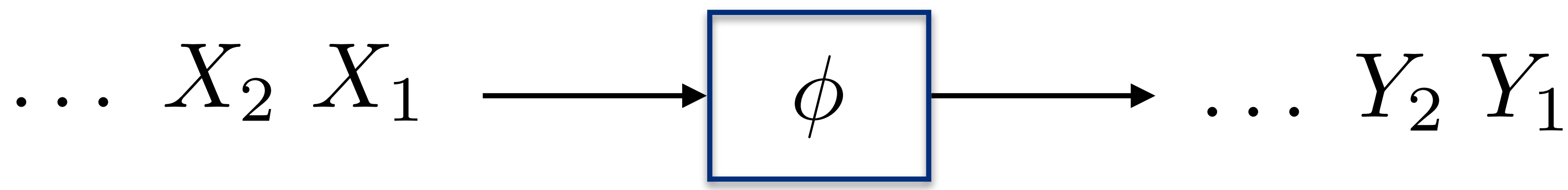
Random Number Generation: Old and New

WPI2019@Univ. Hong Kong
August, 2019

Shun Watanabe

joint work with Te Sun Han

What is Random Number Generation (RNG)



$\boldsymbol{X} = \{X^m = (X_1, \dots, X_m)\}_{m=1}^{\infty}$ coin process

$\boldsymbol{Y} = \{Y^n = (Y_1, \dots, Y_n)\}_{n=1}^{\infty}$ target process

We shall simulate the target process using the output from the coin process exactly.

von Neumann's Algorithm

$\boldsymbol{X}$ is i.i.d. Bernoulli: $\Pr(X_i = 0) = p, \;\; \Pr(X_i = 1) = 1 - p, \quad p \neq 0, \frac{1}{2}, 1$

$\boldsymbol{Y}$ is i.i.d. unbiased Bernoulli

von Neumann's Algorithm

$\boldsymbol{X}$ is i.i.d. Bernoulli: $\Pr(X_i = 0) = p, \ \Pr(X_i = 1) = 1 - p, \quad p \neq 0, \frac{1}{2}, 1$

$\boldsymbol{Y}$ is i.i.d. unbiased Bernoulli

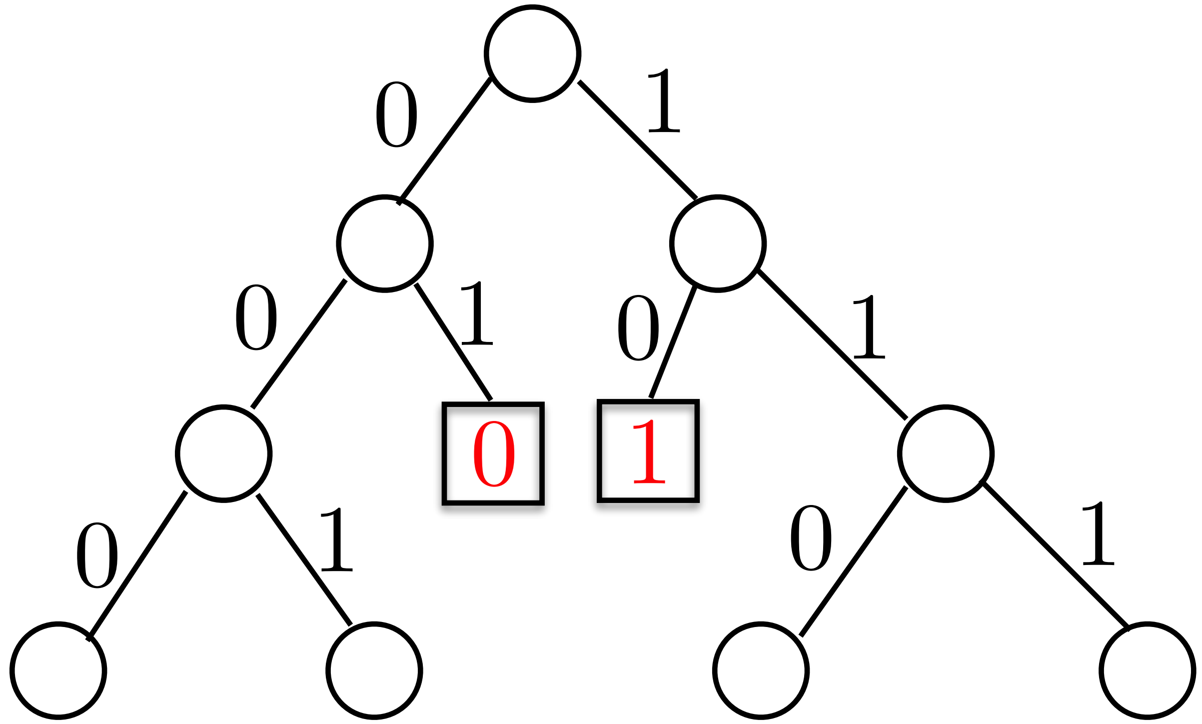
1: Set *i=1*.

2: If

$$X_{2i-1} = 0, \ X_{2i} = 1 \Longrightarrow \text{outputs } 0$$

$$X_{2i-1} = 1, \ X_{2i} = 0 \Longrightarrow \text{outputs } 1$$

Else, set *i=i+1* and repeat Step 2.



von Neumann's Algorithm (cont'd)

T : stopping time (#of coin tosses until the algorithm terminates)

von Neumann's Algorithm (cont'd)

T : stopping time (#of coin tosses until the algorithm terminates)

$T/2$ follows geometric distribution with parameter $2p(1-p)$...

$$\mathbb{E}[T] = \frac{1}{p(1-p)}$$

Generalization by Elias

von Neuman's algorithm was extended in various direction.

(eg. Samuelson '68, Hoeffding-Simons '70, Elias '72, Peres '92)

Generalization by Elias

von Neuman's algorithm was extended in various direction.
(eg. Samuelson '68, Hoeffding-Simons '70, Elias '72, Peres '92)

Let's simulate unbiased n bits Y^n simultaneously.

Generalization by Elias

von Neuman's algorithm was extended in various direction.
(eg. Samuelson '68, Hoeffding-Simons '70, Elias '72, Peres '92)

Let's simulate unbiased n bits Y^n simultaneously.

Set $m = \left\lceil n\left(\frac{1}{H(X)} + \delta\right)\right\rceil$ and let $\{0,1\}^m = \bigcup_{k=0}^{m} \mathcal{T}_k$ $\quad \mathcal{T}_k := \{x^m : w_H(x^m) = k\}$

For each $k \in \mathcal{G} := \{k' : |\mathcal{T}_{k'}| \geq 2^n\}$, take the largest subset $\mathcal{C}_k \subseteq \mathcal{T}_k$ with $|\mathcal{C}_k| = c2^n$

Let $\varphi_k : \mathcal{C}_k \to \{0,1\}^n$ be "balanced" assignment

Generalization by Elias

von Neuman's algorithm was extended in various direction.
(eg. Samuelson '68, Hoeffding-Simons '70, Elias '72, Peres '92)

Let's simulate unbiased n bits Y^n simultaneously.

Set $m = \left\lceil n\left(\frac{1}{H(X)} + \delta\right)\right\rceil$ and let $\{0,1\}^m = \bigcup_{k=0}^{m} \mathcal{T}_k$ $\mathcal{T}_k := \{x^m : w_H(x^m) = k\}$

For each $k \in \mathcal{G} := \{k' : |\mathcal{T}_{k'}| \geq 2^n\}$, take the largest subset $\mathcal{C}_k \subseteq \mathcal{T}_k$ with $|\mathcal{C}_k| = c2^n$

Let $\varphi_k : \mathcal{C}_k \to \{0,1\}^n$ be "balanced" assignment

Upon observing $X^m = (X_1, \ldots, X_m) \in \mathcal{C}_k$ for some $k \in \mathcal{G}$, outputs $\varphi_k(X^m)$

Otherwise, go to the next block

Generalization by Elias

von Neuman's algorithm was extended in various direction.
(eg. Samuelson '68, Hoeffding-Simons '70, Elias '72, Peres '92)

Let's simulate unbiased n bits Y^n simultaneously.

Set $m = \left\lceil n\left(\frac{1}{H(X)} + \delta\right)\right\rceil$ and let $\{0,1\}^m = \bigcup_{k=0}^{m} \mathcal{T}_k$ $\quad \mathcal{T}_k := \{x^m : w_H(x^m) = k\}$

For each $k \in \mathcal{G} := \{k' : |\mathcal{T}_{k'}| \geq 2^n\}$, take the largest subset $\mathcal{C}_k \subseteq \mathcal{T}_k$ with $|\mathcal{C}_k| = c2^n$

Let $\varphi_k : \mathcal{C}_k \to \{0,1\}^n$ be "balanced" assignment

Upon observing $X^m = (X_1, \ldots, X_m) \in \mathcal{C}_k$ for some $k \in \mathcal{G}$, outputs $\varphi_k(X^m)$

Otherwise, go to the next block

T/m follows geometric distribution with parameter $\Pr\left(X^m \in \bigcup_{k \in \mathcal{G}} \mathcal{C}_k\right)$

$$\frac{1}{n}\mathbb{E}[T] = \frac{m/n}{\Pr\left(X^m \in \bigcup_{k \in \mathcal{G}} \mathcal{C}_k\right)} \to \frac{1}{H(X)} \qquad (n \to \infty, \delta \to 0)$$

optimal

Knuth-Yao's Algorithm

X is i.i.d. unbiased Bernoulli

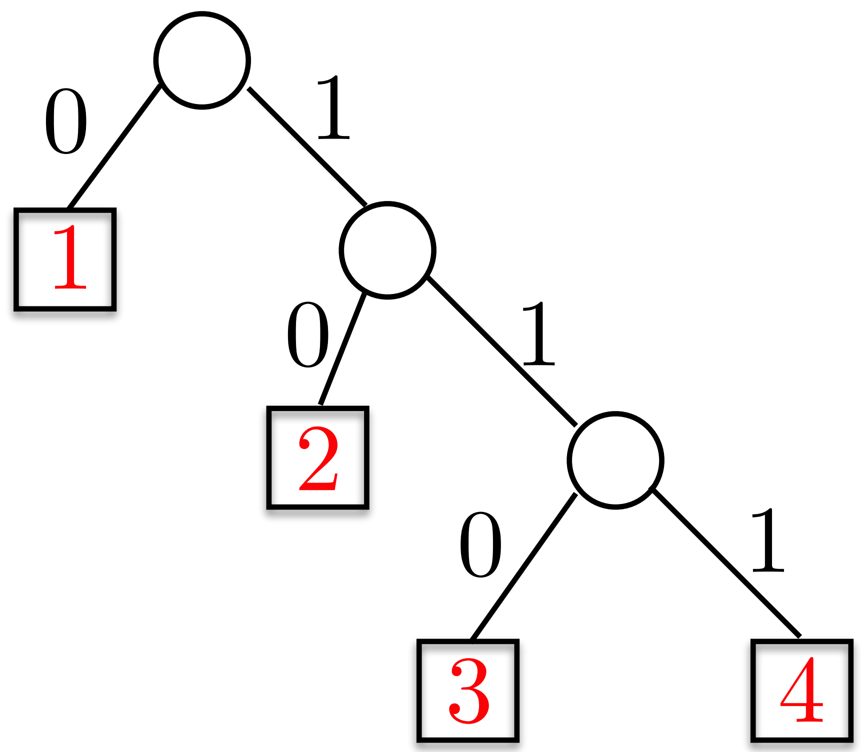
Knuth-Yao's Algorithm

$\boldsymbol{X}$ is i.i.d. unbiased Bernoulli

The case with didactic target distribution.

eg) $P_Y = (1/2, 1/4, 1/8, 1/8)$

Use the Huffman code tree…



Since $\ell(\phi^{-1}(y)) = \log \frac{1}{P_Y(y)}$

$$\mathbb{E}[T] = H(Y)$$

Knuth-Yao's Algorithm (cont'd)

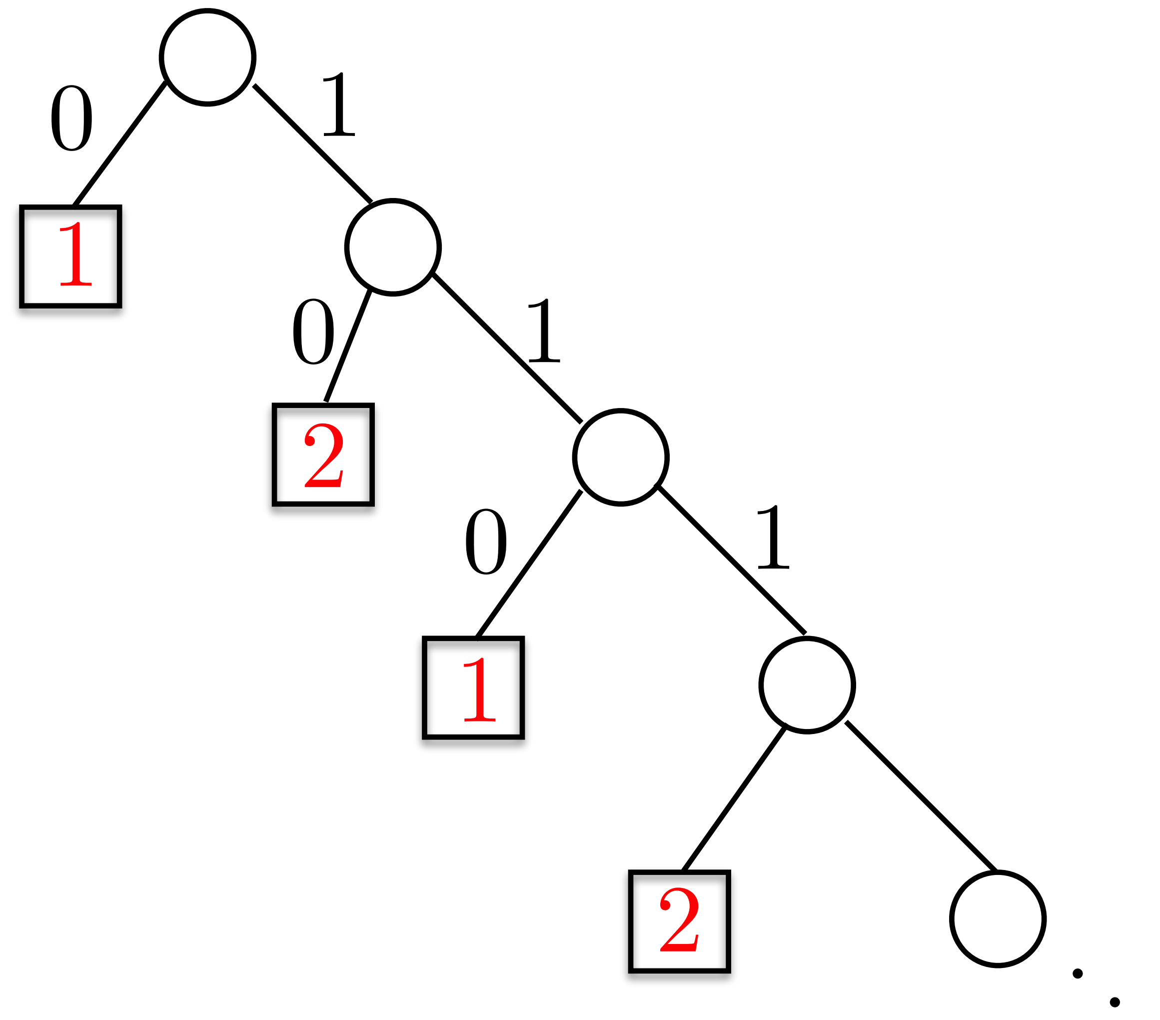
General case:

eg) $P_Y = (2/3, 1/3)$

Consider the binary expansion

$$\frac{2}{3} = \frac{1}{2} + \frac{1}{2^3} + \frac{1}{2^5} + \cdots \qquad \frac{1}{3} = \frac{1}{2^2} + \frac{1}{2^4} + \frac{1}{2^6} + \cdots$$

Then, use the Huffman code tree…



Knuth-Yao's Algorithm (cont'd)

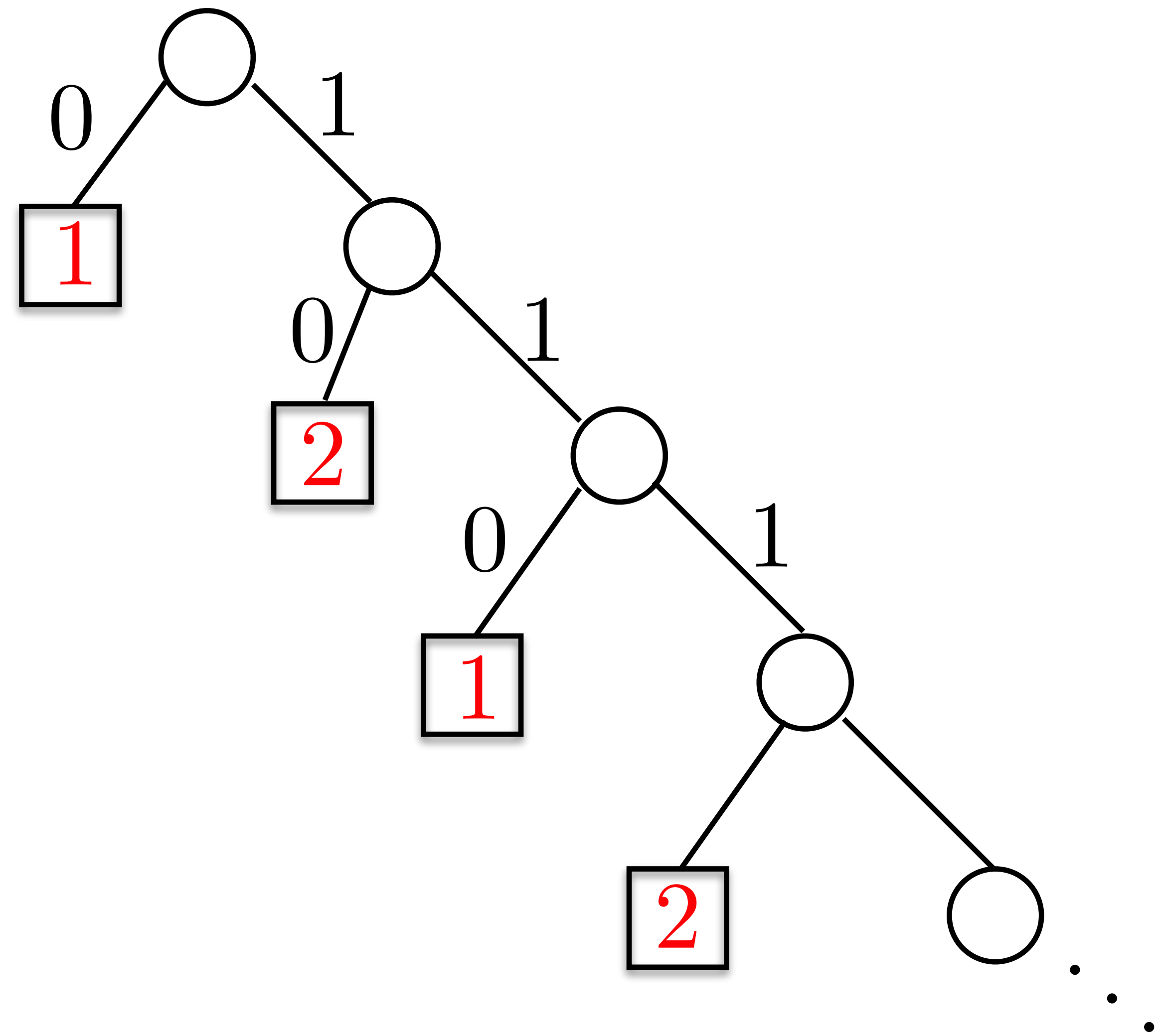
General case:

eg) $P_Y = (2/3, 1/3)$

Consider the binary expansion

$$\frac{2}{3} = \frac{1}{2} + \frac{1}{2^3} + \frac{1}{2^5} + \cdots \qquad \frac{1}{3} = \frac{1}{2^2} + \frac{1}{2^4} + \frac{1}{2^6} + \cdots$$

Then, use the Huffman code tree…



Theorem (Knuth-Yao)

The Knuth-Yao's algorithm satisfies

$$\mathbb{E}[T] \le H(Y) + 2$$

Any RNG algorithm must satisfy

$$\mathbb{E}[T] \ge H(Y)$$

RNG from Biased coin to Biased Target

Roche '91   an algorithm based on arithmetic coding

Abrahams '96   an extension of Knuth-Yao algorithm

Han-Hoshi '97   "interval algorithm"

etc.

A Converse Bound

<u>Proposition</u> (Han-Hoshi)

When the coin process is i.i.d., any RNG algorithm simulating Y^n exactly must satisfy

$$\mathbb{E}[T] \geq \frac{H(Y^n)}{H(X)}$$

A Converse Bound

<u>Proposition</u> (Han-Hoshi)

When the coin process is i.i.d., any RNG algorithm simulating Y^n exactly must satisfy

$$\mathbb{E}[T] \geq \frac{H(Y^n)}{H(X)}$$

Proof sketch)

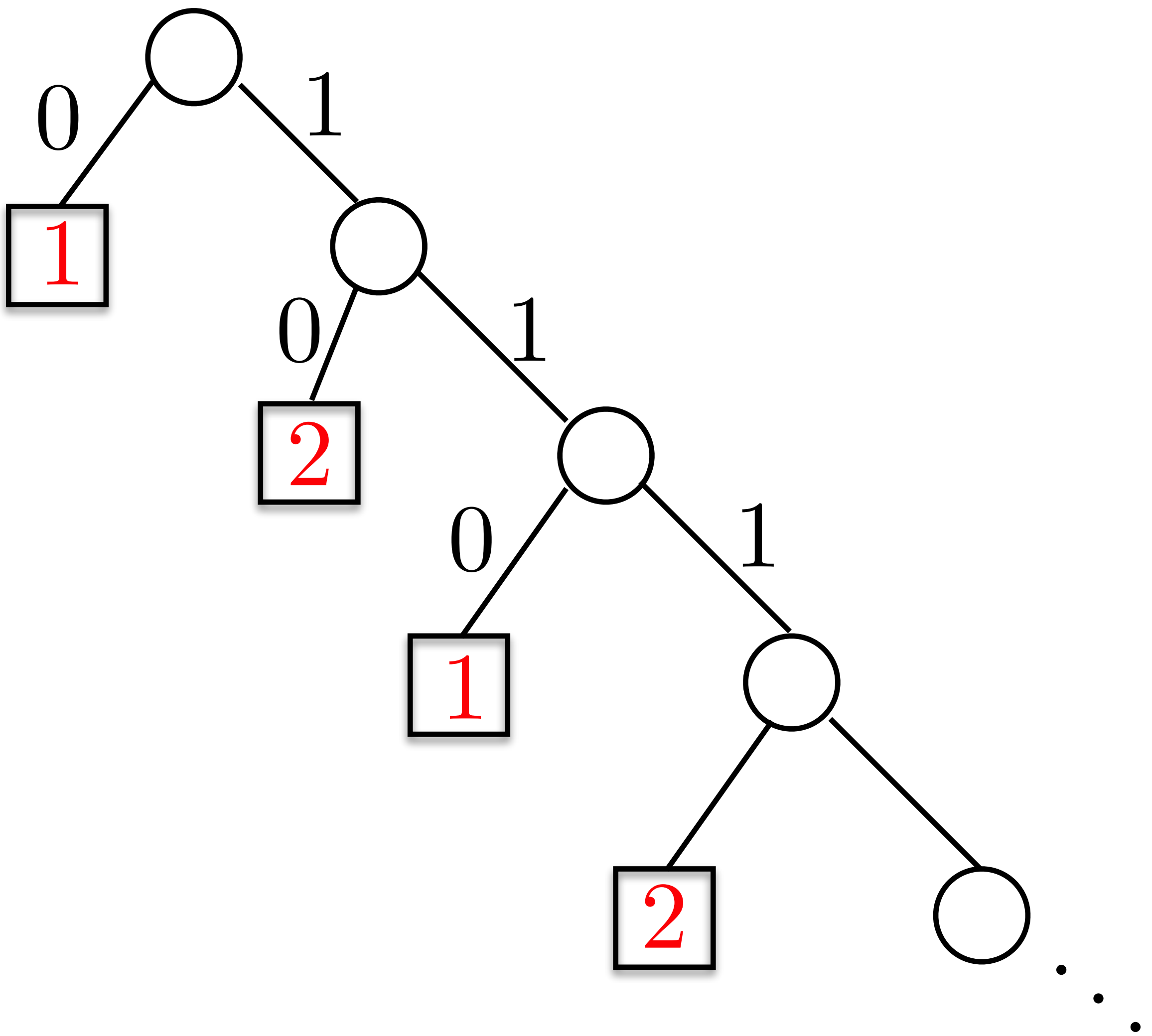
Z : RV describing the leaves induced by $\boldsymbol{X}$

For an i.i.d. coin process,

$$H(Z) = \mathbb{E}[T] \cdot H(X)$$

Since Y^n is a function of Z,

$$H(Y^n) \leq H(Z) = \mathbb{E}[T] \cdot H(X)$$



Interval Algorithm

Basic Idea

For a sequence $x^m \in \mathcal{X}^m$, assign $\mathcal{I}_{x^m} = [\underline{\alpha}_{x^m}, \overline{\alpha}_{x^m}) \subseteq [0,1)$ with $|\mathcal{I}_{x^m}| = P_{X^m}(x^m)$

For a sequence $y^n \in \mathcal{Y}^n$, assign $\mathcal{J}_{y^n} = [\underline{\beta}_{y^n}, \overline{\beta}_{y^n}) \subseteq [0,1)$ with $|\mathcal{J}_{y^n}| = P_{Y^n}(y^n)$

Upon observing x^m, if $\mathcal{I}_{x^m} \subseteq \mathcal{J}_{y^n}$ for some y^n, outputs y^n.

Interval Algorithm

Basic Idea

For a sequence $x^m \in \mathcal{X}^m$, assign $\mathcal{I}_{x^m} = [\underline{\alpha}_{x^m}, \overline{\alpha}_{x^m}) \subseteq [0,1)$ with $|\mathcal{I}_{x^m}| = P_{X^m}(x^m)$

For a sequence $y^n \in \mathcal{Y}^n$, assign $\mathcal{J}_{y^n} = [\underline{\beta}_{y^n}, \overline{\beta}_{y^n}) \subseteq [0,1)$ with $|\mathcal{J}_{y^n}| = P_{Y^n}(y^n)$

Upon observing x^m, if $\mathcal{I}_{x^m} \subseteq \mathcal{J}_{y^n}$ for some y^n, outputs y^n.

More precisely,…

$\underline{\alpha}_s = \underline{\beta}_s = 0$ and $\overline{\alpha}_s = \overline{\beta}_s = 1$ for $s = t = \perp$

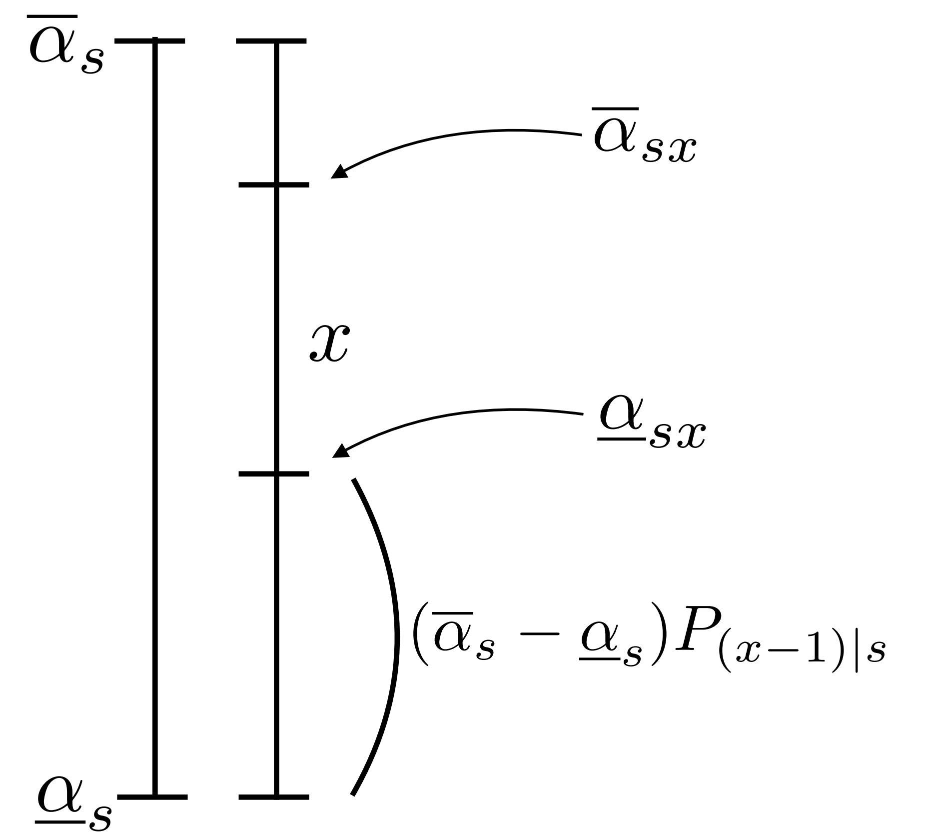
$$\underline{\alpha}_{sx} := \underline{\alpha}_s + (\overline{\alpha}_s - \underline{\alpha}_s) P_{(x-1)|s}$$

$$\overline{\alpha}_{sx} := \underline{\alpha}_s + (\overline{\alpha}_s - \underline{\alpha}_s) P_{x|s}$$

for $s \in \mathcal{X}^i,\ x \in \mathcal{X}$

$$P_{s|x} := \sum_{k=1}^{x} P_{X_{i+1}|X^i}(k|s)$$

$\underline{\beta}_t$ and $\overline{\beta}_t$ are defined similarly by P_{Y^n}.



Interval Algorithm (cont'd)

1: (Initialization) Set $s = t = \perp$, $i = 0$ and $j = 1$.

2: If $[\underline{\alpha}_s, \overline{\alpha}_x) \subseteq [\underline{\beta}_{ty}, \overline{\beta}_{ty})$ for some $y \in \mathcal{Y}$, then output $y_j = y$ and go to Step 3;

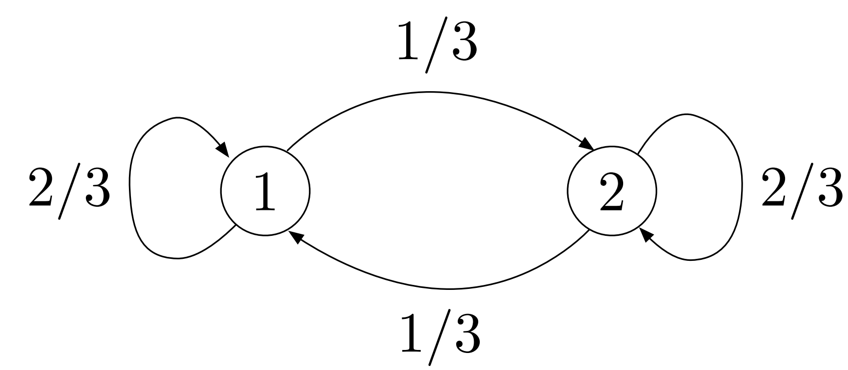
Otherwise, set $i = i + 1$, $s = sx_i$, and repeat Step 2 again.

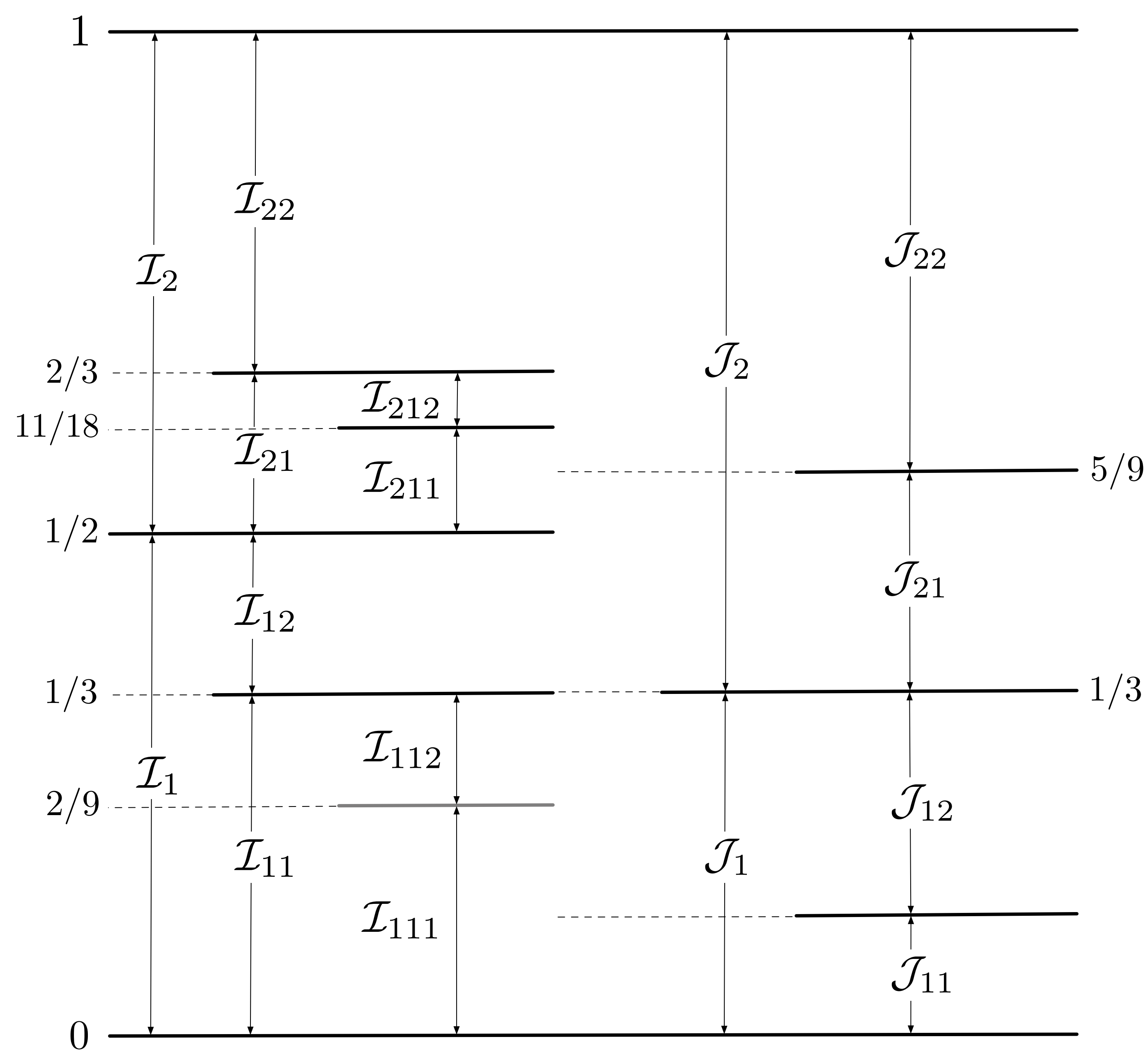
3: If $j = n$, terminates; otherwise, set $t = ty_j$, $j = j + 1$, and go to Step 2.

Interval Algorithm (cont'd)

Example) The coin $\{X^m\}_{m=1}^{\infty}$ is Markov chain.

The target $\{Y^n\}_{y=1}^{\infty}$ is i.i.d with $P_Y = (1/3, 2/3)$.





Interval Algorithm (cont'd)

The algorithm itself is quite simple, but performance analysis is not straightforward…

Theorem (Han-Hoshi '97)

When the coin process is i.i.d., the stopping time of the interval algorithm satisfies

$$\mathbb{E}[T] \leq \frac{H(Y^n)}{H(X)} + \frac{\log(2|\mathcal{Y}| - 1)}{H(X)} + \frac{H(X)}{(1 - p_{\max})H(X)}$$

where $p_{\max} = \max_{x \in \mathcal{X}} P_X(x)$.

Interval Algorithm (cont'd)

The algorithm itself is quite simple, but performance analysis is not straightforward…

Theorem (Han-Hoshi '97)

When the coin process is i.i.d., the stopping time of the interval algorithm satisfies

$$\mathbb{E}[T] \leq \frac{H(Y^n)}{H(X)} + \frac{\log(2|\mathcal{Y}| - 1)}{H(X)} + \frac{H(X)}{(1 - p_{\max})H(X)}$$

where $p_{\max} = \max_{x \in \mathcal{X}} P_X(x)$.

Asymptotically,

$$\limsup_{n \to \infty} \frac{1}{n} \mathbb{E}[T] \leq \frac{H(\boldsymbol{Y})}{H(X)}$$

optimal

where

$$H(\boldsymbol{Y}) = \limsup_{n \to \infty} \frac{1}{n} H(Y^n)$$

sup entropy rate

Work on Interval Algorithm

Oohama ’11: refined analysis for i.i.d. coin process

Uyematsu-Kayana ’00: analysis for ergodic coin/target processes

Uyematsu-Kayana ’99: large deviation analysis for i.i.d. coin/target processes

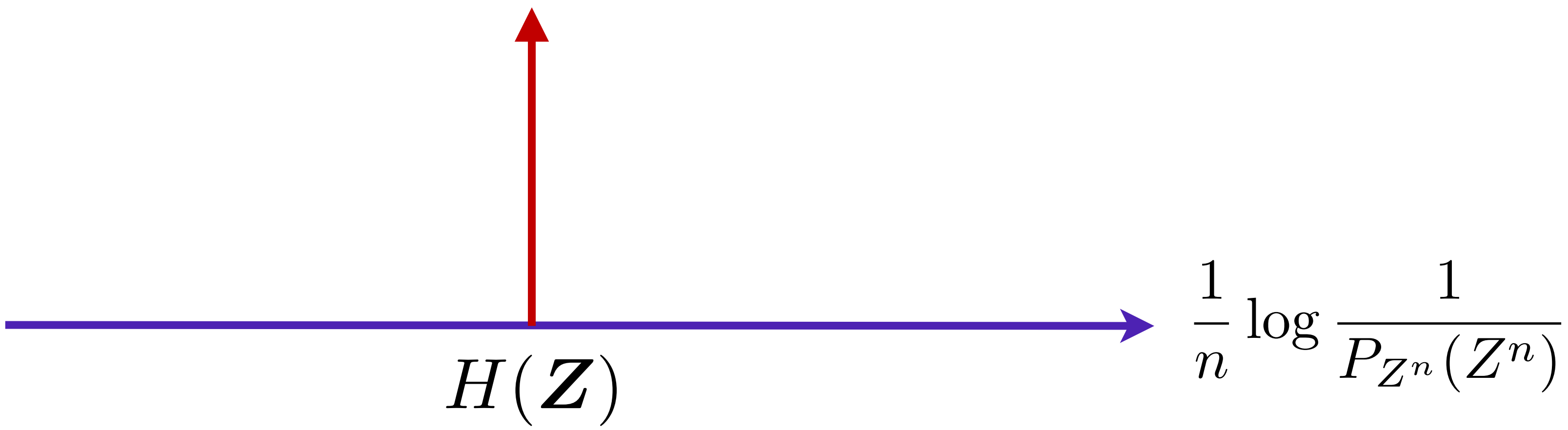
W.-Han ’19: analysis via information spectrum approach

- analysis for general coin/target processes
- optimality of the interval algorithm among any RNG algorithms for wide class of general coin/target processes

Information Spectrum Approach: Brief Review

When $\{Z^n\}_{n=1}^{\infty}$ is ergodic, the AEP states

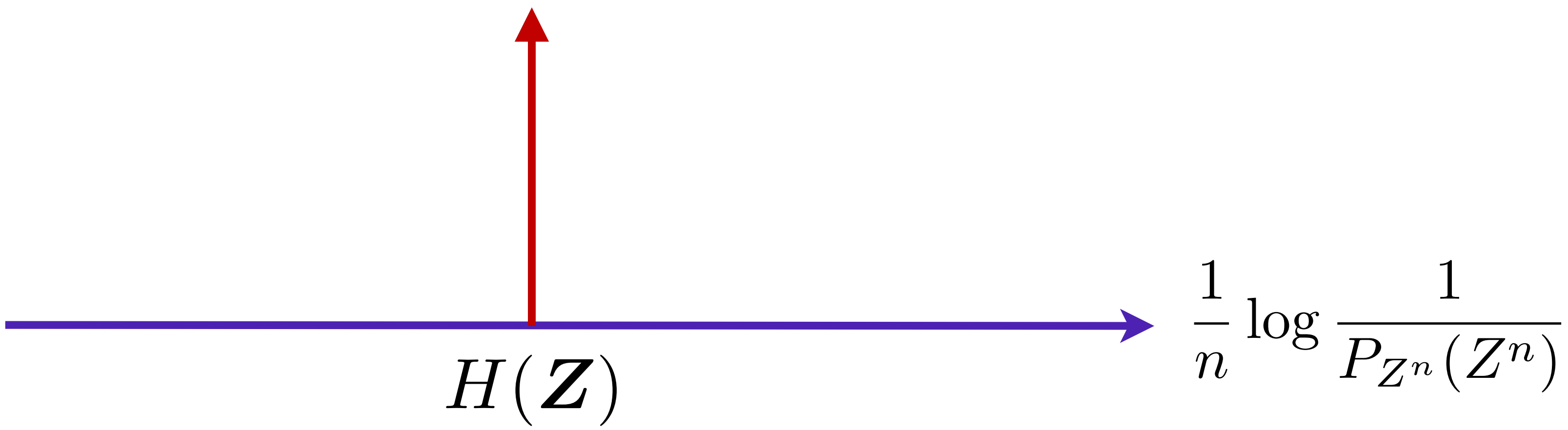
$$\Pr\left(\left|\frac{1}{n}\log\frac{1}{P_{Z^n}(Z^n)} - H(\boldsymbol{Z})\right| \leq \delta\right) \to 1$$



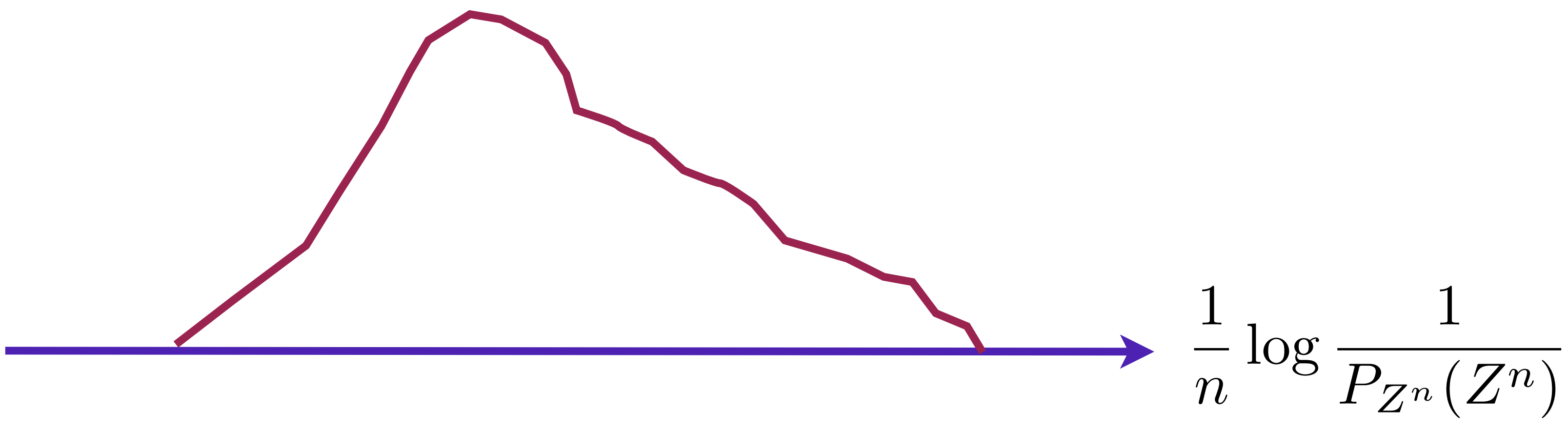
Information Spectrum Approach: Brief Review

When $\{Z^n\}_{n=1}^{\infty}$ is ergodic, the AEP states

$$\Pr\left(\left|\frac{1}{n}\log\frac{1}{P_{Z^n}(Z^n)} - H(\boldsymbol{Z})\right| \leq \delta\right) \to 1$$



In general, spectrum is spreading (eg. reducible Markov chain)



Information Spectrum Approach (cont'd)

To handle spreading spectrum, it is more convenient to define "typical sets" by

$$\mathcal{S}_n(\lambda) := \left\{ z^n : \log \frac{1}{P_{Z^n}(z^n)} \geq \lambda \right\}$$

$$\mathcal{T}_n(\lambda) := \left\{ z^n : \log \frac{1}{P_{Z^n}(z^n)} \leq \lambda \right\}$$

Information Spectrum Approach (cont'd)

To handle spreading spectrum, it is more convenient to define "typical sets" by

$$\mathcal{S}_n(\lambda) := \left\{ z^n : \log \frac{1}{P_{Z^n}(z^n)} \geq \lambda \right\}$$

$$\mathcal{T}_n(\lambda) := \left\{ z^n : \log \frac{1}{P_{Z^n}(z^n)} \leq \lambda \right\}$$

If we define

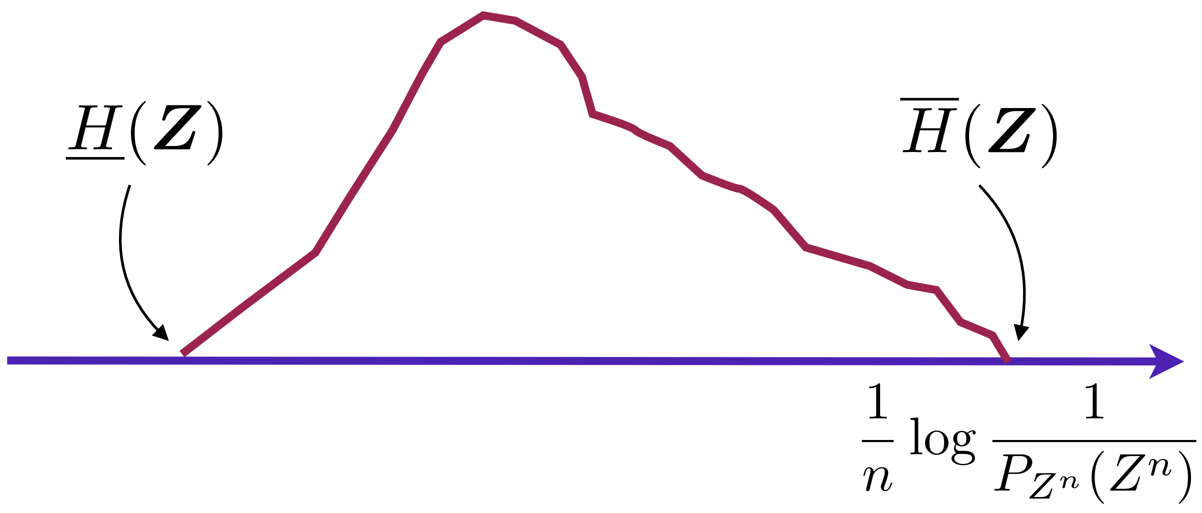
$$\underline{H}(\boldsymbol{Z}) := \sup \left\{ a : \lim_{n\to\infty} \Pr\left(\frac{1}{n} \log \frac{1}{P_{Z^n}(Z^n)} \leq a \right) = 0 \right\}$$ spectral inf-entropy

$$\overline{H}(\boldsymbol{Z}) := \inf \left\{ a : \lim_{n\to\infty} \Pr\left(\frac{1}{n} \log \frac{1}{P_{Z^n}(Z^n)} \geq a \right) = 0 \right\}$$ spectral sup-entropy

then

$$\lambda = n(\underline{H}(\boldsymbol{Z}) - \delta) \Longrightarrow P_{Z^n}(\mathcal{S}_n(\lambda)) \to 1$$

$$\lambda = n(\overline{H}(\boldsymbol{Z}) + \delta) \Longrightarrow P_{Z^n}(\mathcal{T}_n(\lambda)) \to 1$$



Information Spectral Analysis of Interval Algorithm

Theorem (W.-Han '19)

For the interval algorithm, the overflow probability of the stopping time satisfies

$$\Pr(T > m) \leq P_{X^m}(\mathcal{S}_m^c(\lambda)) + P_{Y^n}(\mathcal{T}_n^c(\tau)) + 2^{-\lambda+\tau+1}$$

where

$$\mathcal{S}_m(\lambda) := \left\{ x^m \in \mathcal{X}^m : \log \frac{1}{P_{X^m}(x^m)} \geq \lambda \right\}$$

$$\mathcal{T}_n(\tau) := \left\{ y^n \in \mathcal{Y}^n : \log \frac{1}{P_{Y^n}(y^n)} \leq \tau \right\}$$

If we set $\lambda \simeq m\underline{H}(\boldsymbol{X})$, $\tau \simeq n\overline{H}(\boldsymbol{Y})$, and $m \simeq n\frac{\overline{H}(\boldsymbol{Y})}{\underline{H}(\boldsymbol{X})}$, then

$$\Pr(T > m) \to 0 \qquad (n \to \infty)$$

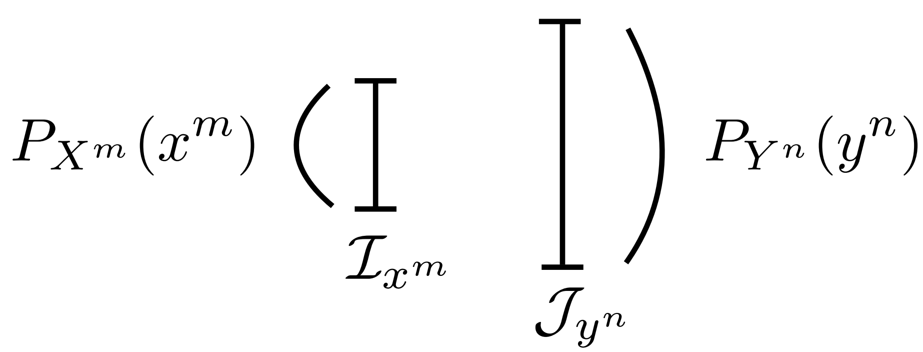
Proof Sketch

$$\Pr(T > m) \le P_{X^m}(\mathcal{S}_m^c(\lambda)) + P_{Y^n}(\mathcal{T}_n^c(\tau)) + 2^{-\lambda+\tau+1}$$

$$\mathcal{S}_m(\lambda) := \left\{ x^m \in \mathcal{X}^m : \log \frac{1}{P_{X^m}(x^m)} \ge \lambda \right\}$$

$$\mathcal{T}_n(\tau) := \left\{ y^n \in \mathcal{Y}^n : \log \frac{1}{P_{Y^n}(y^n)} \le \tau \right\}$$

When x^m is observed, the interval algorithm stops iff.



for some $y^n \in \mathcal{Y}^n$

small $P_{X^m}(x^m)$ and large $P_{Y^n}(y^n)$ are favorable; $\mathcal{S}_m^c(\lambda)$ and $\mathcal{T}_n^c(\tau)$ are handled as exceptions.

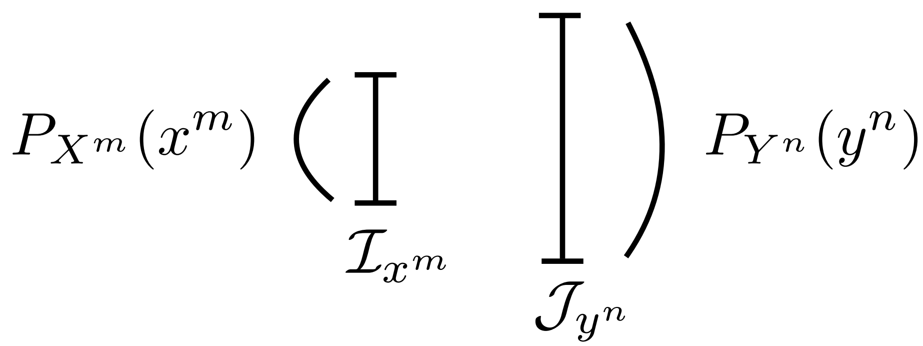
Proof Sketch

$$\Pr(T > m) \le P_{X^m}(\mathcal{S}_m^c(\lambda)) + P_{Y^n}(\mathcal{T}_n^c(\tau)) + 2^{-\lambda+\tau+1}$$

$$\mathcal{S}_m(\lambda) := \left\{ x^m \in \mathcal{X}^m : \log \frac{1}{P_{X^m}(x^m)} \ge \lambda \right\}$$

$$\mathcal{T}_n(\tau) := \left\{ y^n \in \mathcal{Y}^n : \log \frac{1}{P_{Y^n}(y^n)} \le \tau \right\}$$

When x^m is observed, the interval algorithm stops iff.

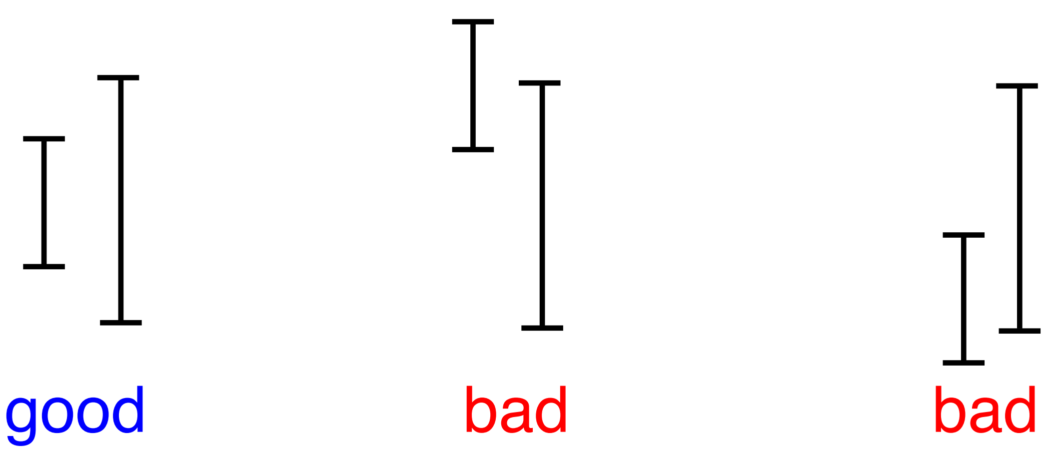


for some $y^n \in \mathcal{Y}^n$

small $P_{X^m}(x^m)$ and large $P_{Y^n}(y^n)$ are favorable; $\mathcal{S}_m^c(\lambda)$ and $\mathcal{T}_n^c(\tau)$ are handled as exceptions.

$$x^m \in \mathcal{S}_m(\lambda) \Longrightarrow P_{X^m}(x^m) \le 2^{-\lambda}$$

$$y^n \in \mathcal{T}_n(\tau) \Longrightarrow P_{Y^n}(y^n) \ge 2^{-\tau}$$



A Converse Bound

Theorem (W.-Han '19)

For any RNG algorithm, the overflow probability of the stopping time satisfies

$$\Pr(T > m) \geq P_{Y^n}(\mathcal{T}_n^c(\tau)) - P_{X^m}(\mathcal{S}_m(\lambda)) - 2^{-\tau+\lambda}$$

$$= P_{X^m}(\mathcal{S}_m^c(\lambda)) - P_{Y^n}(\mathcal{T}_n(\tau)) - 2^{-\tau+\lambda}$$

Set $\lambda \simeq m\underline{H}(\boldsymbol{X})$, $\tau \simeq n\underline{H}(\boldsymbol{Y})$, and $m = nR$. Then, $\Pr(T > m) \to 0$ only if

$$R \geq \frac{\underline{H}(\boldsymbol{Y})}{\underline{H}(\boldsymbol{X})}$$

Similarly,

$$R \geq \frac{\overline{H}(\boldsymbol{Y})}{\overline{H}(\boldsymbol{X})}$$

Optimality of Interval Algorithm

If either the coin or the target process has one point spectrum, i.e.,

$$\underline{H}(\boldsymbol{X}) = \overline{H}(\boldsymbol{X}) = H(\boldsymbol{X}) \qquad \text{or} \qquad \underline{H}(\boldsymbol{Y}) = \overline{H}(\boldsymbol{Y}) = H(\boldsymbol{Y})$$

then $\Pr(T > nR) \to 0$ iff.

$$R \geq \frac{\overline{H}(\boldsymbol{Y})}{H(\boldsymbol{X})} \qquad \text{or} \qquad R \geq \frac{H(\boldsymbol{Y})}{\underline{H}(\boldsymbol{X})}$$

Furthermore, it is attained by the interval algorithm.

Average Stopping Time

By using

$$\mathbb{E}[T] = \int_0^\infty \Pr(T > z) dz$$

$$\lesssim \int_0^\infty \Pr\left(\frac{1}{\underline{H}(\boldsymbol{X})} \log \frac{1}{P_{Y^n}(Y^n)} > z \right) dz$$

$$= \frac{H(Y^n)}{\underline{H}(\boldsymbol{X})}$$

<u>Corollary</u> (W.-Han '19)

Under some regularity condition, the interval algorithm satisfies

$$\limsup_{n\to\infty} \frac{1}{n}\mathbb{E}[T] \le \frac{H(\boldsymbol{Y})}{\underline{H}(\boldsymbol{X})}$$

Any RNG algorithm satisfies

$$\limsup_{n\to\infty} \frac{1}{n}\mathbb{E}[T] \ge \frac{H(\boldsymbol{Y})}{\overline{H}(\boldsymbol{X})}$$

If the coin process has one point spectrum, the interval algorithm is optimal.

Example

Let $\boldsymbol{X} = \{X^m\}_{m=1}^{\infty}$ be a Markov chain induced by irreducible $W(x|x')$

For the stationary distribution π, let

$$H^W(X) = \sum_{x,x'} \pi(x')W(x|x') \log \frac{1}{W(x|x')}$$

Then, $\underline{H}(\boldsymbol{X}) = \overline{H}(\boldsymbol{X}) = H(\boldsymbol{X}) = H^W(X)$

Let $\boldsymbol{Y} = \{Y^n\}_{n=1}^{\infty}$ be a Markov chain induced by reducible $V(y|y')$,

but assume that there is no transient class; then

$$V = \bigoplus_{\xi=1}^{r} V_\xi \qquad\qquad V_\xi \quad \text{is irreducible}$$

For the weight $w(\xi)$ induced by an initial distribution,

$$\overline{H}(\boldsymbol{Y}) = \max\left\{H^{V_\xi}(Y) : 1 \le \xi \le r, w(\xi) > 0\right\}$$

$$H(\boldsymbol{Y}) = \sum_{\xi=1}^{r} w(\xi) H^{V_\xi}(Y)$$

Example (cont'd)

The interval algorithm satisfies

$$\limsup_{n\to\infty} \frac{1}{n}\mathbb{E}[T] \leq \frac{1}{H^W(X)} \sum_{\xi=1}^{r} H^{V_\xi}(Y)$$

and $\Pr(T > nR) \to 0$ if

$$R \geq \frac{1}{H^W(X)} \max\left\{H^{V_\xi}(Y) : 1 \leq \xi \leq r, w(\xi) > 0\right\}$$

These performances are optimal among any RNG algorithms.

Isomorphism Problem

In the ergodic theory, a basic problem is to show if a (two-sided) random process

$\boldsymbol{X} = (\ldots, X_{-1}, X_0, X_1, \ldots)$ is isomorphic to

another random process $\boldsymbol{Y} = (\ldots, Y_{-1}, Y_0, Y_1, \ldots)$.

$S : \mathcal{X}^{\mathbb{Z}} \to \mathcal{X}^{\mathbb{Z}}$ shift operator

For $\boldsymbol{x} = (\cdots, x_{-1}, x_0, x_1, \ldots)$, $(S\boldsymbol{x})_i = x_{i+1}$

<u>Definition</u>

A measurable map ϕ from $\mathcal{X}^{\mathbb{Z}}$ to $\mathcal{Y}^{\mathbb{Z}}$ is termed homomorphism from $(\mathcal{X}^{\mathbb{Z}}, \mathcal{B}_{\mathcal{X}}, \mu, S)$ to $(\mathcal{Y}^{\mathbb{Z}}, \mathcal{B}_{\mathcal{Y}}, \mu, S)$ if $\nu(A) = \mu(\phi^{-1}(A))$ for $A \in \mathcal{B}_{\mathcal{Y}}$ and $\phi(S\boldsymbol{x}) = S\phi(\boldsymbol{x})$ for μ-a.e. $\boldsymbol{x}$.

Furthermore, if ϕ is invertible for μ-a.e. $\boldsymbol{x}$, then it is termed isomorphism.

Isomorphism Problem (cont'd)

Consider i.i.d. processes $\boldsymbol{X}$ and $\boldsymbol{Y}$, which are termed Bernoulli shifts in ergodic theory.

Theorem (Ornstein '70)

For Bernoulli shifts, isomorphism exists iff. $H(X) = H(Y)$.

In 1970s, the isomorphism problem was actively studied in IT community from the viewpoint of source coding (eg. Gray, Neuhoff).

The connection between the isomorphism problem and the RNG problem seems to be not well understood; but there are some work (eg. Harvey-Holroyd-Peres-Romik '07).

Thank you very much.