Markov chains in stationary and ergodic random environments [1]

Attila Lovas[1,2], Miklós Rásonyi[1]

[1] Alfréd Rényi Institute of Mathematics
[2] BUTE, Department of Analysis

8th European Congress of Mathematics

20–26 June 2021, Portorož, Slovenia

[1] Both authors benefited from the support of the "Lendület" grant LP 2015-6 of the Hungarian Academy of Sciences. The second author was also supported by the NKFIH (National Research, Development and Innovation Office, Hungary) grant KH 126505.

1. **Markov chains in random environments**
   - The mathematical setting
   - Assumptions
   - The main result
   - Generalizations
2. **An overview of the proof**
   - The big picture
   - A coupling argument for the stability of Law(X_N)
   - L-mixing approach to the Law of Large Numbers
3. **Applications**
   - Linear systems in random environments
   - Stochastic gradient Langevin algorithm
   - Queuing with dependent inter-arrival times

The mathematical setting

We fix a standard Borel space $(\mathcal{X}, \mathcal{B})$. The set of probability Borel measures on $(\mathcal{X}, \mathcal{B})$ are denoted by $\mathcal{M}_1$. The total variation metric on $\mathcal{M}_1$ is defined by

$$d_{\mathsf{TV}}(\mu_1, \mu_2) = |\mu_1 - \mu_2|(\mathcal{X}), \quad \mu_1, \mu_2 \in \mathcal{M}_1.$$

Let $(\mathcal{Y}, \mathcal{A})$ be a measurable space and $Q : \mathcal{Y} \times \mathcal{X} \times \mathcal{B} \to [0, 1]$ is a parametric family of stochastic kernels:

1. for all $B \in \mathcal{B}$ the function $(y, x) \mapsto Q(y, x, B)$ is $\mathcal{A} \otimes \mathcal{B}$-measurable,
2. for all $(y, x) \in \mathcal{Y} \times \mathcal{X}$, $B \mapsto Q(y, x, B)$ is a probability measure on $\mathcal{B}$.

$Y : \mathbb{Z} \times \Omega \to \mathcal{Y}$ a *strongly stationary* stochastic process being interpreted as *the environment* which influences the evolution our main process of interest.

We assume that we are given the $\mathcal{X}$-valued process X_t, $t \in \mathbb{N}$ such that $X_0 = x_0 \in \mathcal{X}$ is fixed and

$$\mathbb{P}(X_{t+1} \in B \mid \mathcal{F}_t) = Q(Y_t, X_t, B) \quad \mathbb{P}-\text{a.s.},\ t \in \mathbb{N},$$

where the filtration is

$$\mathcal{F}_t = \sigma(X_s,\ 0 \le s \le t;\ Y_s,\ s \le t),\ t \in \mathbb{N}.$$

Let $\mu_t \in \mathcal{M}_1$ denote the law of X_t for $t \in \mathbb{N}$.

We aim to study...

- the convergence of μ_t to a limiting law as $t \to \infty$, and
- the ergodic properties of X_t.

...under various assumptions.

Assumptions

Definition

We define the action of a parametric kernel $Q : \mathcal{Y} \times \mathcal{X} \times \mathcal{B} \to [0, 1]$ on a bounded measurable function $\phi : \mathcal{X} \to \mathbb{R}$ as

$$[Q(y)\phi](x) = \int_{\mathcal{X}} \phi(z)\, Q(y, x, \mathrm{d}z),\ y \in \mathcal{Y},\ x \in \mathcal{X}.$$

This definition makes sense for any non-negative measurable ϕ, too.

The structure of our assumptions will be "A) + B)", where

A) is a parametric set of conditions on the dynamics, where the parameter represents the actual state of the environment,

B) is a condition controlling the environment.

Foster-Lyapunov type drift condition with long-time contractivity

Assumption

A) Let $V : \mathcal{X} \to \mathbb{R}_+$ be a measurable function. We assume that there are measurable functions $K, \gamma : \mathcal{Y} \to [1, \infty)$ such that, for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$,

$$[Q(y)V](x) \leq \gamma(y)V(x) + K(y).$$

B) We assume that

$$\bar{\gamma} := \limsup_{n\to\infty} \mathbb{E}^{1/n}\left(K(Y_0)\prod_{k=1}^{n}\gamma(Y_k)\right) < 1.$$

Minorization on niveau sets

Assumption

A) We assume that for some $0 < \varepsilon < 1/\bar{\gamma}^{1/2} - 1$, *there is a measurable function* $\alpha : \mathcal{Y} \to [0,1)$ *and a probability kernel* $\kappa : \mathcal{Y} \times \mathcal{B} \to [0,1]$ *such that, for all* $y \in \mathcal{Y}$ *and* $A \in \mathcal{B}$,

$$\inf_{x \in V^{-1}([0,R(y)])} Q(y,x,A) \geq (1-\alpha(y))\kappa(y,A),$$

where $R(y) = \frac{2K(y)}{\varepsilon\gamma(y)}$ *and* $V^{-1}([0,R(y)]) \neq \emptyset$.

B) There exists $0 < \theta < 1$ *such that*

$$\lim_{n\to\infty} \mathbb{E}^{1/n^{\theta}} \left(\alpha(Y_0)^n \right) = 0.$$

The main result

Theorem (Convergence to stationary distribution)

Under our standing assumptions, there exists a universal probability law μ_*, independent of x_0, such that

$$\mathit{Law}(X_N) \xrightarrow{d_{TV}} \mu_*, \text{ as } N \to \infty.$$

More precisely, for any $1/2 < \lambda < 1$, there exist $c(\lambda), \nu(\lambda) > 0$ such that

$$d_{TV}(\mu_N, \mu_*) \le 2 \sum_{n=N}^{\infty} \left[\mathbb{E}\left(\max_{0 \le k < \lfloor n^{1/3} \rfloor^3} \alpha(Y_k) \right)^{\lfloor n^{1/3} \rfloor - 1} + \bar{\gamma}^{\left(\lambda - \frac{1}{2}\right)\lfloor n^{1/3} \rfloor^2 - n^{1/3}} + c n^{2/3} e^{-\nu n^{1/3}} \right]$$

holds for all $N \in \mathbb{N}$.

Remark

An important particular case is when $\alpha(\cdot)$ is constant, then we have

$$d_{TV}(\mu_N, \mu_*) = \mathcal{O}(e^{-\zeta n^{1/3}})$$

for some $\zeta > 0$.

Let $S : \mathcal{Y}^{\mathbb{Z}} \to \mathcal{Y}^{\mathbb{Z}}$ be the left shift $(S\mathbf{y})_j = \mathbf{y}_{j+1}$, $j \in \mathbb{Z}$.

The environment is stationary $\Rightarrow$ Law(Y) is S-invariant.

$(\mathcal{Y}^{\mathbb{Z}}, \mathcal{A}^{\otimes\mathbb{Z}}, \mathrm{Law}(Y), S)$ forms a dynamical system. We say that an event $A \in \mathcal{A}^{\otimes\mathbb{Z}}$ is invariant iff $S^{-1}(A) = A$ and the process Y is *ergodic* if the σ-algebra generated by invariant events $\mathcal{I}$, is trivial for Law(Y) i.e. for all $A \in \mathcal{I}$, $\mathrm{Law}(Y)(A) \in \{0, 1\}$.

Theorem (A law of large numbers)

Under our assumptions, if $(Y)_{t\in\mathbb{Z}}$ is ergodic, then for any bounded and measurable $\Phi : \mathcal{X} \to \mathbb{R}$

$$\frac{\Phi(X_1) + \ldots + \Phi(X_N)}{N} \to \int_{\mathcal{X}} \Phi(z)\, \mu_*(\mathrm{d}z),\ N \to \infty$$

holds in L^p, for all $1 \le p < \infty$, where μ_* is as in the previous theorem.

Generalizations

Many stochastic models, including all discretized diffusions (as well as discretely sampled diffusions), fall within the scope of our framework.

Why we need to generalize these results?

- A broad class of processes are not covered in this way.
- For instace, vector autoregressive models are failed to satisfy the one-step drift and minorization conditions.

To alleviate this issue and thus making these techniques applicable to a larger class of models, we generalize our theorems.

We require the drift and minorization conditions only over several steps and with constant additive term.

Generalizations

Many stochastic models, including all discretized diffusions (as well as discretely sampled diffusions), fall within the scope of our framework.

Why we need to generalize these results?

- A broad class of processes are not covered in this way.
- For instace, vector autoregressive models are failed to satisfy the one-step drift and minorization conditions.

To alleviate this issue and thus making these techniques applicable to a larger class of models, we generalize our theorems.

We require the drift and minorization conditions only over several steps and with constant additive term.

Generalizations

Many stochastic models, including all discretized diffusions (as well as discretely sampled diffusions), fall within the scope of our framework.

Why we need to generalize these results?

- A broad class of processes are not covered in this way.
- For instace, vector autoregressive models are failed to satisfy the one-step drift and minorization conditions.

To alleviate this issue and thus making these techniques applicable to a larger class of models, we generalize our theorems.

We require the drift and minorization conditions only over several steps and with constant additive term.

The p-step drift and contractivity condition

Assumption

A) Let $V : \mathcal{X} \to \mathbb{R}_+$ be a measurable function. We assume that there is an integer $p \geq 1$, a measurable function $\gamma : \mathcal{Y}^p \to (0, \infty)$ and a constant $K \geq 1$ such that, for all $x \in \mathcal{X}$ and $(y_1, \ldots, y_p) \in \mathcal{Y}^p$,

$$[Q(y_p)Q(y_{p-1}) \ldots Q(y_1)V](x) \leq \gamma(y_1, \ldots, y_p)V(x) + K$$

and

$$[Q(y)V](x) \leq KV(x) + K.$$

B) We assume that

$$\bar{\gamma} := \limsup_{n\to\infty} E^{1/n}\left(\prod_{i=1}^{n} \gamma(Y_{(i-1)p+1}, \ldots, Y_{ip})\right) < 1.$$

Assumption (The p-step minorization condition)

A) We assume that for some $0 < \epsilon < 1/\bar{\gamma}^{1/2} - 1$, there is a constant $0 \leq \alpha < 1$ and a probability κ such that, for all $(y_1, \ldots, y_p) \in \mathcal{Y}^p$ and $A \in \mathcal{B}(\mathcal{X})$,

$$\inf_{x \in V^{-1}([0, R(y_1, \ldots, y_p)])} [Q(y_p)Q(y_{p-1}) \ldots Q(y_1)](x, A) \geq (1 - \alpha)\kappa(A), \tag{1}$$

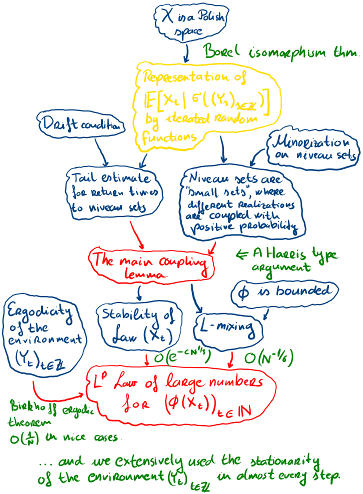
where $R(y_1, \ldots, y_p) = \frac{2K}{\epsilon\gamma(y_1, \ldots, y_p)}$ and we require $V^{-1}([0, R(y_1, \ldots, y_p)]) \neq \emptyset$.

Theorem

Under the new assumptions above, Law(X_N) converges to a limiting law, independent of x_0, at speed $e^{-cN^{1/3}}$.

Furthermore, if $(Y_t)_{t \in \mathbb{Z}}$ is ergodic, then the law of large numbers holds as stated above.

The big picture



Representation by IRFs:

There exists $T_t : \mathcal{Y} \times \mathcal{X} \times \Omega \to \mathcal{X}$, $t \in \mathbb{Z}$ such that

$$\mathbb{E}(X_t \mid \mathbf{Y} = \mathbf{y}) = T_t(y_t) \circ \ldots T_1(y_1)x_0 \stackrel{\text{def}}{=} Z_{0,t}^{x_0,\mathbf{y}}$$

$T_t(y,\cdot,\cdot)$ is *constant* on $\underbrace{\overset{-1}{V}([0,R(y)])}_{\text{small set}}$ with probability at least $1-\alpha(y) \Rightarrow$ coupling effect.

Furthermore, the sigma-algebras $\sigma(T_t(y,x,\cdot),\ x\in\mathcal{X},\ y\in\mathcal{Y})$, $t\in\mathbb{Z}$ are independent.

A coupling argument for the stability of Law(X_N)

Let us start the process from two different initialization x_0, x_0'. The time interval $0, \dots, T$ is cut into disjoint pieces of size $T^{2/3}$ (there are $T^{1/3}$ such pieces).

Idea

Either the "small sets" are visited at least $T^{1/3}$ times or there is a whole piece of size $T^{2/3}$ without visits.

- In the first case, the minorization condition guarantees couplings of the two trajectories outside a set of probability of the order $\max_{0\le k<T} \alpha(y_k)^{T^{1/3}-1}$ $(= \mathcal{O}(e^{-cT^{1/3}})$ if α is constant).
- In the other case the drift condition guarantees that a contractive effect acts for at least $T^{2/3}$ steps. As no returns to the small set occur during this interval, the latter probability is also small.

An extension of this argument guarantees the Cauchy property for the sequence $\mathcal{L}(X_n)$ in total variation norm, which proves convergence, even with a rate.

L-mixing approach to the Law of Large Numbers

Let $(\mathcal{G}_t)_{t\in\mathbb{N}}$, be an increasing, and $(\mathcal{G}_t^+)_{t\in\mathbb{N}}$, be a decreasing sequence of sigma-algebras such that, for each $t\in\mathbb{N}$, $\mathcal{G}_t$ is independent of $\mathcal{G}_t^+$.
For some $p\geq 1$, the process $(W_t)_{t\in\mathbb{N}}$ is called *L-mixing of order p* with respect to $(\mathcal{G}_t,\mathcal{G}_t^+)_{t\in\mathbb{N}}$, if it is adapted to $(\mathcal{G}_t)_{t\in\mathbb{N}}$, and

$$M_p(W)=\sup_{t\in\mathbb{N}}\|W_t\|_p<\infty,$$

$$\Gamma_p(W)=\sum_{\tau=0}^{\infty}\sup_{t\geq\tau}\|W_t-E(W_t|\mathcal{G}_{t-\tau}^+)\|<\infty.$$

Lemma (L. Gerencsér, 1989)

For an L-mixing process W of order $p\geq 2$ satisfying $\mathbb{E}(W_t)=0$, $t\in\mathbb{N}$,

$$\left\|\frac{1}{N}\sum_{i=1}^{N}W_i\right\|_p\leq\frac{C_pM_p^{1/2}(W)\Gamma_p^{1/2}(W)}{\sqrt{N}},\quad N\geq 1$$

with a constant C_p that does not depend either on N or on W.

In the very last step of the proof of LLN,

by the triangle inequality, we can write

$$\left\| \frac{1}{N} \sum_{t=1}^{N} \Phi \left(\underbrace{Z_{0,t}^{x_0,\mathbf{Y}}}_{\sim X_t} \right) - \int_{\mathcal{X}} \Phi(z)\, \mu_*(\mathrm{d}z) \right\|_p \le \underbrace{\left\| \frac{1}{N} \sum_{t=0}^{N-1} \int_{\mathcal{X}} \Phi(z)\, \left[\mu_*(S^t \mathbf{Y}, \mathrm{d}z) - \mu_*(\mathrm{d}z) \right] \right\|_p}_{A_N}$$

$$+ \underbrace{\left\| \frac{1}{N} \sum_{t=1}^{N} \int_{\mathcal{X}} \Phi(z)\, (\mu_t - \mu_*)(S^{t-1}\mathbf{Y}, \mathrm{d}z) \right\|_p}_{B_N} + \underbrace{\left\| \frac{1}{N} \sum_{t=1}^{N} \left(\Phi\left(Z_{0,t}^{x_0,\mathbf{Y}} \right) - \int_{\mathcal{X}} \Phi(z)\, \mu_t(S^{t-1}\mathbf{Y}, \mathrm{d}z) \right) \right\|_p}_{C_N}$$

where

- by the Birkhoff ergodic theorem, $A_N \xrightarrow{N\to\infty} 0$, with no explicit rate,
- by the stability of $\mathrm{Law}(X_t)$, $B_N \xrightarrow{N\to\infty} 0$, $(B_N \ge \mathcal{O}(e^{-cn^{1/3}}))$,
- by the L-mixing lemma, $C_N \xrightarrow{N\to\infty} 0$, $(C_N \ge \mathcal{O}(N^{-1/6}))$.

$\square$

Linear systems in random environments

Fix integers $m, d \geq 1$ and let $A, B : \mathbb{R}^m \to \mathbb{R}^{d\times d}$ be measurable functions.

Operator norm of a matrix $M \in \mathbb{R}^{d\times d}$ will be denoted $|||M|||$.

Let $Y_t \in \mathcal{Y} := \mathbb{R}^m$, $t \in \mathbb{Z}$ be a stationary process.

We consider the process $X_t \in \mathbb{R}^d$ obeying the linear dynamics

$$\begin{aligned} X_{t+1} &= A(Y_t)X_t + B(Y_t)\varepsilon_{t+1}, \ t \in \mathbb{N}, \\ X_0 &= x_0 \in \mathbb{R}^d, \end{aligned}$$

where $\varepsilon_t \in \mathbb{R}^d$, $t \geq 1$ is an i.i.d. sequence, independent of $(Y_t)_{t\in\mathbb{Z}}$.

Theorem: Assume that...

- A, B are bounded; $B(y)$, $A(y)$ are invertible for all $y \in \mathbb{R}^m$ such that $\sup_{y \in \mathbb{R}^m} \left(|||A(y)^{-1}||| + |||B(y)^{-1}||| \right) < \infty$;
- ε_0 has a density $f : \mathbb{R}^d \to \mathbb{R}_+$ bounded away from 0 on compact sets; $E|\varepsilon_0| < \infty$.
- Exists $p \geq 1$ such that $E\left[\ln |||A(Y_p)A(Y_{p-1}) \ldots A(Y_1)|||\right] < 0$.
- There is $\eta > 0$ such that

$$\Gamma(\alpha) = \lim_{n \to \infty} \mathbb{E}^{\frac{1}{n}} \prod_{i=1}^{n} |||A(Y_{ip})A(Y_{ip-1}) \ldots A(Y_{(i-1)p+1})|||^{\alpha}$$

  is differentiable on $\alpha \in (-\eta, \eta)$.

Then Law(X_t) converges to a limiting law, independent of x_0, at speed $e^{-ct^{1/3}}$. Furthermore, if $(Y_t)_{t \in \mathbb{Z}}$ is ergodic, then the law of large numbers holds as stated above.

Stochastic gradient Langevin algorithm

Let $U : \mathbb{R}^d \to \mathbb{R}_+$ be differentiable with $\nabla U = E[H(\theta, Y_0)]$, where $H : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$ a measurable function. The data stream $(Y_n)_{n\in\mathbb{Z}}$ is a $\mathbb{R}^m$-valued strict sense stationary process We consider, for some fixed $\lambda > 0$,

$$
\begin{aligned}
\theta_{n+1} &= \theta_n - \lambda H(\theta_n, Y_n) + \sqrt{\lambda}\xi_{n+1}, \\
\theta_0 &\in \mathbb{R}^d
\end{aligned}
$$

where ξ_n, $n \geq 1$ is an i.i.d sequence of standard d-dimensional Gaussian random variables independent of $(Y_n)_{n\in\mathbb{Z}}$.

An application: sampling from high-dimensional probability distributions

For λ small and n large, $\mathrm{Law}(\theta_n)$ is expected to be close to the probability defined by

$$
\pi(A) = \frac{\int_A e^{-U(\theta)} d\theta}{\int_{\mathbb{R}^d} e^{-U(\theta')} d\theta'}, \; A \in \mathcal{B}(\mathbb{R}^d).
$$

We assume that...

- **A**
  1. There is a bounded and measurable $\Delta : \mathbb{R}^m \to \mathbb{R}_+$, and $b \geq 0$ such that, $\langle H(\theta, y), \theta \rangle \geq \Delta(y)|\theta|^2 - b$ for all $\theta \in \mathbb{R}^d$ and $y \in \mathbb{R}^m$.
  2. There exist K_1, K_2, K_3 such that $|H(\theta, y)| \leq K_1|\theta| + K_2|y| + K_3$.
- **B** Y_0 is bounded, say, $|Y_0| \leq M$ a.s., $E[\Delta(Y_0)] > 0$. Furthermore, there is $\eta > 0$ such that the limit

$$\Gamma(\alpha) := \lim_{n \to \infty} \frac{1}{n} \ln \mathbb{E} e^{\alpha(\Delta(Y_1) + \ldots + \Delta(Y_n))}$$

exists for all $\alpha \in (-\eta, \eta)$ and Γ is continuously differentiable on $(-\eta, \eta)$. (Gärtner–Ellis type condition.)

Theorem

Let $\lambda > 0$ be small enough. Under these Assumptions, there exists a probability law $\mu(\lambda)$, independent of the initial value, such that, for all $0 < \varrho < 1/3$,

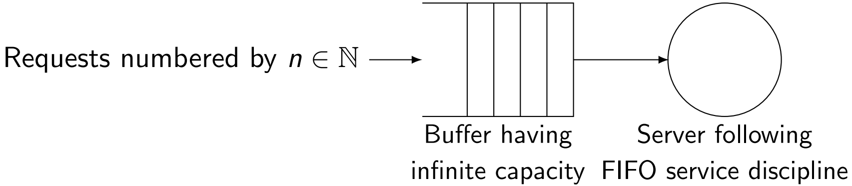
$$d_{TV}(\mathrm{Law}(\theta_n), \mu(\lambda)) \leq c_1(\varrho) \sum_{j=n}^{\infty} e^{-c_2(\varrho) j^{1/3-\varrho}},$$

for some $c_1(\varrho), c_2(\varrho) > 0$ (which also depend on λ). Moreover, if the data stream is ergodic, then for arbitrary bounded measurable $\Phi : \mathbb{R}^d \to \mathbb{R}$,

$$\frac{\Phi(\theta_0) + \ldots + \Phi(\theta_{n-1})}{n} \to \int_{\mathbb{R}^d} \Phi(z)\, \mu(\lambda)(dz),$$

as $n \to \infty$ in L^p, for all $p \geq 1$.

We analyze single-server queuing systems with...



Inter-arrival and service times:

- $Z_n \in \mathbb{R}_+$ is the time between the arrival of n-th and $(n+1)$-th requests,
- $S_n \in \mathbb{R}_+$ denotes the service time of the n-th request.

Starting with an empty queue, the sequence of waiting times $(W_n)_{n\in\mathbb{N}}$ satisfies the Lindley recursion

$$
\begin{aligned}
W_n &= (W_{n-1} + S_{n-1} - Z_n)_+, \ \ n > 0 \\
W_0 &= 0.
\end{aligned}
$$

If $(S_n)_{n\in\mathbb{Z}}$, $(Z_n)_{n\in\mathbb{Z}}$ are i.i.d. sequences, independent of each other, and $\mathrm{E}(S_0) < \mathrm{E}(Z_0)$, then $(W_n)_{n\in\mathbb{N}}$ is a Markov chain satisfying the Doeblin condition.

$\Rightarrow$ Geometric rate of convergence can be established.

However, if $(S_n, Z_n)_{n\in\mathbb{Z}}$ is merely stationary and ergodic, then the only bound which we have found is

$$\mathrm{d}_{TV}(\mathcal{L}(W_n), \mu_*) \leq \mathbb{P}\left(\min_{0<k<n} \sum_{j=1}^{k} \xi_j > \max\left(W_1, \xi_0 + \sup_{k\in\mathbb{N}} \sum_{j=1}^{k} \xi_{-j}\right)\right),$$

where $\xi_n = S_n - Z_{n+1}$, $n \in \mathbb{Z}$.

1. This bound is not a concrete and explicit rate estimate.
2. It is inappropriate for statistical analysis.

If $(S_n)_{n\in\mathbb{Z}}$, $(Z_n)_{n\in\mathbb{Z}}$ are i.i.d. sequences, independent of each other, and $\mathrm{E}(S_0) < \mathrm{E}(Z_0)$, then $(W_n)_{n\in\mathbb{N}}$ is a Markov chain satisfying the Doeblin condition.

$\Rightarrow$ Geometric rate of convergence can be established.

However, if $(S_n, Z_n)_{n\in\mathbb{Z}}$ is merely stationary and ergodic, then the only bound which we have found is

$$\mathrm{d}_{TV}(\mathcal{L}(W_n), \mu_*) \leq \mathbb{P}\left(\min_{0<k<n} \sum_{j=1}^{k} \xi_j > \max\left(W_1, \xi_0 + \sup_{k\in\mathbb{N}} \sum_{j=1}^{k} \xi_{-j}\right)\right),$$

where $\xi_n = S_n - Z_{n+1}$, $n \in \mathbb{Z}$.

1. This bound is not a concrete and explicit rate estimate.
2. It is inappropriate for statistical analysis.

If $(S_n)_{n\in\mathbb{Z}}$, $(Z_n)_{n\in\mathbb{Z}}$ are i.i.d. sequences, independent of each other, and $\mathrm{E}(S_0) < \mathrm{E}(Z_0)$, then $(W_n)_{n\in\mathbb{N}}$ is a Markov chain satisfying the Doeblin condition.

$\Rightarrow$ Geometric rate of convergence can be established.

However, if $(S_n, Z_n)_{n\in\mathbb{Z}}$ is merely stationary and ergodic, then the only bound which we have found is

$$\mathrm{d}_{TV}(\mathcal{L}(W_n), \mu_*) \leq \mathbb{P}\left(\min_{0<k<n} \sum_{j=1}^{k} \xi_j > \max\left(W_1, \xi_0 + \sup_{k\in\mathbb{N}} \sum_{j=1}^{k} \xi_{-j}\right)\right),$$

where $\xi_n = S_n - Z_{n+1}$, $n \in \mathbb{Z}$.

1. This bound is not a concrete and explicit rate estimate.
2. It is inappropriate for statistical analysis.

The non-i.i.d. setting is interesting because...

1. dependence arises often in queuing networks where arrival processes are departure processes of other queues,
2. processing operations like batching or the presence of multiple, different classes of requests can also cause dependence.

What makes it challenging is that $(W_n)_{n\in\mathbb{N}}$ **fails to be Markovian**.

We consider $G_I/G/1/\infty$ queuing:

- $(Z_n)_{n\in\mathbb{Z}}$ is strongly stationary,
- $(S_n)_{n\in\mathbb{Z}}$ is i.i.d and independent of $(Z_n)_{n\in\mathbb{Z}}$.

This case is amenable to analysis because the sequence of waiting times is a MCRE with driving noise $(S_n)_{n\in\mathbb{Z}}$ and random environment $(Z_n)_{n\in\mathbb{Z}}$. The parametric kernel is

$$Q(z,w,A) := \mathbb{P}\left[(w+S_0-z)_+ \in A\right],\ z,w\in\mathbb{R}_+,\ A\in\mathcal{B}(\mathbb{R}_+).$$

Theorem

1. *Assume that the sequence of inter-arrival times* $(Z_n)_{n\in\mathbb{Z}}$ *is*
   - A. *bounded, and strongly stationary,*
   - B. *satisfies the Gärtner–Ellis type condition: there is* $\eta > 0$ *such that the limit*
     $$\Gamma(\alpha) := \lim_{n\to\infty} \frac{1}{n} \ln E\left[e^{\alpha(Z_1+\ldots+Z_n)}\right]$$
     exists for all $\alpha \in (-\eta, \eta)$ *and* Γ *is differentiable on* $(-\eta,\eta)$.
2. *The sequence of service times* $(S_n)_{n\in\mathbb{Z}}$ *is*
   - A. *i.i.d. and independent of* $(Z_n)_{n\in\mathbb{Z}}$.
   - B. $\exists\beta_0 > 0\ E[e^{\beta_0 S_0}] < \infty$,
   - C. $\mathcal{L}(S_0)$ *has a density* $s \mapsto f(s)$ *(w.r.t. the Lebesgue measure) which is bounded away from* 0 *on compact subsets of* $\mathbb{R}_+$.

Then there exists a probability μ_* *on* $\mathcal{B}(\mathbb{R}_+)$ *such that*

$$\mathrm{d}_{TV}(\mathcal{L}(W_n), \mu_*) \le c_1 \exp\left(-c_2 n^{1/3}\right),$$

for some $c_1, c_2 > 0$.

Some remarks on the proof:

- A good candidate for the Lyapunov function is $V_\beta(x) = e^{\beta x} - 1$, $\beta > 0$.
- With this V_β, we have $\gamma_\beta(z) = e^{-\beta z}\mathbb{E}\left(e^{\beta S_0}\right)$, and $K_\beta(z) := \gamma_\beta(z) + 1$.
- The Gärtner–Ellis type condition ensures that exists $\bar{\beta} > 0$ such that

$$\limsup_{n\to\infty} \mathbb{E}^{1/n}\left(K_{\bar{\beta}}(Z_0)\prod_{k=1}^{n}\gamma_{\bar{\beta}}(Z_k)\right) < 1.$$

Thank you for your attention!

References

- A. A. Borovkov, *Asymptotic Methods in Queuing Theory*, Nauka, Moscow, 1980.
- L. Gerencsér. On a class of mixing processes. *Stochastics and Stochastic Reports*, 26(3):165–191, 1989.
- A. Lovas and M. Rásonyi, Markov chains in random environment with applications in queuing theory and machine learning, *Stochastic Processes and their Applications* **137** (2021) 294–326.
- A. Lovas and M. Rásonyi, Ergodic theorems for queuing systems with dependent inter-arrival times, *arXiv preprint:2004.01475* (2021).
- M. Welling and Y. W. Teh. Bayesian learning via stochastic gradient langevin dynamics. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 681–688, 2011.