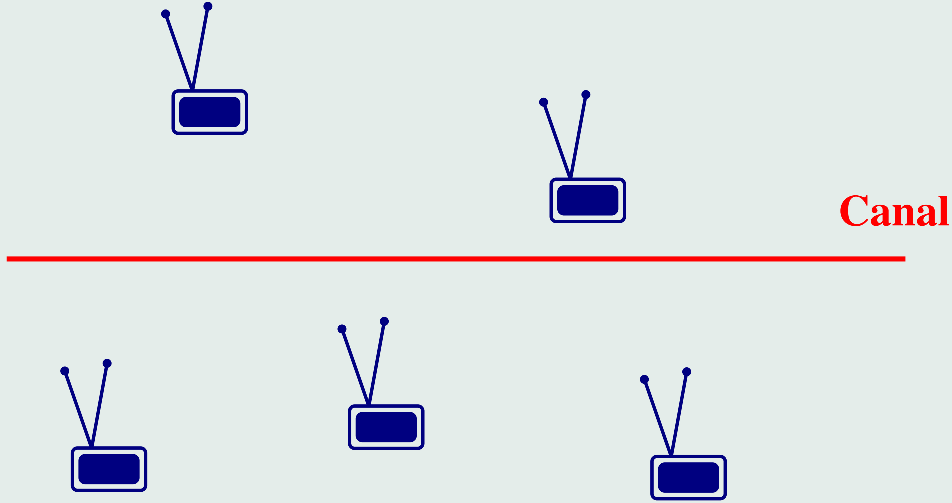
Probabilistic Methods for Algorithms and Stochastic Networks

Lecture 2

Philippe Robert
INRIA Paris-Rocquencourt

A Reminder


Canal

Information Available at a Station

Every time unit
A station can listen the channel and detect:

0 —Silence
no transmission attempt on the channel.

1 —Success
only one attempt.

2 —Collision
at least two attempts.

Ternary information.

The Tree Algorithm

The Algorithm

Each station s has a variable "counter" C_s.

— If $C_s = 0$: transmission attempt.

1. Success: the end.

2. Collision: coin tossing:
   $C_s = 0$ if head, $C_s = 1$ otherwise.

The Algorithm (II)

— If $C_s > 0$, no attempt.

State of Channel:

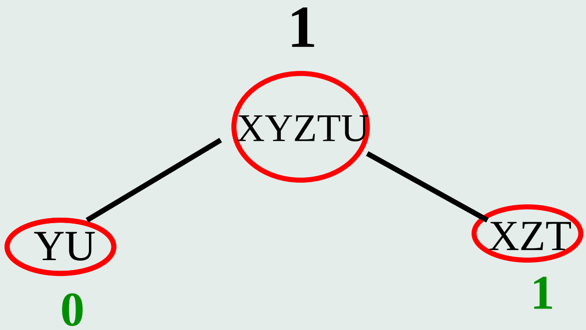
1. Success or Silence $C_s \to C_s - 1$.

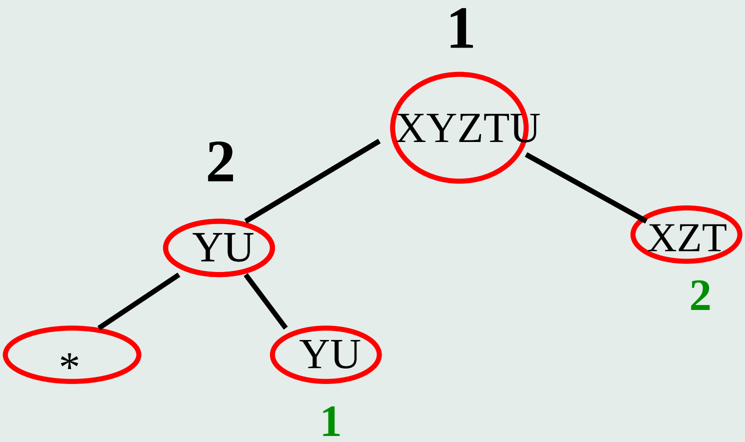
2. Collision, $C_s \to C_s + 1$.

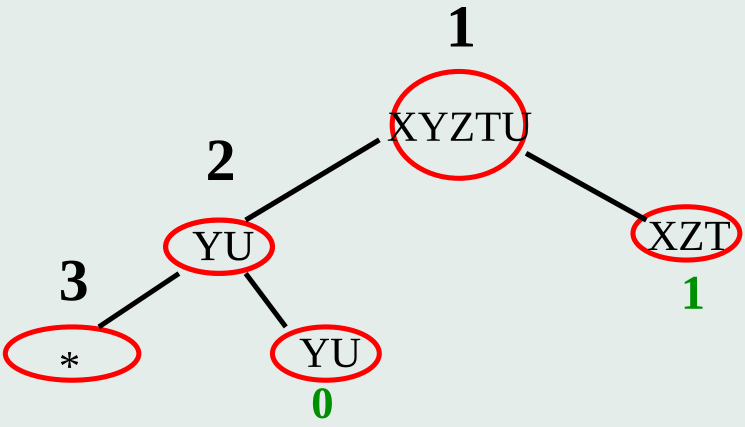
1

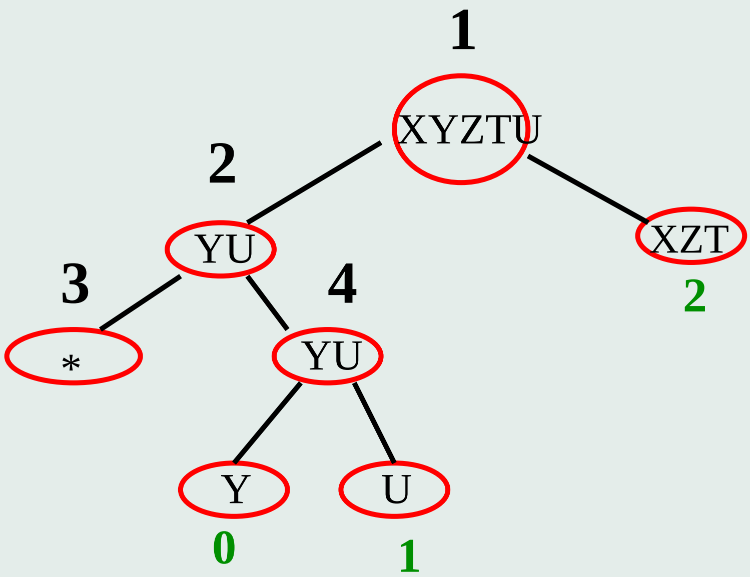
XYZTU

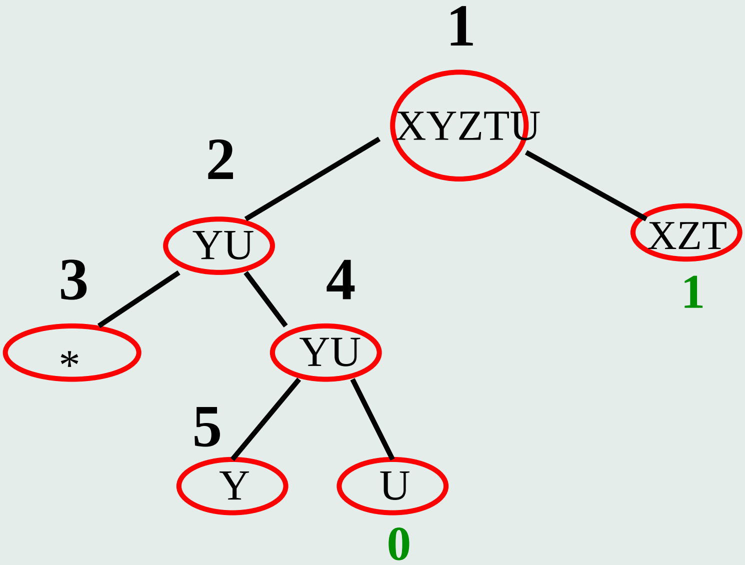
0


1
XYZTU
YU
0
XZT
1

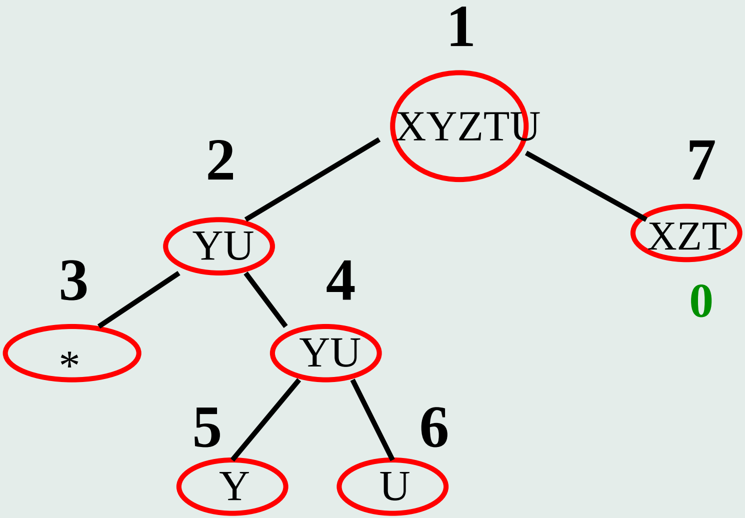

1
XYZTU
2
YU
XZT
2
*
YU
1

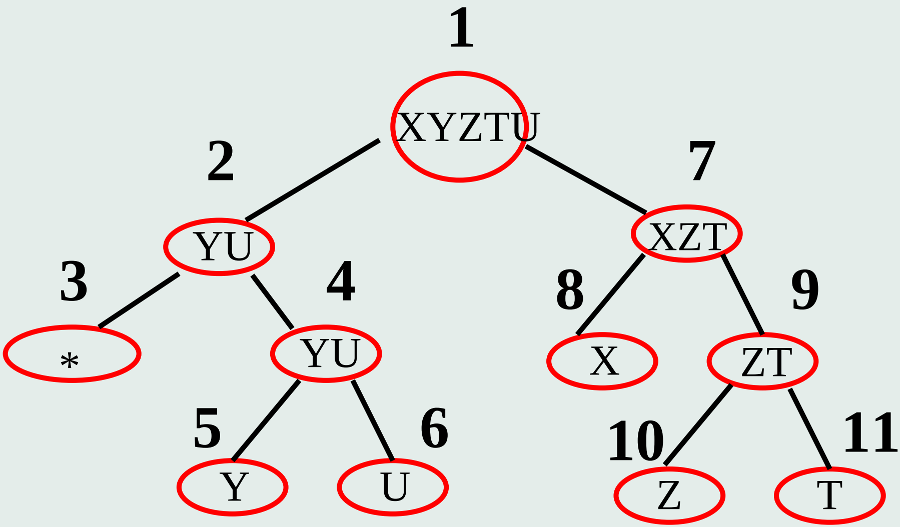

1
XYZTU
2
YU
XZT
1
3
*
YU
0


1
XYZTU
2
YU
XZT
2
3
4
*
YU
Y
U
0
1

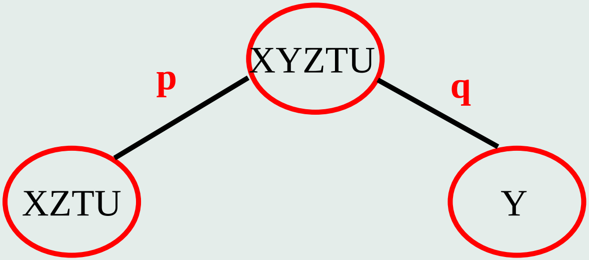


1
XYZTU
2
YU
XZT
1
3
4
*
YU
5
Y
U
0


1
XYZTU
2
7
YU
XZT
3
4
0
*
YU
5
6
Y
U


1
XYZTU
2
YU
7
XZT
3
*
4
YU
8
X
9
ZT
5
Y
6
U
10
Z
11
T

A Non-Symmetrical Tree Algorithm



$$\mathbf{p + q = 1, \quad p > q}$$

Digital Search Trees

A Problem of Data Structure

— N elements $x_1, \ldots, x_N \in \mathcal{S}$

— Pb: if $y \in \mathcal{S}$,
determine **quickly** if $y \in \{x_1, \ldots, x_n\}$:
Small nb of operations.

Hash Function

— $h : \mathcal{S} \longrightarrow [0, 1]$;

— For $x \in \mathcal{S}$, $h(x) = 0.X_1^x X_2^x \ldots X_n^x \ldots$,

  - $X_k^x \in \{0, 1\}$, $k \geq 1$,

Hash Function

— $h : \mathcal{S} \longrightarrow [0,1]$;

— For $x \in \mathcal{S}$, $h(x){=}0.X_1^x X_2^x \dots X_n^x \dots$,

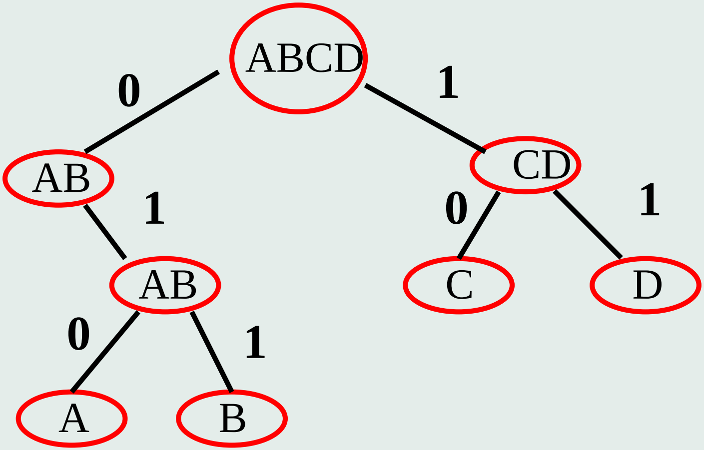
  – $X_k^x \in \{0,1\}$, $k \geq 1$,

  – **Assumption 1**: $(X_n^x, n{\geq}1)$ independent,

$$\mathbb{P}(X_n^x = 0) = \mathbb{P}(X_n^x = 1) = 1/2.$$

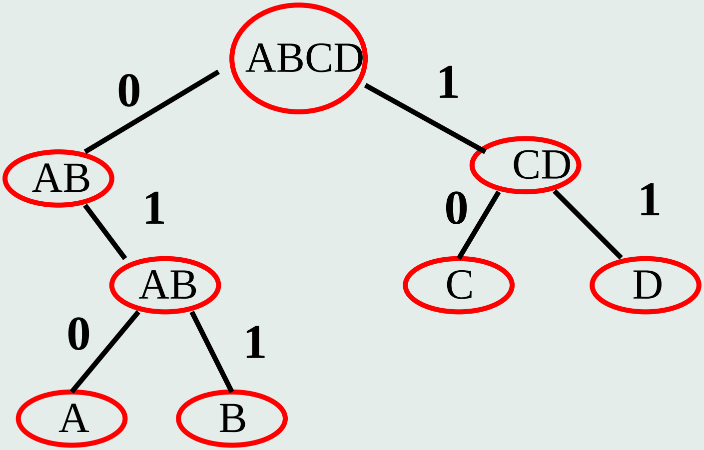
  – **Assumption 2**: $((X_n^x, n{\geq}1), x{\in}\mathcal{S})$ independent,

A Digital Search Tree



$A = h(x_1),\ B = h(x_2),\ C = h(x_3),\ D = h(x_4)$
$A = 0.0101,\ B = 0.0111,\ C = 0.101,\ D = 0.110$

A Digital Search Tree



$A = h(x_1),\ B = h(x_2),\ C = h(x_3),\ D = h(x_4)$
$A = 0.0101,\ B = 0.0111,\ C = 0.101,\ D = 0.110$

Easy to test if $h(y) = 0.y_1y_2\ldots$ is stored

Divide and Conquer Algorithms

Algorithm $\mathcal{A}(n)$:

— Termination Condition $\longrightarrow$ Stop.

— Otherwise: Tree Structure

Split at random into d groups

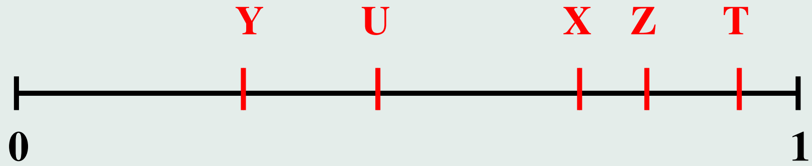
of size $n_1,\ n_2, \ldots n_d,$

$$n_1 + n_2 + \cdots + n_d = n$$

$\Rightarrow$ Apply $\mathcal{A}(n_1)$, $\mathcal{A}(n_2)$, . . . , $\mathcal{A}(n_d)$.

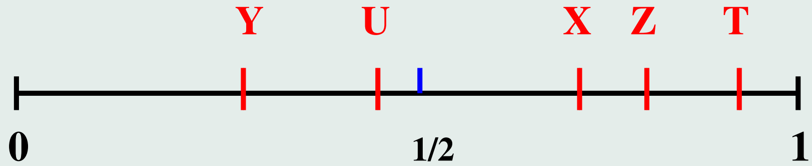
Related Mathematical Models

An Alternative point of View



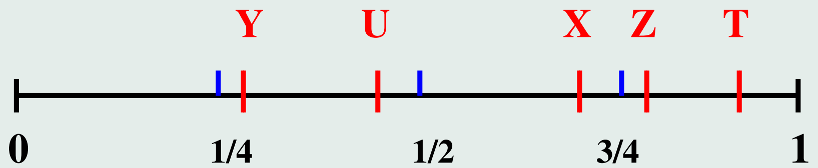
5 random points in $[0, 1]$.

Interval Fragmentation



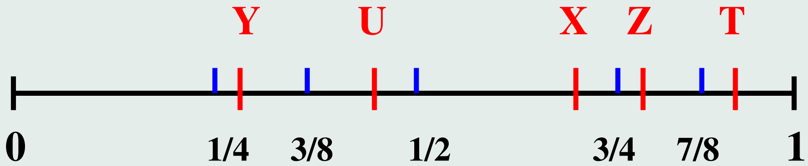
Split any sub-interval with more than 2 points.

Interval Fragmentation



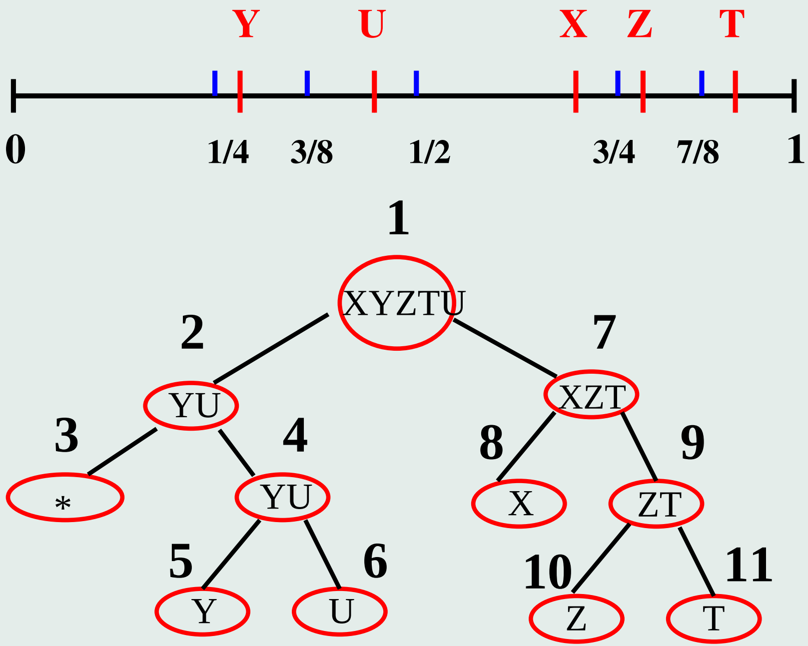
Split any sub-interval with more than 2 points.

Interval Fragmentation



Split any sub-interval with more than 2 points.

Interval Fragmentation



Related Problems I

Random Recursive Decompositions

— Each Step: each interval $I \Rightarrow (I_1, \ldots, I_\sigma)$.

— Geometric properties of:

$$\bigcap_{n \geq 1} \left(\bigcup_{J:\ \text{interval created at step } n} J \right)$$

Related Problems I

Random Recursive Decompositions

— Each Step: each interval $I \Rightarrow (I_1, \ldots, I_\sigma)$.

— Geometric properties of:

$$\bigcap_{n \geq 1} \left(\bigcup_{J:\ \text{interval created at step } n} J \right)$$

Hausdorff dimension, . . .

Mauldin and Williams (1988),
Waymire and Willliams (1996)
Literature on Multiplicative Martingales.

Related Problems II

Fragmentation Processes

At time t a particle of mass $x \Rightarrow \sigma$ particles of size $x_1, \ldots, x_\sigma$.

Pitman (1995), Bertoin (2001)

Related Problems II

Fragmentation Processes

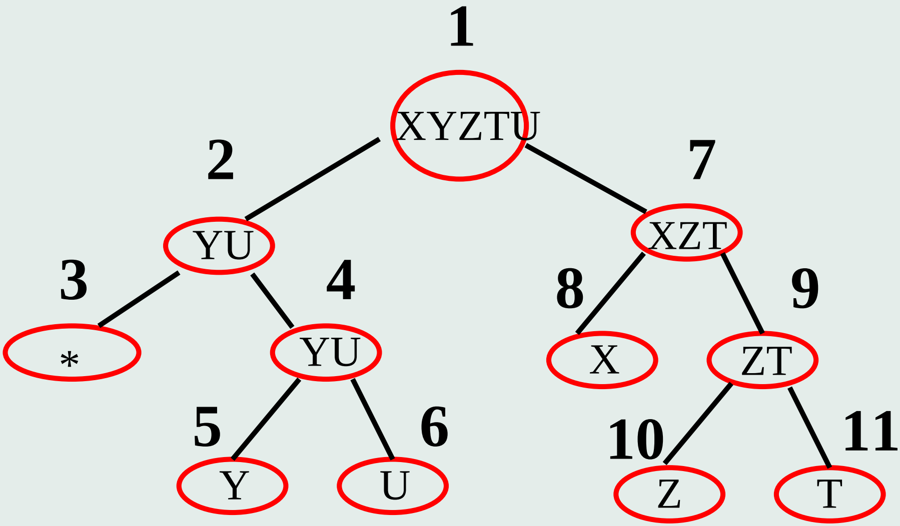
At time t a particle of mass $x \Rightarrow \sigma$ particles of size $x_1, \ldots, x_\sigma$.

Pitman (1995), Bertoin (2001)

The Tree Algorithms "breaks" the integer n into pieces of size 0 and 1.

Asymptotic Behavior

Back to Tree Algorithm



Size of Tree: Time to Transmit all Messages.

Asymptotic behavior

Definition: R_n nb of nodes with n at root.

Asymptotic behavior

Definition: R_n nb of nodes with n at root.

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

Asymptotic behavior

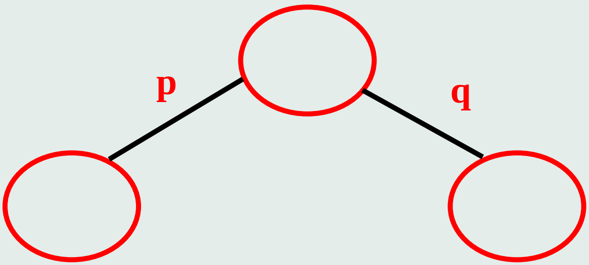
Definition: R_n nb of nodes with n at root.

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

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

A Recursive Approach

Binary Tree



Recurrence Relation

(R_n) is an An Additive Functional
$R_0 = R_1 = 1$.

If $n \geq 2$,

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

with

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

(B_i) i.i.d. Bernoulli parameter p.

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

Recurrence Relation

(R_n) is an An Additive Functional
$R_0 = R_1 = 1$.

If $n \geq 0$,

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

with

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

(B_i) i.i.d. Bernoulli parameter p.

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

A Poisson Process

$(N[0,x], x \geq 0)$ Poisson proc. intensity 1

$(N[0,x]) \Leftrightarrow (t_k)$ with $0 < t_1 \leq \cdots \leq t_k \leq \cdots$

$$N[0,x] = \#\{k : t_k \leq x\}.$$

A Poisson Process

$(N[0,x], x \geq 0)$ Poisson proc. intensity 1

$(N[0,x]) \Leftrightarrow (t_k)$ with $0 < t_1 \leq \cdots \leq t_k \leq \cdots$

$$N[0,x] = \#\{k : t_k \leq x\}.$$

the variables

$$\left(\sum_{i=1}^{N[0,x]} B_i, \sum_{i=1}^{N[0,x]} (1 - B_i) \right)$$

are Poisson with resp. rate px and $qx = (1-p)x$.

A Poisson Process

$(N[0,x], x \geq 0)$ Poisson proc. intensity 1

$(N[0,x]) \Leftrightarrow (t_k)$ with $0 < t_1 \leq \cdots \leq t_k \leq \cdots$

$$N[0,x] = \#\{k : t_k \leq x\}.$$

the variables

$$\left(\sum_{i=1}^{N[0,x]} B_i, \sum_{i=1}^{N[0,x]} (1 - B_i) \right) \overset{\text{dist}}{=} (N_1[0,px], N_2[0,qx])$$

are Poisson with resp. rate px and $qx = (1-p)x$.

Recurrence Relation

$$R_n \stackrel{\text{dist.}}{=} 1 + R_{X_n} + \overline{R}_{n-X_n} - 2_{\{n \le 1\}}$$

Recurrence Relation

$$R_n \overset{\text{dist.}}{=} 1 + R_{X_n} + \overline{R}_{n-X_n} - 2_{\{n\le 1\}}$$

Plug $n = N[0,x]$

$$R_{N[0,x]} \overset{\text{dist.}}{=} 1 + R_{X_{N[0,x]}} + \overline{R}_{N[0,x]-X_{N[0,x]}} - 2_{\{N[0,x]\le 1\}}$$

Recurrence Relation

$$R_n \stackrel{\text{dist.}}{=} 1 + R_{X_n} + \overline{R}_{n-X_n} - 2_{\{n \le 1\}}$$

Plug $n = N[0, x]$

$$R_{N[0,x]} \stackrel{\text{dist.}}{=} 1 + R_{X_{N[0,x]}} + \overline{R}_{N[0,x]-X_{N[0,x]}} - 2_{\{N[0,x] \le 1\}}$$

$$R_{N[0,x]} \stackrel{\text{dist.}}{=} 1 + R_{N_1[0,px]} + \overline{R}_{N_2[0,qx]} - 2_{\{t_2 \ge x\}}$$

Poisson Transform of a Sequence

If (C_n) sequence of $\mathbb{R}_+$,
Poisson transform of (C_n) at $x \geq 0$

$$\phi_C(x) = \sum_{n=0}^{+\infty} C_n \frac{x^n}{n!} e^{-x}$$

$$\phi_C(x) = \mathbb{E}(C_{N[0,x]}).$$

Poisson Transform of a Sequence

If (C_n) sequence of $\mathbb{R}_+$,
Poisson transform of (C_n) at $x \geq 0$

$$\phi_C(x) = \sum_{n=0}^{+\infty} C_n \frac{x^n}{n!} e^{-x}$$

$$\phi_C(x) = \mathbb{E}(C_{N[0,x]}).$$

Property

$$\frac{d\phi_C}{dx} = \phi_{\Delta C}(x),$$

with $\Delta C = (C_{n+1} - C_n)$.

Poisson Transform of $\boldsymbol{R} = (\mathbb{E}(\boldsymbol{R}_n))$

$$\boldsymbol{R}_{N[0,x]} \overset{\text{dist.}}{=} \mathbf{1} + \boldsymbol{R}_{N_1[0,px]} + \overline{\boldsymbol{R}}_{N_2[0,qx]} - \mathbf{2}_{\{t_2 \geq x\}}$$

Poisson Transform of $R = (\mathbb{E}(R_n))$

$$R_{N[0,x]} \overset{\text{dist.}}{=} 1 + R_{N_1[0,px]} + \overline{R}_{N_2[0,qx]} - 2_{\{t_2 \geq x\}}$$

$$\phi_R(x) = \phi_R(px) + \phi_R(qx) + 1 - 2\mathbb{P}(t_2 \geq x)$$

Symmetrical Case $p = q = 1/2$

$$\psi(x) = \phi_R(x) - 1,$$

$$\psi(x) = 2\psi(x/2) + 2\mathbb{P}(t_2 \leq x)$$

Symmetrical Case $p = q = 1/2$

$$\psi(x) = \phi_R(x) - 1,$$

$$\psi(x) = 2\psi(x/2) + 2\mathbb{P}(t_2 \leq x)$$

$$\phi_R(x) = 1 + 2\sum_{k\geq 0} 2^k \mathbb{P}\left(t_2 \leq x/2^k\right)$$

Symmetrical Case $p = q = 1/2$

$$\psi(x) = \phi_R(x) - 1,$$

$$\psi(x) = 2\psi(x/2) + 2\mathbb{P}(t_2 \le x)$$

$$\phi_R(x) = 1 + 2\sum_{k\ge 0} 2^k \mathbb{P}\left(t_2 \le x/2^k\right)$$

$$= 1 + 2\sum_{k\ge 0} 2^k \sum_{n\ge 2} \mathbb{P}\left(t_2 \le \frac{x}{2^k}, N[0,x] = n\right)$$

$$\mathbb{P}\left(t_2 \le \frac{x}{2^k}, N[0,x] = n\right) =$$

$$\mathbb{P}(t_2 \le x/2^k \mid N[0,x] = n)\frac{x^n}{n!}e^{-x}$$

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

$$\mathbb{P}(t_2 \le \frac{x}{2^k}, N[0,x] = n) = \mathbb{P}(t_2 \le x/2^k | N[0,x] = n) \frac{x^n}{n!} e^{-x}$$

Given $\{N[0,x]=n\}$ the n points of Poisson proc.
are like 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 $[0,1]$

$$\mathbb{P}(t_2 \leq \frac{x}{2^k}, N[0,x] = n) = \mathbb{P}(t_2 \leq x/2^k | N[0,x] = n) \frac{x^n}{n!} e^{-x}$$

$$\mathbb{P}(t_2 \leq x/2^k | N[0,x] = n) = \mathbb{P}\left(xU_{2,n} \leq x/2^k\right)$$

Given $\{N[0,x]=n\}$ the n points of Poisson proc. are like 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 $[0,1]$

$$\mathbb{P}(t_2 \leq \frac{x}{2^k}, N[0,x] = n) = \mathbb{P}(t_2 \leq x/2^k | N[0,x] = n) \frac{x^n}{n!} e^{-x}$$

$$\mathbb{P}(t_2 \leq x/2^k | N[0,x] = n) = \mathbb{P}\left(U_{2,n} \leq 1/2^k\right)$$

Expression for $\phi_R(x)$

$$\phi_R(x) = \sum_{n\geq 0} \mathbb{E}(R_n)\frac{x^n}{n!}e^{-x}$$

$$= 1 + 2\sum_{k\geq 0} 2^k \sum_{n\geq 2} \mathbb{P}\left(U_{2,n} \leq 1/2^k\right)\frac{x^n}{n!}e^{-x}$$

Expression for $\phi_R(x)$

$$\phi_R(x) = \sum_{n\geq 0} \mathbb{E}(R_n)\frac{x^n}{n!}\, e^{-x}$$

$$= 1 + 2\sum_{k\geq 0} 2^k \sum_{n\geq 2} \mathbb{P}\left(U_{2,n} \leq 1/2^k\right) \frac{x^n}{n!}\, e^{-x}$$

$$= 1 + 2\sum_{n\geq 2} \left(\sum_{k\geq 0} 2^k \mathbb{P}\left(U_{2,n} \leq 1/2^k\right)\right) \frac{x^n}{n!}\, e^{-x}$$

Expression for $\mathbb{E}(\boldsymbol{R_n})$

$$\mathbb{E}(\boldsymbol{R_n}) = \mathbf{1} + \mathbf{2} \sum_{\boldsymbol{k} \geq \mathbf{0}} \mathbf{2}^{\boldsymbol{k}} \, \mathbb{P}\left(\boldsymbol{U_{2,n}} \leq \frac{\mathbf{1}}{\mathbf{2}^{\boldsymbol{k}}} \right)$$

Expression for $\mathbb{E}(R_n)$

$$\begin{aligned}\mathbb{E}(R_n) &= 1 + 2\sum_{k\geq 0} 2^k\,\mathbb{P}\left(U_{2,n} \leq \frac{1}{2^k}\right)\\ &= 1 + 2\sum_{k\geq 0} 2^k\,\mathbb{E}\left(\mathbf{1}_{\{U_{2,n}\leq 1/2^k\}}\right)\end{aligned}$$

Expression for $\mathbb{E}(R_n)$

$$\begin{aligned}\mathbb{E}(R_n) &= 1 + 2\sum_{k\geq 0} 2^k\,\mathbb{P}\left(U_{2,n} \leq \frac{1}{2^k}\right)\\ &= 1 + 2\sum_{k\geq 0} 2^k\,\mathbb{E}\left(\mathbf{1}_{\{U_{2,n}\leq 1/2^k\}}\right)\\ &= 1 + 2\mathbb{E}\left(\sum_{k=0}^{\lfloor -\log_2(U_{2,n})\rfloor} 2^k\right)\end{aligned}$$

Expression for $\mathbb{E}(R_n)$

$$\begin{aligned}\mathbb{E}(R_n) &= 1 + 2\sum_{k\geq 0} 2^k\,\mathbb{P}\left(U_{2,n} \leq \frac{1}{2^k}\right)\\ &= 1 + 2\sum_{k\geq 0} 2^k\,\mathbb{E}\left(1_{\{U_{2,n}\leq 1/2^k\}}\right)\\ &= 1 + 2\mathbb{E}\left(\sum_{k=0}^{\lfloor -\log_2(U_{2,n})\rfloor} 2^k\right)\\ &= 4\mathbb{E}\left(2^{\lfloor -\log_2(U_{2,n})\rfloor}\right) - 1\end{aligned}$$

Asymptotic expansion for $\mathbb{E}(R_n)$

Proposition. For $n \geq 2$,

$$\frac{\mathbb{E}(R_n)}{n} = 4 \int_0^1 2^{-\{-\log_2(x)\}} (n-1)\,(1-x)^{n-2}\, dx - \frac{1}{n}$$

Asymptotic expansion for $\mathbb{E}(R_n)$

Proposition. For $n \geq 2$,

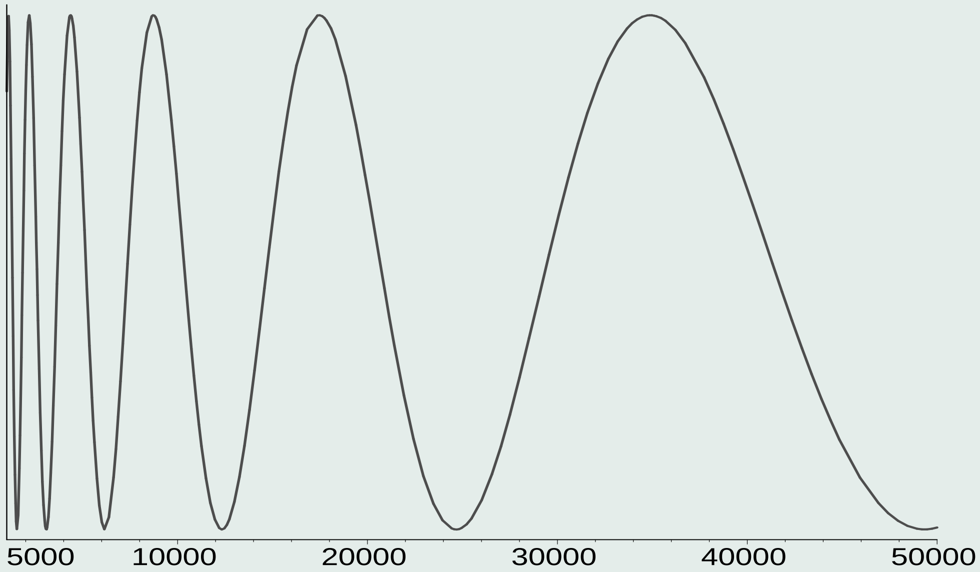
$$\frac{\mathbb{E}(R_n)}{n} = 4 \int_0^1 2^{-\{-\log_2(x)\}} (n-1)\,(1-x)^{n-2}\, dx - \frac{1}{n}$$

$$= F(\log_2(n)) + o(1),$$

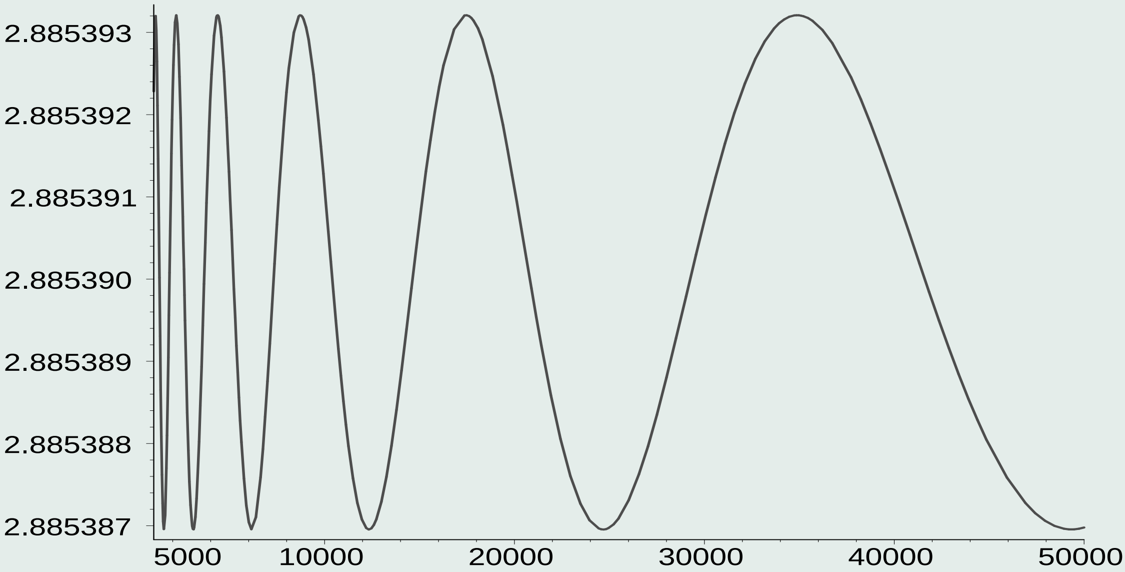
with $F(y) = 4 \int_0^{+\infty} 2^{-\{y-\log_2(x)\}} e^{-x}\, dx.$

and $\{z\} = z - \lfloor z \rfloor$

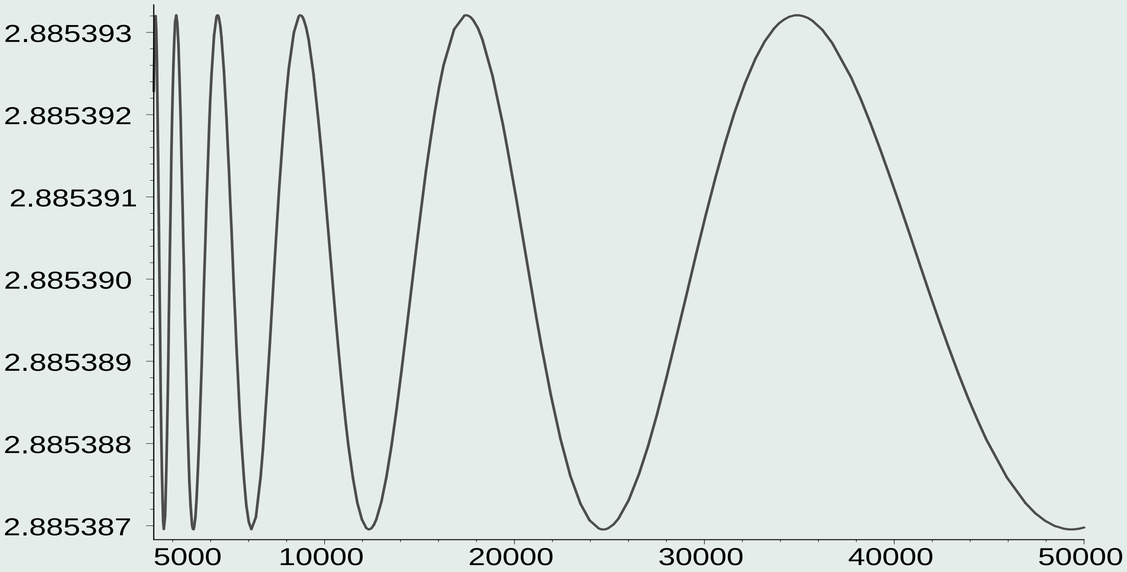
The sequence $n \to \mathbb{E}(R_n)/n$



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



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



Range $\subset \left[\frac{2}{\log 2} - 4.10^{-6}, \frac{2}{\log 2} + 4.10^{-6}\right]$

Non-Symmetrical Case

Non-Symmetrical Case

$$\psi(x) = \phi_R(x) - 1,$$

$$\psi(x) = \psi(px) + \psi(qx) + 2\mathbb{P}(t_2 \leq x)$$

Non-Symmetrical Case: Asymptotic behavior

$$\left(\frac{\mathbb{E}(\boldsymbol{R}_n)}{\boldsymbol{n}}\right) : \begin{cases} \text{oscillates} \\ \\ \text{converges} \end{cases}$$

Non-Symmetrical Case: Asymptotic behavior

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

References

Hanène Mohamed and Philippe Robert
A probabilistic analysis of tree algorithms.
Annals of Applied Probability, (2005).

Hanène Mohamed and Philippe Robert
Dynamic tree algorithms.
Annals of Applied Probability, (2010).

Next Lecture

— Explanation of Periodic/Non-Periodic Behavior.

— Non-additive functionals
depth, insertion cost, . . .

— A larger class of Stochastic Models

The Renewal Theorem

Some History

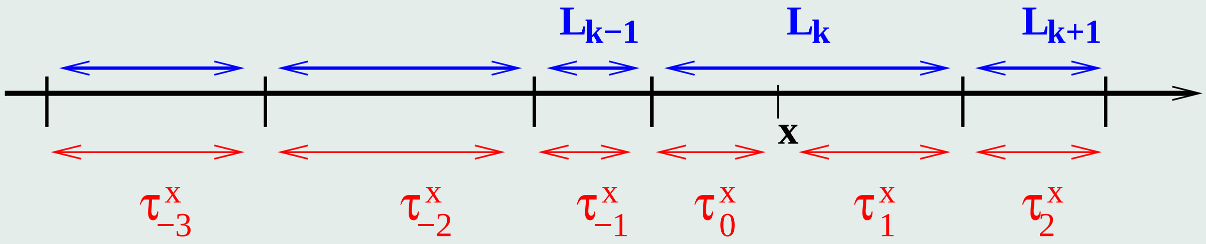
Blackwell (1948):

— A light bulb last two years in average.

— How many are necessary for ten years ?

Doob, Feller, Blackwell, . . .

General Framework



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

Does the landscape around x converge in dist. when $x \to +\infty$?

The Renewal Theorem

Translated Potential Functional

f continuous compact support on $\mathbb{R}_+$,

$$T_x G(f) \stackrel{\text{def.}}{=} \mathbb{E}\left(\sum_{n=0}^{+\infty} f(S_n - x)\right).$$

The Renewal Theorem

Translated Potential Functional

f continuous compact support on $\mathbb{R}_+$,

$$T_x G(f) \stackrel{\text{def.}}{=} \mathbb{E}\left(\sum_{n=0}^{+\infty} f(S_n - x)\right).$$

Theorem. If L_1 is non-lattice, then

$$\lim_{x \to +\infty} T_x G(f) = \frac{1}{\mathbb{E}(L_1)} \int_{\mathbb{R}} f(u)\, du.$$

Lattice Case

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

Lattice Case

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

For $h > 0$**,**

$$\lim_{n \to +\infty} T_{h+n\delta} G(f) = \mathcal{L}_h(f) = \frac{1}{\mathbb{E}(L_1)} \sum_{j \in \mathbb{Z}} f(h + j\delta)$$

Lattice Case

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

For $h > 0$,

$$\lim_{n \to +\infty} T_{h+n\delta} G(f) = \mathcal{L}_h(f) = \frac{1}{\mathbb{E}(L_1)} \sum_{j \in \mathbb{Z}} f(h+j\delta)$$

Periodic Behavior: $\mathcal{L}_h(f) = \mathcal{L}_{h+\delta}(f)$.