Empirical concentration for time series

Pierre Alquier



ESSEC
BUSINESS SCHOOL

Workshop: Optimisation and Uncertainty in Foundation Models, NTU, Singapore
Nov. 23, 2026

1 Introduction

2 Markov chains

3 Chains with infinite memory

1 Introduction

2 Markov chains

3 Chains with infinite memory

Theorem – McDiarmid's inequality

Assume $X_1, \ldots, X_n$ are independent random variables.

Let $f : \mathcal{X}^n \to \mathbb{R}$ satisfy the bounded difference property :

$$\sup_{x_1,\ldots,x_n,x'} |f(x_1, \ldots, x_i, \ldots, x_n) - f(x_1, \ldots, x', \ldots, x_n)| \leq c_i.$$

Then, for any $\varepsilon \in (0, 1)$,

$$\mathbb{P}\left(\left| f(X_1, \ldots, X_n) - \mathbb{E}[f(X_1, \ldots, X_n)] \right| \geq \varepsilon \right) \leq \exp\left(-\frac{2\varepsilon^2}{\sum_{i=1}^n c_i^2} \right).$$



C. McDiarmid (1989). On the method of bounded differences. In *Surveys in Combinatorics, 1989 : Invited Papers at the Twelfth British Combinatorial Conference*, pp. 148–188.

Theorem – McDiarmid's inequality

For independent random variables + bounded difference property,

$$\mathbb{P}\left(\left|f(X_1,\ldots,X_n)-\mathbb{E}[f(X_1,\ldots,X_n)]\right| \geq \varepsilon\right) \leq 2\exp\left(-\frac{2\varepsilon^2}{\sum_{i=1}^n c_i^2}\right).$$

Typical application : $f(X_1,\ldots,X_n) = \frac{1}{n}\sum_{i=1}^n g(X_i)$, $0 \leq g(x) \leq 1$.

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n g(X_i) - \mathbb{E}[g(X_1)]\right| \geq \varepsilon\right) \leq 2\exp\left(-2n\varepsilon^2\right) =: \delta.$$

Solving in δ :

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n g(X_i) - \mathbb{E}[g(X_1)]\right| \geq \sqrt{\frac{\log\frac{2}{\delta}}{2n}}\right) \leq \delta.$$

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} g(X_i) - \mathbb{E}[g(X_1)]\right| \geq \sqrt{\frac{\log\frac{2}{\delta}}{2n}}\right) \leq \delta.$$

Application 1 : generalization bounds for classification

- $X_i = (Z_i, Y_i)$, $Z_i \in \mathcal{Z}$ object, $Y_i \in \mathcal{Y}$ label ;
- classifiers $h_\theta : \mathcal{Z} \to \mathcal{Y}$ indexed by $\theta \in \Theta$;
- $g(X_i) := 1_{\{Y_i \neq f_\theta(Z_i)\}}$, $R_n(\theta) := \frac{1}{n}\sum_{i=1}^{n} g(X_i)$ and $R(\theta) := \mathbb{P}(Y_i \neq f_\theta(Z_i))$.

$$\mathbb{P}\left(|R_n(\theta) - R(\theta)| \geq \sqrt{\frac{\log\frac{2}{\delta}}{2n}}\right) \leq \delta.$$

$$\mathbb{P}\left(|R_n(\theta) - R(\theta)| \geq \sqrt{\frac{\log \frac{2}{\delta}}{2n}}\right) \leq \delta.$$

Can be used as a starting point for MDL, PAC-Bayes, etc. For example, when Θ is countable and π a probability on Θ,

$$\mathbb{P}\left(\forall \theta \in \Theta, R(\theta) \leq R_n(\theta) + \sqrt{\frac{\log \frac{1}{\pi(\theta)} + \log \frac{2}{\delta}}{2n}}\right) \geq 1 - \delta.$$

This motivates :

$$\hat{\theta} := \arg\min_{\theta \in \Theta} \left\{ R_n(\theta) + \sqrt{\frac{\log \frac{1}{\pi(\theta)} + \log \frac{2}{\delta}}{2n}} \right\}.$$

Application 2 : multi-armed bandits



At each round $t \in \mathbb{N}$,

- choose an arm $k_t \in \{1, \dots, K\}$ to sample from ;
- receive a reward $R_t \sim \mu_{k_t}$; the μ_k's are distributions on $[0, 1]$.

At time t, let $N_k(t) \leq t$ denote the number of times we pulled arm k. We have an i.i.d. sample $X_1^{(k)}, \ldots, X_{N_k(t)}^k$. We can estimate its average reward m_k by :

$$\hat{m}_{k,t} := \frac{1}{N_k(t)} \sum_{i=1}^{N_k(t)} X_i^{(k)}.$$



McDiarmid :

$$\mathbb{P}\left(|m_k - \hat{m}_{k,t}| \leq \sqrt{\frac{\log\frac{2}{\delta}}{2N_k(t)}}\right) \geq 1-\delta.$$

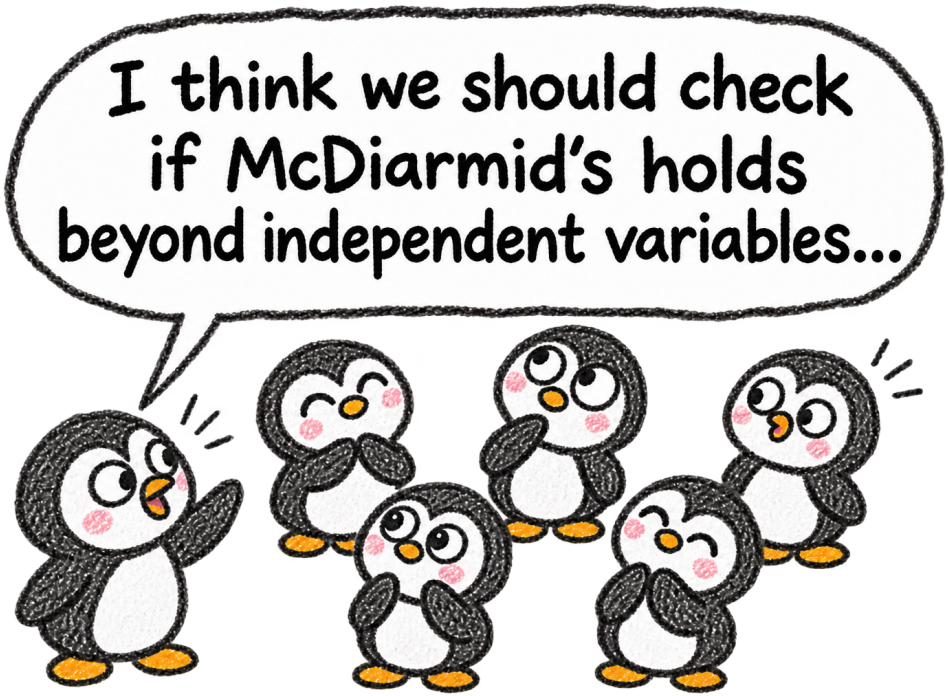
The UCB algorithm uses this approach with a suitable $\delta = \delta(t)$ ensuring an anytime-valid Upper Confidence Bound :

$$k_t := \underset{1\leq k\leq K}{\arg\max}\left\{\hat{m}_{k,t} + \sqrt{\frac{\log\delta(t)}{2N_t(k)}}\right\}.$$



T. Lattimore and C. Szepesvári (2020). *Bandit Algorithms*. Cambridge University Press.


These guys are naive enough to believe the rewards are independent...
μ_1
μ_2
μ_3
μ_4
μ_5
μ_6


I think we should check if McDiarmid's holds beyond independent variables...

Let $\mathcal{A}$ and $\mathcal{B}$ be σ-fields. If $\mathcal{A}$ and $\mathcal{B}$ are independent, then for $A \in \mathcal{A}$ and $B \in \mathcal{B}$, we have

$$\mathbb{P}(B|A) = \mathbb{P}(B).$$

Definition – φ-coefficient

$$\varphi(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}} \Big| \mathbb{P}(B|A) - \mathbb{P}(B) \Big|.$$

φ-mixing coefficients

Let $X = (X_t)_{t \in \mathbb{N}}$ be a sequence of random variables. We put, for any $\ell \in \mathbb{N}$,

$$\varphi_X(\ell) = \sup_{t \in \mathbb{N}} \varphi\Big(\sigma(X_1, \ldots, X_t), \sigma(X_{t+\ell}, X_{t+\ell+1}, \ldots) \Big).$$

Theorem – Rio's version of McDiarmid's inequality

Assume

$$\Sigma := \sum_{\ell=1}^{\infty} \varphi_X(\ell) < +\infty.$$

Let $f : \mathcal{X}^n \to \mathbb{R}$ satisfy the bounded difference property with coefs. $(c_1, \ldots, c_n)$. Put $F := f(X_1, \ldots, X_n)$ for short.

For any $\varepsilon \in (0, 1)$,

$$\mathbb{P}\left(\left| F - \mathbb{E}[F] \right| \geq \varepsilon \right) \leq \exp\left(- \frac{2\varepsilon^2}{(1 + 2\Sigma) \sum_{i=1}^n c_i^2} \right).$$



E. Rio (2000). Inégalités de Hoeffding pour les fonctions lipschitziennes de suites dépendantes. *Compte Rendus de L'Académie des Sciences* **330**(10) :905–908.

Theorem – Rio's version of McDiarmid's inequality

$$\mathbb{P}\left(\left|F-\mathbb{E}[F]\right|\geq\varepsilon\right)\leq\exp\left(-\frac{2\varepsilon^2}{(1+2\Sigma)\sum_{i=1}^n c_i^2}\right).$$

ML application : $F=\frac{1}{n}\sum_{i=1}^n g(X_i)$, $0\leq g(x)\leq 1$.

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n g(X_i)-\mathbb{E}[g(X_1)]\right|\geq\sqrt{\frac{(1+2\Sigma)\log\frac{2}{\delta}}{2n}}\right)\leq\delta.$$

Application 1 : model selection in classification

$$\hat{\theta}:=\arg\min_{\theta\in\Theta}\left\{R_n(\theta)+\sqrt{(1+2\Sigma)\frac{\log\frac{1}{\pi(\theta)}+\log\frac{2}{\delta}}{2n}}\right\}.$$

Unless an oracle provides $(1+2\Sigma)$, this is not feasible!

Application 2 : bandits

$$k_t := \underset{1 \leq k \leq K}{\arg\max} \left\{ \hat{m}_{k,t} + \sqrt{(1+2\Sigma)\frac{\log \delta(t)}{2N_k}} \right\}.$$



Many Hoeffding / Bernstein / McDiarmid's inequalities for dependent variables, under various assumptions. They all depend on a quantity $\mathcal{D}$ quantifying the dependence (like Σ) :

- φ-mixing (Rio's inequality),
- β-mixing,



B. Yu (1994). Rates of convergence for empirical processes of stationary β-mixing sequences. *The Annals of Statistics* **22**(1) :94–116.

- Markov chains (references coming later)...

Current applications in machine learning requires the prior knowledge of an upper bound on Σ. For example :



P. Alquier and O. Wintenberger (2012). Model selection for weakly dependent time series forecasting. *Bernoulli* **18**(3) :883-913.



C. Tekin and M. Liu (2012). Online learning of rested and restless bandits. *IEEE Transactions on Information Theory* **58**(8) :5588–5611.

Roadmap (heuristic)

1. Start from such an inequality :

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} g(X_i) - \mathbb{E}[g(X_1)]\right| \geq \sqrt{\mathcal{D}\frac{\log\frac{2}{\delta}}{2n}}\right) \leq \delta.$$

2. Estimate $\mathcal{D}$ on the same $(X_1, \ldots, X_n)$, with confidence :

$$\mathbb{P}\left(\left|\mathcal{D} - \hat{\mathcal{D}}\right| \geq u_n(\delta)\right) \leq \delta,\ u_n(\delta) \xrightarrow[n\to\infty]{} 0.$$

3. Apply a union bound :

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} g(X_i) - \mathbb{E}[g(X_1)]\right| \geq \sqrt{\left[\hat{\mathcal{D}} + u_n(\delta)\right]\frac{\log\frac{2}{\delta}}{2n}}\right) \leq 2\delta.$$

- Estimation of α and β-mixing coefficients :



A. Khaleghi and G. Lugosi (2023). Inferring the mixing properties of a stationary ergodic process from a single sample-path. *IEEE Transactions on Information Theory* **69**(6) :4014–4026.

- Estimation of φ-mixing coefficients :



S. Yao (2026). *Strongly consistent estimation of the extended ℓ_1-sum of ϕ-mixing coefficients from a single trajectory*. Preprint arXiv :2609.26232.

Beautiful, but no rates of convergence in both cases.
However, in the Markov setting, we have can derive rates of convergence for β-mixing :



S. Grünewälder and A. Khaleghi (2026). Estimating the mixing coefficients of geometrically ergodic Markov processes. *IEEE Transactions on Information Theory* **72**(2) :1305–1318.



G. Wolfer and P. Alquier (2027). Optimistic estimation of convergence in Markov chains with the average-mixing time. *Bernoulli*, to appear.

Context	Inequality	$\mathcal{D}$	Estimator	Paper
Markov chain, finite space	Paulin	pseudo-spectral gap	Wolfer and Kontorovich	Karagulyan and Alquier
Markov chain, compact space	Paulin	pseudo-spectral gap	new	Karagulyan, Kasprzak and Alquier
Chain with infinite memory	new	it's a mess	new	Alquier and Doukhan

D. Paulin (2015). Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electronic Journal of Probability* 20 :1–32.

G. Wolfer and A. Kontorovich (2024). Improved estimation of relaxation time in nonreversible Markov chains. *Annals of Applied Probability* 34(1A) :249–276.

V. Karagulyan and P. Alquier (2026). Empirical PAC-Bayes bounds for Markov chains. *Proceedings of the 29th International Conference on Artificial Intelligence and Statistics*, PMLR 300, pp. 361-369.

V. Karagulyan, M. Kasprzak and P. Alquier (2026). *Minimax estimation of the pseudo-spectral gap of continuous Markov chains, with applications to empirical concentration bounds.* Preprint arXiv :2610.20007.

P. Alquier and P. Doukhan (2026). *Empirical concentration for chains with infinite memory*. Preprint arXiv :2610.32265.

1 Introduction

2 Markov chains

3 Chains with infinite memory

In this project, the penguins were saved by





Mikołaj Kasprzak,
ESSEC Business School

Vahe Karagulyan,
ESSEC Business School







(A1) (X_t) is a stationary Markov chain with transition kernel P :

$$\mathbb{P}(X_{t+1} \in A | X_t = x) = P(x, A)$$

and stationary distribution μ : $\mu P = \mu$.

P defines an operator on functions in $\mathcal{L}^2(\mu)$:

$$Pf(y) = \int P(y, \mathrm{d}x) f(x).$$

Definition

The pseudo-spectral gap (or PSG) of P is given by :

$$\gamma_{ps} := \sup_{k \geq 1} \frac{1 - \|P^k\|^2_{\mathcal{H} \to \mathcal{H}}}{k}$$

where $\mathcal{H} = \{f \in \mathcal{L}^2(\mu) : \int f \mathrm{d}\mu = 0\}$.



(Rough application of) Paulin's version of Bernstein inequality

For $0 \leq g \leq 1$,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} g(X_i) - \mathbb{E}[g(X_1)]\right| \geq \Delta_n(\gamma_{ps}, \delta)\right) \leq \delta$$

where

$$\Delta_n(\gamma_{ps}, \delta) = \frac{40 \log \frac{2}{\delta}}{n\gamma_{ps}} + \sqrt{\frac{8\left(1 + \frac{1}{n\gamma_{ps}}\right)\log \frac{2}{\delta}}{n\gamma_{ps}}}.$$

Finite state space $|\mathcal{X}| = d$: estimator of Wolfer and Kontorovich.

$$N^k(i,j) = \sum_{\ell=1}^{\lfloor \frac{n-1}{k} \rfloor} 1_{X_{1+(\ell-1)k}=i, X_{1+\ell k}=j}, \quad N^k(i) = \sum_{\ell=1}^{\lfloor \frac{n-1}{k} \rfloor} 1_{X_{1+(\ell-1)k}=i},$$

$$\hat{L}^k(i,j) = \frac{N^k(i,j)}{\sqrt{N^k(i)N^k(j)}} \text{ estimates } L^k = \mathrm{diag}(\mu)^{\frac{1}{2}} P^k \mathrm{diag}(\mu)^{-\frac{1}{2}}.$$

$$\text{Put } K = \lceil 2/\epsilon \rceil \text{ and } \hat{\gamma}_{\mathrm{ps},K} = \max\left\{ \frac{1 - \|\hat{L}^k\|_{\mathrm{op}}^2}{k}, 1 \le k \le K \right\}.$$

Proposition (Wolfer and Kontorovich)

$$\mathbb{P}\left(\left| \frac{\hat{\gamma}_{\mathrm{ps},K}}{\gamma_{\mathrm{ps}}} - 1 \right| > \epsilon \right) \le \alpha(n, \epsilon, \gamma_{\mathrm{ps}}) = \alpha,$$

$$\alpha = \frac{dC \min(d, C)}{\epsilon \gamma_{\mathrm{ps}} c} \exp\left(-n\epsilon^2 \gamma_{\mathrm{ps}}^2 c \min\left(\gamma_{\mathrm{ps}}, \frac{1}{d \min(d, C)} \right) \right), c = \min_i \mu_i, C = \frac{\max_j \mu_j}{c}.$$

Case $\mathcal{X} = [0,1]^d$ (our paper).

Assumptions :

- (A2) : $P(x, \mathrm{d}y) = F(x,y)\mathrm{d}y$.
- (A3) : $|F(x,y) - F(x',y')| \leq L(\|x - x'\| + \|y - y'\|)$.
- (A4) : let $F_m(x,y)$ denote the density of $P^m(x, \mathrm{d}y)$; there is $m \geq 1$ and $\pi_* > 0$ such that $\min_{x,y} F_m(x,y) \geq \pi_*$.

Construction of the estimator :

- regular partition of $[0,1]^d$ into H^d small hypercubes $\{A_i\}$.
- $\hat{\alpha}_{i,j,n}^{(k)} = \frac{1}{n-k} \sum_{t=1}^{n-k} 1\{X_t \in A_i, X_{t+k} \in A_j\}$.
- $\hat{\alpha}_n^{(k)} = (\hat{\alpha}_{i,j,n}^{(k)})_{1 \leq i,j \leq K^d}$, $\hat{r}_n^{(k)} = \hat{\alpha}_n^{(k)} 1_n$ and $\hat{c}_n^{(k)} = [\hat{\alpha}_n^{(k)}]^T 1_n$.
- $\hat{B}_n^{(k)} := \mathrm{diag}(\hat{r}_n^{(k)})^{-1/2} \left[\hat{\alpha}_n^{(k)} - \hat{r}_n^{(k)} (\hat{c}_n^{(k)})^T \right] \mathrm{diag}(c_n^{(k)})^{-1/2}$.
- $\hat{\lambda}_n^{(k)} = \mathrm{clip}_{[0,1]} \| \hat{B}_n^{(k)} \|_{\mathrm{op}}^2$.
- $\hat{\gamma}_{ps,n} := \max_{1 \leq k \leq k_n} \frac{1 - \hat{\lambda}_n^{(k)}}{k}$.
- calibration : $H \sim \left(\frac{n}{\log(n)} \right)^{\frac{1}{d+2}}$, $k_n \sim \log(n)$.

Important : $\hat{\gamma}_{ps,n}$ does not depend on (m, π_*, L) from (A2)-(A4).

Theorem

There is a constant $C = C(d, m, \pi_*, L)$ such that, under (A1)-(A4), we have

$$\mathbb{P}\left[|\hat{\gamma}_{ps,n} - \gamma_{ps}| \leq C\left(\frac{\log\frac{64(n+1)}{\delta}}{n}\right)^{\min\left(\frac{1}{2},\frac{2}{d+2}\right)}\right] \geq 1-\delta.$$

Theorem

Let $\mathcal{Q}_d(m, \pi_*, L)$ denote the class of kernels P satisfying (A1)-(A4). There is a constant $c = c(d, m, \pi_*, L)$ such that

$$\inf_{\hat{\gamma}_n} \sup_{P\in\mathcal{Q}_d(m,\pi_*,L)} \mathbb{P}\left[|\hat{\gamma}_n - \gamma_{ps}(P)| > \frac{c}{n^{\min\left(\frac{1}{2},\frac{2}{d+2}\right)}}\right] \geq \frac{1}{4}.$$

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} g(X_i) - \mathbb{E}[g(X_1)]\right| \geq \Delta_n(\gamma_{ps}, \delta)\right) \leq \delta.$$

$$\Delta_n(\gamma_{ps}, \delta) \to \Delta_n\left(\hat{\gamma}_{ps,n} - Cu_n, \delta\right),\ u_n := \left(\frac{\log\frac{64(n+1)}{\delta}}{n}\right)^{\min\left(\frac{1}{2}, \frac{2}{d+2}\right)}$$

not feasible because $C = C(d, m, \pi_*, L)$. Instead, overshoot a little :

$$\widehat{\Delta}_n(\delta) := \Delta_n(\hat{\gamma}_{ps,n} - u_n \log(n), \delta) = \Delta_n(\gamma_{ps}, \delta)(1 + o(1)).$$

Empirical version of Paulin's inequality

For $n \geq \exp(C)$,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} g(X_i) - \mathbb{E}[g(X_1)]\right| \geq \widehat{\Delta}_n(\delta)\right) \leq 2\delta.$$

Application 2 : multi-armed bandits

Put $m^* = \max\{m_k, 1 \leq k \leq K\}$. In the i.i.d. case, there is a version of the UCB algorithm ensuring :

$$\forall T,\ Tm^* - \mathbb{E}\left[\sum_{t=1}^{T} R_t\right] \leq \sqrt{KT\log(T)} + 3K.$$

When each arm k is a Markov chain with respective kernels $P_1, \ldots, P_K$ and PSGs $\gamma_{ps,1}, \ldots, \gamma_{ps,K}$ (unknown), we propose a version of UCB ensuring :

$$\forall T,\ Tm^* - \mathbb{E}\left[\sum_{t=1}^{T} R_t\right] \leq 15\sqrt{\frac{KT\log(KT^3)}{\min_{1\leq k\leq K}\gamma_{ps,k}}} + C(P_1, \ldots, P_K).$$

1 Introduction

2 Markov chains

3 Chains with infinite memory

The penguins are grateful to

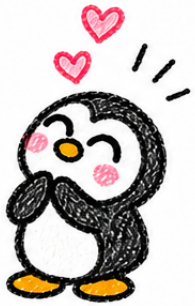


Paul Doukhan,

CY Cergy Paris University







- (A1) $(X_t)_{t\in\mathbb{Z}}$ a stationary process taking values in a metric space (E,ρ). We assume that

$$\mathcal{L}(X_t|X_{t-1}=u_{-1},X_{t-2}=u_{-2},\dots)=P_{(u_{-1},u_{-2},\dots)}=P_u.$$

- (A2) We have

$$W_{1,\rho}(P_u,P_v)\leq\sum_{j\geq 1}\delta_j(P)\rho(u_{-j},v_{-j})$$

and

$$\Delta(P):=1-\sum_{j\geq 1}\delta_j(P)>0.$$

- (A3) E is bounded : $\Delta_E:=\sup_{x,x'\in E}\rho(x,x')<\infty$.

Theorem

Let $f : \mathcal{X}^n \to \mathbb{R}$ satisfy :

$$|f(x_1, \ldots, x_i, \ldots, x_n) - f(x_1, \ldots, x', \ldots, x_n)| \leq c_i \rho(x_i, x').$$

Put $F := f(X_1, \ldots, X_n)$ for short. For any $\varepsilon \in (0, 1)$,

$$\mathbb{P}\left(\left|F - \mathbb{E}[F]\right| \geq \varepsilon\right) \leq \exp\left(-\frac{2\varepsilon^2 [\Delta(P)]^2}{\Delta_E^2 \sum_{i=1}^n c_i^2}\right).$$

For example, the Hamming metric $\rho(x, x') = 1\{x \neq x'\}$ recovers the bounded-difference condition and thus, for $0 \leq g \leq 1$,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n g(X_i) - \mathbb{E}[g(X_1)]\right| \geq \frac{1}{\Delta(P)}\sqrt{\frac{\log\frac{1}{\delta}}{2n}}\right) \leq \delta.$$

Case 1 : E finite, Hamming metric

In this case, we can take :

$$\delta_j(P) = \sup_{\substack{u,v \\ u_{-\ell}=v_{-\ell} \ (\ell \neq j)}} \frac{1}{2} \sum_{x \in E} |p(x \mid u) - p(x \mid v)|$$

and our inequality recovers exactly the one from Chazottes et al.



J.-R. Chazottes, S. Gallo, and D. Y. Takahashi (2023). Gaussian concentration bounds for stochastic chains of unbounded memory. *The Annals of Applied Probability* 33(5) :3321–3350.

Construction of the estimator :

- Fix a number K of lags.
- Put $U_i = (X_{i-1}, \ldots, X_{i-K})$, and

$$C_n(u) = \sum_{i=K+1}^{n} 1_{\{U_i = u\}} \text{ and } \widehat{p}_K(x \mid u) = \frac{\sum_{i=K+1}^{n} 1_{\{U_i = u, X_i = x\}}}{C_n(u)}.$$

- We can then estimate :

$$\widehat{\delta}_{j,K} = \max_{\substack{u,v \\ u_{-\ell} = v_{-\ell} \ (\ell \neq j)}} \frac{1}{2} \sum_{x \in E} |\widehat{p}_K(x \mid u) - \widehat{p}_K(x \mid v)|,$$

- Finally, put :

$$\widehat{\Delta}_{n,K} = \mathrm{clip}_{[0,1]} \left(1 - \sum_{j=1}^{K} \widehat{\delta}_{j,K} \right).$$

Calibration : $K = K_n \sim \sqrt{\log n}$.

Theorem

Under (A1)-(A3), assume in addition that

$$\inf_{x,u} p(x|u) > 0.$$

For any $q > 0$, there is a constant $C = C(|E|, \Delta(P), q)$ such that

$$\mathbb{P}\left[\left|\widehat{\Delta}_{n,K_n} - \Delta(P)\right| \leq C v_n\right] \geq 1 - \frac{1}{n^q}$$

where v_n is given by :

Memory assumption	Rate v_n
$\delta_j(P) \leq a e^{-bj^\gamma}$, $a, b > 0$, $\gamma > 2$	$n^{-1/2+\varepsilon}$
$\delta_j(P) \leq a e^{-bj}$, $a, b > 0$	$(\log n)^{1/2} e^{-b\sqrt{\log n}}$
$\delta_j(P) \leq a j^{-r}$, $a > 0$, $r > 2$	$(\log n)^{-(r-2)/2}$

Case 2 : $E = [0,1]^d$, Euclidean metric

How to satisfy (A2) : $W_{1,\rho}(P_u, P_v) \leq \sum_{j \geq 1} \delta_j(P) \|u_{-j} - v_{-j})\|$?

Assuming P_u has a density $p(\cdot|u)$, Poincaré inequality implies :

$$W_{1,\rho}(P_u, P_v) \leq \frac{1}{\pi} \|p(\cdot|u) - p(\cdot|v)\|_\infty.$$

We propose to assume the stronger :

- (A2') : $p(\cdot|u)$ is differentiable in each u_{-j} and

$$\sup_{x,u} \|\nabla_{u_{-j}} p(x|u)\| \leq L_j \left(= \frac{\delta_j}{\pi} \right)$$

and estimate the coefficients L_j.

Construction of the estimator :

- $\mathcal{K}$ kernel with support $[-1,1]$, $\mathcal{K}_d$ product kernel on $[-1;1]^d$,
- $\widehat{f}_{K,h}(x,u) = \frac{1}{(n-K)h^{d(K+1)}} \sum_{i=K+1}^{n} \mathcal{K}_d\left(\frac{X_i - x}{h}\right) \mathcal{K}_{dK}\left(\frac{U_i - u}{h}\right)$,
- $\widehat{g}_{K,h}(u) = \frac{1}{(n-K)h^{dK}} \sum_{i=K+1}^{n} \mathcal{K}_{dK}\left(\frac{U_i - u}{h}\right)$,
- $\widehat{g}^{\tau}_{K,h} = \widehat{g}_{K,h} \vee n^{-1}$,
- $\widehat{D}_{j,K,h}(x,u) = \frac{\nabla_{u_{-j}} \widehat{f}_{K,h}(x,u)}{\widehat{g}^{\tau}_{K,h}(u)} - \frac{\widehat{f}_{K,h}(x,u) \nabla_{u_{-j}} \widehat{g}_{K,h}(u)}{[\widehat{g}^{\tau}_{K,h}(u)]^2}$,
- $\widehat{L}_{j,K,h} = \sup_{(x,u)\in[h,1-h]^{d+1}} \|\widehat{D}_{j,K,h}(x,u)\|$,
- $\widehat{\Delta}_{K,h} = \mathrm{clip}_{[0,1]}\left(1 - \sum_{j=1}^{K} \widehat{L}_{j,K,h}\right)$.

Calibration : $K = K_n \sim (\log n)^{1/3}$, $h = h_n \sim n^{-\frac{1}{4d(K_n+1)+12}}$.

Theorem

Under (A1), (A2') and (A3), assume in addition that

$$\inf_{x,u} p(x|u) > 0.$$

For any $q > 0$, there is a constant $C = C(d, \Delta(P), q)$ such that

$$\mathbb{P}\left[\left|\widehat{\Delta}_{K_n,h_n} - \Delta(P)\right| \leq C v_n\right] \geq 1 - \frac{1}{n^q}$$

where v_n is given by :

Memory assumption	Rate v_n
$\delta_j(P) = 0$ for $j > K_0$	$e^{-c(d)(\log n)^{2/3}}$
$\delta_j(P) \leq Ce^{-bj}$, $b > 0$	$(\log n)^{1/3} e^{-b(\log n)^{1/3}}$
$\delta_j(P) \leq Cj^{-r}$, $r > 2$	$(\log n)^{-(r-2)/3}$

CY Initiatives Horizon project SafeTime : (187,500€) :
Safe Decisions via Adaptive Learning of Time Series



Initiative of excellence



- time series theory & stochastic processes :
  Paul Doukhan (AGM, CY), Mikołaj Kasprzak (ESSEC, Paris), Daniel Paulin (NTU, Singapore), Lionel Truquet (ENSAI), Geoffrey Wolfer (TUAT, Tokyo), Nicolas Marie (Nanterre), Xiequan Fan (Northeastern University, Qinhuangdao)
- time series & econometrics :
  Guillaume Chevillon (ESSEC, Paris), Jamus Lim (ESSEC, Singapore),
- machine learning theory & algorithms :
  Kamélia Daudel (ESSEC, Paris), Pierre Alquier (ESSEC, Singapore)

1 post-doc starting September 2027 at ESSEC (Paris campus, 2 years)