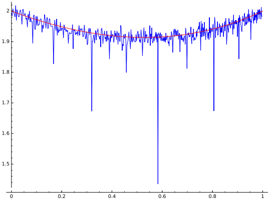
Random Number Generation and Fitting Interval Partitions

Pablo Rotondo[1]



AleaEnAmSud,
Caen, 6 June, 2017.

[1] IRIF, Paris 7 Diderot. Universidad de la República, Uruguay. GREYC, associate

Random Variable Simulation

Objective: given a perfect source of *independent fair bits*

$$\mathcal{X} = X_1, X_2, \ldots$$

simulate a random variable Y with a *prescribed distribution*.

– Generating random bits $\longrightarrow$ costly.
– What if X_i were not random bits? $\longrightarrow$ other distributions.

Classical algorithm in probability courses
– given uniform $U \in [0, 1]$ and a continuous distribution function F consider the so called inverse method

$$Y := F^{-1}(U)\,.$$

– in our context we may consider

$$U := (0.X_1 X_2 \ldots)_2 \ .$$

Interval algorithm: intro

Discrete random variable $Y \in \mathbb{Z}_{>0}$ with distribution vector $p_1, p_2, \ldots$

The inverse method gives intervals

$$I_i(\mathbf{p}) := \left[\sum_{j<i} p_j, \sum_{j\le i} p_j \right),$$

and defines

$$Y = i \iff U \in I_i(\mathbf{p}).$$

Question

How many fair bits $X_1, X_2, \ldots$ do we need to determine Y?

Interval algorithm: intro

Discrete random variable $Y \in \mathbb{Z}_{>0}$ with distribution vector $p_1, p_2, \ldots$

The inverse method gives intervals

$$I_i(\mathbf{p}) := \left[\sum_{j<i} p_j, \sum_{j \leq i} p_j \right),$$

and defines

$$Y = i \iff U \in I_i(\mathbf{p}).$$

Question

How many fair bits $X_1, X_2, \ldots$ do we need to determine Y?

More precisely, given $u \in [0,1]$ in binary $u = (0.x_1 x_2 \ldots)_2$ we define

$$k_{\mathbf{p}}(u) := \inf \left\{ k \geq 0 : \exists i \text{ s.t. } \left(0.x_1 \ldots x_k, 0.x_1 \ldots x_k + 2^{-k}\right) \subset I_i(\mathbf{p}) \right\}$$

Interval algorithm: intro

Discrete random variable $Y \in \mathbb{Z}_{>0}$ with distribution vector $p_1, p_2, \ldots$

The inverse method gives intervals

$$I_i(\mathbf{p}) := \left[\sum_{j<i} p_j, \sum_{j \leq i} p_j \right),$$

and defines

$$Y = i \iff U \in I_i(\mathbf{p}).$$

Question

How many fair bits $X_1, X_2, \ldots$ do we need to determine Y?

More precisely, given $u \in [0,1]$ in binary $u = (0.x_1 x_2 \ldots)_2$ we define

$$k_{\mathbf{p}}(u) := \inf \left\{ k \geq 0 : \exists i \text{ s.t. } \left(0.x_1 \ldots x_k, 0.x_1 \ldots x_k + 2^{-k}\right) \subset I_i(\mathbf{p}) \right\}$$

What is

$$\mathbb{E}[k_{\mathbf{p}}(U)] \quad ?$$

Lower bound: the entropy

Theorem

The expected number of bits is bounded from below by the entropy of $\mathbf{p}$

$$\mathbb{E}[k_{\mathbf{p}}(U)] \geq H(Y)\,,$$

where $H(Y) := \sum_i p_i \log_2(1/p_i)$.

Lower bound: the entropy

Theorem

The expected number of bits is bounded from below by the entropy of $\mathbf{p}$

$$\mathbb{E}[k_{\mathbf{p}}(U)] \geq H(Y)\,,$$

where $H(Y) := \sum_i p_i \log_2(1/p_i)$.

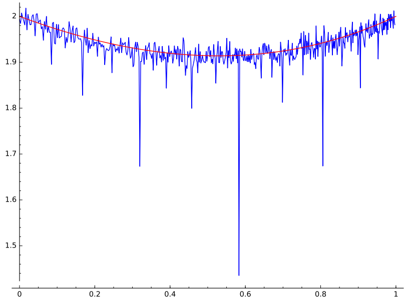
Proof.

The code

$$\begin{aligned}\mathcal{C} := \Big\{ & x_1 \dots x_k \in \{0,1\}^+ : \\ & \exists i \text{ s.t. } \big(0.x_1 \dots x_k, 0.x_1 \dots x_k + 2^{-k}\big) \subset I_i(\mathbf{p})\,, \\ & \nexists j \big(0.x_1 \dots x_{k-1}, 0.x_1 \dots x_{k-1} + 2^{-(k-1)}\big) \subset I_j \Big\}\end{aligned}$$

is prefix free and determines Y. □

The distribution $\mathbf{p}_N := (1/N, 1/N \ldots, 1/N)$



Theorem

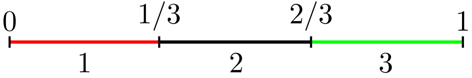
The redundancy $\mathbb{E}[k_{\mathbf{p}_N}(U)] - H(Y)$ equals

$$R(x) = 2^x - x + 1 - \frac{2^{\nu(N)} - 1}{N-1}\, 2^x - \log_2\left(1 + \tfrac{1}{N-1}\right),$$

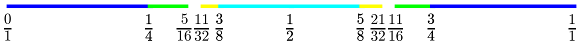
where $x = \{\log_2(N-1)\}$, $\{\cdot\}$ denotes the fractional part and $\nu(N)$ is the greatest t such that 2^t divides N.

Example: fair 3-sided dice

First $\mathbf{p}_3 = (1/3, 1/3, 1/3)$ divides the interval $[0, 1]$ as follows



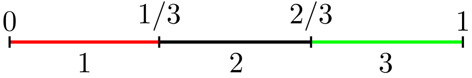
while the subdivision procedure, a binary search for U, gives



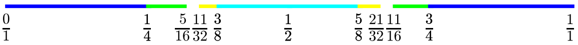
where we remark that the number of bits can be deduced from the denominators.

Example: fair 3-sided dice

First $\mathbf{p}_3 = (1/3, 1/3, 1/3)$ divides the interval $[0,1]$ as follows



while the subdivision procedure, a binary search for U, gives



where we remark that the number of bits can be deduced from the denominators.
In this case we have

$$\mathbb{E}[k_{\mathbf{p}_3}(U)] = 3 \doteq \log_2(3) + 1.41503\ldots .$$

Generic Distribution

For an arbitrary probability vector $\mathbf{p} = (p_1, p_2, \ldots)$

Theorem

The redundancy $\mathbb{E}[k_{\mathbf{p}}(U)] - H(Y)$ is at most 2, i.e.

$$H(Y) \le \mathbb{E}[k_{\mathbf{p}}(U)] \le H(Y) + 2\,.$$

Furthermore, the $+2$ is tight by our example $\mathbf{p}_N$.

Generic Distribution

For an arbitrary probability vector $\mathbf{p} = (p_1, p_2, \ldots)$

Theorem

The redundancy $\mathbb{E}[k_{\mathbf{p}}(U)] - H(Y)$ is at most 2, i.e.

$$H(Y) \leq \mathbb{E}[k_{\mathbf{p}}(U)] \leq H(Y) + 2\,.$$

Furthermore, the $+2$ is tight by our example $\mathbf{p}_N$.

Knuth-Yao proved that the optimal algorithm satisfies these bounds.

- Algorithm seen as binary tree.
- Optimal algorithm obtained by decomposing each

$$p_i = (0.p_1^{(i)} p_2^{(i)} \ldots)_2$$

in binary and assigning a leaf of probability $p_j^{(i)}$ for each (i, j).

Example: continued fractions

Procedures gives way to, given the binary representation of U, decide to which interval among the partition

$$[0, p_1), [p_1, p_1 + p_2), [p_1 + p_2, p_1 + p_2 + p_3), \ldots$$

it belongs to. This can be applied to other partitions.

Example: continued fractions

Procedures gives way to, given the binary representation of U, decide to which interval among the partition

$$[0, p_1), [p_1, p_1 + p_2), [p_1 + p_2, p_1 + p_2 + p_3), \ldots$$

it belongs to. This can be applied to other partitions.

For example, if we consider the convergents $(q_k)_k$ of U, the set

$$I_k(a, b) := \{U \in [0, 1] : (q_{k-1}(U), q_k(U)) = (a, b)\}$$

is an interval of length $\frac{1}{b(a+b)}$ when $\gcd(a, b) = 1$ and $a \leq b$.

Example: continued fractions

Procedures gives way to, given the binary representation of U, decide to which interval among the partition

$$[0, p_1), [p_1, p_1 + p_2), [p_1 + p_2, p_1 + p_2 + p_3), \ldots$$

it belongs to. This can be applied to other partitions.

For example, if we consider the convergents $(q_k)_k$ of U, the set

$$I_k(a, b) := \{U \in [0, 1] : (q_{k-1}(U), q_k(U)) = (a, b)\}$$

is an interval of length $\frac{1}{b(a+b)}$ when $\gcd(a, b) = 1$ and $a \leq b$.

A possible partition (negligible intersection) is given by fixing k

$$\mathcal{I}_k := \{I_k(a, b) : \gcd(a, b) = 1,\ 1 \leq a \leq b\},$$

which determines the value of $(q_{k-1}(U), q_k(U))$.

Example: continued fractions

A different partition, relating to our ANALCO paper is:

- Fix $n \in \mathbb{Z}_{>0}$, and consider

$$\mathbb{I}_n := \left\{(a,b) \in \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} : \gcd(a,b) = 1,\, a \leq n < b\right\}.$$

Example: continued fractions

A different partition, relating to our ANALCO paper is:

- Fix $n \in \mathbb{Z}_{>0}$, and consider

$$\mathbb{I}_n := \left\{(a,b) \in \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} : \gcd(a,b) = 1,\ a \leq n < b\right\}.$$

- Consider $k := k(U,n)$ such that $q_{k-1}(U) \leq n < q_k(U)$ and define

$$I_{a,b} := \left\{U \in [0,1] : \left(q_{k(U,n)-1}, q_{k(U,n)}\right) = (a,b)\right\},$$

where $(a,b) \in \mathbb{I}_n$, again an interval, forming a partition.

Example: continued fractions

A different partition, relating to our ANALCO paper is:

- Fix $n \in \mathbb{Z}_{>0}$, and consider

$$\mathbb{I}_n := \left\{(a,b) \in \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} : \gcd(a,b) = 1,\, a \le n < b\right\}.$$

- Consider $k := k(U,n)$ such that $q_{k-1}(U) \le n < q_k(U)$ and define

$$I_{a,b} := \left\{U \in [0,1] : \left(q_{k(U,n)-1}, q_{k(U,n)}\right) = (a,b)\right\},$$

where $(a,b) \in \mathbb{I}_n$, again an interval, forming a partition.

- Then the number of bits needed to determine

$$\left(q_{k(U,n)-1}, q_{k(U,n)}\right),$$

is roughly

$$\begin{aligned} H\left(q_{k(U,n)-1}, q_{k(U,n)}\right) &= 2\log_2 n + 1 \\ &\quad - \frac{12}{\pi^2} \iint_{0<x<1\le y} \frac{\log_2(y(x+y))}{y(x+y)} dx dy + o(1) \\ &= 2\log_2 n - 2.4263\ldots + o(1) \end{aligned}$$

Interval Algorithm

- Generalization of the previous procedure for fair bits

$$\mathcal{X} = (X_1, X_2, \ldots) .$$

- Each X_i takes values on $[M]$ with vector $(q_1, q_2, \ldots, q_M)$.

Interval Algorithm

- Generalization of the previous procedure for fair bits

$$\mathcal{X} = (X_1, X_2, \ldots).$$

- Each X_i takes values on $[M]$ with vector $(q_1, q_2, \ldots, q_M)$.

Now the procedure goes as follows:

⊛ Let $K_k(\mathbf{q}) = [A_k, B_k)$ be our **working interval** after processing $X_1, \ldots, X_k$.

⊛ Partition $K_k(\mathbf{q})$ into intervals

$$K_{k,j}(\mathbf{q}) := [A_k + Q_{j-1}(B_k - A_k), A_k + Q_j(B_k - A_k)),$$

according to $Q_j(\mathbf{q}) := \sum_{i \le j} q_i$.

⊛ Suppose $X_{k+1}(\mathbf{q}) = j$, then set $K_{k+1}(\mathbf{q}) := K_{k,j}(\mathbf{q})$.

Interval $K_k(\mathbf{q})$ corresponds to what before was

$$[0.x_1 \dots x_k, 0.x_1 \dots x_k + 2^{-k})\,.$$

We continue until

$$k_{\mathbf{q},\mathbf{p}} := \inf \left\{ k \geq 0 : \exists i \text{ s.t. } K_k(\mathbf{q}) \subset I_i(\mathbf{p}) \right\},$$

in which case we return $Y = i$ if $K_k(\mathbf{q}) \subset I_i(\mathbf{p})$

Interval $K_k(\mathbf{q})$ corresponds to what before was

$$[0.x_1 \dots x_k, 0.x_1 \dots x_k + 2^{-k}) \, .$$

We continue until

$$k_{\mathbf{q},\mathbf{p}} := \inf \left\{ k \geq 0 : \exists i \text{ s.t. } K_k(\mathbf{q}) \subset I_i(\mathbf{p}) \right\} ,$$

in which case we return $Y = i$ if $K_k(\mathbf{q}) \subset I_i(\mathbf{p})$

Theorem (Lower bound)

The cost of simulating the random variable Y having prob. vector $\mathbf{p}$ by using the Interval Algorithm with an M-valued "coin flips" according to the prob. vector $\mathbf{q}$ is bounded from below by

$$\frac{H(\mathbf{p})}{H(\mathbf{q})} \leq \mathbb{E}[k_{\mathbf{q},\mathbf{p}}] \, .$$

Interval Algorithm: efficiency

Theorem (Han,Hoshi 95)

For any probability vectors $\mathbf{p} = (p_1, \ldots, p_n)$ *and* $\mathbf{q} = (q_1, \ldots, q_m)$*, the expected number of coin tosses in the interval algorithm is upper-bounded*

$$\mathbb{E}[k_{\mathbf{q},\mathbf{p}}] \leq \frac{H(\mathbf{p})}{H(\mathbf{q})} + \frac{\log 2(M-1)}{H(\mathbf{q})} + \frac{h(q_{\max})}{(1-q_{\max})H(\mathbf{q})} \, .$$

Proof.

Whiteboard (or blackboard). □

Generalization to random processes

We want to simulate a random process

$$\mathcal{Y} = (Y_1, Y_2, Y_3, \ldots),$$

rather than a single Y with a prescribed distribution.

The question now is

what is the asymptotic cost of producing $\mathcal{Y}_n = (Y_1, Y_2, \ldots, Y_n)$?

Generalization to random processes

We want to simulate a random process

$$\mathcal{Y} = (Y_1, Y_2, Y_3, \ldots),$$

rather than a single Y with a prescribed distribution.

The question now is

what is the asymptotic cost of producing $\mathcal{Y}_n = (Y_1, Y_2, \ldots, Y_n)$?

Remark

If the "target" source $Y_1, Y_2, \ldots$ is stationary, $\mathbf{p}^n$ denotes the vector of $(Y_{j+1}, \ldots, Y_{j+n})$ for $j \geq 0$ and $k_n := k_{\mathbf{q},\mathbf{p}^n}$:

$$\lim_{n\to\infty} \frac{\mathbb{E}[k_n]}{n} = \frac{H(\mathcal{Y})}{H(\mathcal{X})}$$

Generalization to random processes

We want to simulate a random process

$$\mathcal{Y} = (Y_1, Y_2, Y_3, \ldots)\,,$$

rather than a single Y with a prescribed distribution.

The question now is

what is the asymptotic cost of producing $\mathcal{Y}_n = (Y_1, Y_2, \ldots, Y_n)$?

Remark

If the "target" source $Y_1, Y_2, \ldots$ is stationary, $\mathbf{p}^n$ denotes the vector of $(Y_{j+1}, \ldots, Y_{j+n})$ for $j \geq 0$ and $k_n := k_{\mathbf{q},\mathbf{p}^n}$:

$$\lim_{n\to\infty} \frac{\mathbb{E}[k_n]}{n} = \frac{H(\mathcal{Y})}{H(\mathcal{X})}$$

Proof.

By independence $H(\mathcal{X}) = H(\mathbf{q})$, while $H(\mathcal{Y}_n)/n \to H(\mathcal{Y})$. $\square$

Generalization

- We may imagine now that the variables $X_1, X_2, \ldots$ are not necessarily independent or identically distributed
  $\rightarrow$ natural if produced by a dynamical system (e.g. Euclid).
- The question can be framed more purely in terms of sequences of interval partitions.

Generalization

- We may imagine now that the variables $X_1, X_2, \ldots$ are not necessarily independent or identically distributed
  $\rightarrow$ natural if produced by a dynamical system (e.g. Euclid).
- The question can be framed more purely in terms of sequences of interval partitions.

Definition (Interval partition)

An interval partition is a finite or countable partition of $[0, 1]$ into intervals. If $\mathcal{P}$ is an interval partition and $x \in [0, 1]$, we let $\mathcal{P}(x)$ denote the interval $I \in \mathcal{P}$ such that $x \in I$.

Generalization

- We may imagine now that the variables $X_1, X_2, \ldots$ are not necessarily independent or identically distributed
  $\rightarrow$ natural if produced by a dynamical system (e.g. Euclid).
- The question can be framed more purely in terms of sequences of interval partitions.

Definition (Interval partition)

An interval partition is a finite or countable partition of $[0,1]$ into intervals. If $\mathcal{P}$ is an interval partition and $x \in [0,1]$, we let $\mathcal{P}(x)$ denote the interval $I \in \mathcal{P}$ such that $x \in I$.

Now our problem is rephrased in terms of

$$k_{\mathcal{Q},\mathcal{P}}(n,x) = \inf\left\{k \geq 0 : \mathcal{Q}_k(x) \subset \mathcal{P}_n(x)\right\};$$

Generalization

- We may imagine now that the variables $X_1, X_2, \dots$ are not necessarily independent or identically distributed
  $\rightarrow$ natural if produced by a dynamical system (e.g. Euclid).
- The question can be framed more purely in terms of sequences of interval partitions.

Definition (Interval partition)

An interval partition is a finite or countable partition of $[0,1]$ into intervals. If $\mathcal{P}$ is an interval partition and $x \in [0,1]$, we let $\mathcal{P}(x)$ denote the interval $I \in \mathcal{P}$ such that $x \in I$.

Now our problem is rephrased in terms of

$$k_{\mathcal{Q},\mathcal{P}}(n,x) = \inf\left\{k \geq 0 : \mathcal{Q}_k(x) \subset \mathcal{P}_n(x)\right\};$$

compare it with $k_{\mathbf{q},\mathbf{p}^n}$:

- $\mathcal{P}_n$ corresponds to the partition according to $(Y_1, \dots, Y_n)$.
- $\mathcal{Q}_k$ corresponds to the partition according to $(X_1, \dots, X_k)$.

Entropy of an interval partition

Definition (Entropy)

Let $\mathcal{P} := (\mathcal{P}_n)_{n=1}^{\infty}$ be a sequence of interval partitions. We say that $\mathcal{P}$ has *entropy* $c \geq 0$ with respect to a measure λ if

$$\lim_{n\to\infty} -\frac{1}{n} \log \lambda(\mathcal{P}_n(x)) = c\,, \quad \lambda - \ a.e.$$

Entropy of an interval partition

Definition (Entropy)

Let $\mathcal{P} := (\mathcal{P}_n)_{n=1}^{\infty}$ be a sequence of interval partitions. We say that $\mathcal{P}$ has *entropy* $c \geq 0$ with respect to a measure λ if

$$\lim_{n\to\infty} -\frac{1}{n} \log \lambda(\mathcal{P}_n(x)) = c\,, \quad \lambda - \ a.e.$$

Let λ be the Lebesgue measure:

- For $\mathcal{Q}_k = \left\{ \left[\frac{i}{2^k}, \frac{i+1}{2^k}\right) : i = 0, \ldots, 2^k - 1 \right\}$ we get
$$-\frac{1}{k} \log \lambda(\mathcal{Q}_k(x)) = \log 2\,, \forall k \geq 1\,.$$

Entropy of an interval partition

Definition (Entropy)

Let $\mathcal{P} := (\mathcal{P}_n)_{n=1}^{\infty}$ be a sequence of interval partitions. We say that $\mathcal{P}$ has *entropy* $c \geq 0$ with respect to a measure λ if

$$\lim_{n\to\infty} -\frac{1}{n} \log \lambda(\mathcal{P}_n(x)) = c\,, \quad \lambda - a.e.$$

Let λ be the Lebesgue measure:

- For $\mathcal{Q}_k = \left\{ \left[\frac{i}{2^k}, \frac{i+1}{2^k}\right) : i = 0, \ldots, 2^k - 1 \right\}$ we get

$$-\frac{1}{k} \log \lambda(\mathcal{Q}_k(x)) = \log 2\,, \forall k \geq 1\,.$$

- For $\mathcal{P}_n = \left\{ I_n(a,b) : \gcd(a,b) = 1,\ 1 \leq a \leq b \right\}$, where

$$I_n(a,b) = \{x \in [0,1] : (q_{n-1}(x), q_n(x)) = (a,b)\}\,,$$

we get (blackboard explanation)

$$\lim_{n\to\infty} -\frac{1}{n} \log \lambda(\mathcal{P}_n(x)) = \frac{\pi^2}{6\log 2}\,, \quad \lambda - a.e.$$

Asymptotic Cost

Theorem (Dajani, Fieldsteel, 2001)

Let $\mathcal{P} := \{\mathcal{P}_n\}_{n=1}^{\infty}$ *and* $\mathcal{Q} := \{\mathcal{Q}_n\}_{n=1}^{\infty}$ *be sequences of interval partitions, and let* λ *be a Borel probability measure on* $[0,1)$*.*

Assume $\mathcal{P}$ *and* $\mathcal{Q}$ *have entropies* $H(\mathcal{P})$ *and* $H(\mathcal{Q})$ *respectively with respect to* λ*, then*

$$\lim_{n\to\infty} \frac{1}{n} k_{\mathcal{Q},\mathcal{P}}(n,x) = \frac{H(\mathcal{P})}{H(\mathcal{Q})}$$

for λ*-a.e.* x*.*

Asymptotic Cost

Theorem (Dajani, Fieldsteel, 2001)

Let $\mathcal{P} := \{\mathcal{P}_n\}_{n=1}^{\infty}$ *and* $\mathcal{Q} := \{\mathcal{Q}_n\}_{n=1}^{\infty}$ *be sequences of interval partitions, and let* λ *be a Borel probability measure on* $[0, 1)$.

Assume $\mathcal{P}$ *and* $\mathcal{Q}$ *have entropies* $H(\mathcal{P})$ *and* $H(\mathcal{Q})$ *respectively with respect to* λ, *then*

$$\lim_{n\to\infty} \frac{1}{n} k_{\mathcal{Q},\mathcal{P}}(n, x) = \frac{H(\mathcal{P})}{H(\mathcal{Q})}$$

for λ*-a.e.* x.

Example: the number of digits required to determine $(q_{n-1}(x), q_n(x))$ from the base 10 expansion of x behaves like $\frac{\pi^2}{6\log 2 \log 10} n$ a.e. x, a result previously proved by Lochs.

Good partition sequences

Theorem (Shannon,McMillan,Breiman)

Let T be an ergodic measure preserving transformation on a probability space $(\Omega, \mathcal{B}, \mu)$ and let P be a finite or countable generating partition for T for which $H_\mu(P) < \infty$. Then for μ-a.e. x,

$$\lim_{n\to\infty} -\frac{\log \mu\left(P_n(x)\right)}{n} = h_\mu(T)\,.$$

Here $H_\mu(P)$ denotes the entropy of the partition P, $h_\mu(T)$ the entropy of T and $P_n(x)$ denotes the element of the partition $\bigvee_{i=0}^{n-1} T^{-i}P$ containing x.

Good partition sequences

Theorem (Shannon,McMillan,Breiman)

Let T be an ergodic measure preserving transformation on a probability space $(\Omega, \mathcal{B}, \mu)$ and let P be a finite or countable generating partition for T for which $H_\mu(P) < \infty$. Then for μ-a.e. x,

$$\lim_{n\to\infty} -\frac{\log \mu\left(P_n(x)\right)}{n} = h_\mu(T)\,.$$

Here $H_\mu(P)$ denotes the entropy of the partition P, $h_\mu(T)$ the entropy of T and $P_n(x)$ denotes the element of the partition $\bigvee_{i=0}^{n-1} T^{-i}P$ containing x.

We recall that

$$h_\mu(T) = \sup\{h_\mu(T, \mathcal{A}) : \mathcal{A} \text{ countable partition of } X\}\,,$$

and

$$h_\mu(T, \mathcal{A}) = \lim_{n\to\infty} \frac{1}{n} H\left(\mathcal{A}(U), \mathcal{A}(TU), \ldots, \mathcal{A}\left(T^{n-1}U\right)\right)\,,$$

where U is distributed according to μ.

Concluding remarks

There are other interesting results/ideas not mentioned in this talk,

- The optimal algorithm for discrete uniform generation from coin flips has a fairly simple implementation, see the note by Jérémie Lumbroso arXiv:1304.1916v1.

Concluding remarks

There are other interesting results/ideas not mentioned in this talk,

- The optimal algorithm for discrete uniform generation from coin flips has a fairly simple implementation, see the note by Jérémie Lumbroso arXiv:1304.1916v1.
- The generalisation of Bosma-Dajani-Kraaikamp of the cost of passing from base 10 to the Continued Fraction Expansion. One considers appropriate maps T and S called "number-theoretic fibered maps" associated with digits and get a limiting result with the quotient of $h(T)$ and $h(S)$.

Concluding remarks

There are other interesting results/ideas not mentioned in this talk,

- The optimal algorithm for discrete uniform generation from coin flips has a fairly simple implementation, see the note by Jérémie Lumbroso arXiv:1304.1916v1.
- The generalisation of Bosma-Dajani-Kraaikamp of the cost of passing from base 10 to the Continued Fraction Expansion. One considers appropriate maps T and S called "number-theoretic fibered maps" associated with digits and get a limiting result with the quotient of $h(T)$ and $h(S)$.
- What about non-discrete random variables? See the original paper by von Neumann, Knuth-Yao and Philippe Duchon.

References

L. Devroye,
Non-Uniform Random Variate Generation,
Springer-Verlag, New York 1986.

T. Cover, J. Thomas
Elements of Information Theory,
Wiley Series in Telecommunications and Signal Processing, Second Edition, 2006 .

D. Knuth and A. Yao,
The complexity of nonuniform random number generation,
Algorithms and Complexity, New Directions and Recent Results, pp. 357–428, 1976.

T. S. Han, and M. Hoshi,
Interval Algorithm for Random Number Generation,
IEEE Transactions on Information Theory, 32 (2), pp. 599-611, March 1997.

K. Dajani, and A. Fieldsteel,
Equipartition of Interval Partitions and an Application to Number Theory,
Proceedings of the American Mathematical Society, vol 129, n. 12, pp. 3453–3460, 2001.