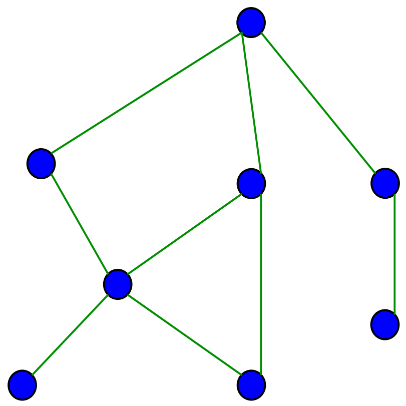
Convergence of finite Markov chains

Philippe Robert
INRIA-Rocquencourt

A random walk (X_n) on a graph $\mathcal{G}$



$x \to y$ with probability $p(x,y) = \frac{1}{d_x}$, if $y \in V_x$.

For n sufficiently large, $P(X_n = x) \sim C d_x$.

In the case of a constant degree, X_n is asymptotically uniform.

$$P(X_n = i) \sim \frac{1}{G}$$

Examples : torus, hypercube,...

More generally, if (X_n) is a Markov chain irreducible and aperiodic on a finite set S,

$$P(X_n = i) \sim \pi_i$$

(π_i) is the equilibrium distributionon S of (X_n).

Moreover (Doeblin), there exists $\alpha \in]0,1[$ such that

$$|P(X_n = i/X_0 = j) - \pi_i| \leq C\alpha^n = Ce^{-\frac{n}{\log(1/\alpha)}}$$

$\tau_r = \min\left\{\frac{1}{\log(1/\alpha)}\right\}$ is *the relaxation time*.

More accurate estimates ?

Motivations

- The equilibrium distributions have, sometimes, nice closed expressions easy to use.

- In practice, the probability at some *finite* time is useful.
  Example:
  X_n is the number of messages at time n, it is required that

$$P(X_n > C) < 10^{-9},$$

  for all $n \geq n_0$, with n_0 as small as possible.

- In general $P(X_n = i)$ is, most of the time, difficult to express.

How to measure the rate of convergence

Distances on probability distributions

- The convergence in distribution distance

$$d_l(P,Q) = \sup_{i \in S}\{|P(\{i\}) - Q(\{i\})|\}$$

- The total variation distance,

$$d_{tv}(P,Q) = \sup_{A \subset S}\{|P(A) - Q(A)|\} = \frac{1}{2}\sum_{i \in S}|P(\{i\}) - Q(\{i\})|$$

- The separation "distance"

$$d_s(P,Q) = \sup_{i \in S}\left\{\left|1 - \frac{P(\{i\})}{Q(\{i\})}\right|\right\}$$

Some elementary properties

- $d_l \leq d_{tv} \leq d_s$,
- $d_l, d_{tv} \leq 1$,
- if P, Q are very far apart, i.e. $P(A) \sim 1$ and $Q(A^c) \sim 1$, for $A \subset S$,
$$d_l(P,Q) \sim 1, \text{ and } d_{tv}(P,Q) \sim 1.$$

The *maximal variation distance* $d(t)$

$$d(t) = \max_{x \in S} \{ d_{tv}(P_x(t), \pi) \}.$$

shall measure the distance at equilibrium.

The time to reach equilibrium

$$\tau_v = \inf\{t \geq 0 / d(t) < e^{-1}\}$$

The framework and the problems

$S_N = \{1, \ldots, N\}$, a Markov chain on S_N, $N \to +\infty$,

Asymptotic behavior of τ_r^N, τ_v^N ?

Explicit bounds on $d^N(t)$?

No general theory, but examples.

"The examples are more interesting and more deep than the theory"

(Aldous-Diaconis)

Methods

- Geometry.
- Fourier analysis (if S is a group).
- Probabilistic technics: coupling.

Assumptions : (X_n) is a Markov chain with transition probabilities $P = (p(x, y))$ and equilibrium probability (π_x) which is *reversible*, i.e. $(\pi_x p(x, y))$ is symmetrical.

Probabilistic point of view : if $X_0 \sim \pi$, then $(X_0, X_1, \ldots, X_t)$ has the same distribution as $(X_t, X_{t-1}, \ldots, X_0)$.

Examples: All the random walks on graphs, all the (recurrent) birth and death processes are reversible.

Some linear Algebra

The eigenvalues of $(p(x,y))$ are such that

$$-1 < \beta_{m-1} \leq \beta_{m-2} \leq \ldots \leq \beta_1 < \beta_0 = 1,$$

A general upper bound for the convergence rate

$$d_{tv}(P_x(n),\pi) \leq \frac{1}{2}\sqrt{\frac{1-\pi(x)}{\pi(x)}}\left(\max\{\beta_1, |\beta_{m-1}|\}\right)^n.$$

The eigenvalue β_1 can be obtained via Rayleigh-Ritz's principle

$$1 - \beta_1 = \inf \left\{ \frac{\mathcal{E}_P(f,f)}{< f, f >} \Big/ f \text{ non constant } \right\} \tag{1}$$

with

$$< f, g >= \sum_{x \in S} f(x) g(x) \pi_x - \left(\sum_x f(x) \pi_x \right) \left(\sum_x g(x) \pi_x \right)$$

and

$$\mathcal{E}_P(f,f) =< f, (I-P) f >= \frac{1}{2} \sum_{x,y \in S} (f(y) - f(x))^2 \pi_x p(x,y),$$

the Dirichlet form associated to p.

A geometric point of view

The relation (1) is a Poincaré inequality,

$$< f, f >\leq (1-\beta_1)^{-1} \mathcal{E}_P(f,f).$$

Examples.

1. The random walk on $\mathbf{Z}_N$.

$$1-\beta_1^N = \frac{1}{2}\inf_f \frac{\sum_0^{N-1}(f(\frac{k+1}{N})-f(\frac{k}{N}))^2}{\sum_0^{N-1} f(\frac{k}{N})^2}, \text{ as } N \to +\infty,$$

$$N^2(1-\beta_1^N) \sim \frac{1}{2}\inf_f \frac{\int_0^1 f'(x)^2 dx}{\int_0^1 f(x)^2 dx} = 2\pi^2,$$

$\beta_1^N \sim 1-\frac{2\pi^2}{N^2}, \tau_r^N \sim \frac{N^2}{2\pi^2}$.

2. The random walk on the hypercube $\{0,1\}^N$.
   $\beta_1^N = 1-\frac{2}{N}$, $\tau_r^N \sim \frac{N}{2}$.

Coupling

Two Markov processes (X_n), $(\tilde{X}_n)$, with transition matrix p, and $X_0 = x \in S$, $\tilde{X}_0 \sim \pi$.

A coupling is a finite random time T such that

$$X_n = \tilde{X}_n, \quad n \geq T,$$

$$|P_x(X_n \in A) - \pi(A)| = \left|E\left(1_{\{X_n \in A\}} - 1_{\{\tilde{X}_n \in A\}}\right)\right| \leq P(T > n)$$

$$d_{tv}(P_x(n), \pi) \leq P(T > n)$$

There exists a coupling T^* such that $d_{tv}(P_x(t), \pi) = P(T^* > t)$.

Examples

The random walk on $\mathbf{Z}_N$. $X_0 = x \in \{0, \ldots, N-1\}$, $\tilde{X}_0$ uniform, X_n, $\tilde{X}_n$, jumps independently as long as they do not meet. If they meet they follow the same path forever.

Until $(Y_n) = (X_n - \tilde{X}_n)$ do not reach 0, (Y_n) is a symmetric random walk. The meeting time T is the hitting time of 0 by (Y_n).

$$d_{tv}(P_x(t), \pi) \leq P(T > n) \leq C\frac{N^2}{n}, \qquad \tau_v \sim N^2.$$

The random walk on the hypercube $\{0,1\}^N$.

$$d_{tv}(P_x(t), \pi) \leq \frac{1}{2}\frac{N\log(N)}{n}, \qquad \tau_v \sim N\log(N).$$

The cutoff property

The random walk on $\mathbf{Z}_N$.

The random walk on the hypercube $\{0,1\}^N$.

Definition

There is a cutoff if there exists $a_n, b_n \to +\infty$, $b_n/a_n \to 0$, such that

$$\lim_{n \to +\infty} d(a_n + tb_n) = H(t),$$

with $lim_{t \to -\infty} H(t) = 1$, $lim_{t \to +\infty} H(t) = 0$.

No cutoff for the random walk on $\mathbf{Z}_N$.

A cutoff occurs at $\frac{1}{4} N \log(N)$ for the random walk on the hypercube.

A queueing example

A node with N places. Messages arrive at rate λ and leave at rate 1. The process $X(t)$ has transition rates

$$\begin{aligned} &x \to x+1, \quad \lambda, \quad x \le N \\ &x \to x-1, \quad x \end{aligned}$$

Equilibrium distribution

$$\pi(x) = C_N \frac{\lambda^x}{x!}, \quad x \le N.$$

A renormalization

$$\begin{aligned} x &\to x+1, \quad \lambda N, \quad x \le N \\ x &\to x-1, \quad\quad x \end{aligned}$$

Stable point at $x = \lambda N$ if $\lambda \le 1$.

Approximation: $dX_N(t) = \lambda N 1_{\{X_N < N\}} dt - X_N(t) dt,$

The convergence

$$\left(\frac{X_N(t)}{N}\right) \longrightarrow \left(\lambda(1-e^{-t}) \wedge 1\right)$$

holds with probability 1.

The case $\lambda > 1$

A second order approximation, $Z_N(t) = N - X_N(t/N)$,

$$\begin{aligned} &z \to z+1, \quad 1, \\ &z \to z-1, \quad \lambda, \quad z > 0 \end{aligned}$$

$Z_N(t)$ is a birth and death process with equilibrium measure $((1-\frac{1}{\lambda})\frac{1}{\lambda^n})$.

Second eigenvalue: $\beta_N/N \to (\sqrt{\lambda}-1)^2$.

Time to reach equilibrium

when

$$\lambda(1-e^{-t}) = 1$$

$\tau_v \sim \log \frac{\lambda}{\lambda-1}$.

The case $\lambda \leq 1$

A second order approximation, $Z_N(t) = \frac{X_N(t)-\lambda N}{\sqrt{N}}$.
Asymptotically, $Z_N(t)$ satisfies

$$dZ_N(t) \sim \sqrt{2\lambda}dB(t) - Z_N(t)dt,$$

where $(B(t))$ is a Brownian motion, $Z_N(t)$ is an Ornstein-Ühlenbeck process.

Second eigenvalue: $\beta_N \to 1$, if $\lambda < 1$; $\beta_N \to 2$, if $\lambda = 1$.

Time to reach equilibrium
when

$$\lambda N(1 - e^{-t}) \sim \lambda N + a\sqrt{N}$$

$\tau_v \sim \frac{\log(N)}{2}$.

Some technical points

The free process $Y_N(t)$, (M/M/∞ queue),

$$x \to x+1, \quad \lambda N,$$
$$x \to x-1, \quad x,$$

Then
$$(\mathcal{E}_c(t)) = \left((1+ce^t)^{Y_N(t)} e^{-\lambda N c e^t}\right)$$

is a martingale.

$\phi : (t,x) \to (1+ce^t)^x e^{-\lambda N c e^t}$ is space-time harmonic

$$\frac{\partial \phi}{\partial t}(t,x) + \lambda N \left(\phi(t,x+1) - \phi(t,x)\right) + x \left(\phi(t,x-1) - \phi(t,x)\right) = 0$$

$$(1+w)^x e^{-\lambda N w} = \sum_{n=0}^{+\infty} C_n^{\lambda N}(x) \frac{w^n}{n!},$$

$C_n^{\lambda N}(x)$ is a Poisson-Charlier polynomial.

Integrating $\mathcal{E}_c(t)$ with respect to $c^{\alpha-1}dc, c > 0$,

$$M(t) = \left(e^{-\alpha t} \int_0^{+\infty} (1+c)^{Y_N(t)} c^{\alpha-1} e^{-\lambda N c} dc \right),$$

is also a martingale.

For $a < b$, if T_{ab} is the hitting time of b starting from a for $(Y_N(t))$.
Martingale property:

$$E(M(T_{ab})) = E(M(0))$$

$$E\left(e^{-\alpha T_{ab}}\right) = \frac{\int_0^{+\infty} (1+c)^a c^{\alpha-1} e^{-\lambda N c} dc}{\int_0^{+\infty} (1+c)^b c^{\alpha-1} e^{-\lambda N c} dc}$$

Proposition 1 *For* $a < b$,

$$E\left(e^{-\alpha T_{ba}}\right) = \frac{\Psi_b(\alpha)}{\Psi_a(\alpha)},$$

with

$$\Gamma_y(\alpha) = \int_0^{+\infty} (1+c)^y c^{\alpha-1} e^{-\lambda N c} dc,$$

and

$$\Psi_y(\alpha) = \Gamma_y(\alpha) \sum_{k=y}^{+\infty} \frac{\Gamma_0(k+1)}{\Gamma_k(\alpha)\Gamma_{k+1}(\alpha)}.$$

The case $\lambda > 1$

Proposition 2 *The hitting time* T_{0N} *has an exponential moment of order* $N(\sqrt{\lambda}-1)^2$. *Furthermore,*

$$E\left(e^{N(\sqrt{\lambda}-1)^2 T_{0N}}\right) \leq \lambda^{\frac{N}{2}}, \tag{2}$$

and consequently

$$d_N(t) \leq P(T_{0N} > t) \leq \lambda^{N/2} e^{-N(\sqrt{\lambda}-1)^2 t}, \quad t \geq 0. \tag{3}$$

The exponential decay in t is correct.

Upper bound on the time to reach equilibrium:

$$\frac{1}{2}\frac{\log\lambda}{(\sqrt{\lambda}-1)^2} > \log\frac{\lambda}{\lambda-1}.$$

Proposition 3

$$\lim_{N \to +\infty} d_N(t) = \begin{cases} 1 & \textit{if } t < \log \frac{\lambda}{\lambda - 1} \\ \\ 0 & \textit{if } t > \log \frac{\lambda}{\lambda - 1}. \end{cases}$$

Proof: Martingale argument:

$$E(e^{T_{0,N-A}}) \sim \frac{\lambda}{\lambda - 1}, \quad E\left(e^{2T_{0,N-A}}\right) \sim \frac{\lambda^2}{(\lambda - 1)^2 - \frac{1}{N}}.$$

The case $\lambda = 1$

Proposition 4 $T_{0N} - \frac{logN}{2}$ *converges in distribution to a variable with density on* $\mathbb{R}$

$$\psi(x) = \sqrt{\frac{2}{\pi}} e^{-x} e^{-\frac{1}{2} e^{-2x}}.$$

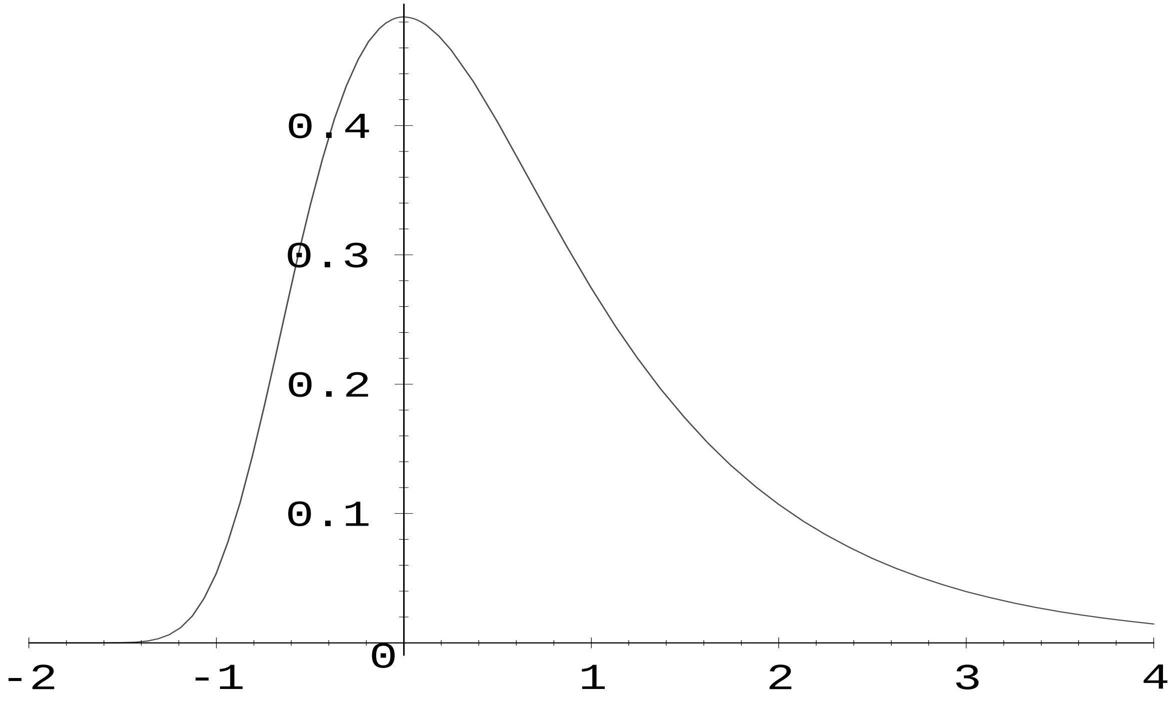


Figure 1: The distribution of $T_N - \frac{log N}{2}$

Proposition 5 *If $\lambda = 1$, then for any sequence $(\phi(N))$ converging to $+\infty$, then*

$$\lim_{N \to +\infty} d_N \left(\frac{\log N}{2} + a\phi(N) \right) = \begin{cases} 1 & \text{if } a < 0 \\ \\ 0 & \text{if } a > 0 \end{cases} ;$$

moreover for any $a \in \mathbb{R}$,

$$\limsup_{N \to +\infty} d_N \left(\frac{\log N}{2} + a \right) \leq \sqrt{\frac{2}{\pi}} \int_0^{e^{-a}} e^{-\frac{x^2}{2}} dx. \tag{4}$$

The case $\lambda < 1$,

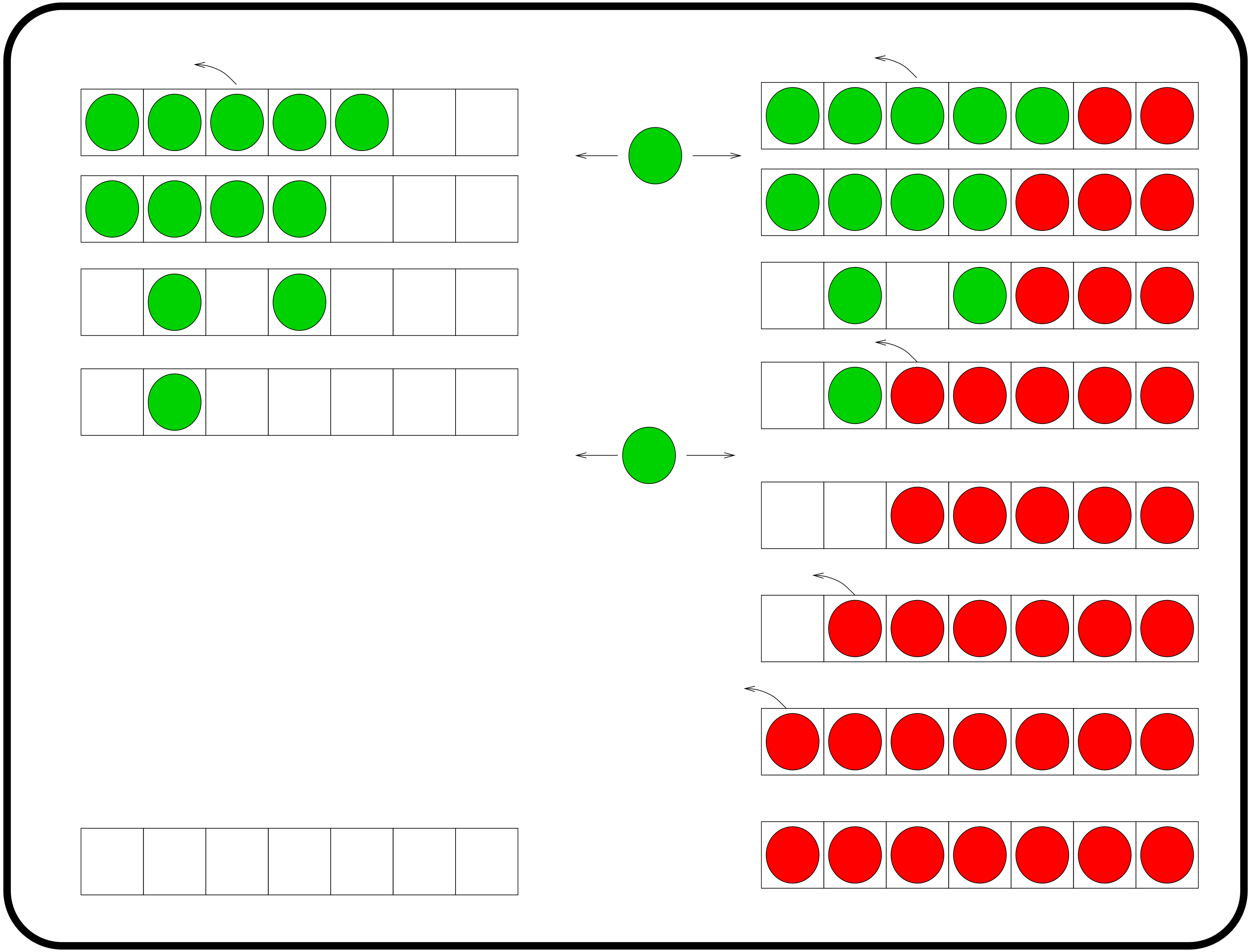
Proposition 6

$$d_N(t) \leq (N+1)e^{-t}, \quad t \geq 0. \tag{5}$$

For any sequence $(\phi(N))$ *such that* $\lim_{N\to+\infty} \phi(N) = +\infty$, *then*

$$\lim_{N\to+\infty} d_N\left(\frac{\log N}{2} + a\phi(N)\right) = \begin{cases} 1 & \textit{if } a < 0 \\ \\ 0 & \textit{if } a > 0. \end{cases}$$

Proof: green and red balls.



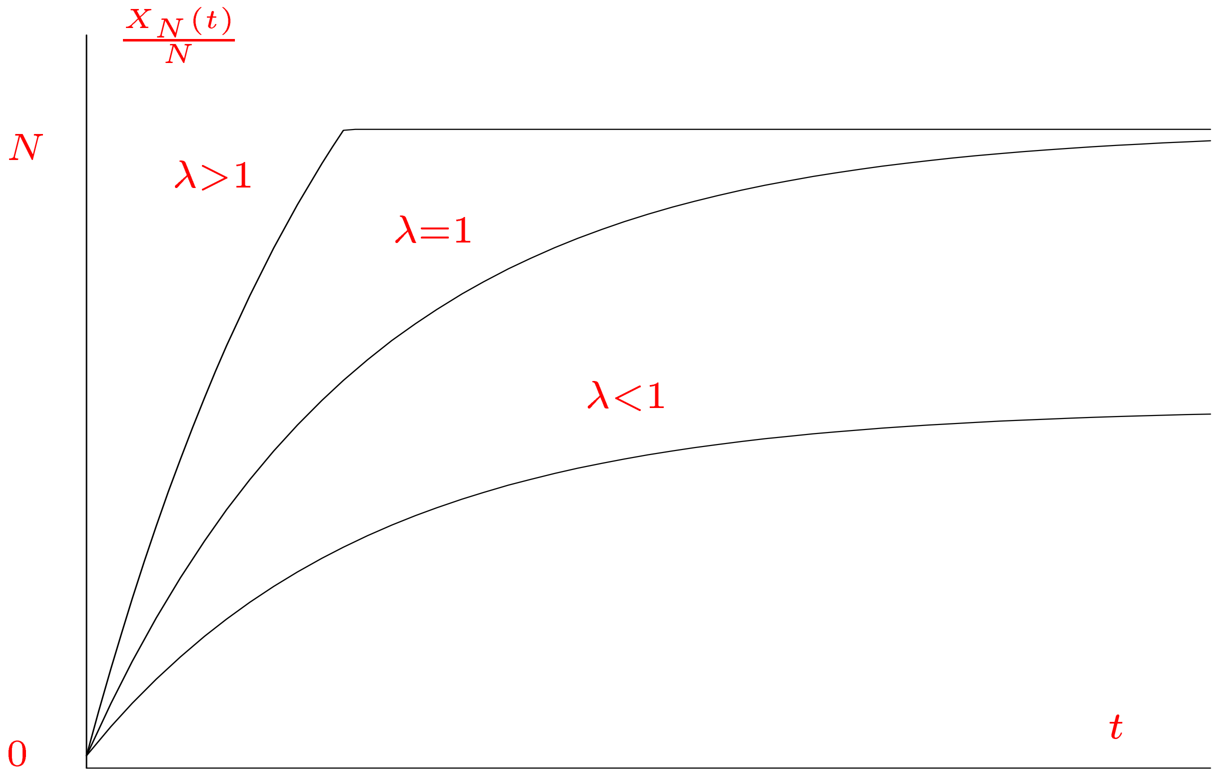


Figure 2: The three different regimes