Markov Chains in Random Environments

Theory and Applications

Attila Lovas

BME Mathematical Modeling Seminar
Budapest University of Technology and Economics
30 September, 2025

HUN-REN Alfréd Rényi Institute of Mathematics
Budapest University of Technology and Economics

Expected outcome, my goals, why I am here...

Motivation

Many real-world stochastic models can be understood as **Markov chains in random environments (MCREs)**. Recently, this theory has made remarkable progress. My goal today is to **share these new insights** and highlight their value for practitioners.

Expected outcome, my goals, why I am here...

Motivation

Many real-world stochastic models can be understood as **Markov chains in random environments (MCREs)**. Recently, this theory has made remarkable progress. My goal today is to **share these new insights** and highlight their value for practitioners.

Opportunities and challenges

- Still many open problems remain: both in theory and in applications (machine learning, stochastic volatility, growth models in economics, etc.).

Expected outcome, my goals, why I am here...

Motivation

Many real-world stochastic models can be understood as **Markov chains in random environments (MCREs)**. Recently, this theory has made remarkable progress. My goal today is to **share these new insights** and highlight their value for practitioners.

Opportunities and challenges

- Still many open problems remain: both in theory and in applications (machine learning, stochastic volatility, growth models in economics, etc.).
- These challenges range from approachable to deep mathematical problems.

Expected outcome, my goals, why I am here...

Motivation

Many real-world stochastic models can be understood as **Markov chains in random environments (MCREs)**. Recently, this theory has made remarkable progress. My goal today is to **share these new insights** and highlight their value for practitioners.

Opportunities and challenges

- Still many open problems remain: both in theory and in applications (machine learning, stochastic volatility, growth models in economics, etc.).
- These challenges range from approachable to deep mathematical problems.
- We welcome motivated students to join research at various levels: **BSc, MSc, or PhD theses**.

Overview

Introduction

Starting point: classical discrete state Markov chains

Markov property:

The future state depends only on the present state, not on the past history.

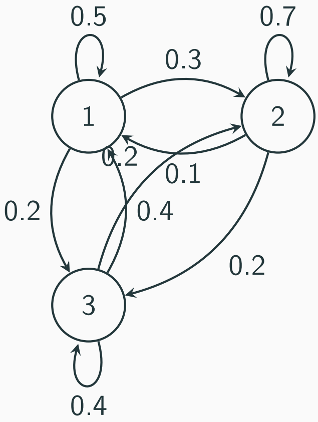
$$P(X_{n+1} = j \mid X_n = i, (X_k)_{k<n}) = P_{ij}$$

$$P_{ij} = P(X_{n+1} = j \mid X_n = i), \; i, j \in \mathcal{X},$$

where $|\mathcal{X}| \leq \aleph_0$.

Example: $\mathcal{X} = \{1, 2, 3\}$:

$$P = \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.1 & 0.7 & 0.2 \\ 0.4 & 0.2 & 0.4 \end{pmatrix}$$



General state spaces Markov chains

State space: complete, separable metric space $(\mathcal{X}, d_{\mathcal{X}\times\mathcal{X}})$, $\mathcal{B}(\mathcal{X})$ refers to its Borel σ-algebra.

Markov property:

$$\mathbb{P}(X_{n+1} \in A \mid X_n = x, (X_k)_{k<n}) = Q(x, A),\ x \in \mathcal{X},\ A \in \mathcal{B}(\mathcal{X}),$$

where the transition kernel

- $x \mapsto Q(x, A)$ is $\mathcal{B}(\mathcal{X})$-measurable, for every $A \in \mathcal{B}(\mathcal{X})$,
- $A \mapsto Q(x, A)$ is a probability on $\mathcal{B}(\mathcal{X})$, for every $x \in \mathcal{X}$.

This framework is general enough to cover:

- Discretizations of homogeneous Itô-diffusions,
- Time series models with dependence,
- and much more.

Markov chains – a different view

Random dynamical system representation

- Any general-state Markov process can be represented as

$$X_{n+1} = f(X_n, \xi_{n+1}),$$

- where $(\xi_n)_{n \geq 1}$ is an i.i.d. noise sequence.
- The representation is not unique: different maps f may lead to the same transition kernel.

Example:

Let $\{(A_n, B_n)\}$ be an i.i.d. sequence with

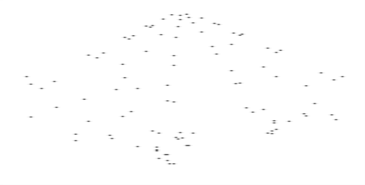
$$\mathbb{P}\big((A_0, B_0) = (a_i, b_i)\big) = w_i, \quad i = 1, 2,$$

where $w_1 = 0.2993$, $w_2 = 1 - w_1$.

$$a_1 = \begin{bmatrix} 0.4 & -0.3733 \\ 0.06 & 0.6 \end{bmatrix}, \quad b_1 = \begin{bmatrix} 0.3533 \\ 0 \end{bmatrix}$$

$$a_2 = \begin{bmatrix} -0.8 & -0.1867 \\ 0.1371 & 0.8 \end{bmatrix}, \quad b_2 = \begin{bmatrix} 1.1 \\ 0.1 \end{bmatrix}$$

$$X_{n+1} = A_{n+1} X_n + B_{n+1}$$



Markov chains in random environments – Definition

Let $\mathcal{Y}$ be a complete separable metric space and $(Y_n)_{n\in I}$, $I = \mathbb{N}, \mathbb{Z}$, a $\mathcal{Y}$-valued stochastic process. We say that the $\mathcal{X}$-valued process $(X_n)_{n\in I}$ is a **Markov chain in the random environment** $(Y_n)_{n\in I}$ if

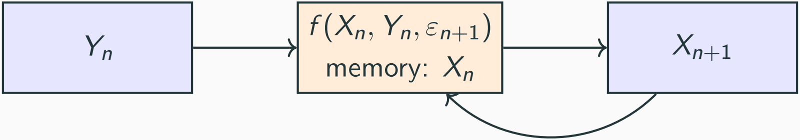
$$\mathbb{P}\big(X_{n+1} \in A \,\big|\, (X_k)_{k\leq n}, (Y_k)_{k\in I}\big) = Q(Y_n, X_n, A), \quad A \in \mathcal{B}(\mathcal{X}).$$

The parametric kernel $Q : \mathcal{Y} \times \mathcal{X} \times \mathcal{B}(\mathcal{X}) \to [0,1]$ satisfies:

- For each fixed $A \in \mathcal{B}(\mathcal{X})$, $(y,x) \mapsto Q(y,x,A)$ is $\mathcal{B}(\mathcal{Y}) \otimes \mathcal{B}(\mathcal{X})$-measurable.
- For each $(y,x) \in \mathcal{Y} \times \mathcal{X}$, $A \mapsto Q(y,x,A)$ is a probability measure on $\mathcal{B}(\mathcal{X})$.

Intuitive meaning: If we freeze one trajectory of the environment (Y_n), then (X_n) behaves like a **time-inhomogeneous Markov chain**.

MCREs as random iterative models



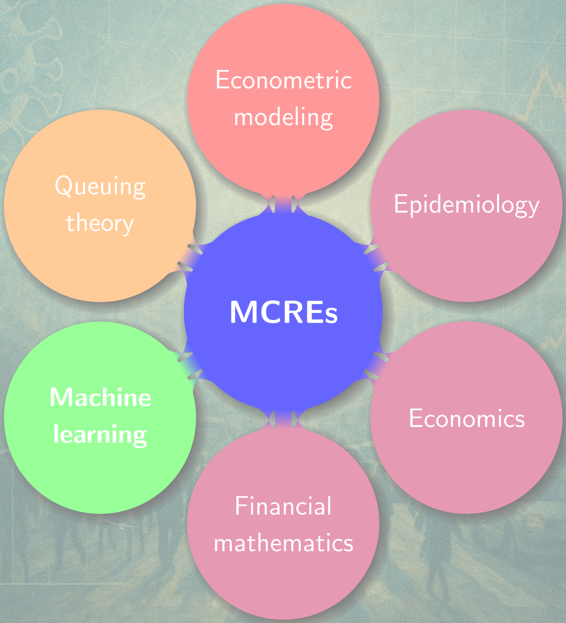
Strict exogeneity condition: $(\varepsilon_n)_{n\geq 1}$ i.i.d., independent of the environment. No feedback exists between $(\varepsilon_n)_{n\geq 1}$ and $(Y_n)_{n\in\mathbb{N}}$.

- Mathematical model of online data processing.
- Environment $(Y_n)_{n\in I}$ can be:
  - Exogenous regressor
  - Data stream

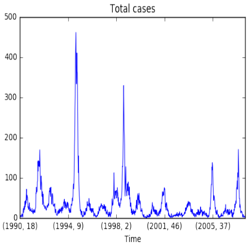
Sequential exogeneity / predetermination: ε_n independent of past $(Y_{n-j}, \varepsilon_{n-j})_{j\geq 1}$, may influence Y_n.

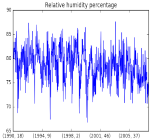
- Useful for more realistic econometric models.
- Not covered in this talk.

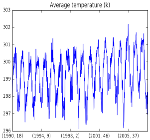
Applications of MCREs across fields

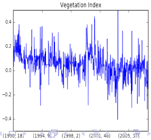


Epidemiology: Weekly dengue fever cases in San Juan





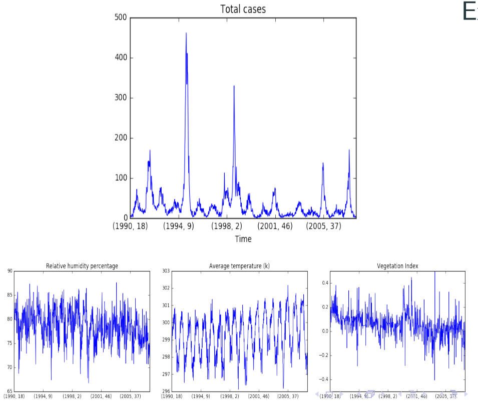




Exogenous regressors:

- **Relative humidity:** Moisture conditions affecting mosquitoes.
- **Average temperature:** Influences breeding and virus spread.
- **Vegetation index:** Habitat quality; dense vegetation provides breeding grounds.

Epidemiology: Weekly dengue fever cases in San Juan



Exogenous regressors:

- **Relative humidity:** Moisture conditions affecting mosquitoes.
- **Average temperature:** Influences breeding and virus spread.
- **Vegetation index:** Habitat quality; dense vegetation provides breeding grounds.

Current questions:

- Can we estimate the parametric kernel Q directly from data?
- Is there a theoretical guarantee that this estimation procedure is correct?

Source: L. Truquet, "Time Series, Exogeneity and Random Environments."

Economics: Theory of optimal economic growth

Hypothetical economy: Produces a single good that can be *consumed* or *invested*.

- Capital k_t and consumption c_t
- Output: $y_t = f(k_{t-1}, r_t)$, with r_t as a productivity shock.

Objective:

$$\max_{\{c_t\}} \mathbb{E}\Big[\sum_{t=0}^{\infty} \delta^t u(c_t)\Big] \quad \text{s.t. } y_t = f(k_{t-1}, r_t),\ c_t + k_t \leq y_t$$

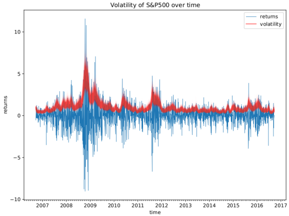
- $u(c)$: instantaneous utility
- δ: discount factor

Intriguing questions

- Does an optimal consumption strategy exist?
- Does the optimal strategy ensure **sustainable growth**, i.e., does the economy converge to a steady state with positive output and consumption in the long run?

Financial mathematics: Stochastic volatility

In financial time series the volatility is time-varying (e.g. volatility clustering):



Stochastic volatility models treat volatility as a random process.

Discrete time model (Gerencsér–Rásonyi, 2021):

$$S_{n+1} = \gamma S_n + \rho e^{Z_n} \eta_{n+1} + \sqrt{1-\rho^2}\, e^{Z_n} \varepsilon_{n+1}, \quad Z_n = \sum_{k=0}^{\infty} a_k \eta_{n-k},$$

where $(\eta_n)_{n\in\mathbb{Z}}$, $(\varepsilon_n)_{n\in\mathbb{Z}}$ are i.i.d. standard Gaussian, independent of each other, $\sum a_k^2 < \infty$.

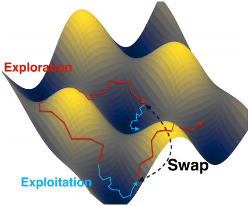
- The price $(S_n)_{n\in\mathbb{Z}}$ is conditionally Markov given the log-volatility process $(Z_n)_{n\in\mathbb{Z}}$.

Machine learning: the SGLD algorithm, (Welling & Teh, 2011)

- Given a dataset $\mathcal{D}$ and parameter θ, we aim to sample from

$$p(\theta \mid \mathcal{D}) \propto p(\mathcal{D} \mid \theta)\, p(\theta).$$

- Combine stochastic gradient descent (SGD) with Langevin dynamics to approximate samples from a posterior distribution.



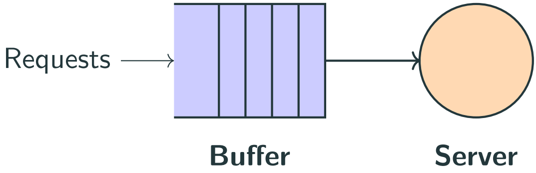
- Update rule (with step size η_t):

$$\theta_{n+1} = \theta_n - \lambda_n H(\theta_n, X_n) + \sqrt{2\lambda_n}\,\xi_{n+1},$$

where the gradient is replaced by a **stochastic estimate**: $\nabla U(\theta) = \mathbb{E}[H(\theta, X_0)]$, $U(\theta) = -\log p(\theta \mid \mathcal{D})$, $\xi_n \sim \mathcal{N}(0, I)$, i.i.d.

Queuing theory

Single-server queuing system with an infinite buffer and a first-in, first-out (FIFO) service discipline:



The waiting times satisfy the *Lindley recursion*:

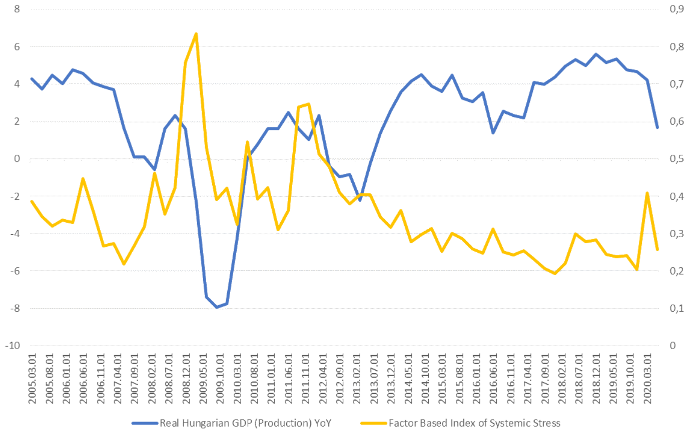
$$W_{n+1} = (W_n + S_n - Z_{n+1})_+, \quad n \in \mathbb{N},$$

with $W_0 := 0$, assuming an initially empty queue, where

- Z_n denotes the time between the arrivals of the n-th and $(n+1)$-th customers,
- S_n is the service time of the n-th customer.

Econometric modeling: time series with exogenous regressors

Hungarian GDP and factor-based index of systematic stress (FISS):



Source: Katalin Varga – Central Bank of Hungary

Questions relevant to practitioners

1. **Stability:** Does $\mathrm{Law}(X_n)$, $n \in \mathbb{N}$, converge in some metric to some limiting distribution?
2. **Ergodicity:** Under what conditions, and in what sense (strong, weak, L^p), does the law of large numbers hold?
3. **Fluctuations:** Does a (functional) central limit theorem apply to the sequence of iterates?

- These questions highlight gaps in the statistical toolkit for nonlinear autoregressive models, where few established tools exist, particularly when the exogeneous covariate sequence $(Y_n)_{n\in\mathbb{N}}$ *is not stationary*.
- Our objective is to contribute to this area by developing theoretical tools to address these questions.

Main theoretical results

Main milestones of MCRE research

Countable spaces
(Cogburn, Orey)

General state spaces,
Doeblin-type conditions
(Kifer, Seppalainen)

Non-compact spaces
strictly contracting dynamics
(Stenflo)

Harris-type conditions
contractive environment
(Gerencsér–Rásonyi, 2018)

Contractivity in average
(Lovas–Rásonyi, 2019)

Stationary solution
existence and unicity
(Truquet, 2021)

Non-stationary environment
transition of mixing
(Lovas–Truquet, 2025)

Standard conditions and their role

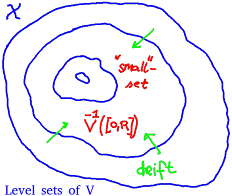
Drift/Lyapunov condition: there exists a function $V : \mathcal{X} \to [0, \infty)$ and functions $\gamma, K : \mathcal{Y} \to (0, \infty)$ such that

$$\int_{\mathcal{X}} V(z)\, Q(y, x, \mathrm{d}z) \leq \gamma(y) V(x) + K(y), \quad (y, x) \in \mathcal{Y} \times \mathcal{X}.$$

- If $\gamma(y) < 1$, the dynamics are *contractive*.
- If $\gamma(Y_n) < 1$ occurs sufficiently often*, then the process X_n visits appropriate level sets of V frequently enough.

Minorization/small set condition: on $\overset{-1}{V}([0, R])$

$$Q(y, x, A) \geq (1 - \beta(R, y))\kappa_R(y, A), \; y \in \mathcal{Y}, \, A \in \mathcal{B}(\mathcal{X}).$$



Level sets of V

- On level sets of V, the process has a positive probability of *"forgetting" its past* at each step.
- The larger $1 - \beta(R, y)$ is, the stronger this effect**.

+ additional assumptions to ensure * and **.

Markov Chains in stationary random environment

Theorem (Lovas-Rásonyi, 2019) Under the *long-term contractivity condition*

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

along with additional technical conditions, we have:

- Convergence in total variation:

  $$\mathrm{Law}(X_n) \xrightarrow{d_{\mathrm{TV}}} \mu_* \quad \text{as} \quad n\to\infty,$$

  with an explicit rate of order $O(c_1 e^{-c_2 n^{1/3}})$.

- If $(Y_n)_{n\in\mathbb{Z}}$ is also ergodic, then for any <u>bounded</u> measurable function $\Phi:\mathcal{X}\to\mathbb{R}$ and $1\le p<\infty$,

  $$\frac{1}{n}\sum_{k=1}^{n}\Phi(X_k) \xrightarrow{L^p} \int_{\mathcal{X}}\Phi(z)\,\mu_*(\mathrm{d}z).$$

The proof relies on an L-mixing result by L. Gerencsér.

Existence and uniqueness of a stationary solution

Theorem (L. Truquet, 2021): Given a stationary process $(Y_n)_{n\in\mathbb{Z}}$ and the conditions

$$\mathbb{E}\left[\log(\gamma(Y_0))_+\right] + \mathbb{E}\left[\log(K(Y_0))_+\right] < \infty,\ \mathbb{E}\left[\log(\gamma(Y_0))\right] < 0,$$

there exists a Markov chain in random environment $(X_n^*)_{n\in\mathbb{Z}}$ such that:

- The process $((Y_t, X_t^*))_{t\in\mathbb{Z}}$ is stationary with a unique invariant distribution.
- If $(Y_t)_{t\in\mathbb{Z}}$ is ergodic, then $((Y_t, X_t^*))_{t\in\mathbb{Z}}$ is also ergodic.

Truquet also showed that

$$\mathrm{Law}(X_n) \xrightarrow{d_{\mathrm{TV}}} \mathrm{Law}(X_0^*) \quad \text{as} \quad n \to \infty$$

when $X_0 = x_0$ is deterministic, though without providing an explicit rate of convergence.

α-mixing processes (Rosenblatt, 1956)

For any sequence of random variables $(W_t)_{t\in\mathbb{Z}}$, we define:

- $\mathcal{F}^W_{t,s} := \sigma(W_k,\ t \leq k \leq s)$ for $-\infty \leq t \leq s \leq \infty$.
- The mixing coefficient $\alpha^W_j(n)$ is defined as:

$$\alpha^W_j(n) = \sup\left\{ |\mathbb{P}(G \cap H) - \mathbb{P}(G)\mathbb{P}(H)| \;\middle|\; F \in \mathcal{F}^W_{-\infty,j},\ G \in \mathcal{F}^W_{j+n,\infty} \right\}.$$

- The overall mixing coefficient of the process W is given by $\alpha^W(n) = \sup_{j\in\mathbb{Z}} \alpha^W_j(n)$.
- The sequence $(\alpha^W(n))_{n\in\mathbb{N}}$ is non-increasing and measures the independence of events occurring at distant times.

Remark: Several mixing concepts exist, such as ϕ, ψ, and ρ. Among these the classical mixing concepts, α-mixing is the weakest.

α-mixing (continued)

The decay rate of $\alpha^W(n)$ is quantified by the summability criterion:

$$\sum_{n=1}^{\infty} \left[\alpha^W(n)\right]^{1/\kappa} < \infty \quad \text{for some } \kappa > 0.$$

Size: $(W_n)_{n\in\mathbb{N}}$ is said to be α-mixing of size $-\kappa_0$ if

$$\alpha^W(n) = O(n^{-\kappa}) \quad \text{for some } \kappa > \kappa_0.$$

A process W is **strongly mixing** if

$$\lim_{n\to\infty} \alpha^W(n) = 0.$$

Folklore: Stationarity + strong mixing $\Rightarrow$ ergodicity $\Rightarrow$ strong law of large numbers.

Many powerful theoretical results are available for α-mixing sequences.

Law of large numbers for α-mixing sequences

Theorem (McLeish, 1975): Let $(W_n)_{n\in\mathbb{N}}$ be a process with finite means $\mu_n = \mathbb{E}[W_n]$ and with $(\alpha(n))_{n\in\mathbb{N}}$ of size $-r/(r-2)$ for $r > 2$.

Furthermore, assume that

$$\sum_{n=1}^{\infty} \left(\frac{\mathbb{E}\,|W_n - \mu_n|^p}{n^p} \right)^{2/r} < \infty \tag{1}$$

for some p such that $r/2 < p \le r < \infty$. Under these conditions,

$$\frac{1}{n}\sum_{k=1}^{n}(W_k - \mu_k) \to 0 \quad \text{a.s. as } n \to \infty.$$

Remark: In the models we study, it holds that

$$\sup_{n\in\mathbb{N}} \mathbb{E}\,|W_n - \mu_n|^p < \infty,$$

so condition (1) is evidently satisfied.

Law of large numbers for α-mixing sequences (continued)

Theorem (Hansen, 2019): Consider a strongly mixing $\mathbb{R}$-valued process $(W_n)_{n\in\mathbb{N}}$, and define the sequence of partial sums $S_n := \sum_{k=1}^n W_k$ for $n \geq 1$. Additionally, suppose that the following uniform integrability condition holds:

$$\lim_{B\to\infty} \sup_{n\geq 1} \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}\left(|W_n| \mathbf{1}(|W_n| \geq B)\right) = 0. \tag{2}$$

Then, we have

$$\frac{S_n}{n} - \frac{\mathbb{E}(S_n)}{n} \xrightarrow{L^1} 0, \quad n \to \infty.$$

Remark: The average uniform integrability condition (2) is satisfied if the process $(W_n)_{n\in\mathbb{N}}$ has a uniformly bounded moment, i.e., $\sup_{n\in\mathbb{N}} \mathbb{E}(|W_n|^r) < \infty$ for some $r > 1$.

A recent CLT-like result for α-mixing variables

Theorem (Ekström, 2014): Let $\{\xi_{n,i} \mid 1 \leq i \leq d_n, n \in \mathbb{N}\}$ be an array of $\mathbb{R}$-valued random variables with $\mathbb{E}\xi_{n,i} = 0$.

Define $S_n = \sum_{k=1}^{d_n} \xi_{n,k}$ for $n \in \mathbb{N}$, and assume:

i. $\sup_{n\in\mathbb{N}} \max_{1\leq i\leq d_n} \mathbb{E}|\xi_{n,i}|^{2+r} < \infty$ for some $r > 0$,
ii. $\sup_{n\in\mathbb{N}} \sum_{k=0}^{\infty}(k+1)^2 \left(\alpha^{\xi_{n,\cdot}}(k)\right)^{\frac{r}{4+r}} < \infty$,

where $\alpha^{\xi_{n,\cdot}}(\cdot)$ is the strong mixing coefficient for the nth row.

Then, the distributions $\mathrm{Law}(d_n^{-1/2} S_n)$ and $\mathcal{N}(0, \mathrm{Var}(d_n^{-1/2} S_n))$ are *weakly approaching* i.e.

$$\mathbb{E}\left[g\left(d_n^{-1/2} S_n\right)\right] - \int_{\mathbb{R}} g\left(\mathrm{Var}(d_n^{-1/2} S_n)^{1/2} t\right) \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt \to 0 \quad \text{as } n \to \infty$$

for any bounded continuous function $g : \mathbb{R} \to \mathbb{R}$.

A CLT for α-mixing sequences

Theorem (White, 2001): Let $\{X_{tn}\}$ be a triangular array of scalars with means $\mu_{tn} = \mathbb{E}[X_{tn}]$ and variances $\sigma_{tn}^2 = \text{Var}(X_{tn})$ such that

$$\sup_{t,n} \mathbb{E}|X_{tn}|^r < \infty$$

for some $r \geq 2$. Suppose that $\{X_{tn}\}$ has α-mixing coefficient of size $-r/(r-2)$ for $r > 2$. Furthermore, assume that

$$\overline{\sigma}_n^2 = \mathrm{Var}\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_{tn}\right) \geq \delta > 0$$

for all sufficiently large n. Under these conditions, as $n \to \infty$,

$$\frac{\sqrt{n}}{\overline{\sigma}_n}\left(\overline{X}_n - \overline{\mu}_n\right) \xrightarrow{d} N(0, 1),$$

where

$$\overline{X}_n = \frac{1}{n} \sum_{t=1}^{n} X_{tn} \quad \text{and} \quad \overline{\mu}_n = \frac{1}{n} \sum_{t=1}^{n} \mu_{tn}.$$

Functional CLT for α-mixing sequences

In non-stationary cases, a major challenge is that $\text{Var}(S_n)$ can diverge at an arbitrary nonlinear rate, complicating the analysis.

Herrndorf (1984): If $\mathbb{E}(W_n) = 0$, $\sup_{n\in\mathbb{N}} \mathbb{E}(W_n^{2+\epsilon}) < \infty$, $\mathbb{E}(S_n^2)/n \to \sigma \geq 0$ as $n \to \infty$, and $\alpha^W(n) \to 0$ "sufficiently fast," then a functional CLT holds.

Note: By assuming $\mathbb{E}(S_n^2)/n \to \sigma \geq 0$ as $n \to \infty$, the variance issue is circumvented. However, this result has been valuable in later research.

Functional CLT for α-mixing sequences

In non-stationary cases, a major challenge is that $\mathrm{Var}(S_n)$ can diverge at an arbitrary nonlinear rate, complicating the analysis.

Herrndorf (1984): If $\mathbb{E}(W_n) = 0$, $\sup_{n\in\mathbb{N}} \mathbb{E}(W_n^{2+\epsilon}) < \infty$, $\mathbb{E}(S_n^2)/n \to \sigma \geq 0$ as $n \to \infty$, and $\alpha^W(n) \to 0$ "sufficiently fast," then a functional CLT holds.

Note: By assuming $\mathbb{E}(S_n^2)/n \to \sigma \geq 0$ as $n \to \infty$, the variance issue is circumvented. However, this result has been valuable in later research.

Merlevède, Peligrad, and Utev (2019): Functional CLT for random variables $\{\xi_{k,n} \mid k \leq n\}$ satisfying the Lindeberg condition. In addition to $\sup_m \alpha^{\xi_{\cdot,m}}(n) \to 0$ sufficiently fast, they required:

$$\sum_{j=1}^{n} \mathrm{Var}(\xi_{j,n}) = O\left(\mathrm{Var}(S_n)\right).$$

Functional CLT for α-mixing sequences

In non-stationary cases, a major challenge is that $\mathrm{Var}(S_n)$ can diverge at an arbitrary nonlinear rate, complicating the analysis.

Herrndorf (1984): If $\mathbb{E}(W_n) = 0$, $\sup_{n\in\mathbb{N}} \mathbb{E}(W_n^{2+\epsilon}) < \infty$, $\mathbb{E}(S_n^2)/n \to \sigma \geq 0$ as $n \to \infty$, and $\alpha^W(n) \to 0$ "sufficiently fast," then a functional CLT holds.

Note: By assuming $\mathbb{E}(S_n^2)/n \to \sigma \geq 0$ as $n \to \infty$, the variance issue is circumvented. However, this result has been valuable in later research.

Merlevède, Peligrad, and Utev (2019): Functional CLT for random variables $\{\xi_{k,n} \mid k \leq n\}$ satisfying the Lindeberg condition. In addition to $\sup_m \alpha^{\xi_{\cdot,m}}(n) \to 0$ sufficiently fast, they required:

$$\sum_{j=1}^{n} \mathrm{Var}(\xi_{j,n}) = O\left(\mathrm{Var}(S_n)\right).$$

For suitable functionals of MCREs, proving the $\geq$ direction is straightforward, while proving $\leq$ required novel techniques.

Main assumptions

We assume that the parametric kernel Q satisfies the **drift** and **minorization** conditions, such that the following hold:

A1

$$r_0 := \sum_{\ell \geq 0} d_\ell < \infty, \qquad d_0 = \sup_{t \geq 0} \mathbb{E}\big[K(Y_t) + \gamma(Y_t)\big] < \infty,$$

$$d_\ell := \sup_{t \geq -1} \mathbb{E}\left[K(Y_t) \prod_{i=1}^{\ell} \gamma(Y_{t+i})\right], \quad \ell \geq 1.$$

A2 For any $R > 0$,

$$\lim_{\overline{\beta} \uparrow 1} \sup_{t \in \mathbb{N}} \mathbb{P}\big(\beta(R, Y_t) > \overline{\beta}\big) = 0.$$

Non-stationaty environments (Lovas–Truquet, 2025)

Suppose that either X_0 is independent from $(Y_t)_{t\in\mathbb{Z}}$ with $\mathbb{E}\left[V(X_0)\right] < \infty$ or $((X_t, Y_t))_{t\in\mathbb{Z}}$ is a stationary process.

1. **Uniformly bounded moments:** Suppose that for some constant $C > 0$ and exponent $p > 1$, we have

$$|\Phi(x)|^p \leq C(1 + V(x)).$$

   Then,

$$\sup_{j\geq 0} \|\Phi(X_j)\|_p < \infty.$$

2. **Transition of mixing:** Set $r_i = \sum_{\ell\geq i} d_\ell$ for some positive integer i. Then there exist $\kappa \in (0, 1)$ and a positive constant c such that

$$\alpha^{X,Y}(n) \leq c \inf_{1\leq i\leq q\leq n} \left\{ r_i + \kappa^{n/q} + \alpha^Y(q + 1 - i) \right\}.$$

Stationary environment – stability, (Lovas–Truquet, 2025)

The processes $(W_n)_{n\in\mathbb{N}}$ and $(W_n')_{n\in\mathbb{N}}$ are said to be **forward coupled**, if there exists an almost surely finite random time τ such that

$$W_n = W_n',\ \ n \geq \tau.$$

Assumptions:

- Suppose that $(Y_n)_{n\in\mathbb{Z}}$ is strongly stationary and α-mixing.
- Let $(X_n^*, Y_n)_{n\in\mathbb{Z}}$ denote the associated stationary process, and let X_0 be a random initial state independent of $(Y_n)_{n\in\mathbb{Z}}$.

Result: There exist versions of $(X_n)_{n\in\mathbb{N}}$ and $(X_n^*)_{n\in\mathbb{N}}$ that are *forward coupled*. Furthermore, the following rate estimate holds:

$$\|\mathrm{Law}(X_n) - \mathrm{Law}(X_n^*)\|_{TV} \leq 2c \inf_{1\leq i\leq q\leq n} \left\{ r_i + \kappa^{n/q} + \alpha^Y(q-i) \right\}$$

Selected applications

Machine learning

Importance of optimization in learning

- Optimization is the standard tool for training machine learning models.
- A *loss function* $U : \mathbb{R}^d \to \mathbb{R}$ measures the discrepancy between predictions and observed data.

Training as an optimization problem

$$\theta^* = \arg\min_{\theta \in \mathbb{R}^d} U(\theta).$$

Challenges

- d is typically very large and U is often non-convex.
- Training deep neural networks involves optimizing trillions of parameters.
- U may not be explicitly available — only ∇U can be estimated from data.

Pitfalls of gradient descent

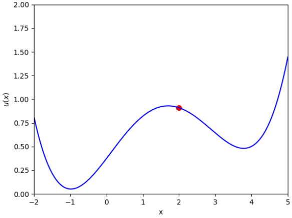
Gradient descent (GD):

$$\theta_{n+1} = \theta_n - \lambda \nabla U(\theta_n)$$

is the earliest and most common optimization method.

- $\lambda > 0$ is called step-size or learning rate depending on the context.

Limitations:



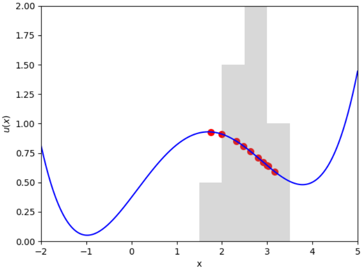
- Computing gradients over the full dataset can be expensive (minibatch SGD mitigates this to some extent).
- GD can get stuck in local minima or saddle points, especially in non-convex, multimodal landscapes:

Would adding some artificial noise at each step help?

The Langevin iteration:

$$\theta_{n+1} = \theta_n - \lambda \nabla U(\theta_n) + \sqrt{\frac{2\lambda}{\beta}}\, \xi_{n+1},$$

- $(\xi_n)_{n\geq 1}$ are i.i.d. standard Gaussian variables,
- $\beta > 0$ is the inverse temperature parameter.

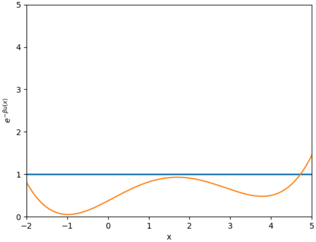


How Langevin iteration actually works?

The Langevin SDE:

$$\mathrm{d}\theta_t = -\nabla U(\theta_t)\mathrm{d}t + \beta^{-1/2}\mathrm{d}B_t,$$

where $(B_t)_{t\geq 0}$ is a standard Brownian motion.



Convergence to unique invariant distribution:

$\mathrm{Law}(\theta_t) \to \pi_\beta,\ \pi_\beta(dx) \propto e^{-\beta U(x)}dx.$

For large β, π_β concentrates to the global minimum x^*.

Error analysis:

- $\mathrm{Law}(\theta_n) \to \pi_{\beta,\lambda}$ exponentially fast;
- $\pi_{\beta,\lambda}$ differs from π_β: $d_{W^1}(\pi_\beta, \pi_{\beta,\lambda}) = O(\sqrt{\lambda})$.

When only an unbiased estimate of ∇U is available

Stochastic Gradient Langevin Dynamics (SGLD):

$$\theta_{n+1} = \theta_n - \lambda H(\theta_n, X_n) + \sqrt{\frac{2\lambda}{\beta}}\,\xi_{n+1},$$

where

- $(X_n)_{n\in\mathbb{N}}$ is an $\mathbb{R}^m$-valued stationary process, interpreted as data stream;
- $H : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$ measurable function, the stochastic gradient of U i.e. $\nabla U(\theta) = \mathbb{E}\left[H(\theta, X_0)\right],\ \theta \in \mathbb{R}^d$.

What we already know?

- The case when $(X_n)_{n\in\mathbb{N}}$ is i.i.d. has been thoroughly studied. In this setting, $(\theta_n)_{n\in\mathbb{N}}$ forms a Markov chain.
- Our focus is on the more challenging situation where $(X_n)_{n\in\mathbb{N}}$ may even be **non-stationary**.

Assumptions on the update function and the data stream

Assumptions on H:

- Dissipativity: there are measurable $\Delta, b : \mathbb{R}^m \to \mathbb{R}$ and $v : \mathbb{R}^m \to [0, \infty)$ such that for all $\theta \in \mathbb{R}^d$ and $x \in \mathbb{R}^m$,

$$\langle \theta, H(\theta, x) \rangle \geq \Delta(x) \|\theta\|^2 - b(x).$$

- At most linear growth: for some $L > 0$

$$\|H(\theta, x)\| \leq L(\|\theta\| + v(x) + 1), \quad \theta \in \mathbb{R}^d, x \in \mathbb{R}^m.$$

Assumptions on $(X_n)_{n\in\mathbb{N}}$:

- There exists $\delta \in (0, 1)$ such that

$$\sup_{n\in\mathbb{N}} \mathbb{E}\left[v(X_n)^{2\delta} + |b(X_n)|^{\delta} \right] < \infty.$$

- $\tilde{\Delta} := \inf_{n\in\mathbb{N}} \mathbb{E}[\Delta(X_n)] > 0.$

The main theorem on SGLD (Lovas–Truquet, 2025)

Assume that there exists $M > 0$ such that for all $t > 0$ and $n \geq 1$,

$$\sup_{j\in\mathbb{N}} \mathbb{P}\left(\sum_{k=1}^{n} \log\gamma(X_{k+j}) > t + \sum_{k=1}^{n} \mathbb{E}[\log\gamma(X_{k+j})]\right) \leq \exp\left(-\frac{t^2}{Mn}\right),$$

where

$$\gamma(x) = 3L^2\lambda^2 - 2\lambda\Delta(x) + 1.$$

Then for step size $0 < \lambda < \frac{2\tilde{\Delta}}{3L^2}$, the following holds:

- There exists $s \in (0,1)$ such that for any $\Phi : \mathbb{R}^d \to \mathbb{R}$ satisfying $|\Phi(\theta)|^p \leq C(1 + \|\theta\|^{2s})$ for some $C > 0$ and $p \geq 1$, we have
$$\sup_{n\in\mathbb{N}} \mathbb{E}[|\Phi(\theta_n)|^p] < \infty.$$
- If $\alpha^X(n) = O(n^{-a})$ for some $a > 1$, then $\alpha^\theta(n) = O\left(\frac{\log^a(n)}{n^a}\right)$. If $\alpha^X(n) = O(\kappa^n)$ for some $\kappa \in (0,1)$, then there exists $\overline{\kappa} \in (0,1)$ such that $\alpha^\theta(n) = O(\overline{\kappa}^{\sqrt{n}})$.

Queuing systems

The queuing model

Recall the Lindley's recursion:

$$W_{n+1} = (W_n + S_n - Z_{n+1})_+, \qquad n \in \mathbb{N},$$

with $W_0 = 0$ (initially empty queue), where

- Z_n = inter-arrival time between the n-th and $(n+1)$-th customers,
- S_n = service time of the n-th customer.

Loynes' classical results (1962)

Assuming the sequence $\{(S_n, Z_{n+1})\}_{n \in \mathbb{Z}}$ is stationary and ergodic:

- **Subcritical case:** $\mathbb{E}[S_0] < \mathbb{E}[Z_1] \Rightarrow$ the queue is **stable**.
- **Supercritical case:** $\mathbb{E}[S_0] > \mathbb{E}[Z_1] \Rightarrow$ the queue is **unstable**.
- **Critical case:** $\mathbb{E}[S_0] = \mathbb{E}[Z_1] \Rightarrow$ the queue may be **stable**, **substable**, or **unstable**.

Theorem (Györfi–Morvai, 1999)

Assume $\{(S_n, Z_{n+1})\}_{n\in\mathbb{Z}}$ is stationary and ergodic, and the system is **subcritical**:

$$\mathbb{E}[S_0] < \mathbb{E}[Z_1].$$

Then $(W_n)_{n\in\mathbb{N}}$ is **forward coupled with** a stationary and ergodic $(W'_n)_{n\in\mathbb{Z}}$, where

$$W'_0 = \sup_{n\in\mathbb{N}} Y_n, \qquad Y_n = \sum_{k=1}^{n}(S_{-k} - Z_{-k+1}), \quad Y_0 = 0.$$

Rate estimate:

$$\left\|\mathrm{Law}(W_n) - \mathrm{Law}(W'_0)\right\|_{TV} \;\le\; \mathbb{P}\left(\min_{0<k<n} \sum_{j=1}^{k}(S_j - Z_{j+1}) \;>\; \max\bigl(W_1,\; W'_0 + S_0 - Z_1\bigr)\right).$$

This bound is **not informative** in practice and of little use for applications.

Stationary service times, stability (Lovas, 2024)

Assumptions: Let $(S_n)_{n\in\mathbb{Z}}$ be stationary, bounded, and independent of i.i.d. (Z_n).

- The system is subcritical $\mathbb{E}[S_0] < \mathbb{E}[Z_1]$.
- Log-moment generating function

  $$\Gamma(t) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E} e^{t(S_1+\cdots+S_n)}$$

  exists and is differentiable near 0.
- $\mathbb{P}(Z_0 \geq \tau) > 0$ for a suitable threshold τ.

Result: There exists a stationary $(S_n, W_n^*)_{n\in\mathbb{Z}}$ satisfying the Lindley recursion such that

- Processes (W_n) and (W_n^*) are forward coupled,
- for some $\kappa \in (0,1)$,

  $$\|\mathrm{Law}(W_n) - \mathrm{Law}(W_n^*)\|_{TV} = O(\kappa^{1/2}).$$

Non-stationary service times (Lovas, 2024)

Under some additional technical assumptions:

A) If (S_n) is **strongly mixing**, then

$$\frac{1}{n}\sum_{k=1}^{n}\left(W_k - \mathbb{E}[W_k]\right) \xrightarrow{L^1} 0.$$

B) If $\alpha^S(n) \leq cn^{-\kappa}$ with $\kappa > 1$, then

$$\frac{1}{n}\sum_{k=1}^{n}\left(W_k - \mathbb{E}[W_k]\right) \xrightarrow{\text{a.s.}} 0.$$

C) If $\sum_n (n+1)^2 \alpha^S(n)^\delta < \infty$ for some $\delta \in (0,1)$, then a **CLT-type bound** holds:

$$\limsup_{n\to\infty} \mathbb{P}\left(\frac{1}{\sqrt{n}}\Big|\sum_{k=1}^{n}(W_k - \mathbb{E}W_k)\Big| \geq a\right) \leq \mathbb{P}(\sigma\xi \geq a),$$

for some $\sigma > 0$ with $\xi \sim \mathcal{N}(0,1)$.

Invariance principle for queues (Lovas, 2024)

Assumptions:

- Arrival density f_{Z_1} satisfying $\int_0^\infty \frac{f'_{Z_1}(z)^2}{f_{Z_1}(z)}\,dz < \infty$.
- Queue stability condition: $\liminf_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{P}(S_k > Z_1)^2 > 0$.
- Mixing: for some $\delta > 0$, $\sum_{n\geq 1} n^\delta \alpha^S(n) < \infty$.

Result: Define $\Sigma_n = \sum_{k=1}^n W_k$, $\xi_{k,n} = \frac{W_k - \mathbb{E}[W_k]}{\sqrt{\mathrm{Var}(\Sigma_n)}}$, and

$$v_n(t) = \min\{1 \leq k \leq n \mid \mathrm{Var}(\Sigma_k) \geq t\mathrm{Var}(\Sigma_n)\}.$$

Then the partial-sum process

$$B_n(t) = \sum_{k=1}^{v_n(t)} \xi_{k,n}, \quad t \in (0,1],$$

converges in distribution (in $D([0,1])$ with uniform topology) to a **standard Brownian motion**.

Econometrics

A simple VAR-X model

Let $Y_n \in \mathbb{R}^d$, $n \in \mathbb{N}$ be an exogenous process. We consider the process $X_n \in \mathbb{R}^d$ obeying the linear dynamics

$$X_{n+1} = AX_n + BY_n + \varepsilon_{n+1}, \tag{3}$$

where

- A and B are fixed $d \times d$ matrices such that the spectral radius of A satisfies $r(A) < 1$;
- $(\varepsilon_n)_{n\geq 1}$ is i.i.d. independent of $(Y_n)_{n\in\mathbb{N}}$;
- $\mathbb{E}[|\varepsilon_0|] < \infty$, Law$(\varepsilon_0)$ is absolutely continuous with respect to the Lebesgue measure, with density bounded away from zero on compact sets;
- X_0 and $(\varepsilon_n)_{n\geq 1}$ are conditionally independent given Y.

Theorem (Lovas–Truquet, 2025)

Under our standing assumptions, if

$$\sup_{n\in\mathbb{N}} \mathbb{E}[\|Y_n\|] < \infty,$$

then there exist constants $c'' > 0$ and $r \in (0, 1)$ such that

$$\alpha^{X,Y}(n) \ \leq \ c''\Big[\, r^{\,n^{1/2}} + \alpha^{Y}\Big(\lfloor n^{1/2}/2 \rfloor\Big)\Big].$$

Moreover, for any function $\Phi : \mathbb{R}^d \to \mathbb{R}$ satisfying

$$|\Phi(x)| \ \leq \ c'(1 + \|x\|),$$

for some $c' > 0$, the moments are uniformly bounded:

$$\sup_{n\in\mathbb{N}} \mathbb{E}[|\Phi(X_n)|] < \infty.$$

Corollary: All classical results available for mixing processes also apply here:

- Law of Large Numbers (LLN), construction of confidence intervals, concentration inequalities, and beyond.

Future research and challenges

Future research and challenges

Theory:

- Refine existing results on the *transition of α-mixing*.
- Establish transfer of mixing properties under **monotonicity conditions** on the parametric kernel $\rightarrow$ essential for applications in *stochastic optimal economic growth*.
- Extend the theory to **sequentially exogenous** random iterative models, motivated by econometrics.

Applications:

- Advanced queuing models (multi-server systems, finite buffer capacity, diverse service disciplines).
- Convergence analysis of Monte Carlo algorithms on data streams.
- Structural design of polymers.

Thank you for your attention!

Questions?

Bibliography

References

B. GERENCSÉR AND M. RÁSONYI, *On the ergodicity of certain Markov chains in random environments*, Journal of Theoretical Probability, 36 (2023), pp. 2093–2125.

A. LOVAS AND L. TRUQUET, *Mixing properties of some Markov chains models in random environments*, arXiv preprint arXiv:2509.16118, (2025).

A. LOVAS, *Transition of α-mixing in Random Iterations with Applications in Queuing Theory*, arXiv preprint arXiv:2410.05056, (2024).

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), pp. 294–326.