Large Deviations, Multiplicative Ergodicity and Spectral Theory for Markov Chains

I. Kontoyiannis

Brown University

joint work with S.P. Meyn

University of Illinois/Urbana-Champaign

Joint Appied Mathematics and Probability Seminar

Stanford University, October 2003

Outline

Motivation

Ergodicity and Lyapunov condition (V3)

Multiplicative ergodicity and condition (DV3)

Large deviations

Spectral Theory

(DV3) $\Leftrightarrow$ Finite-dimensional approximation

$\cdots \Rightarrow$ Discrete spectrum

$\cdots \Rightarrow$ Solutions for the multiplicative Poisson eqn

$\cdots \Rightarrow$ Multiplicative ergodic theory

Probabilistic Results

A strong large deviations principle

[The role of entropy]

Exact large deviations asymptotics

An interesting phenomenon

Applications and Extensions

Background: The Question of Ergodicity

Let

$\boldsymbol{X} = \{X_n : n \geq 0\}$ be a Markov chain (MC), with

general state space Σ

transition kernel $P(x, A) = \Pr\{X_n \in A | X_{n-1} = x\}$

$F : \Sigma \to \mathbb{R}$ be a real-valued functional

S_n be the partial sums $S_n := \sum_{i=0}^{n-1} F(X_i)$

Questions

Is $\boldsymbol{X}$ ergodic?

How fast does it converge?

Does the law of large numbers hold?

How "big" a function F can we take in the LLN?

Recall

As any kernel $Q(x, dy)$ acts of functions $F : \Sigma \to \mathbb{R}$ via

$$QF(x) = \int_{\Sigma} F(y) Q(x, dy)$$

we think of any such kernel as a linear operator

An Elegant Criterion

Define

The *generator* of $\boldsymbol{X}$ as $\mathcal{L} := (P - I)$

Lyapunov condition (V3)

$\boldsymbol{X}$ *satisfies condition* (V3) *with Lyapunov function* V if there exist functions $V, W : \Sigma \to [1, \infty)$ & $\delta > 0$ s.t.

$$\mathcal{L}V(x) \leq -\delta W(x) \qquad \text{(V3)}$$

for "most" $x \in \Sigma$

Consequences

If $\boldsymbol{X}$ is irreducible and aperiodic, (V3) implies the existence of a unique invariant measure π s.t.

$$\pi(W) := \int_{\Sigma} W \, d\pi < \infty$$

Moreover, the Lyapunov function V quantifies the speed of convergence, and W the size of functions for the LLN

Ergodic Results

For an irreducible, aperiodic Markov chain $\boldsymbol{X}$
(V3) is essentially *equivalent* to the following

$\rightsquigarrow$ Convergence

$$\sum_{n=1}^{\infty} |P^n(x,A) - \pi(A)| \le (\text{const})V(x)$$

$\rightsquigarrow$ Mean Ergodic Theorem

for all F s.t. $|F| \le (\text{const})W$
and all $X_0 = x \in \Sigma$

$$\frac{1}{n}\mathsf{E}_x[S_n] \to \pi(F) := \int F d\pi \quad \text{as } n \to \infty \qquad \text{(MET)}$$

$\rightsquigarrow$ Moreover

$$\widehat{F}(x) = \lim_n \mathsf{E}_x\big[S_n - n\pi(F)\big]$$

$\rightsquigarrow$ Poisson Equation

the function $\widehat{F}$ solves

$$\mathcal{L}\widehat{F} = -[F - \pi(F)] \qquad \text{(PE)}$$

Goals

Linear structure: (V3) and the above results can be used as the basis for the ergodic theory of MCs They are based on the semigroup of linear operators $\{P^n\}$

$$P^n : F(x) \mapsto \mathsf{E}_x[F(X_n)]$$

and its generator $\mathcal{L} := (P - I)$

Here we consider: For $\alpha \in \mathbb{C}$, the partial products

$$\prod_{i=0}^{n-1} e^{\alpha F(X_i)} \ = \ e^{\alpha S_n}$$

Goal: Provide optimal or near-optimal conditions for *multiplicative* analogs of these ergodic results

Starting point: Consider the (nonlinear) operators

$$Q_n : F(x) \mapsto \log \mathsf{E}_x[e^{F(X_n)}]$$

and the *nonlinear generator*

$$\mathcal{N} := (Q_1 - I)$$

A New Lyapunov Condition

Nonlinear generator

Since $Q_1F(x) = \log \mathsf{E}_x[e^{F(X_1)}] = \log[Pe^F(x)]$
we have

$$\mathcal{N}F = [Q_1 - I]F = \log(Pe^F) - F = \log(e^{-F}Pe^F)$$

Recall (V3)

There exist V, W, and $\delta > 0$:

$$\mathcal{L}V(x) \leq -\delta W(x) \quad \text{for "most" } x \in \Sigma \qquad \text{(V3)}$$

New Lyapunov condition

$\boldsymbol{X}$ *satisfies condition* (DV3) *with Lyapunov function* V
if there exist functions $V, W : \Sigma \to [1, \infty)$ & $\delta > 0$ s.t.

$$\mathcal{N}V(x) \leq -\delta W(x) \quad \text{for "most" } x \in \Sigma \qquad \text{(DV3)}$$

The Nonlinear Generator in Cont's Time

Analogously

Instead of the semigroup of linear operators $\{P^t\}$

$$P^t : F(x) \mapsto \mathsf{E}_x[F(X_t)]$$

and its generator $\mathcal{L} :=$ "$\lim_{t\downarrow 0} \frac{1}{t}[P^t - I]$"

we consider the nonlinear operators $\{Q_t\}$

$$Q_t : F(x) \mapsto \log \mathsf{E}_x[e^{F(X_t)}]$$

and their generator $\mathcal{N} :=$ "$\lim_{t\downarrow 0} \frac{1}{t}[Q_t - I]$"

Formally

Since $Q_t F = \log(P^t e^F)$, we can calculate

$$\begin{aligned} [Q_t - I]F &= \log(P^t e^F) - F \\ &= \log(e^{-F} P^t e^F) \\ &\approx e^{-F} P^t e^F - I \\ &= e^{-F}[P^t - I]e^F \end{aligned}$$

therefore,

$$\mathcal{N}F = \text{“}\lim_{t\downarrow 0} \frac{1}{t}[Q_t F - F] = e^{-F}\mathcal{L}e^F\text{”}$$

and we **define the nonlinear generator** by

$$\mathcal{N}F = e^{-F}\mathcal{L}e^F$$

Multiplicative Ergodic Results

For an irreducible, aperiodic Markov chain $\boldsymbol{X}$ (DV3) is essentially *equivalent* to the following

$\rightsquigarrow$ Exponential Convergence ['easy']

for some $\rho > 1$

$$\sum_{n=1}^{\infty} \rho^n |P^n(x, A) - \pi(A)| \le \text{(const)} V(x)$$

$\rightsquigarrow$ Multiplicative Mean Ergodic Theorem

for all F s.t. $|F| \le$ (const)W
and all $X_0 = x \in \Sigma$

$$\tfrac{1}{n} \log \mathsf{E}_x[e^{S_n}] \to \Lambda(F) \quad \text{as } n \to \infty \qquad \text{(MMET)}$$

$\rightsquigarrow$ Moreover

$$e^{\widetilde{F}(x)} = \lim_n \ \mathsf{E}_x\left[e^{S_n}\right] e^{-n\Lambda(F)}$$

$\rightsquigarrow$ Multiplicative Poisson Equation

the function $\widetilde{F}$ solves

$$\mathcal{N}\widetilde{F} = -[F - \Lambda(F)] \qquad \text{(MPE)}$$

Probabilistic Results

The above convergence

$$\mathsf{E}_x\left[e^{S_n}\right] e^{-n\Lambda(F)} \to e^{\widetilde{F}(x)} \qquad (*)$$

gives very precise information about the asymptotics of the moment generating fns $m_n(\alpha) := \mathsf{E}_x\left[e^{\alpha S_n}\right]$

i. Large Deviations

$(*)$ leads to a full large deviations principle (LDP) for the empirical measures of $\boldsymbol{X}$ in a topology that includes unbounded functions F

ii. Exact Large Deviations

This LDP can be refined to an expansion *a la* Bahadur-Rao

$$\mathsf{P}_x\{S_n \geq nc\} \sim \frac{C(x)}{\sqrt{n}} e^{-nJ(c)}$$

for $c > \pi(F)$

Basic Assumptions

Let

$\boldsymbol{X} = \{X_n\}$ be a MC on the Polish state space $(\Sigma, \mathcal{S})$

$P(x, A) := \mathsf{P}_x\{X_1 \in A\} := \Pr\{X_n \in A | X_{n-1} = x\}$
be its transition kernel

ψ-irreducibility and aperiodicity

Assume that there exists σ-finite measure ψ on $(\Sigma, \mathcal{S})$
such that $\quad P^n(x, A) > 0 \quad$ eventually
for any $x \in \Sigma$ and any $A \in \mathcal{S}$ with $\psi(A) > 0$

$\boldsymbol{X}$ is multiplicatively regular

if there exists a *Lyapunov function* $V : \Sigma \to [1, \infty)$
and also $W : \Sigma \to [1, \infty)$, $\delta > 0$, $b < \infty$ satisfying:

$$\mathcal{N}V \leq \left\{ \begin{array}{ll} -\delta W & \text{on } C^c \\ \\ -\delta W + b & \text{on } C \end{array} \right\} \quad \text{(DV3)}$$

for a "small" set C

[$\sum_{n\geq 0} e^{-n} P^n(x, A) \geq \nu(A)$ for some positive measure ν, all $x \in C$, all $A \in \mathcal{S}$]

Summary of Assumptions

The following are assumed throughout

$\boldsymbol{X}$ is ψ-irreducible, aperiodic

$\boldsymbol{X}$ is multiplicatively regular with Lyapunov fn V and invariant measure π

A strengthened (DV3):

(DV3+)
- **i.** (DV3) holds with an unbounded W
- **ii.** For some T_0, all r there is a measure β_r:
  $$\mathsf{P}_x\{\Phi(T_0) \in A,\ \ \tau_{C_W(r)} > T_0\} \le \beta_r(A)$$
  for all $x \in C_W(r) := \{W \le r\}$, all A

Note

Part ii. of (DV3+) is a mild continuity assumption, weaker than the Feller property but very important (necessary?) for what follows

Comments on Multiplicative Regularity

Strength

(DV3+) is *weaker* than:

- Donsker-Varadhan conditions
- Stroock's 'uniform' assumption and its variants
- Most (all?) common LDP assumptions

Appeal of (DV3)

Easily verifiable!

Leads to *computable bounds* in specific cases

Satisfied by many important examples

Examples of (DV3+) Chains

1. Discrete-time Ornstein-Uhlenbeck process

$$X_{n+1} - X_n = -\beta X_n + \sigma Z_n\,; \qquad Z_n \sim \text{IID } N(0,1)$$

$\boldsymbol{X}$ satisfies all the above conditions with

ψ=Lebesgue measure on $\mathbb{R}$

$V(x) = 1 + \frac{x^2}{2\sigma^2}$

$W(x) = 1 + \beta^2 x^2$

2. More general discrete-time diffusions

For any 'nice' C^2 potential $H : \mathbb{R} \to (0, \infty)$

$$X_{n+1} - X_n = -H'(X_n) + \sigma Z_n\,; \qquad Z_n \sim \text{IID } N(0,1)$$

$\boldsymbol{X}$ satisfies all the above conditions with

ψ=Lebesgue measure on $\mathbb{R}$

$V(x) = 1 + \frac{H(x)}{\sigma^2}$

$W(x) = 1 + [H'(x)]^2$

3. Many hypoelliptic diffusions in $\mathbb{R}^n$

Main Multiplicative Ergodicity Results

Theorem

Under (DV3+), for any $F : \Sigma \to \mathbb{R}$ s.t. $|F| \le (\text{const})W$:

i. The Multiplicative Poisson Equation

$$\mathcal{N}\widetilde{F} = -[F - \Lambda(F)] \qquad \text{(MPE)}$$

has a "maximal" solution pair $(\widetilde{F}, \Lambda(F))$, where

ii. $\Lambda(F)$ is "analytic," strictly convex
and for all $x \in \Sigma$ it satisfies the
Multiplicative Mean Ergodic Theorem

$$\frac{1}{n} \log \mathsf{E}_x[e^{S_n}] \to \Lambda(F) \quad \text{as } n \to \infty \quad \text{(MMET)}$$

iii. $\widetilde{F}$ satisfies, for some $B_0 < \infty$, $b_0 > 0$,

$$\left| \mathsf{E}_x\left[e^{\alpha S_n - n\Lambda(\alpha F)}\right] - e^{(\widetilde{\alpha F})(x)} \right| \le V(x)|\alpha|B_0 e^{-b_0 n}$$

uniformly over all $x \in \Sigma$, $n \ge 1$
and over $\alpha \in \mathbb{C}$, F in "short compacts"

The Proof Setting

Consider

For an arbitrary 'weighting' function $U : \Sigma \to [1, \infty)$,

$$L_\infty^U := \left\{ G : \Sigma \to \mathbb{R} \text{ s.t. } \|G\|_U := \sup_x \frac{|G(x)|}{U(x)} < \infty \right\}$$

so that in the theorem we consider all $F \in L_\infty^W$

Notation

Write $\|\!|\cdot|\!\|_U$ for the induced operator norm on L_∞^U:
For a linear operator Q on L_∞^U

$$\|\!|Q|\!\|_U := \sup_{G \in L_\infty^U,\, G \neq 0} \frac{\|QG\|_U}{\|G\|_U}$$

Proof Outline

1. State the MPE as an eigenvalue problem for the operators $P_F(x,dy) := e^{F(x)}P(x,dy)$, $F \in L_\infty^W$ on an appropriate space L_∞^U

2. If a kernel P satisfies (DV3+), then it can be approximated *in* L_∞^U*-norm* by a finite-rank operator

3. *All* of the operators P_F can be similarly approximated in norm

4. From 3. $\Rightarrow$ each P_F has discrete spectrum in L_∞^U

5. From 4. we obtain maximal solutions to the MPE equipped with a spectral gap $\Rightarrow$ **i.**

6. a. Write the potential operator $U_z = [Iz - (P_F - \mathbb{I}_C \otimes \nu)]^{-1}$ as an inverse, via Nummelin's inversion formula

$$[Iz - Q]^{-1} = [Iz - (Q - \mathbb{I}_C \otimes \nu)]^{-1}\left(I + \frac{1}{1-\beta}\mathbb{I}_C \otimes \nu\right)$$

b. Expand U_z in a power series

c. Derive bounds on the norm of U_z from the contraction ineq (DV3)

d. Interpreting the bounds $\Rightarrow$ MMET

1. The MEP as an Eigenvalue Problem

Expand the MPE:

$$\begin{aligned}
\mathcal{N}\widetilde{F} &= -[F - \Lambda(F)] \\
\log(e^{-\widetilde{F}} P e^{\widetilde{F}}) &= \Lambda(F) - F \\
e^{-\widetilde{F}} P e^{\widetilde{F}} &= e^{\Lambda(F)} e^{-F} \\
e^{F} P e^{\widetilde{F}} &= e^{\Lambda(F)} e^{\widetilde{F}}
\end{aligned}$$

Defining the new kernel $\boxed{P_{_F}(x,dy) := e^{F(x)} P(x,dy)}$

$\rightsquigarrow$ the eigenvalue equation $\boxed{P_{_F} e^{\widetilde{F}} = e^{\Lambda(F)} e^{\widetilde{F}}}$

From now on

We consider the MPE in L_∞^U

w.r.t. the weighting function $U := e^V$

2. (DV3+) $\rightsquigarrow$ Finite-Rank Approximation

Assume for simplicity

$\Sigma = \mathbb{R}^n$ (and $\psi =$ Lebesgue measure)

(DV3+) holds with V, W continuous

W unbounded w/ compact $C_W(r) := \{x : W(x) \le r\}$

P has 'nice' densities $P(x, dy) = p(x, y)\psi(dy)$

Then

We can approximate

$$\begin{aligned} P(x, dy) &= p(x, y)\psi(dy) \\ &\approx \Big[\sum_{i=1}^{k} s_i(x)p_i(y)\Big]\psi(dy) \\ &= \sum_{i=1}^{k} s_i(x)\nu_i(dy) \ =: \ R(x, dy) \end{aligned}$$

Moreover

a. $R(x, dy) := \sum_{i=1}^{k} s_i(x)\nu_i(dy)$ is a transition kernel

b. $R \approx P$ in that $|||P - R|||_{U} < \epsilon$

c. $\sup_{n\ge1} |||P^n - R^n|||_{U} < \epsilon$

2. Cont'd: Interpretation of R

$$\boxed{R(x,dy) := \sum_{i=1}^{k} s_i(x)\nu_i(dy)}$$

The MC $\boldsymbol{Y} \sim \boldsymbol{R(x,dy)}$ is a finite-state HMM!

A. Define a finite-state MC $\boldsymbol{Z} = \{Z_n\}$ on $\{1,2,\ldots,k\}$ via

$$\begin{cases} \Pr\{Z_n = j \mid Z_{n-1} = i\} = \int s_j \, d\nu_i \\ \\ \Pr\{Z_0 = j\} = s_j(x) \end{cases}$$

$$\begin{array}{cccccccccc} & Z_0 & - & Z_1 & \cdots\cdots & Z_{n-2} & - & Z_{n-1} & \cdots \\ & | & & | & & | & & | & \\ Y_0 = x & Y_1 & & Y_2 & \cdots\cdots & Y_{n-1} & & Y_n & \cdots \end{array}$$

B. From this, define a HMM $\boldsymbol{Y} \sim R$ on Σ via

$$\begin{cases} Y_n \,\Big|\, Y_0,\ldots,Y_{n-1},Z_0,\ldots,Z_{n-1} \sim \nu_{Z_{n-1}} \\ Y_0 = x \end{cases}$$

Rest of the Proof

3. The continuity part of (DV3+)

$$\mathsf{P}_x\{\Phi(T_0) \in A, \ \ \tau_{C_W(r)} > T_0\} \le \beta_r(A)$$

guarantees the existence of densities for *all* $P_{_F}^{T_0+1}$
Similar arguments show that the $P_{_F}$ can be also approximated in norm by finite-rank operators

4. From 3. $\Rightarrow$ each $P_{_F}$ has discrete spectrum in L_∞^U

5. From 4. $\Rightarrow$ maximal solutions to the MPE equipped with a spectral gap

6. From 5. $\Rightarrow$ the precise MMET

$$\Bigl|\mathsf{E}_x\bigl[e^{\alpha S_n - n\Lambda(\alpha F)}\bigr] - e^{(\widetilde{\alpha F})(x)}\Bigr| \ \le \ V(x)|\alpha|B_0 e^{-b_0 n}$$

via Perron-Frobenius arguments
and tools from potential theory

QED

Probabilistic Results

- Under all the earlier assumptions, choose $W_0 \in L_\infty^W$ which grows to infinity slower than W

$$\lim_{r\to\infty} \sup_{x:W(x)>r} \frac{W_0(x)}{W(x)} = 0$$

- On the space of all prob measures ν on Σ define the τ_{W_0}-topology by

$$\nu_n \to \nu \quad \text{iff} \quad \int F d\nu_n \to \int F d\nu \ \text{ for all } F \in L_\infty^{W_0}$$

Theorem: Large Deviations Principle

The empirical measures of $\boldsymbol{X}$ satisfy a large deviations principle in the space of prob measures on Σ equipped with the τ_{W_0}-topology, and with respect to the good, convex rate function

$$I(\nu) := \inf_{\mu} H(\mu(dx,dy)\|\nu(dx)P(x,dy))$$

where the infimum is over all bivariate measures μ both of whose marginals $= \nu$, and $H =$ relative entropy

Proof Ideas

Proof

Step 1. LDP:
The simple MMET

$$\frac{1}{n}\log \mathsf{E}_x\left[e^{S_n}\right] \to \Lambda(F)$$

combined with the smoothness of $\Lambda(\cdot)$ and the Dawson-Gärtner projective limit theorem give the LDP with rate function

$$\Lambda^*(\phi) := \sup_{F\in L_\infty^{W_0}} \Big[(F,\phi) - \Lambda(F)\Big]$$

on the space of all linear functionals ϕ on $L_\infty^{W_0}$

Step 2. Restrict to prob measures:
Use finiteness and continuity

Step 3. Entropy:
Identify the rate function in terms of entropy, by passing to the bivariate chain, and a lot of hard work... $\square$

Exact Large Deviations Asymptotics

Still under the same assumptions, the partial sums $S_n = \sum_{i=0}^{n-1} F(X_i)$ of any $F \in L_\infty^{W_0}$ satisfy:

Theorem: Exact Large Deviations

If $F \in L_\infty^{W_0}$ is non-lattice,
for all $c > \pi(F)$, all $x \in \Sigma$

$$\mathsf{P}_x\{S_n \geq nc\} \;\sim\; \frac{e^{(\widetilde{aF})(x)}}{a\sqrt{2\pi n\sigma^2}} e^{-nJ(c)}$$

where a is chosen s.t. $\frac{d}{da}\Lambda(aF) = c, \;\; \sigma^2 := \frac{d^2}{da^2}\Lambda(aF)$
and $J(c) = \inf\{I(\nu) \;:\; \nu(F) > c\}$

[A corresponding result holds for the lower tail]

Notes

Dependence on initial condition x
is only via the solution to the MPE $(\widetilde{aF})(x)$

An analogous expansion exists for lattice functions F

Proof:
Change of measure: $e^{-\Lambda(\alpha)} e^{-(\widetilde{aF})(x)} e^{\alpha F(x)} P(x, dy) e^{(\widetilde{aF})(y)}$
$+$ Edgeworth expansion in the exponent! $\square$

An Interesting Phenomenon

Example

Consider the following discrete-time Ornstein-Uhlenbeck process in $\mathbb{R}^2$

$$X_{n+1} = \begin{pmatrix} 0 & 1 \\ -a_2 & -a_1 \end{pmatrix} X_n + \begin{pmatrix} 0 \\ Z_{n+1} \end{pmatrix} \qquad Z_n \sim \text{IID } N(0,1)$$

For appropriate a_1, a_2, this MC satisfies (DV3+) with Lyapunov function

$$V(x) = 1 + x^T M x \quad \text{and} \quad W = V$$

for some positive definite matrix M

Consider the function

$$F(x) = F\begin{pmatrix} x^{(1)} \\ x^{(2)} \end{pmatrix} = \mathbb{I}_{\{|x^{(1)}|<1\}} + x^{(2)} - x^{(1)}$$

$\rightsquigarrow$ although F has full support:

$\pi\{F > c\} > 0$ for all c

$\rightsquigarrow$ for $c > 1$, $\Pr\{S_n > nc\}$ decays *super*-exponentially fast:
$J(c) = \infty$ for $c > 1$

Example Details

Let

$$A = \begin{bmatrix} 0 & 1 \\ -a_2 & -a_1 \end{bmatrix}$$

and assume that the roots of the quadratic equation $z^2 + a_1 z + a_2 = 0$ lie within the open unit disk in $\mathbb{C}$

Note that there exist $\gamma < 1$ and a positive definite matrix N satisfying

$$A^T N A \leq \gamma I$$

E.g., take

$$N = \sum_{k=0}^{\infty} \gamma^{-k} (A^k)^T A^k$$

with $\gamma < 1$ chosen so that the sum is convergent

Then, take $M = \epsilon N$ for ϵ small enough

Conclusions

1. So Far

Under (DV3+), a near-complete picture of:

Multiplicative ergodic theory
Large deviations asymptotics

Structure theory:

Multiplicative regularity $\Leftrightarrow$ discrete spectrum
$\Leftrightarrow$ full multiplicative mean ergodic theorem

Intuitively:

$\boldsymbol{X}$ is a finite-state MC in disguise $\Leftrightarrow$ (DV3+)

2. Extensions

Continuous time

Random matrices

3. Current Work/Applications

Stability of linear systems: e.g., LMS algorithm

Time-varying CDMA networks

Metastability in physical models

Nonlinear dynamical systems

References

The above results above can be found in:

[1] I.K. & S.P. Meyn.
"Spectral theory and limit theorems
for geometrically ergodic Markov processes"
Annals Appl Probab, February 2003

[2] I.K. & S.P. Meyn.
"Large deviations asymptotics and the spectral theory
of multiplicatively regular Markov processes"
Preprint, August 2003

These and a number of related results can be found at

`www.dam.brown.edu/people/yiannis`