Musing upon Information Theory

ISIT Austin, June 13-18, 2010 Te Sun HAN

I. Preliminaries---Information-Spectra

Example 0

source 1: $X = \begin{pmatrix} a & b & c \\ 3/8 & 1/2 & 1/8 \end{pmatrix}$ symbols / probabilities $P_X(\cdot)$

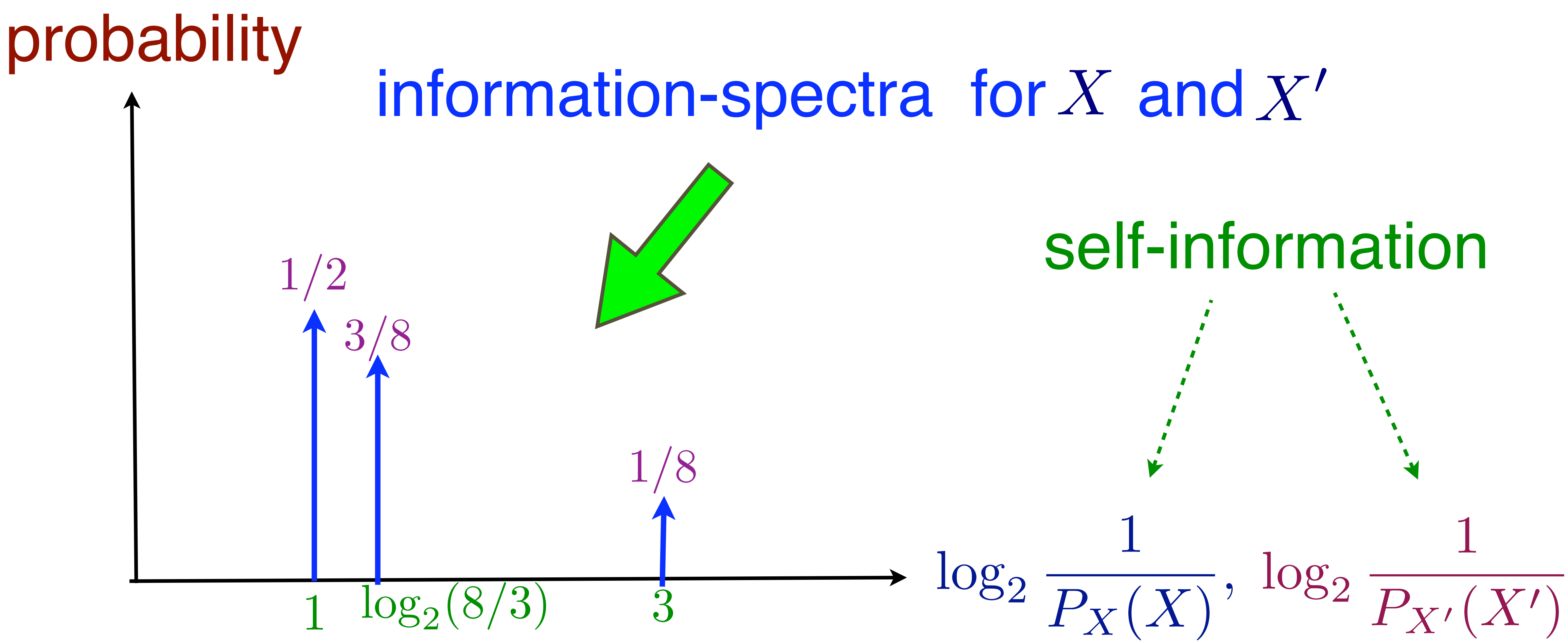
source 2: $X' = \begin{pmatrix} a' & b' & c' \\ 1/2 & 1/8 & 3/8 \end{pmatrix}$ symbols / probabilities $P_{X'}(\cdot)$

Two standpoints:

- $X \neq X'$ (under renaming and permutation)
- $X \equiv X'$

⇒ concept of information- spectra !

What is the information-spectrum?



- One information-spectrum for one source
- It is easy to check that the information-spectra are the same for sources X and X'

i.i.d. sequence	probabilities
0000	$(\frac{1}{3})^4$
0001	$(\frac{1}{3})^3\frac{2}{3}$
0010	$(\frac{1}{3})^3\frac{2}{3}$
0011	$(\frac{1}{3})^2(\frac{2}{3})^2$
0100	$(\frac{1}{3})^3\frac{2}{3}$
0101	$(\frac{1}{3})^2(\frac{2}{3})^2$
0110	$(\frac{1}{3})^2(\frac{2}{3})^2$
0111	$\frac{1}{3}(\frac{2}{3})^3$
1000	$(\frac{1}{3})^3\frac{2}{3}$
1001	$(\frac{1}{3})^2(\frac{2}{3})^2$
1010	$(\frac{1}{3})^2(\frac{2}{3})^2$
1011	$\frac{1}{3}(\frac{2}{3})^3$
1100	$(\frac{1}{3})^2(\frac{2}{3})^2$
1101	$\frac{1}{3}(\frac{2}{3})^3$
1110	$\frac{1}{3}(\frac{2}{3})^3$
1111	$(\frac{2}{3})^4$

$$p(0) = \frac{1}{3},$$
$$p(1) = \frac{2}{3}$$

renaming

$$X \Longrightarrow X'$$

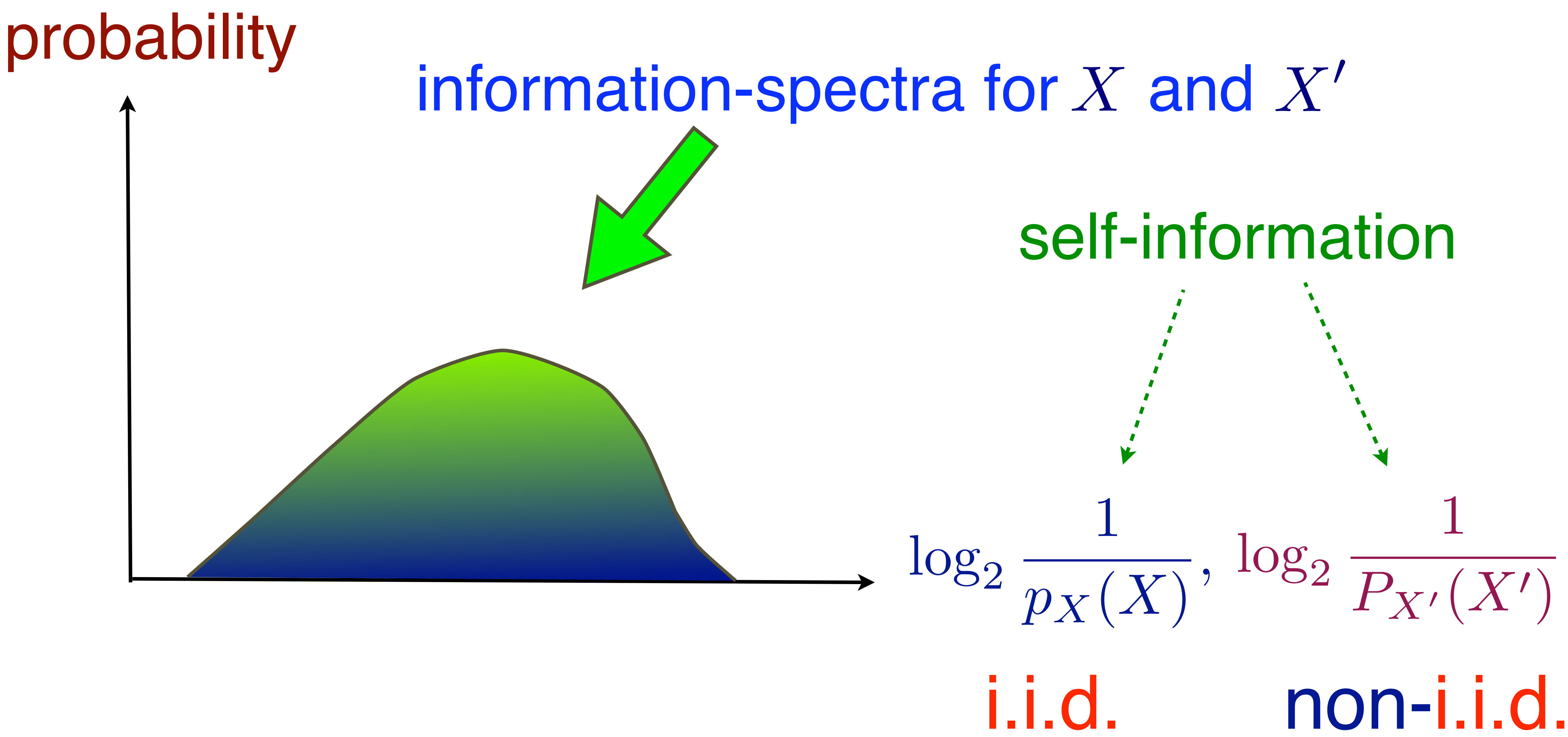
$$\Downarrow$$

What source is X' at all?

Far from standard!

non i.i.d.sequence	probabilities
1010	$(\frac{1}{3})^4$
1101	$(\frac{1}{3})^3\frac{2}{3}$
0100	$(\frac{1}{3})^3\frac{2}{3}$
0110	$(\frac{1}{3})^2(\frac{2}{3})^2$
1011	$(\frac{1}{3})^3\frac{2}{3}$
0000	$(\frac{1}{3})^2(\frac{2}{3})^2$
0001	$(\frac{1}{3})^2(\frac{2}{3})^2$
1000	$\frac{1}{3}(\frac{2}{3})^3$
1111	$(\frac{1}{3})^3\frac{2}{3}$
1001	$(\frac{1}{3})^2(\frac{2}{3})^2$
0111	$(\frac{1}{3})^2(\frac{2}{3})^2$
0010	$\frac{1}{3}(\frac{2}{3})^3$
1110	$(\frac{1}{3})^2(\frac{2}{3})^2$
1100	$\frac{1}{3}(\frac{2}{3})^3$
0101	$\frac{1}{3}(\frac{2}{3})^3$
0011	$(\frac{2}{3})^4$

Let us depict again the information-spectra for these sources X and X':

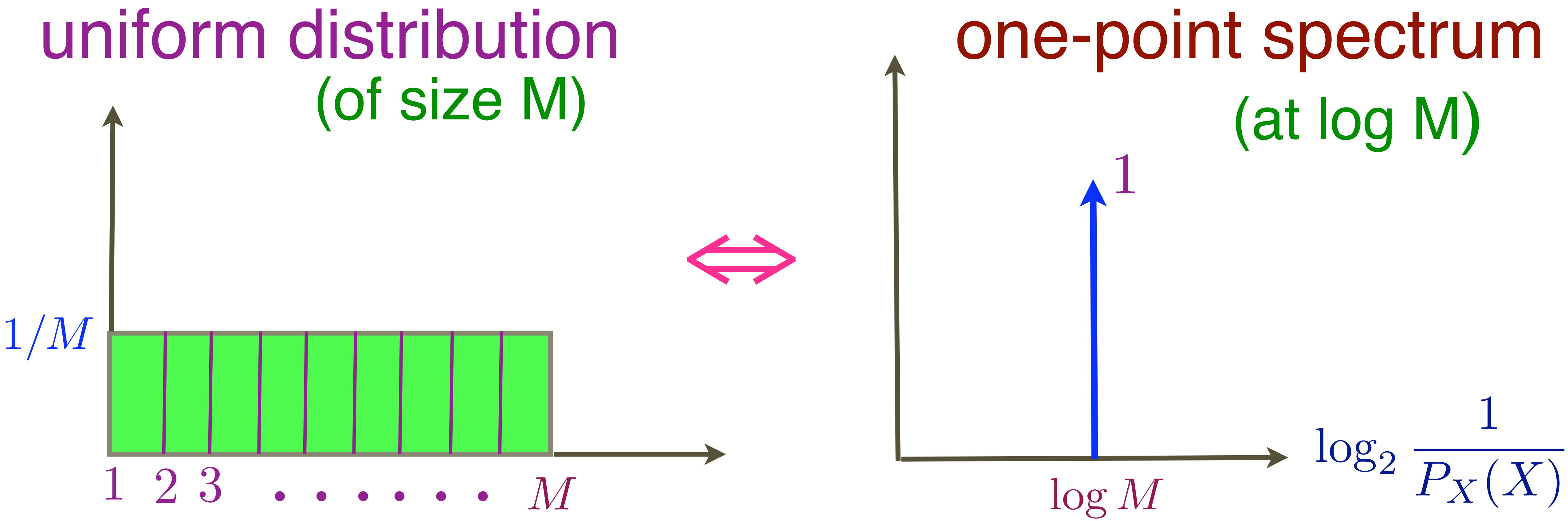


It is easy to check that both coincide again!

$\Downarrow$ Thus

- The information spectra remain the same under renaming and permutation !!

Remark 1(very import observation!):



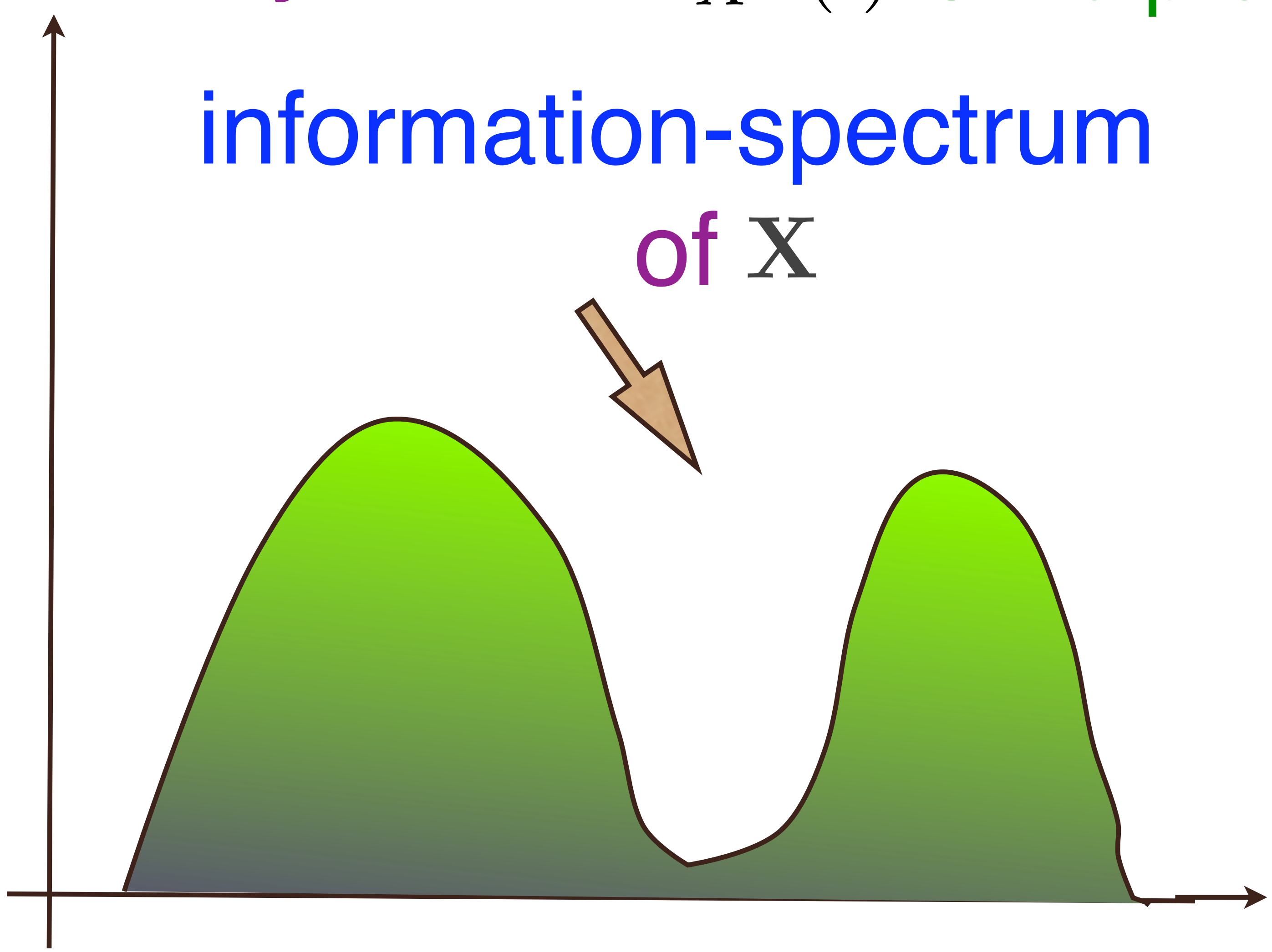
II. Asymptotic Information-Spectra

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

- alphabet $\mathcal{X}$ is countably infinite
- $X^n \in \mathcal{X}^n$ is a block source of length n
- $P_{X^n}(\cdot)$ is the probability dist. of X^n

probability

information-spectrum of $\mathbf{X}$



normalized self-information

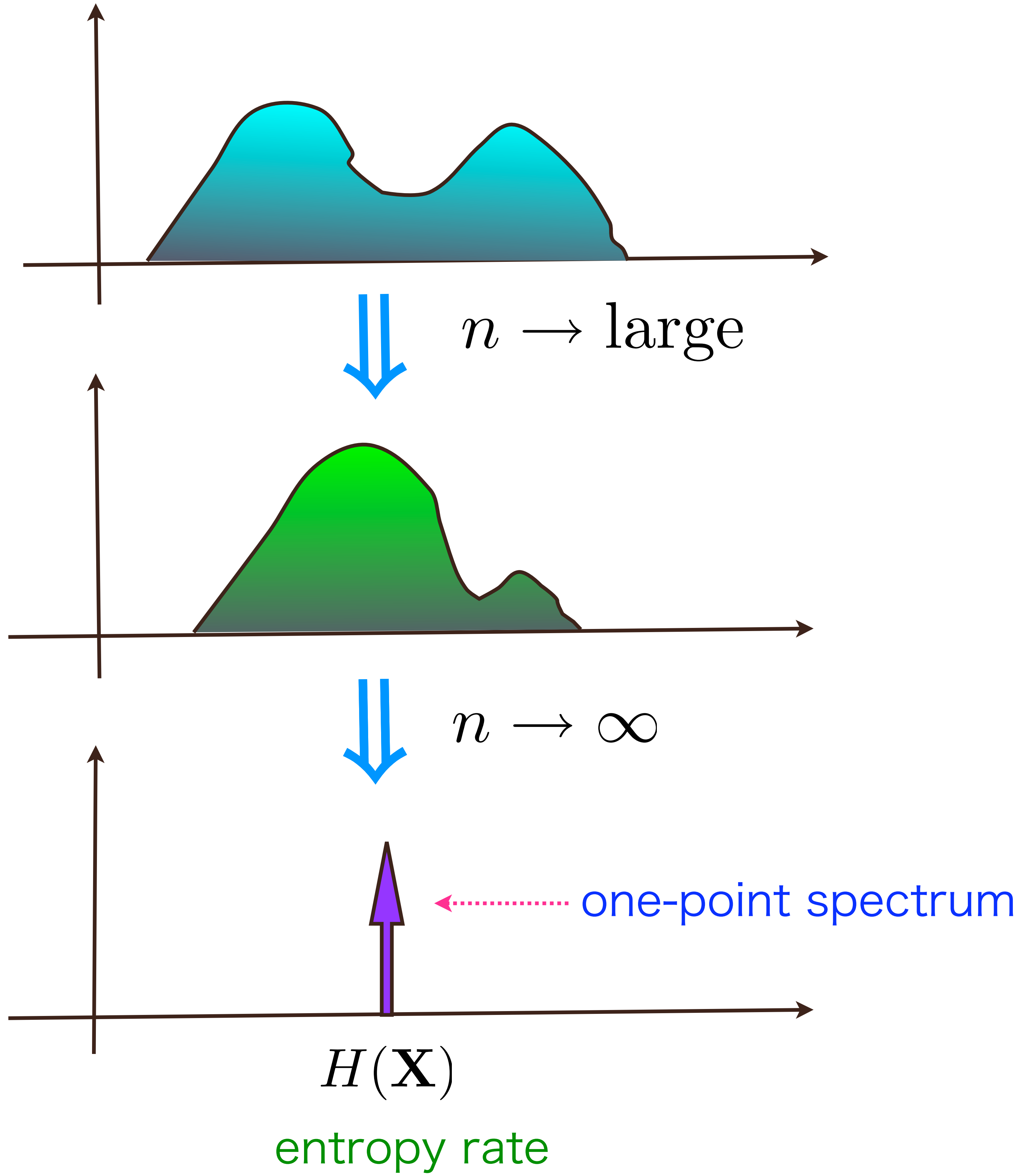
$$\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)}$$

Example 1

1) i.i.d. source
2) Markov source
3) ergordic source

$\Downarrow$

AEP
(Asymptotic Equi-Partition)



Example 2

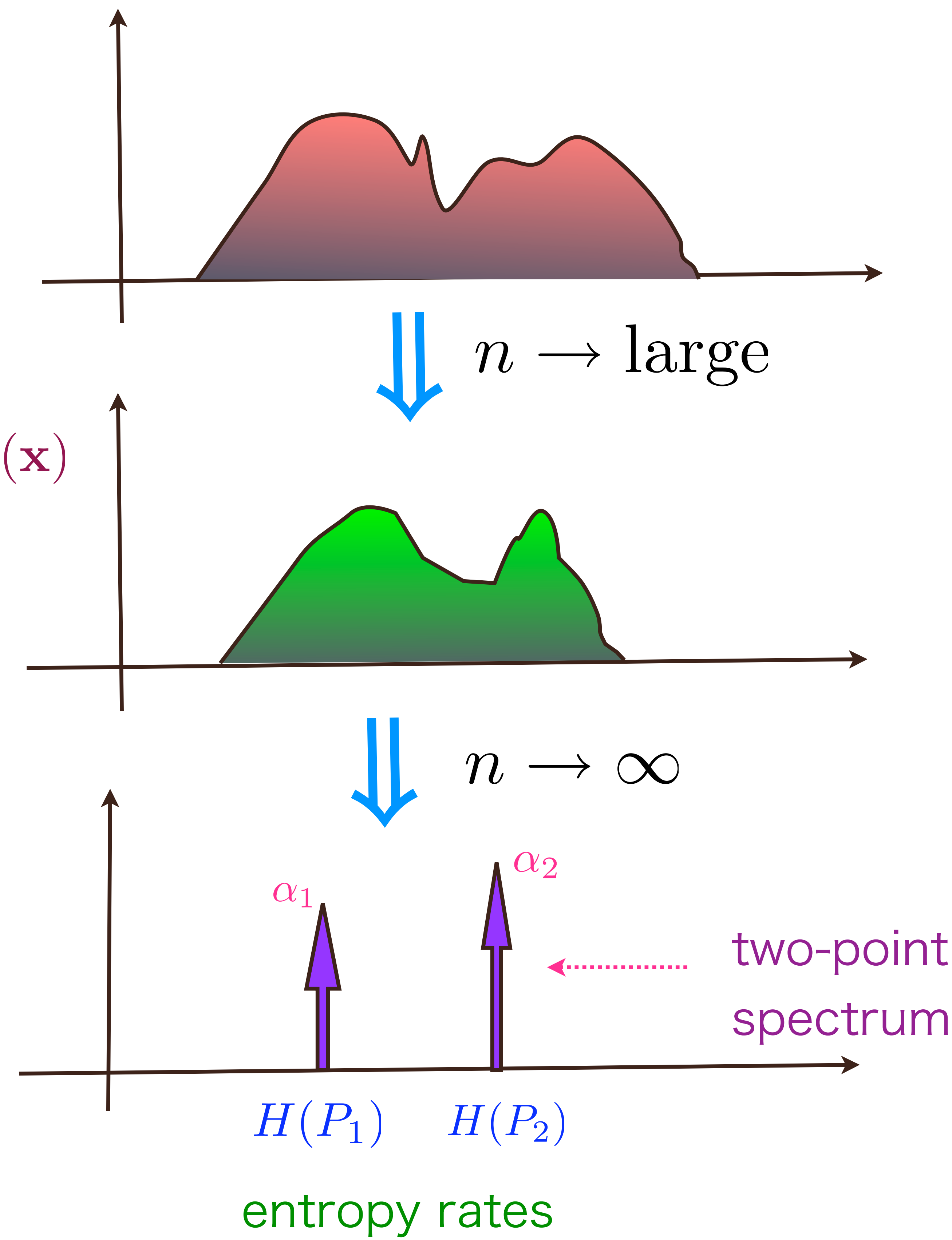
4)mixed source

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

X_1^n : i.i.d. $\sim P_1$

X_2^n : i.i.d. $\sim P_2$

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



non-AEP $\Leftarrow$

Quesiotn: How to define “AEP” for general sources?

↓ to this end

Information-spectrum approach is useful!

↓

Let me try to show it!

Fundamental tools for information-spectra

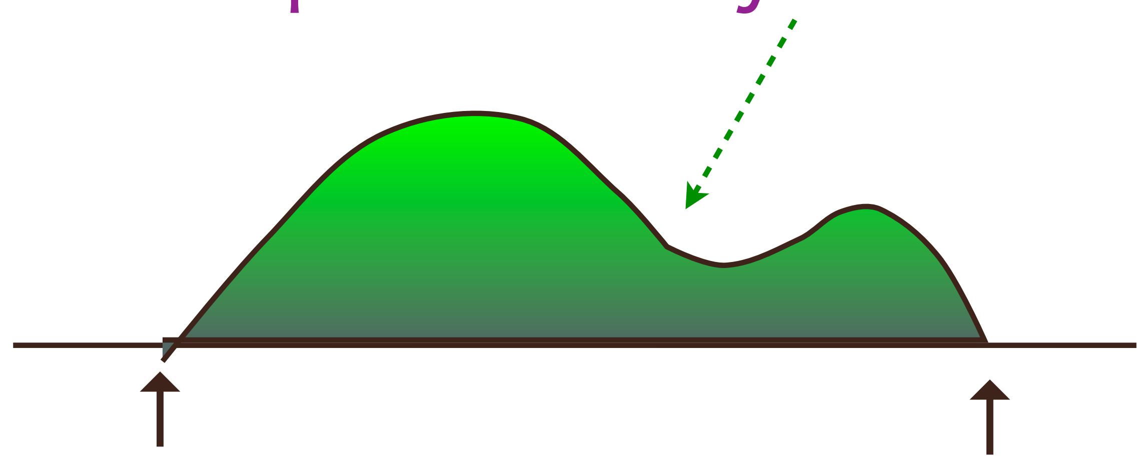
¶ lim-sup in probability

$$\text{p-}\limsup_{n\to\infty} Z_n \equiv \inf\left\{\alpha \mid \lim_{n\to\infty} \Pr\{Z_n > \alpha\} = 0\right\}$$

¶ lim-inf in probability

$$\text{p-}\liminf_{n\to\infty} Z_n \equiv \sup\left\{\beta \mid \lim_{n\to\infty} \Pr\{Z_n < \beta\} = 0\right\}$$

$\{Z_n\}_{n=1}^{\infty}$ is a sequence of real random variables



lim-sup and lim-inf in probabilities

$$\begin{aligned}
\text{p-}\limsup_{n\to\infty}(Z_n+V_n) &\le \text{p-}\limsup_{n\to\infty} Z_n + \text{p-}\limsup_{n\to\infty} V_n \\
\text{p-}\liminf_{n\to\infty}(Z_n+V_n) &\ge \text{p-}\liminf_{n\to\infty} Z_n + \text{p-}\liminf_{n\to\infty} V_n \\
\text{p-}\limsup_{n\to\infty}(Z_n+V_n) &\ge \text{p-}\limsup_{n\to\infty} Z_n + \text{p-}\liminf_{n\to\infty} V_n \\
\text{p-}\liminf_{n\to\infty}(Z_n+V_n) &\le \text{p-}\limsup_{n\to\infty} Z_n + \text{p-}\liminf_{n\to\infty} V_n \\
\text{p-}\liminf_{n\to\infty} Z_n &\le \text{p-}\limsup_{n\to\infty} Z_n \\
\text{p-}\limsup_{n\to\infty}(-Z_n) &= -\text{p-}\liminf_{n\to\infty} Z_n
\end{aligned}$$

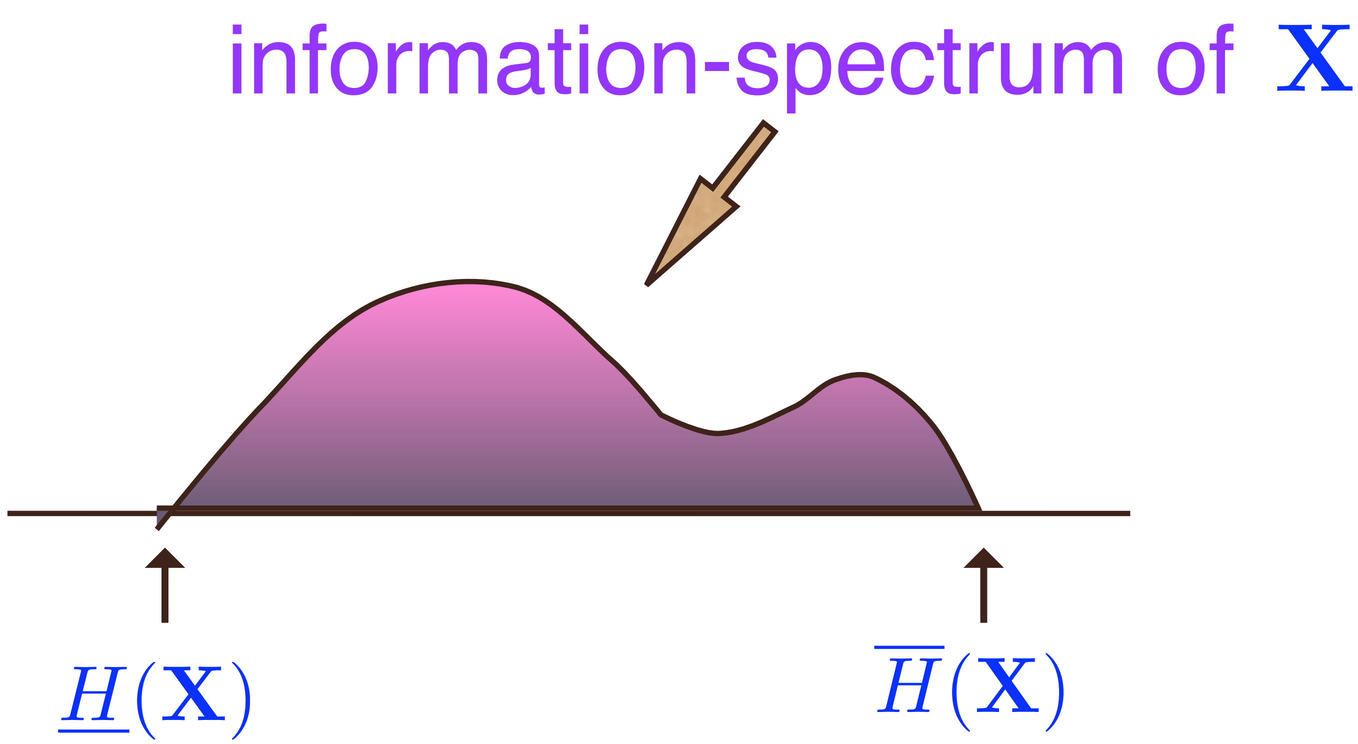
$$\text{p-}\limsup_{n\to\infty} Z_n = \text{p-}\liminf_{n\to\infty} Z_n = c \Longleftrightarrow \text{p-}\lim_{n\to\infty} Z_n = c$$

(limit in probability)

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

¶ spectral lim-sup: $\overline{H}(\mathbf{X}) = \text{p-}\limsup_{n\to\infty} \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)}$

¶ spectral lim-inf: $\underline{H}(\mathbf{X}) = \text{p-}\liminf_{n\to\infty} \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)}$



II. AEP, Strong Converse, and Information-Spectra

¶Definition1¶ (strong converse property)

A general source $\mathbf{X}$ is said to satisfy the strong converse property if, with any given rate R, the optimal coding always yields either $\lim_{n\to\infty} \varepsilon_n = 1$ or $\lim_{n\to\infty} \varepsilon_n = 0$ where ε_n is the decoding error probability.

¶Definition 2 ¶ (generalized AEP)

A general source $\mathbf{X}$ is said to satisfy AEP if there exists a set $T_n^\varepsilon \subset \mathcal{X}^n$ such that $P_{X^n}(T_n^\varepsilon) \to 1$ and $P_{X^n}(\mathbf{x}) \simeq \exp[-n(\underline{H}(\mathbf{X}) \pm \epsilon)]$ for $\forall \mathbf{x} \in T_n^\varepsilon$ ($\forall \epsilon > 0$)

¶Theorem 1¶

one-point spectrum

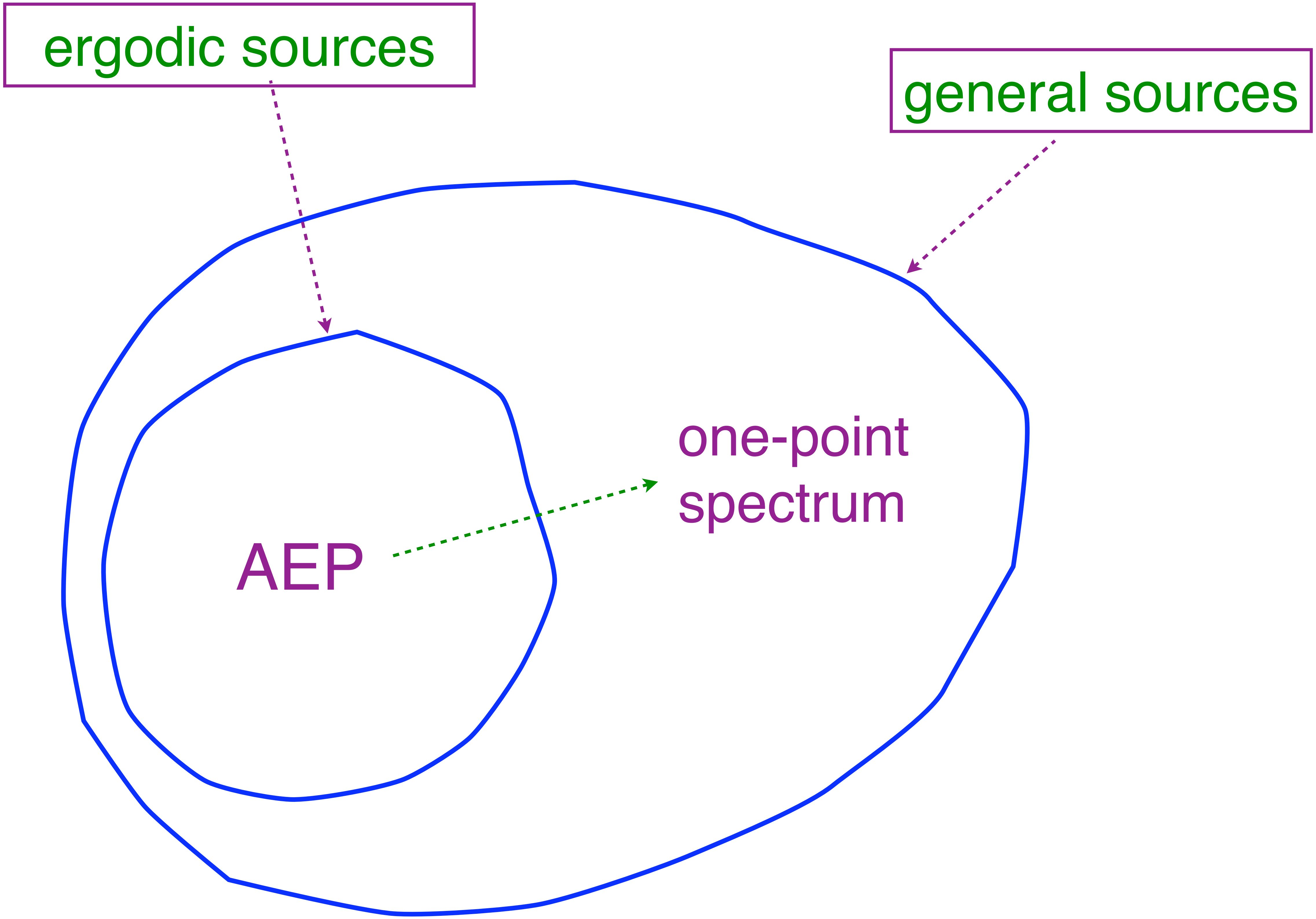
definition ⇕

$$\underline{H}(\mathbf{X}) = \overline{H}(\mathbf{X}) \Leftrightarrow$$ strong converse property of $\mathbf{X}$

theorem ⇕

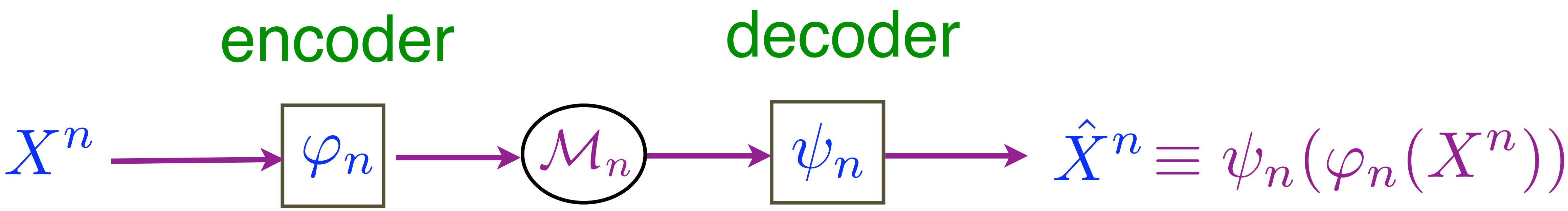
theorem

generalized AEP


ergodic sources
general sources
one-point
spectrum
AEP

III. Source Coding, and Information-Spectra

Fixed-length source coding



$$\mathcal{M}_n \equiv \{1, 2, \cdots, M_n\}$$

- coding rate : $\frac{1}{n} \log M_n$
- error probability : $\varepsilon_n \equiv \Pr\{\hat{X}^n \neq X^n\}$
- R is achievable $\Leftrightarrow$ $\limsup_{n\to\infty} \frac{1}{n} \log M_n \leq R, \quad \lim_{n\to\infty} \varepsilon_n = 0$

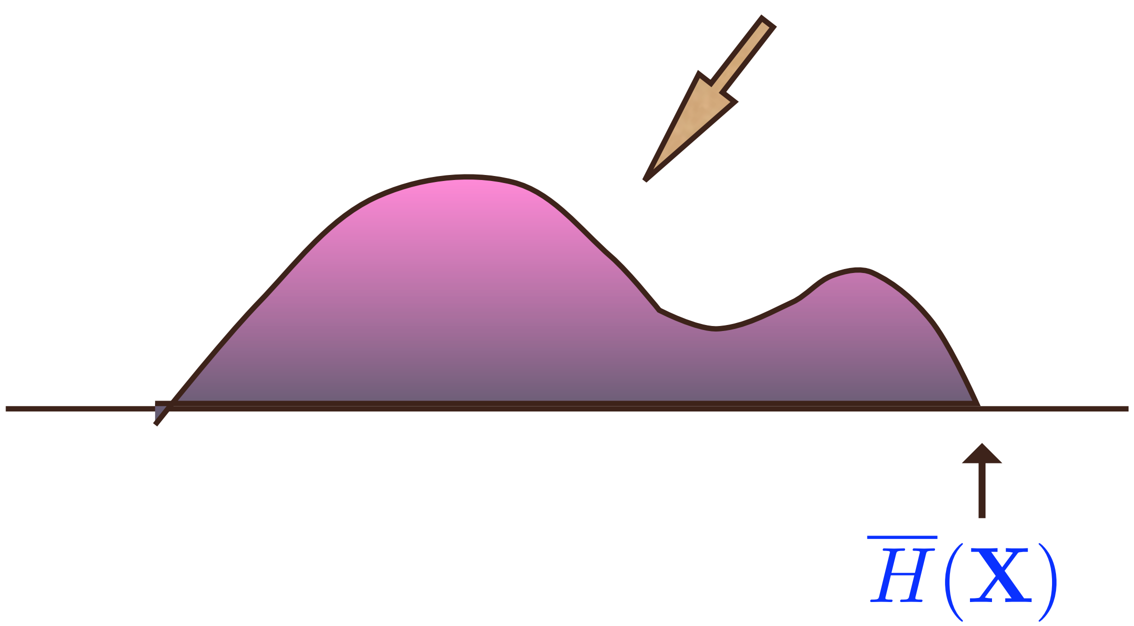
(we want here to make R as small as possible)

¶Definition 3 ¶ $R_f(\mathbf{X}) = \inf\{R | R \text{ is achievable}\}$

¶Thorem 2 ¶ (Han & Verdu,1993)

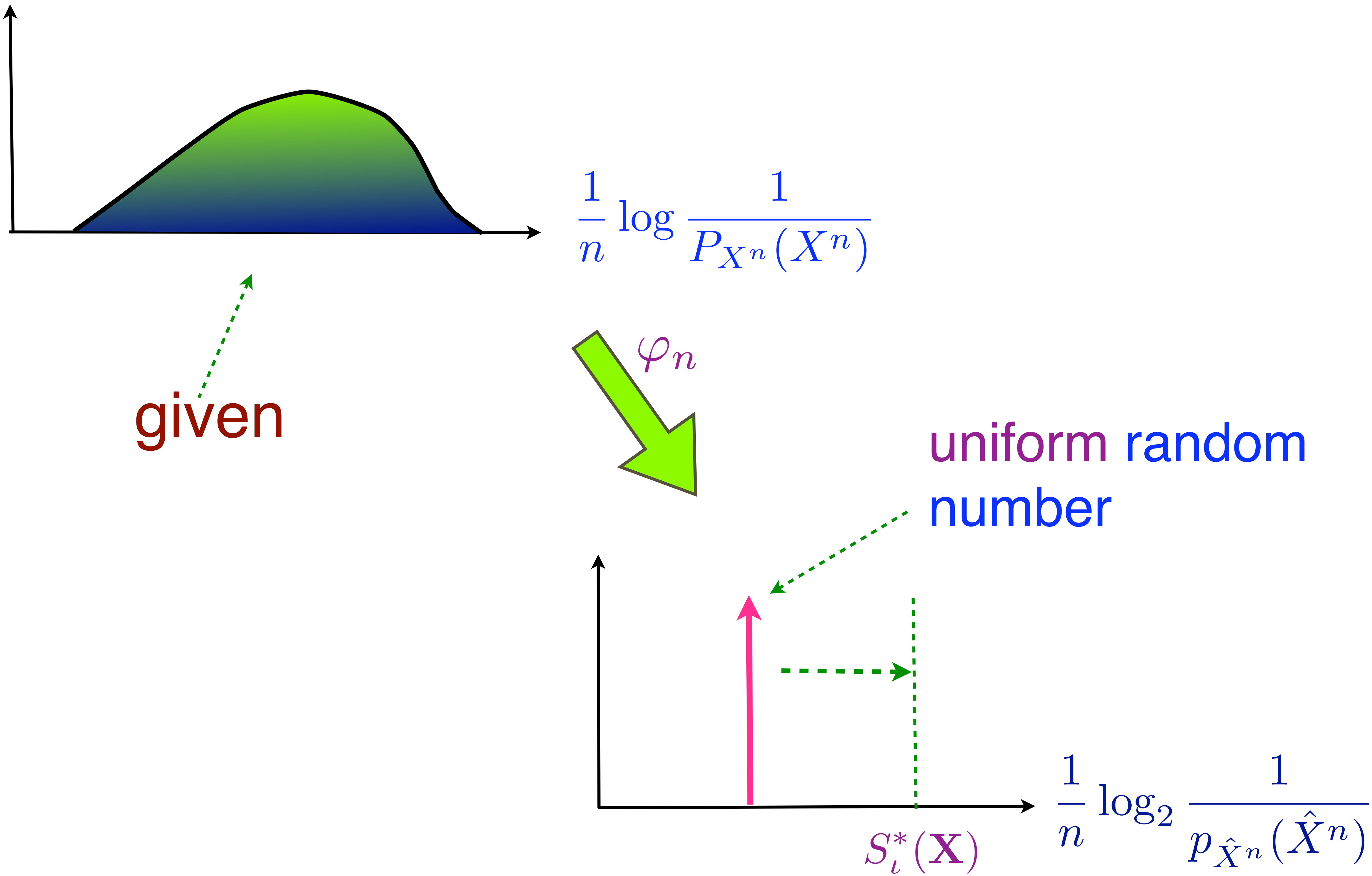
- Optimal coding rate : $R_f(\mathbf{X}) = \overline{H}(\mathbf{X})$

information-spectrum of $\mathbf{X}$



IV. Intrinsic Randomness and Information-Spectra

Uniform random number generation (intrinsic randomness)



Formal description (1): intrinsic randomness

encoder

$\hat{X}^n \equiv \varphi_n(X^n)$

$\mathcal{M}_n \equiv \{1, 2, \cdots, M_n\}$

- U_{M_n} : uniform (on $\mathcal{M}_n$) random variable
- divergence:

$$D(P||Q) = \sum_{i=1}^{M_n} p_i \log \frac{p_i}{q_i}$$

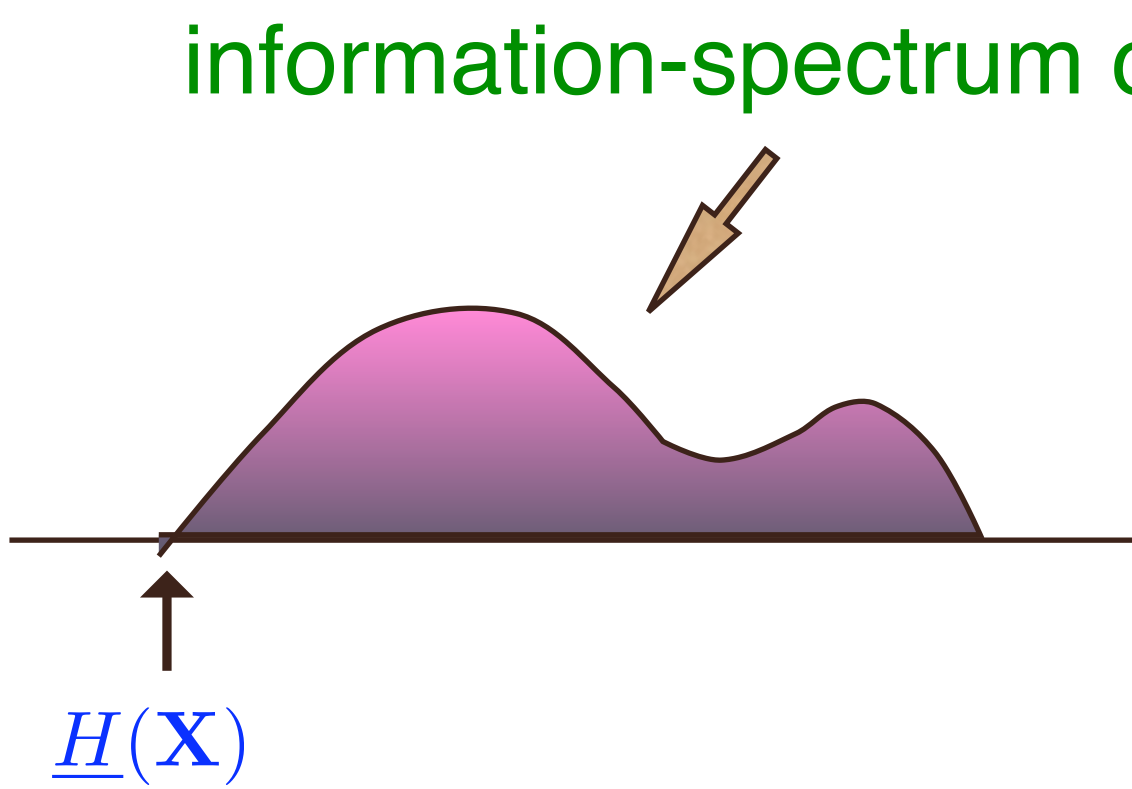
- random number generation rate : $\frac{1}{n} \log M_n$
- randomness distance : $\rho_n \equiv \frac{1}{n} D(\hat{X}^n || U_{M_n})$
- R is achievable $\Leftrightarrow$ $\liminf_{n\to\infty} \frac{1}{n} \log M_n \geq R, \quad \lim_{n\to\infty} \rho_n = 0$

(we want here to make R as large as possible)

¶Definition 4 ¶ $S_\iota^*(\mathbf{X}) = \sup\{R | R \text{ is achievable}\}$

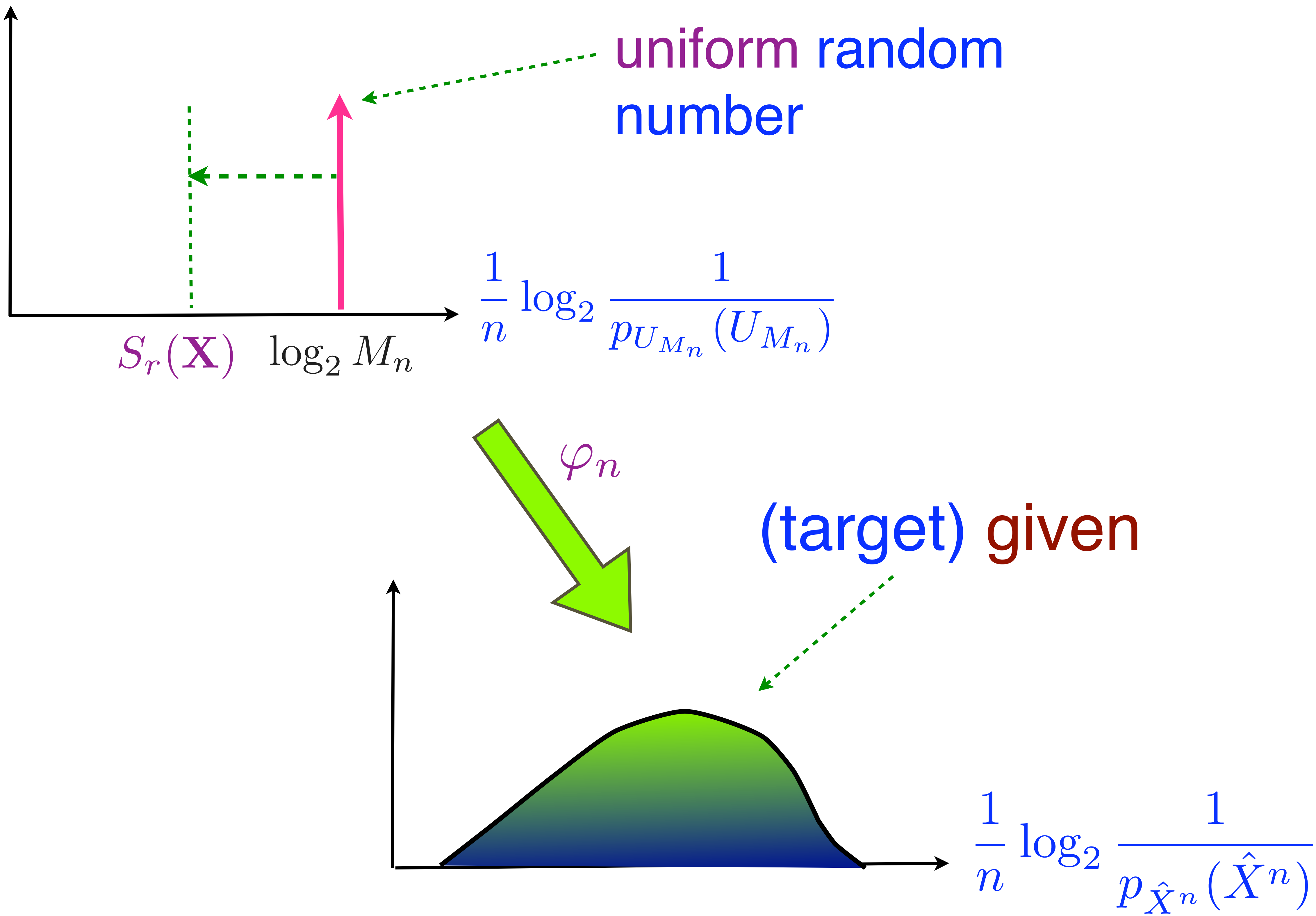
¶Theorem 3 ¶ (Vembu & Verdu,1995)

Intrinsic randomness : $S_\iota^*(\mathbf{X}) = \underline{H}(\mathbf{X})$



V. Resolvability and Information-Spectra

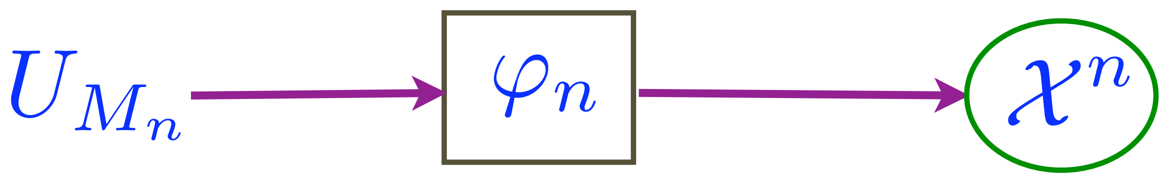
Arbitrary random number generation (resolvability)



2010年7月20日火曜日

Formal description (2): resolvability

encoder $\mathbf{X} \equiv \{X^n\}_{n=1}^{\infty}$ (target) is given



$\hat{X}^n \equiv \varphi_n(U_{M_n})$

◉ U_{M_n} : uniform (on $\mathcal{M}_n$) random variable

◉variational distance

$$d(P, Q) = \sum_{i=1}^{\infty} |p_i - q_i|$$

- random number coding

rate : $\frac{1}{n} \log M_n$

- randomness distance : $\sigma_n \equiv d(\hat{X}^n, X^n)$

- R is achievable $\Leftrightarrow$ $\limsup_{n\to\infty} \frac{1}{n} \log M_n \le R, \quad \lim_{n\to\infty} \sigma_n = 0$

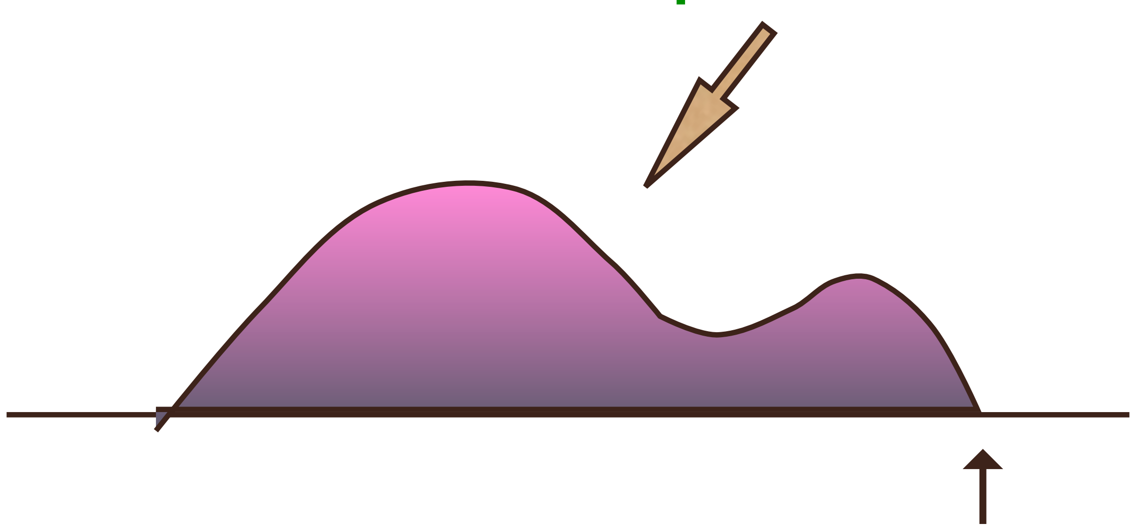
(we want here to make R as small as possible)

¶Definition 5 ¶ $S_r(\mathbf{X}) = \inf\{R | R \text{ is achievable}\}$

¶THeorem 4 ¶ (Han & Verdu,1993)

Resolvability： $S_r(\mathbf{X}) = \overline{H}(\mathbf{X})$

information-spectrum of $\mathbf{X}$



$\overline{H}(\mathbf{X})$

VI. Folklore, and Information-Spectra

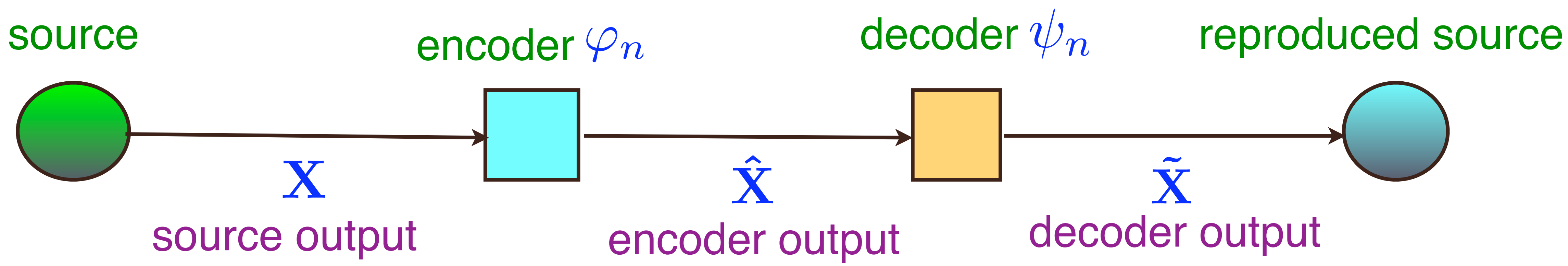
Let us consider the following "folklore:"

The encoder output $\varphi_n(X^n)$ working at the optimal rate for a source $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$ looks like almost completely random with $n \rightarrow \infty$.

- Question: How to formally define the "class of sources" and " notion of almost completely random."

- In order to solve this problem, we need Theorems 1, 2, 3 (previous), and Theorem 5 (next).

¶Three information sources in coding¶



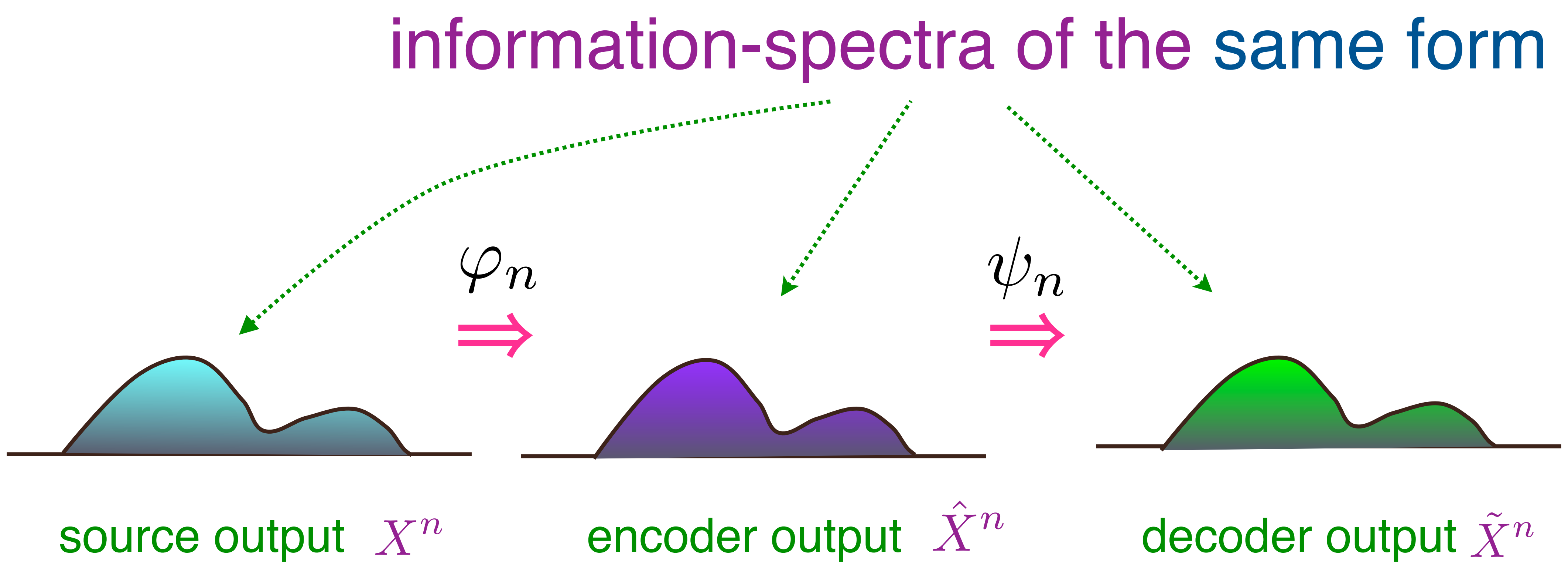
♯ $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$ (source 1)

♯ $\hat{\mathbf{X}} = \{\hat{X}^n\}_{n=1}^{\infty}$ $(\hat{X}^n = \varphi_n(X^n))$ (source 2)

♯ $\tilde{\mathbf{X}} = \{\tilde{X}^n\}_{n=1}^{\infty}$ $(\tilde{X}^n = \psi_n(\varphi_n(X^n)))$ (source 3)

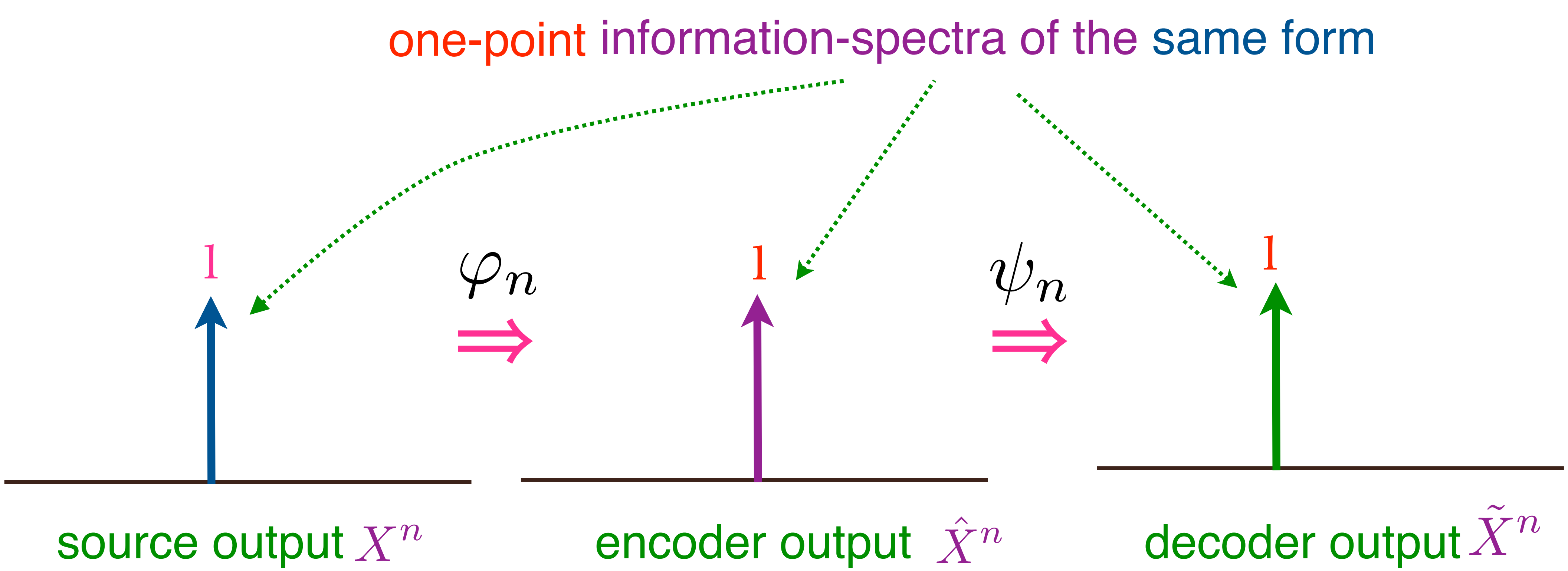
¶Theorem 5 ¶ (invariance of information-spectra)

In the process of source coding with decoding error probability $\varepsilon_n \to 0 \ (n \to \infty)$, three information-spectra of $\mathbf{X}, \hat{\mathbf{X}}, \tilde{\mathbf{X}}$ asymptotically have the same form.



¶Corollary 1 ¶(preservation of strong converse property)

In the process of source coding with decoding error probability $\varepsilon_n \to 0 \ (n \to \infty)$, the strong converse property (= AEP) is preserved at the same position.



¶Definition 6 ¶(Levy distance between prob. dist.s)

The notion “asymptotically the same form” in Theorem 5:

$$\Updownarrow$$ ← definition

$$\lim_{n\to\infty} L\left(\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)}, \frac{1}{n}\log\frac{1}{P_{\hat{X}^n}(\hat{X}^n)}\right) = 0$$

$$\lim_{n\to\infty} L\left(\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)}, \frac{1}{n}\log\frac{1}{P_{\tilde{X}^n}(\tilde{X}^n)}\right) = 0$$

◉ $L(U,V)$ is called Levy distance:= the infimum of all $\mu > 0$ such that, for all reals x,

$$\Pr\{U \le x-\mu\} - \mu \le \Pr\{V \le x\} \le \Pr\{U \le x+\mu\} + \mu$$

¶Definition 7 ¶ (the notion "almost completely random")

A general source $\mathbf{Z} = \{Z^n\}_{n=1}^{\infty}$ is said to be almost completely random with size M_n if it holds that

$$\lim_{n\to\infty} \frac{1}{n} D(Z^n || U_{M_n}) = 0,$$

where $D(\cdot||\cdot)$ is the divergence, and the limit:

$$\lim_{n\to\infty} \frac{1}{n} \log M_n$$

exists.

¶ Theorem 6 ¶ (Folklore: Han, 2005)

The encoder output $\varphi_n(X^n)$ working at the optimal rate for a "general" source $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$ is almost completely random if and only if $\mathbf{X}$ satisfies the strong converse property (=AEP).

◉Remark: A "general" source $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$ affords "any" nonstationary memory structure, which includes, as very special cases, all ergodic sources.

Hint for the proof:

1) part $\Leftarrow$: Use Theorem 1 ($\underline{H}(\mathbf{X}) = \overline{H}(\mathbf{X})$)

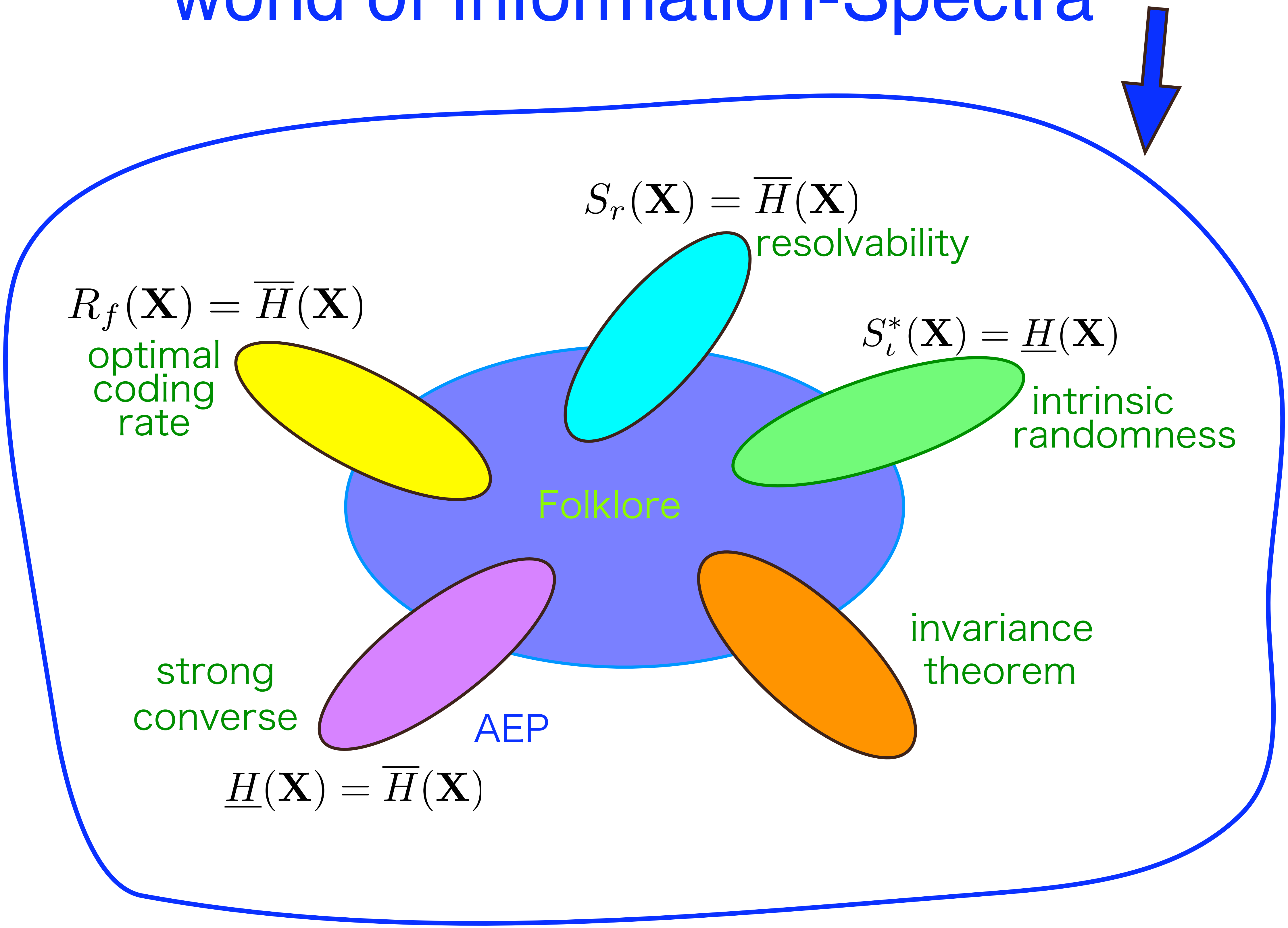
Theorem 2 ($R_f(\mathbf{X}) = \overline{H}(\mathbf{X})$)

Theorem 5 (invariance theorem)

2) part $\Rightarrow$: Use Theorem 2 ($R_f(\mathbf{X}) = \overline{H}(\mathbf{X})$)

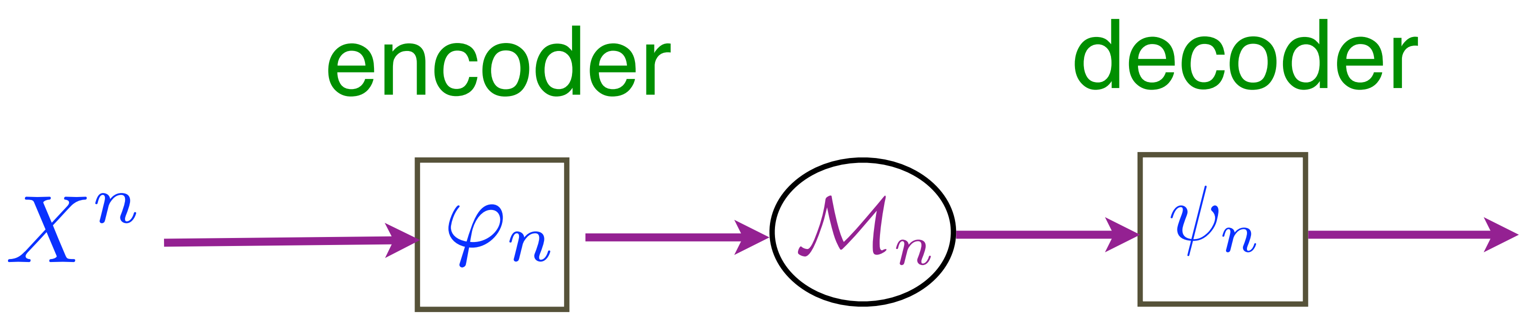
Theorem 3 ($S_\iota^*(\mathbf{X}) = \underline{H}(\mathbf{X})$)

Folklore looks like a navel of the world of Information-Spectra



VII. Reliability Function, and Information Spectra

Fixed-length source coding (cont.)



$\hat{X}^n \equiv \psi_n(\varphi_n(X^n))$

$$\mathcal{M}_n \equiv \{1, 2, \cdots, M_n\}$$

- coding rate : $\frac{1}{n} \log M_n$
- error probability : $\varepsilon_n \equiv \Pr\{\hat{X}^n \neq X^n\}$
- R is r-achievable $\Leftrightarrow$ $\limsup_{n\to\infty} \frac{1}{n} \log M_n \leq R, \lim_{n\to\infty} \frac{1}{n} \log \frac{1}{\varepsilon_n} \geq r$

exponentially decaying error probability

$$\varepsilon_n \lesssim e^{-nr}$$

(we want to make R as small as possible)

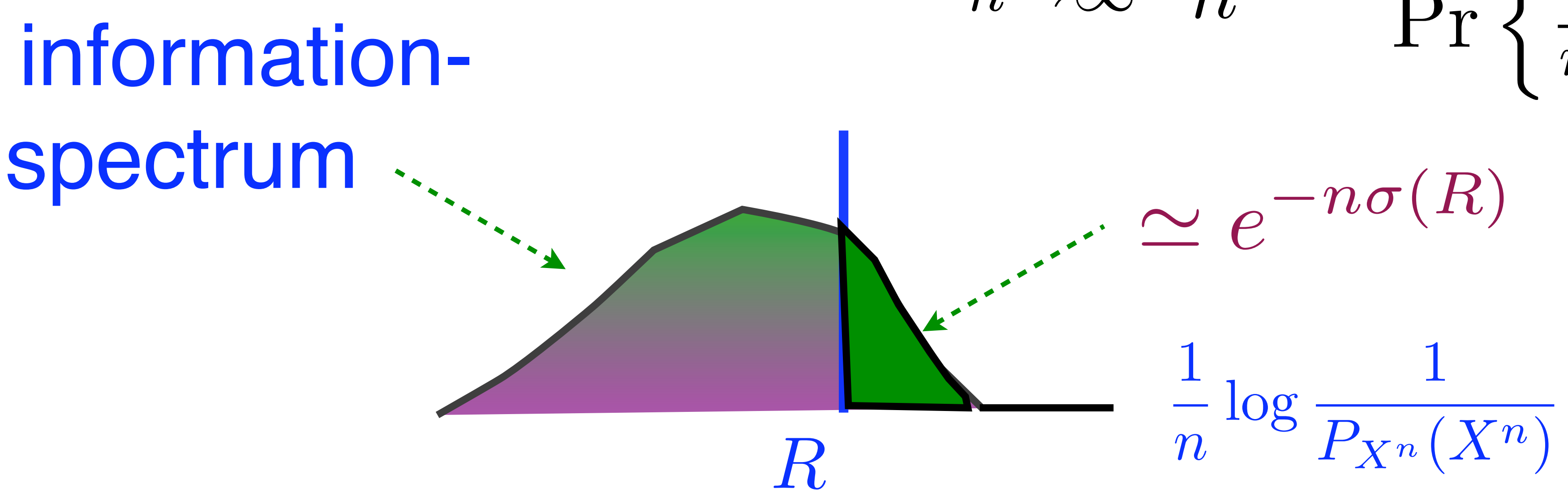
Reliability function for source coding

¶Definition 8 ¶ $R_e(r|\mathbf{X}) = \inf\{R | R \text{ is } r\text{-achievable}\}$

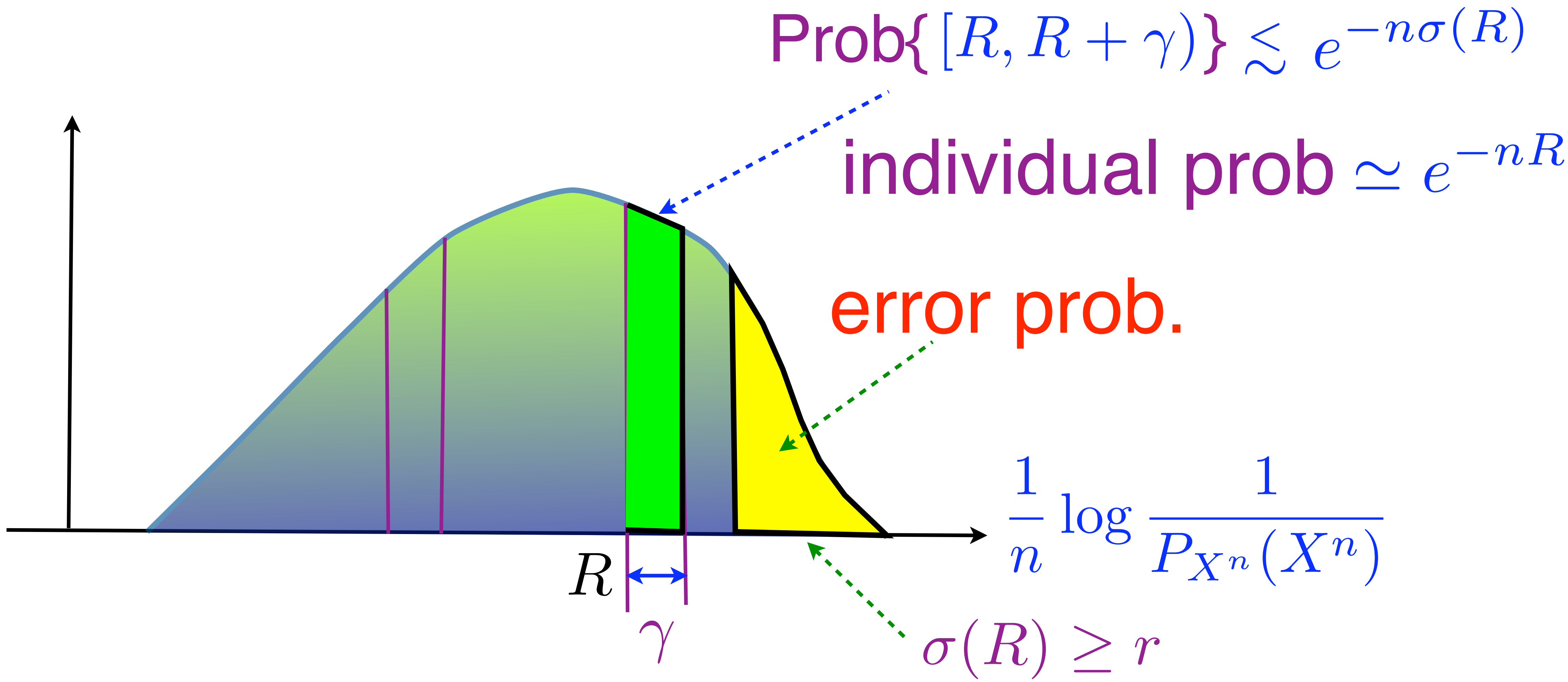
¶Theorem 7 ¶ (Reliability function: Han, 2000)

$$R_e(r|\mathbf{X}) = \sup_R \{R - \sigma(R) | \sigma(R) < r\}$$

where, $\sigma(R) = \liminf_{n\to\infty} \frac{1}{n} \log \frac{1}{\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq R\right\}}$



Hint for the proof (direct part):



- # of sequences in $[R, R+\gamma)$

  $\lesssim e^{-n\sigma(R)}/e^{-nR} \simeq e^{n(R-\sigma(R))}$ $(\sigma(R) < r)$

- #of total sequences $\lesssim \sup_{R:\sigma(R)<r} e^{n(R-\sigma(R))}$

Remark 2 (large deviation):

- The argument used here is called "information-spectrum slicing," which is a kind of typical large deviation techniques; the slicing would be intuitively understandable.

- The "informaton-spectrum slicing" argument is applicable also to the problems in random number generation, rate-distortion theory, and hypothesis testing, channel coding, etc.

Remark 3 :

- It is a tradition, given a rate R, to compute the optimal error probabilty ε

- Here, given an error probability ε, we compute the optimal rate R. This approach would be easier.

Example 3 (Longo & Sgarro, 1979)

- Let $\mathbf{X} = (X_1, X_2, \cdots)$ be an i.i.d. subject to P on a finite alphabet $\mathcal{X}$
- Define $\pi_R = \left\{ Q \in \mathcal{P}(\mathcal{X}) \,\middle|\, \sum_{x \in \mathcal{X}} Q(x) \log \frac{1}{P(x)} \geq R \right\}$ and

$$\inf_{Q \in \pi_R} D(Q||P) = D(Q_R||P)$$

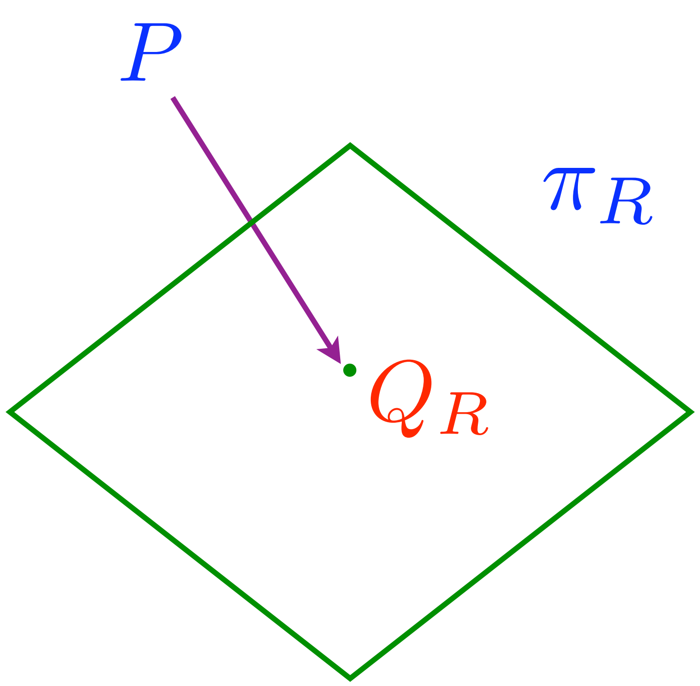
$\Downarrow$ Sanov's theorem

$$\sigma(R) = D(Q_R||P)$$

$\Downarrow$ Theorem 7

$$R_e(r|\mathbf{X}) = \sup_{Q:D(Q||P)<r} H(Q).$$

Q_R : projection of P



2010年7月20日火曜日

Example 4

- Let $\mathbf{X} = (X_1, X_2, \cdots)$ be an i.i.d. subject to $P = (p_1, p_2, \cdots)$ on a countably infinite alphabet $\mathcal{X} = \{1, 2, \cdots\}$, where

$$p_i = \frac{c}{(i+1)(\log(i+1))^4} \quad (i = 1, 2, \cdots).$$

(heavy tail distribution; finite entropy)

⇓ Cramer's theorem (with rate function $I(x)$)

$$\sigma(R) = \inf_{x \geq R} I(x) = 0 \quad (\text{for all } R)$$

⇓ Theorem 7

$$R_e(r|\mathbf{X}) = +\infty \ (\text{for all } r > 0)$$

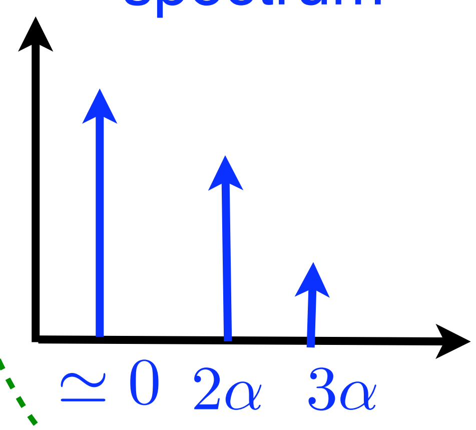
- any finite rate is not enough for exponentially decaying error probability

Example 5 (nonstationary and nonergodic source)

- Consider a subset $S_n \subset \{0,1\}^n$ such that $|S_n| = 2^{\alpha n}$ $(0 < \alpha < 1)$, and fix $\mathbf{x}_0, \mathbf{x}_1 \in \{0,1\}^n - S_n$ with $\mathbf{x}_0 \neq \mathbf{x}_1$
- Define the probability distribution:

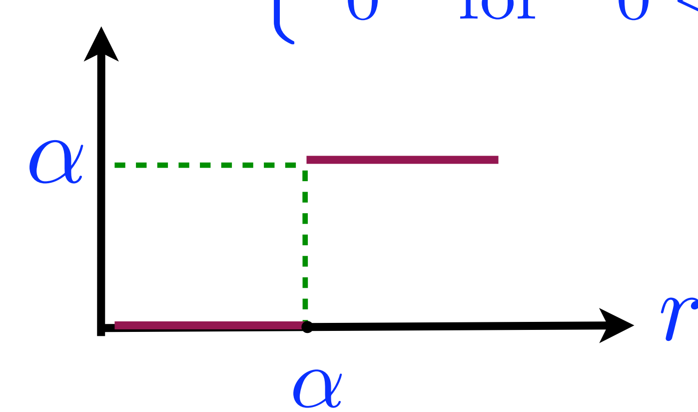
$$P_{X^n}(\mathbf{x}) = \begin{cases} 2^{-2\alpha n} & \text{for} \quad \mathbf{x} \in S_n, \\ 2^{-3\alpha n} & \text{for} \quad \mathbf{x} = \mathbf{x}_1, \\ 1 - 2^{-\alpha n} - 2^{-3\alpha n} & \text{for} \quad \mathbf{x} = \mathbf{x}_0, \\ 0 & \text{for} \quad \mathbf{x} \notin S_n \cup \{\mathbf{x}_1, \mathbf{x}_0\}. \end{cases}$$

three-point spectrum



Cramer's theorem and Sanov theorem do not work !

$$\sigma(R) = \begin{cases} 0 & \text{for} \quad R \le 0, \\ \alpha & \text{for} \quad 0 < R \le 2\alpha, \\ 3\alpha & \text{for} \quad 2\alpha < R \le 3\alpha, \\ +\infty & \text{for} \quad 3\alpha < R \end{cases} \Longrightarrow R_e(r|\mathbf{X}) = \begin{cases} \alpha & \text{for} \quad r > \alpha, \\ 0 & \text{for} \quad 0 < r \le \alpha. \end{cases}$$



Theorem 7

direct calculation

VIII. Hypothesis Testing, and Information-Spectra

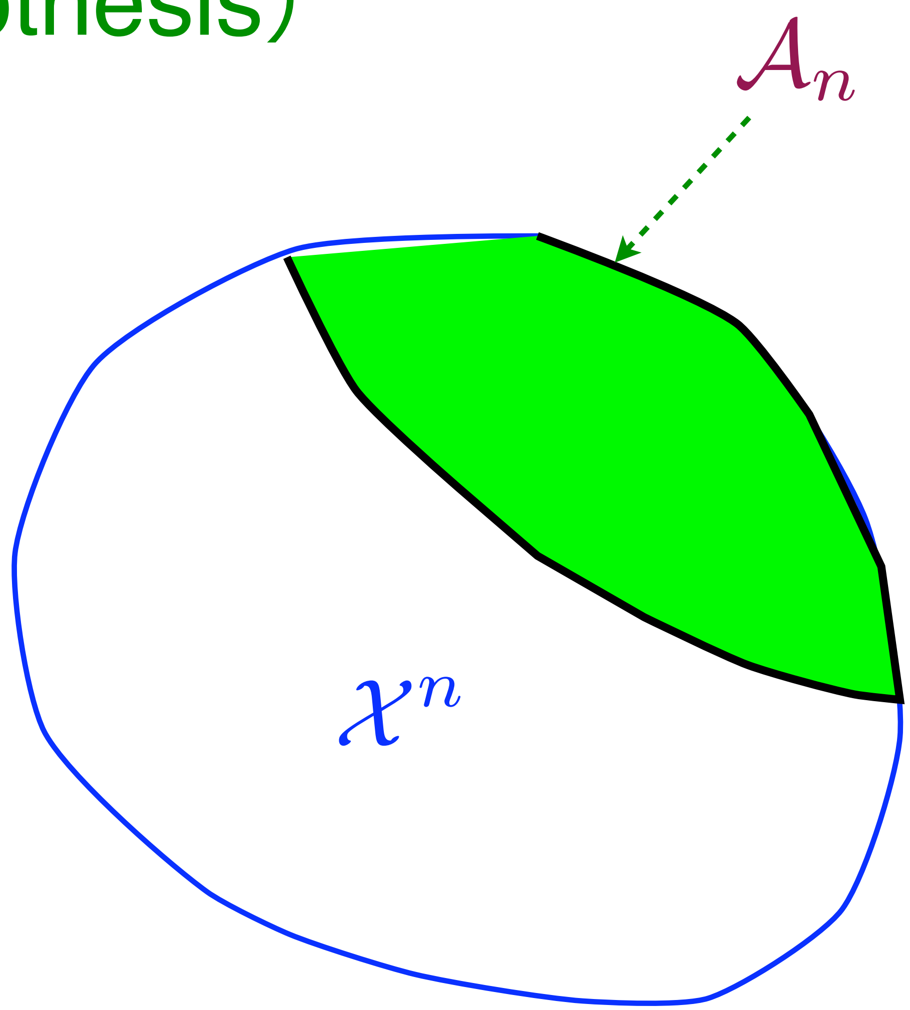
Hypothesis testing

- $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$ and $\overline{\mathbf{X}} = \{\overline{X}^n\}_{n=1}^{\infty}$ are general sources

$$\begin{cases} H : \mathbf{X} & \text{(null hypothesis)} \\ \overline{H} : \overline{\mathbf{X}} & \text{(alternative hypothesis)} \end{cases}$$

- acceptance region $\mathcal{A}_n$:
- decision rule:

$$\begin{cases} \text{if } \mathbf{x} \in \mathcal{A}_n, \text{ output } H : \\ \text{if } \mathbf{x} \notin \mathcal{A}_n, \text{ output } \overline{H} : \end{cases}$$



◉ error probabilities:

$\mu_n \equiv \Pr\{X^n \notin \mathcal{A}_n\}$ (type I error probability)

$\lambda_n \equiv \Pr\{\overline{X}^n \in \mathcal{A}_n\}$ (type II error probability)

◉ Let $B_e(r|\mathbf{X}||\overline{\mathbf{X}})$ denote the supremum of rates R such that $\mu_n \lesssim e^{-nr}, \quad \lambda_n \lesssim e^{-nR}$

(we want to make R as large as possible)

r : fixed

¶Theorem 8 ¶ (Testing power function: Han, 2000)

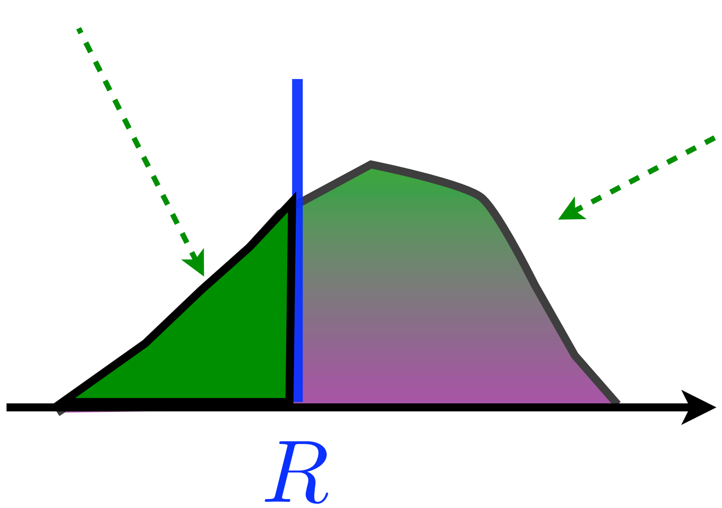
$$B_e(r|\mathbf{X}||\overline{\mathbf{X}}) = \inf_{R}\{R + \eta(R) | \eta(R) < r\}$$

where

$$\eta(R) = \liminf_{n\to\infty} \frac{1}{n} \log \frac{1}{\Pr\left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{P_{\overline{X}^n}(X^n)} \le R \right\}}$$

Radon-Nikodym derivative

$\simeq e^{-n\eta(R)}$

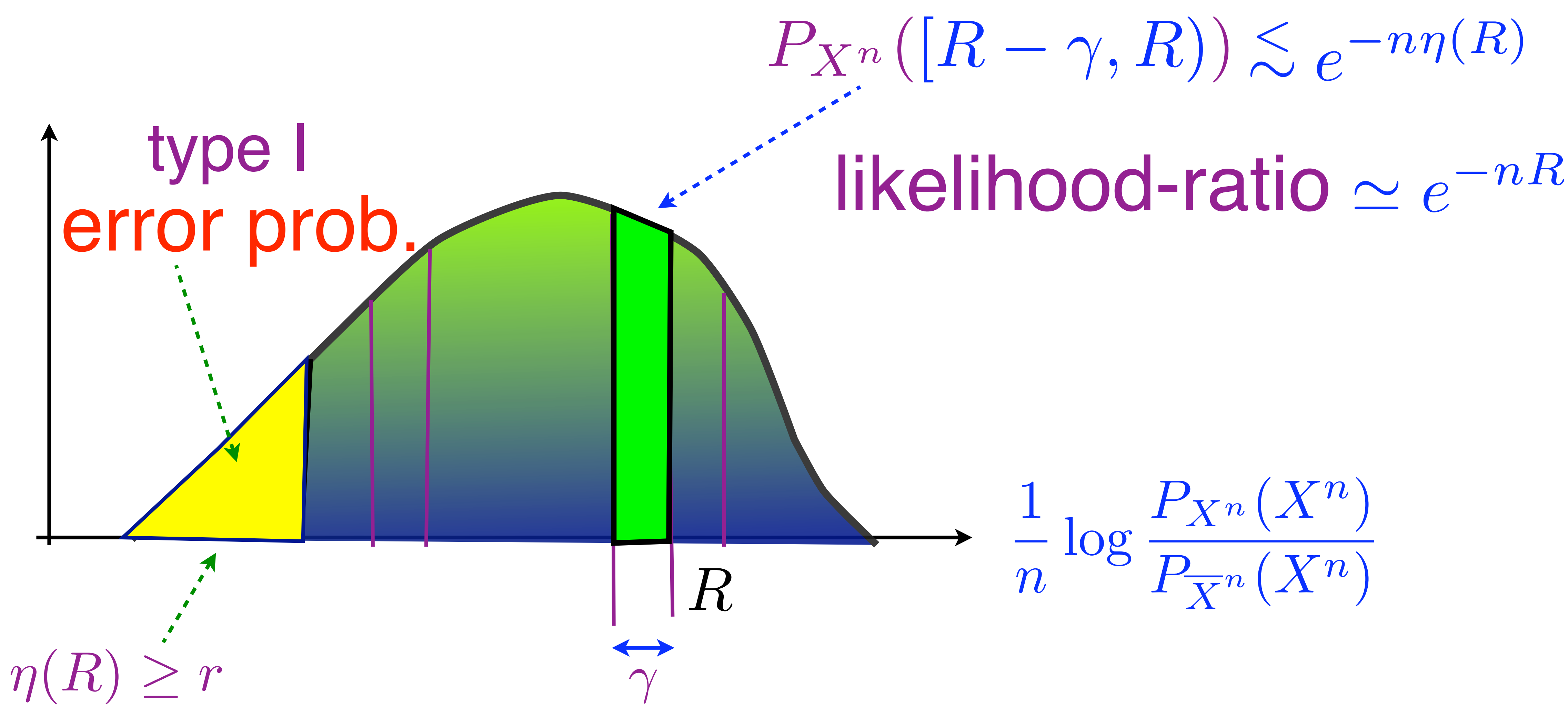


information-spectrum

$$\frac{1}{n} \log \frac{P_{X^n}(X^n)}{P_{\overline{X}^n}(X^n)}$$

normalized
log likelihood-ratio

Hint for the proof (direct part):



◉ $P_{\overline{X}^n}([R-\gamma, R))$

$$\simeq P_{X^n}([R-\gamma, R)) \cdot e^{-nR}$$

$$\lesssim e^{-n\eta(R)} e^{-nR} \simeq e^{-n(R+\eta(R))} \quad (\eta(R) < r)$$

$$\lesssim \sup_{R:\eta(R)<r} e^{-n(R+\eta(R))}$$

Example 6 (Hoeffding, 1965)

- Let $\mathbf{X} = \{X^n\}_{n=1}^{\infty}, \overline{\mathbf{X}} = \{\overline{X}^n\}_{n=1}^{\infty}$ be i.i.d. subject to $P, \overline{P}$, on a finite alphabet $\mathcal{X}$
- Define $\kappa_R = \left\{ Q \in \mathcal{P}(\mathcal{X}) \,\middle|\, \sum_{x \in \mathcal{X}} Q(x) \log \frac{P(x)}{\overline{P}(x)} \le R \right\}$ and

$$\inf_{Q \in \kappa_R} D(Q||P) = D(Q_R||P)$$

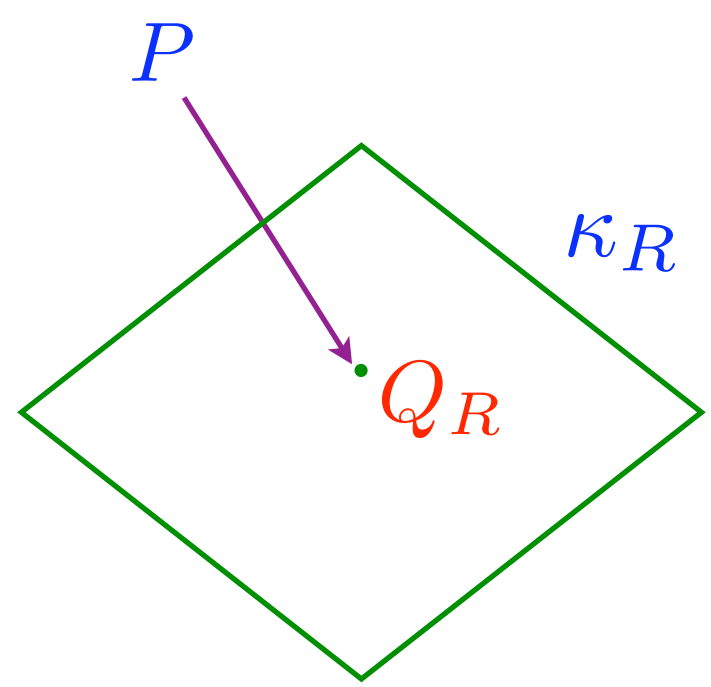
⇓ Sanov's theorem

Q_R : projection of P

$$\eta(R) = D(Q_R||P)$$

⇓ Theorem 8

$$B_e(r|\mathbf{X}||\overline{\mathbf{X}}) = \inf_{Q:D(Q||P)<r} D(Q||\overline{P}).$$



Example 7 (Gaussian sources)

- Let $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$, $\overline{\mathbf{X}} = \{\overline{X}^n\}_{n=1}^{\infty}$ be i.i.d. sources subject to Gaussian distributions $N(\kappa, \sigma^2)$, $N(\overline{\kappa}, \sigma^2)$:

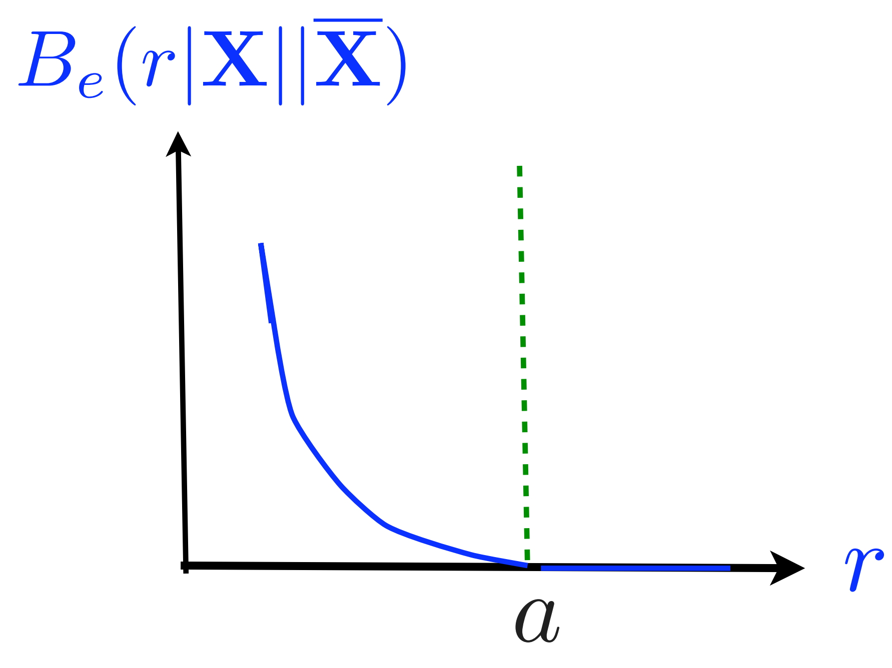
$$P_{\kappa}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\kappa)^2}{2\sigma^2}}, \quad P_{\overline{\kappa}}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\overline{\kappa})^2}{2\sigma^2}}.$$

$\Downarrow$ Cramer's theorem

$$\eta(R) = \inf_{x \le R} I(x) = \min\left\{[a-R]^+, \frac{\sigma^2(R-a)^2}{2(\kappa-\overline{\kappa})^2}\right\} \quad (a = D(P_{\kappa}||P_{\overline{\kappa}}))$$

$\Downarrow$ Theorem 8

$$B_e(r|\mathbf{X}||\overline{\mathbf{X}}) = (\sqrt{r} - \sqrt{a})^2 \mathbf{1}[r \le a]$$



Example 8 (nonstationary and nonergodic source)

- Let P_{X^n} be the source (null) as specified as in Example 6, and $P_{\overline{X}^n}$ be another source (alternative) defined by $P_{\overline{X}^n}(\mathbf{x}) = 2^{-n} \ (\forall \mathbf{x} \in \{0,1\}^n)$.

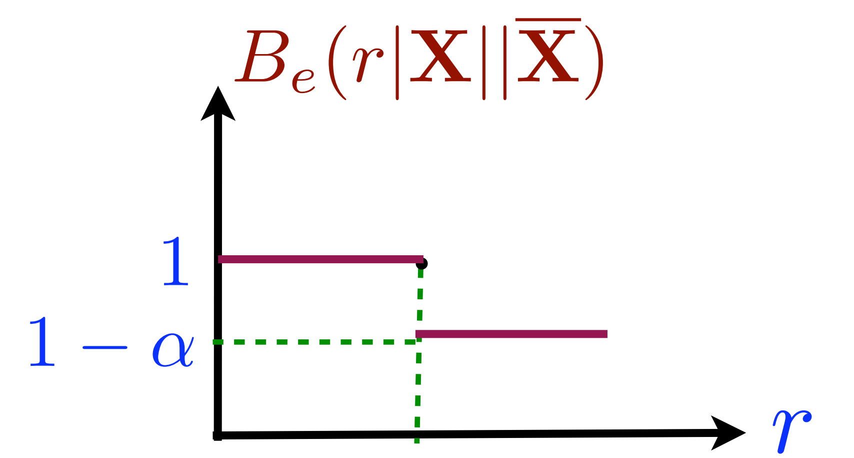
$\Downarrow$

Cramer's theorem and Sanov theorem do not work !
but we can directly calculate:

$$\eta(R) = \begin{cases} +\infty & \text{for} \quad R < 1 - 3\alpha, \\ 3\alpha & \text{for} \quad 1 - 3\alpha \le R < 1 - 2\alpha, \\ \alpha & \text{for} \quad 1 - 2\alpha \le R < 1, \\ 0 & \text{for} \quad 1 \le R. \end{cases}$$

$\Downarrow$ Theorem 8

$$B_e(r|\mathbf{X}||\overline{\mathbf{X}}) = \begin{cases} 1 - \alpha & \text{for} \quad r > \alpha, \\ 1 & \text{for} \quad 0 < r \le \alpha. \end{cases}$$



IX. Generalized Hypothesis Testing, and Information-Spectra

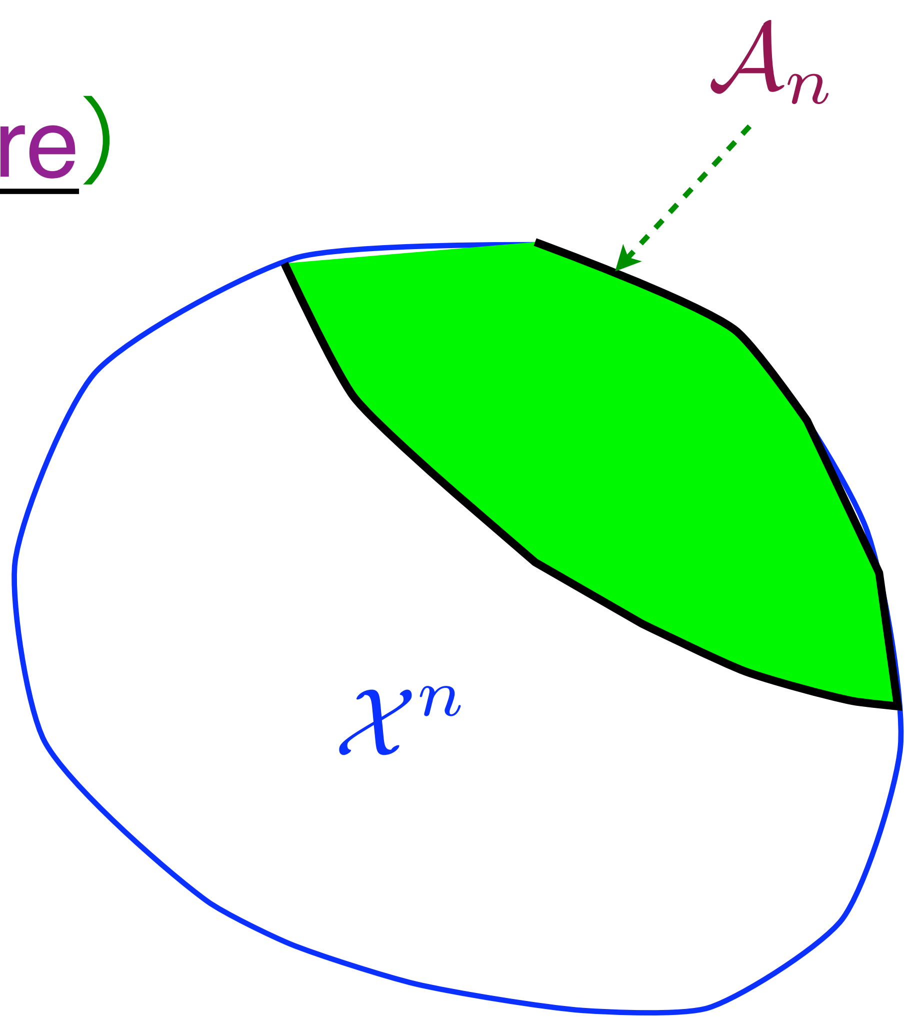
Generalized hypothesis testing

- $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$ is a general source, and $\overline{\mathbf{G}} = \{\overline{G}_n\}_{n=1}^{\infty}$ is a sequence of nonnegative measures

$$\begin{cases} H : \mathbf{X} \text{ (nul hypothesis)} \\ \overline{H} : \overline{\mathbf{G}} \text{ (alternative measure)} \end{cases}$$

- acceptance region $\mathcal{A}_n$:
- decision rule:

$$\begin{cases} \text{if } \mathbf{x} \in \mathcal{A}_n, \text{ output } H : \\ \text{if } \mathbf{x} \notin \mathcal{A}_n, \text{ output } \overline{H} : \end{cases}$$



- error probability (measure):

$\mu_n \equiv \Pr\{X^n \notin \mathcal{A}_n\}$ (type I error probability)

$\lambda_n \equiv \overline{G}_n(\mathcal{A}_n)$ (type II measure)

- Let $B_e(r|\mathbf{X}||\overline{\mathbf{G}})$ denote the supremum of rates R such that $\mu_n \lesssim e^{-nr}, \ \lambda_n \lesssim e^{-nR}$

(we want to make R as large as possible)

r : fixed

¶Theorem 9 ¶ (generalized testing power function: Han,1998)

$$B_e(r|\mathbf{X}||\overline{\mathbf{G}}) = \inf_R \{R + \eta_{\overline{G}}(R) | \eta_{\overline{G}}(R) < r\}$$

where $\eta_{\overline{G}}(R) = \liminf_{n \to \infty} \frac{1}{n} \log \frac{1}{\Pr\left\{\frac{1}{n}\log\frac{P_{X^n}(X^n)}{\overline{G}_n(X^n)} \le R\right\}}$

Radon-Nikodym derivative

Hypothesis testing and source coding

- As an alternative measure $\overline{\mathbf{G}} = \{\overline{G}_n\}_{n=1}^{\infty}$, consider the counting measure $\overline{\mathbf{C}} = \{\overline{C}_n\}_{n=1}^{\infty}$ defined by $\overline{C}_n(\mathcal{A}_n) = |\mathcal{A}_n|$

$\Downarrow$ then

- we can establish one-to-one correspondence between source coding and hypothesis testing

$\Downarrow$

source coding | hypothesis testing

indifferently identify

correctly decoded set = acceptance region

decoding error probability = type I error probability

# of correctly decoded sequences = type II measure

⇓ Thus

◉ one-to-one correspondence between hypothesis testing and source coding holds:

Therefore, we have a formula connecting souece coding and hypothesis testing:

¶Theorem10 (Han,1998) ¶

$$R_e(r|\mathbf{X}) = -B_e(r|\mathbf{X}||\overline{\mathbf{C}})$$

The proof:

$$\lambda_n = \overline{C}_n(\mathcal{A}_n) = |\mathcal{A}_n| = M_n = e^{nr_n} \text{ where } r_n = \frac{1}{n}\log M_n$$

Thus, $\liminf_{n\to\infty} \frac{1}{n}\log\frac{1}{\lambda_n} = -\limsup_{n\to\infty} r_n,$

which means $R_e(r|\mathbf{X}) = -B_e(r|\mathbf{X}||\overline{\mathbf{C}})$

¶ Theorem 11: another proof of Theorem 7 ¶

$$R_e(r|\mathbf{X}) = \sup_R\{R - \sigma(R)|\sigma(R) < r\}$$

The proof:

It is easy to check that $\eta_{\overline{C}}(R) = \sigma(-R)$.

Then, Theorem 9 implies

$$B_e(r|\mathbf{X}||\overline{\mathbf{C}}) = -\sup_R\{R - \sigma(R)|\sigma(R) < r\}$$

Finally, use Theorem 10.

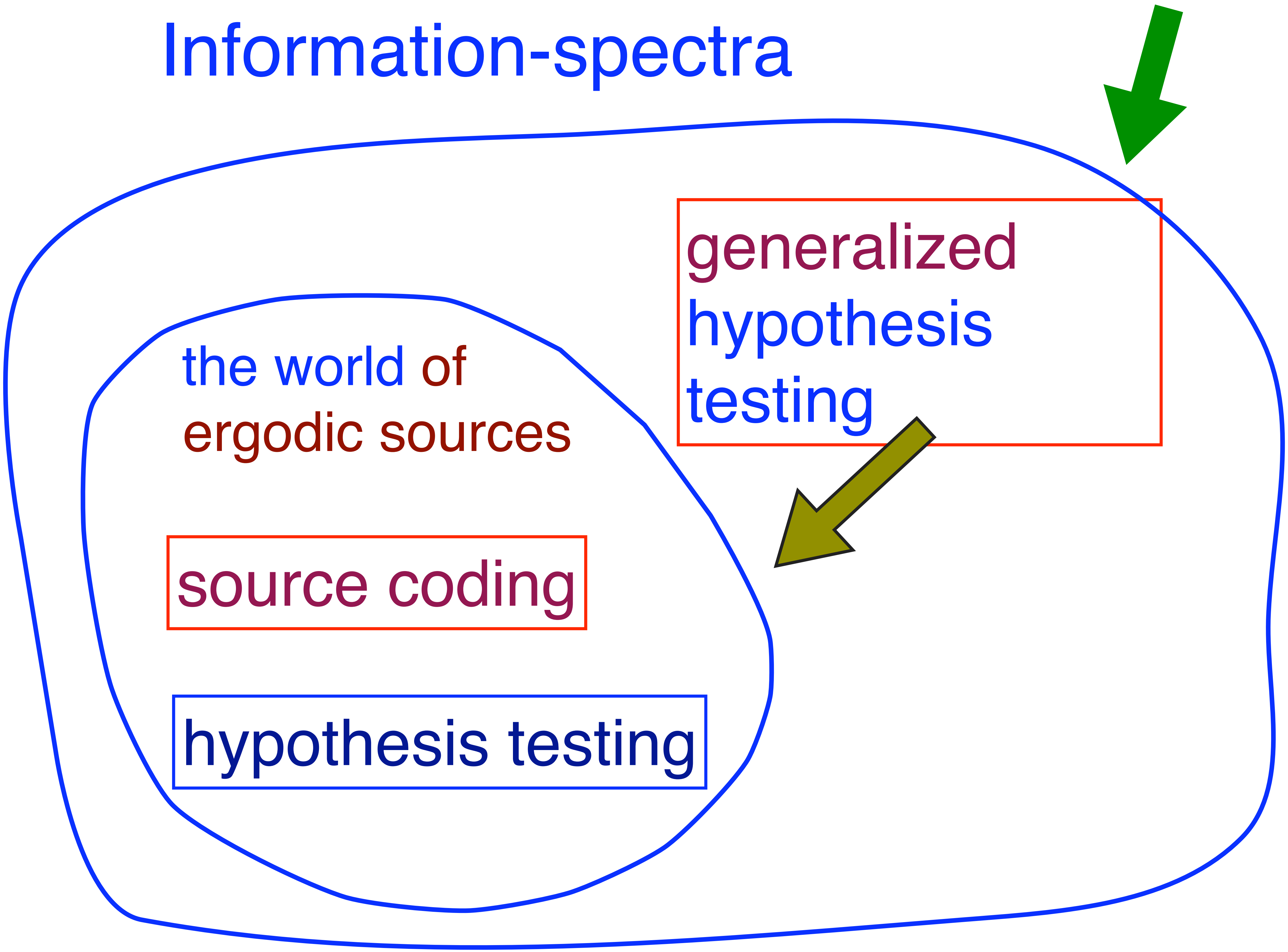
In a similar way, we have formulas connecting source coding and hypothesis testing:

$$R_f(\mathbf{X}) = -B(\mathbf{X}||\overline{\mathbf{C}}) = \overline{H}(\mathbf{X})$$

$$R_f(\varepsilon|\mathbf{X}) = -B_f(\varepsilon|\mathbf{X}||\overline{\mathbf{C}}) = \inf\{R|F(R) \leq \varepsilon\}$$

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

Hypothesis testing and the world of Information-spectra



Conclusion

Information-spectrum approach
provides
conceptually insightful viewpoints
in musing on Information Theory!

Thank you !

reference book:

Te Sun Han

Information-Spectrum Methods
in Information Theory,

Springer, 2003