

Stationary Tries and Renewal Theorem

Philippe Robert

INRIA-Rocquencourt

Contents

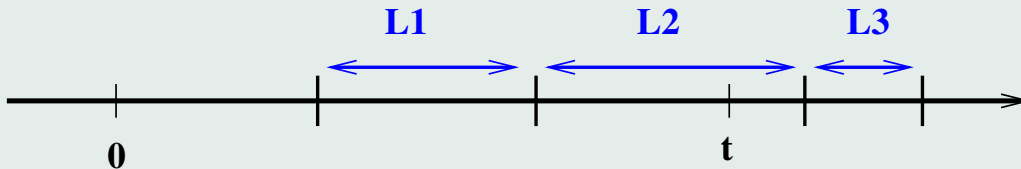
1	Renewal Theorems	2
2	Tree Algorithms	10
3	A Probabilistic Point of View	19
4	Stationary tries	35

1. Renewal Theorems

Some History

Blackwell (1948):

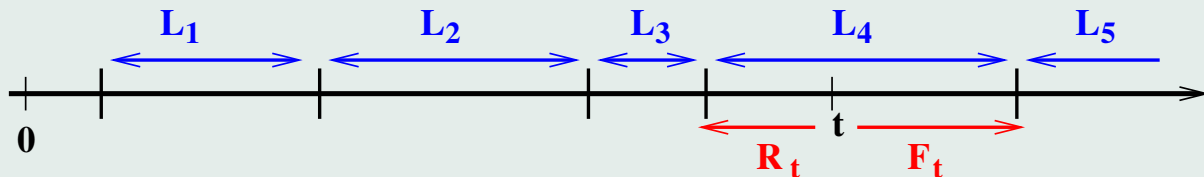
- A light bulb last two years in average.
- How many are necessary for ten years ?



Breiman, Feller, Lindvall, ...

Babillot (Habilitation).

General Framework



(L_i) i.i.d. non-negative random variables.

— $U(a, b)$: average number of points in $[a, b]$,
 U : Renewal measure,

for $h > 0$ $\lim_{t \rightarrow +\infty} U(t, t + h)$?

— Behavior of (R_t, F_t) as $t \rightarrow +\infty$?

Renewal Theorem

Non-Lattice Case: $\forall \delta > 0, \mathbb{P}(L \in \delta\mathbb{N}) < 1$

$$\lim_{t \rightarrow +\infty} U(t, t+h) = \frac{h}{\mathbb{E}(L_1)};$$

$(R_t, F_t) \xrightarrow{\text{dist.}} (R_\infty, F_\infty) :$

$$\mathbb{E}(f(R_\infty, F_\infty)) = \frac{1}{\mathbb{E}(L_1)} \mathbb{E} \left(\int_0^{L_1} f(u, L_1 - u) du \right)$$

F_t has density $\sim x \rightarrow \mathbb{P}(L_1 \geq x) / \mathbb{E}(L_1)$

Proofs

— **Renewal Equation:**

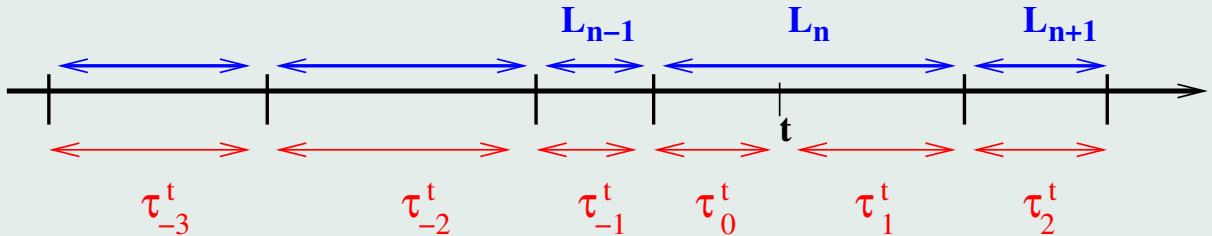
$$U(0, t) = 1 + \int_0^t U(0, t - u) L(du)$$

⇒ **Fourier Analysis (Feller).**

— **Coupling:**

Lindvall, Athreya and Ney, ...

A Point Process Point of View



Renewal Theorem:

$$(\tau_i^t, i \in \mathbb{Z}) \xrightarrow{\text{dist.}} (L_i^*, i \in \mathbb{Z})$$

Stationary renewal process.

1. $(L_i^*, i \leq -1), (L_i^*, i > 1)$, i.i.d. dist. as L_1 ;
2. $(L_0^*, L_1^*) \stackrel{\text{dist.}}{=} (R_\infty, F_\infty)$.

Lattice Case

If $\mathbb{P}(L_1 \in \delta\mathbb{N}) = 1$ and $\mathbb{P}(L_1 = \delta) > 0$:

— $(\tau_i^t, i \in \mathbb{Z})$ does not converge as $t \rightarrow +\infty$

— for $h > 0$,

$$(\tau_i^{h+n\delta}, i \in \mathbb{Z}) \xrightarrow{n \rightarrow +\infty} (L_i^{*h}, i \in \mathbb{Z})$$

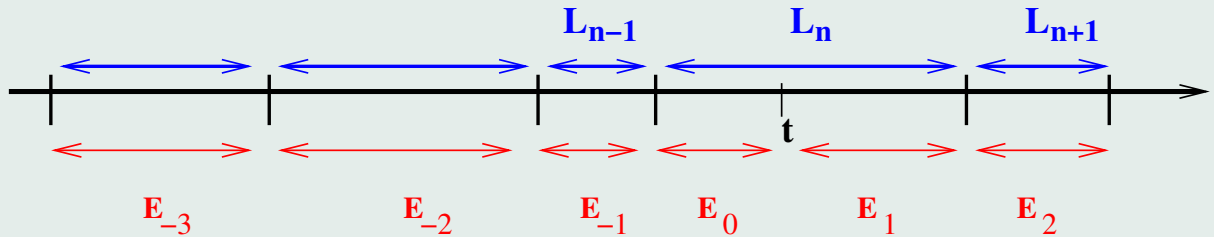
— Periodic Behavior

$$(L_i^{*h}, i \in \mathbb{Z}) \stackrel{\text{dist.}}{=} (L_i^{*(h+\delta)}, i \in \mathbb{Z})$$

The Poisson Process

(L_n) exponential r.v. with parameter **1**.

$$\mathbb{P}(L_1 \geq x) = \exp(-x).$$



(E_n) also exp. with parameter **1**.

Invariant by translation.

Renewal Theorem holds at time $t = 0$.

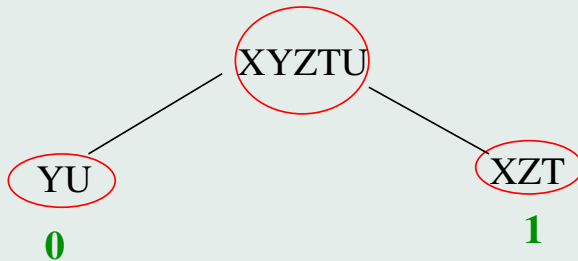
2. Tree Algorithms

XYZTU

— Each item of root node flips a coin X

$X = 0 \Rightarrow$ left node

$X = 1 \Rightarrow$ right node



Each item of root node flips a coin

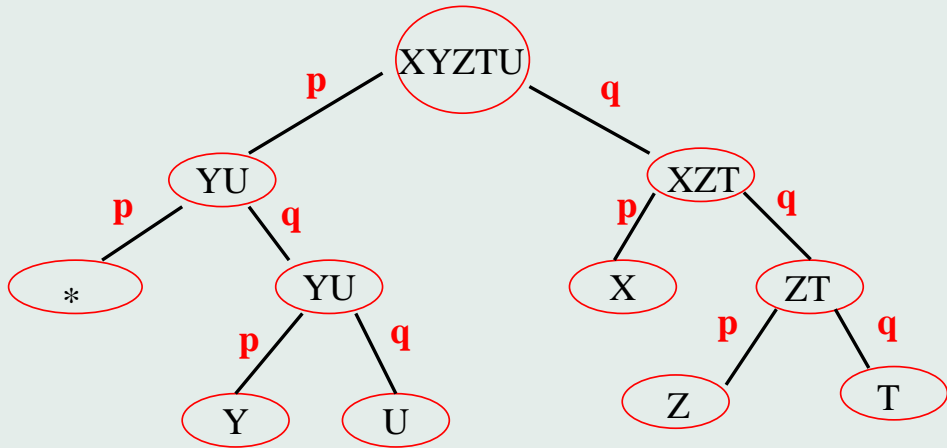
0 \Rightarrow left node

1 \Rightarrow right node

Continue until at most one item per node.

Independent coin tossings.

A Non-Symmetrical Tree Algorithm



$$\mathbb{P}(X = 0) = p, \mathbb{P}(X = 1) = q$$

Applications

- **Communication protocols;**
- **Search and Sort Algorithms;**
- **Leader election problems;**
- **Statistical tests,**
- **...**

Asymptotic behavior

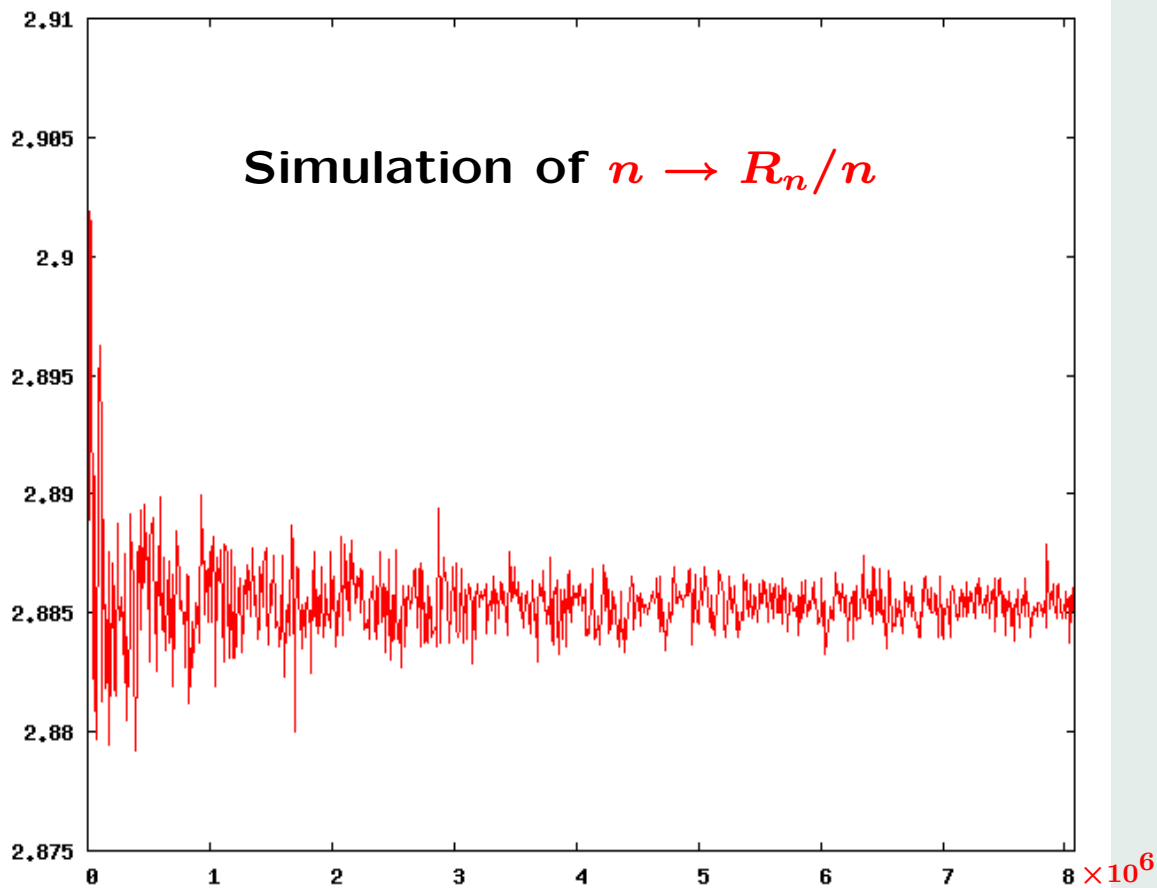
Cost Function. Starting with n items at root
 R_n : nb. of nodes of the tree

$\frac{\mathbb{E}(R_n)}{n}$: Average cost to process 1 item.

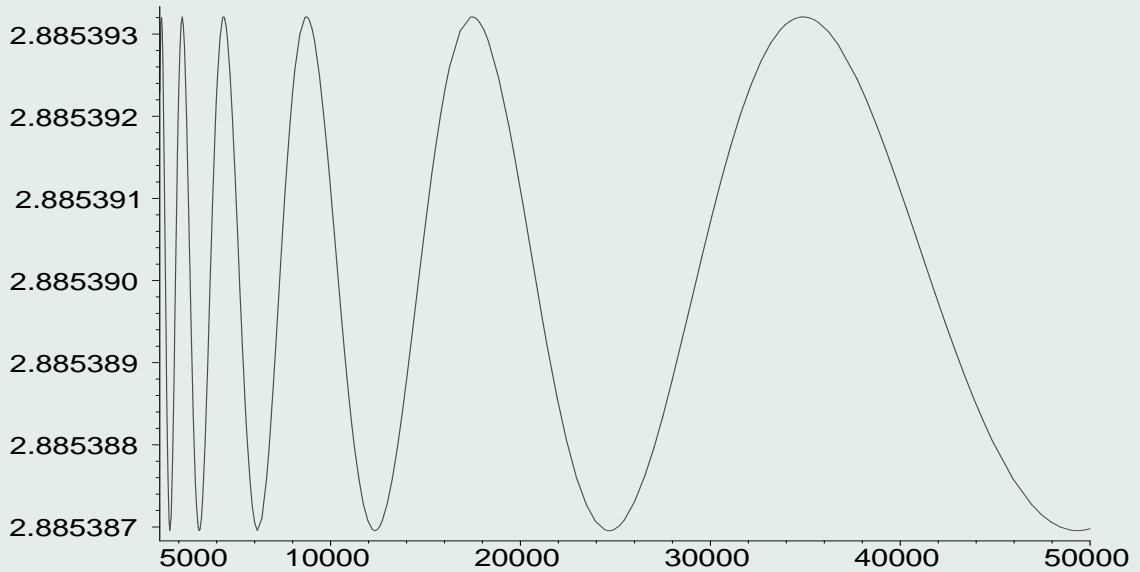
Law of Large Numbers: $\lim_{n \rightarrow +\infty} \frac{\mathbb{E}(R_n)}{n} ?$

the limit does not always exist !

Simulation of $n \rightarrow R_n/n$



The sequence $n \rightarrow \mathbb{E}(R_n)/n$



Asymptotic behavior

$$\left(\frac{\mathbb{E}(R_n)}{n}\right) : \begin{cases} \text{converges} & \text{if } \frac{\log p}{\log q} \notin \mathbb{Q} \\ \text{oscillates} & \text{otherwise} \end{cases}$$

Literature: Complex analysis methods
FLajolet and co-authors.

3. A Probabilistic Point of View

Recurrence Relation

$R_0 = R_1 = 1$. For $n \geq 2$,

$$R_n \stackrel{\text{dist.}}{=} 1 + R_{X_n} + \bar{R}_{n-X_n}$$

with

$$X_n = B_1 + B_2 + \cdots + B_n.$$

$(B_i) \in \{0, 1\}$ i.i.d. Bernoulli parameter p .

(\bar{R}_n) same dist. as (R_n) independent of (R_n)

Equation for the Poisson Transform

$$R_n \stackrel{\text{dist.}}{=} 1 + R_{X_n} + \bar{R}_{n-X_n} - 2 \times 1_{\{n \leq 1\}}$$

If $(N_x, x \geq 0)$ Poisson process with rate 1

$$\mathbb{E}(R_{N_x}) = \mathbb{E}(R_{N_{px}}) + \mathbb{E}(R_{N_{qx}}) + 1 - 2\mathbb{P}(t_2 \geq x)$$

t_2 second point of Poisson process (N_x)

$$\text{If } r(x) = \mathbb{E}(R_{N_x}) - 1$$

$$r(x) = r(px) + r(qx) + 2\mathbb{P}(t_2 < x)$$

Poisson Transform (II)

$$r(x) = r(px) + r(qx) + 2\mathbb{P}(t_2 < x)$$

If $A_1 \in \{p, q\}$ r.v. such that $\mathbb{P}(A_1 = p) = p$

$$r(x) = \mathbb{E} \left(\frac{r(A_1 x)}{A_1} \right) + 2\mathbb{E} (1_{\{t_2 < x\}})$$

Poisson Transform (II)

$$r(x) = r(px) + r(qx) + 2\mathbb{P}(t_2 < x)$$

If $A_1 \in \{p, q\}$ r.v. such that $\mathbb{P}(A_1 = p) = p$

$$\begin{aligned} r(x) &= \mathbb{E} \left(\frac{r(A_1 x)}{A_1} \right) + 2\mathbb{E} \left(\mathbf{1}_{\{t_2 < x\}} \right) \\ &= \mathbb{E} \left(\frac{r(A_2 A_1 x)}{A_2 A_1} \right) + 2\mathbb{E} \left(\frac{1}{A_1} \mathbf{1}_{\{t_2 < x A_1\}} \right) \\ &\quad + 2\mathbb{E} \left(\mathbf{1}_{\{t_2 < x\}} \right) \\ &= 2 \sum_{k=0}^{+\infty} \mathbb{E} \left(\frac{1}{\prod_1^k A_i} \mathbf{1}_{\{t_2 < x \prod_1^k A_i\}} \right) \end{aligned}$$

Poisson Transform (III)

$$\begin{aligned}\mathbb{E}(R_{N_x}) &= 1 + 2 \sum_{k=0}^{+\infty} \mathbb{E} \left(\frac{1}{\prod_1^k A_i} \mathbf{1}_{\{t_2 < x \prod_1^k A_i\}} \right) \\ &= 1 + 2 \sum_{k=0}^{+\infty} \mathbb{E} \left(\sum_{n=2}^{+\infty} \frac{1}{\prod_1^k A_i} \mathbf{1}_{\{t_2 < x \prod_1^k A_i, N_x = n\}} \right)\end{aligned}$$

Given $\{N_x = n\}$ the n points of Poisson proc.:
 n i.i.d. uniformly dist. r.v. on $[0, x]$

$U_{2,n}$: second smallest value of n i.i.d. uniform
r.v. on $[0, 1]$

$$\begin{aligned} \mathbb{P} \left(t_2 \leq \frac{x}{\mathcal{A}}, N_x = n \right) \\ = \mathbb{P} (t_2 \leq x/\mathcal{A} | N_x = n) \frac{x^n}{n!} e^{-x} \end{aligned}$$

$$\mathbb{P} (t_2 \leq x/\mathcal{A} | N_x = n) = \mathbb{P} (U_{2,n} \leq 1/\mathcal{A})$$

Probabilistic de-Poissonization

$$\begin{aligned}\mathbb{E}(R_{N_x}) &= \sum_{n \geq 0} \mathbb{E}(R_n) \frac{x^n}{n!} e^{-x} \\ &= 1 + 2 \sum_{k=0}^{+\infty} \sum_{n=2}^{+\infty} \mathbb{E} \left(\frac{1}{\prod_1^k A_i} \mathbf{1}_{\{t_2 < x \prod_1^k A_i, N_x = n\}} \right) \\ \mathbb{E}(R_{N_x}) &= \sum_{n \geq 0} \frac{x^n}{n!} e^{-x} \\ &\quad + 2 \sum_{n=2}^{+\infty} \sum_{k=0}^{+\infty} \mathbb{E} \left(\frac{1}{\prod_1^k A_i} \mathbf{1}_{\{U_{2,n} < \prod_1^k A_i\}} \right) \frac{x^n}{n!} e^{-x}\end{aligned}$$

An Associated Random Walk

$U_{2,n}$: **2th min.** of n ind. uniform r.v. on $[0, 1]$

$$\mathbb{E}(R_n) = 1 + 2\mathbb{E} \left(\sum_{k \geq 0} \frac{1}{\prod_1^k A_i} \mathbf{1}_{\{U_{2,n} < \prod_1^k A_i\}} \right), \quad n \geq 2.$$

$$\frac{\mathbb{E}(R_n) - 1}{2n} = \mathbb{E} \left(\sum_k e^{-\log n + \sum_{i=1}^k -\log(A_i)} \times \mathbf{1}_{\left\{ -\sum_1^k \log(A_i) \leq -\log U_{2,n} \right\}} \right)$$

The Use of Renewal Theorem

$$L_i = -\log A_i \quad U_{2,n} \sim 1/n$$

$$\Delta_n = \mathbb{E} \left(\sum_k e^{-(\log n - \sum_{i=1}^k L_i)} \mathbf{1}_{\left\{ \sum_1^k L_i \leq \log n \right\}} \right)$$

$$\Delta_n \sim \mathbb{E} \left(\sum_{k \leq 0} e^{-\sum_{i=k}^0 \tau_i^{\log n}} \right) \sim \mathbb{E} \left(\sum_{k \leq 0} e^{-\sum_{i=k}^0 L_i^*} \right)$$

if L_0 non-lattice.

Renewal Theorems

- Dist. of $-\log A$ non-lattice: $\log p / \log q \notin \mathbb{Q}$.
Continuous Renewal Theorem.
Convergence of $\mathbb{E}(R_n) / n$.
- Dist. of $-\log A$ lattice: $\log p / \log q \in \mathbb{Q}$.
Discrete Renewal Theorem.
Periodic Fluctuations of $\mathbb{E}(R_n) / n$.

General Splitting Scheme

Algorithm $\mathcal{A}(n)$

— $n < D \Rightarrow$ STOP.

Otherwise:

— Take a r.v. $G \in \mathbb{N}$; **Branching variable**

— Take a random probability vector

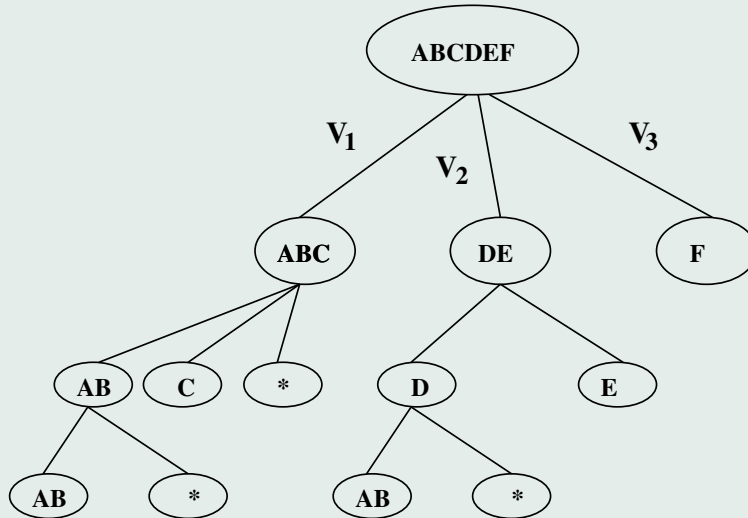
$$\mathcal{V} = (V_1, \dots, V_G), \quad V_1 + \dots + V_G = 1;$$

Weights on arcs

— Split n into G subgroups (n_1, \dots, n_G)
randomly, according to vector \mathcal{V} .

— Apply $\mathcal{A}(n_1), \dots, \mathcal{A}(n_G)$.

General Splitting Algorithms



Splitting Measure

Probability Measure on $[0, 1]$:

$$\int_0^1 f(x) \mathcal{W}(dx) = \mathbb{E} \left(\sum_{i=1}^G V_{i,G} f(V_{i,G}) \right) .$$

Theorem. [Mohamed and R. (2005)] If

$$\int_0^1 \frac{|\log(y)|}{y} \mathcal{W}(dy) < +\infty,$$

— if $\log \mathcal{W}$ non-lattice,

$$\lim_{n \rightarrow +\infty} \mathbb{E}(R_n)/n = \frac{\mathbb{E}(G)}{(D-1) \int_0^1 |\log(y)| \mathcal{W}(dy)}.$$

— $\log \mathcal{W}$ lattice: $\mathbb{E}(R_n)/n \sim F(\log n/\lambda)$

$$F(x) = C \int_0^{+\infty} \exp\left(-\lambda \left\{x - \frac{\log y}{\lambda}\right\}\right) \frac{y^{D-2}}{(D-1)!} e^{-y} dy$$

Some connections

- Fragmentation processes;
- Random Fractal subsets of $[0, 1]$;
- Branching processes:
Multiplicative martingales;
- Dynamical systems: counting problems.

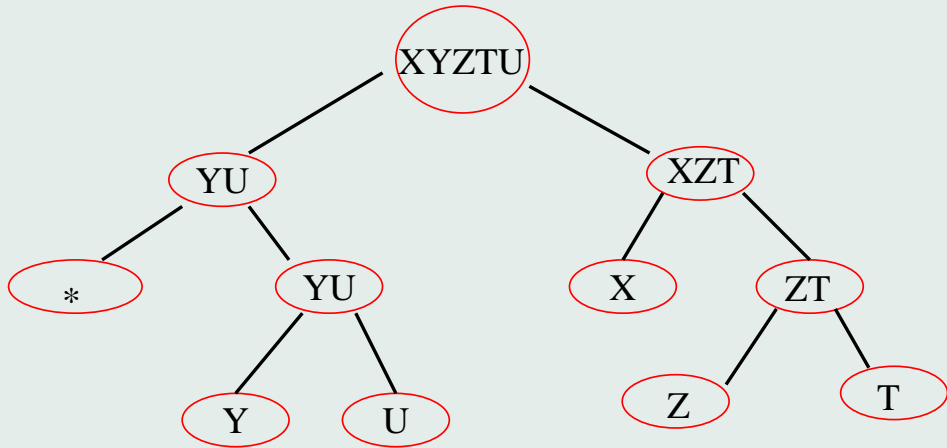
4. Stationary tries

Stationary sequences

- $X = (X_k, k \in \mathbb{Z}) \in \{0, 1\}^{\mathbb{Z}}$
a stationary sequence;
- Each item draws a sequence dist. as X ;
 X_i : “coin” (possibly) used at level i ;
- R_n size of the tree with n items.

Asymptotic behavior of $\mathbb{E}(R_n)/n$?

X = 10, Y = 010, Z = 1110, T = 1111, U = 011.



Functional Analysis Approach

Context

- A function $\phi[0, 1] \rightarrow [0, 1]$;
- A Partition $[0, 1] = \cup_i I_i$;
- $X_n = p$ if $\phi^{(n)}(X_0) \in I_p$.

Methods

- Ruelle's Transfer Operator;
- Functional Transforms.

Baladi, Bourdon, Clément, Flajolet, Vallée, . . .

Cost Function

$$C_n(f) = \sum_{\alpha \in \Sigma^*} f(np_\alpha)$$

Σ^* finite vectors in $\{0, 1, \dots, C\}$

p_α proba. of vector $\alpha \in \Sigma^*$;

$f(0) = 0$.

Exemple: Clément, Flajolet and Vallée (2001)

$$\begin{aligned} R_n &= \sum_{\alpha \in \Sigma^*} (1 - (1 + np_\alpha)(1 - p_\alpha)^n) \\ &\sim \sum_{\alpha \in \Sigma^*} \int_0^{np_\alpha} u e^{-u} du \end{aligned}$$

Probabilistic rewriting of cost function

$$\begin{aligned} C_n(f) &= \sum_{\alpha \in \Sigma^*} f(np_\alpha) \\ &= \sum_{k \geq 1} \sum_{\substack{\alpha \in \Sigma^* \\ |\alpha|=k}} \frac{f(np_\alpha)}{p_\alpha} p_\alpha \\ &= \sum_{k \geq 1} \mathbb{E} \left(\frac{f(np_{\kappa_k})}{p_{\kappa_k}} \right) \end{aligned}$$

κ_k projection on k first coord.

Rewriting of cost function: Fubini's Theorem

$$\begin{aligned} C_n(f) &= \sum_{k \geq 1} \mathbb{E} \left(\frac{f(np_{\kappa_k})}{p_{\kappa_k}} \right) \\ &= \mathbb{E} \left(\sum_{k \geq 1} \frac{1}{p_{\kappa_k}} \int_0^{+\infty} \mathbf{1}_{\{u \leq np_{\kappa_k}\}} f'(u) du \right) \\ &= \int_0^{+\infty} \mathbb{E} (G_n(u)) f'(u) du, \end{aligned}$$

with

$$G_n(u) = \sum_{k \geq 1} \frac{1}{p_{\kappa_k}} \mathbf{1}_{\{u \leq np_{\kappa_k}\}}.$$

A rough estimation

Shannon's Theorem: $-\log p_{\kappa_k} \sim kH$, a.s.
 H entropy.

$$\begin{aligned} G_n(u) &= \sum_{k \geq 1} \frac{1}{p_{\kappa_k}} \mathbf{1}_{\{u \leq n p_{\kappa_k}\}} \\ &= \sum_{k \geq 1} e^{-\log p_{\kappa_k}} \mathbf{1}_{\{-\log p_{\kappa_k} \leq \log(n/u)\}} \\ &\stackrel{\text{"} \sim \text{"}}{\sim} \sum_{k \geq 1} e^{kH} \mathbf{1}_{\{k \leq \log(n/u)/H\}} \\ &= \frac{e^{\lceil \log(n/u)/H \rceil H} - 1}{e^H - 1} \sim \frac{n}{u(e^H - 1)}. \end{aligned}$$

G_n of order n .

More Rigor

$$\begin{aligned} p_{\kappa_n(\omega)} &= \mathbb{P}(X_n = \omega_n, \dots, X_0 = \omega_0 \mid X_k = \omega_k, k < 0) \\ &= \prod_{i=1}^n \mathbb{P}(X_i = \omega_i \mid X_i = \omega_i, \dots, X_0 = \omega_0, X_k = \omega_k, k < 0) \\ -\log p_{\kappa_n} &= \sum_{i=1}^n h \circ \theta^i(\omega) \end{aligned}$$

θ : shift on sequences;

$$h(\omega) = -\log \mathbb{P}(X_0 = \omega_0 \mid X_k = \omega_k, k < 0).$$

h entropy function, $H = \mathbb{E}(h)$

More Rigor (II)

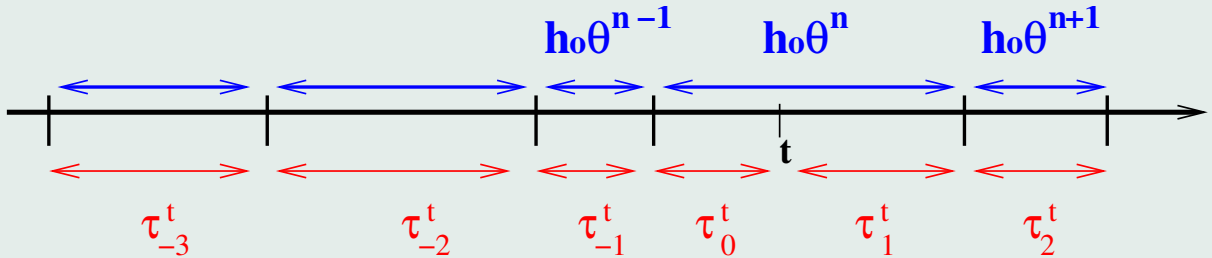
$$-\log p_{\kappa_n} = \sum_{i=1}^n h \circ \theta^i(\omega)$$

$$G_n(u) = \sum_{k \geq 1} \frac{1}{p_{\kappa_k}} \mathbf{1}_{\{u \leq n p_{\kappa_k}\}}$$

$$\frac{G_n(u)}{n} = \frac{1}{u} \sum_{k \geq 1} \exp \left[- \left(\log(n/u) - \sum_1^k h \circ \theta^i \right) \right] \\ \mathbf{1}_{\left\{ \sum_1^k h \circ \theta^i \leq \log(n/u) \right\}}$$

\Rightarrow Renewal Theorem Context.

Renewal Theory for Stationary Sequences



Renewal Theorem: $(h \circ \theta^n)$ stationary ergodic,

$$(\tau_i^t, i \in \mathbb{Z}) \xrightarrow{\text{dist.}} ?$$

Renewal Theory for Stationary Sequences (II)

Blanchard (1976) Delasnerie and Neveu (1977)
Equivalence between

- Renewal thm. for sequence $(h \circ \theta^n)$;
- Mixing property of special flow under h .
- The variable h is non lattice.

Lalley (1989) and Guivarc'h

- Limit theorems for counting measures of stationary sequences.

A convergence result

Entropy function

$$h(x) = -\log(P(X_0 = x \mid X_{-1}, X_{-2}, \dots))$$

Entropy $H = \mathbb{E}(h(X_0))$.

Theorem:

If the distribution of $h(X_0)$ is not lattice,

$$\lim_{n \rightarrow +\infty} \frac{\mathbb{E}(C_n(f))}{n} = \frac{1}{H} \int_0^{+\infty} \frac{f'(u)}{u} du.$$

The End