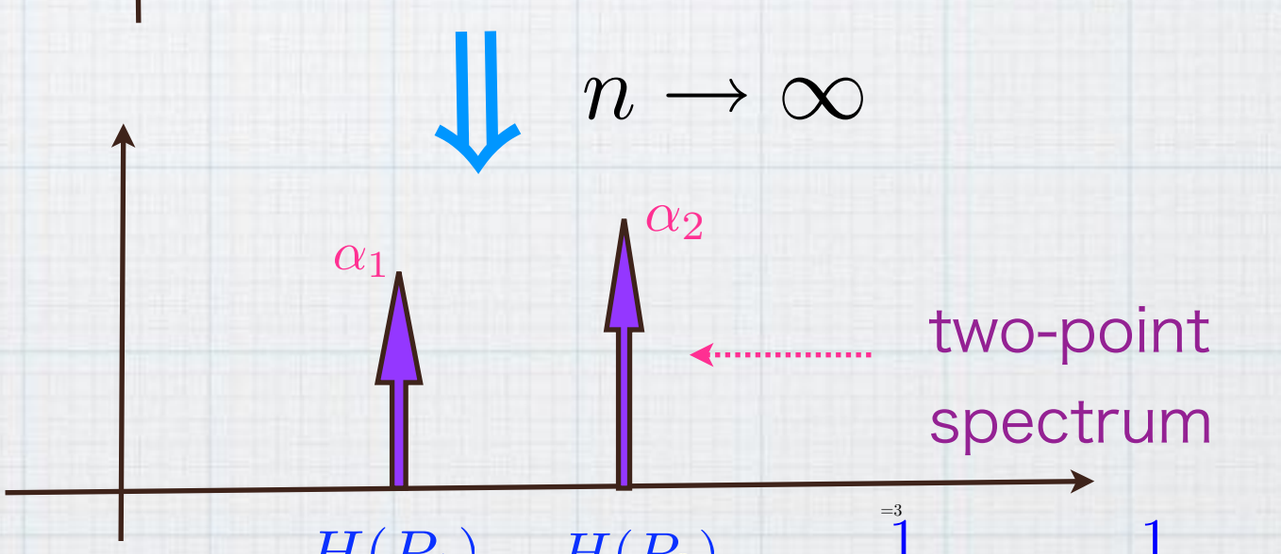
$$P_{X^n}(\mathbf{x}) = \alpha_1 P_{X_1^n}(\mathbf{x}) + \alpha_2 P_{X_2^n}(\mathbf{x})$$

$$X_1^n : \text{i.i.d.} \sim P_1$$

$$X_2^n : \text{i.i.d.} \sim P_2$$

$$H(P_1) \neq H(P_2)$$



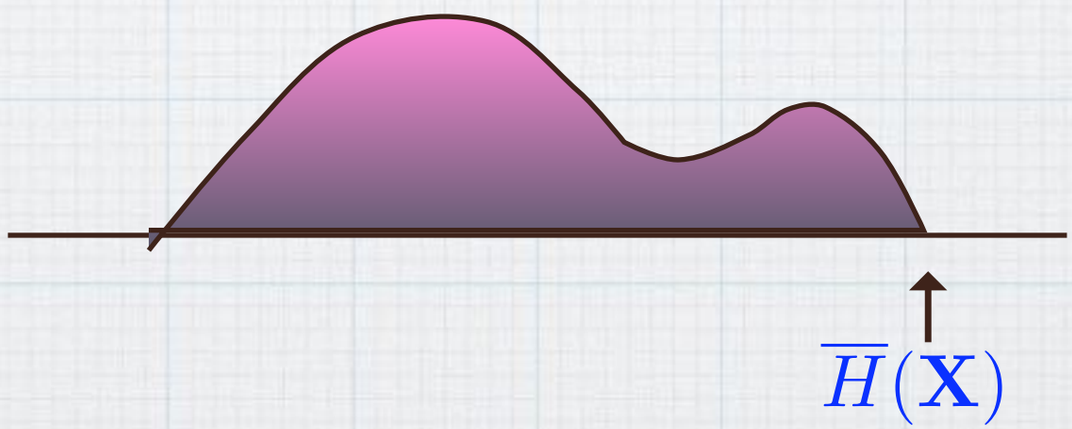


non-AEP

- Generalization of typical sequences

$$T_n = \left\{ \mathbf{x} \in \mathcal{X}^n : \frac{1}{n} \log \frac{1}{P_{X^n}(\mathbf{x})} < \overline{H}(\mathbf{X}) + \gamma \right\}.$$

information spectrum



- Typical sequences $(\varepsilon = 0)$の概念を捨てると

Direct part:

Lemma 1 Let M_n be an arbitrary given positive integer.
Then, for all $n = 1, 2, \cdots$ there exists an (n, M_n, ε_n)-code satisfying

$$\varepsilon_n \leq \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq \frac{1}{n}\log M_n\right\}.$$

Converse:

Lemma 2 For all $n = 1, 2, \cdots$, any (n, M_n, ε_n)-code satisfies

$$\varepsilon_n \geq \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq \frac{1}{n}\log M_n + \gamma\right\} - e^{-n\gamma},$$

where $\gamma > 0$ is an arbitrary constant.

- Csiszar uses these lemmas at his Institute

What is happening?

A kind of “Induction Jump”

帰納的飛躍

bottom-up から top-down へ

- Remark:
- Finite-length analysis: for any n
- Strong converse $(\varepsilon_n \to 0 \quad \text{or} \quad \varepsilon_n \to 1)$ for ergodic sources
- These formulas enabled us to treat the finite error probability case, that is,

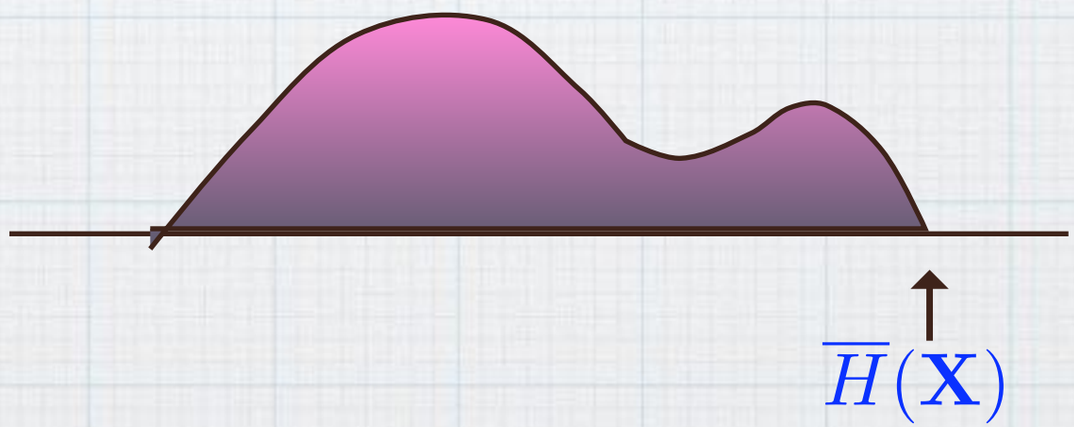
$$\limsup_{n\to\infty} \varepsilon_n \le \varepsilon \quad (0 \le \forall \varepsilon < 1)$$

Theorem 1:

$$R_f \overset{=3}{=} \overline{H}(\mathbf{X})$$

optimal rate

information spectrum



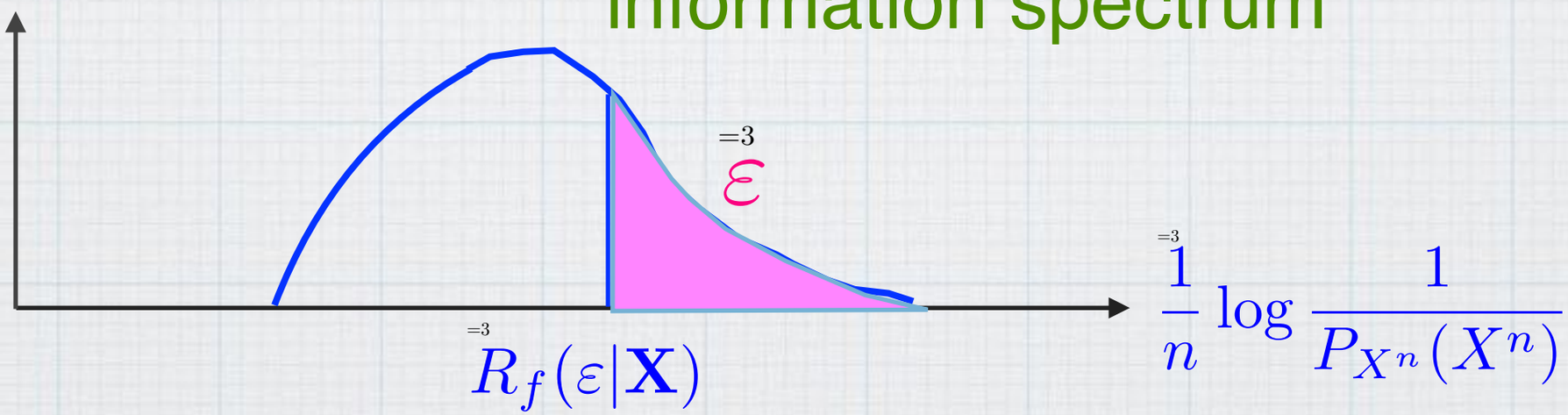
Theorem 2:

$$R_f(\varepsilon|\mathbf{X}) = \inf\{R \mid F(R) \le \varepsilon\}$$

optimal rate

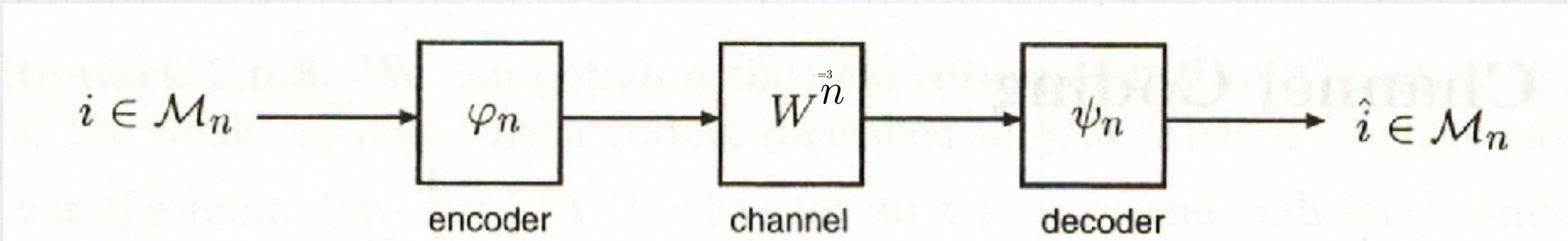
where $F(R) = \limsup_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \ge R\right\}$

information spectrum



II. Channel Coding

- Channel coding system



$$R \simeq \frac{1}{n} \log M_n \to R_{\max} \quad (M_n = |\mathcal{M}_n|)$$

$$\varepsilon_n \longrightarrow 0 \quad (\text{error probability})$$

Channels

memoryless channels,
Markov channels
ergodic channels
non-ergodic channels
non-stationary channels,
general channels

$\Big\}$ infinite length inputs (cf. Gray)

$$\overset{=3}{W^n} = (W_1, W_2, \cdots, W_n), \quad n = 1, 2, \cdots$$

$$\overset{=3}{\mathbf{W}} = \{W^n\}_{n=1}^{\infty}, \quad W^n(\,\cdot|\,\cdot) : \mathcal{X}^n \to \mathcal{Y}^n$$ (stochastic mapping)

Capacity

$$\overset{=3}{R_{\max}} = \max_{X} I(X;Y)$$

for memoryless channels

- Typical sequences (for memoryless channel)

$$T_n = \left\{ (\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \left| \frac{1}{n} \log \frac{W^n(\mathbf{y}|\mathbf{x})}{P_{Y^n}(\mathbf{y})} - I(X;Y) \right| < \gamma \right\},$$

$$P_{X^n Y^n}(\mathbf{x}, \mathbf{y}) = P_{X^n}(\mathbf{x}) W^n(\mathbf{y}|\mathbf{x})$$

mutual information

Then,

$$\Pr\{X^n, Y^n\} \in T_n\} \to 1, \quad n \to \infty$$

分散が無限でも良い

1) Direct part:

- Generate a code $\mathcal{C}_n = \{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_{M_n}\}$ according to P_{X^n}
- Decode $i \in \mathcal{M}_n$ such that $(\mathbf{u}_i, \mathbf{y}) \in T_n$ when the decoder received $\mathbf{y} \in \mathcal{Y}^n$
- Error probability

$$\varepsilon_n \leq Pr\{X^n Y^n \neq T_n\} + 2^{-n\gamma}$$

Leading to Feinstein lemma

- Another definition of typical sequences:
  (cf. Cover-Thomas)

$$T_{1,n} = \left\{ \mathbf{x} \in \mathcal{X}^n : \left| \frac{1}{n} \log \frac{1}{P_{X^n}(\mathbf{x})} - H(X) \right| < \gamma \right\},$$

$$T_{2,n} = \left\{ \mathbf{y} \in \mathcal{Y}^n : \left| \frac{1}{n} \log \frac{1}{P_{Y^n}(\mathbf{y})} - H(Y) \right| < \gamma \right\},$$

$$T_{3,n} = \left\{ (\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \left| \frac{1}{n} \log \frac{1}{P_{X^n Y^n}(\mathbf{x}, \mathbf{y})} - H(X,Y) \right| < \gamma \right\},$$

===> $T_n = T_{1,n} \cap T_{2,n} \cap T_{3,n}$

Not leading to Feinstein lemma

2) Converse:

$\overset{=3}{X^n}$ distributed uniformly on $\mathcal{C}_n$

$\overset{=3}{\hat{X}^n} = \psi_n(Y^n)$, $\varepsilon_n = \Pr\{X^n \neq \hat{X}^n\}$

$$\begin{aligned}
\overset{=3}{\log M_n} &= H(X^n) \\
&= I(X^n; \hat{X}^n) + H(X^n|\hat{X}^n) \\
&\leq I(X^n; Y^n) + H(X^n|\hat{X}^n),
\end{aligned}$$

Fano inequality

$$\overset{=3}{H(X^n|\hat{X}^n)} \leq \varepsilon_n \log M_n + h(\varepsilon_n).$$

$$\overset{=3}{\frac{1}{n}\log M_n} \leq \frac{1}{n} I(X^n; Y^n) + \varepsilon_n \frac{\log M_n}{n} + \frac{h(\varepsilon_n)}{n},$$

- $\overset{=3}{|\mathcal{X}| = \infty, |\mathcal{Y}| = \infty}$ の場合でもよい！

- What is the characteristics of the above arguments (for channels):

information-spectrum

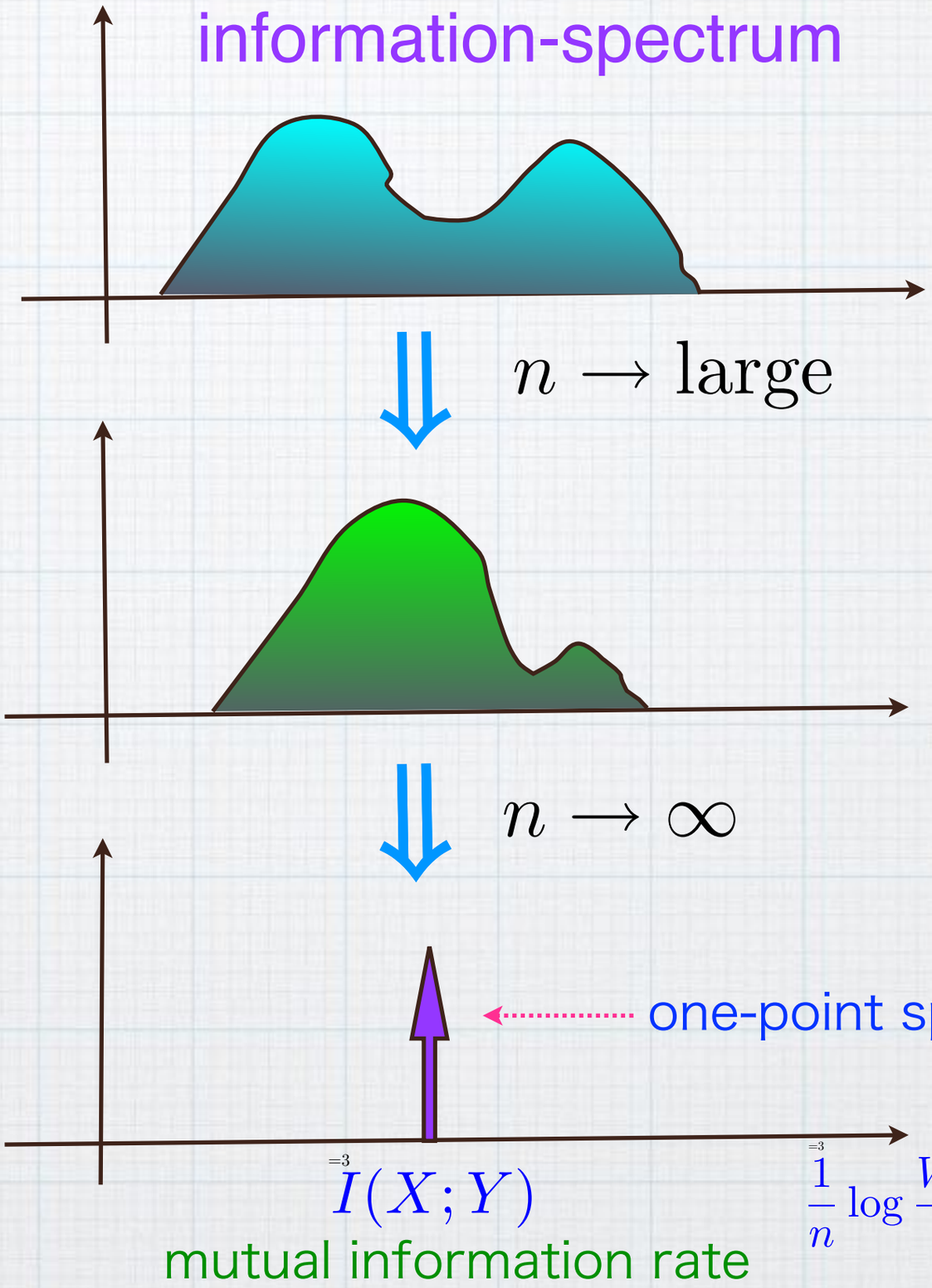
- Let us consider the probability distribution of

$$\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)}$$

(mutual information density)

Example 3

1) memoryless channels
2) “ergodic” channels



⇓

AEP
(Asymptotic
Equi-Partition)

- How about the case of non-ergodic cases

Example 4

3) mixed channels

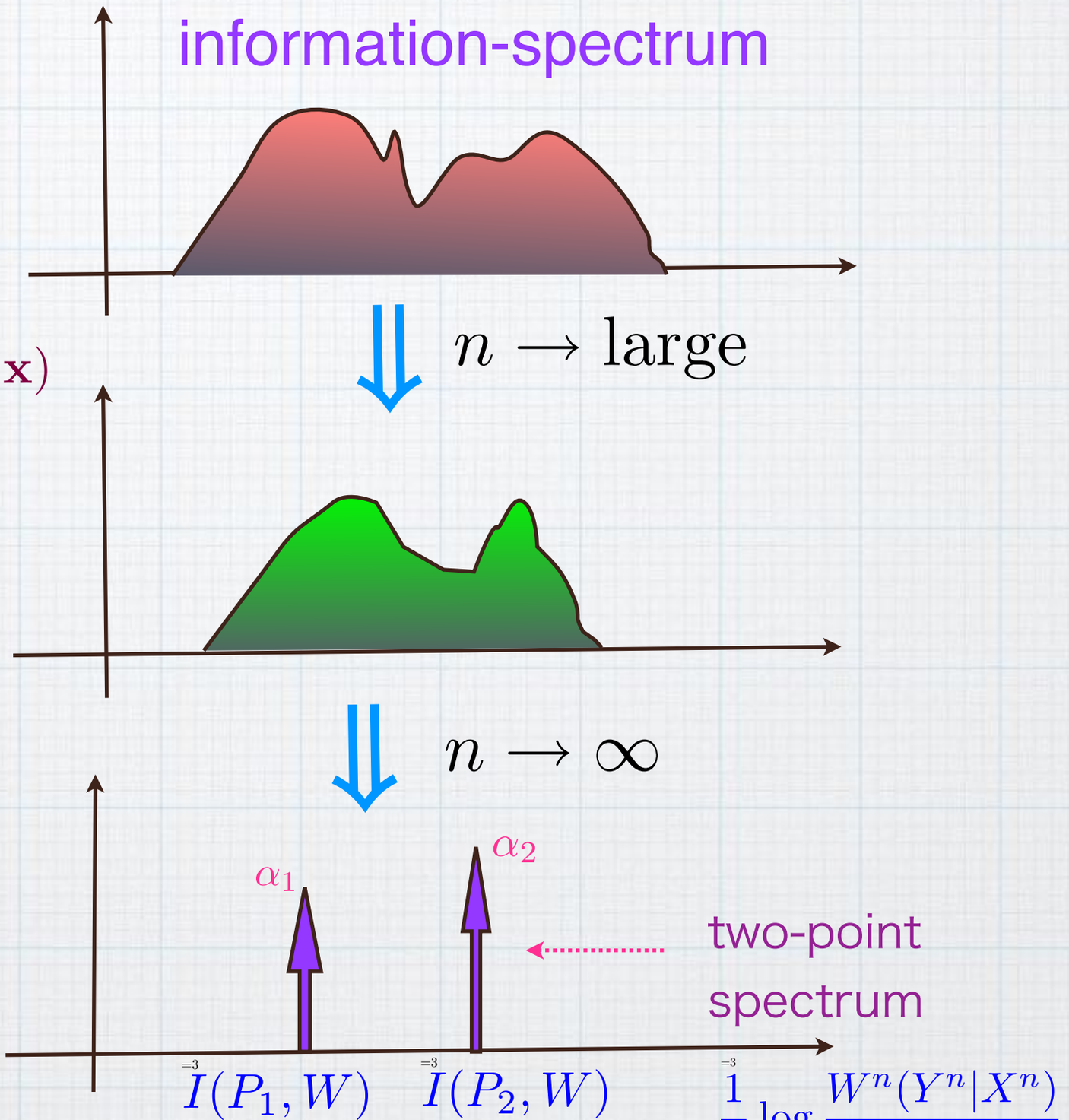
$$\overset{=3}{W}{}^n(\mathbf{y}|\mathbf{x}) = \alpha_1 W_1^n(\mathbf{y}|\mathbf{x}) + \alpha_2 W_2^n(\mathbf{y}|\mathbf{x})$$

$$\overset{=3}{W}{}_1^n : \text{i.i.d.} \sim (P_1, W)$$

$$\overset{=3}{W}{}_2^n : \text{i.i.d.} \sim (P_2, W)$$

$$\overset{=3}{I}(P_1, W) \neq I(P_2, W)$$

non-AEP



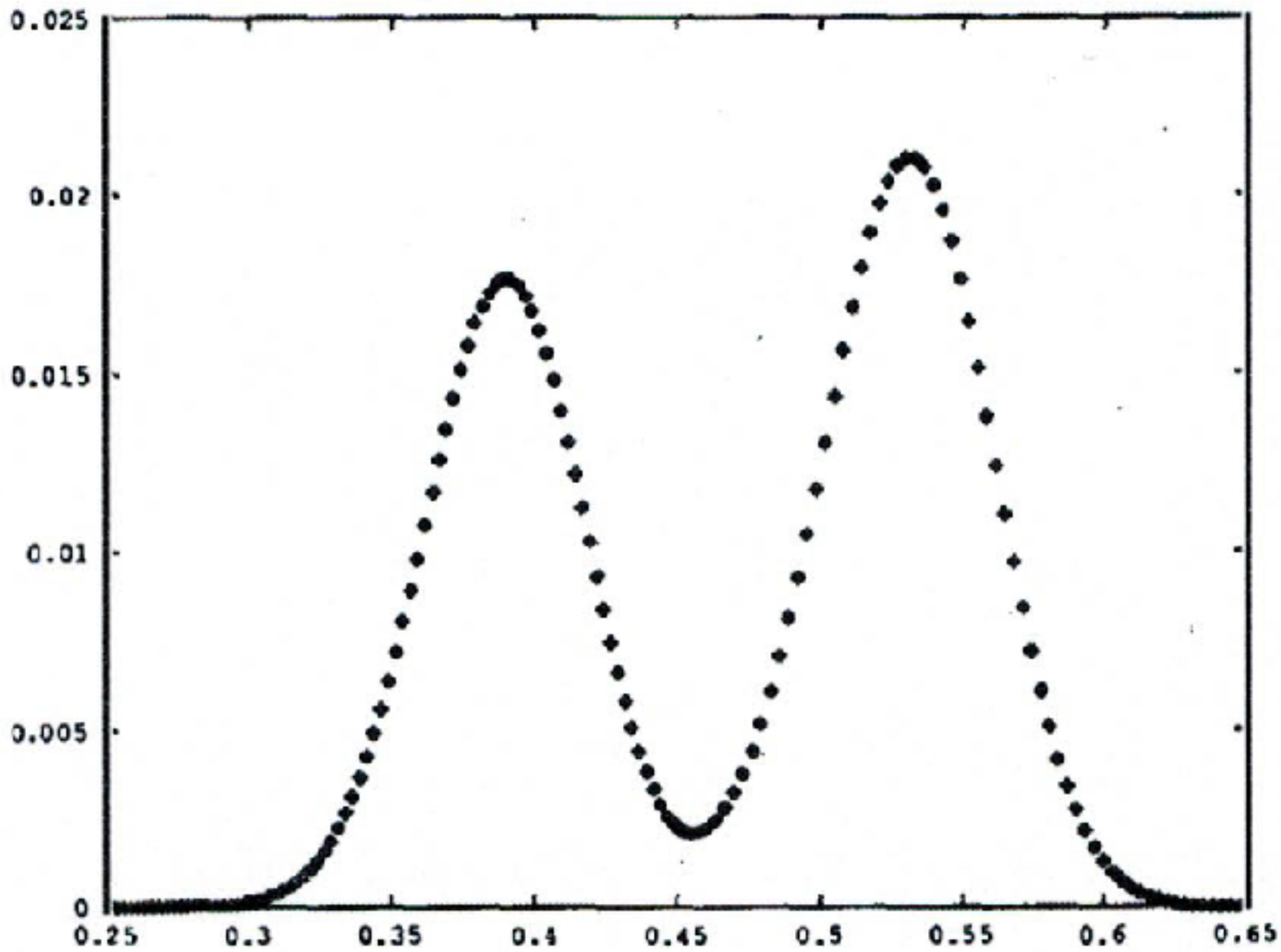
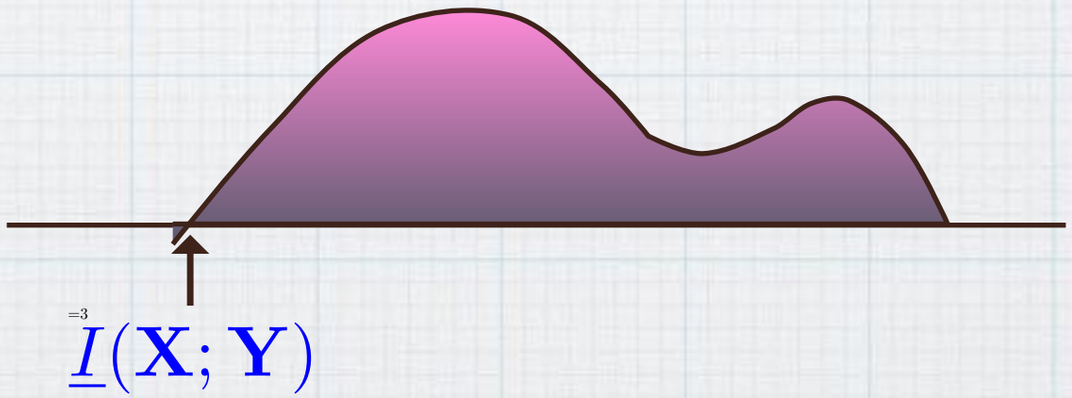


Fig. 3. Information spectrum (probability mass function of the normalized information density) of channel in Example 2 with $\delta_1 = 0.1$, $\delta_2 = 0.15$, $\alpha = 0.5$, $n = 1000$.

- Generalization of typical sequences for channels

$$T_n = \left\{ (\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \frac{1}{n} \log \frac{W^n(\mathbf{y}|\mathbf{x})}{P_{Y^n}(\mathbf{y})} > \underline{I}(\mathbf{X};\mathbf{Y}) - \gamma \right\},$$

information spectrum



- Typical sequences $(\varepsilon = 0)$の概念を捨てると

Lemma 3 (Achievability: Feinstein, 1954)

Let $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$ be an arbitrary input to the channel $\mathbf{W} = \{W^n\}_{n=1}^{\infty}$ and $\mathbf{Y} = \{Y^n\}_{n=1}^{\infty}$ the output from the channel corresponding to $\mathbf{X}$.

Given an arbitrary positive integer M_n , there exists an (n, M_n, ε_n)-code such that

$$\varepsilon_n \leq \Pr\left\{\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} \leq \frac{1}{n}\log M_n + \gamma\right\} + e^{-n\gamma}$$

for all $n = 1, 2, \cdots$, where $\gamma > 0$ an arbitrary const.

Lemma 4 (Converse: Verdu-Han, 1994)

Let X^n be the random variable uniformly distributed on an (n, M_n, ε_n)-code, and Y^n the output of the channel $\mathbf{W} = \{W^n\}_{n=1}^{\infty}$ due to X^n Then, for all $n = 1, 2, \cdots,$ it holds that

$$\varepsilon_n \geq \Pr\left\{\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} \leq \frac{1}{n}\log M_n - \gamma\right\} - e^{-n\gamma}$$

where $\gamma > 0$ is an arbitrary constant.

Lemma 4' (Converse: Hayashi-Nagaoka)

Let X^n be the random variable uniformly distributed on an (n, M_n, ε_n)-code, and Y^n the output of the channel $\mathbf{W} = \{W^n\}_{n=1}^{\infty}$ due to Then, for all $n = 1, 2, \cdots,$ it holds that

$=3$

$$\varepsilon_n \geq \Pr\left\{ \frac{1}{n} \log \frac{W^n(Y^n|X^n)}{Q^n(Y^n)} \leq \frac{1}{n} \log M_n - \gamma \right\} - e^{-n\gamma}$$

where $\gamma > 0$ is an arbitrary constant.

and Q^n is arbitrary.

Remark: This version is used for capacity problem for mixed memoryless channels

<== quantum

Comparison

$$\varepsilon_n \geq \Pr\left\{\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{Q^n(Y^n)} \leq \frac{1}{n}\log M_n - \gamma\right\} - e^{-n\gamma}$$

$$\varepsilon_n \geq \Pr\left\{\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} + \underbrace{\frac{1}{n}\log\frac{P_{Y^n}(Y^n)}{Q^n(Y^n)}} \leq \frac{1}{n}\log M_n - \gamma\right\} - e^{-n\gamma}$$

asymptotically nonnegative in probability

- Hence, inequality is strongest when $Q^n = P_{Y^n}$

What is happening?

A kind of “Induction Jump”

帰納的飛躍

bottom-up から top-down へ

- Remark:
- Finite-length analysis: for any n
- Strong converse $(\varepsilon_n \to 0 \quad \text{or} \quad \varepsilon_n \to 1)$ for "ergodic" sources
- These formulas enabled us to treat the finite error probability case, that is,

$$\limsup_{n\to\infty} \varepsilon_n \le \varepsilon \quad (0 \le \forall \varepsilon < 1)$$

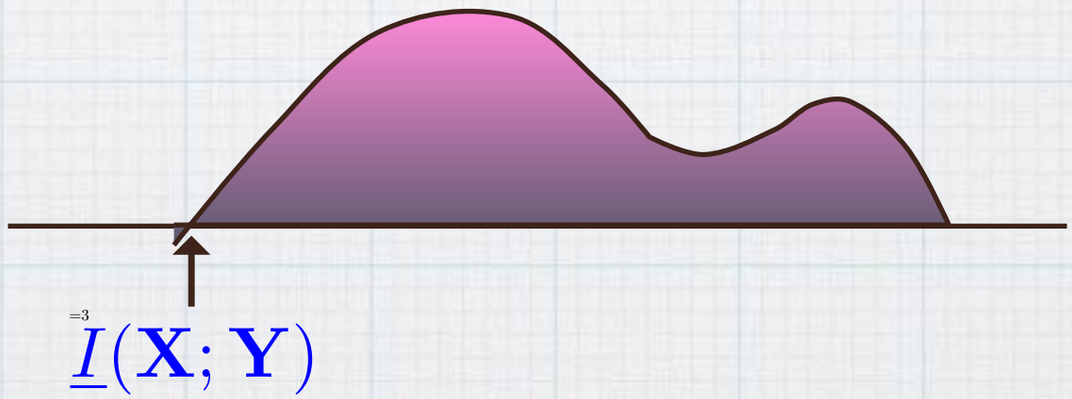
Theorem 3:

$$\overset{=3}{C}(\mathbf{W}) = \sup_{\mathbf{X}} \underline{I}(\mathbf{X};\mathbf{Y})$$

$\overset{=3}{Q}^n$ は現れない

capacity

information spectrum

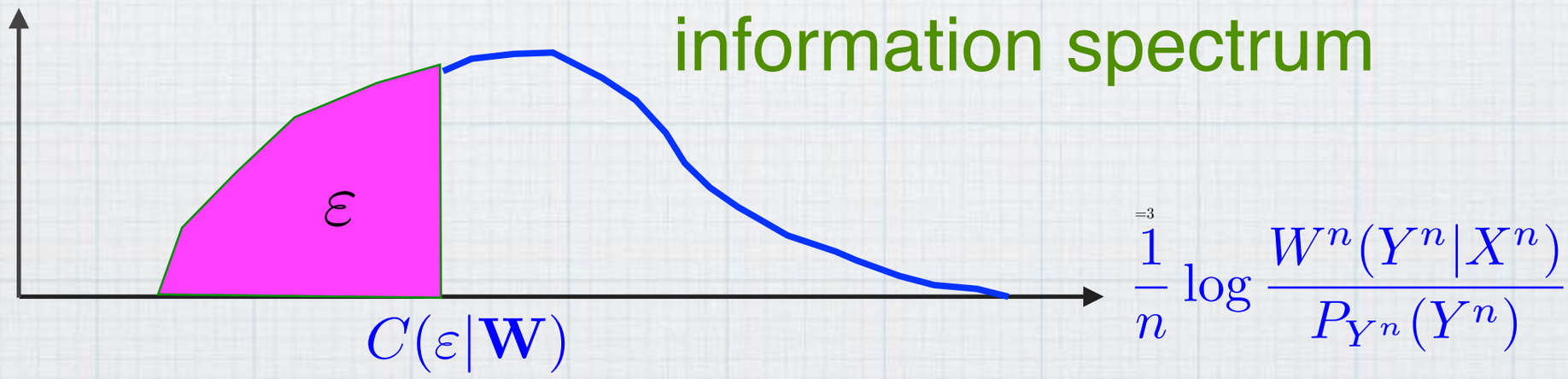


Theorem 4:

$$C(\varepsilon|\mathbf{W}) = \sup\{R \mid G(R) \le \varepsilon\}$$

capacity

where $$G(R) = \limsup_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} \le R\right\}$$



Example 5: Mixed memoryless channel

Yagi-Han-Nomura 2016

$$W^n(\mathbf{y}|\mathbf{x}) = \int_{\Theta} W_\theta^n(\mathbf{y}|\mathbf{x}) dw(\theta)$$

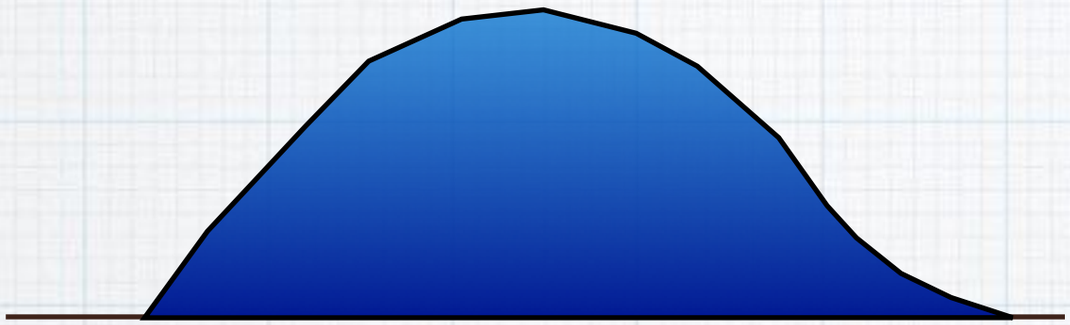
W_θ^n : memoryless channel indexed by $\theta \in \Theta$

Corollary 1: Let $\boldsymbol{W}$ be a general mixed memoryless channel with measure w. For any fixed $\varepsilon \in [0, 1)$, the first-order ε-capacity is given by

$$C_\varepsilon = \sup_P \sup \left\{ R \,\middle|\, \int_{\{\theta|\, I(P,W_\theta) < R\}} dw(\theta) \le \varepsilon \right\}, \qquad (11)$$

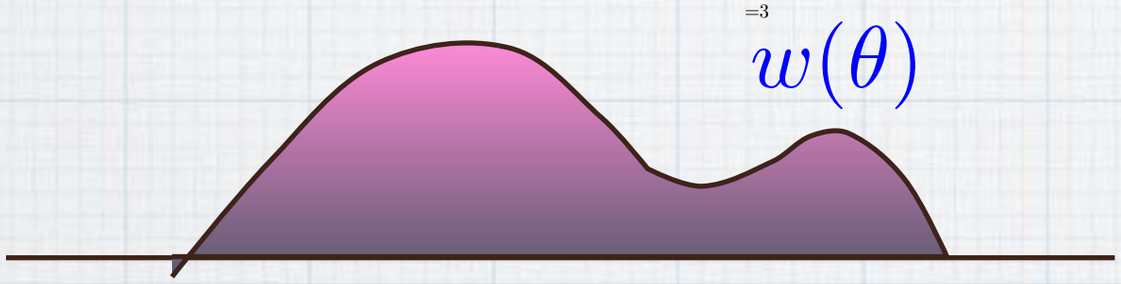
where $\sup_P$ denotes the supremum over the set $\mathcal{P}(\mathcal{X})$ of all probability distributions on $\mathcal{X}$. □

information spectrum



$$\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)}$$

$$\Downarrow\ n\to\infty$$



- information spectrum の分布は
  の極限では $w(\theta)$ に収束する $n\to\infty$
- Examples 1~ 4 はその特別の場合

Remark:

公式 (11) の converse を証明するには,

Lemma 4 ではなく Lemma 4' (Q^n を含む) が必要

一旦, 最適性を緩めて必要な計算を行い,

そのあと最適性を議論する

Lemma 4 (Upper Decomposition Lemma): Let $\boldsymbol{W}$ be a general mixed memoryless channel with measure ω. Then, it holds that

$$\Pr\left\{\frac{1}{n}\log\frac{W^n(Y_\theta^n|X^n)}{P_{Y^n}(Y_\theta^n)} \le z_n\right\}$$

$$\le \Pr\left\{\frac{1}{n}\log\frac{W_\theta^n(Y_\theta^n|X^n)}{P_{Y_\theta^n}(Y_\theta^n)} \le z_n + \frac{\gamma}{\sqrt{n}} + \frac{1}{\sqrt[4]{n^3}}\right\} + e^{-\sqrt{n}\gamma}$$

$$(\forall\theta \in \Theta_n^*), \qquad (36)$$

where $\gamma > 0$ and $z_n > 0$ are arbitrary numbers, and Y_θ^n indicates the output variable due to the input X^n via channel W_θ^n.

Lemma 5 (Lower Decomposition Lemma): Let $\boldsymbol{W}$ be a general mixed memoryless channel with measure w. Given a set of arbitrary i.i.d. product probability distributions Q_θ^n on $\mathcal{Y}^n$, let Q^n be defined by (30). Then, it holds that

$$\Pr\left\{\frac{1}{n}\log\frac{W^n(Y_\theta^n|X^n)}{Q^n(Y_\theta^n)} \le z_n\right\}$$
$$\ge \Pr\left\{\frac{1}{n}\log\frac{W_\theta^n(Y_\theta^n|X^n)}{Q_\theta^n(Y_\theta^n)} \le z_n - \frac{\gamma}{\sqrt{n}} - \frac{1}{\sqrt[4]{n^3}}\right\} - e^{-\sqrt{n}\gamma}$$
$$(\forall\theta \in \Theta_n^*), \qquad (37)$$

where $\gamma > 0$ and $z_n > 0$ are arbitrary numbers, and Y_θ^n indicates the output variable due to the input X^n via channel W_θ^n.

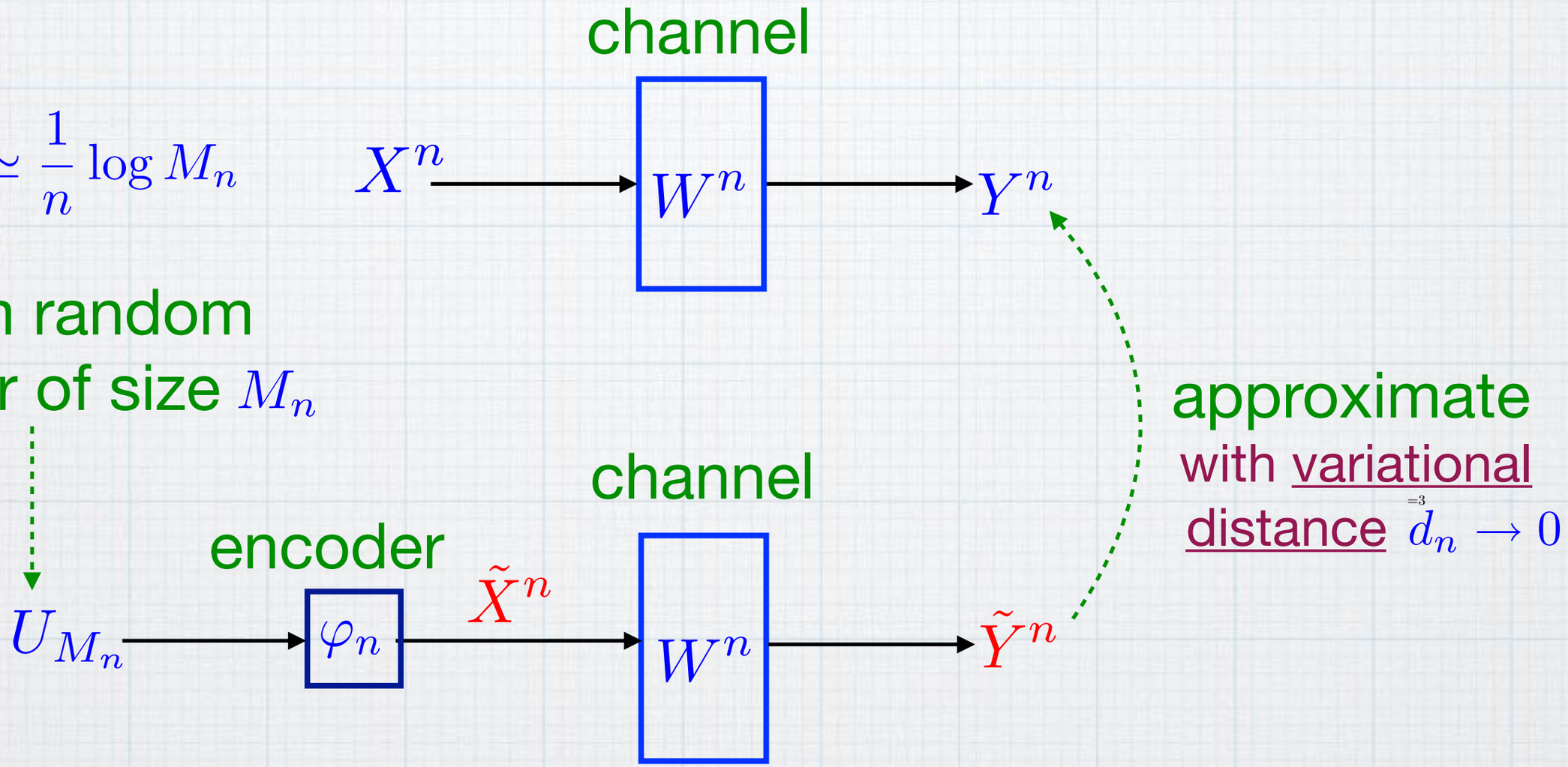
$$Q^n(\boldsymbol{y}) := \int_\Theta Q_\theta^n(\boldsymbol{y})\,dw(\theta) \quad (\forall \boldsymbol{y} \in \mathcal{Y}^n)$$ (mixed memoryless)

Corollary of Corollary 1: $\varepsilon \overset{=3}{=} 0$ の場合

$$C_{\varepsilon \overset{=3}{=} 0} = w\text{-ess.inf}\ I(P, W_\theta)$$

Ahlswede 1968
Han 2003

III. Channel resolvability



- We want to approximate by using a random number of as small size M_n as possible

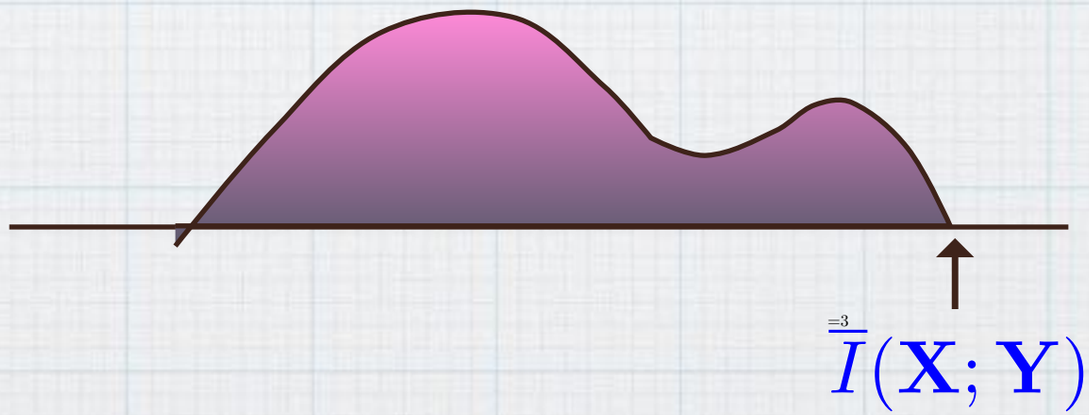
Theorem 5:

$$\overset{=3}{S_{\mathrm{f}}}(\mathbf{W}) = \sup_{\mathbf{X}} \overline{I}(\mathbf{X};\mathbf{Y})$$

for finite input alphabet

resolvability

information spectrum



IV. Resolvability for DMC

Theorem 6:

$$S_{\mathrm{f}}(\varepsilon|\mathbf{X},\mathbf{W}) = \min_{q:qW=pW} I(q;W)$$

Han 1993

Watanabe-Hayashi 2014

$d_n \le \varepsilon$ $(0 \le \varepsilon < 1)$

i.i.d. $\sim p$ given

not depends on ε (strong converse)

i.i.d. $\sim W$ given

VL resolvablelity for DMS Yagi-Han 2017

using a variable-length uniform random number

$$R \simeq \frac{1}{n}\mathrm{E}(L_n)$$

Theorem 7:

$$S_{\mathrm{v}}(\varepsilon|\mathbf{X},\mathbf{W}) = (1-\varepsilon) \min_{q:qW=pW} I(q;W)$$

$$d_n \le \varepsilon$$

$$(0 \le \varepsilon < 1)$$

V. Smooth Renyi entropy (divergence)

1) $H_0^{\delta}(P) \stackrel{\triangle}{=} \min_{\substack{A\subset\mathcal{Z}:\\ P(A)\geq 1-\delta}} \log|A| \quad \rightarrow \quad \overline{H}(\mathbf{X}) = \lim_{\delta\downarrow 0}\limsup_{n\to\infty}\frac{1}{n}H_0^{\delta}(P_n)$

2) $D_0^{\delta}(P\|Q) \stackrel{\triangle}{=} \sup_{\substack{\Phi:\Omega\to[0,1]\\ \int_\Omega \Phi\, dP\geq 1-\delta}} \left\{-\log\int_\Omega \Phi\, dQ\right\}$ (仮説検定の type II error)

$$\rightarrow \quad \lim_{\delta\downarrow 0}\underline{\lim_{n\to\infty}}\frac{1}{n}D_0^{\delta}(P_n\|Q_n) = \{P_n\}\text{-}\underline{\lim_{n\to\infty}}\frac{1}{n}\log\frac{dP_n}{dQ_n} \stackrel{\triangle}{=} \underline{I}(\mathbf{X};\mathbf{Y})$$

cf. Renner, Datta, Wolf, Warsi, Wang, Colbeck, Uyematsu (2004~2014)

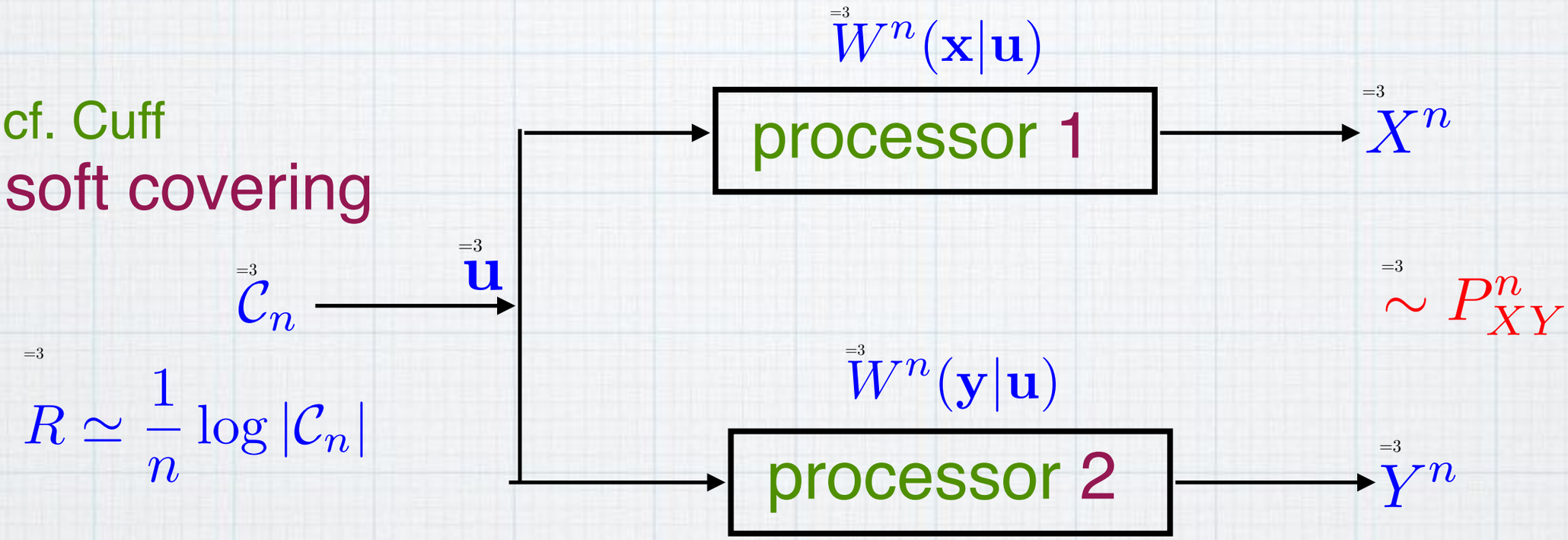
Motivated from quantum theory, cryptography

VI. Common information

(Wyner 1975)

cf. Cuff

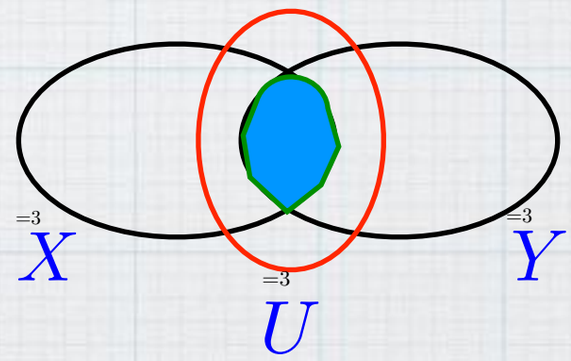
soft covering



$$R \simeq \frac{1}{n} \log |\mathcal{C}_n|$$

$$\frac{1}{n} D(P||Q) \to 0$$

$$C(X;Y) = \max_{X \to U \to Y} I(U;XY)$$



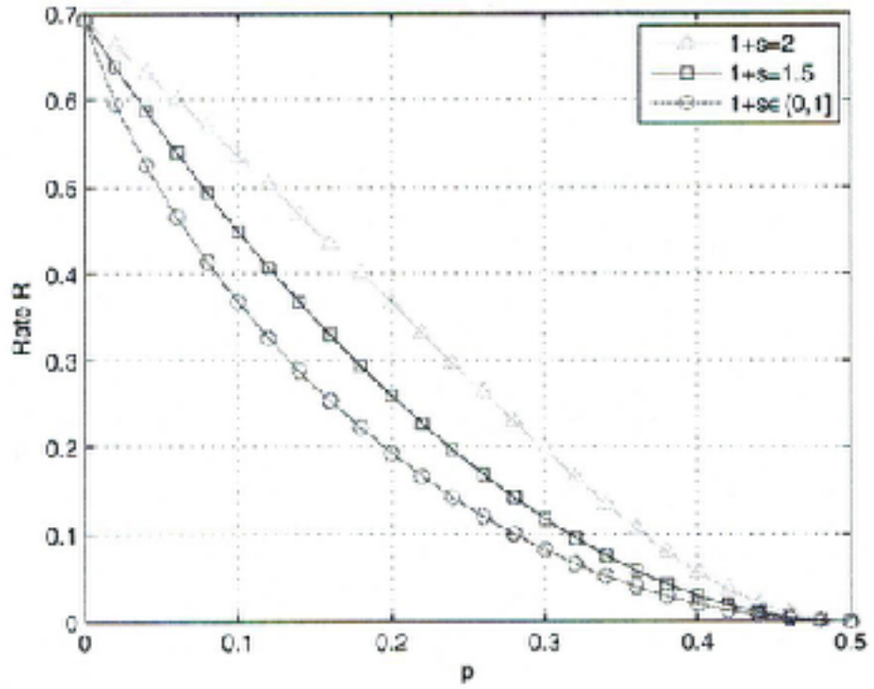
■ 分布間の距離

- variational distance $d(P,Q) = \frac{1}{2}\sum_{x\in\mathcal{X}} |P(x) - Q(x)|$
- Kullback-Leibler divergence $D(P||Q),\ \frac{1}{n}D(P||Q)$
- E_γ -resolvability $E_\gamma(P||Q) = \max_{A\subset\mathcal{X}}[P(A) - \gamma Q(A)], \quad (\gamma \geq 1)$
- Renyi-resolvability (Chernoff distance)

$$D_{1+s}D(P||Q) = \frac{1}{s}\log\sum_{x\in\mathcal{X}} P(x)^{1+s}Q(x)^{-s}, \quad s \geq -1$$

- resolvability with feedback
- variable-length resolvability

Renyi-resolvability



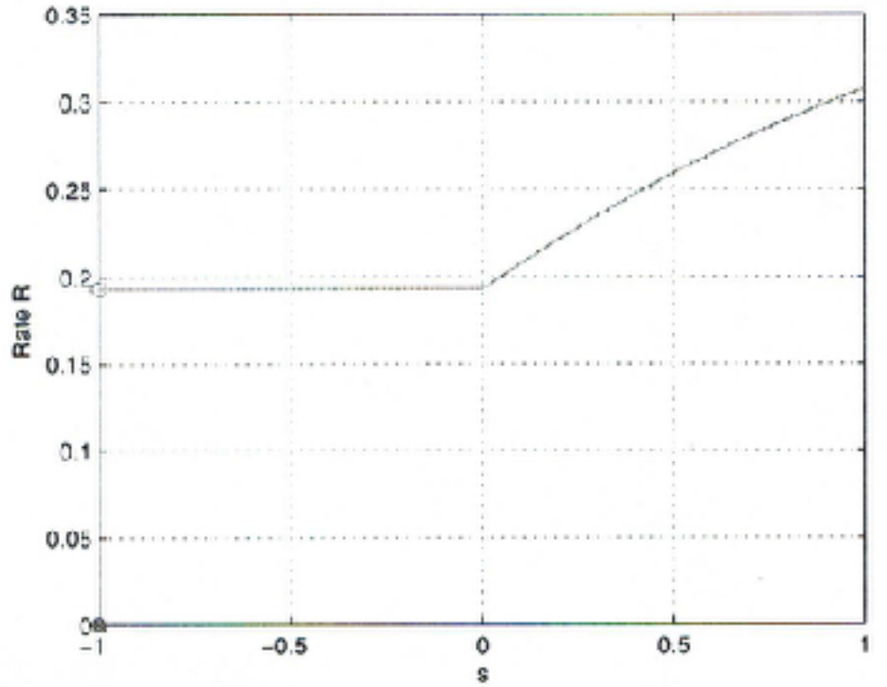
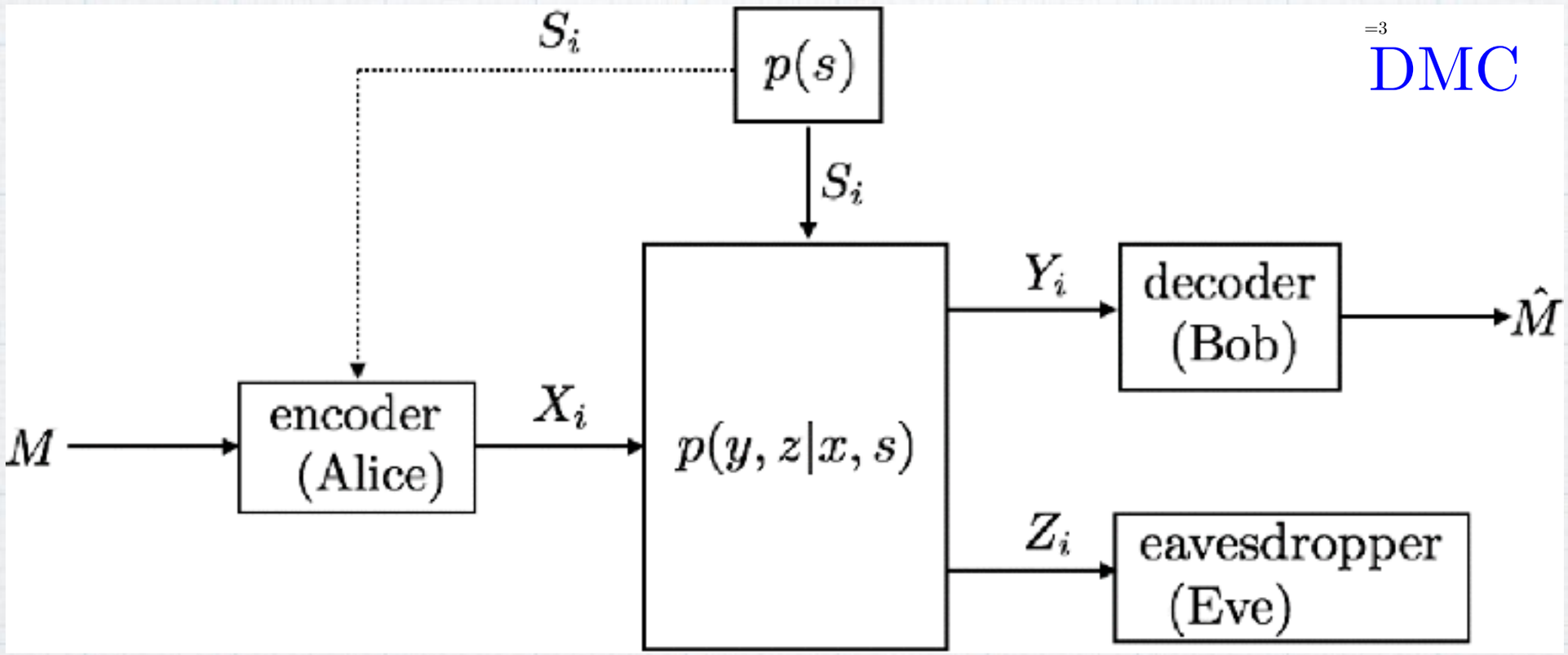


Fig. 3. Illustration of the optimal rates for vanishing resolvability in (17) for the BSC $Y = X \oplus V, V \sim \text{Bern}(p)$ and $Q_Y = \text{Bern}(0.5)$. For the bottom subfigure, $p = 0.2$.

L. Yu and V. Y. F. Tan "Rényi Resolvability and Its Applications to the Wiretap Channel," arXiv:1707.00810, 2017.

VII. Wiretap channel (Wyner 1975)



=3 DMC

Secrecy criterion: (strong)

=3 $\lim_{n\to\infty} D_{1+s}(P_{MZ^n}||P_M Q_{Z^n}) = 0$

=3 $s = 0$ のとき =3 $\lim_{n\to\infty} D(P_{MZ^n}||P_M Q_{Z^n}) = 0$

strong secrecy capacity (Lu-Tan 2018)

$$\overset{=3}{C_{1+s}(Q_Z)} = \max_{P_{UX}:P_X W_{Z|X}=Q_Z} (I(U;Y) - \tilde{R}_{1+s}(P_{UX}, W_{Z|X}, Q_Z))$$

where

$$\tilde{R}_{1+s}(P_{UX}, W_{Z|X}, Q_Z)) \overset{=3}{=} \begin{cases} \max_{P_{Z|UX}} \left[-\frac{1+s}{s} D(P_{Z|UX}||W_{Z|X}|P_{UX}) \right. & s \in (0,1] \\ \qquad \left. + D(P_{Z|U}||Q_Z|P_U) \right], & \\ I(U;Z), & s \in (-1,0] \\ 0, & s = -1 \end{cases}$$

$s \overset{=3}{=} 0$ のとき $C_{1+s}(Q_Z) \overset{=3}{=} \max_{P_{UX}:P_X W_{Z|X}=Q_Z} (I(U;Y) - I(U;Z))$

(Hayashi 2006)

Strong soft covering lemma (Cuff 2015)

$$\overset{=3}{\Pr}\left\{ D(P_{Z^n|\mathcal{C}_n} || P_{Z^n}) > 2^{-\gamma_1 n} \right\} \leq 2^{-2^{\gamma_2 n}}$$

$\overset{=3}{\text{if}}\ R > I(X;Z)$

$\overset{=3}{(W^n : X^n \to Z^n)}$

with some $\overset{=3}{\gamma_1 > 0, \gamma_2 > 0}$ and $\overset{=3}{\mathcal{C}_n \sim P_{X^n}}$

does not necessarily mean

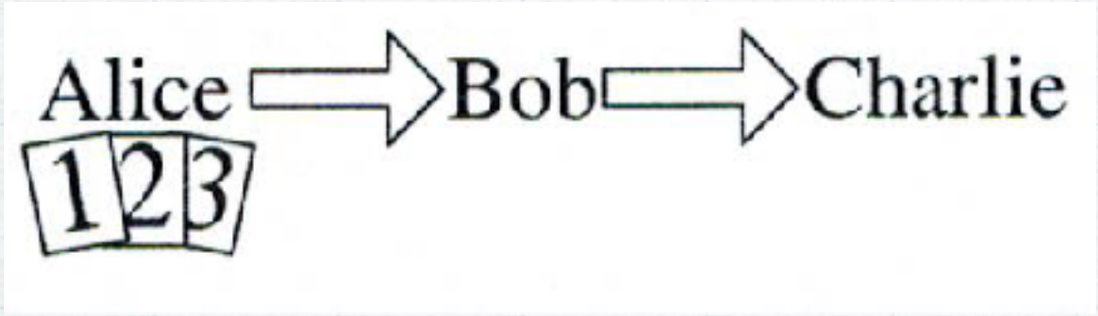
$$\overset{=3}{\mathrm{E}}\left\{ D(P_{Z^n|\mathcal{C}_n} || P_{Z^n}) \right\} \to 0, \quad (n \to \infty)$$ (weak soft covering)

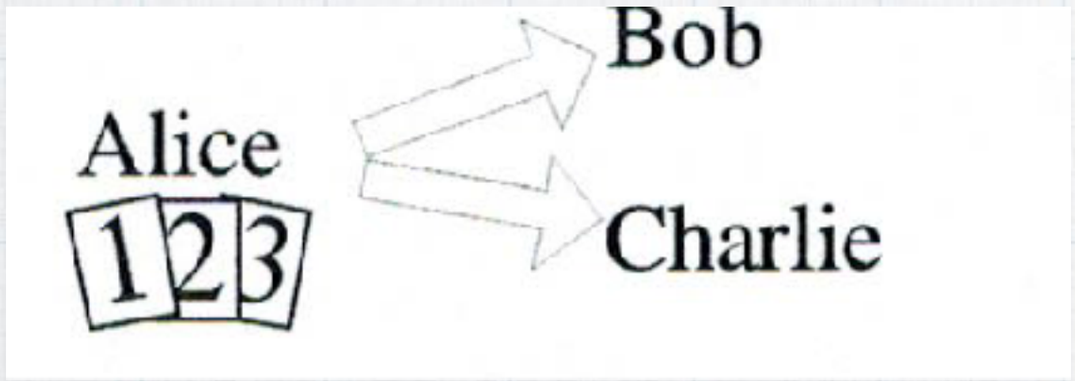
$\overset{=3}{\bullet}$ 「Semantic secrecy」 を保証 ($\overset{=3}{P_{Z^n|\mathcal{C}_n}}$ が $\overset{=3}{M}$ に寄らない)

「indistinguishability」 (data processing lemma を使う)

VIII. Coordination

Cover-Permuter
2007

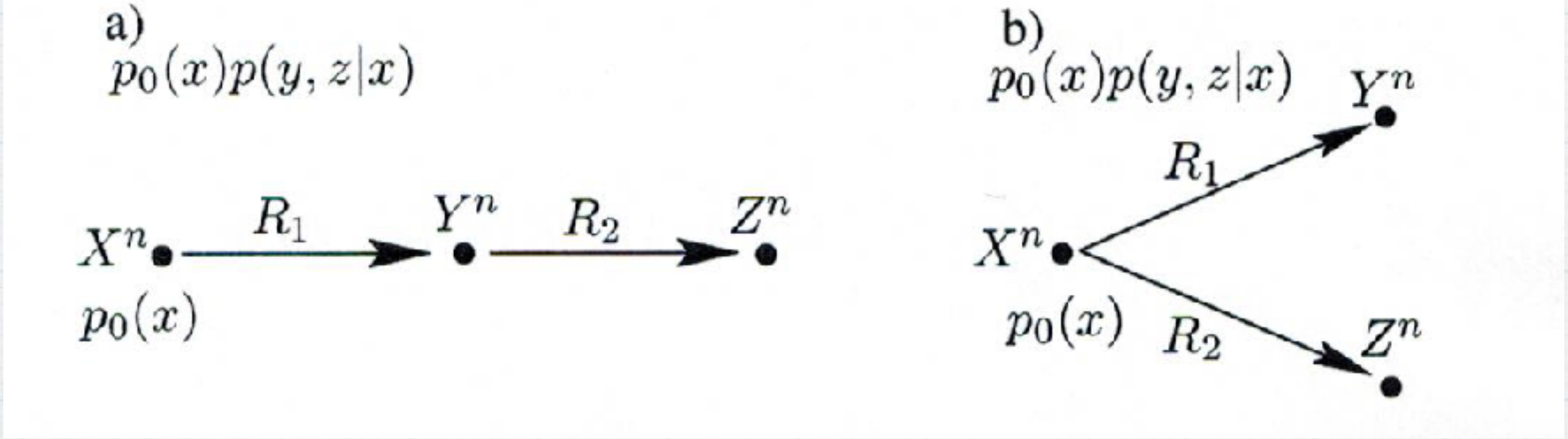




T. Cover and H. Permuter, “Capacity of Coordinated Actions,”
Proc. 2007 IEEE Int. Symp. Information Theory, pp. 2701-2705, 2007.

■ joint distribution を empirical joint distribution で近似

$p(x,y,z)$ $\tilde{p}(x,y,z)$



Theorem 1*

$$R_1 > I(X;YZ),$$
$$R_2 > I(X;Z)$$

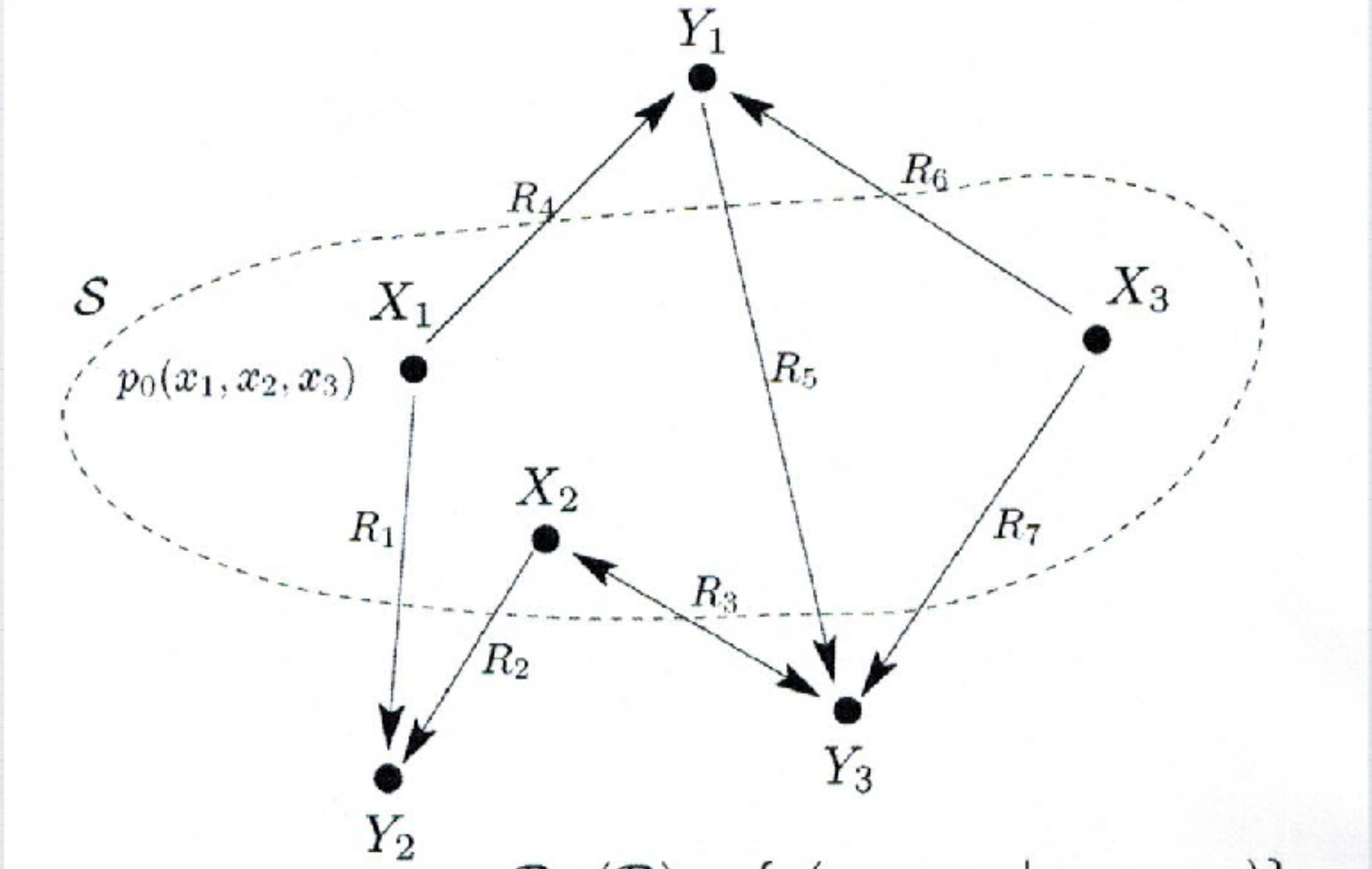
Theorem 2*

$$R_1 > I(X;UY),$$
$$R_2 > I(X;UZ)$$

where

$$X \to UX \to Z$$

T. Cover and H. Permuter, "Capacity of Coordinated Actions," Proc. 2007 IEEE Int. Symp. Information Theory, pp. 2701-2705, 2007.



P. Cuff, H. Permuter and T. Cover, "Coordination Capacity," IEEE Trans. Information Theory, vol. 56, no. 9, pp. 4181-4206, 2010.

２５歳になった情報スペクトル理論
25 Years of Information Spectrum Theory

韓太舜
Te Sun Han
情報通信研究機構
National Institute of Information and Communications Technology

今年（２０１８年）は，Claude E. Shannon が１９４８年に革命的論文 “A Mathematical Theory of Communication” (*Bell Syst. Tech. Journal*, vol.27, pp.379 - 423, pp.623 - 656, 1948) を発表した時から数えて丁度７０年目に当たり，しかも２年前の２０１６年には Shannon 生誕１００周年（Centennial）を迎えていたこともあって，世界の情報理論 コミュニティはこれら２つのビッグ・エポックを祝って数々の祝賀行事 (Workshops) で沸き立っていた．「情報理論」は今や止まることを知らない巨大な潮流の中で隆盛を極めつつあるかのようにも見える．しかし，その反面，これはもしかしたら「バブル」ではないのかと思うと「不安」に身の引き締まる思いもする．筆者も，長年にわたり情報理論の研究に携わって来た者として，その様変わりに深い感慨を禁じ得ないが，その間の個人に関わることで最も心に残る事柄を挙げよと言われれば，それはやはり「情報スペクトル理論誕生」にまつわる苦労話しである．本稿では，これについて少し振り返って見よう．

筆者が，本稿のタイトルになっている「情報スペクトル理論」の嚆矢となった Verdú 教授との共著論文 “Approximation Theory of Output Statistics” (*IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 752 -772, 1993) を出版したのは，今から丁度２５年前の１９９３年のことであった．この論文は，情報源符号化（source coding），通信路符号化（channel coding），同定符号化（identification coding），乱数生成 (resolvability) などの基本的諸問題に対する「情報スペクトル的アプローチ」の雛形を提示して見せたものであった．

それから５年後の１９９８年には，これらの成果を拡張・発展させると同時に，さらに仮説検定（hypothesis testing），レート・歪み理論（rate-distortion theory), 多端子情報理論（multi-user information theory）などの諸問題に対しても「情報スペクトル的アプローチ」を発展させ展開したものをまとめて，単行本「情報理論における情報スペクトル的方法」（培風館，1998)[*1]として出版することが出来た．これは未だ十分開拓されたとは言えない新分野を，曲がりなりにも初めて体系的に扱った試みであった．[*2]しかし，まだ評価が未確定の研究成果をこのような形で速やかに出版できたのには，幾つもの「僥倖」が関与していた．先の見込みも成算も何もないまま，まるで何かに取り憑かれたかのように，ただひたすらに不眠不休，一心不乱にメモ書きと TeX うちを繰り返していた頃，電通大に所用のあった培風館編集部の木村博信氏がたまたま私の研究室にも立ち寄られた．同氏は異様な（？）小生の姿を見て「いま何をしているのか」と問うので，「情報スペクトルの研究です」と答えて説明すると，同氏は「それが完成したらウチでいただきます」と即決．原稿も見ずに（もっとも，その頃まとまった原稿というものはなかったが）こんなことを決めていいのか，と多少の疑問が頭をよぎったが，嫌も応もない，とにかくこちらにとっては渡りに船，その日は喜びに満ちた最高の日となった．これが契機になって TeX 打ちがさらに加速し，最終原稿のページ数は予定より２００ページほど増えた．木村博信氏が「たまたま」電通大にやって来て，「たまたま」こちらに立ち寄って，「たまたま」TeX 打ちの場面に遭遇するというような幾つもの偶然が重ならなかったならば，「情報スペクトル理論」が纏まった形で世に出ることは無かったのである．それはまさに運命的な出会いであった．後日，出版された「情報理論における情報スペクトル的方法」を共立出版のある編集者に見せたところ，「このような “標準的テキストとして売れる見込みが立たない本” はウチではやり

[*1] 培風館からの推薦により，幸運にも，この年の大川出版賞を受賞した．
[*2] 諸々の情報を「１点スペクトル」としてではなく，広がりを持った「情報スペクトル」という視点から捉えようというという考え方の萌芽は，Shannon [1], Winkelbauer [2] などに見られる（例えば，“quantile” など）．

ません」とのことであった．これは何も共立出版だけが特別だという訳ではなく，他の出版社に持ち込んでも同様の結果になったに違いない．その意味で，木村博信氏の存在は，「情報理論における情報スペクトル的方法」日本語版が世に産声を上げるために，確かに決定的な役割を果たしたのである．“特異な研究”が “特異な編集者”に出会った結果であった．同氏に深謝する所以である．

さて，本書が出版されてみると，そのことが SITA[*3] を中心とした情報理論関係の若手研究者達の間に口コミで伝わり，ある程度の部数は売れたのでホッとしたのを覚えている．数学系の研究者には「大変わかりやすい」と評判が良かったが，工学系の研究者には「もう少し具体的なイメージが欲しい」との注文を戴いた．本書の執筆にあたっては，出来る限り「論理の連鎖」に全てを語らせる，という方針を貫いたので，後者の注文は尤もなことであった．一方，さすがに，本書の「モノグラフ」としての性格上，大学の講義でテキストとして使われることは無かったが，それはそれで本望であった．とにかく，「情報スペクトル理論」の存在自体をこの世に知らしめたい，という強い思いが全てに優先していたからである．

そうこうするうちに，海外の Alajaji 教授 や Chen 教授から，「お前は情報スペクトルとかいうものに関する本を出版したらしいが，ここには “培風館”が発行する日本語の本を注文するための代理店がない．そちらで何とかしてこちらに届くようにしてくれないか」というメールが来た．驚いて「日本語が読めるのか」と聞くと，「数式をたどっていけば基本的なことは分かる．分からないときは，日本からの留学生が助けてくれる」という具合であった．そこで，小生は早速，培風館から小生の立て替えで 2 部購入しそれを両教授に送った．その後本代の清算をどうしたか今や記憶が定かでないが，何れにしても，その当時の筆者にとっては，大変 pleasing で encouraging なことであった．

また，この頃，本書を “in Japanese” として引用する英語論文がいくつか国際専門誌に現れ始めた．このことも同様に pleasing で encouraging なことではあった．しかし，この事は同時に，これらの英語論文を読んだ研究者が情報スペクトル理論に関連した論文を書くときには，原典の本書ではなくそれを引用した英語論文を引用して済ませざるを得ないであろうことを意味していた．何故なら，大抵の場合，それらの研究者には日本語が読めないであろうからである．要するに，「情報理論における情報スペクトル的方法」は日本国内では確かに「存在する」が，国際的には未だ「存在していない」に等しいという当たり前だが厳然たる事実に衝撃を受けたのである．これは，日本語で研究活動をする者にとって共通する宿弊でもあった．

しかし，本書の存在を国際的に認知させるには 2 つの関門があった．1 つ目は，筆者が本書執筆のために持てる力の全てを費やしてしまったために，結局「燃え尽き症候群 (burned out)」状態に陥入り（５００ページ超の）一つの本を丸々「英訳」するなどの力は到底残っていなかったことであり，2 つ目は，海外の出版社に全くコネがないことであった．

1 つ目については，まず「英語の native で日本語にも堪能な若手の情報理論研究者」という “理想的” 条件で翻訳者の候補を色々探してみたが，全く思い浮かばない．しかも，できれば，翻訳作業は開始したら「1 年で終了」という条件も満たして欲しいとの思いがあったので，尚更であった．今から考えたら，こんな無茶な条件を全て満たす研究者などはこの世に存在するはずがなかった．そこで，条件を「日本語の native で英語にも堪能な若手の情報理論研究者」というものに変えてみたが，これもなかなか思うように行かない．そうこうするうちに，１９９９年夏，筆者はギリシャの Metsovo で開催された IEEE Information Theory Workshop に参加する機会を得たが，そこに，なんと，筑波大の古賀弘樹教授（当時，講師）がいたのである．思わぬ再会を喜び合う間にも，一瞬「この人だ！」という強い思いに打たれて筆者の心が決まった．これも運命的出会いであった．後は，同氏をどう説得するかであったが，会議の終了後にに立ち寄った Athens のとある寿司屋で「旅情あふれる雰囲気の中」で情報理論的会話を楽しみながら，説得に成功した．「旅情」と言うものは，人の心にロマンを響かせてくれるのである．同氏はこのあと 1 年後に翻訳作業に取り掛かり，それをほぼ 1 年で見事に完成させたのである．同氏に深謝する所以である．それにしても，このとき食べた Athens の寿司がなんとおいしかったことか．それは日本で食べたどの寿司よりも美味

[*3] 情報理論とその応用学会 (Society of Information Theory and its Applications)

だったのである．

２つ目についても，出版社探しは難航した．最初は，闇雲に２，３の出版社に当たって見たがうまく行く筈もなかった．困って，当時親交のあったドイツの Vinck 教授に相談してみると，オランダの Shannon Foundation という財団を紹介してくれた．この財団は，いわゆる大手出版社ではないが，出版事業も手がけているという．これで大いにホッとしたが，課題も残った．この財団から本を出版するにしても，いわゆる大手出版社ではないので，研究者がそれを実際に購入するには，まず Vinck 教授の研究室に注文を出し，同教授が研究者と財団とを仲介するというシステムによることが判明したからである　要するに，この財団は「販売網」というものを持っていなかったのである．これでは，「仏作って魂入れず」ではないか．やはり，大手の海外出版社を探さなければならないのかと，気を取直してあちこち当たって見てもやはりうまく行かない．

困り果てて，理化研の甘利俊一先生に経緯の全てを話して advise を求めたところ，なんと，即座に Springer Verlag の応用数学部門の女編集長を紹介してくれたではないか．この女編集長は，たまたま，甘利先生に脳科学関係の本の執筆を依頼中で，その関係で「情報理論における情報スペクトル的方法」英語版の出版を快く受け入れてくれたのである．これで，英語版出版に関わる全ての関門がクリアされたではないか．長い間追い求められていたジグソーパズルの最後のピースがついにはめ込まれた瞬間であった．天にも昇る心地であった．ここで最後の「たまたま」をもたらしてくれた甘利先生に深謝する所以である．

その後しばらくすると，女編集長から英語原稿の見本を送れという連絡が来たので，情報源符号化，乱数生成，仮説検定に関する部分の英訳原稿を送って見たところ，２ヶ月ほどしてから，詳細なコメントが送られて来た．それは，Springer Verlag によって選ばれた情報理論関係の研究者３人による「査読結果 (peer review)」であった．正直に言うと，これには多少の驚きを禁じ得なかった．何故なら，筆者の知る限り，日本では，理科系の書物を出版するとき，執筆者と出版社との間で合意さえすれば，「原稿を査読する」などと言う慣習は全くなかったからである．

しかし，よく考えて見ると，このような仕組みは，執筆者と出版社の両方にとってなかなか有益なものであった．執筆者にとっては，不注意な思い込みや致命的な誤りを正すことができるばかりでなく，原稿の全体的な「質的向上」のための貴重なコメントが得られるからである．実際，上記の「査読結果」は，おおむね好意的ではあったが批判的提言も含んでいたので，これらは原稿の最終版の改善 に大いに役立てられた．一方，出版社にとっては，新書の出版過程で，内容の学問的レベルを確保すると同時に営業リスクの低減も計ることができる．

以上のように，幾つもの「たまたま」に導かれて，最終的には，２００３年に Springer Verlag から，英訳版 *Information Spectrum Methods in Information Theory* が出版されるに至った．すなわち，情報スペクトル理論」は，ここに至ってついに「存在する」ことが国際的にも認知されることになったのである．喜びこれに勝るものはなかった．

一方，よく考えて見ると，本稿で「たまたま」と言って来たものは要するに「偶然」である．これほど不確かなものはない．しかし，この「偶然」を最初から最後まで順に並べてみると，そこには何か「必然」なるものが背景に作用しているようにも感じられる．そして，「偶然」を「必然」たらしめたものは，やはり「ひたすらに真理を追い求めて努力し続ける行為」ではなかろうかと思える．筆者の信条を表すものとして，「永遠に真理なるもの，我を惹きて行かしむ」（ファウスト）と言うのがあるが，以上のようなスペクトル本の出版過程を振り返って見ると，ファウストの言葉は確かにそうであるかも知れないと言う感懐を強く抱かせてくれる．筆者の研究生活もすでに５０年を超えて研究能力にも様々な困難を覚えるようになっているが，これからもあわよくば，ファウストの言葉で道を照らしながらさらなる歩みを進めて行きたいと思うのである．

文献

[1] C. E. Shannon, “Certain results in coding theory for noisy channels,” *Information and Control*, vol.1, pp.6-25, 1957

[2] K. Winkelbauer, “On the coding theorem for decomposable discrete information channels I,” *Kybernetika*, vol.7, no.2, pp.109-123, 1971