



Bootstrap and uniform bounds for Harris Markov chains

Gabriela Ciolek

► To cite this version:

Gabriela Ciolek. Bootstrap and uniform bounds for Harris Markov chains. Statistics [math.ST]. Université Paris Saclay (COMUE), 2018. English. NNT : 2018SACLT024 . tel-01988595

HAL Id: tel-01988595

<https://pastel.hal.science/tel-01988595>

Submitted on 21 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT

de

L'UNIVERSITÉ PARIS-SACLAY

École doctorale de mathématiques Hadamard (EDMH, ED 574)

Établissement d'inscription : Télécom ParisTech

Laboratoire d'accueil : Laboratoire traitement et communication de l'information

Spécialité de doctorat : Mathématiques appliquées

Gabriela CIOLEK

Bootstrap and uniform bounds for Harris Markov chains

Date de soutenance : 14 Décembre 2018

Après avis des rapporteurs : KENGO KATO (Cornell University)
OLIVIER WINTENBERGER (Sorbonne Université)

	PATRICE BERTAIL	(Université Paris Nanterre) Codirecteur de thèse
	STEPHAN CLÉMENÇON	(Télécom ParisTech) Codirecteur de thèse
	RANDAL DOUC	(Télécom SudParis) Examineur
<i>Jury de soutenance :</i>	CÉCILE DUROT	(Université Paris Nanterre) Examineur
	KENGO KATO	(Cornell University) Rapporteur
	SANA LOUHICHI	(Université de Grenoble) Président
	OLIVIER WINTENBERGER	(Sorbonne Université) Rapporteur

Titre : Bootstrap et bornes uniformes pour des chaînes de Markov Harris récurrentes

Mots Clefs : Chaînes de Markov, processus régénératifs, bootstrap, robustesse, inégalités de concentration, apprentissage statistique

Résumé : Cette thèse se concentre sur certaines extensions de la théorie des processus empiriques lorsque les données sont Markoviennes. Plus spécifiquement, nous nous concentrons sur plusieurs développements de la théorie du bootstrap, de la robustesse et de l'apprentissage statistique dans un cadre Markovien Harris récurrent positif. Notre approche repose sur la méthode de régénération qui s'appuie sur la décomposition d'une trajectoire de la chaîne de Markov atomique régénérative en blocs d'observations indépendantes et identiquement distribuées (i.i.d.). Les blocs de régénération correspondent à des segments de la trajectoire entre des instants aléatoires de visites dans un ensemble bien choisi (l'atome) formant une séquence de renouvellement. Dans la première partie de la thèse nous proposons un théorème fonctionnel de la limite centrale de type bootstrap pour des chaînes de Markov Harris récurrentes, d'abord dans le cas de classes de fonctions uniformément bornées puis dans un cadre non borné. Ensuite, nous utilisons les résultats susmentionnés pour obtenir un théorème de la limite centrale pour des fonctionnelles Fréchet différentiables dans un cadre Markovien. Motivés par diverses applications, nous discutons la manière d'étendre certains concepts de robustesse à partir du cadre i.i.d. à un cas Markovien. En particulier, nous considérons le cas où les données sont des processus Markoviens déterministes par morceaux. Puis, nous proposons des procédures d'échantillonnage résiduel et wild bootstrap pour les processus périodiquement autorégressifs et établissons leur validité. Dans la deuxième partie de la thèse, nous établissons des versions maximales d'inégalités de concentration de type Bernstein, Hoeffding et des inégalités de moments polynomiales en fonction des nombres de couverture et des moments des temps de retour et des blocs. Enfin, nous utilisons ces inégalités sur les queues de distributions pour calculer des bornes de généralisation pour une estimation d'ensemble de volumes minimum pour les chaînes de Markov régénératives.

Title : Bootstrap and uniform bounds for Harris Markov chains

Keys words : Markov chains, regenerative processes, bootstrap, robustness, concentration inequalities, statistical learning

Abstract : This thesis concentrates on some extensions of empirical processes theory when the data are Markovian. More specifically, we focus on some developments of bootstrap, robustness and statistical learning theory in a Harris recurrent framework. Our approach relies on the regenerative methods that boil down to division of sample paths of the regenerative Markov chain under study into independent and identically distributed (i.i.d.) blocks of observations. These regeneration blocks correspond to path segments between random times of visits to a well-chosen set (the atom) forming a renewal sequence. In the first part of the thesis we derive uniform bootstrap central limit theorems for Harris recurrent Markov chains over uniformly bounded classes of functions. We show that the result can be generalized also to the unbounded case. We use the aforementioned results to obtain uniform bootstrap central limit theorems for Fréchet differentiable functionals of Harris Markov chains. Propelled by vast applications, we discuss how to extend some concepts of robustness from the i.i.d. framework to a Markovian setting. In particular, we consider the case when the data are Piecewise-deterministic Markov processes. Next, we propose the residual and wild bootstrap procedures for periodically autoregressive processes and show their consistency. In the second part of the thesis we establish maximal versions of Bernstein, Hoeffding and polynomial tail type concentration inequalities. We obtain the inequalities as a function of covering numbers and moments of time returns and blocks. Finally, we use those tail inequalities to derive generalization bounds for minimum volume set estimation for regenerative Markov chains.



*High in the sky
There can be seen towering
A tall mountain,
Were one but wish to climb it
A path of ascent exists.*

Emperor Meiji (1852-1912), Japan

Acknowledgements

First and foremost, I want to thank my thesis advisors Patrice Bertail and Stephan Cl  men  on for giving me the opportunity to do my PhD in T  l  com ParisTech. I would like to express my gratitude to Patrice Bertail for introducing me to the field of Markov chains, bootstrap and concentration inequalities. I am very grateful for all your remarks, advice and research directions I have received from both of you. I thank Stephan Cl  men  on for his encouragement and opportunity to present my research results outside the school.

I am thankful to Kengo Kato and Olivier Wintenberger to have accepted to review my thesis manuscript and their valuable remarks and future research directions suggestions.

I also want to thank the other members of jury, Sana Louhichi, C  cile Durot, and Randal Douc for participation in my defense and insightful remarks concerning my work.

I am very grateful to Kengo Kato for his mentorship in the University of Tokyo during my research stay in the summer term 2018. It was a real privilege and an honour to work with you. Thank you for all the discussions, constructive suggestions and encouragement. I have learned from you a lot about empirical processes theory.

I want to express my gratitude to my co-authors Charles Tillier, Pawe   Potorski and Balamurugan Palaniappan. It was a great pleasure to work with you. I am also grateful to Fran  ois Portier for many scientific discussions.

I thank Professor Szymon Peszat for making it possible to join PhD programme in the University of Science and Technology in Poland, and all his encouragement and support through all the years.

I thank Taiji Suzuki for giving me the opportunity to do my internship in RIKEN. I also want to thank all my Japanese colleagues for very warm welcome and hospitality when I was in Tokyo.

I am very grateful to Mainak Jas and Balamurugan Palaniappan for sharing with me your coding expertise!

I am very thankful to Kam  lia Daudel, Jean Baptiste Schiratti and Amaury Durand for all your suggestions concerning my French scientific writing. Your remarks were very

valuable.

I thank all my friends from the laboratory, Aakanksha, Kimsy, Giorgia, Kamélia, Safa, Jean-Baptiste, Mainak, Pavlo, Nilesh, Amaury, Mastane, Bala, Charles, Lucas, Karim, Robert, Robin, Pierre, Umut and many others. It was a great pleasure to spend time with you and have you around!

I want to give my special thanks to Jonathan. You are amazing friend and like a brother to me.

I thank all my family for all their love and support through all the years.

Finally I want to thank all other persons I did not mention here but made it possible for me to grow as a person and researcher through all the years.

List of publications

Published articles, chapters and conference papers

G. Ciolek. Bootstrap uniform central limit theorems for Harris recurrent Markov chains. *Electronic Journal of Statistics*, 10:2157–2178, 2016.

G. Ciolek and P. Potorski. Bootstrapping periodically autoregressive models. *ESAIM: Probability and Statistics*, 21:2157–2178, 2017.

P. Bertail, G. Ciolek, and C. Tillier. Robust estimation for Markov chains with application to PDMP. In *Statistical Inference for Piecewise-Deterministic Markov Processes*. Wiley, 2018.

P. Bertail, G. Ciolek, and S. Cléménçon. Generalization bounds for minimum volume set estimation based on Markovian data. *ISAIM, International Symposium on Artificial Intelligence and Mathematics Proceedings Book*, pages 1–7, 2018.

P. Bertail and G. Ciolek. New Bernstein and Hoeffding type inequalities for regenerative Markov chains. Accepted for publication in *Latin American Journal of Probability and Mathematical Statistics*. November, 2018.

G. Ciolek. Bootstrapping Harris recurrent Markov chains. In *Nonparametric Statistics*, 3rd ISNPS, Avignon, France, June 2016, Springer Proceedings in Mathematics and Statistics. Due to December, 2018.

Articles in revision

P. Bertail, G. Ciolek, and S. Cléménçon. Statistical learning based on Markovian data. Maximal deviation inequalities and learning rates. Submitted, 2018.

Contents

1	Introduction	12
1	Motivation	12
2	Basic properties of regenerative and Harris recurrent Markov chains	16
2.1	Regenerative Markov chains	17
2.2	General Harris Markov chains and the splitting technique	19
2.3	Regenerative blocks for dominated families	21
2.4	A few examples of regenerative and Harris recurrent Markov chains	24
3	A preview of contributions and future perspectives	28
2	Bootstrap uniform central limit theorems for Harris recurrent Markov chains	39
1	Preliminaries	39
2	Non-parametric bootstrap for regenerative and Harris recurrent Markov chains	42
3	Main asymptotic results	47
4	Bootstrap uniform central limit theorems for Fréchet differentiable functionals of Markov chains	52
4.1	Preliminary assumptions and remarks	52
4.2	Main asymptotic results	54
5	Conclusion	57
3	Robust estimation for Markov chains with applications to PDMPs	59
1	Definition	60
2	Robust functional parameter estimation for Markov Chains	61
2.1	The influence function on the torus	61
2.2	Example 1: Sample means	63
2.3	Example 2: M-estimators.	64
2.4	Example 3: Quantiles.	65

3	Fréchet differentiability of functionals of Markov chains	65
4	A Markov view for estimators in PDMPs	68
4.1	Example: Sparre-Andersen model with barrier	69
4.2	Example: Kinetic Dietary Exposure Model	72
5	Robustness for risk PDMP models	74
6	Stationary measure	74
6.1	Ruin probability	77
6.2	Extremal index	81
7	Expected shortfall	84
8	Simulations	85
9	Conclusion	90
4	Residual and wild bootstrap methods for periodically autoregressive processes	91
1	Preliminaries and Markovian form of PAR(p) processes	91
2	The least squares estimation for model's parameters	97
3	Residual bootstrap for PAR processes	99
4	Wild bootstrap for PAR(p) time series	104
5	Simulations	106
6	Conclusion	107
5	Maximal concentration inequalities for regenerative and Harris recurrent Markov chains	116
1	Preliminaries	116
2	Bernstein and Hoeffding type deviation inequalities for Markov chains	122
2.1	Bernstein and Hoeffding type bounds for atomic regenerative Markov chains	122
2.2	Maximal concentration inequalities under uniform entropy	132
2.3	Bernstein and Hoeffding type tail inequalities for Harris recurrent Markov chains	136
3	Polynomial tail maximal concentration inequality for Markov chains	138
4	Bound of the expectation of the supremum of an empirical process in a Markovian setting	141
5	Conclusion	144

6	Minimum volume set estimation for Markovian data	146
1	Preliminaries	146
2	Minimum Volume Set Estimation	148
3	Minimum Volume Set Estimation - Generalization Results	151
3.1	Harris recurrent case	156
4	Simulations	157
5	Conclusion	159
7	Résumé substantiel	164
1	Motivation	164
2	Propriétés fondamentales des chaînes de Markov régénératives atomiques et Harris récurrentes	168
2.1	Chaînes de Markov régénératives atomiques	169
2.2	Chaînes de Markov Harris récurrentes et scission de Nummelin	171
2.3	Blocs régénératifs pour les familles dominées	173
3	Un aperçu des contributions et des perspectives futures	176
4	Liste des travaux	188
5	Liste des communications orales	188
	Bibliography	202

Chapter 1

Introduction

This thesis concentrates on some extensions of empirical processes theory when the data are Markovian. We focus on some developments of bootstrap, robustness and statistical learning theory in atomic regenerative and Harris recurrent framework.

1 Motivation

The theory of empirical processes plays a crucial role in modern statistics. It delivers necessary tools that allow to tackle many statistical problems in various fields, e.g. spectral analysis, extreme value theory, bootstrap and statistical learning. The theory of empirical processes for the i.i.d. data is well-studied, see for instance [87], [119] and [141].

However, in practice, the i.i.d. assumption often seems to be unrealistic. The data coming from applications in fields such as climatology, genetics, finance, geology or telecommunication are inherently temporal by nature and consequently not i.i.d. processes. This motivates researchers to extend the concepts of theory of empirical processes from the i.i.d. case into dependent framework (see for instance [54] for an exhaustive survey on such developments for stationary sequences and long-range dependent data).

The reason for investigation of regenerative atomic and Harris recurrent Markov chains is driven both by theoretical and applicative considerations. Firstly, the class of Markov chains (including chains with infinite memory) is very general and may approach a lot of time series (including non-stationary processes); it is used in many econometric models involving dependent data. The special regenerative structure of Markov chains (see for instance [112], [12] and [111]) makes them ideal tools for extending some results from the i.i.d. setting to the dependent case. Indeed, it is known since the work of [111], that it is possible to cut (theoretically) regular Harris recurrent Markov chains into independent blocks by using

an adequate probabilistic extension of the chain. The theory of aforementioned class of Markov chains is well-studied, one can refer to [106], [136] and [111]. More specifically, one can also look into [98] and [44] to find central limit theorems for such processes, some bootstrap developments in [26], [122], [75] and [44], deviation inequalities in [2], [28], [22] and applications to statistical learning in [23], [24].

There are many real world situations when the data exhibit regenerative atomic and Harris recurrent structure. The classical examples involve storage and queuing systems, as well as many models in finance, insurance or food risk assessment (see [106], [27], [31] and [25]). Many results for Markov chains has been established under mixing properties which are difficult to verify in practice. This was an additional encouragement to make use of regenerative techniques in order to work under return time and block moment conditions (which in many cases may be more tractable) instead.

The first part of the thesis is devoted to bootstrap considerations. The naive bootstrap scheme was proposed by Efron in his seminal paper [59]. Given n i.i.d. observations X_1, \dots, X_n distributed according to unknown distribution F , one may want to estimate the sampling distribution of some functional $R_n(X_1, \dots, X_n, F)$. Here, the simple non-parametric bootstrap method becomes handy. Firstly, we construct the empirical version of F , i.e. $F_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$. The next step involves drawing n times from F_n bootstrap observations X_i^* , $i = 1, \dots, n$, which are i.i.d. conditionally on F_n . Now, we approximate the sampling distribution of some functional of interest $R_n(X_1, \dots, X_n, F)$ by the distribution of $R_n^*(X_1^*, \dots, X_n^*, F_n)$ conditionally on F_n . One may simply ask why we will not use central limit theorems in order to study the behaviour of $R_n(X_1, \dots, X_n, F)$. This approach is however not always possible since the closed form of limiting distribution may be difficult to obtain (refer for instance to [121] for more details and references). Moreover, it is quite common that the distribution of limiting process depends on some unknown parameters. Briefly speaking, bootstrap methods remedy the aforementioned problems, we elaborate on it more in Chapter 2.

The naive non-parametric bootstrap method for the i.i.d. data has gradually evolved and new types of bootstrap schemes in both i.i.d. and dependent settings were established (see [76], [93] and [43], [42] for results in high dimensions). This led to tremendously vast number of applications in almost all fields of statistics. Under conditions that hold in a wide variety of econometric applications, bootstrap provides approximations to distributions of statistics, coverage probabilities of confidence intervals, and rejection probabilities of hypothesis tests that are more accurate than the approximations of first-order asymptotic distribution theory (refer to [80] and [76] for details).

With raising interest in the statistical inference in dependent framework, new bootstrap procedures have been developed. Most of the schemes in the dependent setting rely on block techniques. These approaches essentially boil down to resampling block segments of observations so that dependence structure is captured. There are many variants of block bootstrap methods for dependent data such as moving block bootstrap (MBB), non-overlapping block bootstrap (NBB) or circular block bootstrap (CBB) to name just a few (see for instance [93] for an exhaustive overview of the aforementioned procedures). Regrettably, as indicated by many authors (see for instance [46] and [93]), these procedures struggle with many problems. The large drawback is that block bootstrap methods are very sensitive to the choice of the length of the blocks. Indeed, the optimal length of the blocks heavily depend on the sample size and the data generating processes. Moreover, popular MBB method requires the stationarity for observations that usually results in failure of this method in non-stationary setting (see [93] for more details). Furthermore, the asymptotic behaviour of MBB method is highly dependent on the estimation of the bias and of the asymptotic variance of the statistic of interest that is a significant drawback when considering practical applications. Finally, it is noteworthy, that the rate of convergence of the MBB distribution is slower than that's of bootstrap distribution in the i.i.d. setting.

We focused on recalling block bootstrap methods when dealing with dependent data since the bootstrap procedures we consider in this thesis are also based on dividing the data into block segments. However, there are many other bootstrap schemes one can use in dependent setting, such as residual, wild or sieve bootstrap to name just a few. We refer to [89] or [93] for more details concerning applicability of those schemes and their limitations.

Taking into consideration the limitations of block bootstrap methods, we decided to focus on regenerative techniques for atomic and Harris recurrent Markov chains. In the seminal paper [26] the regenerative block bootstrap (RBB) and approximate regenerative block bootstrap method (ARBB) are introduced. The aforementioned procedures do not require choice of the length of the blocks, moreover, in atomic case, the division of data into blocks is completely data driven. It is also shown in [26] that bootstrap central limit theorems for the mean in Markovian setting hold. We developed this theory further by establishing uniform bootstrap CLT's over not necessarily bounded classes of functions. The uniform bootstrap central limit theorems are helpful when proving the validity of bootstrap procedures (see [122], [44] and [71]) and may be used in many statistical applications, one may be interested in getting bootstrap versions of the results in [110], [68] and [69].

One can use bootstrap central limit theorems stated in [44] in order to establish bootstrap limit results for Fréchet differentiable functionals of Harris Markov chains. Fréchet

differentiability is a vital concept in robust statistics since it guarantees the existence of influence function which allows to detect the outliers in the data (see [140] for details). In fact, it is shown in [29] that some of the concepts of robustness can be naturally extended to a Harris Markovian case. More specifically, one can detect outliers and construct robust plug-in estimators by eliminating blocks having either too large contribution to the statistics of interest or having too large length resulting in an important bias on the statistics (instead of consideration of the impact of a single observation on a given statistic). A further development of these ideas has been done in Chapter 3 (see also [25]) with a special focus on applications to piecewise-deterministic Markov processes. Robust statistical methods are applied to the solutions of many problems such as estimation of regression parameters, estimation of scale and location or in statistical learning.

Our second direction concerning bootstrap developments in a dependent framework is a study periodically autoregressive processes (PAR) which are an example of Harris recurrent Markov chains. We propose residual and wild bootstrap methods and prove their validity. The aforementioned methods are data-driven and do not need any block length calibration which may be attractive for practitioners.

The second part of the thesis concentrates on applications of empirical processes theory to statistical learning. Not surprisingly, the statistical learning theory is mostly studied in the i.i.d. case (see [38], [142], [144] and [64]). However, there is a huge demand mostly propelled by the field of big data for some extensions to the dependent framework. As mentioned in [149], applications such as market prediction, system diagnosis, and speech recognition are inherently temporal in nature, and consequently not i.i.d. processes. Machine learning theory for dependent processes has been intensively investigated in the last years, see for instance [3], [5], [132] or [78] for some results stated in a very general setting.

In statistical learning theory, numerous works established non-asymptotic bounds assessing the generalization capacity of empirical risk minimizers under a large variety of complexity assumptions for the class of decision rules over which optimization is performed, by means of sharp control of uniform deviation of i.i.d. averages from their expectation, while fully ignoring the possible dependence across training data in general. The sharp control of the supremum distance between averages of random variables and their expectation is managed by concentration inequalities that give an upper bound on the tail probability for suprema of empirical processes. Those, at first glance, very theoretical probabilistic results are the crucial tools when investigating the learning capacity of statistical learning algorithms. There are many concentration results for dependent data, in (pseudo-) regenerative Markovian setting one should mention [28], [2] and [1]. Despite of various concentration results we establish

maximal type concentration inequalities tailor-made for our applications, i.e. they are established for non-stationary Markov processes and hold for unbounded classes of functions \mathcal{F} and involve easy to interpret parameters in the bound.

In this thesis we are interested in establishing generalization bounds (so-called error bounds) for statistical learning algorithms when the data are Markovian. We obtain such results via empirical risk minimization approach. Our strategy essentially boils down to 3 steps.

- We obtain concentration inequalities for bounded and unbounded classes of functions of Markov chains (see [22], [23] and [24]). Exponential (for instance Bernstein and Hoeffding) and polynomial tail inequalities are an essential tool when one wants to conduct empirical risk minimization or derive the rate of convergence of a statistical learning algorithm.
- We investigate the performance of the learning algorithms (when dealing with Harris Markov chain samples) via empirical risk minimization. It is noteworthy that the analysis of the ERM algorithm and consistency and properties of statistical learning procedures are very significant and urgent problems to solve.
- We investigate the generalization properties of selected statistical learning algorithm (in atomic regenerative and Harris Markovian setting).

In this thesis we present generalization bounds for minimum volume (MV) set estimation problem. The concept of MV-set estimation was for the first time introduced in [120] and extends the concept of quantile for multivariate probability distributions (see [120] for details). This method offers a non-parametric framework for (unsupervised) anomaly/novelty detection. As observed in [126], MV-set estimation can be cast in a learning framework very similarly to empirical risk minimization in (supervised) classification. Generalization bounds for MV-set estimation problem were established in [126] in the i.i.d. setting. Some developments were also made in a dependent case in [55]. We extended the aforementioned results to atomic regenerative and Harris recurrent setting in [23] and [24].

2 Basic properties of regenerative and Harris recurrent Markov chains

In this section we introduce some notations and recall the key concepts of Markov chains theory (we refer to [106] and [136], [27] and [112] for exhaustive reviews). The results

and statements provided in this section have informative character. An interested reader is advised to look into aforementioned literature for proofs of theorems stated for Markov chains on general (non-countable) state space. Throughout all this section \mathbb{I}_A is the indicator function of the event A .

Let $X = (X_n)_{n \in \mathbb{N}}$ be a homogeneous Markov chain on a countably generated state space (E, \mathcal{E}) with transition probability Π and initial probability ν . Note that for any $B \in \mathcal{E}$ and $n \in \mathbb{N}$, we have

$$X_0 \sim \nu \text{ and } \mathbb{P}(X_{n+1} \in B | X_0, \dots, X_n) = \Pi(X_n, B) \text{ a.s.}$$

In what follows, \mathbb{P}_x (resp. \mathbb{P}_ν) is the probability measure such that $X_0 = x$ and $X_0 \in E$ (resp. $X_0 \sim \nu$), and we write $\mathbb{E}_x(\cdot)$ for the \mathbb{P}_x -expectation (resp. $\mathbb{E}_\nu(\cdot)$ is the \mathbb{P}_ν -expectation). The following definitions formalize the idea of a communication structure of Markov chains we consider (see [106] for more details).

Definition 1. We say that $X = (X_n)_{n \in \mathbb{N}}$ is ψ -irreducible if there exists a measure ψ on \mathcal{E} such that whenever $\psi(B) > 0$, we have

$$\mathbb{P}_x(\tau_B < \infty) > 0 \text{ for all } x \in E$$

and τ_B is the first time when X hits B .

Definition 2. We say that a Markov chain X is *periodic* if there exist $d' > 0 \in \mathbb{N}$ and disjoint sets $D_1, D_2, \dots, D_{d'}$ (with convention $D_{d'+1} = D_1$) weighted by ψ such that

$$\psi(E \setminus \cup_{1 \leq i \leq d'} D_i) = 0 \text{ and } \forall x \in D_i \Pi(x, D_{i+1}) = 1.$$

The period of the chain is the greatest common divisor d (g.c.d. d) of such integers. In case $d = 1$ we say that X is *aperiodic*.

In the following, we assume that X is ψ -irreducible and aperiodic, unless it is specified otherwise.

2.1 Regenerative Markov chains

In this thesis, our particular interest concentrates on atomic structure of Markov chains due to its abilities to extend the theory of empirical processes (vital for developments in the field of bootstrap and statistical learning) from the i.i.d. case to a Markovian framework.

Definition 3. Suppose that X is aperiodic and ψ -irreducible. A set $A \in \mathcal{E}$ is an accessible atom if for all $x, y \in A$ we have $\Pi(x, \cdot) = \Pi(y, \cdot)$ and $\psi(A) > 0$. In that case we call X atomic.

Intuitively speaking, the atom is a set from which all the transition probabilities of X are the same. Consequently, whenever X hits A , it forgets its past and starts afresh (regenerates).

The strong Markov property (see [106] for rigorous justification) guarantees that given any initial law ν , the sample paths can be split into i.i.d. blocks corresponding to the consecutive visits of the chain to the atom A . The segments of data are of the form:

$$B_j = (X_{1+\tau_A(j)}, \dots, X_{\tau_A(j+1)}), \quad j \geq 1$$

and take values in the torus $\mathbb{T} = \cup_{k=1}^{\infty} E^k$.

Being ensured that our chain possesses the atomic structure, we define the sequence of regeneration times $(\tau_A(j))_{j \geq 1}$. The sequence consists of the successive points of time when the chain forgets its past. Let

$$\tau_A = \tau_A(1) = \inf\{n \geq 1 : X_n \in A\}$$

be the first time when the chain hits the regeneration set A and

$$\tau_A(j) = \inf\{n > \tau_A(j-1), X_n \in A\} \text{ for } j \geq 2.$$

We introduce few more pieces of notation: throughout the thesis we write

$$l_n = \sum_{i=1}^n \mathbb{I}\{X_i \in A\}$$

for the total number of consecutive visits of the chain to the atomic set A , thus we have $l_n + 1$ data blocks. We make the convention that $B_{l_n}^{(n)} = \emptyset$ when $\tau_A(l_n) = n$. Furthermore, we denote by

$$l(B_j) = \tau_A(j+1) - \tau_A(j), \quad j \geq 1,$$

the length of regeneration blocks.

In order to make this exposition concise, we provide the blocks' construction scheme below. Step 3 can be omitted depending on application. We assume that we observe the sample $\mathbb{X}_n = (X_1, \dots, X_n)$.

Algorithm 1 Regeneration blocks construction

Step 1 Count the total number of visits $l_n = \sum_{i=1}^n \mathbb{I}\{X_i \in A\}$ to atom A up to time n .

Step 2 Divide the data \mathbb{X}_n into $l_n + 1$ blocks according to the consecutive visits of the trajectory to the atom A , i.e.

$$B_0 = (X_1, \dots, X_{\tau_A(1)}), \dots, B_j = (X_{\tau_A(j)+1}, \dots, X_{\tau_A(j+1)}), \dots, \\ B_{l_n-1} = (X_{\tau_A(l_n-1)+1}, \dots, X_{\tau_A(l_n)}), B_{l_n}^{(n)} = (X_{\tau_A(l_n)+1}, \dots, X_n).$$

Step 3 Discard the first block B_0 and the last one $B_{l_n}^{(n)}$ if $\tau_A(l_n) < n$.

In our framework, we also are interested in the asymptotic behaviour of positive recurrent Harris Markov chains.

Definition 4 (Harris recurrent Markov chain). Assume that X is a ψ -irreducible Markov chain. Chain X is *Harris recurrent* iff, starting from any point $x \in E$ and any set such that $\psi(A) > 0$, we have

$$\mathbb{P}_x(\tau_A < +\infty) = 1.$$

Note that the property of Harris recurrence ensures that X visits set A infinitely often a.s..

In our framework, we are interested in the steady-state analysis of Markov chains. More specifically, when an invariant measure is finite, then we can normalize it to a stationary probability measure.

Theorem 1 (Kac's theorem). Assume that Markov chain X is ψ -irreducible and admits an atom A . Then, X is positive recurrent if and only if $\mathbb{E}_A(\tau_A) < \infty$. The unique invariant probability distribution μ is the Pitman's occupation measure given by

$$\mu(B) = \frac{1}{\mathbb{E}_A(\tau_A)} \mathbb{E}_A \left(\sum_{i=1}^{\tau_A} \mathbb{I}\{X_i \in B\} \right), \forall B \in \mathcal{E}.$$

Note that the by the Kac's theorem we have that

$$\mathbb{E}(l(B_j)) = \mathbb{E}_A(\tau_A) = \frac{1}{\mu(A)}.$$

Consider μ -integrable function $f : E \rightarrow \mathbb{R}$. By

$$u_n(f) = \frac{1}{\tau_A(l_n) - \tau_A(1)} \sum_{i=1+\tau_A(1)}^{\tau_A(l_n)} f(X_i)$$

we denote the estimator of the unknown asymptotic mean $\mathbb{E}_\mu(f(X_1))$.

2.2 General Harris Markov chains and the splitting technique

In this section we explain how the regenerative techniques can be stretched from atomic case to a Harris recurrent framework due to work of [12] and [111]. In particular, Nummelin [111] proposed the so-called *splitting technique* which, in simple words, allows to extend the probabilistic structure of any Harris chain in order to artificially construct a regeneration set. In this section, unless specified otherwise, X is assumed to be a Harris recurrent Markov chain with transition kernel Π . In this section we follow closely the notation from [27], we also refer therein for more details concerning the theory presented here.

Definition 5. We say that a set $S \in \mathcal{E}$ is small if there exists a parameter $\delta > 0$, a positive probability measure Φ supported by S and an integer $m \in \mathbb{N}^*$ such that

$$\forall x \in S, A \in \mathcal{E} \quad \Pi^m(x, A) \geq \delta \Phi(A), \quad (1.1)$$

where Π^m denotes the m -th iterate of the transition probability Π .

When (1.1) is satisfied, we say that the chain satisfies the minorization condition $\mathcal{M} = \mathcal{M}(m, S, \delta, \Phi)$. For the simplicity's sake, throughout the rest of this thesis, the minorization condition \mathcal{M} is satisfied with $m = 1$, unless specified otherwise. In order to generalize the results to case $m > 1$ one can replace the chain $(X_n)_{n \in \mathbb{N}}$ by the chain $((X_{nm}, \dots, X_{n(m+1)-1}))_{n \in \mathbb{N}}$.

Remark 1. In general case, it is not obvious that small sets having positive irreducible measure exist. As pointed out in [82] they do exist for any irreducible kernel Π under the assumption that the state space is countably generated.

In what follows, we expand the sample space in order to define a sequence $(Y_n)_{n \in \mathbb{N}}$ of independent r.v.'s with parameter δ . Let $\mathbb{P}_{\nu, \mathcal{M}}$ be the joint distribution of $X^{\mathcal{M}} = (X_n, Y_n)_{n \in \mathbb{N}}$. The construction is based on the mixture representation of Π on S . It can be retrieved by the following randomization of the transition probability Π each time the chain X visits the set S . If $X_n \in S$ and

- if $Y_n = 1$ (which occurs with probability $\delta \in]0, 1[$), then X_{n+1} is distributed according to the probability measure Φ ,
- if $Y_n = 0$ (that occurs with probability $1 - \delta$), then X_{n+1} is distributed according to the probability measure $(1 - \delta)^{-1}(\Pi(X_n, \cdot) - \delta\Phi(\cdot))$.

In what follows we introduce one more piece of notation. Let

$$Ber_\delta(\beta) = \delta\beta + (1 - \delta)(1 - \beta)$$

for $\beta \in \{0, 1\}$. The bivariate Markov chain $X^{\mathcal{M}}$ is called the *split chain*. Note that it takes its values in $E \times \{0, 1\}$ and possesses transition kernel $\Pi_{\mathcal{M}}$ given by

- for any $x \notin S, B \in \mathcal{E}, \beta$ and β' in $\{0, 1\}$,

$$\Pi_{\mathcal{M}}((x, \beta), B \times \{\beta'\}) = Ber_\delta(\beta') \times \Pi(x, B),$$

- for any $x \in S, B \in \mathcal{E}, \beta'$ in $\{0, 1\}$,

$$\Pi_{\mathcal{M}}((x, 1), B \times \{\beta'\}) = Ber_\delta(\beta') \times \Pi(x, B) \times \Phi(B),$$

$$\Pi_{\mathcal{M}}((x, 0), A \times \{\beta'\}) = Ber_\delta(\beta') \times (1 - \delta)^{-1}(\Pi(x, B) - \delta\Phi(B)).$$

Note that $X^{\mathcal{M}}$ possesses an atom $S \times \{1\}$. Observe that the split chain $X^{\mathcal{M}}$ inherits all the stability and communication properties of the chain X (refer to [106] and [112] for a rigorous treatment).

Remark 2. It should be noted that the blocks created via splitting technique are i.i.d. in case when $m = 1$ in minorization condition (1.1)). If the chain X satisfies $\mathcal{M}(m, S, \delta, \Phi)$ for $m > 1$, then the blocks of data are 1-dependent. In many cases, however, it is easy to adapt the theory from the case when $m = 1$ by considering sums of odd and even blocks in order to deal with dependence between B'_j s (see for instance [98] or [95]) or by vectorizing the chain (see [106]).

2.3 Regenerative blocks for dominated families

We suppose that the family of the conditional distributions $\{\Pi(x, dy)\}_{x \in E}$ and the initial distribution ν are dominated by a σ -finite measure λ of reference, so that

$$\nu(dy) = f(y)\lambda(dy) \quad \text{and} \quad \Pi(x, dy) = p(x, y)\lambda(dy),$$

for all $x \in E$. The minorization condition requests that Φ is absolutely continuous with respect to λ and that

$$p(x, y) \geq \delta\phi(y), \quad \lambda(dy) \quad \text{a.s. for any } x \in S$$

with $\Phi(dy) = \phi(y)dy$. In what follows, let Y be a binary random sequence obtained via the Nummelin's technique from the parameters given by condition \mathcal{M} . Observe that the distribution of $Y^{(n)} = (Y_1, \dots, Y_n)$ conditionally to $X^{(n+1)} = (x_1, \dots, x_{n+1})$ is the tensor product of Bernoulli distributions given by: for all $\beta^{(n)} = (\beta_1, \dots, \beta_n) \in \{0, 1\}^n$, $x^{(n+1)} = (x_1, \dots, x_{n+1}) \in E^{n+1}$,

$$\mathbb{P}_\nu(Y^{(n)} = \beta^{(n)} \mid X^{(n+1)} = x^{(n+1)}) = \prod_{i=1}^n \mathbb{P}_\nu(Y_i = \beta_i \mid X_i = x_i, X_{i+1} = x_{i+1}),$$

with, for $1 \leq i \leq n$,

- if $x_i \notin S$, $\mathbb{P}_\nu(Y_i = 1 \mid X_i = x_i, X_{i+1} = x_{i+1}) = \delta$,
- if $x_i \in S$, $\mathbb{P}_\nu(Y_i = 1 \mid X_i = x_i, X_{i+1} = x_{i+1}) = \delta\phi(x_{i+1})/p(x_i, x_{i+1})$.

Note that given $X^{(n+1)}$, from $i = 1$ to n , Y_i is distributed according to the Bernoulli distribution with parameter δ , unless X has hit the small set S at time i : then, Y_i is

drawn from the Bernoulli distribution with parameter $\delta\phi(X_{i+1})/p(X_i, X_{i+1})$. We denote by $\mathcal{L}^{(n)}(p, S, \delta, \phi, x^{(n+1)})$ this probability distribution. If we were able to generate Y_1, \dots, Y_n , so that

$$X^{\mathcal{M}(n)} = ((X_1, Y_1), \dots, (X_n, Y_n))$$

be a realization of the split chain $X^{\mathcal{M}}$, then we could do the block decomposition of the sample path $X^{\mathcal{M}(n)}$ leading to asymptotically i.i.d. blocks. Note that this procedure requires knowledge of the transition density $p(x, y)$ in order to generate random variables (Y_1, \dots, Y_n) . However, in practice the transition density is unknown and needs to be estimated. As a consequence, we can not directly use the procedure stated above and need to apply its approximated version which was proposed in [26]. The construction consists of two steps, firstly, build an estimator $p_n(x, y)$ of $p(x, y)$ based on $X^{(n+1)}$, i.e. $p_n(x, y)$ which fulfills

$$p_n(x, y) \geq \delta\phi(y), \quad \lambda(dy) - \text{a.s. and } p_n(x, y) > 0, 1 \leq i \leq n. \quad (1.2)$$

In the second step, generate random vector $\hat{\mathbb{Y}}_n = (\hat{Y}_1, \dots, \hat{Y}_n)$ conditionally to $X^{(n+1)}$ from distribution $\mathcal{L}^{(n)}(p_n, S, \delta, \gamma, X^{(n+1)})$ which is an approximation of the conditional distribution $\mathcal{L}^{(n)}(p, S, \delta, \gamma, X^{(n+1)})$ of (Y_1, \dots, Y_n) for given $X^{(n+1)}$. The validity of this approximation has been shown in [26].

In this setting, we define the successive hitting times of $A_{\mathcal{M}} = S \times \{1\}$ as

$$\hat{\tau}_{A_{\mathcal{M}}}(i), \quad i = 1, \dots, \hat{l}_n, \quad \text{where } \hat{l}_n = \sum_{i=1}^n \mathbb{I}\{X_i \in S, \hat{Y}_i = 1\}$$

is the total number of visits of the split chain to $A_{\mathcal{M}}$ up to time n . Below we provide approximated block construction scheme. Let $\mathbb{X}_{n+1} = (X_1, X_2, \dots, X_{n+1})$ be random sample drawn from Harris chain X . We assume that X fulfills assumptions stated previously in this section. Step 5 may be omitted depending on application.

Algorithm 2 Approximate regeneration blocks construction

Step 1 Construct an estimator $p_n(x, y)$ of the transition density using sample \mathbb{X}_{n+1} . The estimator $p_n(x, y)$ must fulfill the conditions in (1.2)

Step 2 Conditioned on \mathbb{X}_{n+1} , draw $(\hat{Y}_1, \dots, \hat{Y}_n)$ from $\mathcal{L}^{(n)}(p_n, S, \delta, \gamma, \mathbb{X}_{n+1})$. In practice, one draws \hat{Y}_i 's only at those time points when $X_i \in S$ (see [27] for details). At such time point i , draw \hat{Y}_i from the Bernoulli distribution with parameter $\delta\gamma(X_{i+1})/p_n(X_i, X_{i+1})$.

Step 3 Count the number of visits

$$\hat{l}_n = \sum_{i=1}^n \mathbb{I}\{X_i \in S, \hat{Y}_i = 1\}$$

to the atom $S_1 = S \times \{1\}$ up to time n .

Step 4 Cut the trajectory \mathbb{X}_{n+1} into $\widehat{l}_n + 1$ approximate regeneration blocks which correspond to successful consecutive visits of (X, \widehat{Y}) to S_1 . Approximated blocks are of the form

$$\widehat{B}_0 = (X_1, \dots, X_{\widehat{\tau}_{A_{\mathcal{M}}}(1)}), \dots, \widehat{B}_j = (X_{\widehat{\tau}_{A_{\mathcal{M}}}(j)+1}, \dots, X_{\widehat{\tau}_{A_{\mathcal{M}}}(j+1)}), \dots, \\ \widehat{B}_{\widehat{l}_n-1} = (X_{\widehat{\tau}_{A_{\mathcal{M}}}(\widehat{l}_n-1)+1}, \dots, X_{\widehat{\tau}_{A_{\mathcal{M}}}(\widehat{l}_n)}), \widehat{B}_{\widehat{l}_n}^{(n)} = (X_{\widehat{\tau}_{A_{\mathcal{M}}}(\widehat{l}_n)+1}, \dots, X_{n+1}).$$

Step 5 Discard the first block \widehat{B}_0 and the last one $\widehat{B}_{\widehat{l}_n}^{(n)}$ if $\widehat{\tau}_{S_1}(\widehat{l}_n) < n$.

In what follows, we denote by

$$\widehat{n}_{A_{\mathcal{M}}} = \widehat{\tau}_{A_{\mathcal{M}}}(\widehat{l}_n) - \widehat{\tau}_{A_{\mathcal{M}}}(1) = \sum_{i=1}^{\widehat{l}_n-1} l(\widehat{B}_i)$$

the total number of observations after the first and before the last pseudo-regeneration times. Let

$$\sigma_f^2 = \frac{1}{\mathbb{E}_{A_{\mathcal{M}}}(\tau_{A_{\mathcal{M}}})} \mathbb{E}_{A_{\mathcal{M}}} \left(\sum_{i=1}^{\tau_{A_{\mathcal{M}}}} \{f(X_i) - \mu(f)\}^2 \right)$$

be the asymptotic variance. Furthermore, we set that

$$\widehat{\mu}_n(f) = \frac{1}{\widehat{n}_{A_{\mathcal{M}}}} \sum_{i=1}^{\widehat{l}_n-1} f(\widehat{B}_i), \quad \text{where } f(\widehat{B}_j) = \sum_{i=1+\widehat{\tau}_{A_{\mathcal{M}}}(j)}^{\widehat{\tau}_{A_{\mathcal{M}}}(j+1)} f(X_i)$$

and

$$\widehat{\sigma}_n^2(f) = \frac{1}{\widehat{n}_{A_{\mathcal{M}}}} \sum_{i=1}^{\widehat{l}_n-1} \left\{ f(\widehat{B}_i) - \widehat{\mu}_n(f) \right\}^2.$$

As noted in [27], the splitting technique relies heavily on a minorization condition (1.1) and small set chosen. We find this information important and recall the details in the remark below. The choice of size of small set and sharpness of the uniform bound from below on the transition density of X in the minorization condition occur to be critical in order to obtain enough blocks. The following observation comes from [27] (see the aforementioned paper for examples of the choice of a small set for different time series).

Remark 3. Observe that if the size of a small set is increased, the uniform bound from below for transition density of X decreases. Thus, in practice, one has to manoeuvre the minorization conditions taking into consideration the following, for a given realization of the trajectory, increasing the size of the small set S causes an increment of the number of points of the trajectory that are candidates for determining a block. However, simultaneously, the probability of dividing the trajectory decreases as size of S is larger (since the uniform lower bound for $\{p(x, y)\}_{(x, y) \in S^2}$ decreases).

Finally, we briefly mention that there exists a relation between α -mixing coefficients and regeneration times for Harris recurrent Markov chains. In this framework we will work under moment conditions imposed on τ_A and block moment conditions instead of making use of mixing properties mainly due to the fact that mixing conditions are difficult to verify in practice. However, taking into consideration the huge number of works when the dependence between data is expressed in terms of mixing conditions, we provide few comments below.

Let \mathcal{F}_a^b be the σ -algebra generated by X_a, \dots, X_b . The strong α -mixing coefficient between σ -fields \mathcal{A} and \mathcal{B} is defined as

$$\alpha(\mathcal{A}, \mathcal{B}) := \sup_{(A, B) \in \mathcal{A} \times \mathcal{B}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|.$$

The strong mixing coefficients related to a sequence of random variables are defined by

$$\alpha(k) = \sup_n \sup_{A \in \mathcal{F}_{-\infty}^n} \sup_{B \in \mathcal{F}_{n+k}^{+\infty}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|.$$

Remark 4. Theorem 2 from [36] states that for stationary Harris chains if for some $\lambda \geq 0$ the sum $\sum_m m^\lambda \alpha(m) < \infty$, then for all $B \in \mathcal{E}$ such that $\mu(B) > 0$ we have

$$\mathbb{E}_\mu(\tau_B^{1+\lambda}) < \infty, \text{ where } \tau_B = \inf\{n \geq 1 : X_n \in B\}.$$

This result guarantees that the rate of decay of strong mixing coefficients is polynomial. This is a weaker condition, because usually the exponential rate of decay is assumed.

2.4 A few examples of regenerative and Harris recurrent Markov chains

In Section 1 we mentioned that many time series can be seen as atomic regenerative and Harris recurrent Markov chains. In what follows we will provide few simple examples in order to show that such processes appear quite naturally in the real world. The models we present here come from [106] and [25].

Example 1 (Storage model). We consider

- L_1, L_2, L_3, \dots , which are input times into storage systems
- inter-arrival times ΔL_i which are i.i.d. and distributed according to

$$G(\infty, t] = \mathbb{P}(\Delta L_i \leq t)$$

- random variables S_n which quantify the n -th input to the storage. S_n possesses a distribution

$$H(\infty, t] = \mathbb{P}(S_n \leq t).$$

The sequence S_1, \dots, S_n, \dots are random variables which are independent of each other as well as of the inter-arrival times.

In a simple storage model, between inputs, we observe steady removal from the storage system, at rate r (in more complicated storage systems the removal can be governed by different rules): in the period $[x, x+t]$ there is a decrease of the stored content by an amount t so that contents of a storage drop by an amount rt since we do not register next input. When the storage process reaches zero it remains on this level until the new input is observed (the storage model does not take negative values). Thus, we consider the process

$$\Phi_{n+1} = [\Phi_n + S_n - J_n]^+,$$

where $x^+ = \max(x, 0)$ and J_n are i.i.d. random variables with

$$\mathbb{P}(J_n \leq x) = G(\infty, x/r]$$

and $r > 0$. In the above setting $\Phi = \{\Phi_n\}$ is a storage process and atomic regenerative Markov chain (due to the fact that S_{n+1} does not depend on S_{n-1}, S_{n-2}, \dots etc., and since the sequence of ΔL_i 's is i.i.d.). When $\mathbb{E}[S_n] < \mathbb{E}[J_n]$, the chain returns infinitely often to 0 and 0 is an atom. Thus, whenever Φ reaches 0 we cut the trajectory and the new bock is created.

One of the most classical examples of Harris recurrent Markov chains is a class of autoregressive processes. Harris recurrence property of an autoregressive model has been firstly shown in [13]. We advise to look to the paper of [6], [106] and [27] for more examples of Markov chains exhibiting Harris recurrent structure.

Example 2 (Autoregressive process of order p). Consider AR(1) process

$$X_n = \rho X_{n-1} + \theta_n, \quad n \geq 1$$

on state space $E = \mathbb{R}$ and where the noise sequence is given by $\theta_1, \theta_2, \dots$, which are i.i.d. and distributed according to \mathbb{G} . We suppose that \mathbb{G} possesses an absolutely continuous component. Under assumptions that

$$|\rho| < 1 \quad \text{and} \quad \mathbb{E} \log |\theta| < \infty$$

the conditions of Theorem 2.1 from [6] are fulfilled and AR(1) process is a Harris Markov chain.

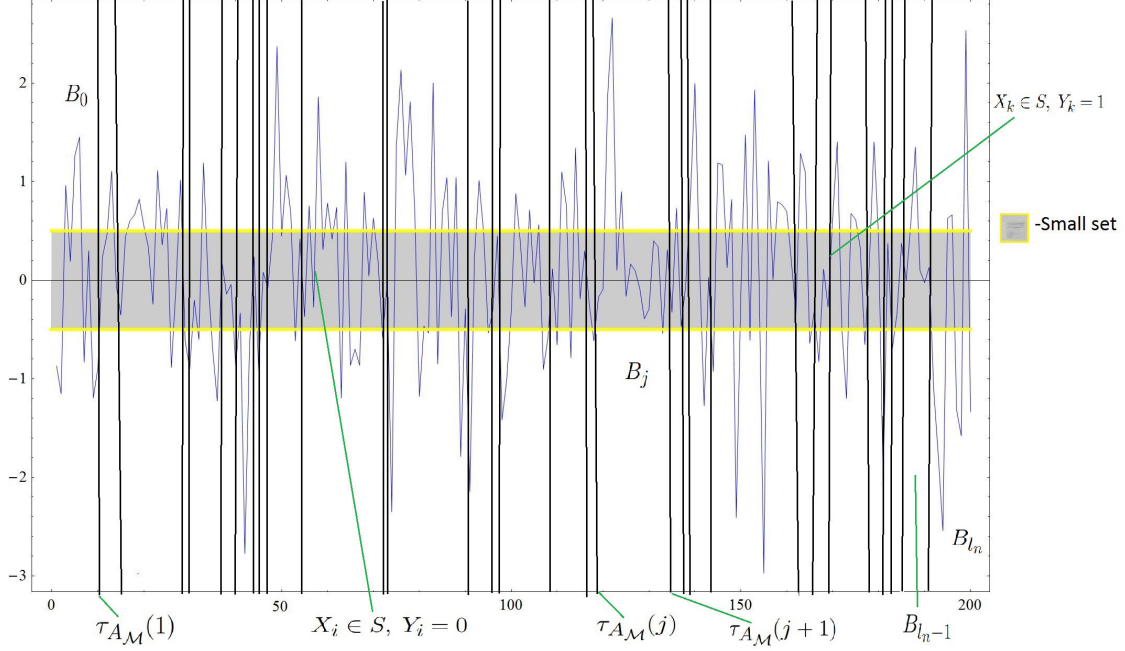


Figure 1.1: Regeneration block construction for AR(1) model.

As shown in [106] one can write the AR(p) process

$$X_n = \rho_1 X_{n-1} + \rho_2 X_{n-2} + \cdots + \rho_k X_{n-k} + \theta_n$$

in a Markovian form by constructing the multivariate sequence

$$Y_n = (X_n, \dots, X_{n-k+1})'$$

and considering the process $Y = \{Y_n, n \geq 0\}$. Indeed, Y is a Markov chain whose first component has exactly the sample paths of the autoregressive process (see [106] for more details) and by Theorem 2.1 in [6] is a Harris recurrent Markov chain (see Example 2.6 in [6] for more details). Figure 2 illustrates the splitting technique for a trajectory of AR(1) process.

As the last example we present the Kinetic Exposure Model introduced in [31]. This process is an example of a piecewise deterministic Markov process (PDMP) (see also [25] for other examples how to relate the properties of the PDMPs (stationary distribution) with the properties of embedded chains possessing regenerative structure).

Example 3 (Kinetic Dietary Exposure Model). The Kinetic Dietary Exposure Model (KDEM) is a stochastic process that expresses the evolution of a contaminant in the human body along time. Let

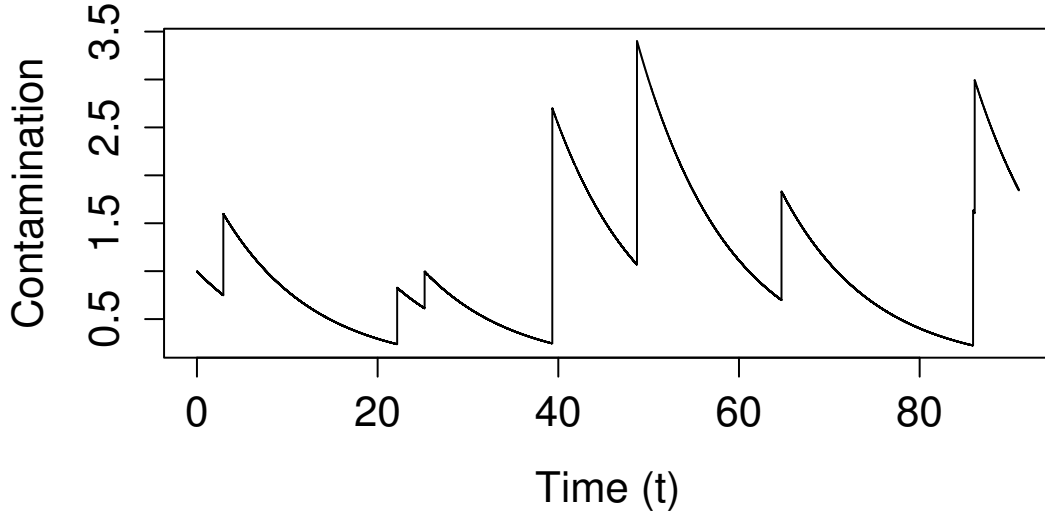


Figure 1.2: Kinetic Dietary Exposure Model which describes evolution of a contaminant in the human body.

- T_1, \dots, T_n be the contaminant's intake times
- $\Delta T_{n+1} = T_{n+1} - T_n$ be the inter-intake time
- U_1, \dots, U_n be the contaminant's intakes time at T_1, \dots, T_n

The counting process $\{(N(t))_{t \geq 0}$ is defined as

$$N(t) := \#\{i \in \mathbb{N}^* : T_i \leq t\}$$

for $t \geq 0$ is a *renewal process* and

$$A(t) = t - T_{N(t)}$$

is the backward recurrence time. We consider exposure process $X(t)$ which dynamics is given by the first order ordinary differential equation

$$dX(t) = -\omega X(t)dt \tag{1.3}$$

and $\omega > 0$ is a fixed parameter called elimination rate and corresponds to body's metabolism dealing with the chemical elimination of contaminants. The total body burden of a given

dietary contaminant at time t is :

$$X(t) = X(0) + \sum_{n=1}^{N(t)} U_n - \sum_{n=1}^{N(t)+1} \int_{T_{n-1}}^{T_n \wedge t} \omega X(s) ds.$$

By solving (1.3), the exposure process is given by

$$X(t) = X_{T_{N(t)}} e^{-\omega A(t)}$$

for any $t \geq 0$. The process $\{X(t)\}_{t \geq 0}$ (that is $X_n = X(T_n)$) is a PDMP. Let $\tilde{X} = (X_n)_{n \in \mathbb{N}}$, be an embedded chain of X . The evolution of the process \tilde{X} is given by the following stochastic recurrence equation

$$X_{n+1} = X_n \times e^{-\omega \Delta T_{n+1}} + U_{n+1}, \quad n \geq 0$$

which is an autoregressive process with a random coefficient.

Process \tilde{X} plays an important role in the analysis of X , i.e. it describes the exposure process immediately after each intake (refer to [31] for details). Under some additional assumptions (specified for instance in [31]) it is possible to relate the continuous-time process X with the embedded chain \tilde{X} . Figure 3 shows the trajectory of evolution of the contaminant in a human body along time.

3 A preview of contributions and future perspectives

As indicated in Section 1, this thesis focuses on developments in the field of bootstrap and statistical learning when the data are Markovian. We start with uniform bootstrap central limit theorems for Harris recurrent Markov chains. Next, we generalize the wild and residual bootstrap procedures for autoregressive processes of order p to periodically autoregressive processes (PAR). PAR sequences can be written in a Markovian form and are an example of Harris recurrent Markov chains. The second part of the thesis focuses on applications of empirical processes theory to statistical learning. We derive exponential and polynomial type maximal inequalities in order to control the supremum distance between stationary distribution μ of the chain X and its empirical counterpart. Concentration inequalities of this type are crucial tool when proving the generalization bounds for statistical learning algorithms via empirical risk minimization approach (we encourage to look into [142], [144] and [38] for detailed treatment). In this thesis, we use this strategy in order to obtain generalization bounds for minimum volume set estimation problem.

Having given a very general outline of the thesis, we proceed with more detailed preview of contributions.

1. Uniform bootstrap central limit theorems for Harris recurrent Markov chains

The first contribution involves asymptotic results for Harris recurrent Markov chains and was published in [44], [45] and [25]. We extend the bootstrap central limit theorem for the mean established in [26]. We show that uniform bootstrap CLT holds over uniformly bounded classes of functions \mathcal{F} as well as in unbounded case when we only require second order moment conditions imposed on the envelope F of \mathcal{F} . In this framework we measure the complexity of \mathcal{F} by its *covering number* $N_p(\epsilon, Q, \mathcal{F})$ which is interpreted as the minimal number of balls with radius ϵ needed to cover \mathcal{F} in the norm $L_p(Q)$ and Q is a measure on E with finite support. In what follows, we impose the finiteness of the *uniform entropy integral* of \mathcal{F} , namely

$$\int_0^\infty \sqrt{\log N_2(\epsilon, \mathcal{F})} d\epsilon < \infty, \quad \text{where } N_2(\epsilon, \mathcal{F}) = \sup_Q N_2(\epsilon, Q, \mathcal{F}).$$

More specifically, we show that under some technical conditions (specified in [44] and the second chapter of this thesis) imposed on X , f , class \mathcal{F} and small set S and under the assumption

$$\int_0^\infty \sqrt{\log N_2(\epsilon, \mathcal{F})} d\epsilon < \infty$$

we obtain that

$$\mathbb{Z}_n^* = n_{A_M}^{*1/2} \left[\frac{1}{n_{A_M}^*} \sum_{i=1}^{l_n^*-1} f(B_i^*) - \frac{1}{\hat{n}_{A_M}} \sum_{i=1}^{\hat{l}_n-1} f(\hat{B}_i) \right] \quad (1.4)$$

converges in probability under \mathbb{P}_ν to a Gaussian process G indexed by \mathcal{F} whose sample paths are bounded and uniformly continuous with respect to the metric $L_2(\mu)$. To understand Equation 1.4 we briefly mention that $n_{A_M}^*$ and l_n^* are bootstrap equivalents of the quantities n_{A_M} and l_n introduced in Subsection 2 and B_i^* are bootstrap blocks that is a sequence of i.i.d. blocks drawn with replacement from empirical distribution function based on the regenerative blocks or the approximate ones (see page 19, Algorithm 2).

Our theorems are bootstrap versions of uniform central limit theorems for Harris chains for bounded classes of functions presented in [98] and in unbounded case (see [137]). Moreover, our proof techniques allow to apply our reasoning into regenerative case that enables simplification of the proof of the uniform bootstrap central limit theorem for regenerative Markov chains established in [122]. Uniform central limit theorems (and their bootstrap versions) are an useful tool in non-parametric maximum likelihood estimation, kernel density estimation or wavelet density estimation (refer for instance to [110], [68] and [69] for details).

We are using the aforementioned results to establish bootstrap uniform central limit theorems for Fréchet differentiable functionals of Harris Markov chains. Under some technical conditions (uniform entropy condition, assumptions on a small set S , etc.) we have for a Fréchet differentiable functional at μ (with respect to a metric indexed by a class of functions, see [17] and Chapter 2 for details) that the bootstrap is asymptotically valid. We have chosen to work with Fréchet differentiable functionals since it guarantees the existence of influence function useful when detecting the outliers in the data (see [140] for details).

2. Robust estimation for Markov chains with applications to PDMPs In our next contribution we propose a method to construct robust estimators for atomic regenerative and Harris recurrent Markov chains with a special focus on Piecewise-Deterministic Markov Processes (PDMPs). In this framework we rely on a renewal theory for Markov chains and further developments of the approximate regenerative block bootstrap method. The main idea is to eliminate blocks having either too much contribution to the statistics of interest, or having a too large length.

It is known (see [29]) that some classical concepts of robust statistics can be naturally extended to a Markovian framework. For instance, one can define an influence function on the torus $\mathbb{T} = \cup_{k=1}^{\infty} E^k$. Indeed, let $\mathcal{P}_{\mathbb{T}}$ denote the set of all probability measures on the torus \mathbb{T} and for any $b \in \mathbb{T}$, set

$$L(b) = k \text{ if } b \in E^k, k \geq 1.$$

Then the influence function on the torus can be defined as follows.

Definition . (Influence function on the torus) Let $(\mathcal{V}, \|\cdot\|)$ be a separable Banach space. Let $T : \mathcal{P}_{\mathbb{T}} \rightarrow \mathcal{V}$ be a functional on $\mathcal{P}_{\mathbb{T}}$. If, for some \mathcal{L} in $\mathcal{P}_{\mathbb{T}}$,

$$t^{-1}(T((1-t)\mathcal{L} + t\delta_b) - T(\mathcal{L}))$$

has a finite limit as $t \rightarrow 0$ for any $b \in \mathbb{T}$, the influence function $T^{(1)} : \mathcal{P}_{\mathbb{T}} \rightarrow \mathcal{V}$ of the functional T at \mathcal{L} is then said to be well-defined, and, by definition, one set for all b in \mathbb{T} ,

$$T^{(1)}(b, \mathcal{L}) = \lim_{t \rightarrow 0} \frac{T((1-t)\mathcal{L} + t\delta_b) - T(\mathcal{L})}{t}. \quad (1.5)$$

Now it is straightforward to define Fréchet differentiability of functionals of Markov chains.

Definition . (Fréchet differentiability of functionals of Markov chains) The functional $T : \mathcal{P}_{\mathbb{T}} \rightarrow \mathbb{R}$ is Fréchet differentiable at $\mathcal{L}_A \in \mathcal{P}_{\mathbb{T}}$ for a metric d , if there exists a continuous linear operator $DT_{\mathcal{L}_A}$ (from the set of signed measures of the form $\mathcal{L} - \mathcal{L}_A$ in $(\mathbb{R}, \|\cdot\|)$) and a function

$$\epsilon^{(1)}(\cdot, \mathcal{L}_A) : \mathbb{R} \rightarrow (\mathbb{R}, \|\cdot\|),$$

which is continuous at 0 and $\epsilon^{(1)}(0, \mathcal{L}_A) = 0$ such that

$$\forall \mathcal{L} \in \mathcal{P}_{\mathbb{T}}, \quad T(\mathcal{L}) - T(\mathcal{L}_A) = DT_{\mathcal{L}_A}(\mathcal{L} - \mathcal{L}_A) + R^{(1)}(\mathcal{L}, \mathcal{L}_A),$$

where

$$R^{(1)}(\mathcal{L}, \mathcal{L}_A) = d(\mathcal{L}_A, \mathcal{L})\epsilon^{(1)}(d(\mathcal{L}_A, \mathcal{L}), \mathcal{L}_A).$$

Furthermore, assume that T admits the following representation,

$$\forall \mathcal{L}_A \in \mathcal{P}_{\mathbb{T}}, \quad DT_{\mathcal{L}_A}(\mathcal{L} - \mathcal{L}_A) = \int T^{(1)}(b, \mathcal{L}_A)\mathcal{L}(db),$$

where $T^{(1)}(b, \mathcal{L}_A)$ is the influence function at \mathcal{L}_A .

Fréchet differentiability is a standard tool for obtaining central limit theorems for plug-in estimators. Indeed, suppose that $T : \mathcal{P}_{\mathbb{T}} \rightarrow \mathbb{R}$ is a Fréchet differentiable functional at \mathcal{L}_A for some metric $d_{\mathcal{F}}$ (see Chapter 3 for details), where \mathcal{F} is a permissible class of functions with an envelope F , satisfying the uniform entropy condition

$$\int_0^\infty \sqrt{\sup_Q \log N_2(\epsilon, Q, \mathcal{F})} d\epsilon < \infty.$$

Assume in addition, in the regenerative atomic case

$$\mathbb{E}_A \left[\sum_{1 \leq j \leq \tau_A} F(X_j) \right]^2 < \infty, \quad \mathbb{E}_\nu(\tau_A) < \infty, \quad \mathbb{E}_A(\tau_A^2) < \infty.$$

Then,

$$n^{1/2}(T(\mathcal{L}_n) - T(\mathcal{L}_A)) \rightarrow \mathcal{N} \left(0, \frac{\text{Var}(T^{(1)}(B_i, \mathcal{L}_A))}{E_A(\tau_A)} \right).$$

We establish similar results for Harris recurrent Markov chains.

In this framework, we consider in particular the construction of robust estimators for the embedded Markov chains associated to the PDMP (which is a process whose

behaviour is determined by random jumps at points in time and its evolution is deterministically ruled by an ordinary differential equation between those times).

We consider robust estimators of several risk indicators such as the ruin probability, the expected shortfall and the extremal index of two PDMPs: the Cramér-Lundberg with a dividend barrier and the Kinetic Dynamic Exposure Model (KDEM) used in modeling pharmacokinetics of contaminants (see [31] for instance).

3. Wild and residual bootstrap methods for periodically autoregressive processes

This contribution is published in [46]. We consider periodically autoregressive process (PAR) of the form

$$X_{nT+v} = \sum_{k=1}^p \phi_k(v) X_{nT+v-k} + \epsilon_{nT+v}, \quad (1.6)$$

where

$$\Phi' = [\phi_1(1), \phi_2(1), \dots, \phi_p(1), \phi_1(2), \dots, \phi_p(2), \dots, \phi_1(T), \dots, \phi_p(T)]$$

designates the vector of parameters and $'$ is a transpose. The $\{X_{nT+v}\}$ denotes the series during the n -th cycle ($0 \leq n \leq N-1$) during v -th season ($1 \leq v \leq T$). The $\{\epsilon_{nT+v}\}$ is the mean zero white noise with variance of the form $\text{Var}(\epsilon_{nT+v}) = \sigma_v^2 > 0$ for all seasons v . The process in (1.6) can be written in a Markovian form using analogous vectorization trick as in Example 2 since PAR process may be written as T -variate autoregressive model (AR), refer to [20] for details. We obtained the least squares estimators of model's parameters

$$\hat{\Phi}' = [\hat{\phi}_1(1), \hat{\phi}_2(1), \dots, \hat{\phi}_p(1), \hat{\phi}_1(2), \dots, \hat{\phi}_p(2), \dots, \hat{\phi}_1(T), \dots, \hat{\phi}_p(T)]$$

in order to generate their bootstrap equivalents.

We propose wild and residual bootstrap procedures which are data driven (since they do not require a choice of the length of bootstrap blocks), thus can be attractive for practical use. The residual bootstrap procedure for PAR processes consists of the following steps.

Residual bootstrap method for PAR processes

Step 1 Compute the ordinary least squares estimator $\hat{\Phi}$ of Φ .

Step 2 Compute the residuals of the estimated model

$$\hat{\epsilon}_{nT+v} = X_{nT+v} - \sum_{k=1}^p \hat{\phi}_k(v) X_{nT+v-k},$$

where $1 \leq v \leq T$, $0 \leq n \leq N-1$.

Step 3 Compute the centred residuals

$$\bar{\eta}_{nT+v} = \frac{\hat{\epsilon}_{nT+v}}{\sigma_v} - \frac{1}{NT} \sum_{n=0}^{N-1} \sum_{v=1}^T \frac{\hat{\epsilon}_{nT+v}}{\sigma_v},$$

where NT is the number of all observations in the model.

Step 4 Generate bootstrap variables η_{nT+v}^* by drawing randomly with replacement from

$$\{\bar{\eta}_1, \dots, \bar{\eta}_{NT}\}.$$

Step 5 Generate the bootstrap version of the model (1.6)

$$X_{nT+v}^* = \sum_{k=1}^p \hat{\phi}_k(v) X_{nT+v-k} + \sigma_v \eta_{nT+v}^*, \quad 1 \leq v \leq T.$$

Step 6 Calculate the bootstrap estimators of parameters for each season v , $1 \leq v \leq T$

$\hat{\Phi}^*(v)$, where

$$z^*(v) = [X_v^*, \dots, X_{(N-1)T+v}^*]', \quad 1 \leq v \leq T.$$

The second method we propose is wild bootstrap procedure for PAR processes.

Wild bootstrap method for PAR processes

Step 1 Compute the ordinary least squares estimator $\hat{\Phi}$ of Φ .

Step 2 Compute the residuals of the estimated model

$$\hat{\epsilon}_{nT+v} = X_{nT+v} - \sum_{k=1}^p \hat{\phi}_k(v) X_{nT+v-k},$$

where $1 \leq v \leq T$, $0 \leq n \leq N-1$.

Step 3 Generate the bootstrap process X_{nT+v}^\dagger for each season v , $1 \leq v \leq T$

$$X_{nT+v}^\dagger = \sum_{k=1}^p \hat{\phi}_k(v) X_{nT+v-k} + \epsilon_{nT+v}^\dagger$$

and

$$\epsilon_{nT+v}^\dagger = \hat{\epsilon}_{nT+v} \eta_{nT+v}^\dagger,$$

where $\eta_{nT+v}^\dagger \sim \mathcal{N}(0, 1)$ and $(\eta_{nT+v}^\dagger)_{nT+v \in \mathbb{R}}$ is independent of $\hat{\epsilon}_{nT+v}$.

Step 4 Calculate the bootstrap estimator of parameters, namely $\hat{\Phi}^\dagger(v)$ for each season v , $1 \leq v \leq T$, where

$$z^\dagger(v) = [X_v^\dagger, \dots, X_{(N-1)T+v}^\dagger]'$$

Next, we proved weak consistency for both methods. More specifically, we showed for a causal periodic autoregressive series X_{nT+v} defined in (1.6) with finite fourth moment that, the residual bootstrap procedure we proposed is weakly consistent, i.e.

$$\sqrt{N} \left(\widehat{\Phi}^* - \widehat{\Phi} \right) \xrightarrow{P^*} \mathcal{N} \left(0, R^{-1} \right),$$

where

$$\widehat{\Phi}^{*'} = \left[\widehat{\phi}_1^*(1), \widehat{\phi}_2^*(1), \dots, \widehat{\phi}_p^*(1), \widehat{\phi}_1^*(2), \dots, \widehat{\phi}_p^*(2), \dots, \widehat{\phi}_1^*(T), \dots, \widehat{\phi}_p^*(T) \right]$$

is a vector of bootstrap estimators of parameters for each season v obtained from residual bootstrap algorithm and R is specified in Chapter 4. Similar consistency results are obtained for wild bootstrap procedure (see Chapter 4). Finally, we illustrate our theoretical considerations by simulations.

4. Exponential and polynomial type maximal inequalities for Harris Markov chains

Motivated by applications in statistical learning we establish bounds for the tail probability for suprema of empirical processes in a Markovian framework. These contributions are presented in [22], [23] and [24].

Let $f : E \rightarrow \mathbb{R}$ be a measurable function. Since our inequalities are maximal type, we control the class of functions via uniform entropy number. Under exponential moments imposed on τ_A and $f(B_j)$ we establish Bernstein and Hoeffding type maximal inequalities as a function of uniform entropy number and moments of time returns and blocks. One of main difficulties when deriving such bounds is that even if we assume that f is bounded, $f(B_j)$ may be unbounded over a whole block of observations. In order to derive the inequalities we rely heavily on Montgomery-Smith's inequality from [108] and symmetrization techniques from [119].

We also show that under weaker conditions imposed on time returns and $f(B_j)$'s, the polynomial bounds can be established. Interestingly, the conditions imposed on a Markov chain X are satisfied by sub-geometrically ergodic Markov chains, to which our polynomial tail inequality can be applied.

Furthermore, we establish bounds for the expectation of the supremum of empirical process in a Markovian setting since they occur to be particularly useful when one wants to select a model via some penalization criterion with penalty term depending on a complexity of the whole collection of models.

We present the aforementioned results in a detailed form in the subsequent sections. However, to give a reader a general overview of obtained results we provide the bounds

in a general (and somewhat simplified) form below. The detailed conditions imposed on chain X are omitted here and stated in further sections. For the sake of simplicity we provide the results solely in the atomic regenerative case (we formulate the inequalities for Harris recurrent Markov chains in further sections). Let

$$\sigma_m^2 = \max_{f \in \mathcal{F}} \sigma^2(f) > \eta > 0.$$

- **Bernstein type maximal inequality** Assume that $N_1(\epsilon, \mathcal{F}) < \infty$. Then, under exponential block moment conditions and exponential moments of return times to set A , we have for any $x > 0$, $0 < \epsilon < x/2$ and for all $n \geq 1$

$$\mathbb{P}_\nu \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(X_i) - \mu(f) \right| \geq x \right] \leq N_1(\epsilon, \mathcal{F}) K_1 \left\{ \exp \left[\frac{-n(x - 2\epsilon)^2}{K_2(\sigma_m^2 + K_3(x - 2\epsilon))} \right] \right\},$$

where K_1 , K_2 and K_3 are positive parameters specified in Chapter 5.

- **Hoeffding type maximal inequality** Assume that $N_1(\epsilon, \mathcal{F}) < \infty$. Suppose further that the class of functions \mathcal{F} is uniformly bounded. Then, under exponential block moment conditions and exponential moments of return times to set A , we have for any $x > 0$, $0 < \epsilon < x/2$ and for all $n \geq 1$

$$\mathbb{P}_\nu \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(X_i) - \mu(f) \right| \geq x \right] \leq N_1(\epsilon, \mathcal{F}) L_1 \left\{ \exp \left[-\frac{n(x - 2\epsilon)^2}{L_2 D^2} \right] \right\},$$

where D is a constant such that $\forall f \in \mathcal{F} |f| < D$ and L_1 and L_2 are positive parameters specified later in Chapter 5.

- **Polynomial tail maximal inequality** Assume that $N_1(\epsilon, \mathcal{F}) < \infty$. Suppose further that the p -th block moment and p -th moment of return times to the atom A are finite. Then, we have for any $x > 0$, $0 < \epsilon < x/2$ and for all $n \geq 1$

$$\mathbb{P}_\nu \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(X_i) - \mu(f) \right| \geq x \right] \leq C_1 \frac{N_1(\epsilon, \mathcal{F})}{(x - 2\epsilon)^p n^{p/2}}$$

and C_1 is a positive parameter specified in Chapter 5.

- **Bound for expectation of supremum of empirical processes** Assume that

$$\mathbb{E}_A[l(B_1)]^2 < \infty \text{ and } \mathbb{E}_A[F(B_1)]^2 < \infty,$$

where F is an envelope for \mathcal{F} . Moreover, suppose that uniform entropy number $N_1\left(\frac{\epsilon}{R_1}, \mathcal{F}\right) < \infty$. Then, for any $\epsilon > 0$ we have

$$\begin{aligned} & \mathbb{E}_A \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^{l_n} (f(B_i) - \mu(f(B_1))) \right| \right] \\ & \leq R_2 \left[\epsilon + \mathcal{N}\left(\frac{\epsilon}{R_1}, \mathcal{F}\right) \times \mathbb{E}_A[F(B_1)^2]^{1/2} \sqrt{\frac{2 \log N_1\left(\frac{\epsilon}{R_1}, \mathcal{F}\right)}{n}} \right], \end{aligned}$$

where R_1 and R_2 are positive constants that can be explicitly computed.

The above results may be easily generalized to a Harris recurrent case.

5. Generalization bounds for minimum volume set estimation problem The last contributions presented in this thesis are generalization bounds for minimum volume set (MV-set) for regenerative and Harris recurrent Markov chains. The results have been presented in [23] and [24]. The MV-set estimation problem was firstly proposed in [120] in the i.i.d. setting.

Let μ be a probability distribution on a measurable space (E, \mathcal{E}) . Let $\alpha \in (0, 1)$ and λ be a σ -finite measure of reference on (E, \mathcal{E}) , any solution of the minimization problem (1.7)

$$\min_{\Omega \in \mathcal{E}} \lambda(\Omega) \text{ subject to } \mu(\Omega) \geq \alpha \quad (1.7)$$

is called a *MV-set* of level α . The distribution μ is assumed to be absolutely continuous w.r.t. λ and denote by

$$f(x) = (d\mu/d\lambda)(x)$$

the related density.

Under some technical assumptions on f , for any $\alpha \in (0, 1)$, [120] showed that the set

$$\Omega_\alpha^* = \{x \in E : f(x) > \gamma_\alpha\},$$

where γ_α is the unique number such that

$$\int_{f(x) > \gamma_\alpha} f(x) d\lambda(x) = \alpha$$

is the unique solution of the MV-set estimation problem (1.7). Minimum volume sets can be interpreted as follows, for small values of the mass level α , *MV-sets* enable to

recover the modes of the distribution, while their complementary sets correspond to *rare observations* when α is large.

In practice, distribution μ is unknown and is replaced by its empirical counterpart μ_n . Then finding a minimum volume set of level α boils down to solving the following minimization problem

$$\min_{\Omega \in \mathcal{E}} \lambda(\Omega) \text{ subject to } \hat{\mu}_n(\Omega) \geq \alpha - \psi_n \quad (1.8)$$

with ψ_n being a tolerance parameter (see Chapter 6 for more details).

The minimum volume set estimation technique can be used as an unsupervised anomaly detection algorithm since when dealing with unlabelled data we consider anomaly as a rare event.

Scott and Nowak [126] established generalization bounds for MV-set estimation problem in the i.i.d. setting. In order to establish generalization bounds in a Markovian setting we retract to the framework of empirical risk minimization and thus, heavily rely on concentration inequalities which allows us to control the suprema of empirical processes involved. Our approach boils down to decomposition of empirical distribution function of interest into: $\forall \Omega \in \mathcal{E}$,

$$\hat{\mu}_n(\Omega) = \frac{1}{n} \sum_{i=1}^{\tau_A} \mathbb{I}\{X_i \in \Omega\} + \frac{l_n - 1}{n} \left(\frac{1}{l_n - 1} \sum_{j=1}^{l_n - 1} S_j(\Omega) \right) + \frac{1}{n} \sum_{i=1 + \tau_A(l_n)}^n \mathbb{I}\{X_i \in \Omega\}, \quad (1.9)$$

where $l_n = \sum_{i=1}^n \mathbb{I}\{X_i \in A\}$ denotes the number of visits to set A (*regenerations*), the occupation time of the set Ω between the j -th and $(j + 1)$ -th regeneration times is denoted by $S_j(\Omega) = \sum_{\tau_A(j) < i \leq \tau_A(j+1)} \mathbb{I}\{X_i \in \Omega\}$. Decomposition (1.9) along with application of polynomial tail maximal inequality for Markov chains allows us to extend the result in [126] to the atomic regenerative and Harris recurrent case.

Let $p \geq 2$. In order to establish generalization bounds for minimum volume set estimation problem we assume that

$$\mathbb{E}_A[\tau_A^p] < \infty \text{ and } \mathbb{E}_\nu[\tau_A^p] < \infty.$$

Let $r \geq 1$. The collection of indicator functions on E ,

$$\mathcal{F} = \{\mathbb{I}\{x \in \Omega\} : \Omega \in \mathcal{G}\}$$

is a uniform Donsker class (relative to L_1) with polynomial uniform covering numbers, *i.e.* there exists a constant $c > 0$ s.t. $\forall \zeta > 0$,

$$N_1(\zeta, \mathcal{F}) \stackrel{def}{=} \sup_Q N(\zeta, \mathcal{F}, L_1(Q)) \leq c(1/\zeta)^r,$$

where the supremum is taken over the set of finitely discrete probability measures on (E, \mathcal{E}) .

Under the preceding assumptions imposed on moments of return of X to the atom A and on complexity of class \mathcal{F} and for all $\delta \in (0, 1)$, and if ψ_n (we specify ψ_n in Chapter 6) is a well-chosen penalty for class \mathcal{G} , then, with probability at least $1 - \delta$,

$$\lambda(\widehat{\Omega}_n) \leq \lambda(\Omega_\alpha^*) + \left\{ \inf_{\Omega \in \mathcal{G}: \mu(\Omega) \geq \alpha} \lambda(\Omega) - \lambda(\Omega_\alpha^*) \right\}$$

and

$$\mu(\widehat{\Omega}_n) \geq \alpha - 2\psi_n(\delta).$$

The organization of this thesis corresponds to the order of contributions listed above. In Chapter 2 we recall non-parametric bootstrap procedures for regenerative and Harris recurrent Markov chains established in [26] and next proceed to the exposition of our uniform bootstrap central limit theorems for Harris recurrent Markov chains. In the last part of the chapter, we use the aforementioned results to obtain bootstrap uniform central limit theorems for Fréchet differentiable functionals in a Markovian setting. In Chapter 3 some tools to detect outliers in a Harris Markovian framework are presented. This chapter focus on applications to piecewise-deterministic Markov processes. Chapter 4 presents wild and residual bootstrap procedures for PAR processes that we formulated in [46]. Next, we show that both methods are weakly consistent and illustrate our theoretical considerations by simulations. In Chapter 5 we present Bernstein and Hoeffding type maximal deviation inequalities for Markov chains. Next, we discuss how the conditions imposed on the chain can be relaxed in order to obtain polynomial tail maximal bound. We also show how to bound the expectation of the supremum of an empirical process in a Markovian framework. In Chapter 6 we show how to use the bounds from Chapter 5 in order to obtain generalization bounds for statistical learning algorithms. We illustrate this approach with minimum volume set estimation procedure. Finally, Chapter 7 is aimed to shortly summarize the state of the art as well as highlight main scientific contributions of this thesis. This chapter is written only in French.

Chapter 2

Bootstrap uniform central limit theorems for Harris recurrent Markov chains

The main objective of this chapter is to establish bootstrap uniform functional central limit theorem for Harris recurrent Markov chains over uniformly bounded classes of functions. We show that the result can be generalized also to the unbounded case. To avoid some complicated mixing conditions, we make use of the well-known regeneration properties of Markov chains. We show that in the atomic case the proof of the bootstrap uniform central limit theorem for Markov chains for functions dominated by a function in L_2 space proposed by Radulović [122] can be significantly simplified. Finally, we prove bootstrap uniform central limit theorems for Fréchet differentiable functionals in a Markovian setting.

This chapter gathers results published in [44] and [45]. Some additional remarks and explanations (in addition to the material included in [44]) are added in order to make this exposition more comprehensive.

1 Preliminaries

The bootstrap method was introduced by Efron in [59] for i.i.d. data. The intuition behind this procedure is rather simple. Let X_1, \dots, X_n be a sequence of i.i.d. random variables. We are interested in studying the behaviour of some statistic $H_n(X_1, \dots, X_n)$. In particular, one may want to compute from H_n a confidence interval for some unknown population parameter. By \xrightarrow{w} we denote weak convergence and by $\mathcal{L}(H)$ limiting distribution of H . Usually, in this

situation, weak central limit theorems are handy, i.e. as $n \rightarrow \infty$

$$H_n(X_1, \dots, X_n) \xrightarrow{W} H$$

which implies that

$$\mathcal{L}(H_n(X_1, \dots, X_n)) \approx \mathcal{L}(H). \quad (2.1)$$

Weak central theorems often offer a closed form of the distribution $\mathcal{L}(H)$ which makes the approximate confidence interval computable.

As indicated in [54] this approach struggles with some challenges. Firstly, the closed form of $\mathcal{L}(H)$ may be very difficult to obtain. Secondly, one can construct statistics H_n for which the limiting process does not exist. Lastly, it often occurs that $\mathcal{L}(H)$ depends on unknown parameters, thus confidence regions can not be computed (see much broader discussion in [121]).

It appears that bootstrap procedures may be handy in such situations. Indeed, we build the empirical distribution function from the data $\mathbb{P}_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$. Next we draw bootstrap observations

$$X_1^*, X_2^*, \dots, X_n^* \text{ from } \mathbb{P}_n.$$

Such bootstrap observations are i.i.d. conditionally on X_1, \dots, X_n . Next, we conclude that $\mathbb{P} \approx \mathbb{P}_n$ essentially means that $\mathcal{L}(\mathbb{P}_n) \approx \mathcal{L}(\mathbb{P})$. Thus, one can hope that

$$\mathcal{L}(H_n(X_1, \dots, X_n)) \approx \mathcal{L}(H_n^*(X_1^*, \dots, X_n^*)). \quad (2.2)$$

Now the crucial observation is that $\mathcal{L}(H_n^*(X_1^*, \dots, X_n^*))$ does not depend on any unknown parameters and importantly, can be computed. Let d be some metric that metrizes weak convergence (see [141] for more details). The bootstrap procedure is valid when we can show that

$$d(\mathcal{L}(H_n), \mathcal{L}(H_n^*)) \rightarrow 0 \text{ a.s. or in probability.}$$

It is well-known that the bootstrap methods work in many cases when the limiting process is Gaussian. Here the theory of empirical processes kicks in. More specifically, the uniform central limit theorems justify (2.2) for many statistics H_n (refer also to [140] and [141]). Let

$$\mathbb{P}_n f = \frac{1}{n} \sum_{i=1}^n f(X_i) \text{ and } \mathbb{P} f = \int f d\mathbb{P}.$$

Showing that (under regularity assumptions),

$$\sqrt{n}(\mathbb{P}_n f - \mathbb{P} f)_{f \in \mathcal{F}} \xrightarrow{W} G_p(f) \quad (2.3)$$

(where G_p is a Gaussian process) enables to prove weak convergence $\mathcal{L}(H_n) \rightarrow \mathcal{L}(G)$ for many statistics H_n of interest. Moreover it is shown in [71] that essentially (2.3) holds if and only if

$$\sqrt{n}(\mathbb{P}_n^* f - \mathbb{P}_n f)_{f \in \mathcal{F}} \xrightarrow{W} G_p(f)_{f \in \mathcal{F}} \text{ in probability.}$$

and \mathbb{P}^* is conditional distribution on the data.

The above reasoning motivated us to study uniform central limit theorems and their bootstrap versions in a regenerative atomic and Harris recurrent case. The uniform central limit theorems have also many applications in modern statistics, an interested reader is referred to the series of papers [110], [68] and [69]. The preceding summary is based on a survey written by Radulović [121]. We also direct to the aforementioned work for more details concerning developments of bootstrap in dependent case.

Interestingly, the regenerative bootstrap theory for Markov chains has received relatively limited attention given very wide number of results concerning various bootstrap methods for both i.i.d. and dependent data. We will give a brief overview of how the regenerative bootstrap methods developed through last 26 years.

It is noteworthy that one of the first bootstrap results for Markov chains were obtained in the series of papers [10] and [11]. The proposed methods rely on the renewal properties of Markov chains when a (recurrent) state is visited infinitely often. The idea behind such procedures is to resample a deterministic number of data blocks which are corresponding to regeneration cycles. More specifically, it was established in [10] and [11] that the distribution of naive bootstrap of the pivot $\sqrt{n}(\hat{\Pi}_n - \Pi)$, where $\hat{\Pi}_n$ is the maximum likelihood estimator of a transition probability matrix Π , approximates that of the pivot as $n \rightarrow \infty$. The approach is based on a consideration of a double array of Markov chains for which a central limit theorem is established.

Datta and Mc Cormick [50] studied bootstrapping the distribution of the sample mean of a fixed real function of a Markov chain. Bertail and Cléménçon [26] proposed two effective methods for bootstrapping Markov chains: Regenerative block bootstrap (RBB) method for atomic chains and Approximate block bootstrap method (ARBB) for general Harris recurrent Markov chains. The main intuition behind these procedures is to mimic the renewal (pseudo-renewal in general Harris case) structure of the chain by drawing regeneration data blocks, until the length of the reconstructed bootstrap sample is larger than the length of the original data. We recall these procedures in Section 2. In the regenerative atomic setting, the RBB method has the uniform rate of convergence of order $O_{\mathbb{P}}(n^{-1})$ which is the optimal rate of convergence in the i.i.d. case. Moreover, it is proven in [26] that the second-order correctness of the ARBB procedure in the unstudentized stationary case,

the rate of convergence is close to that in the i.i.d. setting.

One of main advantages of aforementioned methods is the fact that the division of the data into blocks is completely data-driven in atomic regenerative case and the ARBB procedure relies only on the parameters of minorization condition (1.1), which is a significant advantage in comparison to block bootstrap methods. It is noteworthy that in parallel to [26], the Markov chains bootstrap CLT for the mean under no additional assumptions was proposed in [122].

Bootstrap results for Markov chains in [122] and [26] showed how natural tool the regenerative technique is in order to extend the bootstrap theory to empirical processes indexed by classes of functions in a Markovian setting.

To the best of our knowledge, Radulović was the first who proved the bootstrap uniform central limit theorem over uniformly bounded classes of functions \mathcal{F} in [122]. In this work, Radulović studies countable regenerative Markov chains and indicates that with an additional uniform entropy condition the bootstrap result can be extended to the uncountable case. Gorst-Rasmussen and Bøgsted [75] have proved bootstrap uniform central limit theorem over classes of functions with an envelope in L^2 . They have considered regenerative case which was motivated by their study of queuing systems with abandonment.

We generalized in [44] the bootstrap result for empirical processes for Markov chains obtained in [122]. We established the bootstrap uniform functional central limit theorem over a permissible uniformly bounded classes of functions in general Harris case. We also show that by arguments from [137], the condition of uniform boundedness of \mathcal{F} can be weakened and it is sufficient to require only that \mathcal{F} has an envelope F in L^2 . Some applications to Fréchet differentiable functionals are presented in Section 3

2 Non-parametric bootstrap for regenerative and Harris recurrent Markov chains

In this subsection we recall non-parametric bootstrap procedures for regenerative and Harris recurrent Markov chains. The algorithms were proposed in [26].

Remark 5. In order to avoid large bias of the estimators based on the regenerative blocks we discard the data before the first and after the last pseudo-regeneration times (for more details refer to [26], page 693).

We will start with Regenerative block bootstrap (RBB) procedure when the atom A in known *a priori*. The algorithm allows to compute the estimate of the sample distribution

of some statistic $T_n = T(B_1, \dots, B_{l_n-1})$ with standarization $S_n = S(B_1, \dots, B_{l_n-1})$. The procedure is formalized in Algorithm 3.

Algorithm 3 Regenerative block bootstrap method

Step 1 Draw sequentially bootstrap data blocks B_1^*, \dots, B_k^* (we denote the length of the blocks by $l(B_j^*)$, $j = 1, \dots, k$) independently from the empirical distribution function

$$\mathcal{L}_n = \frac{1}{l_n - 1} \sum_{i=1}^{l_n-1} \delta_{B_i},$$

where B_i , $i = 1, \dots, l_n - 1$ are regeneration blocks. We generate bootstrap blocks until the joint length of bootstrap blocks $l^*(k) = \sum_{i=1}^k l(B_i^*)$ exceeds n . We set

$$l_n^* = \inf\{k : l^*(k) > n\}.$$

Step 2 Bind bootstrap blocks from Step 1 and construct the RBB bootstrap sample $X^{*(n)} = (X_1^*, \dots, X_{l_n^*-1}^*)$.

Step 3 Compute the RBB statistic and its RBB distribution, namely

$$T_n^* = T(X^{*(n)}) = T(B_1^*, \dots, B_{l_n^*-1}^*)$$

and its standarization

$$S_n^* = S(X^{*(n)}) = S(B_1^*, \dots, B_{l_n^*-1}^*).$$

Step 4 The RBB distribution is given by

$$H_{ARBB}(x) = \mathbb{P}^*(S_n^{*-1}(T_n^* - T_n) \leq x),$$

where \mathbb{P}^* is conditional probability given the data.

Taking a closer look at the steps of the RBB procedure, one can see that it is a very natural generalization of non-parametric bootstrap method in the i.i.d. setting. The crucial difference is that the empirical distribution function is built from blocks of data instead of single observations (see Step 1 of Algorithm 3). The validity of RBB method has been proved in [26] for the mean.

In this chapter we will state the uniform bootstrap central limit theorems for Harris recurrent Markov chains. Thus, in this framework, we are more interested in non-parametric bootstrap procedure formulated in a Harris recurrent case. Similarly as in RBB method, the Approximate regenerative block bootstrap (ARBB) aims to compute the estimate of the sample distribution of some statistic $T_n = T(\hat{B}_1, \dots, \hat{B}_{\hat{l}_n-1})$ with standarization

$S_n = S(\widehat{B}_1, \dots, \widehat{B}_{\widehat{l}_n-1})$. Note, that since we deal with Harris recurrent Markov chains, in order to split the data into segments, in practice we need to apply the approximate Nummelin splitting technique. Thus, in consequence, we deal with approximated blocks $\widehat{B}_1, \dots, \widehat{B}_{\widehat{l}_n-1}$. For the completeness of exposition, we recall the ARBB bootstrap procedure below. The algorithm proceeds as follows.

Algorithm 4 Approximate block bootstrap

Step 1 Draw sequentially bootstrap data blocks B_1^*, \dots, B_k^* (we denote the length of the blocks by $l(B_j^*)$, $j = 1, \dots, k$) independently from the empirical distribution function

$$\widehat{\mathcal{L}}_n = \frac{1}{\widehat{l}_n - 1} \sum_{i=1}^{\widehat{l}_n-1} \delta_{\widehat{B}_i},$$

where \widehat{B}_i , $i = 1, \dots, \widehat{l}_n - 1$ are initial pseudo-regeneration blocks. We generate bootstrap blocks until the joint length of bootstrap blocks $l^*(k) = \sum_{i=1}^k l(B_i^*)$ exceeds n . We set

$$l_n^* = \inf\{k : l^*(k) > n\}.$$

Step 2 Bind bootstrap blocks from Step 1 and construct the ARBB bootstrap sample

$$X^{*(n)} = (X_1^*, \dots, X_{l_n^*-1}^*).$$

Step 3 Compute the ARBB statistic and its ARBB distribution, namely

$$T_n^* = T(X^{*(n)}) = T(B_1^*, \dots, B_{l_n^*-1}^*)$$

and its standarization

$$S_n^* = S(X^{*(n)}) = S(B_1^*, \dots, B_{l_n^*-1}^*).$$

Step 4 The ARBB distribution is given by

$$H_{ARBB}(x) = \mathbb{P}^*(S_n^{*-1}(T_n^* - T_n) \leq x),$$

where \mathbb{P}^* is conditional probability given the data.

Observe, that analogously to the RBB method, the main innovation is that we construct the empirical distribution function from data blocks instead of single observations. Throughout this chapter, we denote by

$$n_{A_{\mathcal{M}}}^* = \sum_{i=1}^{l_n^*-1} l(B_i^*)$$

the length of the bootstrap sample,

$$\mu_n^*(f) = \frac{1}{n_{\mathcal{AM}}^*} \sum_{i=1}^{l_n^*-1} f(B_i^*) \quad \text{and} \quad \sigma_n^{*2}(f) = \frac{1}{n_{\mathcal{AM}}^*} \sum_{i=1}^{l_n^*-1} \{f(B_i^*) - \mu_n^*(f)\}^2$$

are bootstrap versions of empirical mean and variance.

In what follows, we state necessary conditions under which the ARBB procedure is consistent. The assumptions have been formulated in [26] (see page 700 therein). We list them below since our bootstrap uniform central limit theorems also require these assumptions to be satisfied.

Let (X_n) be a positive recurrent Harris Markov chain. We assume that $(\alpha_n)_{n \in \mathbb{N}}$ is a sequence of non-negative numbers that converges to zero. We impose the following conditions on the chain:

1. S is chosen so that $\inf_{x \in S} \phi(x) > 0$.
2. Transition density p is estimated by p_n at the rate α_n (when the chain is very smooth generally the best rate is of order $\alpha_n = \frac{\log(n)}{n}$) for the mean squared error (MSE) when error is measured by the L^∞ loss over S^2 (see [9] and [26] for details and additional references).

Moreover, we assume the following conditions (we advise to look into [26] for a comprehensive treatment on these assumptions). Let $k \geq 2$ be a real number.

$\mathcal{H}_1(f, k, \nu)$. The small set S is such that

$$\sup_{x \in S} \mathbb{E}_x \left[\left(\sum_{i=1}^{\tau_S} |f(X_i)| \right)^k \right] < \infty$$

and

$$\mathbb{E}_\nu \left[\left(\sum_{i=1}^{\tau_S} |f(X_i)| \right)^k \right] < \infty.$$

$\mathcal{H}_2(k, \nu)$. The set S is such that

$$\sup_{x \in S} \mathbb{E}_x(\tau_S^k) < \infty \quad \text{and} \quad \mathbb{E}_\nu(\tau_S^k) < \infty.$$

\mathcal{H}_3 . The density $p(x, y)$ is estimated by $p_n(x, y)$ at the rate α_n for the MSE when error is measured by the L^∞ loss over $S \times S$:

$$\mathbb{E}_\nu \left(\sup_{(x, y) \in S \times S} |p_n(x, y) - p(x, y)|^2 \right) = O(\alpha_n), \quad \text{as } n \rightarrow \infty.$$

\mathcal{H}_4 . The density ϕ is such that $\inf_{x \in S} \phi(x) > 0$.

\mathcal{H}_5 . The transition density $p(x, y)$ and its estimate $p_n(x, y)$ are bounded by a constant $R < \infty$ over S^2 .

In what follows, we recall two theorems from [26] that essentially guarantee the consistency of the ARBB procedure for pseudo-regeneration blocks.

Theorem 2. *Suppose that the conditions [1] and [2] are satisfied by the chain and $\mathcal{H}_1(f, \rho, \nu)$, $\mathcal{H}_2(\rho, \nu)$ with $\rho \geq 4$, $\mathcal{H}_3, \mathcal{H}_4$ and \mathcal{H}_5 hold. Then, as $n \rightarrow \infty$ we have*

$$\hat{\sigma}_n^2(f) \rightarrow \sigma_f^2 \quad \text{in } \mathbb{P}_\nu - \text{probability}$$

and

$$\hat{n}_{A\mathcal{M}}^{1/2} \frac{\hat{\mu}_n(f) - \mu(f)}{\hat{\sigma}_n(f)} \rightarrow \mathcal{N}(0, 1) \quad \text{in distribution under } \mathbb{P}_\nu.$$

Throughout the thesis, we denote by $BL_1(l^\infty(\mathcal{F}))$ the set of all 1-Lipschitz bounded functions on $l^\infty(\mathcal{F})$. We define the bounded Lipschitz metric on $l^\infty(\mathcal{F})$ as

$$d_{BL_1}(X, Y) = \sup_{b \in BL_1(l^\infty(\mathcal{F}))} |\mathbb{E}b(X) - \mathbb{E}b(Y)|; \quad X, Y \in l^\infty(\mathcal{F}).$$

It is noteworthy that d_{BL_1} metrizes weak convergence of empirical processes (refer to [141] for more details). Expectations of non-measurable elements are understood as outer expectations (see [141], Chapter 1) and weak convergence holds in sense of Hoffmann-Jørgensen.

Definition 6. We say that \mathbb{Z}_n^* is weakly consistent if $d_{BL_1}(\mathbb{Z}_n^*, \mathbb{Z}_n) \xrightarrow{P} 0$. Analogously, \mathbb{Z}_n^* is strongly consistent if $d_{BL_1}(\mathbb{Z}_n^*, \mathbb{Z}_n) \xrightarrow{\text{a.s.}} 0$.

Theorem 3. *Under the hypotheses of Theorem 2, we have the following convergence in probability under \mathbb{P}_ν :*

$$\Delta_n = \sup_{x \in \mathbb{R}} |H_{ARBB}(x) - H_\nu(x)| \rightarrow 0, \quad \text{as } n \rightarrow \infty,$$

where

$$H_\nu(x) = \mathbb{P}_\nu(x) \left(\hat{n}_{A\mathcal{M}}^{1/2} \sigma_f^{-1} (\hat{\mu}_n(f) - \mu(f)) \leq x \right)$$

and

$$H_{ARBB}(x) = \mathbb{P}^* \left(n_{A\mathcal{M}}^{*1/2} \hat{\sigma}_n^{-1}(f) (\mu_n^*(f) - \hat{\mu}_n(f)) \leq x | X^{(n+1)} \right).$$

In order to make this exposition clear, we briefly mention that the convergence $X_n \xrightarrow{P^*} X$ in \mathbb{P}_ν -probability (\mathbb{P}_ν -a.s.) along the sample is understood as

$$\mathbb{P}^*(|X_n - X| > \epsilon | X^{(n+1)}) \xrightarrow{n \rightarrow \infty} 0 \quad \text{in } \mathbb{P}_\nu - \text{probability} \quad (\mathbb{P}_\nu - \text{a.s.}).$$

3 Main asymptotic results

The aim of this section is to present uniform bootstrap CLT over permissible, uniformly bounded classes of functions \mathcal{F} . The results has been established in [44] and are one of the main contributions of this thesis.

The CLT's that we obtained hold uniformly over the class of functions \mathcal{F} . In this framework we measure the complexity of \mathcal{F} by its *covering number* $N_p(\epsilon, Q, \mathcal{F})$ which is interpreted as the minimal number of balls with radius ϵ needed to cover \mathcal{F} in the norm $L_p(Q)$ and Q is a measure on E with finite support. In what follows, we impose the finiteness of the *uniform entropy integral* of \mathcal{F} , namely

$$\int_0^\infty \sqrt{\log N_2(\epsilon, \mathcal{F})} d\epsilon < \infty, \quad \text{where } N_2(\epsilon, \mathcal{F}) = \sup_Q N_2(\epsilon, Q, \mathcal{F}).$$

We advise to refer to [141], [119] and [87] for more details regarding entropy conditions as well as for many examples how this assumption may be verified for a specific class \mathcal{F} .

Since main results of this section are bootstrap versions of uniform central limit theorems established by Levental [98] (see Theorem 5.9 therein), we recall it in its original form below.

Theorem 4. *Let (X_n) be a positive recurrent Harris chain taking values in (E, \mathcal{E}) . Let μ be the invariant probability measure for (X_n) . Assume further that \mathcal{F} is a uniformly bounded class of measurable functions on E and*

$$\int_0^\infty \sqrt{\log N_2(\epsilon, \mathcal{F})} d\epsilon < \infty.$$

If $\sup_{x \in A} \mathbb{E}_x(\tau_A)^{2+\gamma} < \infty$ ($\gamma > 0$ fixed), where A is atomic set for the chain, then the empirical process $Z_n(f) = n^{1/2}(\mu_n - \mu)(f)$ converges weakly as a random element of $l^\infty(\mathcal{F})$ to a Gaussian process G indexed by \mathcal{F} whose sample paths are bounded and uniformly continuous with respect to the metric $L_2(\mu)$.

We are aware of the fact that permissibility condition imposed on \mathcal{F} may be unfamiliar. This regularity condition is rather technical and thus its definition will be omitted here, nonetheless, the permissibility assumption allows to avoid measurability problems when dealing with suprema of empirical processes indexed by classes of functions. A rigorous definition can be found in Appendix C in [119].

In what follows, we present bootstrap uniform central limit theorem over permissible, uniformly bounded classes of functions which fulfill the uniform entropy condition.

Theorem 5. *Suppose that (X_n) is positive recurrent Harris Markov chain. Assume that the conditions [1] and [2] are satisfied by the chain and $\mathcal{H}_1(f, \rho, \nu)$, $\mathcal{H}_2(\rho, \nu)$ with $\rho \geq 4$, $\mathcal{H}_3, \mathcal{H}_4$*

and \mathcal{H}_5 hold. Suppose further that \mathcal{F} is a permissible, uniformly bounded class of functions and the following uniformity condition holds

$$\int_0^\infty \sqrt{\log N_2(\epsilon, \mathcal{F})} d\epsilon < \infty. \quad (2.4)$$

Then the process

$$\mathbb{Z}_n^* = n_{A\mathcal{M}}^{*1/2} \left[\frac{1}{n_{A\mathcal{M}}^*} \sum_{i=1}^{l_n^*-1} f(B_i^*) - \frac{1}{\widehat{n}_{A\mathcal{M}}} \sum_{i=1}^{\widehat{l}_n-1} f(\widehat{B}_i) \right] \quad (2.5)$$

converges in probability under \mathbb{P}_ν to a Gaussian process G indexed by \mathcal{F} whose sample paths are bounded and uniformly continuous with respect to the metric $L_2(\mu)$.

Proof. The proof relies on bootstrap central limit theorem from [72]. In order to prove weak convergence of process \mathbb{Z}_n^* we have to verify:

1. Finite dimensional convergence of distributions of \mathbb{Z}_n^* to G .
2. Stochastic asymptotic equicontinuity in probability under \mathbb{P}_ν with respect to the totally bounded semimetric

$$\rho_1(f, g) = \mathbb{E}[(f - g)^2]$$

on \mathcal{F} .

Firstly, we prove that $(\mathbb{Z}_n^*(f_1), \dots, \mathbb{Z}_n^*(f_k))$ converges weakly in probability to $(G(f_1), \dots, G(f_k))$ for every fixed finite collection of functions $\{f_1, \dots, f_k\} \subset \mathcal{F}$. Denote by \xrightarrow{L} weak convergence in law in sense of Hoffmann-Jørgensen. We want to show that for any fixed collection $(a_1, \dots, a_k) \in \mathbb{R}$ we have

$$\sum_{j=1}^k a_j \mathbb{Z}_n^*(f_j) \xrightarrow{L} \mathcal{N}(0, \gamma^2) \quad \text{in probability under } \mathbb{P}_\nu,$$

where

$$\gamma^2 = \lim_{n \rightarrow \infty} \sum_{j=1}^k a_j^2 \text{Var}(\mathbb{Z}_n(f_j)) + \sum_{s \neq r} a_s a_r \text{Cov}(\mathbb{Z}_n(f_s), \mathbb{Z}_n(f_r)).$$

Let $h = \sum_{j=1}^k a_j f_j$. By linearity of h and Theorem 4 we conclude that

$$\mathbb{Z}_n(h) \xrightarrow{L} G(h). \quad (2.6)$$

The above convergence of $\mathbb{Z}_n(h)$ coupled with Theorems 2 and 3 guarantee that

$$\mathbb{Z}_n^*(h) \xrightarrow{L} G(h) \quad \text{in probability under } \mathbb{P}_\nu.$$

Thus, the finite dimensional convergence for the $\mathbb{Z}_n^*(f)$, $f \in \mathcal{F}$ is established.

In order to show [2] we need to check if for every $\epsilon > 0$

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P}^*(\|\mathbb{Z}_n^*\|_{\mathcal{F}_\delta} > \epsilon) = 0 \text{ in probability under } \mathbb{P}_\nu, \quad (2.7)$$

where

$$\|R\|_{\mathcal{F}_\delta} := \sup\{|R(f) - R(g)| : \rho_1(f, g) < \delta\} \text{ and } R \in l^\infty(\mathcal{F}).$$

Moreover, we request that \mathcal{F} is totally bounded in $L_2(\mu)$. In fact, the latter was shown in [98]. For the completeness of exposition of the proof, we repeat the reasoning from the aforementioned paper.

Consider class of functions

$$\mathcal{H} = \{|f - g| : f, g \in \mathcal{F}\}.$$

Denote by Q_n the n -th empirical measure of an i.i.d. process whose law is μ . Using basic properties of covering numbers we conclude that

$$N_1(\epsilon, \mathcal{G}, Q_n) \leq \left(N_2\left(\frac{\epsilon}{2}, \mathcal{F}\right)\right)^2 < \infty$$

and thus by the SLLN for Q_n (see Theorem 3.6 in [98]) we have that

$$\sup_{h \in \mathcal{H}} |(Q_n - \mu)(h)| \rightarrow 0 \text{ a.s.}(\mu).$$

Since \mathcal{F} is totally bounded in $L_1(Q)$ for every measure Q with finite support it follows that is totally bounded in $L_1(\mu)$. Moreover, one can show that if an envelope of \mathcal{F} is in $L_2(\mu)$, then \mathcal{F} is totally bounded in $L_2(\mu)$.

In order to show (2.7), firstly, we replace the random numbers $n_{A_{\mathcal{M}}}^*$ and l_n^* by their deterministic equivalents. By the same arguments as in the proof of Theorem 3 (see [26], page 710 for details) we have the following convergences

$$\frac{l(B_j^*)}{n} \xrightarrow{P^*} 0 \text{ and } \frac{n_{A_{\mathcal{M}}}^*}{n} \xrightarrow{P^*} 1$$

in \mathbb{P}_ν -probability along the sample path as $n \rightarrow \infty$ and

$$\frac{l_n^*}{n} - \mathbb{E}_{A_{\mathcal{M}}}(\tau_{A_{\mathcal{M}}})^{-1} \xrightarrow{P^*} 0$$

in \mathbb{P}_ν -probability along the sample path as $n \rightarrow \infty$. Thus, we conclude that

$$\begin{aligned}\mathbb{Z}_n^*(f) &= \sqrt{n_{A_{\mathcal{M}}}^*} \left[\frac{1}{n_{A_{\mathcal{M}}}^*} \sum_{i=1}^{l_n^*-1} f(B_i^*) - \frac{1}{\widehat{n}_{A_{\mathcal{M}}}} \sum_{i=1}^{\widehat{l}_n-1} f(\widehat{B}_i) \right] \\ &= \frac{1}{\sqrt{n_{A_{\mathcal{M}}}^*}} \left[\sum_{i=1}^{l_n^*-1} \{f(B_i^*) - \widehat{\mu}_n(f)l(B_i^*)\} \right] \\ &= \frac{1}{\sqrt{n}} \left[\sum_{i=1}^{1+\left\lfloor \frac{n}{\mathbb{E}_{A_{\mathcal{M}}}(\tau_A)} \right\rfloor} \{f(B_i^*) - \widehat{\mu}_n(f)l(B_i^*)\} \right] + o_{\mathbb{P}^*}(1),\end{aligned}$$

where $\lfloor x \rfloor$ is an integer part of $x \in \mathbb{R}$. The last line of above reasoning is deduced from [26], page 710, line 15 (see also proof of Theorem 17.2.2, page 429 in [106]). The preceding reasoning allows us to switch to the process

$$\mathbb{U}_n^*(f) = \frac{1}{\sqrt{n}} \left[\sum_{i=1}^{1+\left\lfloor \frac{n}{\mathbb{E}_{A_{\mathcal{M}}}(\tau_A)} \right\rfloor} \{f(B_i^*) - \widehat{\mu}_n(f)l(B_i^*)\} \right].$$

Observe, that

$$\{f(B_i^*) - \widehat{\mu}_n(f)l(B_i^*)\}_{i \geq 1}$$

forms the sequence of i.i.d. random variables. We emphasize the fact, that the number of observations $1 + \left\lfloor \frac{n}{\mathbb{E}_{A_{\mathcal{M}}}(\tau_A)} \right\rfloor$ is *deterministic*.

Now we use the same arguments as in [122], Step 7 of the proof of Theorem 2.2. We are dealing with i.i.d. bootstrap blocks. It is known from [72] (see the proof of Theorem 3.1 therein, page 863) that stochastic asymptotic equicontinuity of the bootstrap process \mathbb{Z}_n^* is implied by stochastic asymptotic equicontinuity of the original process \mathbb{Z}_n and provided that

$$\mathbb{E}[\sup_{f \in \mathcal{F}} |f(B_i)|^2] \leq \mathbb{E} \left[\left(\sum_{i=1}^{\tau_A} F(X_i) \right)^2 \right] < +\infty$$

and since by result of [98] (see Theorem 4.9 therein) the original process \mathbb{Z}_n is equicontinuous, we conclude the stochastic asymptotic equicontinuity of \mathbb{Z}_n^* .

The above reasoning implies that (2.7) holds. We have checked that both conditions [1] and [2] are fulfilled by \mathbb{Z}_n^* . Thus, we can apply the bootstrap CLT from [72] which yields the desired result. \square

Remark 6. It is noteworthy that Theorem 5 is a generalization of Theorem 2.2 from [122] for countable Markov chains. One can apply the reasoning from the above proof to Radulović's setting. As a consequence, the part concerning verification of the asymptotic stochastic equicontinuity of the bootstrap version of the empirical process indexed by uniformly bounded class of functions \mathcal{F} may be significantly simplified. As shown in the proof of Theorem 5, we can switch from process

$$\mathbb{Z}_n^*(f)_{f \in \mathcal{F}} := \sqrt{n^*} \{ \mu_n^*(f) - \mu_{n_A}(f) \},$$

where $n_A = \tau_A(l_n) - \tau_A$ to process

$$\mathbb{U}_n^*(f) = \frac{1}{\sqrt{n}} \left[\sum_{i=1}^{1 + \lfloor \frac{n}{\mathbb{E}_A(\tau_A)} \rfloor} \{ f(B_i^*) - \mu_{n_A}(f) l(B_i^*) \} \right]$$

and standard probability inequalities applied to the i.i.d. blocks of data yield the result.

In the following, we show that we can relax the assumption of uniform boundedness imposed on class \mathcal{F} . Due to results from [137], we can only require that \mathcal{F} has an envelope in $L_2(\mu)$. That assumed, one can easily yield uniform bootstrap central limit theorem in the unbounded case.

Theorem 6. *Suppose that (X_n) is a positive Harris recurrent Markov chain. Suppose that the conditions [1] and [2] are satisfied by the chain and $\mathcal{H}_1(f, \rho, \nu)$, $\mathcal{H}_2(\rho, \nu)$ with $\rho \geq 4$, $\mathcal{H}_3, \mathcal{H}_4$ and \mathcal{H}_5 hold. Assume further that \mathcal{F} is a permissible class of functions and such that the envelope F satisfies*

$$\mathbb{E}_{A_{\mathcal{M}}} \left[\sum_{\tau_{A_{\mathcal{M}}} < j \leq \tau_{A_{\mathcal{M}}}(2)} F(X_j) \right]^2 < \infty. \quad (2.8)$$

Suppose, that the following uniformity condition holds

$$\int_0^\infty \sqrt{\log N_2(\epsilon, \mathcal{F})} d\epsilon < \infty. \quad (2.9)$$

Then the process

$$\mathbb{Z}_n^* = n_{A_{\mathcal{M}}}^{*1/2} \left[\frac{1}{n_{A_{\mathcal{M}}}^*} \sum_{i=1}^{l_n^*-1} f(B_i^*) - \frac{1}{\hat{n}_{A_{\mathcal{M}}}} \sum_{i=1}^{\hat{l}_n-1} f(\hat{B}_i) \right] \quad (2.10)$$

converges in probability under \mathbb{P}_ν to a Gaussian process G indexed by \mathcal{F} whose sample paths are bounded and uniformly continuous with respect to the metric $L_2(\mu)$.

Proof. The proof of Theorem 6 goes analogously to the proof of Theorem 5 with few natural modifications. We indicate the critical points where changes are necessary. The notation remains in the agreement with the previous theorem.

- Theorem 4.3 from [137] guarantees weak convergence

$$\mathbb{Z}_n(h) \xrightarrow{L} G(h).$$

- Due to [26], Theorem 3.3 (see page 706 therein) is also valid when f is unbounded. Thus, the finite dimensional convergence of distributions of \mathbb{Z}_n^* to the right Gaussian process holds.
- It is shown in [137] that \mathcal{F} is totally bounded in $L_2(\mu)$ when \mathcal{F} fulfills only the condition that the envelope F is in $L_2(\mu)$ (see [137], page 9 for details).

Proof of stochastic asymptotic equicontinuity of process \mathbb{Z}_n^* is exactly the same as in Theorem 5. □

4 Bootstrap uniform central limit theorems for Fréchet differentiable functionals of Markov chains

We have shown in [44], that one can combine bootstrap uniform central limit Theorems 5 and 6 with results of Levental [98] in order to obtain bootstrap uniform central limit theorems for general differentiable functionals over uniformly bounded classes (and with an envelope in $L_2(\mu)$) of functions \mathcal{F} when applying limit results from [137].

4.1 Preliminary assumptions and remarks

As previously mentioned, influence function plays a crucial role to detect outliers in the data. More specifically, functions and estimators which have an unbounded influence function should be carefully investigated, because the small proportion of the observations could have too much influence on the estimator.

In what follows, we make our considerations rigorous. We denote by \mathcal{P} the set of all probability measures on E . We keep the notation in agreement with notation introduced in [29].

The classical definition of influence function (which comes from [140]) is provided below.

Definition 7. Let $(\vartheta, \|\cdot\|)$ be a separable Banach space. Let $T : \mathcal{P} \rightarrow \vartheta$ be a functional on \mathcal{P} . If the limit

$$\frac{T((1-t)\mu + t\delta_x) - T(\mu)}{t}, \quad \text{as } t \rightarrow 0$$

exists at $\mu \in \mathcal{P}$ and for any $x \in E$, then the influence function $T^{(1)}(\cdot, \mu) : \mathcal{E} \rightarrow \vartheta$ of the functional T at μ is given by, for all $x \in E$,

$$T^{(1)}(x, \mu) = \lim_{t \rightarrow 0} \frac{T((1-t)\mu + t\delta_x) - T(\mu)}{t}.$$

In what follows, we provide the definition of Fréchet derivative which is a significant concept in robust statistics. More specifically, Fréchet differentiability guarantees the existence of influence function. We denote by d some metric on \mathcal{P} .

Definition 8. We say that the functional $T : \mathcal{P} \rightarrow \mathbb{R}$ is Fréchet differentiable at $\mu_0 \in \mathcal{P}$ for a metric d , if there exists a continuous linear operator DT_{μ_0} (from the set of signed measures of the form $\mu - \mu_0$ in $(\vartheta, \|\cdot\|)$) and a function $\epsilon^{(1)}(\cdot, \mu_0) : \mathbb{R} \rightarrow (\vartheta, \|\cdot\|)$, which is continuous at 0 and $\epsilon^{(1)}(0, \mu_0) = 0$ such that

$$\forall \mu \in \mathcal{P}, \quad T(\mu) - T(\mu_0) = DT_{\mu_0}(\mu - \mu_0) + R^{(1)}(\mu, \mu_0),$$

where

$$R^{(1)}(\mu, \mu_0) = d(\mu, \mu_0)\epsilon^{(1)}(d(\mu, \mu_0), \mu_0).$$

Furthermore, we assume that T admits the integral representation

$$\forall \mu \in \mathcal{P}, \quad DT_{\mu_0}(\mu - \mu_0) = \int_E T^{(1)}(x, \mu_0)\mu(dx).$$

It is noteworthy that when establishing uniform central limit theorems for generally differentiable functionals the appropriate choice of metric is the crucial point. To be more specific, the 'right' choice of metric allows to control the distance $d(\mu_n, \mu)$ and the remainder $R^{(1)}(\mu_n, \mu)$ in a very precise way. In this framework we measure the distance $d(\mu_n, \mu)$ using a generalization of the Kolmogorov's distance. We provide the definition of this metric (in the same form as in [17]) below.

Definition 9. Let \mathcal{H} be a class of real-valued functions (we do not assume measurability conditions as one can work with outer measures and the Hoffmann-Jørgensen convergence). We define a distance

$$d_{\mathcal{H}}(P, Q) := \sup_{h \in \mathcal{H}} \left| \int h d(P - Q) \right| \quad (2.11)$$

for any $P, Q \in \mathcal{P}$.

We have chosen metric defined in (2.11) after careful analysis of the arguments contained in the discussion in [17] and [58]. Essentially, metric $d_{\mathcal{H}}$ enables a very precise control of the distance $d(\mu_n, \mu)$ as we previously required. As a further matter, in many cases we can find a class of functions \mathcal{H} , which makes the functionals Fréchet differentiable for $d_{\mathcal{H}}$. The latter is a significant advantage since choice of metric that guarantees Fréchet differentiability of functionals is usually challenging (see [17] and [58] for exhaustive discussions on this subject).

We make an observation that permissible, uniformly bounded (or with an envelope in $L_2(\mu)$) classes of functions \mathcal{F} satisfy the conditions assumed on class \mathcal{H} . Thus, we can ease the notation and write $d_{\mathcal{F}}$ for the distance defined by (2.11).

4.2 Main asymptotic results

In this section, we show how uniform central limit theorems from [98] and [137] combined with results from previous section may be used to obtain bootstrap limit results for Fréchet differentiable functionals. For the reader's convenience we briefly recall the notation we will be using. Recall that

$$\mu_n^* = \frac{1}{n_{A_{\mathcal{M}}}^*} \sum_{i=1}^{l_n^*-1} f(B_i^*) \quad \text{and} \quad \hat{\mu}_n = \frac{1}{\hat{n}_{A_{\mathcal{M}}}} \sum_{i=1}^{\hat{l}_n-1} f(\hat{B}_i),$$

where \hat{B}_i , $i = 1, \dots, \hat{l}_n - 1$ are pseudo-regeneration blocks. In atomic regenerative case, the empirical mean is of the form

$$\mu_n = \frac{1}{n_A} \sum_{i=1}^{l_n-1} f(B_i).$$

Observe that the key point to obtain bootstrap uniform CLTs for Fréchet differentiable functionals is a very precise command of distance $d_{\mathcal{F}}(\mu_n^*, \hat{\mu}_n)$ (we require it would be sufficiently small) so that we can control the remainder term $R^{(1)}(\mu_n^*, \hat{\mu}_n)$. Next, the uniform central limit theorem guarantees that the linear part of the $T(\mu_n^*) - T(\hat{\mu}_n)$ is converging weakly to a desired normal distribution which yields our result.

Theorem 7. *Let \mathcal{F} be a permissible, uniformly bounded class of functions, such that*

$$\int_0^\infty \sqrt{\log N_2(\epsilon, \mathcal{F})} d\epsilon < \infty.$$

Suppose that the conditions of Theorem 5 hold and $T : \mathcal{P} \rightarrow \mathbb{R}$ is Fréchet differentiable functional at μ with respect to metric $d_{\mathcal{F}}$. Then, in a general Harris positive recurrent case, we have that $n^{1/2}(T(\mu_n^) - T(\hat{\mu}_n))$ converges weakly to a normal distribution $\mathcal{N}(0, \text{Var}(T^{(1)}))$.*

Remark 7. It is a straightforward observation that the above theorem is valid also in the regenerative case. Indeed, replace $\mathcal{A}_{\mathcal{M}}$ and the $\hat{\mu}_n$ for the split chain by \mathcal{A} and μ_n respectively. Then, under the assumptions of Theorem 7, we obtain weak convergence to a desired normal distribution.

Proof. Without loss of generality, we assume that $\mathbb{E}_{\mu} T^{(1)}(x, \mu) = 0$. By Fréchet differentiability formulated in Definition 8 we have

$$T(\hat{\mu}_n) - T(\mu) = DT_{\mu}(\hat{\mu}_n - \mu) + d_{\mathcal{F}}(\hat{\mu}_n, \mu)\epsilon^{(1)}(d_{\mathcal{F}}(\hat{\mu}_n, \mu), \mu) \quad (2.12)$$

and

$$T(\mu_n^*) - T(\mu) = DT_{\mu}(\mu_n^* - \mu) + d_{\mathcal{F}}(\mu_n^*, \mu)\epsilon^{(1)}(d_{\mathcal{F}}(\mu_n^*, \mu), \mu). \quad (2.13)$$

Thus,

$$\begin{aligned} \sqrt{n}(T(\mu_n^*) - T(\hat{\mu}_n)) &= \sqrt{n}(DT_{\hat{\mu}_n}(\mu_n^* - \hat{\mu}_n)) + \sqrt{n}(d_{\mathcal{F}}(\hat{\mu}_n, \mu)\epsilon^{(1)}(d_{\mathcal{F}}(\hat{\mu}_n, \mu), \mu)) \\ &\quad + \sqrt{n}(d_{\mathcal{F}}(\mu_n^*, \mu)\epsilon^{(1)}(d_{\mathcal{F}}(\mu_n^*, \mu), \mu)). \end{aligned}$$

Next, we prove that $d_{\mathcal{F}}(\hat{\mu}_n, \mu)$ and $d_{\mathcal{F}}(\mu_n^*, \mu)$ are of order $O_{\mathbb{P}_{\nu}}(n^{-1/2})$. Theorem 4 guarantees that

$$\sqrt{n}d_{\mathcal{F}}(\hat{\mu}_n, \mu) \xrightarrow{L} \sup_{f \in \mathcal{F}} |G(f)|, \text{ as } n \rightarrow \infty,$$

where G is a Gaussian process whose sample paths are bounded and uniformly continuous with respect to the metric $L_2(\mu)$. Thus, $d_{\mathcal{F}}(\hat{\mu}_n, \mu) = O_{\mathbb{P}_{\nu}}(n^{-1/2})$.

Next, observe that

$$d_{\mathcal{F}}(\mu_n^*, \mu) \leq d_{\mathcal{F}}(\mu_n^*, \hat{\mu}_n) + d_{\mathcal{F}}(\hat{\mu}_n, \mu).$$

From Theorem 5 we conclude that

$$\sqrt{n}d_{\mathcal{F}}(\mu_n^*, \hat{\mu}_n) \xrightarrow{L^*} \sup_{f \in \mathcal{F}} |G(f)|, \text{ as } n \rightarrow \infty.$$

Thus, $d_{\mathcal{F}}(\mu_n^*, \hat{\mu}_n) = O_{P^*}(n^{-1/2})$. We show that $d_{\mathcal{F}}(\mu_n^*, \hat{\mu}_n) = O_{\mathbb{P}_{\nu}}(n^{-1/2})$. Following similar reasoning as in [128], we consider the sequence S_n of order $O_{P^*}(1)$ in \mathbb{P}_{ν} -probability along the sample, i.e.

$$\lim_{T \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}^*\{|S_n| \geq T\} \rightarrow 0 \text{ in } \mathbb{P}_{\nu} \text{ - probability along the sample.}$$

Then,

$$\begin{aligned} \lim_{T \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}_{\nu}\{|S_n| \geq T\} &= \lim_{T \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{E}_{\nu} [\mathbb{P}^*\{|S_n| \geq T\}] \\ &\leq \lim_{T \rightarrow \infty} \mathbb{E}_{\nu} \left[\limsup_{n \rightarrow \infty} \mathbb{P}^*\{|S_n| \geq T\} \right] \\ &= \mathbb{E}_{\nu} \left[\lim_{T \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}^*\{|S_n| \geq T\} \right] = 0 \end{aligned}$$

by the dominated convergence theorem and the Fatou's lemma. Thus, $d_{\mathcal{F}}(\mu_n^*, \hat{\mu}_n) = O_{\mathbb{P}_\nu}(n^{-1/2})$ and $d_{\mathcal{F}}(\mu_n^*, \mu) = O_{\mathbb{P}_\nu}(n^{-1/2})$.

Next, we scale (2.12) by \sqrt{n} :

$$\sqrt{n}(T(\hat{\mu}_n) - T(\mu)) = \sqrt{n}(DT_\mu(\hat{\mu}_n - \mu)) + o_{\mathbb{P}_\nu}(1)$$

and apply Theorem 5. Note that the linear part in the above equation is Gaussian as long as $0 < \mathbb{E}_\mu T^{(1)}(X_i, \mu)^2 \leq C_1^2(\mu) \mathbb{E}_\mu F^2(X) < \infty$ (see [17], Chapter 1 for details), but that assumption is of course fulfilled since \mathcal{F} is uniformly bounded. Thus, the following weak convergence in $l^\infty(\mathcal{F})$ holds:

$$\begin{aligned} \sqrt{n}(T(\hat{\mu}_n) - T(\mu)) &= \sqrt{n}(DT_\mu(\hat{\mu}_n - \mu)) + o_{\mathbb{P}_\nu}(1) \\ &= \sqrt{n} \int_E T^{(1)}(x, \mu)(\hat{\mu}_n - \mu) d(x) \\ &= \sqrt{n} \left[\frac{1}{\hat{n}_{\mathcal{A}_{\mathcal{M}}}} \sum_{i=1}^{\hat{n}_{\mathcal{A}_{\mathcal{M}}}} T^{(1)}(X_i, \mu) - 0 \right] + o_{\mathbb{P}_\nu}(1) \xrightarrow{L} \mathcal{N}(0, \text{Var}(T^{(1)})(x, \mu)). \end{aligned}$$

By the previous discussion, we also have

$$\begin{aligned} \sqrt{n}(T(\mu_n^*) - T(\mu)) &= \sqrt{n}(DT_\mu(T(\mu_n^*) - \mu)) + o_{\mathbb{P}_\nu}(1) \\ &= \sqrt{n} \int_E T^{(1)}(x, \mu)(\mu_n^* - \mu) d(x) \\ &= \sqrt{n} \left[\frac{1}{n_{\mathcal{A}_{\mathcal{M}}}^*} \sum_{i=1}^{n_{\mathcal{A}_{\mathcal{M}}}^*} T^{(1)}(X_i^*, \mu) - 0 \right] + o_{\mathbb{P}_\nu}(1). \end{aligned}$$

The above convergences yield

$$\begin{aligned} \sqrt{n}[T(\mu_n^*) - T(\hat{\mu}_n)] &= \sqrt{n} \left[\frac{1}{n_{\mathcal{A}_{\mathcal{M}}}^*} \sum_{i=1}^{n_{\mathcal{A}_{\mathcal{M}}}^*} T^{(1)}(X_i^*, \mu) - \frac{1}{\hat{n}_{\mathcal{A}_{\mathcal{M}}}} \sum_{i=1}^{\hat{n}_{\mathcal{A}_{\mathcal{M}}}} T^{(1)}(X_i, \mu) \right] + o_{\mathbb{P}_\nu}(1) \\ &\xrightarrow{L} \mathcal{N}(0, \text{Var}(T^{(1)})(x, \mu)) \end{aligned}$$

converges weakly conditionally on the data in \mathbb{P}_ν -probability along the sample as $n \rightarrow \infty$ (by Theorem 3) and this completes the proof. \square

Theorem 7 can be easily extended to case when \mathcal{F} is unbounded and has an envelope in $L_2(\mu)$.

Theorem 8. *Let \mathcal{F} be a permissible class of functions such that the envelope F satisfies*

$$\mathbb{E}_{A_{\mathcal{M}}} \left[\sum_{\tau_{A_{\mathcal{M}}} < j \leq \tau_{A_{\mathcal{M}}}(2)} F(X_j) \right]^2 < \infty. \quad (2.14)$$

Suppose, that the following uniformity condition holds

$$\int_0^\infty \sqrt{\log N_2(\epsilon, \mathcal{F})} d\epsilon < \infty. \quad (2.15)$$

Assume further that the conditions of Theorem 6 hold and that $T : \mathcal{P} \rightarrow \mathbb{R}$ is Fréchet differentiable functional at μ with respect to metric $d_{\mathcal{F}}$. Then, in general Harris positive recurrent case, we have that $n^{1/2}(T(\mu_n^) - T(\hat{\mu}_n))$ converges weakly to a normal distribution $\mathcal{N}(0, \text{Var}(T^{(1)}))$.*

The proof of Theorem 8 follows analogously to the proof of Theorem 7 and thus is omitted. We only make a brief observation that in order to prove Theorem 8 in the unbounded case we apply limit results from [137] instead of uniform central limit theorem obtained by Levental (where uniform boundedness of \mathcal{F} is assumed). Moreover, we need Theorem 6 to control the remainder term. With aforementioned changes taken into account, the reasoning goes line by line as in the proof of Theorem 7.

Remark 8. It is noteworthy that Theorem 8 is also true in atomic regenerative case. Indeed, replace $\hat{\mu}_n$ and $A_{\mathcal{M}}$ by μ_n and A . The proof goes analogously as in the preceding theorems.

5 Conclusion

In this chapter, we have shown how the regenerative properties of Markov chains can generalize some concepts in non-parametric statistics from i.i.d. to dependent case. We have shown that uniform bootstrap functional central limit theorem holds over permissible, uniformly bounded classes of functions. We have proved that the uniform boundedness assumption imposed on \mathcal{F} can be weakened and it is feasible to require that \mathcal{F} has an envelope in $L_2(\mu)$. We have worked with Markov chains on the general state space, but our results can be directly applied to Markov chains on countable state space. Thus, some proofs of the already existing results for the countable case, can be simplified when just applying the methodology introduced in this paper.

The bootstrap asymptotic results for empirical processes indexed by \mathcal{F} naturally lead to bootstrap central limit theorems for Fréchet differentiable functionals. We have shown

that bootstrap uniform CLTs hold in the bounded and the unbounded case over \mathcal{F} . Similar approach can be also applied to Hadamard differentiable functionals in order to establish analogous asymptotic results to presented in this paper.

It is noteworthy that the aforementioned results can be extended to case when the blocks are 1-dependent (when $m > 1$ in minorization condition (1.1) in Chapter 1). Indeed, one can consider sums of even and odd blocks as explained in [98] and in Remark 29 later in Chapter 5. Note that Levental established uniform central limit theorem for Harris Markov chains in case $m > 1$ in Theorem 5.9 in [98]. Thus, (given the results in [98] for 1-dependent blocks) it should not be difficult to establish bootstrap versions of Theorems 5, 6, 7 and 8 in general case when minorization condition (1.1) is satisfied with $m > 1$.

Chapter 3

Robust estimation for Markov chains with applications to PDMPs

The purpose of this chapter is to propose a method to build robust estimators for piecewise-deterministic Markov processes (PDMPs) which is of particular interest when the underlying process is contaminated by some outliers. The aforementioned method relies on a renewal theory for Markov chains and further developments of the approximate regenerative block bootstrap method recalled in Chapter 1. The idea is to eliminate blocks having either too much contribution to the statistics of interest, or having a too large length, and to build efficient and robust estimators for the embedded Markov chain associated to the PDMP. Relating the properties of the underlying process and its embedded chain, this leads to robust estimators for the PDMP. This chapter develops further some concepts of robustness in a Markovian setting presented in [29].

To highlight the applicability of the method, we consider robust estimators of several risk indicators such as the ruin probability, the expected shortfall and the extremal index of two PDMPs: the Cramér-Lundberg with a dividend barrier and the Kinetic Dynamic Exposure Model (KDEM) used in modeling pharmacokinetics of contaminants (see [31] for instance).

Our approach boils down to seeing most of the interesting parameters in a Markov chain framework as functionals of the stationary measure which may itself be seen as a functional of the distribution of the regeneration blocks. It is then possible to introduce the influence function and the tools of robustness for such parameters. In opposition to the i.i.d. case, many parameters are not robust (including quantiles, L-estimates, *etc*) in sense that a single outlier may have a disastrous effect on the whole process because of its dynamic.

This chapter gathers material published in [25], which is Chapter 5 in [14].

1 Definition

In this section, we define a PDMP process. We briefly mention that PDMP is a process whose behaviour is determined by random jumps at points in time and its evolution is *deterministically* ruled by an ordinary differential equation between those times (see [51]). In what follows we provide a brief characteristic of a general PDMP process. Our considerations are based on [14], Chapter 1 and we keep the same notations. We also refer to the aforementioned chapter for more details concerning theory of PDMPs.

In what follows, we will give a general definition of a PDMP process. Assume that $(Z_t)_{t \geq 0}$ is a time-homogeneous Markov process with càdlàg sample paths a.s.. The process Z takes its values in an open subset $\mathcal{X} \subseteq \mathbb{R}^n$, for some $n \geq 1$. Moreover, $\partial\mathcal{X}$ designates the frontier of \mathcal{X} and $\bar{\mathcal{X}}$ is the closure of \mathcal{X} . The Markov semigroup of Z is defined as the family of operators $(P_t)_{t \geq 0}$ acting on a bounded measurable functions f such that, for $t > 0$,

$$P_t f(z) = \mathbb{E}[f(Z_t) | Z_0 = z].$$

A PDMP is specified by three characteristics: a deterministic flow, a jump rate and a jump kernel. We denote by $R : \mathcal{X} \rightarrow \mathcal{X}$ a vector field. We suppose further that the ordinary differential equation

$$\begin{cases} \partial_t \Phi(z, t) = R(\Phi(z, t)) \\ \Phi(z, 0) = z \end{cases} \quad (3.1)$$

has a unique solution. Then, $\Phi : \mathcal{X} \times \mathbb{R} \rightarrow \mathcal{X}$ is the flow of Z , which propels the dynamics of the process between the jumps. We make the convention $\inf \emptyset = +\infty$. The hitting times of the frontier $\partial\mathcal{X}$ starting from z are of the form:

$$\begin{aligned} t^+(x) &= \inf\{t > 0 : \Phi(z, t) \in \partial\mathcal{X}\}, \\ t^-(x) &= \inf\{t > 0 : \Phi(z, -t) \in \partial\mathcal{X}\}. \end{aligned}$$

Next, we specify the jump rate $\lambda : \mathcal{X} \rightarrow \mathbb{R}_+$ which sets off the jump dynamics of the process (observe that the larger $\lambda(z)$, the higher the probability to jump). Furthermore, Q associates a probability kernel¹ over $\mathcal{X} \times \mathcal{B}(\mathcal{X})$, which will governs the direction of the jumps of Z .

In what follows, we define the sample paths of Z in a recursive way. Provided that $T_0 = 0$ and $Z_0 \in \mathcal{X}$, assume ΔT_1 is a positive random variable such that, for all $t > 0$,

$$\mathbb{P}(\Delta T_1 \geq t) = \exp\left(-\int_0^t \lambda(\Phi(Z_0, s)) ds\right) \mathbb{I}_{\{t < t^+(Z_0)\}}. \quad (3.2)$$

¹ Q associates to each point $z \in \mathcal{X}$ a probability measure $Q(z, \cdot)$, see [85]. Q can also be assigned to as a Markov kernel.

Next, we denote by $T_1 = T_0 + \Delta T_1$ the first jump epoch and U_1 is such that

$$\mathbb{P}_{U_1} = Q(\Phi(Z_0, \Delta T_1), \cdot).$$

The sample paths for Z for $t \in [T_0, T_1]$ are constructed according to the following scheme

$$Z_t = \begin{cases} \Phi(Z_0, t) & \text{if } t \in [T_0, T_1) \\ U_1 & \text{if } t = T_1. \end{cases} \quad (3.3)$$

Now, (3.3) can be recursively expanded for $t \in [T_n, T_{n+1}]$ in order to construct the sample paths of Z assuming that ΔT_i are independent. Observe that the exponential factor in (3.2) sets off random jumps at rate λ . Example 3 in Chapter 1 is a particular case of a PDMP where the dynamic can be understood in Figure 3, we will further study this model in this chapter.

2 Robust functional parameter estimation for Markov Chains

Throughout rest of this chapter we consider the embedded chain $(X_n)_{n \in \mathbb{N}}$ with $X_n = Z(T_n)$, $n = 1, \dots, \infty$. We assume that X_n is an atomic regenerative or Harris recurrent Markov chain with initial probability ν and with transition probability Π .

The concepts of *influence function* and/or *robustness* in the i.i.d. setting provide tools to detect *outliers* among the data or influential observations. It also allows to generalize the important notion of *efficient estimation* in semiparametric frameworks; see [35]. Extending the notion of *influence function* and/or *robustness* to the general time series framework is a difficult task; see [90] or [103].

Measuring the influence of a single observation hardly makes sense, due to the dependence structure across time. An extension to the Markovian setting based on martingale approximation has been proposed in [109]. Alternatively, the regenerative approach gives an opportunity to extend in much more natural way the notion of the influence function based on the (approximate) regeneration blocks construction. As shown in the subsequent analysis, this approach immediately leads to central limit and convolution theorems.

2.1 The influence function on the torus

Just like the stationary probability distribution $\mu(dx)$, most parameters of interest related to Harris positive chains are functionals of the distribution \mathcal{L} of the regenerative blocks on the

torus $\mathbb{T} = \cup_{n \geq 1} E^n$, namely the distribution of (X_1, \dots, X_{τ_A}) conditioned on $X_0 \in A$ when the chain possesses an accessible atom A , or the distribution of $(X_1, \dots, X_{\tau_{A_{\mathcal{M}}}})$ conditioned on $(X_0, Y_0) \in A_{\mathcal{M}}$ in the general case when one considers the split chain.

For simplicity, we shall omit the subscript \mathcal{M} and make no notational distinction between the regenerative and pseudo-regenerative cases unless specified otherwise. Indeed, the probability distribution \mathbb{P}_ν of the Markov chain X starting from ν can be factorized as follows:

$$\mathbb{P}_\nu((X_n)_{n \geq 1}) = \mathcal{L}_\nu((X_1, \dots, X_{\tau_{A(1)}})) \prod_{k=1}^{\infty} \mathcal{L}((X_{1+\tau_A(k)}, \dots, X_{\tau_A(k+1)})),$$

where \mathcal{L}_ν means the conditional distribution of (X_1, \dots, X_{τ_A}) given that $X_0 \sim \nu$. Any functional of the law of the discrete-time process $(X_n)_{n \geq 1}$ can be thus expressed as a functional of the pair $(\mathcal{L}_\nu, \mathcal{L})$. In the time-series asymptotic framework, since the distribution of \mathcal{L}_ν cannot be estimated in general, only functionals of \mathcal{L} are of practical interest.

We propose a notion of influence function for such statistics. Let $\mathcal{P}_{\mathbb{T}}$ denote the set of all probability measures on the torus \mathbb{T} and for any $b \in \mathbb{T}$, set

$$L(b) = k \text{ if } b \in E^k, \ k \geq 1.$$

We then have the following natural definition, which straightforwardly extends the classical notion of influence function in the i.i.d. case, with the important novelty that distributions on the torus are considered here.

Definition 10. (INFLUENCE FUNCTION ON THE TORUS) Let $(\mathcal{V}, \|\cdot\|)$ be a separable Banach space. Let $T : \mathcal{P}_{\mathbb{T}} \rightarrow \mathcal{V}$ be a functional on $\mathcal{P}_{\mathbb{T}}$. If, for some \mathcal{L} in $\mathcal{P}_{\mathbb{T}}$,

$$t^{-1}(T((1-t)\mathcal{L} + t\delta_b) - T(\mathcal{L}))$$

has a finite limit as $t \rightarrow 0$ for any $b \in \mathbb{T}$, the influence function $T^{(1)} : \mathcal{P}_{\mathbb{T}} \rightarrow \mathcal{V}$ of the functional T at \mathcal{L} is then said to be well-defined, and, by definition, one set for all b in \mathbb{T} ,

$$T^{(1)}(b, \mathcal{L}) = \lim_{t \rightarrow 0} \frac{T((1-t)\mathcal{L} + t\delta_b) - T(\mathcal{L})}{t}. \quad (3.4)$$

Definition 11. (GROSS-ERROR SENSITIVITY) A functional T is said to be Markov-robust iff its influence function $T^{(1)}(b, \mathcal{L})$ is bounded on the torus \mathbb{T} . The gross-error sensitivity to block contamination is then defined as

$$\gamma^*(T, \mathcal{L}) = \sup_{b \in \mathbb{T}} \|T^{(1)}(b, \mathcal{L})\|.$$

The angle on robustness embraced in this paper is not the one deriving from the sample break point definition (recall incidentally that it may be defined in the time series context in several ways). The concepts we develop here serve to decide whether a specific (*pseudo-regenerative data block*) has an important influence on the value of some given estimate or not, and/or whether it may be considered as an *outlier*.

The notion of robustness is related to blocks of observations, instead of individual observations. Heuristically, one may consider that, given the regenerative dependence structure of the process, a single suspiciously outlying value at some time point n may have a strong impact on the whole trajectory, until the (split) chain regenerates again, so that not only this particular observation but also the whole "contaminated" segment of observations (corresponding to a block outlier) should be eventually removed.

Roughly stated, it turns out that examining (approximate) regeneration blocks and its impact on the functional of interest, allows to identify more accurately outlying data in the sample path, as well as their nature. In the time series framework, different type of outliers may occur, such as additive or innovative outliers. By comparing the data blocks this way (their length, as well as the values of the functional of interest on these blocks), one may detect those which should be preferably removed from subsequent computations.

As illustrated by the three examples listed below, standard computations of the influence function in the i.i.d. context can be straightforwardly extended to the Markovian framework.

2.2 Example 1: Sample means

Suppose that X is positive recurrent with stationary distribution μ . Let $f : E \rightarrow \mathbb{R}$ be a μ -integrable real function and consider the parameter $\mu(f) := \mathbb{E}_\mu[f(X)]$. Denote by B a r.v. valued in \mathbb{T} with distribution \mathcal{L} and observe that

$$\mu(f) = \mathbb{E}_\mathcal{L}[\mathbf{f}(B)] / \mathbb{E}_\mathcal{L}[L(B)] = T(\mathcal{L}),$$

with the notation

$$\mathbf{f}(b) := \sum_{i=1}^{L(b)} f(b_i) \quad \text{for any } b = (b_1, \dots, b_{L(b)}) \in \mathbb{T}.$$

A classical calculation for the influence function of ratios yields

$$T^{(1)}(b, \mathcal{L}) = \frac{d}{dt}(T((1-t)\mathcal{L} + tb))|_{t=0} = \frac{\mathbf{f}(b) - \mu(f)L(b)}{\mathbb{E}_\mathcal{L}[L(B)]}.$$

Notice that $\mathbb{E}_\mathcal{L}[T^{(1)}(B, \mathcal{L})] = 0$.

In the i.i.d. setting it is known that, if f is bounded by some constant $M_f < \infty$, the corresponding functional is robust and may be simply estimated by its empirical counterpart. In the Markovian situation, even in the bounded case, $T^{(1)}(b, \mathcal{L})$ is generally not bounded and $\gamma^*(T, \mathcal{L}) = \infty$. This point has also been stressed in [103], with a different definition of the influence function however.

A robustified version of this parameter can be defined as

$$\tilde{T}_M(\mathcal{L}) = \frac{\mathbb{E}_{\mathcal{L}} [\mathbf{f}(B) \mathbb{I}\{L(B) \leq M\}]}{\mathbb{E}_{\mathcal{L}} [L(B) \mathbb{I}\{L(B) \leq M\}]},$$

where M is some constant larger than 1. In this case, the influence function is given by

$$\frac{(\mathbf{f}(b) - \tilde{T}_M(\mathcal{L})L(b)) \mathbb{I}\{L(b) \leq M\}}{\mathbb{E}_{\mathcal{L}} [L(B) \mathbb{I}\{L(B) \leq M\}]}$$

and the plug-in estimator becomes

$$\frac{\sum_{i=1}^{l_n-1} \mathbf{f}(B_i) \mathbb{I}\{B_i \leq M\}}{\sum_{i=1}^{l_n-1} L(B_i) \mathbb{I}\{B_i \leq M\}}.$$

This simply consists in getting rid of the blocks (or the pseudo-blocks) whose lengths are too large compared to M .

2.3 Example 2: M-estimators.

Suppose that $E \subset \mathbb{R}$ for simplicity. Let θ be the unique solution of the equation:

$$\mathbb{E}_{\mu} [g(X, \theta)] = 0, \tag{3.5}$$

where $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ belongs to class \mathcal{C}^2 . Equipped with the notation

$$\mathbf{g}(b, \theta) := \sum_{i=1}^{L(b)} g(b_i, \theta) \quad \text{for all } b \in \mathbb{T},$$

the score equation is equivalent to $\mathbb{E}_{\mathcal{L}} [\mathbf{g}(B, \theta)] = 0$. A computation analogous to that carried out in the i.i.d. setting (provided that differentiating inside the expectation is authorized) gives

$$T_g^{(1)}(b, \mathcal{L}) = - \frac{\mathbf{g}(b, \theta)}{\mathbb{E}_{\mathcal{L}} \left[\frac{\partial \mathbf{g}(B, \theta)}{\partial \theta} \right]},$$

where

$$\partial \mathbf{g}(b, \theta) / \partial \theta = \sum_{i=1}^{L(b)} \partial g(b_i, \theta) / \partial \theta.$$

By definition of θ , we naturally have

$$\mathbb{E}_{\mathcal{L}} [T_g^{(1)}(B, \mathcal{L})] = 0.$$

2.4 Example 3: Quantiles.

We place ourselves in the case $E \subset \mathbb{R}$. Assume that the stationary distribution has a continuous c.d.f. $F_\mu(x) = \mu(-\infty, x]$ and density $f_\mu(x)$. Consider the α -quantile $\tilde{T}_\alpha(\mu) = F_\mu^{-1}(\alpha)$. This parameter can also be viewed as a functional of \mathcal{L} , $T_\alpha(\mathcal{L})$ i.e. it is the unique solution of the equation

$$\mathbb{E}_{\mathcal{L}} \left[\sum_{i=1}^{L(b)} \{ \mathbb{I}_{\{b_i \leq \theta\}} - \alpha \} \right] = 0.$$

A straightforward computation following in the footsteps of those carried out in the i.i.d. case (see [124] for further details) shows that, if $f_\mu(\tilde{T}_\alpha(\mu)) \neq 0$, the influence function is given by

$$T_\alpha^{(1)}(b, \mathcal{L}) = \frac{\sum_{i=1}^{L(b)} (\alpha - \mathbb{I}_{\{b_i \leq T_\alpha(\mu)\}})}{\mathbb{E}_{\mathcal{L}}[L(B)]f_\mu(\tilde{T}_\alpha(\mu))}.$$

It follows that the gross-error sensitivity of a quantile in a dependent framework is

$$\gamma^*(T_\alpha(\mu), \mathcal{L}) = \infty.$$

Thus, an empirical quantile is generally not robust in the Markovian framework. Indeed, a possibly large excursion of a Markov chain may arise for instance from an innovative outlier with some distribution which is very different from the stationary distribution. Such an outlier may perturb the behavior of the Markov chain over a possibly very large stretch of observations, making the empirical quantile inaccurate. Just like in the previous example, the length of the blocks (or excursions) must be controlled in order to build robust estimators, even in the case of the quantile.

3 Fréchet differentiability of functionals of Markov chains

As mentioned in the previous chapter, Fréchet differentiability is an important concept in robust statistics as it guarantees the existence of the influence function. Let d be some metric on $\mathcal{P}_{\mathbb{T}}$. In what follows, we define Fréchet differentiability of functionals of blocks of data.

Definition 12. The functional $T : \mathcal{P}_{\mathbb{T}} \rightarrow \mathbb{R}$ is Fréchet differentiable at $\mathcal{L}_A \in \mathcal{P}_{\mathbb{T}}$ for a metric d , if there exists a continuous linear operator $DT_{\mathcal{L}_A}$ (from the set of signed measures of the form $\mathcal{L} - \mathcal{L}_A$ in $(\mathbb{R}, \|\cdot\|)$) and a function

$$\epsilon^{(1)}(\cdot, \mathcal{L}_A) : \mathbb{R} \rightarrow (\mathbb{R}, \|\cdot\|),$$

which is continuous at 0 and $\epsilon^{(1)}(0, \mathcal{L}_A) = 0$ such that

$$\forall \mathcal{L} \in \mathcal{P}_{\mathbb{T}}, \quad T(\mathcal{L}) - T(\mathcal{L}_A) = DT_{\mathcal{L}_A}(\mathcal{L} - \mathcal{L}_A) + R^{(1)}(\mathcal{L}, \mathcal{L}_A),$$

where

$$R^{(1)}(\mathcal{L}, \mathcal{L}_A) = d(\mathcal{L}_A, \mathcal{L})\epsilon^{(1)}(d(\mathcal{L}_A, \mathcal{L}), \mathcal{L}_A).$$

Furthermore, assume that T admits the following representation,

$$\forall \mathcal{L}_A \in \mathcal{P}_{\mathbb{T}}, \quad DT_{\mathcal{L}_A}(\mathcal{L} - \mathcal{L}_A) = \int T^{(1)}(b, \mathcal{L}_A) \mathcal{L}(db),$$

where $T^{(1)}(b, \mathcal{L}_A)$ is the influence function at \mathcal{L}_A .

Fréchet differentiability is a standard tool for obtaining central limit theorems for plug-in estimators. The idea is simply that if we consider the plug-in estimator of the distribution of blocks defined by the empirical distribution of complete blocks

$$\mathcal{L}_n = \frac{\sum_{i=1}^{l_n-1} \delta_{B_i}}{\sum_{i=1}^{l_n-1} l(B_i)},$$

then T may be linearized as

$$\begin{aligned} T(\mathcal{L}_n) - T(\mathcal{L}) &= \int T^{(1)}(b, \mathcal{L}_A) \mathcal{L}_n(db) + r_n \\ &= \frac{\sum_{i=1}^{l_n-1} T^{(1)}(B_i, \mathcal{L}_A)}{\sum_{i=1}^{l_n-1} l(B_i)} + r_n, \end{aligned}$$

where r_n is a remainder (which will be controlled with some specific metric). In particular, it is obvious that if $r_n = o(n^{-1/2})$. Then, by the central limit theorem for independent random variables, we will have

$$n^{1/2}(T(\mathcal{L}_n) - T(\mathcal{L})) \rightarrow N\left(0, \frac{\text{Var}(T^{(1)}(B_i, \mathcal{L}_A))}{E_A \tau_A}\right)$$

as soon as

$$0 < \text{Var}(T^{(1)}(B_i, \mathcal{L}_A)) < \infty.$$

Moreover, when the influence $T^{(1)}(B_i, \mathcal{L}_A)$ is bounded, this will hold automatically without imposing any additional block moment assumptions.

Thus, to obtain Fréchet differentiability of functionals as well as a precise control of the remainder $r_n = o(n^{-1/2})$, a careful choice of metric must be conducted. In the following, we work with a generalization of the Kolmogorov's distance which is defined as follows (see also [18] and [58]).

Definition 13. Let \mathcal{F} be a class of real-valued functions with envelope $F > 0$ (we work with outer measures and the Hoffmann-Jørgensen convergence; see for instance [141] for more details concerning measurability issues). We define metric

$$d_{\mathcal{F}}(\mathcal{L}, \mathcal{L}_A) := \sup_{f \in \mathcal{F}} \left| \int \left(\sum_{i=1}^{l(b)} f(b_i) \right) (\mathcal{L} - \mathcal{L}_A)(db) \right|. \quad (3.6)$$

In what follows, we state a CLT for Fréchet differentiable functionals of blocks which combines arguments from [26] and [44].

Theorem 9. Assume that $T : \mathcal{P}_{\mathbb{T}} \rightarrow \mathbb{R}$ is a Fréchet differentiable functional at \mathcal{L}_A for some metric $d_{\mathcal{F}}$, where \mathcal{F} is a permissible class of functions (see [119]) with an envelope F , satisfying the uniform entropy condition

$$\int_0^\infty \sqrt{\sup_Q \log N_2(\epsilon, Q, \mathcal{F})} d\epsilon < \infty.$$

Assume in addition, in the regenerative atomic case

$$\mathbb{E}_A \left[\sum_{1 \leq j \leq \tau_A} F(X_j) \right]^2 < \infty, \quad \mathbb{E}_\nu(\tau_A) < \infty, \quad \mathbb{E}_A(\tau_A^2) < \infty.$$

Then, we have

$$n^{1/2}(T(\mathcal{L}_n) - T(\mathcal{L}_A)) \rightarrow N \left(0, \frac{\text{Var}(T^{(1)}(B_i, \mathcal{L}_A))}{E_A(\tau_A)} \right).$$

Alternatively, in the general Harris recurrent case with a small set S , assume

$$\sup_{x \in S} \mathbb{E}_x \left[\sum_{1 \leq j \leq \tau_{A, \mathcal{M}}} F(X_j) \right]^2 < \infty, \quad \mathbb{E}_\nu(\tau_{A, \mathcal{M}}) < \infty, \quad \sup_{x \in S} \mathbb{E}_x(\tau_x)^2 < \infty.$$

Then, we have

$$n^{1/2}(T(\widehat{\mathcal{L}}_n) - T(\mathcal{L}_{A, \mathcal{M}})) \rightarrow N \left(0, \frac{\text{Var}(T^{(1)}(B_i, \mathcal{L}_{A, \mathcal{M}}))}{E_{A, \mathcal{M}}(\tau_{A, \mathcal{M}})} \right).$$

Proof. Fréchet differentiability with respect to the metric $d_{\mathcal{F}}$ implies that the influence function $T^{(1)}(b, \mathcal{L}_A)$ is bounded (up to some constant depending on \mathcal{L}_A), by $\sum_{1 \leq j \leq l(b)} F(b_j)$ (see [18]). Assumed moment conditions imply that $\text{Var}(T^{(1)}(B_i, \mathcal{L}_A)) < \infty$ both in the regenerative and general Harris case. The proof follows from the fact that the remainder

$$r_n = d_{\mathcal{F}}(\mathcal{L}_n, \mathcal{L}_A) \epsilon^{(1)}(d_{\mathcal{F}}(\mathcal{L}_n, \mathcal{L}_A), \mathcal{L}_A) = o_P(n^{-1/2}),$$

since by the results of [44] conditions, $n^{1/2}d_{\mathcal{F}}(\mathcal{L}_n, \mathcal{L}_A)$ converges to the supremum of a gaussian process. Similarly, the moment conditions ensure that $d_{\mathcal{F}}(\mathcal{L}_n, \widehat{\mathcal{L}}_n) = o_P(n^{1/2})$ yielding $d_{\mathcal{F}}(\widehat{\mathcal{L}}_n, \mathcal{L}_{A_{\mathcal{M}}}) = O_P(n^{-1/2})$ and the result follows analogously. \square

Remark 9. Observe that with extensions of the concepts of influence function, Fréchet differentiability or Kolmogorov's distance to the blocks of data, one can reformulate Theorems 7 and 8 and state uniform bootstrap central limit theorems for Fréchet differentiable functionals defined on the torus.

4 A Markov view for estimators in PDMPs

In this part, we describe two stochastic models which are particular cases of PDMPs and appear naturally in the risk theory. The first one comes from non-life insurance mathematics and the second one (KDEM model) comes from dietary risk assessment. Note that we described the KDEM model in Chapter 1 in Example 3. However, for the reader's convenience we provide this example again with slightly different notation we introduce in this chapter.

Throughout this section, all the random variables are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Moreover, we assume that

(H1): $(W_i)_{i \in \mathbb{N}}$ are i.i.d. nonnegative r.v.'s with a common mean γ and the c.d.f. F_W .

(H2): $(\Delta T_i)_{i \in \mathbb{N}^*}$ is an i.i.d. sequence of a.s. positive r.v.'s with the c.d.f. H independent of the sequence $(W_i)_{i \in \mathbb{N}}$. We assume

$$\lambda = \mathbb{E}[\Delta T_1] < \infty \quad \text{and} \quad \text{Var}[\Delta T_1] < \infty.$$

(H3): $(T_i)_{i \in \mathbb{N}}$, defined for all $i \geq 1$ by

$$T_i = \sum_{k=1}^i \Delta T_k$$

forms an increasing sequence of r.v.'s. By convention, we set $T_0 = 0$.

(H4): The counting process $\{(N(t))_{t \geq 0}$ defined by

$$N(t) := \# \{i \in \mathbb{N}^* : T_i \leq t\}$$

for $t \geq 0$ is a renewal process and

$$A(t) = t - T_{N(t)}$$

is the backward recurrence time.

4.1 Example: Sparre-Andersen model with barrier

The Sparre-Andersen (SA) model is a generalization of the Cramér-Lundberg (CL) model, also called Poisson Point Process, and has been proposed in the 1950's. It aims to represent the evolution of the reserve of an insurance company. In this context, the W_i 's are the *claims* that the company face, which arise at the *claim arrivals* or *claim instants* T_0, T_1, \dots and the ΔT_i 's are the periods between claims, called *inter-arrivals* or *inter-claims*; see [107] and [62] for a review of ruin models in a non-life insurance framework.

In what follows, we denote by $X(t)$ the reserve of an insurance company at time $t \geq 0$ and $S(t)$ the total claim amount at time t , which is defined by

$$S(t) = \sum_{i=1}^{N(t)} W_i, \quad t \geq 0.$$

Recall that the SA process evolves as follows

$$X(t) = ct + u - S(t), \quad t \geq 0, \quad (3.7)$$

where $c > 1$ is the fixed premium rate and $u := X(0)$ is the initial reserve at time $t = 0$. Denote $X_{T_{N(t)}} = X(T_{N(t)})$ the reserve of the insurance company at the latest claim before t . It is clear that the SA model (3.7) may be rewritten as

$$X(t) = X_{T_{N(t)}} + cA(t), \quad t \geq 0. \quad (3.8)$$

Hereinafter, we propose to introduce a dividend barrier $d > 0$ so that dividends are paid out whenever the surplus level attains the threshold d . Then, the SA model with a dividend barrier is defined by

$$X(t) = \min(d, X_{T_{N(t)}} + cA(t)), \quad t \geq 0. \quad (3.9)$$

A trajectory of the continuous-time process $X \equiv \{X(t)\}_{t \geq 0}$ defined in (3.9) is displayed in Figure 3.1. Observe the PDMP type-behavior: the deterministic motion (between claims) punctuated by discrete random jumps (claims) appearing as a discontinuity in the trajectory and occurring at random epochs (claim instants). Between two claims, the process grows until it reaches the dividend barrier d and stays at this level until the next claim unless the latest arises before attaining the barrier. Note that this model may be seen as an extension of the "Growth-fragmentation" model presented in Section 1.1.3.1 of Chapter 1 in [14].

The analysis of the long-term behavior of the PDMP X boils down to investigating the properties of the embedded Markov chain $\tilde{X} = (X_n)_{n \geq 1}$, which corresponds to the PDMP

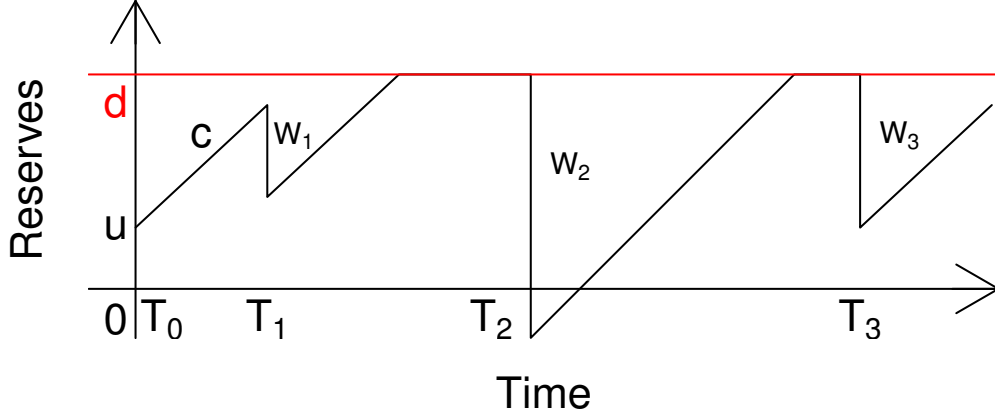


Figure 3.1: Trajectory of the Sparre-Andersen model with a barrier. The horizontal red line corresponds to a dividend barrier.

X evaluated on the claim instants $X_n = X(T_n)$ for all $n \geq 1$. By construction, \tilde{X} satisfies the following autoregressive structure

$$X_{n+1} = (X_n + Z_{n+1})\mathbb{I}_{(X_n + c\Delta T_{n+1} < d)} + (d - W_{n+1})\mathbb{I}_{(X_n + c\Delta T_{n+1} \geq d)}, \quad (3.10)$$

where $(Z_i)_{i \in \mathbb{N}}$ is defined for $i \geq 0$ by

$$Z_i = c\Delta T_i - W_i.$$

It is a Markov chain with an atom at d , at which the chain regenerates by independence of the W_i 's.

In what follows, we assume the usual "net profit condition" $\mathbb{E}[Z_i] > 0$ which ensures that the average reserve of the insurance company is on the positive side: more premium flows into the portfolio than claim sizes flow out in average. Under (H2), the r.v.'s Z_i 's are still i.i.d. since the ΔT_i 's and the W_i 's are two sequences of i.i.d. r.v.'s which are mutually independent.

The limiting behavior of the PDMP X is represented by a stationary probability measure μ that describes the equilibrium state to which the process settles as time goes to infinity.

Then, the stationary distribution describes asymptotic properties of the PDMPs. Nevertheless, μ is hardly tractable in general. One solution is to find a link between μ and $\tilde{\mu}$, the stationary measure of the embedded chain \tilde{X} which is easier to handle. This is the purpose of the following proposition.

Proposition 1 (Stationary measure of the SA with a barrier). *Let (H1)-(H4) hold. Then, $X(t)$ has an absolutely continuous limiting probability distribution μ given by*

$$\mu([-\infty, v]) = \lambda^{-1} \int_{-\infty}^v \int_0^\infty \left(t \wedge \frac{v-x}{c} \right) \tilde{\mu}(dx) H(dt), \quad -\infty < v \leq d, \quad (3.11)$$

where $(a \wedge b)$ is the minimum between a and $b \in \mathbb{R}$.

This proposition will be useful later to robustify estimates of risk indicators such as the ruin probability and the expected shortfall in Section 5.

Proof. Firstly, we consider the reversed PDMP of X , denoted by $Y \equiv \{Y(t)\}_{t \geq 0}$ and defined by

$$Y(t) := d - X(t)$$

for $t \geq 0$. Its embedded chain $\tilde{Y} \equiv (Y_n)_{n \geq 1}$ is defined for $n \geq 0$ by

$$Y_{n+1} = \begin{cases} Y_n - Z_{n+1}, & \text{if } Y_n - c\Delta T_{n+1} > 0, \\ W_{n+1} & \text{otherwise.} \end{cases} \quad (3.12)$$

Denote μ_1 (respectively $\tilde{\mu}_1$) the stationary measure of Y (respectively of \tilde{Y}). Now observe that for $u \geq 0$

$$\frac{1}{t} \int_{s=0}^{s=t} \mathbb{I}_{\{Y(s) > u\}} ds = \frac{1}{t} \sum_{k=1}^{N(t)} \int_{s=T_{k-1}}^{s=T_k} \mathbb{I}_{\{Y(s) > u\}} ds + \frac{1}{t} \int_{s=T_{N(t)}}^{s=t} \mathbb{I}_{\{Y(s) > u\}} ds. \quad (3.13)$$

Besides, for $k \geq 0$, we have

$$\int_{s=T_k}^{s=T_{k+1}} \mathbb{I}_{\{Y(s) > u\}} ds = \mathbb{I}_{\{Y(k) > u\}} \left(\Delta T_{k+1} \wedge \frac{Y_k - u}{c} \right).$$

When $n \rightarrow \infty$, by the strong law of large numbers, it follows that for $u \geq 0$

$$\frac{1}{n} \sum_{k=1}^n \int_{s=T_k}^{s=T_{k+1}} \mathbb{I}_{\{Y(s) > u\}} ds \longrightarrow \int_u^\infty \int_{t=0}^\infty \left(t \wedge \frac{x-u}{c} \right) \tilde{\mu}_1(dx) H(dt).$$

Note that under (H4) and since $\{N(t)\}_{t \geq 0}$ is a renewal process, we have

$$N(t)/t \rightarrow 1/\lambda \quad \text{a.s. as } t \rightarrow \infty.$$

Therefore, the second term in the right-hand side in (5.4) tends to 0 when $t \rightarrow \infty$ and

$$\frac{1}{t} \int_{s=0}^{\infty} \mathbb{I}_{\{Y(s) > u\}} ds \xrightarrow[t \rightarrow \infty]{} \mu_1([u, \infty)), \quad u \geq 0,$$

where μ_1 is given by

$$\mu_1([u, \infty)) = \int_u^{\infty} \int_{t=0}^{\infty} \left(t \wedge \frac{x-u}{c} \right) \tilde{\mu}(dx) H(dt), \quad u \geq 0. \quad (3.14)$$

It follows that for $v \in]-\infty, d[$

$$\mu([-\infty, v]) = \lambda^{-1} \int_{-\infty}^v \int_0^{\infty} \left(t \wedge \frac{v-x}{c} \right) \tilde{\mu}(dx) H(dt),$$

which yields the proof. \square

Remark 10. We could have extended the PDMP (3.9) to the case when the premium rate is not constant. For instance, one may consider that $c \equiv C$ is a non-negative r.v. with c.d.f. F_C . In that case, the stationary measure is given by

$$\mu([-\infty, v]) = \lambda^{-1} \int_{-\infty}^v \int_0^{\infty} \int_0^{\infty} \left(t \wedge \frac{v-x}{c} \right) \tilde{\mu}(dx) H(dt) F_C(dc), \quad \infty < v \leq d.$$

4.2 Example: Kinetic Dietary Exposure Model

The KDEM is a stochastic process that aims at representing the evolution of a contaminant in the human body through time. It has been proposed few years ago in [31].

In a context of dietary risk assessment, the W_i 's correspond to the intakes of contaminated food and occur at *intake instants* T_i 's and we call ΔT_i 's *inter-arrivals* understood as the durations between the $(i-1)$ -th and the i -th intake and $N(t)$ counts the number of intakes that occurred until time $t \geq 0$.

We keep the same notation as in Example 1 except now $X(t)$ is the total body burden of a chemical at the instant $t \geq 0$. Following [31], between two intakes, we consider that the exposure process $X = (X(t))_{t \geq 0}$ evolves deterministically according to the first order differential equation

$$dX(t) = \omega \times X(t) dt \quad (3.15)$$

with a fixed parameter $\omega > 0$ called an *elimination rate*, that describes the metabolism dealing with the chemical elimination.

By solving (3.15), one may straightforwardly see that the exposure process can be written for any $t \geq 0$ as

$$X(t) = X_{T_{N(t)}} \times e^{-\omega A(t)}. \quad (3.16)$$

Note that the bivariate process $\{(X(t), A(t))\}_{t \geq 0}$ is a PDMP.

The embedded chain of X , again denoted $\tilde{X} \equiv (X_n)_{n \in \mathbb{N}}$, which is the process on the intake instants T_0, T_1, \dots plays a leading role in the analysis of X and describes the exposure process immediately after each intake; see Section 5. It is defined by the following stochastic recurrence equation

$$X_{n+1} = X_n \times e^{-\omega \Delta T_{n+1}} + W_{n+1}, \quad n \geq 0. \quad (3.17)$$

Equation (3.17) is an autoregressive process with a random coefficient. Under (H1)-(H4) and the additional assumption

$$\mathbb{E}[\log(\max(1, W_1))] < \infty,$$

the work in [31] has related the continuous-time process X with the embedded chain \tilde{X} . They show that the limiting distribution μ and $\tilde{\mu}$ are linked by the following equation

$$\mu([u, \infty]) = \lambda^{-1} \int_u^\infty \int_0^\infty (t \wedge \omega^{-1} \log(x/u)) \tilde{\mu}(dx) H(dt), \quad u > 0. \quad (3.18)$$

Remark 11. One may allow the elimination parameter $\omega \sim F(d\omega)$ to be random. In this case, the limiting distribution μ and $\tilde{\mu}$ are linked by the following equation

$$\mu([u, \infty]) = \lambda^{-1} \int_0^\infty \int_u^\infty \int_0^\infty (t \wedge \omega^{-1} \log(x/u)) \tilde{\mu}(dx) H(dt) F(d\omega), \quad u > 0.$$

Remark 12. We may also consider a linear elimination. In this case, the study of the PDMP X boils down to Example 1 with barrier at $d = 0$, except that it is reversed. Indeed, with the same notation, the continuous-time process is defined by

$$X(t) = \max(0, X_{T_{N(t)}} - \omega A(t))$$

and the embedded chain is given by

$$X_{n+1} = \begin{cases} X_n - Z_{n+1}, & \text{if } X_n - c\Delta T_{n+1} > 0, \\ W_{n+1} & \text{otherwise.} \end{cases} \quad (3.19)$$

where for $i \geq 0$, $Z_i = c\Delta T_i - W_i$ and ω is the constant elimination parameter.

The stationary distribution of this PDMP is given in equation (3.14) in the proof of Proposition 1.

5 Robustness for risk PDMP models

We suppose that $E \subset \mathbb{R}$ and, for simplicity, place ourselves in the atomic case, where the chain under study possesses a Harris atom A . In this section, we focus on the robustification of estimators of risk indicators of PDMP.

6 Stationary measure

As we have seen in Section 4, in most of the risk models, the stationary distribution μ of the PDMP X -the continuous-time process- is itself a bounded functional of the stationary distribution $\tilde{\mu}$ of its embedded chain; see Proposition 1 for instance. Then it is easy to construct a robust estimator of μ by just plugging the expression of a robust estimator of $\tilde{\mu}$ for which we can most of the time easily get an explicit formulae. This is the purpose of this section.

Recall that if $\tilde{\mu}$ is the stationary measure of the embedded chain \tilde{X} , from the Kac's theorem, it can be written as a functional of the distribution of the blocks:

$$F_{\tilde{\mu}}(y) = \frac{\mathbb{E}_A \left(\sum_{i=1}^{\tau_A} \mathbb{I}_{\{X_i \leq y\}} \right)}{\mathbb{E}_A[\tau_A]}.$$

Let $M > 0$. Consider the robustified version of this c.d.f., which is simply obtained by eliminating too large blocks, and given by

$$\tilde{F}_{\mathcal{L},M}(y) = \frac{\mathbb{E}_A \left[\left(\sum_{i=1}^{\tau_A} \mathbb{I}_{\{X_i \leq y\}} \right) \mathbb{I}_{\{\tau_A \leq M\}} \right]}{\mathbb{E}_A[\tau_A \mathbb{I}_{\{\tau_A \leq M\}}]}.$$

A straightforward computation of the influence function of $\tilde{F}_{\mathcal{L},M}(y)$ leads to the expression

$$\tilde{F}_M^{(1)}(b, y, \mathcal{L}) = \frac{\sum_{i=1}^{L(b)} (\mathbb{I}_{\{b_i \leq y\}} - F_{\mu}(y)) \mathbb{I}_{\{L(b) \leq M\}}}{\mathbb{E}_A[\tau_A \mathbb{I}_{\{\tau_A \leq M\}}]}, \quad \forall b \in \mathbb{T}.$$

From this expression, we deduce that

$$\|\tilde{F}_{\mathcal{L},M} - F_{\mu}\|_{\infty} = \sup_y |F_{\mathcal{L},M}(y) - F_{\mu}(y)| \rightarrow 0$$

a.s. when $M \rightarrow \infty$.

With the norm $\|\cdot\|_{\infty}$, its gross-error sensitivity is bounded by $M/\mathbb{E}_A[\tau_A \mathbb{I}_{\{\tau_A \leq M\}}]$. Notice that $\mathbb{E}_A[\tau_A \mathbb{I}_{\{\tau_A \leq M\}}]$ may also be written as $\mathbb{E}_{\mathcal{L}}[L(B) \mathbb{I}_{\{L(B) \leq M\}}]$ with $B \sim \mathcal{L}$. It follows that the plug-in estimator of this quantity is given by

$$\tilde{F}_{\mathcal{L},M,n}(y) = \frac{\sum_{i=1}^{l_n-1} \sum_{j=\tau_A(i)+1}^{\tau_A(i+1)} \mathbb{I}_{\{X_j \leq y\}} \mathbb{I}_{\{\tau_A(i+1) - \tau_A(i) \leq M\}}}{\sum_{i=1}^{l_n-1} (\tau_A(i+1) - \tau_A(i)) \mathbb{I}_{\{\tau_A(i+1) - \tau_A(i) \leq M\}}}.$$

A robust estimator of μ is obtained by plugging $\tilde{F}_{\mathcal{L},M,n}$ as it is illustrated in the following two examples.

Application to the Sparre-Andersen model with a barrier

We go back to Subsection 4.1 and recall that the stationary distribution μ of the Sparre-Andersen process with a barrier d is given by

$$\mu([-\infty, v]) = \lambda^{-1} \int_{-\infty}^v \int_0^\infty \left(t \wedge \frac{v-x}{c} \right) \tilde{\mu}(dx) H(dt), \quad -\infty < v \leq d,$$

which may be also rewritten as a functional of the blocks :

$$\begin{aligned} \mu([-\infty, v]) &= \lambda^{-1} \int_{-\infty}^v \int_0^\infty \left(t \wedge \frac{v-x}{c} \right) \frac{\mathbb{E}_A [\sum_{i=1}^{\tau_A} \delta_{X_i}(dx)]}{\mathbb{E}_A[\tau_A]} H(dt) \\ &= \frac{1}{\lambda \mathbb{E}_A[\tau_A]} E_A \left(\sum_{i=1}^{\tau_A} \int_{-\infty}^\infty \mathbb{I}_{\{x \leq v\}} \int_0^\infty \left(t \wedge \frac{v-x}{c} \right) \delta_{X_i}(dx) H(dt) \right) \\ &= \frac{1}{\lambda \mathbb{E}_A[\tau_A]} E_A \left(\sum_{i=1}^{\tau_A} \mathbb{I}_{\{X_i \leq v\}} \int_0^\infty \left(t \wedge \frac{v-X_i}{c} \right) H(dt) \right). \end{aligned}$$

Its robustified version is given by

$$\begin{aligned} \mu([-\infty, v]) &= \lambda^{-1} \int_{-\infty}^v \int_0^\infty \left(t \wedge \frac{v-x}{c} \right) \frac{\mathbb{E}_A [(\sum_{i=1}^{\tau_A} \delta_{X_i}(dx)) \mathbb{I}_{\{\tau_A \leq M\}}]}{\mathbb{E}_A[\tau_A \mathbb{I}_{\{\tau_A \leq M\}}]} H(dt) \\ &= \frac{1}{\lambda \mathbb{E}_A[\tau_A \mathbb{I}_{\{\tau_A \leq M\}}]} \mathbb{E}_A \left(\sum_{i=1}^{\tau_A} \mathbb{I}_{\{X_i \leq v\}} \int_0^\infty \left(t \wedge \frac{v-X_i}{c} \right) \mathbb{I}_{\{\tau_A \leq M\}} H(dt) \right) \end{aligned}$$

and can be estimated by the robust plug-in estimator

$$\begin{aligned} \hat{\mu}_n([-\infty, v]) &= \lambda^{-1} \int_{-\infty}^v \int_0^\infty \left(t \wedge \frac{v-x}{c} \right) \tilde{F}_{\mathcal{L},M,n}(dx) H(dt) \\ &= \lambda^{-1} \int_{-\infty}^v \int_0^\infty \left(t \wedge \frac{v-x}{c} \right) \frac{\sum_{i=1}^{l_n-1} \sum_{j=\tau_A(i)+1}^{\tau_A(i+1)} \delta_{X_j}(dx) \mathbb{I}_{\{\tau_A(i+1)-\tau_A(i) \leq M\}}}{\sum_{i=1}^{l_n-1} (\tau_A(i+1) - \tau_A(i)) \mathbb{I}_{\{\tau_A(i+1)-\tau_A(i) \leq M\}}} H(dt), \\ &= \frac{\sum_{i=1}^{l_n-1} \sum_{j=\tau_A(i)+1}^{\tau_A(i+1)} \mathbb{I}_{\{\tau_A(i+1)-\tau_A(i) \leq M\}} \mathbb{I}_{\{X_j \leq v\}} \int_0^\infty \left(t \wedge \frac{v-X_j}{c} \right) H(dt)}{\lambda \sum_{i=1}^{l_n-1} (\tau_A(i+1) - \tau_A(i)) \mathbb{I}_{\{\tau_A(i+1)-\tau_A(i) \leq M\}}}. \end{aligned}$$

Notice that in these expressions we have

$$\int_0^\infty \left(t \wedge \frac{v-X_j}{c} \right) H(dt) = \int_0^{\frac{v-X_j}{c}} t H(dt) + \frac{v-X_j}{c} \int_{\frac{v-X_j}{c}}^\infty H(dt).$$

In particular, if we consider the Cramér-Lundberg model with an exponential inter-arrival then we get an explicit estimator by using the fact that

$$\begin{aligned} \int_0^\infty \left(t \wedge \frac{v - X_j}{c} \right) H(dt) &= -\frac{(v - X_j)}{c} \exp \left(-\frac{(v - X_j)}{c\lambda} \right) \\ &\quad + \lambda - \lambda \exp \left(-\frac{(v - X_j)}{\lambda c} \right) \\ &\quad + \left(\frac{(v - X_j)}{c} \right) \exp \left(-\frac{(v - X_j)}{c\lambda} \right) \\ &= \lambda \left(1 - \exp \left(-\frac{(v - X_j)}{\lambda c} \right) \right). \end{aligned}$$

It follows that the estimator is essentially a mean of $\lambda \left(1 - \exp \left(-\frac{(v - X_j)}{\lambda c} \right) \right)$ over the X'_i 's lower than ν which belongs to blocks with length smaller than M that is

$$\hat{\mu}_n([-\infty, v]) = \frac{\sum_{i=1}^{l_n-1} \sum_{j=\tau_A(i)+1}^{\tau_A(i+1)} \mathbb{I}_{\{\tau_A(i+1) - \tau_A(i) \leq M\}} \mathbb{I}_{\{X_i \leq v\}} \left(1 - \exp \left(-\frac{(v - X_j)}{\lambda c} \right) \right)}{\sum_{i=1}^{l_n-1} (\tau_A(i+1) - \tau_A(i)) \mathbb{I}_{\{\tau_A(i+1) - \tau_A(i) \leq M\}}}.$$

Notice that, in that case, the plug-in (non robust) estimator of the stationary measure is simply given by

$$\mu_n([-\infty, v]) = 1 - n^{-1} \sum_{i=1}^n \exp \left(-\frac{(v - X_i)}{\lambda c} \right) \mathbb{I}_{\{X_i \leq v\}}.$$

It is clear by a straightforward computation that this estimator is not robust, due to the presence of a large contaminated block.

Application to the Kinetic Dietary Exposure Model

Similarly, for the KDEM model, using Kac's representation, we have the expression of the stationary measure of the continuous process given by

$$\mu([u, \infty]) = \lambda^{-1} \int_u^\infty \int_0^\infty (t \wedge \omega^{-1} \log(x/u)) \frac{\mathbb{E}_A(\sum_{i=1}^{\tau_A} \delta_{X_i}(dx))}{\mathbb{E}_A[\tau_A]} H(dt), \quad u > 0,$$

and the robust estimator is thus given by

$$\begin{aligned} \hat{\mu}_n([u, \infty]) &= \lambda^{-1} \int_u^\infty \int_0^\infty (t \wedge \omega^{-1} \log(x/u)) \tilde{F}_{\mathcal{L}, M, n}(dx) H(dt) \\ &= \frac{\sum_{i=1}^{l_n-1} \sum_{j=\tau_A(i)+1}^{\tau_A(i+1)} \mathbb{I}_{\{\tau_A(i+1) - \tau_A(i) \leq M\}} \int_u^\infty \int_0^\infty (t \wedge \omega^{-1} \log(x/u)) \delta_{X_j}(dx) H(dt)}{\lambda \sum_{i=1}^{l_n-1} (\tau_A(i+1) - \tau_A(i)) \mathbb{I}_{\{\tau_A(i+1) - \tau_A(i) \leq M\}}} \end{aligned}$$

$$= \frac{\sum_{i=1}^{l_n-1} \sum_{j=\tau_A(i)+1}^{\tau_A(i+1)} \mathbb{I}_{\{\tau_A(i+1)-\tau_A(i) \leq M\}} \mathbb{I}_{\{X_j \geq u\}} \int_0^\infty (t \wedge \omega^{-1} \log(X_j/u)) H(dt)}{\lambda \sum_{i=1}^{l_n-1} (\tau_A(i+1) - \tau_A(i)) \mathbb{I}_{\{\tau_A(i+1)-\tau_A(i) \leq M\}}}.$$

Similarly to the Sparre-Andersen case, in the exponential inter-arrival case, we have the expression

$$\begin{aligned} \int_0^\infty (t \wedge \omega^{-1} \log(X_j/u)) H(dt) &= \int_0^{\omega^{-1} \log(X_j/u)} t \lambda^{-1} \exp(-\lambda^{-1} t) dt \\ &\quad + \omega^{-1} \log(X_j/u) \int_{\omega^{-1} \log(X_j/u)}^\infty \lambda^{-1} \exp(-\lambda^{-1} t) dt \\ &= \lambda \left(1 - \exp\left(-\frac{\log(X_j/u)}{\omega \lambda}\right) \right) \\ &= \lambda \left(1 - (X_j/u)^{-1/(\omega \lambda)} \right). \end{aligned}$$

Notice that in that case, the (non robust) plug-in estimator of $\mu([u, \infty])$ is of the form

$$\mu_n([u, \infty]) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{X_i \geq u\}} \left(1 - (X_i/u)^{-1/(\omega \lambda)} \right).$$

The robust estimator is simply the version of its mean only over the X_i 's which do not belong to large blocks, i.e.

$$\hat{\mu}_n([u, \infty]) = \frac{\sum_{i=1}^{l_n-1} \sum_{j=\tau_A(i)+1}^{\tau_A(i+1)} \mathbb{I}_{\{\tau_A(i+1)-\tau_A(i) \leq M\}} \mathbb{I}_{\{X_j \geq u\}} \left(1 - (X_j/u)^{-1/(\omega \lambda)} \right)}{\sum_{i=1}^{l_n-1} (\tau_A(i+1) - \tau_A(i)) \mathbb{I}_{\{\tau_A(i+1)-\tau_A(i) \leq M\}}}.$$

6.1 Ruin probability

In many PDMP models, especially when modelling reserves in insurance or level of contaminants in the body in pharmacokinetics or more generally in risk theory, one is interested not only on the stationary measure but also on the ruin probability, that is the probability that the maximum - or minimum, depending on the application field- of the process over a given time window exceeds a threshold $d \in \mathbb{R}$. In non-life insurance mathematics for instance, the ruin probability plays a leading role to estimate the risk an insurance company is exposed at; see [8] for an exhaustive review. In this subsection we focus on this type of models under the assumption that the deterministic part of the PDMP X is monotone. This is in particular the case for treated in Subsections 4.1 and 4.2.

Consider a PDMP $X \equiv \{X(s)\}_{s \in \mathbb{R}^+}$ which is decreasing between each jumps (if it is increasing consider the minimum instead of the maximum) with embedded chain $\tilde{X} = (X_i)_{i \in \mathbb{N}}$.

Under the assumption, that the deterministic part is monotone, the maxima of X are necessarily reached on the embedded chain \tilde{X} . Then the probability of ruin denoted by Ψ for a threshold $d \in \mathbb{R}$ over an interval $[0, T]$, $T > 0$ is defined by

$$\Psi(d, T) = \mathbb{P} \left(\sup_{s \in [0, T]} X(s) > d \right) = \mathbb{P} \left(\max_{0 \leq i \leq N(T)} X_i > d \right),$$

where $N(T) := \text{card}\{i \geq 0 : T_i \leq T\}$.

We assume that $N(T)$ defines a renewal process (see Chapter 1 in [14] for further details on renewal processes) so that

$$\frac{N(T)}{T} \rightarrow 1/\lambda \quad a.s.$$

when $T \rightarrow \infty$. The jump instants T'_i s are such that the sequence

$$\Delta T_{i+1} = T_{i+1} - T_i$$

are independent with $\mathbb{E}[\Delta T_{i+1}] = \lambda$. It is then clear in that case that the study of the ruin probability reduces to the study of the submaxima. The purpose of this subsection is to show how it is possible with the preceding tools to obtain a robust estimator of the ruin probability with our tools.

For this, consider the case when X possesses a known accessible atom A . For instance in Subsection 4.1, the atom is given by value d of the barrier. In Subsection 4.2, the atom can be constructed via the Nummelin splitting technique by taking as a small set any interval $[0, \varepsilon]$ for some small $\varepsilon > 0$ (provided that the observed chain visits this set a large number of times); see Section 8 for some numerical examples.

For $j \geq 1$, define the *submaximum* over the j -th cycle of the sample path as

$$\zeta_j = \max_{1+\tau_A(j) \leq i \leq \tau_A(j+1)} X_i. \quad (3.20)$$

We use the usual convention that $\max \emptyset = -\infty$. Note that

- $\zeta_0 = \max_{1 \leq i \leq \tau_A} X_i$ has a distribution which depends on the distribution of initial value $X_0 \sim \nu$.
- $\zeta_{l_n}^{(n)} = \max_{1+\tau_A(l_n) \leq i \leq n} X_i$, then the maximum over the last non-regenerative block is based on a data block which may be an incomplete block.

For $n \geq 0$, define

$$M_n := \max_{1 \leq i \leq n} X_i.$$

Then we have

$$M_n = \max \left\{ \zeta_0, \max_{1 \leq j \leq l_n-1} \zeta_j, \zeta_{l_n}^{(n)} \right\}. \quad (3.21)$$

Since the blocks are independent, the ζ_j 's are i.i.d. random variables with common distribution function

$$G_A(x) = \mathbb{P}_A(\max_{1 \leq i \leq \tau_A} X_i \leq x).$$

Moreover, by Harris recurrence property, the number of blocks is of order

$$l_n \sim n/E_A \tau_A \quad \mathbb{P}_\nu\text{-a.s. as } n \rightarrow \infty.$$

Thus, M_n behaves like the maximum of $n/E_A \tau_A$ i.i.d. r.v.'s.

Similarly, since $N(T)$ is a renewal process, we expect to have $N(T) \sim T/\lambda$ so that we expect $M_{N(T)}$ to behave like the maximum of $\frac{T}{\lambda E_A \tau_A}$ independent submaxima of distribution G_A . The following result shows that the limiting distribution of the sample maximum of X is entirely determined by the tail behavior of G_A and relies on this crucial asymptotic independence of the blocks.

Proposition 2. *Assume that the first block does not affect the extremal behavior, i.e.*

$$\mathbb{P}_\nu \left(\zeta_0 > \max_{1 \leq k \leq l} \zeta_k \right) \rightarrow 0 \text{ as } l \rightarrow \infty. \quad (3.22)$$

Then, we have

$$\sup_{x \in \mathbb{R}} |\mathbb{P}_\nu(M_n \leq x) - G_A(x)^{n/E_A \tau_A}| \rightarrow 0 \text{ as } n \rightarrow \infty. \quad (3.23)$$

Moreover, the survival distribution of the ruin probability converges uniformly to $G_A(d)^{T/(\lambda E_A \tau_A)}$. Precisely,

$$\sup_{d \in \mathbb{R}} |1 - \Psi(d, T) - G_A(d)^{T/(\lambda E_A \tau_A)}| \rightarrow 0 \text{ as } T \rightarrow \infty. \quad (3.24)$$

Proof. The first result (3.23) has been proved in [125]. The second result (3.24) is an adaptation of Theorem 3.1 in [125] with $n = N(T)$ (see also [32]). Recalling

$$N(T)/T \rightarrow 1/\lambda \quad \text{a.s.},$$

similar arguments hold. □

As a consequence, the limiting behavior of the maximum of these PDMPs may be deduced by using Fischer-Typett-Gnedenko theorem (see [123], Chapter 1) with the marginal distribution replaced by the marginal distribution of a regenerative block, namely

$$F_A(x) := G_A(x)^{1/\mathbb{E}_A \tau_A}.$$

Then, the asymptotic behavior of the sample maximum is entirely determined by the tail properties of F_A . In particular, the limiting distribution of M_n (for a suitable normalization) that is the distribution of the maximum when observing n jumps, or $\Psi(d, T)$ correctly normalized is the generalized extreme value distribution; see [30] for more details.

To simplify the notation, we assume that n jumps have been observed if the process is observed only during a period $[0, T]$. Simply replace n by T/λ to obtain similar results.

Just as before,

$$G_A(x) = \mathbb{P}_A \left(\max_{1 \leq i \leq \tau_A} X_i \leq x \right) = \mathbb{E}_{\mathcal{L}_A} \left[\mathbb{I}_{\{\max_{1 \leq i \leq \tau_A} X_i \leq x\}} \right]$$

is a functional of the distribution of the block with the influence function given by

$$G_A^{(1)}(b, x) = \mathbb{I}_{\{\max_{1 \leq i \leq l(b)} b_i \leq x\}} - G_A(x)$$

which is a bounded function for each block. It follows that, in the atomic case, a robust plug-in estimator of G_A is given by its empirical counterpart, from the observation of a random number $l_n - 1$ of complete regenerative cycles, namely

$$G_{A,n}(x) = \frac{1}{l_n - 1} \sum_{j=1}^{l_n-1} \mathbb{I}_{\{\zeta_j(f) \leq x\}} \quad (3.25)$$

with $G_{f,n} \equiv 0$ by the convention when $l_n \leq 1$.

Observe that we have dropped the first and the last (non-regenerative blocks) because they are independent but with a different distribution. Actually, thanks to the robustness of the estimator, we could have included them with no asymptotic changes.

Applying Glivenko-Cantelli's theorem it follows that

$$\Delta_n = \sup_{x \in \mathbb{R}} |G_{A,n}(x) - G_A(x)| \rightarrow 0, \quad \mathbb{P}_\nu \text{ a.s..}$$

Moreover, by the law of iterated logarithm, we also have

$$\Delta_n = O(\sqrt{\log \log(n)/n}) \text{ a.s..}$$

Notice, however, that the real quantity of interest in this setting is rather the phantom distribution

$$F_A(x) = G_A(x)^{1/E_A \tau_A}.$$

A Markovian influence function is given by

$$\begin{aligned} F_A^{(1)}(b, x) &= \frac{\partial}{\partial t} \left(\left(\mathbb{E}_{(1-t)\mathcal{L}_A + tb}(\mathbb{I}_{\{\max_{1 \leq i \leq \tau_A} X_i \leq x\}}) \right)^{1/((1-t)\mathbb{E}_A \tau_A + t l(b))} \right) \\ &= F_A(x) \left(\frac{\left(\mathbb{I}_{\{\max_{1 \leq i \leq l(b)} b_i \leq x\}} - G_A(x) \right)}{\mathbb{E}_A \tau_A G_A(x)} - \frac{\log(G_A(x)(l(b) - \mathbb{E}_A \tau_A))}{(\mathbb{E}_A \tau_A)^2} \right). \end{aligned}$$

Observe that this influence function strongly depends on the length of the blocks (which was not the case for the robust empirical counterpart estimator $G_{A,n}(x)$ of the distribution function $G_A(x)$) and thus is non robust. A robustified version of the phantom distribution can thus be obtained by excluding blocks of large size, considering for instance

$$F_{n,A}^M(x) = \left(\frac{1}{l_n - 1} \sum_{j=1}^{l_n-1} \mathbb{I}\{\zeta_j \leq x\} \right)^{\frac{\sum_{i=1}^{l_n-1} 1_{\{l(B_i) \leq M\}}}{\sum_{i=1}^{l_n-1} l(B_i) 1_{\{l(B_i) \leq M\}}}}.$$

The following theorem, which is an adaptation of [30] shows that if we observe the process over an interval $[0, T]$, then typically the number of observations is

$$n = N(T) = O(T/\lambda)$$

so that we can predict (with the plug-in rule) the ruin probability over an interval $[0, T^*]$, such that T^* satisfies

$$T^* = o \left(\frac{T^{1/2}}{\log \log(T)^{1/2}} \right). \quad (3.26)$$

Theorem 10 (Robust estimator for the ruin probability). *Let X be a PDMP with a decreasing deterministic motion and Harris recurrent embedded chain with the atom A . Assume that N is a renewal process with a constant intensity function λ and observe the process on an interval $[0, T]$. Then, considering T^* as in (3.26), we have*

$$\lim_{M \rightarrow \infty} \sup_{d \in \mathbb{R}} |1 - \Psi(d, T^*) - F_{N(T),A}^M(d)^{T^*/\lambda}| \rightarrow 0 \quad a.s. \quad T \rightarrow \infty.$$

6.2 Extremal index

We turn our attention to the extremal index (see [32] and [33] for details how to estimate extremal index when the data are Harris Markov chain), and keep the same notation as in

the preceding section. The inverse of this index is an indicator of the average size of the cluster of the extreme values of the embedded chain so that $\frac{\lambda}{\theta}$ can be interpreted as the average duration of the PDMP over a high threshold.

Again we assume that the embedded chain $(X_i)_{i \in \mathbb{N}}$ of the PDMP $\{X(s)\}_{s \in \mathbb{R}^+}$ is Harris recurrent. Without loss of the generality we assume that $(X_i)_{i \in \mathbb{N}}$ possesses an atom A and denote by $\tilde{\mu}$ its stationary measure. It has been shown (see [125] or [96]) that there exists some index $\theta \in [0, 1]$, called the *extremal index* of the sequence $(X_i)_{i \in \mathbb{N}}$ such that

$$\mathbb{P}_{\tilde{\mu}} \left(M_n = \max_{0 \leq i \leq n} (X_i) \leq u_n \right) \underset{n \rightarrow \infty}{\sim} F_{\tilde{\mu}}(u_n)^{n\theta}, \quad (3.27)$$

for any sequence $u_n = u_n(\eta)$ such that $n(1 - F_{\tilde{\mu}}(u_n)) \rightarrow \eta$ when $n \rightarrow \infty$, see also [117].

Now observe that using a straightforward Taylor expansion, we have

$$\begin{aligned} \theta &= \lim_{n \rightarrow \infty} \frac{\log(G_A(u_n))}{\mathbb{E}_A \tau_A \log(F_{\tilde{\mu}}(u_n))} \\ &= \lim_{n \rightarrow \infty} \frac{\log(1 - \overline{G}_A(u_n))}{\mathbb{E}_A \tau_A \log(1 - \overline{F}_{\tilde{\mu}}(u_n))} \\ &= \lim_{n \rightarrow \infty} \frac{\overline{G}_A(u_n)}{\mathbb{E}_A \tau_A \overline{F}_{\tilde{\mu}}(u_n)}. \end{aligned}$$

Define for some fixed level u

$$\theta(u) = \frac{\overline{G}_A(u)}{\mathbb{E}_A \tau_A \overline{F}_{\tilde{\mu}}(u)} = \frac{\mathbb{E}_A \left(\mathbb{I}_{\{\max_{1 \leq i \leq \tau_A} X_i \geq x\}} \right)}{\mathbb{E}_A \left(\sum_{i=1}^{\tau_A} \mathbb{I}_{\{X_i \geq x\}} \right)}.$$

Its influence function is given by

$$\theta^{(1)}(b, u) = \frac{\mathbb{I}_{\{\max_{1 \leq i \leq l(b)} b_i \geq x\}} - \overline{G}_A(u)}{\mathbb{E}_A \tau_A \overline{F}_{\tilde{\mu}}(u)} - \frac{\overline{G}_A(u) \left(\sum_{i=1}^{l(b)} \mathbb{I}_{\{X_i \geq x\}} - \mathbb{E}_A \tau_A \overline{F}_{\tilde{\mu}}(u) \right)}{(\mathbb{E}_A \tau_A \overline{F}_{\tilde{\mu}}(u))^2}.$$

Notice again that because of the presence of $F_{\tilde{\mu}}$, the denominator of this influence function is not bounded due to the length of the blocks.

In [30] it is proposed to estimate θ by its empirical counterpart defined by

$$\theta_n(u) = \frac{\sum_{j=1}^{l_n-1} \mathbb{I}_{\{\zeta_j > u\}}}{\sum_{i=1}^n \mathbb{I}_{\{X_i > u\}}},$$

with the convention that $\theta_n(u) = 0$ if $M_n < u$.

However, if we want a robust estimator of this quantity, it is more appropriate to eliminate large blocks in the denominator, leading to the robust estimator

$$\theta_n(u, M) = \frac{\sum_{j=1}^{l_n-1} \mathbb{I}_{\{\zeta_j > u\}}}{\sum_i^{l_n-1} \sum_{j=\tau_A(i)+1}^{\tau_A(i+1)} \mathbb{I}_{\{X_j \leq y\}} \mathbb{I}_{\{\tau_A(i+1) - \tau_A(i) \leq M\}}}$$

with the convention that $\theta_n(u, M) = 0$ if $M_n < u$. Finally, we have the following result.

Theorem 11. *Under the assumptions of Theorem 10, let $(r_n)_{n \in \mathbb{N}}$ be increasing to infinity in a way that*

$$r_n = o(\sqrt{n/\log \log n}) \quad \text{as } n \rightarrow \infty.$$

Consider $(v_n)_{n \in \mathbb{N}}$ such that

$$r_n(1 - G_A(v_n)) \rightarrow \eta < \infty \quad \text{as } n \rightarrow \infty.$$

Suppose that $\mathbb{E}_v \tau_A < \infty$ and $\mathbb{E}_A \tau_A^2 < \infty$ are fulfilled. Then,

$$\theta_n(v_n, M) \rightarrow \theta \quad \mathbb{P}_\nu\text{-a.s.}, \quad \text{as } n \rightarrow \infty \text{ and } M \rightarrow \infty. \quad (3.28)$$

Moreover, we have

$$\sqrt{n/r_n} (\theta_n(v_n, M) - \theta(v_n, M)) \Rightarrow \mathcal{N}(0, \theta^2/\eta), \quad \text{as } n \rightarrow \infty \text{ and } M \rightarrow \infty, \quad (3.29)$$

where

$$\theta(u, M) = \frac{\mathbb{E}_A(\mathbb{I}\{\max_{1 \leq i \leq \tau_A} X_i \geq x\})}{\mathbb{E}_A((\sum_{i=1}^{\tau_A} \mathbb{I}\{X_i \geq x\})\mathbb{I}\{\tau_A \leq M\})}.$$

Notice that the recentering for the central limit theorem is not completely satisfactory since it depends both on the threshold M and the level u . We would be more interested in centering by θ . As discussed in [30], this can be done by controlling the bias $\sqrt{n/r_n}(\theta(v_n, M) - \theta)$ with some higher second order technical conditions.

From a practical point of view, the choices of the thresholds M and ν_n are obtained by plotting the values of $\theta(v_n, M)$ and by detecting an area of stability of the estimator; see the discussion on $\theta(v_n)$ in [30]. Observe that v_n is defined as an upper quantile of the true underlying submaximum distribution, which is unknown in practice. An empirical choice can also be obtained by taking r_n equal to the quantile $G_{A,n}^{-1}(1 - \eta/r_n)$ which is automatically robust. Because of the condition

$$r_n = o(\sqrt{n/\log \log n}),$$

notice that the best attainable rate with our method is close to $n^{1/4}$.

Similar results are obtained in the general non-regenerative case in [30], with an adequate choice of the level ν_n depending on the rate of convergence of the estimator of the transition density estimator.

7 Expected shortfall

The ruin probability and the extremal index studied in Subsections 6.1 and 6.2 provide information about the probability that the PDMP goes above a threshold or about the replication of extremal events of the PDMP, but give no information on the exceedences themselves. To fill this gap, we finally focus on the *expected shortfall*, which takes into account the mean of the PDMP when the process has already reached the reference threshold d . It is in particular a Gerber-Shiu measure; see [92] for more details.

Let X be a PDMP with stationary measure μ . The *expected shortfall* of X is defined, for $u \in \mathbb{R}$, by

$$S_{\mathbb{E}}(u, \mu) = \mathbb{E}_{\mu}[X - u \mid X > u] = \int_{x=u}^{\infty} (x - u) \mu(dx) / \mu([u, \infty[).$$

Since the expected shortfall is defined as a functional of the probability function of the stationary measure, using the Kac's representation, we can express it as a functional of the distribution of the blocks, i.e.

$$S_{\mathbb{E}}(u, \mathcal{L}_A) = \frac{\int_{x=u}^{\infty} (x - u) \mathbb{E}_A(\sum_{i=1}^{\tau_A} \delta_{X_i}(dx))}{\mathbb{E}_A(\sum_{i=1}^{\tau_A} \mathbb{I}_{\{X_i \geq u\}})}.$$

The influence function of this quantity is defined by

$$\begin{aligned} S_{\mathbb{E}}^{(1)}(b, u, \mathcal{L}_A) &= \frac{\partial}{\partial t} \left[\frac{(1-t) \int_{x=u}^{\infty} (x - u) \mathbb{E}_A(\sum_{i=1}^{\tau_A} \delta_{X_i}(dx)) + t \sum_{i=1}^{l(b)} (b_i - u)}{(1-t) \mathbb{E}_A(\sum_{i=1}^{\tau_A} \mathbb{I}_{\{X_i \geq u\}}) + t \sum_{i=1}^{l(b)} \mathbb{I}_{\{b_i \geq u\}}} \right]_{t=0} \\ &= \frac{\sum_{i=1}^{l(b)} (b_i - u) - \int_{x=u}^{\infty} (x - u) \mathbb{E}_A(\sum_{i=1}^{\tau_A} \delta_{X_i}(dx))}{\mathbb{E}_A(\sum_{i=1}^{\tau_A} \mathbb{I}_{\{X_i \geq u\}})} \\ &\quad - \frac{\int_{x=u}^{\infty} (x - u) \mathbb{E}_A(\sum_{i=1}^{\tau_A} \delta_{X_i}(dx)) \left(\sum_{i=1}^{l(b)} \mathbb{I}_{\{b_i \geq u\}} - \mathbb{E}_A(\sum_{i=1}^{\tau_A} \mathbb{I}_{\{X_i \geq u\}}) \right)}{(\mathbb{E}_A(\sum_{i=1}^{\tau_A} \mathbb{I}_{\{X_i \geq u\}}))^2}. \end{aligned}$$

Again this influence function is not robust and a robust estimator is simply obtained by truncating the blocks with large value and large size.

Consider for this the robustified functional

$$S_{\mathbb{E}}(u, \mathcal{L}_A, M_1, M_2) = \frac{\int_0^{\infty} (x - u) \mathbb{I}_{\{0 < x - u \leq M_1\}} \mathbb{E}_A(\sum_{i=1}^{\tau_A} \delta_{X_i}(dx) \mathbb{I}_{\{\tau_A \leq M_2\}})}{\mathbb{E}_A(\sum_{i=1}^{\tau_A} \mathbb{I}_{\{X_i \geq u\}} \mathbb{I}_{\{\tau_A \leq M_2\}})}.$$

It is easy to see that

$$S_{\mathbb{E}}(u, \mathcal{L}_A, M_1, M_2) \rightarrow S_{\mathbb{E}}(u, \mathcal{L}_A) \quad \text{when } M_1, M_2 \rightarrow \infty.$$

Now its influence function is given by

$$\begin{aligned}
& S_{\mathbb{E}}^{(1)}(b, u, \mathcal{L}_A, M_1, M_2) \\
&= \frac{(\sum_{i=1}^{l(b)} (b_i - u) \mathbb{I}_{\{0 < b_i - u \leq M_1\}} \mathbb{I}_{\{l(b) \leq M_2\}})}{\mathbb{E}_A \left(\sum_{i=1}^{\tau_A} \mathbb{I}_{\{X_i \geq u\}} \mathbb{I}_{\{\tau_A \leq M_2\}} \right)} \\
&- \frac{\int_0^\infty (x - u) \mathbb{I}_{\{0 < x - u \leq M_1\}} \mathbb{E}_A \left(\sum_{i=1}^{\tau_A} \delta_{X_i}(dx) \mathbb{I}_{\{\tau_A \leq M_2\}} \right)}{\mathbb{E}_A \left(\sum_{i=1}^{\tau_A} \mathbb{I}_{\{X_i \geq u\}} \mathbb{I}_{\{\tau_A \leq M_2\}} \right)} \\
&- \int_0^\infty (x - u) \mathbb{I}_{\{0 < x - u \leq M_1\}} \mathbb{E}_A \left(\sum_{i=1}^{\tau_A} \delta_{X_i}(dx) \mathbb{I}_{\{\tau_A \leq M_2\}} \right) \times \\
&\left[\frac{(\sum_{i=1}^{l(b)} \mathbb{I}_{\{b_i \geq u\}} \mathbb{I}_{\{l(b) \leq M_2\}} - \mathbb{E}_A \left(\sum_{i=1}^{\tau_A} \mathbb{I}_{\{X_i \geq u\}} \mathbb{I}_{\{\tau_A \leq M_2\}} \right))}{\left(\mathbb{E}_A \left(\sum_{i=1}^{\tau_A} \mathbb{I}_{\{X_i \geq u\}} \mathbb{I}_{\{\tau_A \leq M_2\}} \right) \right)^2} \right. \\
&\left. - \frac{\mathbb{E}_A \left(\sum_{i=1}^{\tau_A} \mathbb{I}_{\{X_i \geq u\}} \mathbb{I}_{\{\tau_A \leq M_2\}} \right)}{\left(\mathbb{E}_A \left(\sum_{i=1}^{\tau_A} \mathbb{I}_{\{X_i \geq u\}} \mathbb{I}_{\{\tau_A \leq M_2\}} \right) \right)^2} \right]
\end{aligned}$$

which is rather complicated, but obviously bounded.

As a consequence, the plug-in estimator is simply obtained by plugging the distribution

$$\frac{\sum_{i=1}^{l_n-1} \mathbf{f}(B_i) \mathbb{I}_{\{L(B_i) \leq M_2\}}}{\sum_{i=1}^{l_n-1} L(B_i) \mathbb{I}_{\{L(B_i) \leq M_2\}}}$$

in the functional $S_{\mathbb{E}}(u, \mathcal{L}_A, M_1, M_2)$.

Proposition 3 (Expected shortfall of PDMPs). *A robust estimator of the expected shortfall is given by*

$$\widehat{S}_{\mathbb{E},n}(u, \mathcal{L}_A, M_1, M_2) = \frac{\sum_{i=1}^{l_n-1} \sum_{\tau_A(i)+1}^{\tau_A(i+1)} (X_i - u) \mathbb{I}_{\{0 < \sum_{\tau_A(i)+1}^{\tau_A(i+1)} (X_i - u) \leq M_1\}} \mathbb{I}_{\{L(B_i) \leq M_2\}}}{\sum_{i=1}^{l_n-1} L(B_i) \mathbb{I}_{\{L(B_i) \leq M_2\}}}.$$

8 Simulations

In this section we present simulation results on the standard Cramér-Lundberg model with a barrier. We have chosen to simulate this model because of its simplicity and also to highlight some practical problems and difficulties that can be encountered in practice.

We have chosen $c = 1, d = 3$, ΔT_n i.i.d. with exponential distribution with a parameter 0.6 (with mean $\lambda = 1/0.6$) and W_n having exponential distribution with mean 1/0.8. As a

consequence, the chain will return an infinite number of times to the barrier, which is an atom of the chain. Figure 8 shows how the (embedded) chain can be split into independent blocks. When the chain stays at the atom then blocks of size one, consisting in the atom itself, are constructed. As a consequence, the mass at the atom is simply obtained by computing the frequency that the embedded chain visits the atom, here 0.21 (the true value obtained by simulation is actually close to 0.244). We choose $n = 1000$ in all our simulations but we only consider chains of size 200. The average length of a block is 4.25 so that the average number of blocks is close to 235.

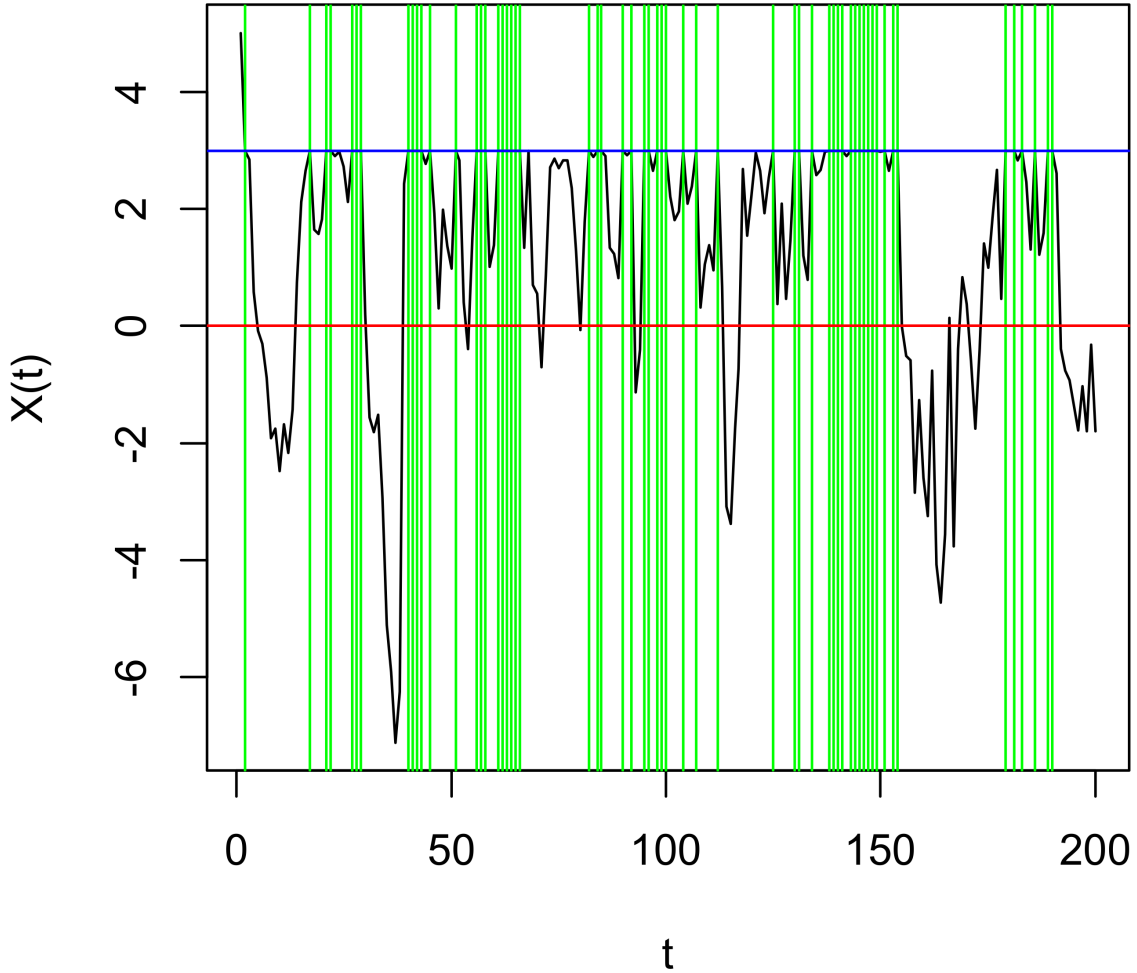


Figure 3.2: An illustration of the splitting technique on the C.L. model with a barrier. Observations between two vertical green lines correspond to independent blocks. The barrier is the horizontal blue line, here at $d = 3$ and the process starts at $X(0) = 5$.

When there is no contamination, the plug-in estimator is rather close to the true distribution (except for a few simulations exhibiting large blocks which also perturb the estimation procedure). In Figure 3.3, we represent the continuous part of the stationary distribution of the associated PDMP.

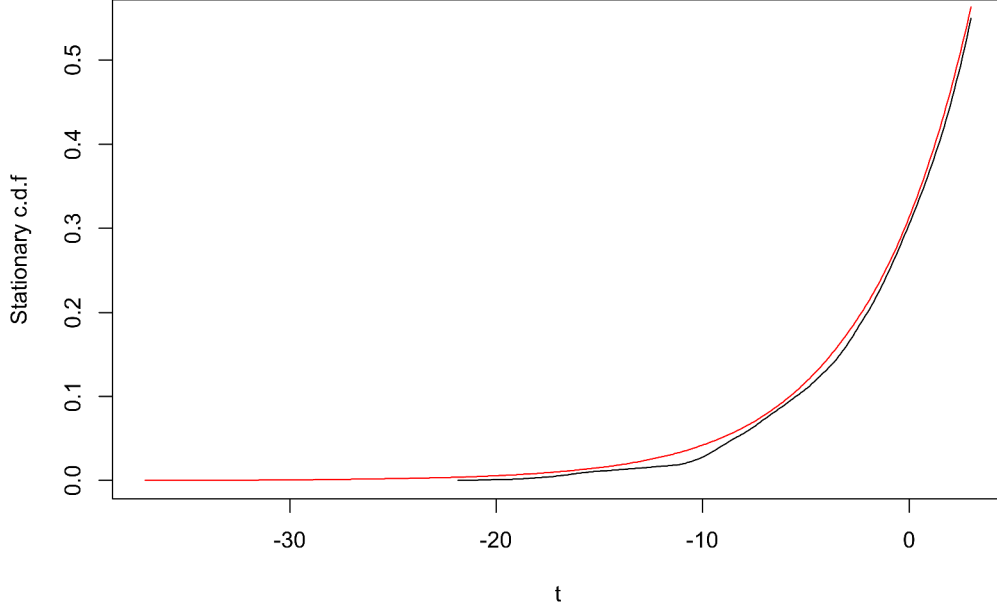


Figure 3.3: Comparison between the true distribution (the red line) and the plug-in estimator (the black line) when there is no contamination.

On the contrary, a few outliers may completely perturb the trajectory of the PDMP and of the embedded chain and destroy the properties of the plug-in estimator. In our simulations, we choose to introduce a Dirac measure at -25 (a fixed loss) for the distribution of W_n for 4 successive times (around the date 75 for a series of length 200, more generally $[3/8n]$ in all our simulations). In most simulations, this induces a very large block as shown in Figure 3.4. This single block completely perturbs the estimators at different levels: first the number of blocks is clearly underestimated (as a consequence the mass of the stationary distribution is totally underestimated), second this induces an important bias on the continuous part of the stationary distribution, both in term of the values of the estimator itself and its standardization.

The robustified estimator $\hat{\mu}_n([-\infty, v])$ gives a reasonable estimator for an adequate choice of the truncation; see Figure 3.5. We choose a fixed level equal to 30 (up to the truncation level 45, this does not change the results) which is of the order of magnitude of the largest

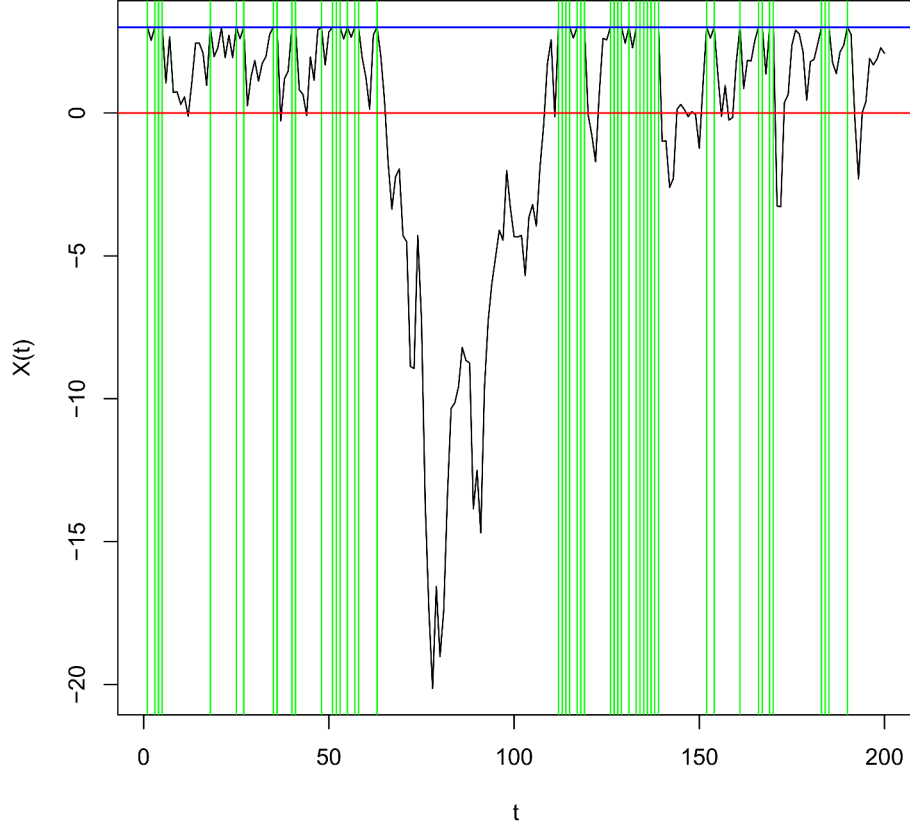


Figure 3.4: Trajectory of the C.L. model when a contamination is added (around 75) thanks to a Dirac measure.

excursion when there is no outlier (on a stretch of size $n = 1000$). Notice that the contamination is of the same order. When choosing a truncation which is too small, the robust estimator is based on too few blocks and gives poor results. We recommend to select the level of the truncation as a quantile of the distribution of the length of the blocks, typically when using the 95% quantile we obtain 35 and efficiently get rid of the largest blocks. Other simulation results on specific functionals (robustified quantiles, robustified means, winsorized means, R-estimators) may be found in [29]. Their examples also show the importance of robustifying bootstrap estimators as well, since resampling large blocks may also lead to the inconsistency of the bootstrap procedure.

Finally, we would like to conclude on a specific problem encountered in the general Harris recurrent case. When dealing with non-regenerative processes, the procedure is more complex and requires a preliminary estimator of the kernel density of the embedded chain. This actually may cause some difficulty in the procedure if the kernel estimator is based on

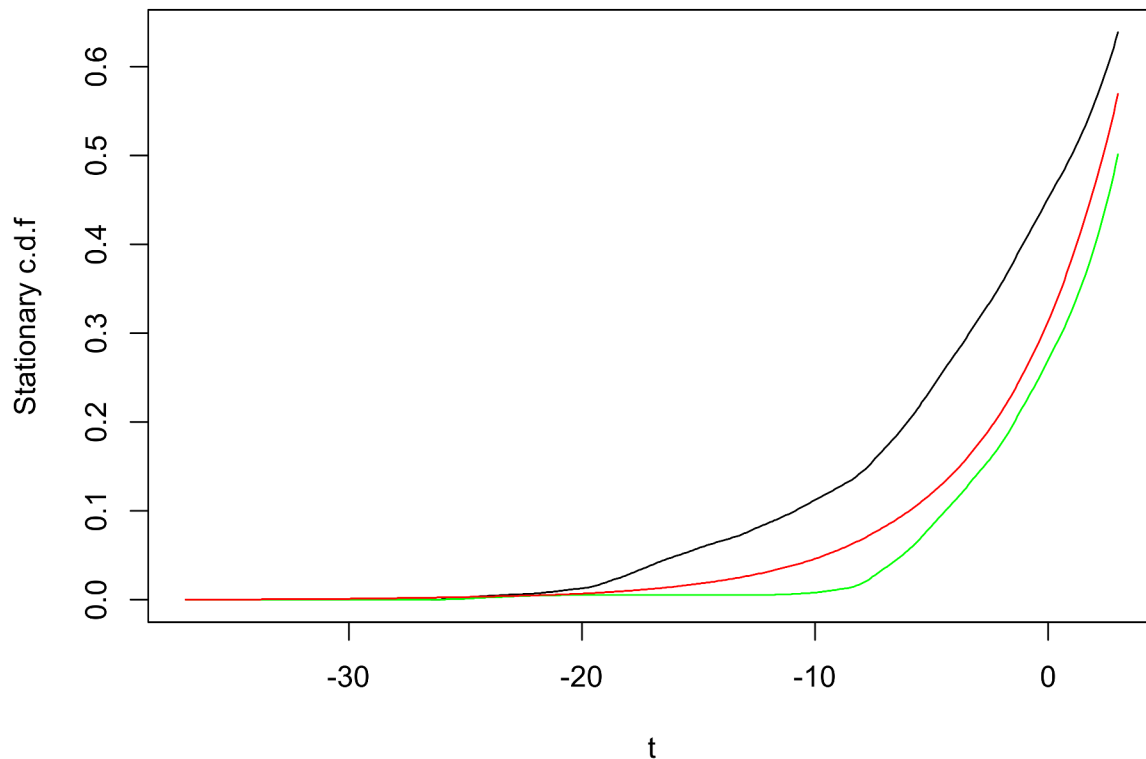


Figure 3.5: Comparison between the true distribution (the red line), the plug-in estimator (the black line) and the robustify estimator (the green line) in the presence of contamination.

all observations. Indeed, kernel density estimators are not robust to the outliers, so that the splitting procedure may be considerably perturbed in this step. We propose the following empirical rule to avoid such problem:

1. Estimate the kernel using observations with very few long excursions from the mean.
2. Compute the blocks with this estimator to split the time series into approximate blocks.
3. Eliminate large blocks (eliminate for instance the $\alpha\%$ largest blocks).
4. Recompute the kernel estimator on the remaining observations, and split again the time series with this new estimator.

The procedure may be iterated until the estimator of the density is stable. The preliminary empirical results that we obtain with this algorithm are encouraging, but the theoretical aspects require further investigation.

9 Conclusion

In this chapter we studied the properties of Markov chains of the embedded chain associated to PDMPs. In particular, we explored regenerative (or approximately regenerative) approach in the framework of PDMPs and showed how to relate the properties of PDMPs (stationary distribution) to the properties of embedded chains, for which we can build efficient robust estimators based on regeneration blocks. In this work, we have developed some tools to detect outliers and build robust plug-in estimators by eliminating blocks having either too much contribution to the statistics of interest or having a length too large, resulting in an important bias on the statistics. We illustrated our approach on two stochastic models which are PDMPs: the Sparre-Andersen model with barrier and the Kinetic Dietary Exposure Model.

Chapter 4

Residual and wild bootstrap methods for periodically autoregressive processes

The main objective of this chapter is to establish the residual and the wild bootstrap procedures for periodically autoregressive models. We use the least squares estimators of model's parameters and generate their bootstrap equivalents. We prove that the bootstrap procedures for causal periodic autoregressive time series with finite fourth moments are weakly consistent. Finally, we confirm our theoretical considerations by simulations.

This chapter gathers results published in [46]. Some additional remarks and explanations are added in order to make this exposition more comprehensive.

1 Preliminaries and Markovian form of PAR(p) processes

As mentioned in Chapters 1 and 2, since the 'naive' bootstrap algorithm for the i.i.d data was proposed by Efron [59], many bootstrap schemes for dependent data have been introduced. A very natural idea is to try to switch from time dependence in the data back to an i.i.d. framework by fitting a parametric model. One of the approaches is to apply so called residual bootstrap. This method was primarily developed for an autoregressive stationary process with finite order, i.e.

$$A(L)y_t = \epsilon_t,$$

where $\{\epsilon_t\}$ is an i.i.d. sequence with zero mean and finite second moment. By

$$Y = [y_1, \dots, y_t]'$$

we denote vector of observations and $A(L)$ is an invertible polynomial in lag operator (see for instance [39], page 84).

Autoregressive process of order one (AR(1)) was bootstrapped for the first time in [60] and [147] and AR(p) in [134]. Since then, the residual bootstrap procedure has been adapted to many other processes, for instance Shimizu [131] bootstraps AR-ARCH or ARMA-GARCH models. We also refer to a survey [21] treating bootstrapping time series for more details and references. The classical residual bootstrap procedure for AR(p) process consists of 6 steps.

Algorithm 5 Residual bootstrap for AR(p) process

Step 1 Compute the estimator $\hat{A}(L)$ of $A(L)$.

Step 2 Compute centred residuals of the estimated model

$$\bar{\epsilon}_t = \hat{A}(L)y_t.$$

Step 3 Generate bootstrap variables ϵ_t^* by resampling with replacement from $\bar{\epsilon}_t$.

Step 4 Generate the bootstrap version of Y i.e.

$$\hat{A}(L)y_t^* = \epsilon_t^*$$

conditioned on Y_0^* .

Step 5 Compute bootstrap estimates $\hat{A}^*(L)$ of the model.

Step 6 Repeat steps 3 – 6 many times in order to construct the targeted empirical distribution function.

It is noteworthy that the residual bootstrap method is primarily designed for homoscedastic models. In case when the model is heteroscedastic, one may apply wild bootstrap procedure instead. Main innovation when comparing to residual bootstrap presented in Algorithm 0 is to compute variables $\epsilon_t^\dagger = \hat{\epsilon}_t w_t^\dagger$, where $w_t^\dagger \sim \mathcal{N}(0, 1)$ is an i.i.d. sequence. Next, we generate bootstrap version Y^\dagger of Y in Step 5 such that $\hat{A}(L)y_t^\dagger = \epsilon_t^\dagger$, see [93] and [131] for details.

Furthermore, residual and wild bootstrap methods may be useful when considering periodically correlated (PC) times series. PC sequences were for the first time considered by Gladyshev [73] and since then have received much attention due to the fact that they can model many phenomena in the real world. Standard time series analysis often relies on the stationarity assumption. However, this condition is not satisfied by many processes in the

real world applications. In this case, one may want to work with PC time series since they are non-stationary processes with many properties of stationary processes. Thus, this motivated many researchers to model many physical phenomena using PC processes. PC time series were primarily applied to the field of hydrology in [135] where they have been used to model monthly stream flow of the rivers. Since then, PC processes are widely used in numerous fields, for instance in climatology (see [56], [66], [84]), finance and econometrics (see [34], [40], [81]) or analysis of genome and biological signals (see [7], [67], [101]). In this chapter we focus on PAR processes which are periodic versions of well-studied autoregressive processes. PAR processes can model phenomena in many fields such as hydrology (river flows), see [116] or in finance (see for example [40]) and more generally, are useful in any applications when one needs to analyze a process consisting of AR sub-models across the seasons.

It is noteworthy that most of the schemes in the dependent setting are block techniques. These procedures rely on resampling block segments of observations so that dependence structure is preserved. There are many types of block bootstrap methods for dependent data such as moving block bootstrap (MBB), non-overlapping block bootstrap (NBB), circular block bootstrap (CBB) or stationary bootstrap (SB) to name just a few (see for instance [93] for a detailed survey of the aforementioned procedures). MBB, NBB and CBB methods resample blocks of data choosing a block length which is non-random (see [93] and [94]) which causes many problems. One of main drawbacks is that block bootstrap procedures are very sensitive to the choice of the length of the blocks. Indeed, the optimal length of the blocks depends on the sample size and the data generating processes. As mentioned in [94] bias and the variance of a block bootstrap estimator are highly dependent on the length of the block (which plays the role smoothing parameter in this case) and either of them can be the leading term in the expansion for the mean-squared error (MSE) of a block bootstrap estimator. The SB method resamples blocks of random length, however the same amount of bias is asymptotically the same as in case of MBB, NBB and CBB methods. As mentioned in [94] the variances of the SB estimators are always at least twice larger compared to variances of the respective NBB estimators and at least three times larger comparing to those of the MBB and CBB estimators. Moreover, popular MBB method requires the stationarity for observations that usually results in failure of this method in non-stationary setting (see [93] for more details). Finally, it is noteworthy, that the rate of convergence of the MBB distribution is slower than that's of bootstrap distribution in the i.i.d. setting. Bootstrap methods proposed in this chapter do not require block length calibration and are data-driven.

In this chapter we study periodic autoregressive (PAR) models which are extensions of autoregressive (AR) models, i.e. autoregressive parameters vary with the seasons. It essen-

tially means that we define a different AR model for each season in a period. PAR processes are fairly well-studied. Bresford [83] investigated moment estimation for PAR sequences. Pagano [113] proved the asymptotic normality of parameter estimates based on periodic Yule-Walker equations. Vecchia [145] analyzed the correlation structure and the parameter estimation of the periodic autoregression moving average processes (PARMA) with a focus on the moment estimation and maximum likelihood estimation. Basawa and Lund [20] have studied large sample properties of parameter estimates for PARMA models and derived a general limit result for coefficients of such models. Shao and Ni [130] have studied the least-squares estimation and ANOVA for PAR series. They showed that the limit results for PARMA from [20] also hold exclusively for PAR sequences. Ursu and Duchesne [138] have generalized limit results from [20] and [130] to vector periodic autoregressive time series model (PVAR).

In what follows we make our considerations formal. For the reader's convenience we recall the notation introduced for PAR processes in Chapter 1. The periodic autoregressive model of interest is given by

$$X_{nT+v} = \sum_{k=1}^p \phi_k(v) X_{nT+v-k} + \epsilon_{nT+v}, \quad (4.1)$$

where

$$\Phi' = [\phi_1(1), \phi_2(1), \dots, \phi_p(1), \phi_1(2), \dots, \phi_p(2), \dots, \phi_1(T), \dots, \phi_p(T)]$$

denotes the vector of parameters and $'$ is a transpose. The $\{X_{nT+v}\}$ designates the series during the n -th cycle ($0 \leq n \leq N-1$) during v -th season ($1 \leq v \leq T$). The $\{\epsilon_{nT+v}\}$ is the mean zero white noise with variance of the form $\text{Var}(\epsilon_{nT+v}) = \sigma_v^2 > 0$ for all seasons v .

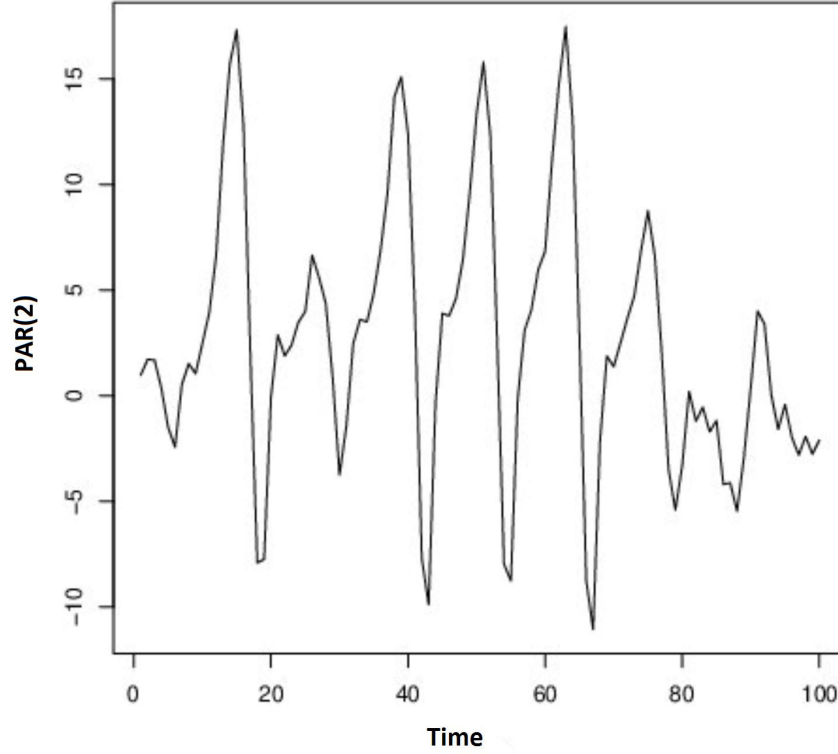
In the following, we slightly abuse the notation, i.e. the periodic notations will be used interchangeably with the non-periodic ones, namely $\{X_t\}$, $\{\epsilon_t\}$. There is no loss of generality in considering the autoregressive model order equal to p instead of $p(v)$ (for each season) as we can take

$$p = \max_{1 \leq v \leq T} p(v)$$

and set

$$\phi_k(v) = 0 \text{ for } p(v) < k \leq p.$$

Figure 4.1: Trajectory of PAR(2) process with 12 seasons and 100 observations.



Note that other periodic notations for the process (4.1) are used in the literature (see for instance [19] and references therein). We have adopted form (4.1) in order to emphasize that the process (4.1) is AR difference equation with periodically varying coefficients. Observe that for $T = 1$, (1.6) is an AR process. There is a strong relation between PAR and AR processes. We formalize this connection in the following remark (we also refer to [20] for further details). We keep our notation consistent with [20].

Remark 13. It is noteworthy that PAR process given in (4.1) may be written as a T -variate AR process. Indeed, the T -variate AR is given by the difference equation

$$\Phi_0 \vec{X}_n - \sum_{k=1}^{p^*} \Phi_k \vec{X}_{n-k} = \vec{\epsilon}_n, \quad (4.2)$$

where

$$\vec{X}_n = [X_{nT+1}, \dots, X_{nT+T}]' \quad \text{and} \quad \vec{\epsilon}_n = [\epsilon_{nT+1}, \dots, \epsilon_{nT+T}]'.$$

We denote by $p^* = \lceil p/T \rceil$ the order of T -variate AR and write $\lceil y \rceil$ for the smallest integer greater than or equal to y . The $T \times T$ autoregressive coefficients of (4.2) are of the form

$$(\Phi_0)_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i < j \\ -\phi_{i-j}(i) & \text{if } i > j \end{cases}$$

and

$$(\Phi_k)_{i,j} = \phi_{kT+i-j}(i), \quad 1 \leq k \leq p^*$$

with the convention $\phi_k(v) = 0$ for $k > p$.

Remark 13 leads us to next conclusion.

Remark 14 (Markovian structure of PAR process). Recall that AR process is a Harris recurrent Markov chain (see Example 2 in Chapter 1). This combined with the fact that PAR process is T -variate AR sequence implies that PAR is also a Harris recurrent Markov chain. Indeed, we just construct multivariate sequence

$$\vec{Y}_n = (\vec{X}_n, \dots, \vec{X}_{n-k})'$$

and consider process $\vec{Y} = \{\vec{Y}_n, n \geq 0\}$.

We assume that the roots of characteristic function

$$z^{p^*} - \sum_{k=1}^{p^*} \Phi_k z^{p^*-k} = 0$$

are less than one in absolute value. This fact is equivalent to PAR model of interest being causal (see also [131], Remark 3.1, page 22).

It is known that when the model is causal, then the stationary T -variate solution $\{\vec{X}_n\}$ to (4.2) is unique. Process $\{X_t\}$ is periodically correlated since we have periodic stationarity of the moments, namely

$$\mathbb{E}(X_{n+T}) = \mathbb{E}(X_n) \text{ and } \text{Cov}(X_{n+T}, X_{m+T}) = \text{Cov}(X_n, X_m).$$

Models X_t which fulfill above moment conditions are called *periodically correlated* or *cyclostationary* (see [19] and [73] for further details).

2 The least squares estimation for model's parameters

In order to generate valid bootstrap approximations we have to obtain the least squares estimator of Φ , namely

$$\widehat{\Phi}' = \left[\widehat{\phi}_1(1), \widehat{\phi}_2(1), \dots, \widehat{\phi}_p(1), \widehat{\phi}_1(2), \dots, \widehat{\phi}_p(2), \dots, \widehat{\phi}_1(T), \dots, \widehat{\phi}_p(T) \right].$$

In what follows, we carry out an analogous analysis as in [138], where the least squares estimators were obtained for PVAR processes. Equation (1.6) can be written as

$$\epsilon_{nT+v} = X_{nT+v} - \sum_{k=1}^p \phi_k(v) X_{nT+v-k}. \quad (4.3)$$

Next, we minimize the sum of squared errors

$$S(\Phi) = \sum_{v=1}^T e'(v)e(v),$$

where

$$e(v) = [\epsilon_v, \epsilon_{T+v}, \dots, \epsilon_{(N-1)T+v}]$$

is $1 \times N$ vector of errors. In order to compute the least squares estimator of Φ we differentiate $S(\Phi)$ with respect to each parameter $\phi_k(v)$, $k = 1, \dots, p(v)$; $v = 1, \dots, T$. Thus, we get

$$\frac{\partial S(\Phi)}{\partial \phi_k(v)} = -2 \sum_{n=0}^{N-1} X_{nT+v-k} \epsilon_{nT+v}.$$

Next, we set the derivatives equal to zero (for $k = 1, \dots, p(v)$) and obtain for a given season v

$$\sum_{n=0}^{N-1} W_n(v) \epsilon_{nT+v} = \vec{0}, \quad (4.4)$$

where $\vec{0}$ is $p(v) \times 1$ vector of zeros and

$$W_n(v) = [X_{nT+v-1}, \dots, X_{nT+v-p(v)}]'. \quad (4.5)$$

are $p(v) \times 1$ random vectors, $n = 0, 1, \dots, N-1$. Vectors defined in (4.5) form a $N \times p(v)$ random matrix

$$W(v) = [W_0(v), \dots, W_{N-1}(v)]'.$$

We consider

$$\epsilon_{nT+v} = X_{nT+v} - W_n'(v)\Phi(v), \quad (4.6)$$

where $\Phi(v) = [\phi_1(v), \dots, \phi_{p(v)}(v)]'$. The normal equations for (4.6) at the season v are of the form

$$\sum_{n=0}^{N-1} W_n(v) X_{nT+v} = \left(\sum_{n=0}^{N-1} W_n(v) W_n'(v) \right) \Phi(v).$$

The least squares estimators of $\Phi(v)$ fulfill the following relation

$$\hat{\Phi}(v) = \left((W'(v)W(v))^{-1} W'(v) \right) z(v),$$

where

$$z(v) = [X_v, X_{T+v}, \dots, X_{(N-1)T+v}]'$$

is $N \times 1$ random vector and

$$\hat{\Phi}(v) = [\phi_1(v), \dots, \phi_{p(v)}(v)]'$$

is $p(v) \times 1$ vector of the least squares estimates of parameters of model (4.6) at the season v , $v = 1, \dots, T$. The invertibility of $W'(v)W(v)$ is ensured by Proposition 4.1 from [19]. For the sake of completeness, we provide it below.

Remark 15. If $\sigma_v > 0$ for each $1 \leq v \leq T$, then for a causal PAR model, $W'(v)W(v)$ is invertible for each $n \geq 1$.

Thus, the residuals are of the form

$$\hat{\epsilon}_{nT+v} = X_{nT+v} - \left(W_n'(v) \hat{\Phi}(v) \right). \quad (4.7)$$

In what follows we discuss the asymptotic normality of $\hat{\Phi}(v)$, $1 \leq v \leq T$. This limit result is inevitable for our bootstrap theory to work. We are heavily relying on central limit theorems presented in [130] and [20]. For the reader's convenience, we provide the theorem of Shao and Ni from [130] which is a general limit result for PAR(p) process. We denote by \xrightarrow{D} the convergence in distribution.

Theorem 12. Assume that a periodic autoregressive series $\{X_{nT+v}\}$ defined in (1.6) is causal and has finite fourth moment. Then,

$$\sqrt{N} \left(\hat{\Phi} - \Phi \right) \xrightarrow{D} \mathcal{N}(0, R^{-1}) \text{ as } n \rightarrow \infty,$$

where $R = R_\phi(v_1, v_2)$ and is defined in the following way: for any $k \geq 1$, $l \leq p$ and $v_1 \geq 1$, $v_2 \leq T$

$$R_\phi(v_1, v_2) = \mathbb{E} \left(\left(\frac{\partial \tilde{\epsilon}_n}{\partial \phi_k(v_1)} \right) \left(\frac{\partial \tilde{\epsilon}_n}{\partial \phi_l(v_2)} \right)' \right),$$

where

$$\tilde{\epsilon}' = [\tilde{\epsilon}'_0, \tilde{\epsilon}'_1, \dots, \tilde{\epsilon}'_{N-1}], \quad \tilde{\epsilon}'_n = \left[\frac{\epsilon_{nT+1}}{\sigma_1}, \frac{\epsilon_{nT+2}}{\sigma_2}, \dots, \frac{\epsilon_{nT+T}}{\sigma_T} \right]$$

and σ_v is the standard deviation of ϵ_{nT+v} at the season v ($1 \leq v \leq T$).

Remark 16. Note that in the above theorem σ_v is assumed to be known for each season v ($1 \leq v \leq T$). However, in case when σ_v is unknown, it can be replaced by \sqrt{N} -consistent estimator. The limiting distribution of $\hat{\Phi}$ remains unchanged (see [20] for details).

3 Residual bootstrap for PAR processes

In what follows we present residual bootstrap method for PAR(p) processes. As mentioned before, residual method allows to avoid problems of block bootstrap methods for time series, i.e. we do not need to struggle with the choice of the length of the block. We recall briefly that residual bootstrap method is originally designed for heteroscedastic models, however it is possible to adapt it in order to make it applicable for PAR sequences. The second method we present is the wild bootstrap which is tailor-made for heteroscedastic models. Both methods we propose require to obtain the ordinary least squares estimator $\hat{\Phi}$ of Φ and secondly to compute residuals

$$\hat{\epsilon}_{nT+v} = X_{nT+v} - \sum_{i=1}^p \hat{\phi}_i(v) X_{nT+v-i}. \quad (4.8)$$

As in the case of bootstrapping AR(p) processes, the difference between residual and wild bootstrap procedures boils down to a different way of computing bootstrap version of residuals. In residual bootstrap one draws randomly with replacement centred and scaled residuals in order to get bootstrap random variables $\bar{\eta}_{nT+v}^*$. In order to preserve the periodic structure of the bootstrap version of X_{nT+v} , we multiply $\bar{\eta}_{nT+v}^*$ by standard deviation of corresponding season and X_{nT+v} . Finally, we generate the bootstrap version X_{nT+v}^* of the PAR process and compute bootstrap estimates $\hat{\Phi}^*$. In the wild bootstrap method we obtain bootstrap version of residuals (4.8) by multiplying $\hat{\epsilon}_{nT+v}$ by random variables drawn from normal distribution with zero mean and variance 1. From now, we generate bootstrap version of process X_{nT+v} with bootstrap residuals ϵ_{nT+v}^\dagger . Thus, the wild bootstrap method results in obtaining bootstrap estimates $\hat{\Phi}^\dagger$.

Typically, one uses the wild bootstrap when there is heteroscedasticity in the model. Residual bootstrap method is designed for models that are homoscedastic. PAR models are heteroscedastic, thus a natural way is to use the wild bootstrap method. However, we

adjusted residual bootstrap procedure (by dividing residuals by σ_v in Step 3 and multiplying residuals by σ_v in Step 5) such that it can be used for PAR sequences.

It is noteworthy that from a second order theory point of view (see [76] for details) the residual bootstrap outperforms the wild bootstrap (generated by Gaussian noise). Indeed, the second one can not correct adequately the skewness of the distribution. However, in practice, this difference between the two methods is difficult to distinguish (except for very small samples, see our simulation studies at the end of this chapter).

The bootstrap procedure for PAR(p) processes is formulated in Algorithm 6.

Algorithm 6 Residual bootstrap method for PAR processes

Step 1 Compute the ordinary least squares estimator $\hat{\Phi}$ of Φ .

Step 2 Compute the residuals of the estimated model

$$\hat{\epsilon}_{nT+v} = X_{nT+v} - \sum_{k=1}^p \hat{\phi}_k(v) X_{nT+v-k},$$

where $1 \leq v \leq T$, $0 \leq n \leq N-1$.

Step 3 Compute the centred residuals

$$\bar{\eta}_{nT+v} = \frac{\hat{\epsilon}_{nT+v}}{\sigma_v} - \frac{1}{NT} \sum_{n=0}^{N-1} \sum_{v=1}^T \frac{\hat{\epsilon}_{nT+v}}{\sigma_v},$$

where NT is the number of all observations in the model.

Step 4 Generate bootstrap variables η_{nT+v}^* by drawing randomly with replacement from

$$\{\bar{\eta}_1, \dots, \bar{\eta}_{NT}\}.$$

Step 5 Generate the bootstrap version of the model (1.6)

$$X_{nT+v}^* = \sum_{k=1}^p \hat{\phi}_k(v) X_{nT+v-k} + \sigma_v \eta_{nT+v}^*, \quad 1 \leq v \leq T.$$

Step 6 Calculate the bootstrap estimators of parameters for each season v , $1 \leq v \leq T$

$$\hat{\Phi}^*(v) = (W'(v)W(v))^{-1} W'(v) z^*(v),$$

where

$$z^*(v) = [X_v^*, \dots, X_{(N-1)T+v}^*]', \quad 1 \leq v \leq T.$$

Remarks 13. In what follows we will give few remarks concerning some steps of Algorithm 6.

1. In Step 3 in the case when σ_v^2 are known, we divide $\widehat{\epsilon}_{nT+v}$ by the true standard deviation σ_v . When σ_v^2 are unknown, we replace it by a \sqrt{N} -consistent estimator of σ_v^2 ; e.g., one can use the sample variance of the residuals obtained in Step 2, that is from sample $\{\widehat{\epsilon}_{nT+v}\}_{n=0}^{N-1}$.
2. In Step 4 variables η_{nT+v}^* , $0 \leq n \leq N-1$, $1 \leq v \leq T$ are conditionally independent on the data. Their common distribution is defined as

$$\mathbb{P}^*(\eta_1^* = \bar{\eta}_{nT+v}) = \frac{1}{NT}, \quad (4.9)$$

where \mathbb{P}^* is the distribution of η^* conditionally on the observations.

3. Note that in Step 6 vector of bootstrap estimates $\widehat{\Phi}^*$ can be written as

$$\begin{aligned} \widehat{\Phi}^*(v) &= (W'(v)W(v))^{-1}W'(v)(W(v)\widehat{\Phi}(v) + g^*(v)) \\ &= \widehat{\Phi}(v) + (W'(v)W(v))^{-1}W'(v)g^*(v), \end{aligned}$$

where

$$g^*(v) = [\epsilon_v^*, \dots, \epsilon_{(N-1)T+v}^*]' = [\sigma_v \eta_v^*, \dots, \sigma_v \eta_{(N-1)T+v}^*]'$$

In what follows we show that the bootstrap procedure proposed in Algorithm 6 for a periodic autoregressive series X_{nT+v} is weakly consistent. Our proof of the validity of bootstrap is kept in spirit of the proof of Theorem 3.1 from [131]. We recall that $\xrightarrow{P^*}$ denotes weak convergence conditionally on the data in probability.

Theorem 14. *Suppose that a periodic autoregressive series X_{nT+v} defined in (4.1) is causal and has finite fourth moment. Then, the residual bootstrap procedure given by the Algorithm 6 is weakly consistent in probability, i.e.*

$$\sqrt{N}(\widehat{\Phi}^* - \widehat{\Phi}) \xrightarrow{P^*} \mathcal{N}(0, R^{-1}).$$

Proof. Central limit theorem for PAR processes from [130] guarantees that

$$\sqrt{N}(\widehat{\Phi} - \Phi) \xrightarrow{D} \mathcal{N}(0, R^{-1}),$$

where

$$R = R_\phi(v_1, v_2) = \mathbb{E} \left(\frac{\partial \tilde{\epsilon}_n}{\partial \phi_k(v_1)} \cdot \frac{\partial \tilde{\epsilon}_n}{\partial \phi_l(v_2)} \right), \quad 1 \leq v_1, v_2 \leq T$$

and

$$\tilde{\epsilon}_n = \left[\frac{\epsilon_{nT+1}}{\sigma_1}, \frac{\epsilon_{nT+2}}{\sigma_2}, \dots, \frac{\epsilon_{nT+T}}{\sigma_T} \right]$$

and σ_v is the standard deviation of ϵ_{nT+v} for $1 \leq v \leq T$.

Moreover, Shao and Ni [130] proved that the estimates of $\Phi(v_1)$ and $\Phi(v_2)$ are asymptotically independent if $v_1 \neq v_2$. It follows from point 3 from Remarks 13 that

$$\widehat{\Phi}^*(v) - \widehat{\Phi}(v) = (W'(v)W(v))^{-1}W'(v)g^*(v). \quad (4.10)$$

Observe that the analysis of the PAR(p) model is equivalent to the analysis of AR($p(v)$) model for each fixed v ($1 \leq v \leq T$), respectively. This fact enables us to apply directly the asymptotic results from [77]. Note however, that our method is more general and the proof based on [130] also allows us to use dependent heteroscedastic residuals.

By Hamilton's remarks (see [77], Chapter 8, Page 215), we have for a given v , $1 \leq v \leq T$,

$$\frac{1}{N}W'(v)W(v) \xrightarrow{P} R, \quad (4.11)$$

where \xrightarrow{P} denotes convergence in probability. Thus, by the virtue of (4.11) combined with the Slutsky theorem, it suffices to prove that

$$\frac{1}{\sqrt{N}}W'(v)g^*(v) \xrightarrow{P^*} \mathcal{N}(0, \sigma_v^2 R).$$

Note that for a given season v we have

$$\frac{1}{\sqrt{N}}W'(v)g^*(v) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \epsilon_{nT+v}^* W_n(v).$$

In order to keep the exposition simple, we introduce one more piece of notation. We write

$$y_n^* = \frac{\epsilon_{nT+v}^* W_n(v)}{\sqrt{N}}.$$

In order to show the consistency of bootstrap procedure formulated in Algorithm 6, we verify if the conditions of Lyapunov's central limit theorem hold. Firstly, observe that

$$\mathbb{E}^*(y_n^*) = \frac{1}{\sqrt{N}}W_n(v)\mathbb{E}(\epsilon_{nT+v}^*) = 0,$$

because

$$\epsilon_{nT+v}^* = \sigma(v)\eta_{nT+v}^* \text{ and } \mathbb{E}(\eta_{nT+v}^*) = 0$$

since $\{\eta_{nT+v}^*\}_{0 \leq n \leq N-1}$ are drawn from the empirical distribution function given by (4.9).

Next, we have

$$\sum_{n=0}^{N-1} \mathbb{E}^*(y_n^* y_n^{*'}) = \frac{1}{N} \sum_{n=0}^{N-1} W_n(v) W_n'(v) \mathbb{E}^*(\epsilon_{nT+v}^{*2}).$$

For a given season v we obtain

$$\begin{aligned}
\mathbb{E}^* (\epsilon_{nT+v}^{*2}) &= \frac{1}{N} \sum_{n=0}^{N-1} \sigma_v^2 \eta_n^{*2}(v) \\
&= \frac{1}{N} \sum_{n=0}^{N-1} (\hat{\epsilon}_n(v) - \bar{\epsilon})^2 \\
&= \frac{1}{N} \sum_{n=0}^{N-1} \hat{\epsilon}_n^2(v) - \bar{\epsilon}^2 \\
&= \mathbb{E} (\epsilon_n^2(v)) - \mathbb{E}^2(\epsilon_n(v)) + o_{\mathbb{P}}(1),
\end{aligned} \tag{4.12}$$

where

$$\eta_n^*(v) = [\eta_{nT+v-1}^*, \dots, \eta_{nT+v-p(v)}^*]'$$

and

$$\hat{\epsilon}_n(v) = [\hat{\epsilon}_{nT+v-1}, \dots, \hat{\epsilon}_{nT+v-p(v)}]'$$

Thus, by (4.11) and (4.12) we can deduce that

$$\sum_{n=0}^{N-1} \mathbb{E}^* (y_n^* y_n^{*'}) = \frac{1}{N} \sum_{n=0}^{N-1} W_n(v) W_n'(v) \mathbb{E}^* (\epsilon_{nT+v}^{*2}) \xrightarrow{P} \sigma_v^2 R. \tag{4.13}$$

Let $c \in \mathbb{R}^p$ and

$$s_N^2 = \sum_{n=0}^{N-1} \mathbb{E}^* (c' y_n^*)^2 \xrightarrow{P} c' \sigma_v^2 R c$$

by (4.13) and the Cramér-Wold device. We check the Lyapunov's condition for $\gamma = 1$.

$$\begin{aligned}
\frac{1}{s_N^3} \sum_{n=0}^{N-1} \mathbb{E}^* |c' y_n^*|^3 &= \frac{1}{s_N^3} \sum_{n=0}^{N-1} \frac{1}{N^{3/2}} \mathbb{E}^* |\epsilon_{nT+v}^*|^3 |c' W_n(v)|^3 \\
&= \frac{1}{s_N^3} \sum_{n=0}^{N-1} \frac{1}{\sqrt{N}} \mathbb{E}^* |\epsilon_{nT+v}^*|^3 \frac{1}{N} |c' W_n(v)|^3.
\end{aligned} \tag{4.14}$$

Notice, that by the similar arguments as in (4.12) we have for a given v

$$\mathbb{E}^* |\epsilon_{nT+v}^*|^3 = \frac{1}{N} \sum_{n=0}^{N-1} |\hat{\epsilon}_n(v) - \bar{\epsilon}|^3 = o_{\mathbb{P}}(1). \tag{4.15}$$

Moreover,

$$\frac{1}{N} |c' W_n(v)|^3 = O_{\mathbb{P}}(1) \tag{4.16}$$

since we assumed that X_{nT+v} has finite fourth moment.

In the view of above discussion, we conclude that

$$\frac{1}{s_N^3} \sum_{n=0}^{N-1} \mathbb{E}^* |c' y_n^*|^3 = o_{\mathbb{P}}(1).$$

We have checked that the Lyapunov's conditions for the central limit theorem are satisfied. In what follows, we consider the sequence $c'_0 y_0^*, \dots, c'_{N-1} y_{N-1}^*$. We apply to that sequence the Cramér-Wold theorem and the central limit theorem for triangular arrays. Thus, we obtain the following convergence

$$\frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} y_n^* \xrightarrow{P^*} \mathcal{N}(0, \sigma_v^2 R). \quad (4.17)$$

Combining (4.11) and (4.17) with the Slutsky theorem we obtain immediately that

$$\sqrt{N} (\hat{\Phi}^* - \hat{\Phi}) \xrightarrow{P^*} \mathcal{N}(0, R^{-1})$$

taking scaled versions of $\{\epsilon'_n\}_{0 \leq n \leq N-1}$ and $\{\epsilon_n^{*'}\}_{0 \leq n \leq N-1}$, namely

$$\tilde{\epsilon}'_n = \left[\frac{\epsilon_{nT+1}}{\sigma_1}, \dots, \frac{\epsilon_{nT+T}}{\sigma_T} \right]$$

and

$$\tilde{\epsilon}_n^{*'} = \left[\frac{\epsilon_{nT+1}^*}{\sigma_1}, \dots, \frac{\epsilon_{nT+T}^*}{\sigma_T} \right]$$

which completes the proof. \square

It is easy to see that Theorem 14 generalizes the results for AR processes (compare with Chapter 8 in [93]). At the end of this chapter we present simulation study which demonstrates the large sample properties of the residual bootstrap for PAR sequences.

4 Wild bootstrap for PAR(p) time series

In this section we briefly present the wild bootstrap method for periodically autoregressive models as an alternative to the residual bootstrap procedure from the previous section. We investigate the behaviour of those two methods through simulations study in Section 5. As mentioned before, in the wild bootstrap procedure one obtains the bootstrap version of residuals ϵ_{nT+v}^\dagger by multiplying the residuals $\hat{\epsilon}_{nT+v}$ by realizations from normal distribution $\mathcal{N}(0, 1)$. This procedure is described in details for example in [88] or [100]. The wild bootstrap procedure for PAR(p) processes is presented in Algorithm 7.

Algorithm 7 Wild bootstrap method for PAR processes

Step 1 Compute the ordinary least squares estimator $\widehat{\Phi}$ of Φ .

Step 2 Compute the residuals of the estimated model

$$\widehat{\epsilon}_{nT+v} = X_{nT+v} - \sum_{k=1}^p \widehat{\phi}_k(v) X_{nT+v-k},$$

where $1 \leq v \leq T$, $0 \leq n \leq N-1$.

Step 3 Generate the bootstrap process X_{nT+v}^\dagger for each season v , $1 \leq v \leq T$

$$X_{nT+v}^\dagger = \sum_{k=1}^p \widehat{\phi}_k(v) X_{nT+v-k} + \epsilon_{nT+v}^\dagger$$

and

$$\epsilon_{nT+v}^\dagger = \widehat{\epsilon}_{nT+v} \eta_{nT+v}^\dagger,$$

where $\eta_{nT+v}^\dagger \sim \mathcal{N}(0, 1)$ and $(\eta_{nT+v}^\dagger)_{nT+v \in \mathbb{R}}$ is independent of $\widehat{\epsilon}_{nT+v}$.

Step 4 Calculate the bootstrap estimator of parameters, namely

$$\widehat{\Phi}^\dagger(v) = (W'(v)W(v))^{-1}W'(v)z^\dagger(v)$$

for each season v , $1 \leq v \leq T$, where

$$z^\dagger(v) = \left[X_v^\dagger, \dots, X_{(N-1)T+v}^\dagger \right]'$$

Remark 17. In Step 4, note that $\widehat{\Phi}^\dagger$ can be written as

$$\begin{aligned} \widehat{\Phi}^\dagger(v) &= (W'(v)W(v))^{-1}W'(v)(W(v)\widehat{\Phi}(v) + g^\dagger(v)) \\ &= \widehat{\Phi}(v) + (W'(v)W(v))^{-1}W'(v)g^\dagger(v), \end{aligned}$$

where

$$g^\dagger(v) = \left[\epsilon_v^\dagger, \dots, \epsilon_{(N-1)T+v}^\dagger \right]' = \left[\widehat{\epsilon}_v \eta_v^\dagger, \dots, \widehat{\epsilon}_{(N-1)T+v} \eta_{(N-1)T+v}^\dagger \right]'$$

In what follows we show weak consistency of the wild bootstrap procedure for PAR sequences formulated in Algorithm 7. We present this theorem in the same form as we did it in [46].

Theorem 15. *Assume that a periodic autoregressive series X_{nT+v} defined in (4.1) is causal and has finite fourth moment. Then, the wild bootstrap procedure defined in Algorithm 7 is weakly consistent in probability, i.e.*

$$\sqrt{N}(\widehat{\Phi}^\dagger - \widehat{\Phi}) \xrightarrow{P^*} \mathcal{N}(0, R^{-1}),$$

where

$$R = R_\phi(v_1, v_2) = \mathbb{E} \left(\frac{\partial \tilde{\epsilon}_n}{\partial \phi_k(v_1)} \cdot \frac{\partial \tilde{\epsilon}_n}{\partial \phi_l(v_2)} \right), \quad 1 \leq v_1, v_2 \leq T$$

and

$$\tilde{\epsilon}_n = \left[\frac{\epsilon_{nT+1}}{\sigma_1}, \frac{\epsilon_{nT+2}}{\sigma_2}, \dots, \frac{\epsilon_{nT+T}}{\sigma_T} \right]$$

and σ_v is the standard deviation of ϵ_{nT+v} for $1 \leq v \leq T$.

Proof. The proof of Theorem 15 is analogous to the proof of Theorem 14 and boils down to verifying conditions of Lyapunov's central limit theorem. \square

5 Simulations

The purpose of these simulations is to analyze the performance of the residual and the wild bootstrap for PAR processes. Firstly, we compare densities of $\sqrt{N}(\hat{\Phi}^* - \hat{\Phi})$ and asymptotic density of $\sqrt{N}(\hat{\Phi} - \Phi)$. Secondly, we compare densities $\sqrt{N}(\hat{\Phi}^\dagger - \hat{\Phi})$ and asymptotic density of $\sqrt{N}(\hat{\Phi} - \Phi)$. Next, we compute actual coverage probabilities (ACP) for simulated PAR sequences. We perform simulations for three different PAR(p) processes in order to show that both methods work well for simple and more complicated models with different numbers of periods and seasons. In the residual bootstrap procedure in the third step of Algorithm 6 we used the sample variance of the residuals obtained in the second step of Algorithm 6 since for simulations we assume that σ_v^2 are unknown (see Remarks 13). We consider the following PAR models:

M1 PAR(1), $T = 3$, $\sigma_1 = 1$, $\sigma_2 = 1$, $\sigma_3 = 1$, $\phi_1(1) = 0.2$, $\phi_1(2) = 0.3$, $\phi_1(3) = 0.2$,

M2 PAR(3), $T = 3$, $\sigma_1 = 1$, $\sigma_2 = 1$, $\sigma_3 = 1$, $\phi_1(1) = 0.2$, $\phi_2(1) = 0.1$, $\phi_3(1) = 0.05$,
 $\phi_1(2) = 0.3$, $\phi_2(2) = 0.2$, $\phi_3(2) = 0.1$, $\phi_1(3) = 0.2$, $\phi_2(3) = 0.1$, $\phi_3(3) = 0.05$,

M3 PAR(2), $T = 2$, $\sigma_1 = 1$, $\sigma_2 = 2$, $\phi_1(1) = 0.2$, $\phi_2(1) = 0.1$, $\phi_1(2) = 0.3$, $\phi_2(2) = 0.2$.

In our simulations we take 599 bootstrap repetitions. We observe that both bootstrap methods work well for $n = 100$ and larger samples, below we illustrate the case when $n = 300$. Our simulations revealed that the residual and the wild bootstrap methods struggle when sample size $n < 100$. We show that the wild bootstrap method behaves worse than the residual bootstrap when the samples are small, below we give an example for $n = 30$.

Figures 4.2-4.7 show that for all considered models tails of distributions are well estimated (when applying the residual and the wild bootstrap procedures). We observe that especially

for the residual bootstrap the centers of the distributions are not very well estimated, however this problem occurs also when bootstrapping i.i.d. data.

When a sample is small ($n = 30$) the distributions are not well estimated (see Figures 4.8 and 4.9). Moreover, in this case we observe worse behaviour of the wild bootstrap method.

To determinate whether $\sqrt{N}(\hat{\phi}_k^*(v) - \hat{\phi}_k(v))$ and $\sqrt{N}(\hat{\phi}_k^\dagger(v) - \hat{\phi}_k(v))$ follow a normal distribution we use the Lilliefors test. The received p -values for model M3 are gathered in Table 4.7. One can see that the p -values are larger than the significance level, the decision is to fail to reject the null hypothesis because we do not have enough evidence to conclude that the data do not follow a normal distribution. The Lilliefors test confirm that the distribution of considered statistics is approximately normal when $n = 300$, and we reject that hypothesis when $n = 30$.

In our simulations we have considered 95% confidence intervals. The whole procedure was repeated 500 times and the ACPs were calculated. Results gathered in Tables 4.1-4.3 show that the actual coverage probabilities for simulated PAR models with $n = 300$ are very well estimated. The results for both bootstrap methods are very similar but in most of the cases the ACP for all coefficients is closer to the nominal confidence level in the residual bootstrap method, especially when sample size is small (see Table 4.1). For model M2, where we estimate relatively many parameters and additionally number of cycles is smaller comparing with model M1 or model M3 of the same size $n = 300$, the ACP is 3.8% lower than the nominal confidence level for the wild bootstrap method for $\phi_2(2)$ and 2.8% lower than the nominal confidence level for residual bootstrap method for coefficient $\phi_1(1)$. We obtain similar result for $\phi_3(3)$, this time ACPs are too large, but not larger by 2.8% than the nominal level. However, obtained results for model M2 are still satisfactory.

6 Conclusion

In this chapter, we have formulated two bootstrap procedures for periodically autoregressive time series. We have established a general limit results for bootstrap estimates of models' coefficients. We have shown that the residual and the wild bootstrap methods for periodic sequences is a natural generalization of bootstrap procedure for simple autoregressive processes given in Lahiri [93]. Finally, we have illustrated our theoretical results by simulations. The simulations confirmed the weak consistency of the residual and the wild bootstrap for PAR sequences.

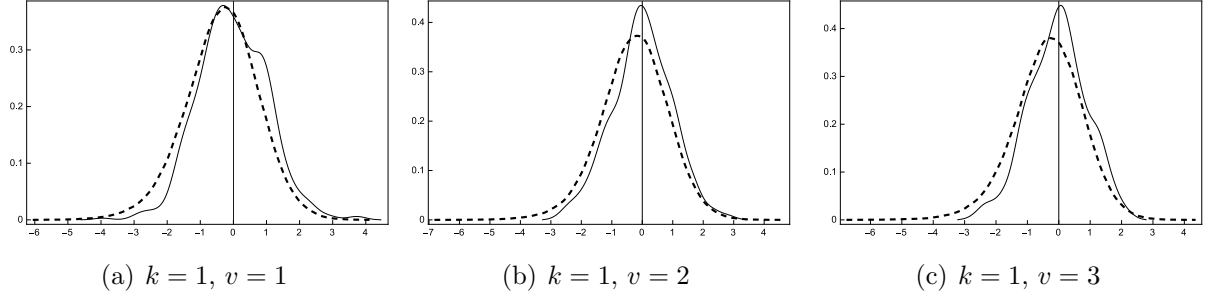


Figure 4.2: The residual bootstrap distribution (**thick dashed lines**: $\sqrt{N}(\hat{\phi}_k^*(v) - \hat{\phi}_k(v))$) and the least squares estimators asymptotic distribution (thin lines: $\sqrt{N}(\hat{\phi}_k(v) - \phi_k(v))$) for model **M1** and sample size $n = 300$.

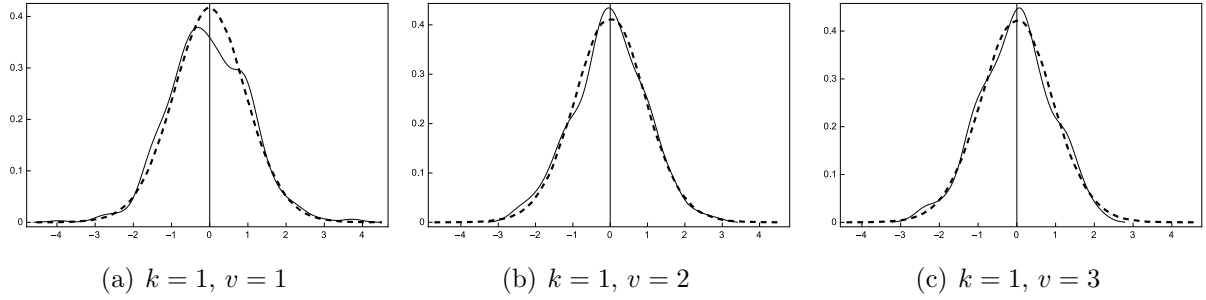


Figure 4.3: The wild bootstrap distribution (**thick dashed lines**: $\sqrt{N}(\hat{\phi}_k^\dagger(v) - \hat{\phi}_k(v))$) and the least squares estimators asymptotic distribution (thin lines: $\sqrt{N}(\hat{\phi}_k(v) - \phi_k(v))$) for model **M1** and sample size $n = 300$.

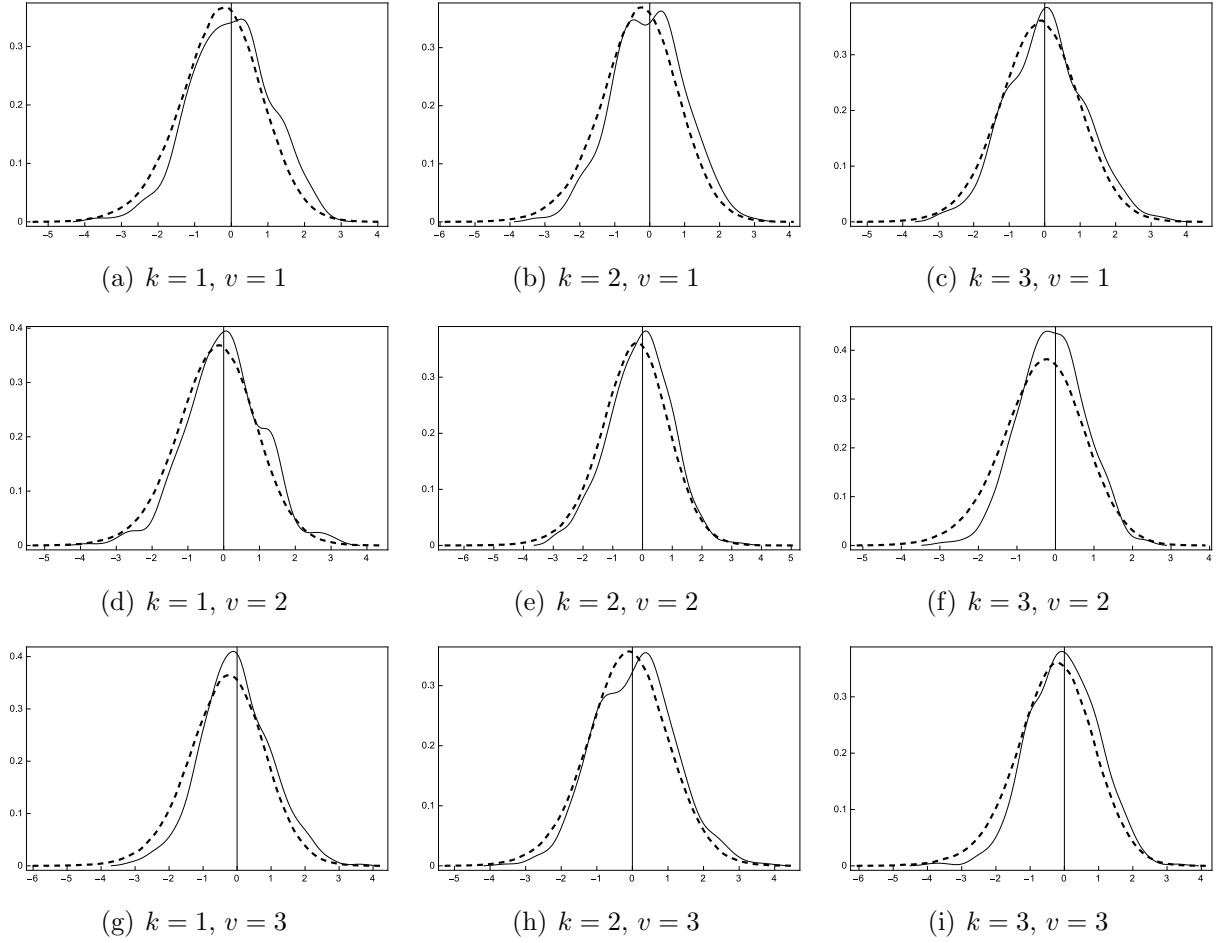


Figure 4.4: The residual bootstrap distribution (**thick dashed lines**: $\sqrt{N}(\hat{\phi}_k^*(v) - \hat{\phi}_k(v))$) and the least squares estimators asymptotic distribution (**thin lines**: $\sqrt{N}(\hat{\phi}_k(v) - \phi_k(v))$) for **M2** and sample size $n = 300$.

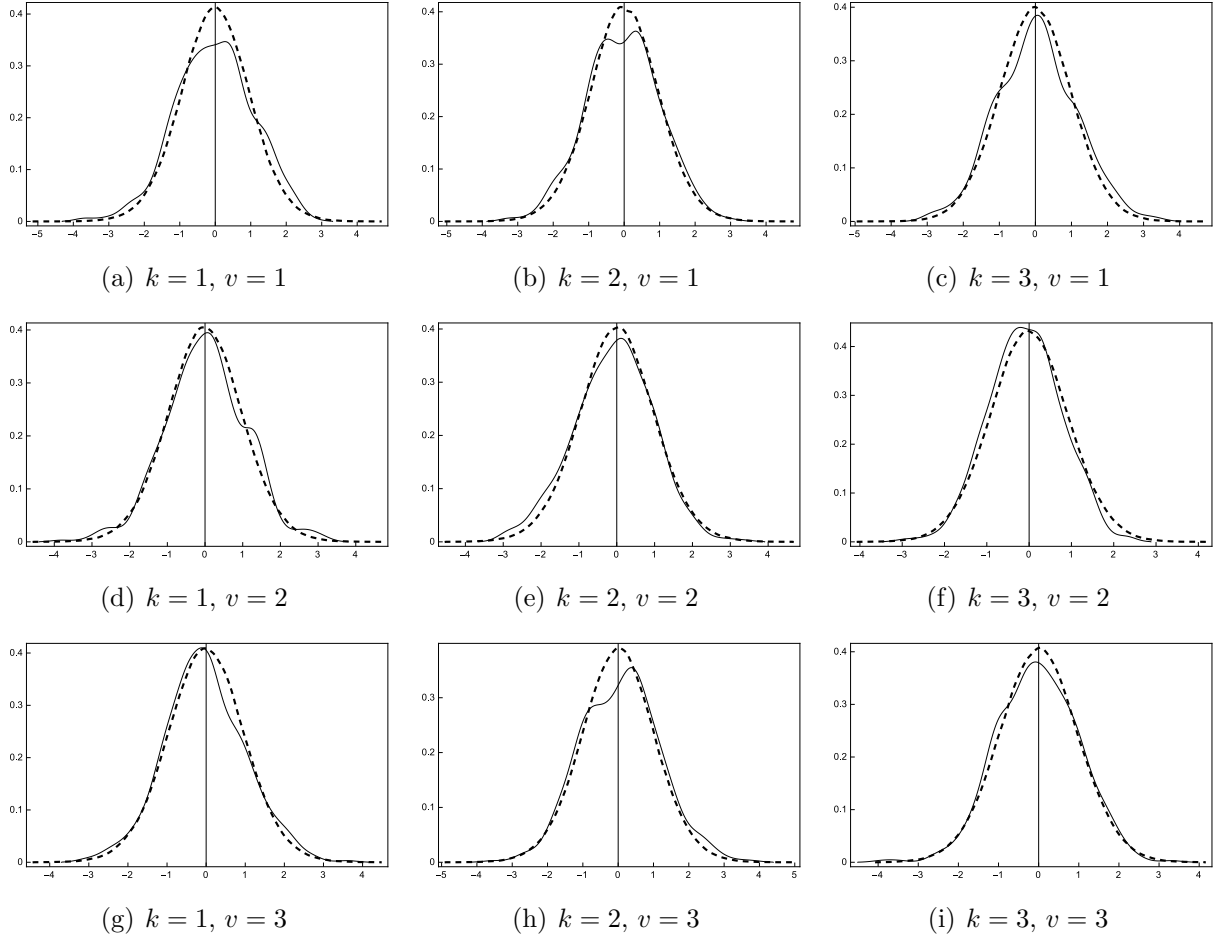


Figure 4.5: The wild bootstrap distribution (**thick dashed lines**: $\sqrt{N}(\hat{\phi}_k^\dagger(v) - \hat{\phi}_k(v))$) and the least squares estimators asymptotic distribution (thin lines: $\sqrt{N}(\hat{\phi}_k(v) - \phi_k(v))$) for **M2** and sample size $n = 300$.

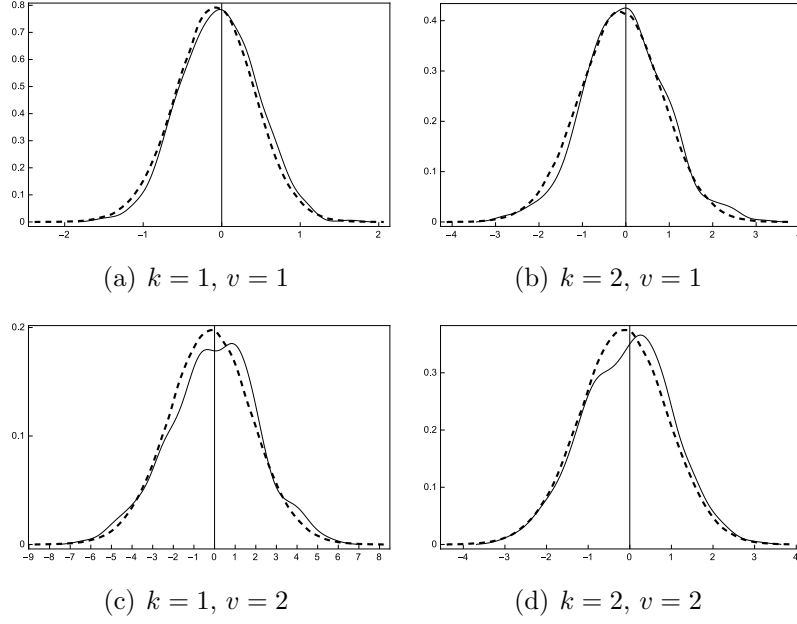


Figure 4.6: The residual bootstrap distribution (**thick dashed lines:** $\sqrt{N}(\hat{\phi}_k^*(v) - \hat{\phi}_k(v))$) and the least squares estimators asymptotic distribution (thin lines: $\sqrt{N}(\hat{\phi}_k(v) - \phi_k(v))$) for model **M3** and sample size $n = 300$.

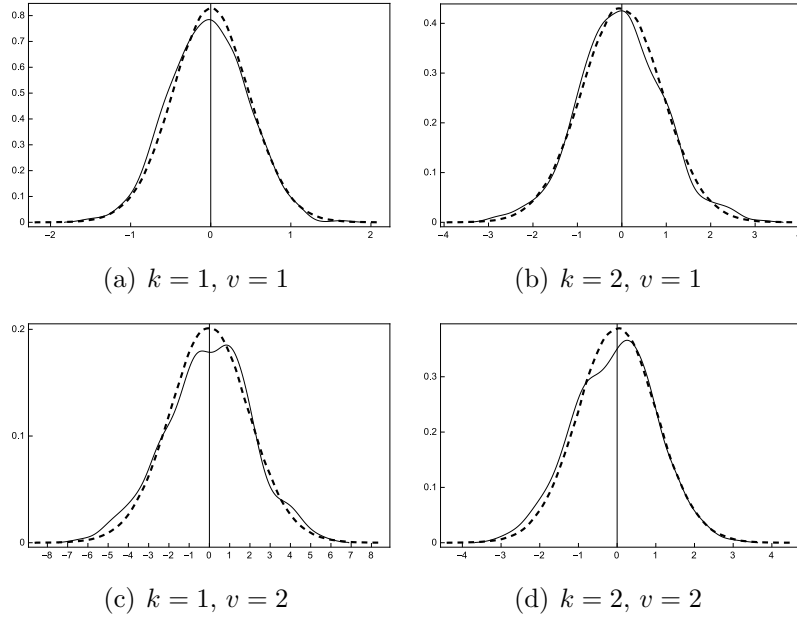


Figure 4.7: The wild bootstrap distribution (**thick dashed lines:** $\sqrt{N}(\hat{\phi}_k^\dagger(v) - \hat{\phi}_k(v))$) and the least squares estimators asymptotic distribution (thin lines: $\sqrt{N}(\hat{\phi}_k(v) - \phi_k(v))$) for model **M3** and sample size $n = 300$.

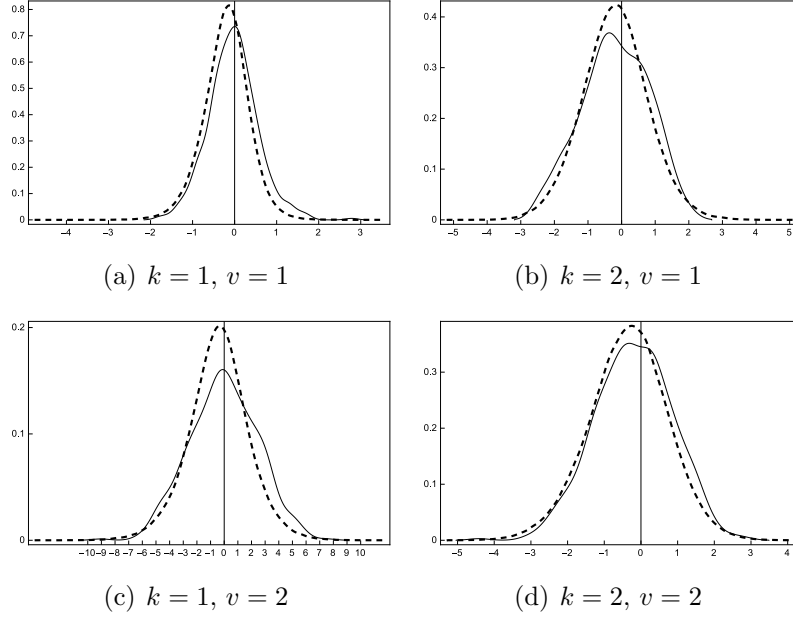


Figure 4.8: The residual bootstrap (**thick dashed lines:** $\sqrt{N}(\hat{\phi}_k^*(v) - \hat{\phi}_k(v))$) and the least squares estimators asymptotic distribution (**thin lines:** $\sqrt{N}(\hat{\phi}_k(v) - \phi_k(v))$) for model **M3** and sample size $n = 30$.

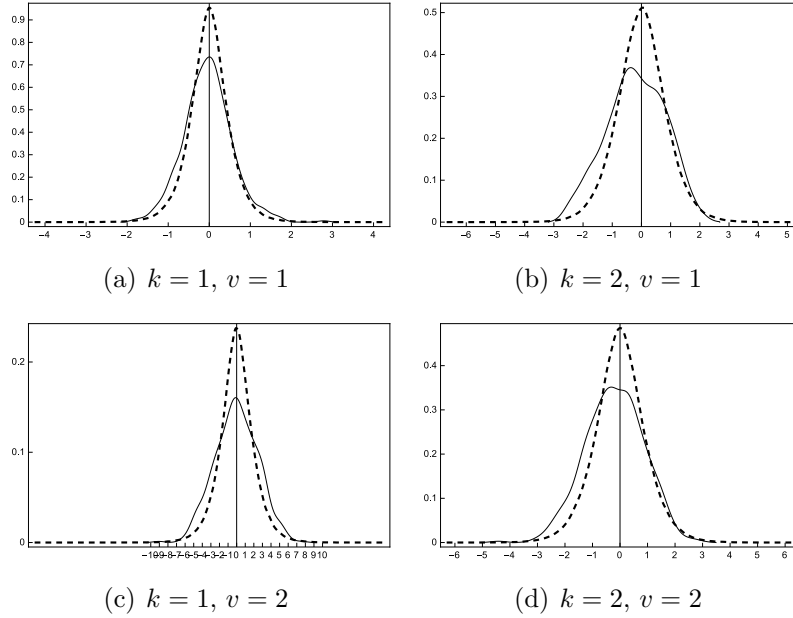


Figure 4.9: The wild bootstrap (**thick dashed lines:** $\sqrt{N}(\hat{\phi}_k^\dagger(v) - \hat{\phi}_k(v))$) and the least squares estimators asymptotic distribution (**thin lines:** $\sqrt{N}(\hat{\phi}_k(v) - \phi_k(v))$) for model **M3** and sample size $n = 30$.

		ACP		
		Model M1		
bootstrap	n	$\phi_1(1)$	$\phi_2(1)$	$\phi_3(1)$
residual	300	94.0%	95.0%	96.8%
wild	300	93.8%	94.6%	95.8%

Table 4.1: Actual coverage probabilities for model **M1**. Rows 4 and 5 contain results for the residual and the wild bootstrap, respectively. For both methods ACPs when sample sizes $n = 300$ are presented.

		ACP								
		Model M2								
bootstrap	n	$\phi_1(1)$	$\phi_2(1)$	$\phi_3(1)$	$\phi_1(2)$	$\phi_2(2)$	$\phi_3(2)$	$\phi_1(3)$	$\phi_3(3)$	$\phi_3(3)$
residual	300	92.2%	93.8%	92.8%	93.2%	92.6%	93.2%	92.2%	97.8%	95.4%
wild	300	91.6%	94.0%	92.6%	93.6%	91.2%	92.2%	92.4%	97.0%	94.2%

Table 4.2: Actual coverage probabilities for model **M2**. Rows 4 and 5 contain results for the residual and the wild bootstrap, respectively. For both methods ACPs when sample sizes $n = 300$ are presented.

		ACP			
		Model M3			
bootstrap	n	$\phi_1(1)$	$\phi_2(1)$	$\phi_1(2)$	$\phi_2(2)$
residual	30	85.8%	86.0%	87.4%	92.8%
	300	95.0%	92.2%	94.0%	94.8%
wild	30	85.4%	83.6%	85.4%	84.0%
	300	95.0%	92.8%	93.2%	94.0%

Table 4.3: Actual coverage probabilities for model **M3**. Rows 4-5 and 6-7 contain results for the residual and the wild bootstrap, respectively. For both methods ACPs for two sample sizes $n = 30$ and $n = 300$ are presented.

		average lengths of CI		
		Model M1		
bootstrap	n	$\phi_1(1)$	$\phi_2(1)$	$\phi_3(1)$
residual	300	0.38057	0.38084	0.37138
wild	300	0.38456	0.38297	0.37376

Table 4.4: Average lengths of confidence intervals for model **M1**. Rows 4 and 5 contain results for the residual and the wild bootstrap, respectively. For both methods the average lengths of confidence intervals when sample sizes $n = 300$ are presented.

		average lengths of CI								
		Model M2								
bootstrap	n	$\phi_1(1)$	$\phi_2(1)$	$\phi_3(1)$	$\phi_1(2)$	$\phi_2(2)$	$\phi_3(2)$	$\phi_1(3)$	$\phi_3(3)$	$\phi_3(3)$
residual	300	0.38665	0.39142	0.38653	0.38419	0.39462	0.40491	0.39923	0.36842	0.39067
wild	300	0.38390	0.39003	0.38654	0.38650	0.39608	0.40737	0.39400	0.36586	0.38926

Table 4.5: Average lengths of confidence intervals for model **M2**. Rows 4 and 5 contain results for the residual and the wild bootstrap, respectively. For both methods the average lengths of confidence intervals when sample sizes $n = 300$ are presented.

		average lengths of CI			
		Model M3			
bootstrap	n	$\phi_1(1)$	$\phi_2(1)$	$\phi_1(2)$	$\phi_2(2)$
residual	30	0.47100	1.94341	0.86581	0.95234
	300	0.15566	0.63470	0.29701	0.33056
wild	30	0.48486	1.97151	0.85201	0.91483
	300	0.15580	0.63781	0.29656	0.32909

Table 4.6: Average lengths of confidence intervals for model **M3**. Rows 4-5 and 6-7 contain results for the residual and the wild bootstrap, respectively. For both methods the average lengths of confidence intervals for two sample sizes $n = 30$ and $n = 300$ are presented.

		p -values of normality test			
		Model M3			
bootstrap	n	$\phi_1(1)$	$\phi_2(1)$	$\phi_1(2)$	$\phi_2(2)$
residual	30	0.0138	0.5276	0.1332	0.0312
	300	0.4796	0.1746	0.9999	0.5738
wild	30	0.5314	0.8883	0.5915	0.7716
	300	0.8925	0.0899	0.5294	0.9267

Table 4.7: p -values of the Lilliefors test that determinate normality of $\sqrt{N}(\hat{\phi}_k^*(v) - \hat{\phi}_k(v))$ and $\sqrt{N}(\hat{\phi}_k^\dagger(v) - \hat{\phi}_k(v))$ for model M3.

Chapter 5

Maximal concentration inequalities for regenerative and Harris recurrent Markov chains

The purpose of this chapter is to present Bernstein's, Hoeffding's and polynomial tail functional inequalities for regenerative Markov chains. Furthermore, we generalize these results and establish exponential and polynomial bounds for suprema of empirical processes over a class of functions \mathcal{F} whose size is controlled by its uniform entropy number. All constants involved in the bounds of the considered inequalities are given in an explicit form which can be advantageous in practical considerations. Finally, we establish bound of the expectation of the supremum of an empirical process in a Markovian setting.

This chapter gathers material presented in [22]; and partially [23] and [24]. Some additional remarks and explanations are added in order to make this exposition more comprehensive.

1 Preliminaries

Tail inequalities are a powerful tool that allows to control the probability that a random variable X exceeds some prescribed value t . They have been extensively investigated by many researchers due to the fact that they are a crucial step in deriving many results in numerous fields such as statistics, learning theory, discrete mathematics, statistical mechanics, information theory or convex geometry. There is a vast literature that provides a comprehensive overview of the theory of tail inequalities in the i.i.d. setting. We direct an interested reader to the classical books regarding this topic [15], [37] and [141].

In this thesis we are interested in exponential type tail inequalities. Such results are

important when studying limit results such as central limit theorems and law of iterated logarithm (refer to [15] for details). Exponential rate is also desired when considering convergence rates in probability. In this thesis, as mentioned in Chapter 1, we focus on applications of those probabilistic results to statistical learning theory. There is plenty of exponential results such as Bennet, Bernstein, Hoeffding, Prohorov or Petrov type exponential inequalities to name just a few (see [15] and [118]). In this thesis we focus on studying Bernstein and Hoeffding type inequalities. For completeness of exposition the aforementioned bounds in the i.i.d. framework are provided. Firstly, we recall Hoeffding's inequality that was established in [79].

Theorem 16 (Hoeffding's inequality). *Let X_1, X_2, \dots, X_n be independent random variables such that $a_i \leq X_i \leq b_i$ ($i = 1, \dots, n$), then for $t > 0$*

$$\mathbb{P} \left(\left| \sum_{i=1}^n (X_i - \mathbb{E}X_i) \right| \geq t \right) \leq 2 \exp \left(- \frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2} \right).$$

In what follows, we provide Bernstein's inequality which can be applied to unbounded random variables (see [37]).

Theorem 17 (Bernstein's inequality). *Let X_1, \dots, X_n be independent random variables such that, for all integers $p \geq 2$, there exists σ_l^2 such that*

$$\mathbb{E}|X_l|^p \leq p! R^{p-2} \sigma_l^2 / 2 \text{ for all } l \in \{1, \dots, n\}.$$

Then, for all $t > 0$,

$$\mathbb{P} \left(\left| \sum_{i=1}^n (X_l - \mathbb{E}X_l) \right| \geq t \right) \leq 2 \exp \left(- \frac{t^2}{2(\sigma^2 + Rt)} \right),$$

where $\sigma^2 = \sum_{l=1}^n \sigma_l^2$.

In this thesis we also investigate tail inequalities in the case when the random variables X_1, \dots, X_n do not possess exponential moments. We are particularly interested in the Rosenthal type inequalities which essentially require the finiteness of moments $\mathbb{E}|X_k|^p$. For the reader's convenience Rosenthal's inequality for independent random variables is recalled below (see Theorem 2.9 in [118]).

Theorem 18 (Rosenthal's inequality). *Let X_1, \dots, X_n be independent and centered random variables and assume that $\mathbb{E}|X_i|^p < \infty$ for all $p \geq 2$. Then,*

$$\mathbb{E} \left(\left| \sum_{i=1}^n X_i \right|^p \right) \leq c_p \left(\sum_{i=1}^n \mathbb{E}|X_i|^p + \left(\sum_{i=1}^n \mathbb{E}X_i^2 \right)^{p/2} \right)$$

for constant $c_p = \max(p^p, p^{p/2+1} e^p \int_0^\infty x^{p/2-1} (1-x)^{-p} dx)$.

The proof of the above theorem can be found in [118]. Below we provide a slight modification of original Rosenthal's inequality for i.i.d. random variables. We provide the polynomial tail inequality in the similar form as in Theorem 2.10 in [118].

Corollary 1 (Polynomial tail inequality). *Let X_1, \dots, X_n be i.i.d. and centered random variables and assume that $\mathbb{E}|X_i|^p < \infty$ for all $p \geq 2$. Then,*

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n X_i\right| \geq x\right) \leq \frac{C_p \mathbb{E}|X_1|^p}{x^p n^{p/2}}$$

for constant $C_p = 2 \max(p^p, p^{p/2+1} e^p \int_0^\infty x^{p/2-1} (1-x)^{-p} dx)$.

Proof. A proof of Corollary 1 can be deduced from the proof of Theorem 2.10 in [118]. Indeed, one can observe that

$$\left(\sum_{i=1}^n \mathbb{E}[X_i]^2\right)^{p/2} \leq \sum_{i=1}^n \mathbb{E}[X_i]^p$$

and constant $C_p = 2 \max(p^p, p^{p/2+1} e^p \int_0^\infty x^{p/2-1} (1-x)^{-p} dx)$. A reader is directed to the proof of Theorem 2.10 in [118] for details. \square

In this thesis we are interested in maximal versions of Bernstein's, Hoeffding's and polynomial tail inequalities. When establishing tail bounds for suprema of empirical processes over a class of functions \mathcal{F} , one needs to control the size of \mathcal{F} . In this framework we measure the complexity of \mathcal{F} via covering number and uniform entropy number. We briefly recall that covering number $N_p(\epsilon, Q, \mathcal{F})$ is interpreted as the minimal number of balls

$$\{g : \|g - f\|_{L_p(Q)} < \epsilon\}$$

of radius ϵ needed to cover \mathcal{F} in the norm $L_p(Q)$ and Q is a discrete probability measure. Moreover, uniform entropy number is defined as

$$N_p(\epsilon, \mathcal{F}) = \sup_Q N_p(\epsilon, Q, \mathcal{F}),$$

where the supremum is taken over all discrete probability measures Q . Refer to [87] and [141] to learn more about the properties of the aforementioned quantities.

In this chapter we also study bounds for expectation of suprema of empirical processes over classes of functions \mathcal{F} whose size is controlled via covering and uniform entropy number. Such bounds are particularly useful when one wants to select a model via some penalization criterion with penalty term depending on a complexity of the whole collection of models. We recall below a result for the i.i.d. processes proposed in [70] (see Theorem 3.1 therein).

Theorem 19. Assume that \mathcal{F} has an envelope $F < 1$. Suppose further that

$$\log N_2(\epsilon, Q, \mathcal{F}) \leq H(\|F\|_{L^2(Q)}/\epsilon)$$

for some non-decreasing function H independent of Q and fulfilling some mild conditions. Let X_1, \dots, X_n be a sequence of i.i.d. random variables with an arbitrary distribution \mathbb{P} . Then we have

$$\mathbb{E} \left[\sup_{f \in \mathcal{F}} \left| \sum_{i=1}^n f(X_i) - \mathbb{E} f(X_i) \right| \right] \leq C(H) \left[\sigma \sqrt{n H(2\sigma^{-1} \|F\|_{L^2(\mathbb{P})})} + H(2\sigma^{-1} \|F\|_{L^2(\mathbb{P})}) \right],$$

where $C(H) > 0$ is a constant depending on H and $\sigma \in (0, 1]$ is such that

$$\sup_{f \in \mathcal{F}} \text{Var}(f(X_1)) \leq \sigma^2.$$

The wealth of possible applications of tail inequalities has naturally led to development of this theory in a dependent setting.

Remark 18. We emphasize that when deriving tail inequalities for Markov chains (or any other process with some dependence structure) one can not expect to recover fully the classical results from the i.i.d. case. The goal is then to get some counterparts of the inequalities for i.i.d. random variables with some extra terms that appear in the bound as a consequence of a Markovian structure of the considered process.

In the recent years such (non-)asymptotic results have been obtained for Markov chains via many approaches: martingale arguments (see [74], where Hoeffding's inequality for uniformly ergodic Markov chains has been presented), coupling techniques (see [41] and [53]). In fact, it is proved in [53] that Hoeffding's inequality holds when a Markov chain is geometrically ergodic and thus weakened the assumptions imposed on the Markov chain in [74]. Winterberger [148] has generalized the result in [53] by showing that Hoeffding's inequality is also valid for unbounded functions of geometrically ergodic Markov chains provided that the sum is correctly self-normalized. Paulin [114] has presented McDiarmid's inequality for Markov chains using Merton coupling and spectral methods. Adamczak [1], Adamczak and Bednorz [2], Cléménçon [47] and Bertail and Cléménçon [28] have obtained exponential inequalities for ergodic Markov chains via regenerative techniques.

There is a fair amount of results generalizing Rosenthal type bound from the independent setting to dependent case. Shao [129] established Rosenthal bound for ρ -mixing processes, Peligrad [115] and Utev [139] obtained the extensions for ϕ -mixing sequences, Viennet [146] studied the case when processes are β -mixing. Merlevède and Peligrad [105] established

Rosenthal type bound for stationary sequences including martingales (see also the aforementioned work for more exhaustive literature review). To the best of our knowledge there is no Rosenthal type inequalities for Markov chains obtained via regenerative techniques that precede our work presented in [23] and [24].

Most of the existing bounds for expected value of supremum of empirical process are obtained for Gaussian or sub-Gaussian processes (see [37], Chapter 13 for an extensive overview). Baraud [16] provided an upper bound for the expectation of supremum of the empirical process over elements of class \mathcal{F} having a small variance. This result can be applied when \mathcal{F} is a VC subgraph or a VC major class (see [16] for more details). The bound involves explicit constants and it is not expressed in terms of entropy of \mathcal{F} . Ready to use result was obtained in [97], where a bound for more general class of empirical processes admitting weak moment conditions imposed of functions that belong to \mathcal{F} is derived.

We point out that the upper bounds for tail inequalities and the bound for expectation of suprema of empirical processes in a Markovian setting presented in this thesis depend on many parameters (which are fairly easy to estimate/compute) such as moments of return times to (pseudo-)atom A , moments of length of blocks, etc. We present them in a detailed form in the subsequent sections. However, to give a reader a general overview of obtained results we provide the bounds in a general (and somewhat simplified) form below. The detailed conditions imposed on chain X are omitted here and stated in further sections. For the sake of simplicity we provide the results solely in the atomic regenerative case (we formulate the inequalities for Harris recurrent Markov chains in further sections). Let $\sigma_m^2 = \max_{f \in \mathcal{F}} \sigma^2(f) > \eta > 0$.

- **Bernstein type maximal inequality** Assume that $N_1(\epsilon, \mathcal{F}) < \infty$. Then, under exponential block moment conditions and exponential moments of return times to set A , we have for any $x > 0$, $0 < \epsilon < x/2$ and for all $n \geq 1$

$$\mathbb{P}_\nu \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(X_i) - \mu(f) \right| \geq x \right] \leq N_1(\epsilon, \mathcal{F}) K_1 \left\{ \exp \left[\frac{-n(x - 2\epsilon)^2}{K_2(\sigma_m^2 + K_3(x - 2\epsilon))} \right] \right\},$$

where K_1 , K_2 and K_3 are positive parameters specified later in this chapter.

- **Hoeffding type maximal inequality** Assume that $N_1(\epsilon, \mathcal{F}) < \infty$. Suppose further that the class of functions \mathcal{F} is uniformly bounded. Then, under exponential block moment conditions and exponential moments of return times to set A , we have for any $x > 0$, $0 < \epsilon < x/2$ and for all $n \geq 1$

$$\mathbb{P}_\nu \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(X_i) - \mu(f) \right| \geq x \right] \leq N_1(\epsilon, \mathcal{F}) L_1 \left\{ \exp \left[-\frac{n(x - 2\epsilon)^2}{L_2 D^2} \right] \right\},$$

where D is a constant such that $\forall f \in \mathcal{F} |f| < D$ and L_1 and L_2 are positive parameters specified later in this chapter.

- **Polynomial tail maximal inequality** Assume that $N_1(\epsilon, \mathcal{F}) < \infty$. Suppose further that the p -th block moment and p -th moment of return times to the atom A are finite. Then, we have for any $x > 0$, $0 < \epsilon < x/2$ and for all $n \geq 1$

$$\mathbb{P}_\nu \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(X_i) - \mu(f) \right| \geq x \right] \leq C_1 \frac{N_1(\epsilon, \mathcal{F})}{(x - 2\epsilon)^p n^{p/2}}$$

and C_1 is a positive parameter specified later in this chapter.

- **Bound for expectation of supremum of empirical processes** Assume that

$$\mathbb{E}_A[l(B_1)]^2 < \infty \text{ and } \mathbb{E}_A[F(B_1)]^2 < \infty,$$

where F is an envelope for \mathcal{F} . Moreover, suppose that uniform entropy number $N_1\left(\frac{\epsilon}{R_1}, \mathcal{F}\right) < \infty$. Then, for any $\epsilon > 0$ we have

$$\begin{aligned} & \mathbb{E}_A \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^{l_n} (f(B_i) - \mu(f(B_1))) \right| \right] \\ & \leq R_2 \left[\epsilon + \mathcal{N}\left(\frac{\epsilon}{R_1}, \mathcal{F}\right) \times \mathbb{E}_A[F(B_1)^2]^{1/2} \sqrt{\frac{2 \log N_1\left(\frac{\epsilon}{R_1}, \mathcal{F}\right)}{n}} \right], \end{aligned}$$

where R_1 and R_2 are positive constants that can be explicitly computed.

Remark 19. We indicated many times that our results are tailor-made for statistical learning applications (see next chapter). However, we want to emphasize that the above bounds stand on their own and can be used in other statistical problems.

All our results rely heavily on Montgomery-Smith's inequality. For the reader's convenience we provide Theorem 1.1.5 from [52] (page 6 therein) below (see also [108]).

Theorem 20 (Montgomery-Smith's (1993) inequality). *If X_i , $i \in \mathbb{N}$ are independent and identically distributed random variables, then for $1 \leq k \leq n < \infty$ and all $t > 0$ we have*

$$\mathbb{P} \left(\max_{1 \leq k \leq n} \left\| \sum_{i=1}^k X_i \right\| > t \right) \leq 9 \mathbb{P} \left(\left\| \sum_{i=1}^n X_i \right\| > t/30 \right).$$

Montgomery-Smith's inequality is particularly useful when dealing with random number of blocks l_n .

2 Bernstein and Hoeffding type deviation inequalities for Markov chains

In this section we firstly present simple tail inequalities for atomic regenerative Markov chains. We rely on the following decomposition of the sum $\sum_{i=1}^n f(X_i)$:

$$\sum_{i=1}^n f(X_i) = \sum_{i=1}^{l_n} f(B_i) + \Delta_n,$$

where

$$\Delta_n = \sum_{i=1}^{\tau_A} f(X_i) + \sum_{i=\tau_A(l_n-1)}^n f(X_i).$$

Furthermore, recall that

$$\sigma^2(f) = \frac{1}{\mathbb{E}_A(\tau_A)} \mathbb{E}_A \left(\sum_{i=1}^{\tau_A} \{f(X_i) - \mu(f)\} \right)^2$$

is the asymptotic variance.

2.1 Bernstein and Hoeffding type bounds for atomic regenerative Markov chains

In the following, we denote $\bar{f}(x) = f(x) - \mu(f)$. Moreover, we write respectively

$$\bar{f}(B_1) = \sum_{i=1}^{\tau_A} \bar{f}(X_i) \quad \text{and} \quad |\bar{f}|(B_1) = \sum_{i=1}^{\tau_A} |\bar{f}|(X_i).$$

We will work under following conditions.

- A1. (Bernstein's block moment condition) There exists a positive constant M_1 such that for any $p \geq 2$ and for every $f \in \mathcal{F}$

$$\mathbb{E}_A |\bar{f}(B_1)|^p \leq \frac{1}{2} p! \sigma^2(f) M_1^{p-2}. \quad (5.1)$$

- A2. (Non-regenerative block exponential moment assumption) There exists $\lambda_0 > 0$ such that for every $f \in \mathcal{F}$ we have

$$\mathbb{E}_\nu \left[\exp \left[\lambda_0 \left| \sum_{i=1}^{\tau_A} \bar{f}(X_i) \right| \right] \right] < \infty.$$

A3. (Exponential block moment assumption) There exists $\lambda_1 > 0$ such that for every $f \in \mathcal{F}$ we have

$$\mathbb{E}_A [\exp [\lambda_1 |\bar{f}| (B_1)]] < \infty.$$

Remark 20. Observe that assumption A1 implies the existence of an exponential moment of $\bar{f}(B_1)$:

$$\mathbb{E}_A [\exp(\lambda \bar{f}(B_1))] \leq \exp \left(\frac{\lambda^2/2}{1 - M_1|\lambda|} \right) \text{ for all } \lambda < \frac{1}{M_1}.$$

In what follows, we present two Bernstein type inequalities for Markov chains. Both results has been established in [22]. The crucial tool to obtain the bounds is Montgomery-Smith's inequality (see Theorem 20) which results in larger constants comparing to the i.i.d. setting.

Before we state the theorems, we will give a short discussion on already existing results for exponential inequalities for Markov chains.

Remarks 21. *Since there is plenty of results concerning exponential inequalities for Markov chains under many assumptions, it may be difficult to compare their strength (measured by assumptions imposed on the chain) and applicability. Thus, before we present the proofs of Theorem 22 and Theorem 23 , we make a short comparison of our result to already existing inequalities for Markov chains. We also strongly recommend seeing an exhaustive overview on the recent results of that type in [2].*

1. *The bounds obtained in this paper are related to the Fuk and Nagaev sharp bound inequality obtained in [28] which is based on the regenerative properties and decomposition of the chain. However, our techniques of proof differ and allow us to obtain a better rate in the main sub-Gaussian part of the inequality under the hypotheses. The proofs of the inequalities are simplified and do not require the partitioning arguments which were used in [28].*
2. *It is noteworthy that we do not impose stationarity condition of the considered Markov chain as in [41] and [53] or any restrictions on the starting point of the chain as in [53]. Moreover, Adamczak and Bednorz [2] use the assumption of strong aperiodicity for Harris Markov chain. We state a remark that this condition can be relaxed and we can only assume that Harris Markov chain is aperiodic (see Remark 29).*
3. *Many results concerning exponential inequalities for Markov chains are established for bounded functions f (see for instance [1], [47] and [53]). Our inequalities work for unbounded functions satisfying Bernstein's block moment condition. Adamczak and Bednorz [2] established exponential bound in case when f is unbounded, however our*

bounds allow faster rate of growth of functions from \mathcal{F} . More specifically, Adamczak and Bednorz [2] expressed the condition on the growth of the function f in terms of drift function. This condition is very strict (even simple AR(1) model does not satisfy it). Moreover, all terms involved in our inequalities are given by explicit formulas. Thus, the results can be directly used in practical considerations. Note also that all the constants are given in simple, easy to interpret form and they do not depend on other underlying parameters.

4. Winterberger [148] has established exponential inequalities in unbounded case extending the result of [53] to the case when the chain can start from any $x \in E$. However, the constant involved in the bound of the Theorem 2.1 in [148] (obtained for bounded and unbounded functions) is very large.
5. As indicated in [1], there are many exponential inequalities that satisfy spectral gaps (see for instance [65] and [99]). Spectral gap inequalities allow to recover Bernstein type inequality at its full strength. We need to mention that the geometric ergodicity assumption does not ensure in the non-reversible case that considered Markov chains admit a spectral gap (see Theorem 1.4 in [86]).

In what follows we present our Bernstein type inequality for Markov chains in the same form as in [22].

Theorem 22. *Assume that $X = (X_n)_{n \in \mathbb{N}}$ is a regenerative positive recurrent Markov chain. Then, under assumptions A1 – A3, we have for any $x > 0$ and for all $n \geq 1$*

$$\mathbb{P}_\nu \left[\left| \sum_{i=1}^n f(X_i) - \mu(f) \right| \geq x \right] \leq 18 \exp \left[- \frac{x^2}{2 \times 90^2 (n\sigma^2(f) + M_1 x / 90)} \right] \\ + C_1 \exp \left[- \frac{\lambda_0 x}{3} \right] + C_2 \exp \left[- \frac{\lambda_1 x}{3} \right],$$

where

$$C_1 = \mathbb{E}_\nu \left[\exp \left| \lambda_0 \sum_{i=1}^{\tau_A} \bar{f}(X_i) \right| \right] \quad \text{and} \quad C_2 = \mathbb{E}_A \left[\exp[\lambda_1 |\bar{f}| (B_1)] \right].$$

Remark 21. Observe that we do not impose a moment condition on $\mathbb{E}_A[\tau_A]^p < \infty$ for $p \geq 2$. At the first glance, this might be surprising since one usually assumes the existence of $\mathbb{E}_A[\tau_A]^2 < \infty$ when proving central limit theorem for regenerative Markov chains. A simple analysis of the proof of the central limit theorem in a Markovian case (see for instance [106]) reveals that it is sufficient to require only $\mathbb{E}_A[\tau_A] < \infty$ when we consider centered function \bar{f} instead of f .

Proof. Firstly, we consider the sum of random variables of the following form

$$Z_n(\bar{f}) = \sum_{i=1}^{l_n} \bar{f}(B_i). \quad (5.2)$$

Furthermore, we have that

$$S_n(\bar{f}) = Z_n(\bar{f}) + \Delta_n(\bar{f}).$$

Next, recall that l_n is random and correlated with blocks itself. In order to apply Bernstein's inequality for the i.i.d. random variables we apply Montgomery-Smith's inequality (see Theorem 20). It follows easily that

$$\mathbb{P}_A \left[\left| \sum_{i=1}^{l_n} \bar{f}(B_i) \right| \geq x/3 \right] \leq \mathbb{P}_A \left[\max_{1 \leq k \leq n} \left| \sum_{i=1}^k \bar{f}(B_i) \right| \geq x/3 \right] \leq 9 \mathbb{P}_A \left[\left| \sum_{i=1}^n \bar{f}(B_i) \right| \geq x/90 \right]$$

and under Bernstein's condition A1 we obtain

$$\mathbb{P}_A \left[\left| \sum_{i=1}^n \bar{f}(B_i) \right| \geq x/90 \right] \leq 2 \exp \left[- \frac{x^2}{2 \times 90^2 (M_1 x/90 + n \sigma^2(f))} \right].$$

Next, we want to control the remainder term Δ_n .

$$\Delta_n = \sum_{i=1}^{\tau_A} \bar{f}(X_i) + \sum_{i=\tau_A(l_n-1)}^n \bar{f}(X_i).$$

The control of Δ_n is simply guaranteed by Markov's inequality, i.e.

$$\mathbb{P}_\nu \left[\left| \sum_{i=1}^{\tau_A} \bar{f}(X_i) \right| \geq \frac{x}{3} \right] \leq \mathbb{E}_\nu \left[\exp \left| \lambda_0 \sum_{i=1}^{\tau_A} \bar{f}(X_i) \right| \right] \exp \left[- \frac{\lambda_0 x}{3} \right].$$

We deal similarly with the last term of Δ_n . We complement the data $1 + \tau_A(l_n) + 1$ by observations up to the next regeneration time $1 + \tau_A(l_n + 1)$ and obtain

$$\begin{aligned} \mathbb{P}_\nu \left[\left| \sum_{i=1+\tau_A(l_n)+1}^n \bar{f}(X_i) \right| \geq \frac{x}{3} \right] &\leq \mathbb{P}_\nu \left[\sum_{i=1+\tau_A(l_n)+1}^n |\bar{f}|(X_i) \geq \frac{x}{3} \right] \\ &\leq \mathbb{P}_\nu \left[\sum_{i=1+\tau_A(l_n)+1}^{1+\tau_A(l_n+1)} |\bar{f}|(X_i) \geq \frac{x}{3} \right] \\ &\leq \mathbb{E}_A \left[\exp[\lambda_1 |\bar{f}|(B_1)] \right] \exp \left[- \frac{\lambda_1 x}{3} \right]. \end{aligned}$$

□

It is noteworthy that although Montgomery-Smith's inequality allows to obtain easily Bernstein's bound for Markov chains, the constants are rather large. Interestingly, under an additional assumption on $\mathbb{E}_A[\tau_A]^p$ we can obtain Bernstein type inequality for regenerative Markov chains with much smaller constants for the dominating counterpart of the bound.

A4. (Block length moment assumption) There exists a positive constant M_2 such that for any $p \geq 2$

$$\mathbb{E}_A[\tau_A]^p \leq p! M_2^{p-2} \mathbb{E}_A[\tau_A^2] \quad \text{and} \quad \mathbb{E}_\nu[\tau_A]^p \leq p! M_2^{p-2} \mathbb{E}_\nu[\tau_A^2].$$

Before we formulate Bernstein's inequality for regenerative Markov chains we introduce a lemma which provides a bound for tail probability of $\sqrt{n} \left(\frac{l_n}{n} - \frac{1}{\mathbb{E}_A[\tau_A]} \right)$ which will be crucial for the proof of Bernstein's bound but also may be of independent interest.

Lemma 1. *Suppose that condition A4 holds. Then*

$$\mathbb{P}_\nu \left(n^{1/2} \left(\frac{l_n}{n} - \frac{1}{\mathbb{E}_A[\tau_A]} \right) \geq x \right)$$

is bounded by

$$\exp \left(-\frac{1}{2} \frac{(\mathbb{E}_A[\tau_A] x \sqrt{n} - 2\mathbb{E}_A[\tau_A])^2}{\left(E_\nu \tau_A^2 + \left(\frac{n}{\mathbb{E}_A[\tau_A]} + x \sqrt{n} \right) E_A \tau_A^2 \right) + (\mathbb{E}_A[\tau_A] x \sqrt{n} + E_\nu \tau_A) M_2 \left(E_\nu \tau_A^2 + \left(\frac{n}{\mathbb{E}_A[\tau_A]} + x \sqrt{n} \right) E_A \tau_A^2 \right)^{1/2}} \right).$$

Remark 22. Note that when $n \rightarrow \infty$, the dominating part in the exponential term is of order

$$\begin{aligned} & \frac{1}{2} \frac{\mathbb{E}_A[\tau_A]^2 x^2}{E_A \tau_A^2 / \mathbb{E}_A[\tau_A] + \mathbb{E}_A[\tau_A]^{1/2} x M_2 (E_A \tau_A^2)^{1/2}} + O(n^{-1/2}) \\ &= \frac{1}{2} \frac{\mathbb{E}_A[\tau_A]^2 x^2}{E_A \tau_A^2 / \mathbb{E}_A[\tau_A] (1 + \mathbb{E}_A[\tau_A] x M_2 (E_A \tau_A^2 / \mathbb{E}_A[\tau_A])^{-1/2})} + O(n^{-1/2}) \\ &= \frac{1}{2} \frac{(\mathbb{E}_A[\tau_A] x)^2 / (E_A \tau_A^2 / \mathbb{E}_A[\tau_A])}{(1 + \mathbb{E}_A[\tau_A] x M_2 (E_A \tau_A^2 / \mathbb{E}_A[\tau_A])^{-1/2})} + O(n^{-1/2}), \end{aligned}$$

thus we have a Gaussian tail with the right variance for moderate x and an exponential tail for large x and, in consequence, the constants are asymptotically optimal.

Proof of Lemma 1. Let τ_k be the time of the k -th visit to the atom A ($S \times \{1\}$ in the general case).

In the following we make use of the argument from [53] and observe that we have for any $k \leq n$

$$\begin{aligned}\mathbb{P}_\nu(l_n \geq k) &= \mathbb{P}_\nu(\tau_k \leq n) = \mathbb{P}_\nu\left(\sum_{i=1}^k \Delta\tau_i \leq n\right) \\ &= \mathbb{P}((\Delta\tau_1 - E_\nu\tau_A) + \sum_{i=2}^k (\Delta\tau_i - \alpha) \leq n - (k-1)\alpha - E_\nu\tau_A).\end{aligned}$$

It follows that if $x > \sqrt{n}(1 - \alpha^{-1})$ ($\alpha = E_A\tau_A$), then

$$\mathbb{P}_\nu\left(n^{1/2}\left(\frac{l_n}{n} - \frac{1}{\alpha}\right) \geq x\right) = 0$$

and if $0 < x \leq \sqrt{n}(1 - \alpha^{-1})$, then

$$\begin{aligned}\mathbb{P}_\nu\left(n^{1/2}\left(\frac{l_n}{n} - \frac{1}{\alpha}\right) \geq x\right) &= \mathbb{P}_\nu\left(l_n \geq \frac{n}{\alpha} + x\sqrt{n}\right) \\ &\leq \mathbb{P}_\nu\left(l_n \geq \left\lceil \frac{n}{\alpha} + x\sqrt{n} \right\rceil\right) \\ &\leq \mathbb{P}((\Delta\tau_1 - E_\nu\tau_A) + \sum_{i=2}^{\lceil \frac{n}{\alpha} + x\sqrt{n} \rceil} (\Delta\tau_i - \alpha) \leq n - (\lceil \frac{n}{\alpha} + x\sqrt{n} \rceil - 1)\alpha - E_\nu\tau_A),\end{aligned}$$

where $\lceil \cdot \rceil$ is the integer part.

Since $\frac{n}{\alpha} + x\sqrt{n} - 1 \leq \lceil \frac{n}{\alpha} + x\sqrt{n} \rceil \leq \frac{n}{\alpha} + x\sqrt{n}$, we get

$$\begin{aligned}n - (\lceil \frac{n}{\alpha} + x\sqrt{n} \rceil - 1)\alpha - E_\nu\tau_A &\leq n - (\frac{n}{\alpha} + x\sqrt{n} - 2)\alpha - E_\nu\tau_A \\ &= -\alpha x\sqrt{n} + 2\alpha - E_\nu\tau_A.\end{aligned}$$

It follows that

$$\mathbb{P}_\nu\left(n^{1/2}\left(\frac{l_n}{n} - \frac{1}{\alpha}\right) \geq x\right) \leq \mathbb{P}\left((\Delta\tau_1 - E_\nu\tau_A) + \sum_{i=2}^{\lceil \frac{n}{\alpha} + x\sqrt{n} \rceil} (\Delta\tau_i - \alpha) \leq -\alpha x\sqrt{n} + 2\alpha - E_\nu\tau_A\right),$$

where $\lceil \cdot \rceil$ is the integer part

Now, we can apply any Bennett's or Bernstein's inequality on these centered i.i.d. random variables to get an exponential bound. This can be done since we assumed A4. Note that other bounds (polynomial for instance) can be obtained under appropriate modifications of A4. In our case we get

$$\begin{aligned}&\mathbb{P}((\Delta\tau_1 - E_\nu\tau_A) + \sum_{i=2}^{\lceil \frac{n}{\alpha} + x\sqrt{n} \rceil} (\Delta\tau_i - \alpha) \leq -\alpha x\sqrt{n} + 2\alpha - E_\nu\tau_A) \\ &\leq \exp\left(-\frac{\frac{1}{2}(\alpha x\sqrt{n} - 2\alpha + E_\nu\tau_A^2)/S_n^2}{1 + (\alpha x\sqrt{n} - 2\alpha + E_\nu\tau_A)M_2/S_n}\right),\end{aligned}$$

where

$$S_n^2 = E_\nu \tau_A^2 + \left(\left\lfloor \frac{n}{\alpha} + x\sqrt{n} \right\rfloor - 1\right) E_A \tau_A^2.$$

The above bound can be reduced to

$$\exp \left(-\frac{1}{2} \frac{(\alpha x \sqrt{n} - 2\alpha)^2}{(E_\nu \tau_A^2 + (\frac{n}{\alpha} + x\sqrt{n}) E_A \tau_A^2) + (\alpha x \sqrt{n} + E_\nu \tau_A) M_2 (E_\nu \tau_A^2 + (\frac{n}{\alpha} + x\sqrt{n}) E_A \tau_A^2)^{1/2}} \right).$$

□

In what follows we present an alternative Bernstein type inequality for regenerative Markov chains, where under additional condition on the length of the blocks we can obtain much better inequality in terms of constants.

Theorem 23. *Assume that $X = (X_n)_{n \in \mathbb{N}}$ is a regenerative positive recurrent Markov chain. Then, under assumptions A1-A4 we have for any $a > 0$, any $x > 0$, any $N > 0$ and for all $n \geq 1$ that*

$$\begin{aligned} \mathbb{P}_\nu \left[\left| \sum_{i=1}^n f(X_i) - \mu(f) \right| \geq x \right] &\leq 2 \exp \left[\frac{-x^2}{2 \times 3^2 (1+a)^2 \left(\left\lfloor \frac{n}{\mathbb{E}_A[\tau_A]} \right\rfloor \sigma^2(f) + \frac{M_1}{3} \frac{x}{1+a} \right)} \right] \\ &+ 18 \exp \left[\frac{-a^2 x^2}{2 \times 90^2 (1+a)^2 \left(N \sqrt{n} \sigma^2(f) + \frac{M_1}{90} \frac{ax}{1+a} \right)} \right] \\ &+ \mathbb{P}_\nu \left(n^{1/2} \left[\frac{l_n}{n} - \frac{1}{\mathbb{E}_A[\tau_A]} \right] > N \right) + C_1 \exp \left[-\frac{\lambda_0 x}{3} \right] + C_2 \exp \left[-\frac{\lambda_1 x}{3} \right], \end{aligned} \quad (5.3)$$

where

$$C_1 = \mathbb{E}_\nu \left[\exp \left| \lambda_0 \sum_{i=1}^{\tau_A} \bar{f}(X_i) \right| \right] \quad \text{and} \quad C_2 = \mathbb{E}_A \left[\exp[\lambda_1 |\bar{f}|(B_1)] \right].$$

Remark 23. In the proof of Theorem 23 we are interested in bounding for some $t > 0$ the probability

$$\mathbb{P}_A \left[\left| \sum_{i=\min(\lfloor \frac{n}{\mathbb{E}_A[\tau_A]} \rfloor, l_n)}^{\max(\lfloor \frac{n}{\mathbb{E}_A[\tau_A]} \rfloor, l_n)} \bar{f}(B_i) \right| \geq t \right].$$

We control this quantity by using truncation argument for the total number of regeneration times l_n , i.e. for some $N > 0$ we cut $\sqrt{n} \left[\frac{l_n}{n} - \frac{1}{\mathbb{E}_A[\tau_A]} \right] \leq N$. The magnitude of N (we want it to be relatively small) is significant since it appears in the final bound for

$\mathbb{P}_\nu [|\sum_{i=1}^n f(X_i) - \mu(f)| \geq x]$. Observe that if we choose $N = \log(n)$, then by Lemma 1 we can see that

$$\mathbb{P}_\nu \left(n^{1/2} \left[\frac{l_n}{n} - \frac{1}{\mathbb{E}_A[\tau_A]} \right] \geq \log(n) \right) = o\left(\frac{1}{n}\right)$$

and in that case the second term in (5.3) remains small uniformly in x .

Proof. We start by the obvious observation that

$$\begin{aligned} \mathbb{P}_\nu \left[\left| \sum_{i=1}^n f(X_i) - \mu(f) \right| \geq x \right] &\leq \mathbb{P}_A \left[\left| \sum_{i=1}^{l_n} \bar{f}(B_i) \right| \geq x/3 \right] \\ &+ \mathbb{P}_\nu \left[\left| \sum_{i=1}^{\tau_A} \bar{f}(X_i) \right| \geq x/3 \right] + \mathbb{P}_A \left[\left| \sum_{i=\tau_A(l_n-1)}^n \bar{f}(X_i) \right| \geq x/3 \right]. \end{aligned} \quad (5.4)$$

Remark 24. Note that instead of dividing x by 3 in (5.4), one can use a different splitting to improve a little bit the final constants.

The bounds for the first and the last non-regenerative blocks can be handled the same way as in Theorem 22. Next, we observe that, for any $a > 0$, we have

$$\mathbb{P}_A \left[\left| \sum_{i=1}^{l_n} \bar{f}(B_i) \right| \geq x/3 \right] \leq \mathbb{P}_A \left[\left| \sum_{i=1}^{\lfloor \frac{n}{\mathbb{E}_A[\tau_A]} \rfloor} \bar{f}(B_i) \right| \geq \frac{x}{3(1+a)} \right] + \mathbb{P}_A \left[\left| \sum_{i=l_{n_1}}^{l_{n_2}} \bar{f}(B_i) \right| \geq \frac{ax}{3(1+a)} \right],$$

where $l_{n_1} = \min(\lfloor \frac{n}{\mathbb{E}_A[\tau_A]} \rfloor, l_n)$ and $l_{n_2} = \max(\lfloor \frac{n}{\mathbb{E}_A[\tau_A]} \rfloor, l_n)$. We observe that $\sum_{i=1}^{\lfloor \frac{n}{\mathbb{E}_A[\tau_A]} \rfloor} \bar{f}(B_i)$ is a sum of independent, identically distributed and sub-exponential random variables. Thus, we can directly apply Bernstein's bound and obtain

$$\mathbb{P}_A \left[\left| \sum_{i=1}^{\lfloor \frac{n}{\mathbb{E}_A[\tau_A]} \rfloor} \bar{f}(B_i) \right| \geq \frac{x}{3(1+a)} \right] \leq 2 \exp \left[\frac{-x^2}{2 \times 3^2(1+a)^2 \left(\lfloor \frac{n}{\mathbb{E}_A[\tau_A]} \rfloor \sigma^2(f) + M_1 x / 3(1+a) \right)} \right]. \quad (5.5)$$

The control of $\sum_{i=l_{n_1}}^{l_{n_2}} \bar{f}(B_i)$ is slightly more challenging due to the fact that l_n is random and correlated with the blocks itself. In what follows, we will rely on Montgomery-Smith's inequality. Notice however, that since we expect the number of terms in this sum to be at most of the order \sqrt{n} , this term will be much smaller than the leading term (5.5) and will

be asymptotically negligible. We have

$$\begin{aligned} \mathbb{P}_A \left(\left| \sum_{i=l_{n_1}}^{l_{n_2}} \bar{f}(B_i) \right| \geq \frac{ax}{3(1+a)} \right) &\leq \mathbb{P}_A \left[\left| \sum_{i=l_{n_1}}^{l_{n_2}} \bar{f}(B_i) \right| \geq \frac{ax}{3(1+a)}, \sqrt{n} \left[\frac{l_n}{n} - \frac{1}{\mathbb{E}_A[\tau_A]} \right] \leq N \right] \\ &\quad + \mathbb{P}_\nu \left(\sqrt{n} \left[\frac{l_n}{n} - \frac{1}{\mathbb{E}_A[\tau_A]} \right] > N \right) = I + II. \end{aligned} \quad (5.6)$$

Firstly, we will bound term I in (5.6) using Montgomery-Smith's inequality and the fact that if

$$\sqrt{n} \left[\frac{l_n}{n} - \frac{1}{\mathbb{E}_A[\tau_A]} \right] \leq N, \quad \text{then} \quad l_{n_2} - l_{n_1} \leq \sqrt{n}N.$$

Note that it is sufficient to consider just the case when $\left\lfloor \frac{n}{\mathbb{E}_A[\tau_A]} \right\rfloor < l_n$. In what follows we rely on the following observation

$$l_n = \sup \{ s : \sum_{i=1}^s l(B_i) \leq n \}.$$

Thus,

$$\begin{aligned} &\mathbb{P}_A \left[\left| \sum_{i=l_{n_1}}^{l_{n_2}} \bar{f}(B_i) \right| \geq \frac{ax}{3(1+a)}, \sqrt{n} \left[\frac{l_n}{n} - \frac{1}{\mathbb{E}_A[\tau_A]} \right] \leq N \right] \\ &= \sum_{k=1}^{N\sqrt{n}} \mathbb{P}_A \left[\left| \sum_{i=\left\lfloor \frac{n}{\mathbb{E}_A[\tau_A]} \right\rfloor}^{\left\lfloor \frac{n}{\mathbb{E}_A[\tau_A]} \right\rfloor + k} \bar{f}(B_i) \right| \geq \frac{ax}{3(1+a)}, l_n = \left\lfloor \frac{n}{\mathbb{E}_A[\tau_A]} \right\rfloor + k \right] \\ &= \sum_{k=1}^{N\sqrt{n}} \mathbb{P}_A \left[\left| \sum_{i=1}^k \bar{f}(B_i) \right| \geq \frac{ax}{3(1+a)}, \sum_{i=1}^{\left\lfloor \frac{n}{\mathbb{E}_A[\tau_A]} \right\rfloor + k} l(B_i) \leq n < \sum_{i=1}^{\left\lfloor \frac{n}{\mathbb{E}_A[\tau_A]} \right\rfloor + k+1} l(B_i) \right] \end{aligned}$$

and by exchangeability of the blocks we have

$$\begin{aligned}
& \sum_{k=1}^{N\sqrt{n}} \mathbb{P}_A \left[\left| \sum_{i=1}^k \bar{f}(B_i) \right| \geq \frac{ax}{3(1+a)}, \sum_{i=1}^{\lfloor \frac{n}{\mathbb{E}_A[\tau_A]} \rfloor + k} l(B_i) \leq n < \sum_{i=1}^{\lfloor \frac{n}{\mathbb{E}_A[\tau_A]} \rfloor + k+1} l(B_i) \right] \\
&= \sum_{k=1}^{N\sqrt{n}} \mathbb{P}_A \left[\left| \sum_{i=1}^k \bar{f}(B_i) \right| \geq \frac{ax}{3(1+a)}, l_n = \left\lfloor \frac{n}{\mathbb{E}_A[\tau_A]} \right\rfloor + k \right] \\
&= \sum_{k=1}^{N\sqrt{n}} \mathbb{P}_A \left[\left| \sum_{i=1}^{l_n - \lfloor \frac{n}{\mathbb{E}_A[\tau_A]} \rfloor} \bar{f}(B_i) \right| \geq \frac{ax}{3(1+a)}, l_n - \left\lfloor \frac{n}{\mathbb{E}_A[\tau_A]} \right\rfloor = k \right] \\
&= \mathbb{P}_A \left[\left| \sum_{i=1}^{l_n - \lfloor \frac{n}{\mathbb{E}_A[\tau_A]} \rfloor} \bar{f}(B_i) \right| \geq \frac{ax}{3(1+a)}, l_n - \left\lfloor \frac{n}{\mathbb{E}_A[\tau_A]} \right\rfloor \leq N\sqrt{n} \right].
\end{aligned}$$

Now, we use Montgomery-Smith's inequality and obtain

$$\begin{aligned}
& \mathbb{P}_A \left[\left| \sum_{i=1}^{l_n - \lfloor \frac{n}{\mathbb{E}_A[\tau_A]} \rfloor} \bar{f}(B_i) \right| \geq \frac{ax}{3(1+a)}, l_n - \left\lfloor \frac{n}{\mathbb{E}_A[\tau_A]} \right\rfloor \leq N\sqrt{n} \right] \\
&= \mathbb{P}_A \left(\max_{1 \leq k \leq N\sqrt{n}} \left| \sum_{i=1}^k \bar{f}(B_i) \right| \geq \frac{ax}{3(1+a)} \right) \\
&\leq 9\mathbb{P}_A \left(\left| \sum_{i=1}^{N\sqrt{n}} \bar{f}(B_i) \right| \geq \frac{ax}{90(1+a)} \right) \\
&\leq 18 \exp \left[\frac{-a^2 x^2}{2 \times 90^2 (1+a)^2 (N\sqrt{n} \sigma^2(f) + \frac{M_1}{90} \frac{ax}{1+a})} \right].
\end{aligned}$$

Lemma 1 allows to control term II . □

Remark 25. The use of Montgomery-Smith's inequality is generating much larger constants when comparing to the i.i.d. case (however, our bounds preserve the same rates as in the i.i.d. setting in terms of n). We do believe that one can use different technique to deal with the random number of blocks l_n to obtain smaller constants in the final bound and simultaneously keep the i.i.d. rates. Moreover, there should be emphasized that using Lemma 1, the large constants appear only in front of terms which are nonetheless asymptotically negligible.

2.2 Maximal concentration inequalities under uniform entropy

In this section we present maximal versions of Bernstein and Hoeffding type inequalities for atomic regenerative Markov chains. We control the size of class \mathcal{F} via covering and uniform entropy numbers.

In what follows, we state assumptions on the size of considered class of functions \mathcal{F} . Rather than considering assumptions A2 and A3, we impose the conditions on the first and the last non-regenerative blocks for an envelope F of \mathcal{F} .

A2'. (Non-regenerative block exponential moment assumption) There exists $\lambda_0 > 0$ such that

$$\mathbb{E}_\nu \left[\exp \left[2\lambda_0 \left| \sum_{i=1}^{\tau_A} \bar{F}(X_i) \right| \right] \right] < \infty.$$

A3'. (Exponential block moment assumption) There exists $\lambda_1 > 0$ such that

$$\mathbb{E}_A [\exp [2\lambda_1 |\bar{F}| (B_1)]] < \infty.$$

A5. (Uniform entropy number condition) $N_1(\epsilon, \mathcal{F}) < \infty$.

Remark 26. Assumption A5 is a typical condition imposed on class \mathcal{F} in empirical processes theory. This assumption is fulfilled by many classes of functions, for instance Besov or Sobolev type (see [141] for details).

Before we present Bernstein concentration type inequality for unbounded classes of functions, we introduce one more piece of notation. Let

$$\sigma_m^2 = \max_{f \in \mathcal{F}} \sigma^2(f) > \eta > 0.$$

Theorem 24. Assume that $X = (X_n)_{n \in \mathbb{N}}$ is a regenerative positive recurrent Markov chain. Then, under assumptions A1, A2', A3' and A5 and for any $x > 0$, $0 < \epsilon < x/2$ and for all $n \geq 1$ we have

$$\begin{aligned} \mathbb{P}_\nu \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(X_i) - \mu(f) \right| \geq x \right] &\leq N_1(\epsilon, \mathcal{F}) \left\{ 18 \exp \left[-\frac{(x - 2\epsilon)^2 n}{2 \times 90^2 (\sigma_m^2 + M_1(x - 2\epsilon)/90)} \right] \right. \\ &\quad \left. + C_1 \exp \left[-\frac{\lambda_0(x - 2\epsilon)n}{3} \right] + C_2 \exp \left[-\frac{\lambda_1(x - 2\epsilon)n}{3} \right] \right\}, \end{aligned} \quad (5.7)$$

where

$$C_1 = \mathbb{E}_\nu \left[\exp \left| 2\lambda_0 \sum_{i=1}^{\tau_A} F(X_i) \right| \right] \quad \text{and} \quad C_2 = \mathbb{E}_A [\exp[2\lambda_1 |F| (B_1)]]$$

and F is an envelope function for \mathcal{F} .

Remark 27. Notice that our bound depends on a notion of uniform entropy number over a certain class of probability measures. However, for some classes of functions, this uniformity holds naturally, see for instance [149].

Observe that if \mathcal{F} belongs to a ball of a Hölder space $\mathbb{C}^P(E')$ on a compact set E' of an Euclidean space endowed with the norm

$$\|f\|_{\mathbb{C}^P(E')} = \sup_{x \in E'} |f(x)| + \sup_{x_1 \in E', x_2 \in E'} \left(\frac{|f(x_1) - f(x_2)|}{d(x_1, x_2)^P} \right)$$

then we have

$$M = \sup_{x \in X} F(x) < \infty$$

as well as

$$L = \sup_{f, g \in \mathcal{F}, f \neq g} \sup_z \frac{|f(z) - g(z)|}{\|f - g\|_{\mathbb{C}^P(E')}} < \infty$$

so that we can directly control the empirical sum by the obvious inequality

$$\sup_{f, g \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(X_i) - g(X_i) \right| \leq L \|f - g\|_{\mathbb{C}^P(E')}.$$

We refer to [49] and [149] for more details. See also examples of such classes of functions used in statistical learning in the latter. It follows that if we replace the notion of uniform covering number $N_1(\varepsilon, \mathcal{F})$ with respect to the norm $\|\cdot\|_{L^1(Q)}$ by the covering numbers $N_{\mathbb{C}^P}(\varepsilon, \mathcal{F})$ with respect to $\|\cdot\|_{\mathbb{C}^P(E')}$ (which does not depend on underlying probability) provided that $N_1(\varepsilon, \mathcal{F})$ is replaced by $N_{\mathbb{C}^P}(\frac{\varepsilon}{L}, \mathcal{F})$ in the inequality, Theorem 24 still holds.

Proof of Theorem 24. We choose functions g_1, g_2, \dots, g_M , where $M = N_1(\epsilon, \mathcal{F})$ such that

$$\min_j \mathbb{Q} |f - \mu(f) - g_j + \mu(g_j)| \leq 2\epsilon \text{ for each } f \in \mathcal{F},$$

where Q is any discrete probability measure. We also assume that g_1, g_2, \dots, g_M belong to \mathcal{F} and satisfy conditions $A1, A2', A3'$. We write f^* for the g_j , where the minimum is achieved.

Next, by definition of uniform covering numbers we obtain

$$\begin{aligned} & \mathbb{P}_\nu \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n (f(X_i) - \mu(f)) \right| \geq x \right] \\ & \leq \mathbb{P}_\nu \left\{ \sup_{f \in \mathcal{F}} \left[\left| \frac{1}{n} \sum_{i=1}^n |f(X_i) - \mu(f) - f^*(X_i) + \mu(f^*)| \right| + \left| \frac{1}{n} \sum_{i=1}^n |f^*(X_i) - \mu(f^*)| \right| \right] \geq x \right\} \\ & \leq \mathbb{P}_\nu \left[\max_{j \in \{1, \dots, N_1(\epsilon, \mathcal{F})\}} \left| \frac{1}{n} \sum_{i=1}^n g_j(X_i) - \mu(g_j) \right| \geq x - 2\epsilon \right] \\ & \leq N_1(\epsilon, \mathcal{F}) \max_{j \in \{1, \dots, N_1(\epsilon, \mathcal{F})\}} \mathbb{P}_\nu \left\{ \left| \frac{1}{n} \sum_{i=1}^n g_j(X_i) - \mu(g_j) \right| \geq x - 2\epsilon \right\}. \end{aligned}$$

We set the notation that

$$\bar{g}_j = g_j - \mu(g_j).$$

In what follows, our reasoning is analogous as in the proof of Theorem 22. Instead of taking any $f \in \mathcal{F}$, we work with the functions $\bar{g}_j \in \mathcal{F}$. Thus, we consider now the processes

$$Z_n(\bar{g}_j) = \sum_{i=1}^{l_n} \bar{g}_j(B_i) \quad (5.8)$$

and

$$S_n(\bar{g}_j) = Z_n(\bar{g}_j) + \Delta_n(\bar{g}_j).$$

Under the assumptions $A1, A2'$ and $A3'$ for \bar{g}_j , we get the analogous to that from Theorem 22 Bernstein's bound for $Z_n(\bar{g}_j)$, namely

$$\mathbb{P}_A \left[\left| \frac{1}{n} \sum_{i=1}^{l_n} \bar{g}_j(B_i) \right| \geq x - 2\epsilon \right] \leq 18 \exp \left[- \frac{(x - 2\epsilon)^2 n}{2 \times 90^2 (\sigma^2(g_j) + M_1(x - 2\epsilon)/90)} \right]. \quad (5.9)$$

We find the upper bound for the remainder term $\Delta_n(\bar{g}_j)$ applying the same reasoning as in Theorem 22. Thus,

$$\mathbb{P}_\nu \left[\left| \frac{1}{n} \sum_{i=1}^{\tau_A} \bar{g}_j(X_i) \right| \geq \frac{x - 2\epsilon}{3} \right] \leq C_1 \exp \left[- \frac{\lambda_0(x - 2\epsilon)n}{3} \right] \quad (5.10)$$

and

$$\mathbb{P}_A \left[\left| \frac{1}{n} \sum_{i=\tau_A(l_n-1)}^n \bar{g}_j(X_i) \right| \geq \frac{x - 2\epsilon}{3} \right] \leq C_2 \exp \left[- \frac{\lambda_1(x - 2\epsilon)n}{3} \right], \quad (5.11)$$

where

$$C_1 = \mathbb{E}_\nu \left[\exp \left| \lambda_0 \sum_{i=1}^{\tau_A} \bar{g}_j(X_i) \right| \right] \quad \text{and} \quad C_2 = \mathbb{E}_A [\exp[\lambda_1 |\bar{g}_j| (B_1)]] .$$

Finally, observe that

$$\mathbb{E}_\nu \left[\exp \left| \lambda_0 \sum_{i=1}^{\tau_A} \bar{g}_j(X_i) \right| \right] \leq \mathbb{E}_\nu \left[\exp \left| 2\lambda_0 \sum_{i=1}^{\tau_A} F(X_i) \right| \right] < \infty$$

and

$$\mathbb{E}_A [\exp[\lambda_1 |\bar{g}_j| (B_1)]] \leq \mathbb{E}_A [\exp[2\lambda_1 |F| (B_1)]] < \infty$$

and insert them into (5.10) and (5.11) which yields the proof. \square

Below we will formulate a maximal version of Theorem 23.

Theorem 25. *Assume that $X = (X_n)_{n \in \mathbb{N}}$ is a regenerative positive recurrent Markov chain. Then, under assumptions A1, A2', A3', A4 – A5 and for any $x > 0$, any $0 < \epsilon < x/2$, any $N > 0$ and for all $n \geq 1$ we have*

$$\begin{aligned} & \mathbb{P}_\nu \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(X_i) - \mu(f) \right| \geq x \right] \\ & \leq N_1(\epsilon, \mathcal{F}) \left\{ 2 \exp \left[\frac{-(x - 2\epsilon)^2 n}{2 \times 3^2 (1+a)^2 \left(\frac{\sigma^2(f)}{\mathbb{E}_A[\tau_A]} + \frac{M_1}{3} \frac{x-2\epsilon}{1+a} \right)} \right] \right. \\ & \quad + 18 \exp \left[\frac{-a^2(x - 2\epsilon)^2 n^2}{2 \times 90^2 (1+a)^2 \left(N\sqrt{n}\sigma^2(f) + \frac{M_1}{90} \frac{a(x-2\epsilon)}{1+a} \right)} \right] \\ & \quad \left. + \mathbb{P}_\nu \left(n^{1/2} \left[\frac{l_n}{n} - \frac{1}{\mathbb{E}_A[\tau_A]} \right] > N \right) + C_1 \exp \left[-\frac{\lambda_0(x - 2\epsilon)n}{3} \right] + C_2 \exp \left[-\frac{\lambda_1(x - 2\epsilon)n}{3} \right] \right\}, \end{aligned}$$

where

$$C_1 = \mathbb{E}_\nu \left[\exp \left| 2\lambda_0 \sum_{i=1}^{\tau_A} F(X_i) \right| \right] \quad \text{and} \quad C_2 = \mathbb{E}_A [\exp[2\lambda_1 |F| (B_1)]] .$$

Proof. The proof is a combination of the proofs of Theorem 23 and Theorem 24. We deal with the supremum over \mathcal{F} the same way as in Theorem 24. Then we apply Theorem 23. \square

We can obtain even sharper upper bound when class \mathcal{F} is uniformly bounded. In the following, we will show that it is possible to get a Hoeffding type inequality and have a stronger control of moments of the sum $S_n(f)$ which is a natural consequence of uniform boundedness assumption imposed on \mathcal{F} .

A6. The class of functions \mathcal{F} is uniformly bounded, i.e. there exists a constant D such that

$$\forall f \in \mathcal{F} \quad |f| < D.$$

Theorem 26. *Assume that $X = (X_n)_{n \in \mathbb{N}}$ is a regenerative positive recurrent Markov chain. Then, under assumptions A1, A2', A3', A5 – A6 and for any $x > 0$, any $0 < \epsilon < x/2$ and for all $n \geq 1$ we have*

$$\begin{aligned} & \mathbb{P}_\nu \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(X_i) - \mu(f) \right| \geq x \right] \leq N_1(\epsilon, \mathcal{F}) \left\{ 18 \exp \left[-\frac{(x - 2\epsilon)^2 n}{2 \times 90^2 D^2} \right] \right. \\ & \quad \left. + C_1 \exp \left[-\frac{\lambda_0(x - 2\epsilon)n}{3} \right] + C_2 \exp \left[-\frac{\lambda_1(x - 2\epsilon)n}{3} \right] \right\}, \end{aligned} \tag{5.12}$$

where

$$C_1 = \mathbb{E}_\nu \exp |2\lambda_0 \tau_A D| \quad \text{and} \quad C_2 = \mathbb{E}_A \exp |2\lambda_1 l(B_1) D|.$$

Proof. The proof bears resemblance to the proof of Theorem 24, with a few natural modifications which are a consequence of the uniform boundedness of \mathcal{F} . \square

Remark 28. Under condition A4 and applying Lemma 1 we can obtain easily the bound with smaller constants, we follow the analogous way as in Theorem 25.

2.3 Bernstein and Hoeffding type tail inequalities for Harris recurrent Markov chains

It is a purpose of this subsection to show that Theorems 22, 24, 26 are also valid in Harris recurrent case under slightly modified assumptions. In what follows we will apply the regenerative approach also in a Harris recurrent case via the Nummelin's splitting technique. In order to generalize the results from previous subsections to a Harris recurrent framework, we impose necessary conditions on Harris chain $X_{\mathcal{M}}$.

AH1. (Bernstein's block moment condition) There exists a positive constant M_1 such that for any $p \geq 2$ and for every $f \in \mathcal{F}$

$$\sup_{y \in S} \mathbb{E}_y |\bar{f}(B_1)|^p \leq \frac{1}{2} p! \sigma^2(f) M_1^{p-2}. \quad (5.13)$$

AH2. (Non-regenerative block exponential moment assumption) There exists a constant $\lambda_0 > 0$ such that for every $f \in \mathcal{F}$ we have

$$\mathbb{E}_\nu \left[\exp \left| \lambda_0 \sum_{i=1}^{\tau_S} \bar{f}(X_i) \right| \right] < \infty.$$

AH3. (Exponential block moment assumption) There exists a constant $\lambda_1 > 0$ such that for every $f \in \mathcal{F}$ we have

$$\sup_{y \in S} \mathbb{E}_y [\exp[\lambda_1 |\bar{f}|(B_1)]] < \infty.$$

In order to shorten the notation we designate

$$\sup_{y \in S} \mathbb{E}_y [\tau_S] = \alpha_{\mathcal{M}} < \infty.$$

In what follows we will formulate Bernstein type inequality for unbounded classes of functions in the Harris recurrent case (equivalent of Theorem 22). Theorems 24 and 26 can be reformulated for Harris chains in a similar way.

Theorem 27. Assume that $X_{\mathcal{M}}$ is a Harris recurrent, strongly aperiodic Markov chain. Then, under assumptions AH1-AH3, for any $x > 0$ and for all $n \geq 1$ we have that

$$\begin{aligned} \mathbb{P}_{\nu} \left[\left| \sum_{i=1}^n f(X_i) - \mu(f) \right| \geq x \right] &\leq 18 \exp \left[-\frac{x^2}{2 \times 90^2 (n\sigma^2(f) + M_1 x/90)} \right] \\ &\quad + C_1 \exp \left[-\frac{\lambda_0 x}{3} \right] + C_2 \exp \left[-\frac{\lambda_1 x}{3} \right], \end{aligned} \quad (5.14)$$

where

$$C_1 = \mathbb{E}_{\nu} \left[\exp \left| \lambda_0 \sum_{i=1}^{\tau_S} \bar{f}(X_i) \right| \right] \quad \text{and} \quad C_2 = \sup_{y \in S} \mathbb{E}_y \left[\exp[\lambda_1 |\bar{f}|] (B_1) \right].$$

The proof of Theorem 27 is analogous to the proof of Theorem 22. We can obtain a bound with much smaller constants under an extra block moment condition.

AH4. (Block length moment assumption) There exists a positive constant M_2 such that for any $p \geq 2$

$$\sup_{y \in S} \mathbb{E}_y [\tau_S]^p \leq p! M_2^{p-2} \sup_{y \in S} E_y \tau_S^2 \quad \text{and} \quad E_{\nu} [\tau_S]^p \leq p! M_2^{p-2} E_{\nu} \tau_S^2.$$

Theorem 28. Assume that $X_{\mathcal{M}}$ is a Harris recurrent, strongly aperiodic Markov chain. Then, under assumptions AH1-AH4, we have for any $x > 0$, any $N > 0$ and for all $n \geq 1$

$$\begin{aligned} \mathbb{P}_{\nu} \left[\left| \sum_{i=1}^n f(X_i) - \mu(f) \right| \geq x \right] &\leq 2 \exp \left[\frac{-x^2}{2 \times 3^2 (1+a)^2 \left(\lfloor \frac{n}{\alpha} \rfloor \sigma^2(f) + \frac{M_1}{3} \frac{x}{1+a} \right)} \right] \\ &\quad + 18 \exp \left[\frac{-a^2 x^2}{2 \times 90^2 (1+a)^2 \left(N \sqrt{n} \sigma^2(f) + \frac{M_1}{90} \frac{ax}{1+a} \right)} \right] \\ &\quad + \mathbb{P}_{\nu} \left[n^{1/2} \left(\frac{l_n}{n} - \frac{1}{\alpha} \right) > N \right] \\ &\quad + C_1 \exp \left[-\frac{\lambda_0 x}{3} \right] + C_2 \exp \left[-\frac{\lambda_1 x}{3} \right], \end{aligned}$$

where

$$C_1 = \mathbb{E}_{\nu} \left[\exp \left| \lambda_0 \sum_{i=1}^{\tau_S} \bar{f}(X_i) \right| \right] \quad \text{and} \quad C_2 = \sup_{y \in S} \mathbb{E}_y \left[\exp[\lambda_1 |\bar{f}|] (B_1) \right].$$

Remark 29. In the Theorem 27 we assumed that $X_{\mathcal{M}}$ is strongly aperiodic. It is easy, however, to relax this assumption and impose only the aperiodicity condition on Harris chain by using the same trick as in [98]. Note that if $X_{\mathcal{M}}$ satisfies $\mathcal{M}(m, S, \delta, \Phi)$ for $m > 1$, then the blocks of data are 1-dependent. Denote by $\bar{S} = S \cup \{*\}$, where $\{*\}$ is an ideal point

which is not in S . Next, we define a pseudo-atom $\alpha_{\overline{\mathcal{M}}} = \overline{S} \times \{1\}$. In order to impose only aperiodicity in this case it is sufficient to consider two processes $\{E_i\}$ and $\{O_i\}$ such that

$$O_i = f(X_i) \text{ if } \tau_{\alpha_{\overline{\mathcal{M}}}}(2k+1) < i \leq \tau_{\alpha_{\overline{\mathcal{M}}}}(2k+2)$$

and $O_i = *$ otherwise

$$E_i = f(X_i) \text{ if } \tau_{\alpha_{\overline{\mathcal{M}}}}(2k) < i \leq \tau_{\alpha_{\overline{\mathcal{M}}}}(2k+1),$$

for some $k \geq 0$ and $E_i = *$. Every function $f : S \rightarrow \mathbb{R}$ will be considered as defined on \overline{S} with identification $f(*) = 0$ (see also [98] for more details concerning those two processes). Then, we prove Bernstein type inequality similarly as we prove Theorems 22 and 27 applying all the reasoning to $\{E_i\}$ and $\{O_i\}$ separately, yielding to a similar inequality up to an additional multiplicative constant 2.

3 Polynomial tail maximal concentration inequality for Markov chains

In this section we show polynomial tail concentration inequality for atomic regenerative Markov chains. We will relax the assumptions that exponential block moments and exponential moments of return times to the atom A must be finite. Then, using the same techniques as in the previous section it is possible to obtain polynomial bound for tail probability $\mathbb{P}_\nu \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n \bar{f} B_i \right| \geq x \right]$.

Let $p \geq 2$. Before we formulate polynomial tail inequality, we state the following assumptions for chain X .

Assumption 1. *We have:*

$$\mathbb{E}_A[\tau_A^p] < \infty \text{ and } \mathbb{E}_\nu[\tau_A^p] < \infty.$$

Recall that F is an envelope of class \mathcal{F} . We need also the following condition.

Assumption 2. *We have $\mathbb{E}_A[F(B_1)^p] < \infty$.*

We also impose a condition on the size of class \mathcal{F} .

Assumption 3. *We assume $\mathcal{N}_1(\epsilon, \mathcal{F}) < \infty$.*

Now we are ready to state our concentration result.

Theorem 29 (Polynomial tail maximal inequality for regenerative Markov chains). *Assume that Assumptions 1, 2 and 3 are satisfied by chain $X = (X_n)_{n \in \mathbb{N}}$. Then, we have for any $x > 0$, any $0 < \epsilon < x/2$, any $N > 0$ and for all $n \geq 1$ that*

$$\mathbb{P}_\nu \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n \bar{f}(B_i) \right| \geq x \right] \leq N_1(\epsilon, \mathcal{F}) \left\{ \frac{3^p \mathbb{E}_\nu [|\sum_{i=1}^{\tau_A} F(X_i)|]^p}{n^p(x-2\epsilon)^p} + \frac{3^p \mathbb{E}_A [F(B_1)]^p}{n^p(x-2\epsilon)^p} \right. \\ \left. + \frac{6^p C_p \mathbb{E}_A |F(B_1)|^p}{n^{p/2}(x-2\epsilon)^p} + \frac{6^p C_p \mathbb{E}_A |F(B_1)|^p N^p}{n^{3p/4}(x-2\epsilon)^p} + \mathbb{P}_\nu \left(n^{1/2} \left(\frac{l_n}{n} - \frac{1}{\mathbb{E}_A[\tau_A]} \right) \geq N \right) \right\},$$

where $C_p = 24 \max(p^p, p^{p/2+1} e^p \int_0^\infty x^{p/2-1} (1-x)^{-p} dx)$.

Remark 30. Note that

$$\mathbb{P}_\nu \left(n^{1/2} \left(\frac{l_n}{n} - \frac{1}{\mathbb{E}_A[\tau_A]} \right) \geq N \right)$$

may be controlled by Lemma 1. It is noteworthy that under Assumption 1, one can obtain polynomial bound

$$\mathbb{P}_\nu \left(n^{1/2} \left(\frac{l_n}{n} - \frac{1}{\mathbb{E}_A[\tau_A]} \right) \geq N \right) \leq \frac{4^p[2^p+1]}{\mathbb{E}_A[\tau_A]^p N^p} + \frac{4^p[2^p+1]}{N^{p/2} n^{p/4}}$$

by slightly modifying the proof of Lemma 1 (see Appendix in [22] for details).

The proof of Theorem 29 relies on the same techniques as in the previous section. We briefly recall main steps below.

Proof of Theorem 29. Let g_1, g_2, \dots, g_M be a collection of functions such $M = N_1(\epsilon, \mathcal{F})$ and

$$\min_j \mathbb{Q} |f - \mu(f) - g_j + \mu(g_1)| \leq 2\epsilon \text{ for each } f \in \mathcal{F},$$

where \mathbb{Q} is any discrete probability measure. We also assume that g_1, g_2, \dots, g_M belong to \mathcal{F} and satisfy Assumptions 1 and 2.

By the same arguments as in Theorem 24 we immediately obtain that

$$\mathbb{P}_\nu \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n (f(X_i) - \mu(f)) \right| \geq x \right] \leq N_1(\epsilon, \mathcal{F}) \max_{j \in \{1, \dots, N_1(\epsilon, \mathcal{F})\}} \mathbb{P}_\nu \left\{ \frac{1}{n} \left| \sum_{i=1}^n g_j(X_i) - \mu(g_j) \right| \geq x - 2\epsilon \right\}.$$

We set the notation that

$$\bar{g}_j = g_j - \mu(g_j).$$

In what follows, we obtain the bound for

$$\mathbb{P}_\nu \left\{ \frac{1}{n} \left| \sum_{i=1}^n \bar{g}_j(X_i) \right| \geq x - 2\epsilon \right\}. \quad (5.15)$$

We rely on the following decomposition of (5.15)

$$\begin{aligned} \mathbb{P}_\nu \left\{ \frac{1}{n} \left| \sum_{i=1}^n \bar{g}_j(X_i) \right| \geq x - 2\epsilon \right\} &\leq \mathbb{P}_\nu \left[\frac{1}{n} \left| \sum_{i=1}^n \bar{g}_j(X_i) \right| \geq (x - 2\epsilon)/3 \right] \\ &+ \mathbb{P}_A \left[\frac{1}{n} \left| \sum_{i=1}^{l_n} \bar{g}_j(B_i) \right| \geq (x - 2\epsilon)/3 \right] + \mathbb{P}_A \left[\frac{1}{n} \left| \sum_{i=\tau_A(l_n-1)}^n \bar{g}_j(X_i) \right| \geq (x - 2\epsilon)/3 \right]. \end{aligned}$$

The bounds for the first and the last non-regenerative blocks can be easily obtained using Markov's inequality. In order to control

$$\mathbb{P}_A \left[\frac{1}{n} \left| \sum_{i=1}^{l_n} \bar{g}_j(B_i) \right| \geq (x - 2\epsilon)/3 \right]$$

we apply the same reasoning as in Theorem 23, i.e.

$$\mathbb{P}_A \left[\frac{1}{n} \left| \sum_{i=1}^{l_n} \bar{g}_j(B_i) \right| \geq (x - 2\epsilon)/3 \right] \leq \mathbb{P}_A \left[\frac{1}{n} \left| \sum_{i=1}^{\lfloor \frac{n}{\mathbb{E}_A[\tau_A]} \rfloor} \bar{g}_j(B_i) \right| \geq x/6 \right] + \mathbb{P}_A \left[\frac{1}{n} \left| \sum_{i=l_{n_1}}^{l_{n_2}} \bar{g}_j(B_i) \right| \geq x/6 \right],$$

where $l_{n_1} = \min(\lfloor \frac{n}{\mathbb{E}_A[\tau_A]} \rfloor, l_n)$ and $l_{n_2} = \max(\lfloor \frac{n}{\mathbb{E}_A[\tau_A]} \rfloor, l_n)$. Next, since $\sum_{i=1}^{\lfloor \frac{n}{\mathbb{E}_A[\tau_A]} \rfloor} \bar{g}_j(B_i)$ are i.i.d., we apply Corollary 1 and get

$$\mathbb{P}_A \left[\frac{1}{n} \left| \sum_{i=1}^{\lfloor \frac{n}{\mathbb{E}_A[\tau_A]} \rfloor} \bar{g}_j(B_i) \right| \geq x/6 \right] \leq N_1(\epsilon, \mathcal{F}) \frac{C_p \mathbb{E}_A |F(B_1)|^p}{n^{p/2} (6(x - 2\epsilon))^p}.$$

Next, we proceed the same way as in Theorem 23. Indeed, for any $N > 0$ we have

$$\begin{aligned} \mathbb{P}_A \left(\frac{1}{n} \left| \sum_{i=l_{n_1}}^{l_{n_2}} \bar{g}_j(B_i) \right| \geq 6(x - 2\epsilon) \right) &\leq \mathbb{P}_A \left[\frac{1}{n} \left| \sum_{i=l_{n_1}}^{l_{n_2}} \bar{g}_j(B_i) \right| \geq 6(x - 2\epsilon), \sqrt{n} \left[\frac{l_n}{n} - \frac{1}{\mathbb{E}_A[\tau_A]} \right] \leq N \right] \\ &+ \mathbb{P}_\nu \left(\sqrt{n} \left[\frac{l_n}{n} - \frac{1}{\mathbb{E}_A[\tau_A]} \right] > N \right). \end{aligned} \quad (5.16)$$

Using analogous reasoning as in Theorem 24, we get that

$$\mathbb{P}_A \left[\frac{1}{n} \left| \sum_{i=l_{n_1}}^{l_{n_2}} \bar{g}_j(B_i) \right| \geq 6(x - 2\epsilon), \sqrt{n} \left[\frac{l_n}{n} - \frac{1}{\mathbb{E}_A[\tau_A]} \right] \leq N \right] \leq \frac{C_p \mathbb{E}_A |F(B_1)|^p N^p}{6^p n^{3p/4} (x - 2\epsilon)^p}.$$

□

It is noteworthy that under Assumption 1, the ergodicity rate of the chain X is at least subgeometric, polynomial namely, in the sense that $\sup_{h: \|h\|_\infty \leq 1} |h(X_n) - \mu(h)| = O(1/n^{p-1})$.
Polynomial tail maximal concentration inequality for Harris recurrent Markov chains

It is straightforward to obtain polynomial tail maximal inequality for Harris recurrent Markov chains. In order to do it we slightly modify conditions for chain X due to the fact that one must consider split chain $X^{\mathcal{M}}$ which can be decomposed into i.i.d. blocks. In what follows, let $p \geq 2$.

Assumption 4. *We have:*

$$\sup_{y \in S} \mathbb{E}_y[\tau_S^p] < \infty, \mathbb{E}_\nu[\tau_S^p] < \infty \quad \text{and} \quad \sup_{y \in S} \mathbb{E}_y[l(B_1)^p] < \infty.$$

We denote by F an envelope of class \mathcal{F} . We need also the following condition.

Assumption 5. *We have for all $f \in \mathcal{F}$:*

$$\sup_{y \in S} \mathbb{E}_y[f(B_1)^p] < \infty \quad \text{and} \quad \sup_{y \in S} \mathbb{E}_y[F(B_1)^p] < \infty.$$

Theorem 30 (Polynomial tail functional inequality for Harris Markov chains). *Let $(X_n)_{n \in \mathbb{N}}$ be a regenerative Markov chain taking its values in (E, \mathcal{E}) and Assumptions 4, 5 and 3 hold. Then, we have for any $x > 0$, $0 < \epsilon < x/2$, any $N > 0$ and for all $n \geq 1$*

$$\begin{aligned} \mathbb{P}_\nu \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n \bar{f}(B_i) \right| \geq x \right] &\leq N_1(\epsilon, \mathcal{F}) \left\{ \frac{3^p \mathbb{E}_\nu [|\sum_{i=1}^{\tau_S} F(X_i)|]^p}{n^p(x-2\epsilon)^p} + \frac{3^p \sup_{y \in S} \mathbb{E}_y [F(B_1)]^p}{n^p(x-2\epsilon)^p} \right. \\ &\quad \left. + \frac{6^p C_p \sup_{y \in S} \mathbb{E}_y |F(B_1)|^p}{n^{p/2}(x-2\epsilon)^p} + \frac{6^p C_p \sup_{y \in S} \mathbb{E}_y |F(B_1)|^p N^p}{n^{3p/4}(x-2\epsilon)^p} + \mathbb{P}_\nu \left(n^{1/2} \left(\frac{l_n}{n} - \frac{1}{\alpha} \right) \geq N \right) \right\} \end{aligned}$$

where $C_p = 24 \max(p^p, p^{p/2+1} e^p \int_0^\infty x^{p/2-1} (1-x)^{-p} dx)$ is a constant.

The proof of polynomial tail inequality for Harris chains is analogous to the proof of Theorem 29 and thus omitted.

4 Bound of the expectation of the supremum of an empirical process in a Markovian setting

Control of the fluctuation of empirical process is an important tool when one wants to obtain generalization bounds for learning algorithms. Bounds for the expectation of the supremum

of empirical process are particularly useful when one wants to select a model via some penalization criterion with penalty term depending on a complexity of the whole collection of models (see for instance [142], [143] and [144]).

To the best of our knowledge, there is no results using regeneration techniques that extend the bounds from the i.i.d. setting into atomic regenerative or Harris recurrent case. In this section we will present the bound of the expectation of the supremum of an empirical process which we established in [24]. Our result is stated for general classes of functions (not necessarily VC), class \mathcal{F} does not have to be bounded.

Theorem 31. *Assume that $(X_n)_{n \in \mathbb{N}}$ is a regenerative Markov chain, such that*

$$\mathbb{E}_A[l(B_1)]^2 < \infty \text{ and } \mathbb{E}_A[F(B_1)]^2 < \infty,$$

where F is an envelope for \mathcal{F} . Moreover, suppose that $N_1\left(\frac{\epsilon}{M\mathbb{E}_A[\tau_A]}, \mathcal{F}\right) < \infty$. Then, for any $\epsilon > 0$ and $M > 1$ we have

$$\begin{aligned} \mathbb{E}_A \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^{l_n} (f(B_i) - \mu(f(B_1))) \right| \right] &\leq 540 \mathbb{E}_A[F(B_1)^2]^{1/2} \frac{[\text{Var}[l(B_1)]]^{1/2}}{n^{1/2} \mathbb{E}_A[\tau_A^2]^{1/2} \sqrt{M-1}} \\ &+ 540 \left[\epsilon + N_1\left(\frac{\epsilon}{M\mathbb{E}_A[\tau_A]}, \mathcal{F}\right) \times \mathbb{E}_A[F(B_1)^2]^{1/2} \sqrt{\frac{2 \log N_1\left(\frac{\epsilon}{M\mathbb{E}_A[\tau_A]}, \mathcal{F}\right)}{n}} \right]. \end{aligned}$$

Proof. In order to deal with the random number of blocks l_n we use Montgomery-Smith's inequality

$$\begin{aligned} \mathbb{P}_\nu \left(\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^{l_n} (f(B_i) - \mu(f(B_1))) \right| \geq x \right) &\leq \mathbb{P}_A \left(\max_{k \leq n} \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^k (f(B_i) - \mu(f(B_1))) \right| \geq x \right) \\ &\leq 9 \mathbb{P}_A \left(\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(B_i) - \mu(f(B_1)) \right| \geq \frac{x}{30} \right) \end{aligned} \quad (5.17)$$

combined with formula $\mathbb{E}[|X|] = \int_0^\infty \mathbb{P}(X > t) dt$:

$$\mathbb{E}_A \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^{l_n} f(B_i) - \mu(f(B_1)) \right| \right] \leq 270 \mathbb{E}_A \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(B_i) - \mu(f(B_1)) \right| \right].$$

Ghost sample of regeneration blocks and randomization.

In the following, we denote by $B' = (B'_1, \dots, B'_n)$ an independent copy of $B = (B_1, \dots, B_n)$ (so called 'ghost') sample. Let $(\epsilon_1, \dots, \epsilon_n)$ be independent Rademacher variables. Let

$$\|l\|_{P_{B_n}} = \frac{\sum_{i=1}^n l(B_i)}{n \mathbb{E}_A[\tau_A]}.$$

Note that for any $M > 0$ we have that

$$\begin{aligned}
270 \mathbb{E}_A \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(B_i) - \mu(f(B_1)) \right| \right] &\leq 540 \mathbb{E}_\epsilon \mathbb{E}_B \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(B_i) \epsilon_i \right| \right] \\
&\leq 540 \mathbb{E}_\epsilon \mathbb{E}_B \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(B_i) \epsilon_i \right| \mathbb{I}_{\|\iota\|_{P_{B_n}} \leq M \mathbb{E}_A[\tau_A]} \right] \\
&\quad + 540 \mathbb{E}_\epsilon \mathbb{E}_B \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(B_i) \epsilon_i \right| \mathbb{I}_{\|\iota\|_{P_{B_n}} > M \mathbb{E}_A[\tau_A]} \right] \\
&= I + II.
\end{aligned}$$

Uniform covering for \mathcal{F} .

We define an uniform covering g_1, \dots, g_W , where $W = N_1 \left(\frac{x}{M \mathbb{E}_A[\tau_A]}, \mathcal{F} \right)$ and

$$\min_j \mathbb{Q} |f - \mu(f) - g_j + \mu(g_j)| \leq \epsilon \text{ for each } f \in \mathcal{F}$$

and \mathbb{Q} is any discrete probability measure. We also assume that g_1, g_2, \dots, g_W belong to \mathcal{F} and satisfy Assumption 1. We write f^* for g_j , where the minimum is achieved. Then,

$$\begin{aligned}
I &\leq 540 \mathbb{E}_\epsilon \mathbb{E}_B \left\{ \sup_{f \in \mathcal{F}} \left[\left| \frac{1}{n} \sum_{i=1}^n (f(B_i) - \mu(f) - f^*(B_i) + \mu(f^*)) \epsilon_i \right| + \left| \frac{1}{n} \sum_{i=1}^n (f^*(B_i) - \mu(f^*)) \epsilon_i \right| \right] \right\} \\
&\leq 540 \left[\epsilon + \mathbb{E}_\epsilon \mathbb{E}_B \left[N_1 \left(\frac{\epsilon}{M \mathbb{E}_A[\tau_A]}, \mathcal{F} \right) \max_{j \in 1, \dots, N_1 \left(\frac{\epsilon}{M \mathbb{E}_A[\tau_A]}, \mathcal{F} \right)} \left| \frac{1}{n} \sum_{i=1}^n g_j(B_i) \epsilon_i \right| \mathbb{I}_{\|\iota\|_{P_{B_n}} \leq M \mathbb{E}_A[\tau_A]} \right] \right].
\end{aligned} \tag{5.18}$$

In what follows we will use Massart's Finite Class Lemma (see [104], Lemma 5.2, page 300). For the reader's convenience, we recall this result below.

Lemma 2. *Let \mathcal{A} be some finite subset of \mathbb{R}^n . Let N denote the cardinality of \mathcal{A} and let $R = \sup_{a \in \mathcal{A}} [\sum_{i=1}^n a_i^2]^{1/2}$, then*

$$\mathbb{E} \left[\sup_{a \in \mathcal{A}} \sum_{i=1}^n a_i \epsilon_i \right] \leq R \sqrt{2 \log N}. \tag{5.19}$$

We bound (5.18) by applying directly (5.19):

$$\begin{aligned}
(5.18) &\leq \mathbb{E}_A \left[\epsilon + \max_{j \in \{1, \dots, N_1\left(\frac{\epsilon}{M\mathbb{E}_A[\tau_A]}, \mathcal{F}\right)\}} \left(\frac{1}{n} \sum_{i=1}^n (g_j(B_i)^2) \right)^{1/2} \times \sqrt{\frac{2 \log N_1\left(\frac{\epsilon}{M\mathbb{E}_A[\tau_A]}, \mathcal{F}\right)}{n}} \right] \\
&\leq 540 \left[\epsilon + N_1\left(\frac{\epsilon}{M\mathbb{E}_A[\tau_A]}, \mathcal{F}\right) \times \mathbb{E}_A[F(B_1)^2]^{1/2} \sqrt{\frac{2 \log N_1\left(\frac{\epsilon}{M\mathbb{E}_A[\tau_A]}, \mathcal{F}\right)}{n}} \right].
\end{aligned}$$

In what follows, we derive an upper bound for II .

$$\begin{aligned}
II &\leq 540 \mathbb{E}_\epsilon \mathbb{E}_B \left[\left(\frac{1}{n} \sum_{i=1}^n F(B_i) \right)^2 \right]^{1/2} \left[\mathbb{P}(\|l\|_{P_{B_n}} > M\mathbb{E}_A[\tau_A]) \right]^{1/2} \\
&= 540 \mathbb{E}_A[F(B_1)^2]^{1/2} \times \left[\mathbb{P}(\|l\|_{P_{B_n}} > M\mathbb{E}_A[\tau_A]) \right]^{1/2}.
\end{aligned}$$

Note that for $M > 1$ and by the fact that $\|l\|_{P_{B_n}} = \frac{\sum_{i=1}^n l(B_i)}{n\mathbb{E}_A[\tau_A]}$ we have

$$\begin{aligned}
(\mathbb{P}[\|l\|_{P_{B_n}} - 1 \geq M - 1])^{1/2} &= \left(\mathbb{P} \left[\frac{1}{n} \frac{\sum_{i=1}^n l(B_i) - \mathbb{E}_A[\tau_A]}{\mathbb{E}_A[\tau_A]} - 1 \geq M - 1 \right] \right)^{1/2} \\
&\leq \frac{[Var[l(B_1)]]^{1/2}}{n^{1/2} \mathbb{E}_A[\tau_A^2]^{1/2} \sqrt{M - 1}}
\end{aligned}$$

by Markov's inequality and the fact that $\mathbb{E}_A[l(B_1)]^2 < \infty$. □

5 Conclusion

In this chapter we presented Bernstein, Hoeffding and polynomial tail maximal inequalities for atomic regenerative and Harris recurrent Markov chains. We established the aforementioned bounds having in mind applications to statistical learning, however the results state on their own and may be applied to various statistical problems where the data exhibit Markovian structure. Moreover, we established bound of the expectation of the supremum of an empirical process in a Markovian setting.

Despite of many concentration results we felt the need to establish maximal type concentration inequalities tailor-made for our applications, i.e. they are established for non-stationary Markov processes and hold for unbounded classes of functions \mathcal{F} , recover rates from the i.i.d. framework and involve easy to interpret parameters in the bound.

One of the main difficulties when obtaining both asymptotic and non-asymptotic results for regenerative Markov chains is the random number of blocks l_n when we control random sum of blocks $\sum_{i=1}^{l_n} f(B_i)$. We dealt with this problem via Montgomery-Smith's inequality. However, this approach resulted in obtaining large constants in the final bounds comparing to the i.i.d. versions of studied inequalities. It is noteworthy that thanks to Lemma 1, the large constants are involved in terms which are asymptotically negligible in a final bound. However, we believe that using other techniques instead of Montgomery-Smith's bound may give smaller constants in the bounds.

Chapter 6

Minimum volume set estimation for Markovian data

In statistical learning theory, numerous works established non-asymptotic bounds assessing the generalization capacity of empirical risk minimizers under a large variety of complexity assumptions for the class of decision rules over which optimization is performed, by means of sharp control of uniform deviation of i.i.d. averages from their expectation, while fully ignoring the possible dependence across training data in general. It is the purpose of this chapter to show that similar results can be obtained when statistical learning is based on a data sequence drawn from a (Harris positive) Markov chain X , through the running example of estimation of *minimum volume sets* (*MV*-sets) related to X 's stationary distribution, an unsupervised statistical learning approach to anomaly/novelty detection. Based on maximal concentration inequalities from Chapter 5, we prove using the *regenerative method*, that learning rate bounds depend not only on the complexity of the class of candidate sets but also on the ergodicity rate of the chain X , expressed in terms of tail conditions for the length of the regenerative cycles.

1 Preliminaries

Empirical Risk Minimization (ERM in abbreviated form) is the main paradigm of statistical learning (see for instance *e.g.* [144]). It has been studied in several contexts, stipulating various complexity assumptions (involving *e.g.* VC dimension, Rademacher averages or metric entropies) for the collection of decision rules over which optimization is performed.

In a classical i.i.d. framework we consider i.i.d. sample $((x_1, y_1), \dots, (x_n, y_n))$ from a distribution $P(x, y)$. We want to investigate the unknown functional dependency between x_i

and $y_i \in \{0, 1\}$ (more generally, one can also consider $y_i \in \{0, 1, \dots, k\}$). More specifically, we would like to build a classifier function h which given a new input x' delivers best available approximation of y' by assessing the loss, or discrepancy

$$L(h) = P(h(x) \neq y)$$

between y and $h(x)$. In general, with the help of data, a good classifier is searched from some class of classifiers \mathcal{H} . In statistical learning framework, each classifier function h from \mathcal{H} is scored across all data points via risk functional

$$R(h) = \mathbb{E}_P[L(h(x), y)] = \int L(h(x), y) dP.$$

The task of finding the “best” classifier boils down to finding a function which minimizes the true risk. In practice however, the risk function can not be computed since P is unknown. The empirical risk minimization paradigm boils down to replacing the unknown distribution $P(x, y)$ by its empirical version $P_n(x, y)$. Then we consider empirical risk

$$R_n(h) = \mathbb{E}_n[L(h(x), y)] = \int L(h(x), y) dP_n(x, y) = \frac{1}{n} \sum_{i=1}^n L(h(x_i), y_i).$$

We choose a classifier h_n that minimizes empirical risk over \mathcal{H} , i.e.

$$h_n = \arg \inf_{h \in \mathcal{H}} R_n(h).$$

Whereas empirical risk minimization approach has been vastly investigated in the i.i.d. framework (see for instance [142] and [143]), very little attention has been paid to situations where training observations are dependent. Most works devoted to this issue assumes in general that the training dataset used to compute the empirical risk functional is composed of i.i.d. data and rely on accurate control of uniform deviation of estimated risks from their true values.

When empirical risk takes the form of a sample mean statistic, just like in classification, the flagship problem in statistical learning theory, concentration inequalities for empirical processes can be directly used to establish generalization bounds for empirical risk minimizers. Among the seldom works coping with learning theory in situations where training data exhibit a (possibly complex) dependence structure, one may mention [133] for statistical guarantees of ERM based on *weakly dependent data* (see also [4] for analogous results in the online framework) and [48] when the data used to train the decision rule are drawn using sampling/survey schemes.

The goal of this chapter is to show that the generalization ability of decision rules based on (Harris positive) Markovian data can be established. More precisely, as a case in point, we establish here generalization bounds for empirical minimum volume sets related to the stationary distribution $\mu(dx)$ of a positive recurrent Markov chain $X = (X_n)_{n \in \mathbb{N}}$ at a given level $\alpha \in (0, 1)$, based on a sample path of length n . Originally introduced in the seminal contribution [61], minimum volume sets (*MV*-sets in a shortened version) generalize the concept of quantile for multivariate probability distributions (see also [120]) and offers a non-parametric framework for (unsupervised) anomaly/novelty detection. As observed in [126], *MV*-set estimation can be cast in a learning framework very similarly to ERM in (supervised) classification. In this chapter we rely on concentration results from the previous chapter. More precisely, we use maximal concentration bounds in order to obtain generalization bounds for minimum volume set estimation problem. We point out that the regenerative method we rely on is far from being the sole technique to obtain learning rate bounds in *M*-estimation problems based on Markovian data. However, although such results can be obtained for instance by means of coupling results under appropriate decay rate assumptions for mixing coefficients (see *e.g.* [133], [3], [5], [132], [78], [91] or [55] in the context of *MV*-set estimation), the regenerative method imposes much less restrictions on the ergodicity properties of the chain than most alternative techniques. As shall be seen below, this approach is entirely tailored to the Markovian framework.

Additionally, another advantage of the approach (based on regenerative properties of Markov chain X) we develop here lies in its frequentist interpretability, in contrast with alternative methods: the learning rate is directly linked to the average number of cycles observed over a finite length trajectory, the latter playing the role of number of training examples in the i.i.d. statistical framework.

2 Minimum Volume Set Estimation

The notion of minimum volume set (*MV*-set) has been proposed in [61] in order to extend the definition of quantile for 1-dimensional probability distributions. Consider a probability distribution μ on a measurable space (E, \mathcal{E}) . Let $\alpha \in (0, 1)$ and λ be a σ -finite measure of reference on (E, \mathcal{E}) , any solution of the minimization problem (6.1)

$$\min_{\Omega \in \mathcal{E}} \lambda(\Omega) \text{ subject to } \mu(\Omega) \geq \alpha \tag{6.1}$$

is called a *MV*-set of level α . Throughout this chapter, the distribution μ is assumed to be absolutely continuous w.r.t. λ and denoted by

$$f(x) = (d\mu/d\lambda)(x)$$

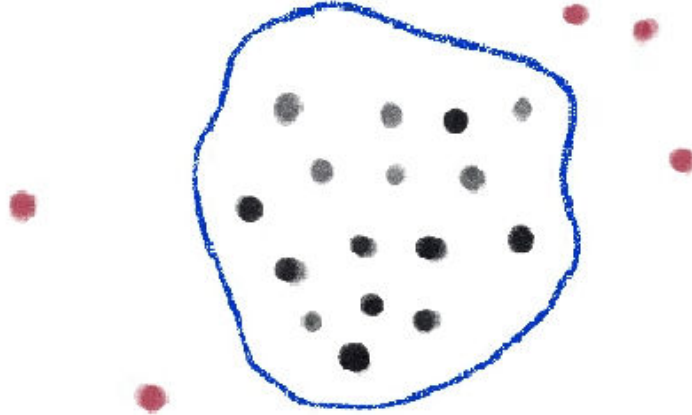
the related density. For any $\alpha \in (0, 1)$, under the assumptions that the density f is bounded and that the image of μ by f , denoted by μ_f , is a continuous probability on \mathbb{R}_+ , it is shown in [120] that the set

$$\Omega_\alpha^* = \{x \in E : f(x) > \gamma_\alpha\},$$

where γ_α is the unique number such that

$$\int_{f(x) > \gamma_\alpha} f(x) d\lambda(x) = \alpha$$

is the unique solution of the *MV*-set estimation problem (6.1). The *MV*-set estimation is a useful tool in unsupervised learning setting when one wants to detect anomalies (recall that in unsupervised learning framework an anomaly is a rare event).



The intuition behind minimum volume sets is as follows:

- for small values of α one captures the modes
- for large values of α ; samples that are in the *MV*-set will be considered as normal; and samples are not in the *MV*-set will be considered as anomalies.

In other words, we want to find the region where the data are mostly concentrated and anomalies which are assumed to be rare events, are located in the tail of the distribution.

Empirical MV-sets in the i.i.d. setting.

A level $\alpha \in (0, 1)$ being preliminarily fixed, a natural way of building estimates of the set Ω_α^* from i.i.d. data

$$X_1, \dots, X_n \sim \mu(dx)$$

consists in solving a statistical version of the constrained optimization problem (6.1)

$$\min_{\Omega \in \mathcal{E}} \lambda(\Omega) \text{ subject to } \hat{\mu}_n(\Omega) \geq \alpha - \psi_n \quad (6.2)$$

where the empirical distribution

$$\hat{\mu}_n = (1/n) \sum_{i=1}^n \delta_{X_i}$$

(or a smoothed counterpart of the latter) replaces the unknown probability measure μ , minimization is restricted to a subset \mathcal{G} of \mathcal{E} , expected to be sufficiently rich to include a reasonable approximation of Ω_α^* , and ψ_n plays the role of a *tolerance parameter* that controls how much the empirical mass can deviate from the targeted value α .

This approach, that essentially boils down to replacing the true (unknown) probability measure $\mu(dx)$ by its statistical counterpart is referred to as MV-ERM in [126]. The class \mathcal{G} is ideally made of sets $\Omega \in \mathcal{E}$ whose *volume* $\lambda(\Omega)$ can be efficiently computed or estimated, *e.g.* by Monte-Carlo simulation. Under usual complexity assumptions on the class \mathcal{G} combined with an appropriate choice of ψ_n , non-asymptotic statistical guarantees for solutions $\hat{\Omega}_\alpha$ of (6.2) have been obtained in [126]. Below we provide the definition of penalty for class \mathcal{G} . We provide it in the same form as in [126].

Definition 14 (Scott and Nowak [126], Definition 2). We say that ψ_n is a (distribution free) complexity penalty for class \mathcal{G} if and only if all distributions P and all $\delta \in (0, 1)$ we have with probability $1 - \delta$

$$\sup_{\Omega \in \mathcal{G}} (|P(\Omega) - P_n(\Omega)| - \psi_n(\delta)) \leq 0.$$

Taking a closer look at the above definition, one can notice that ψ allows to control the rate of uniform convergence of empirical distribution P_n to P for $\Omega \in \mathcal{G}$ (see [126], page 670 for more details). For the completeness of exposition we recall Theorem 3 from [126] that states a generalization bound for minimum volume set estimation problem in the i.i.d. framework.

Theorem 32 (Scott and Nowak [126], Theorem 3). *If ψ_n is a complexity penalty for class \mathcal{G} , then with probability $1 - \delta$ we have*

$$\lambda(\widehat{\Omega}_n) \leq \lambda(\Omega_\alpha^*) + \left\{ \inf_{\Omega \in \mathcal{G}: \mu(\Omega) \geq \alpha} \lambda(\Omega) - \lambda(\Omega_\alpha^*) \right\}$$

and

$$\mu(\widehat{\Omega}_n) \geq \alpha - 2\psi_n(\delta).$$

As indicated by Scott and Nowak [126], penalties for many classes \mathcal{G} are known.

Example 4 (VC classes). Assume that \mathcal{G} is a class of sets with VC dimension V . Penalty ψ for class \mathcal{G} is given by

$$\psi_n(\delta) = \sqrt{32 \frac{V \log n + \log(8/\delta)}{n}}.$$

In this chapter we aim to show that similar generalization bounds can be obtained when training data take the form of a sample path segment of a positive recurrent Markov chain with stationary distribution μ . A reader interested in investigating methods for solving practically the optimization problem (6.2) is referred to [63] and the references therein. In this chapter rate bounds involving the complexity of class \mathcal{G} in terms of covering numbers and the ergodicity rate of the chain in terms of speed at which the (extended) chain regenerates are established. The main tool used to derive these results is a functional polynomial tail inequality for positive Harris chains obtained in the previous chapter.

3 Minimum Volume Set Estimation - Generalization Results

Based on the results established in the previous chapter, we investigate the performance of solutions of the problem (6.2) when the empirical probability estimates $\widehat{\mu}_n(\Omega)$ are based on a Markovian trajectory of length $n \geq 1$. Firstly, we assume that Markov chain X is atomic regenerative (with the same communication properties as specified in Chapter 1). For simplicity, we assume that $E \subset \mathbb{R}^d$ with $d \geq 1$ and that $\lambda(dx)$ is the restriction of Lebesgue measure on E equipped with its Borel σ -algebra.

The stationary distribution is then given by

$$\mu(B) = \frac{1}{\mathbb{E}_A[\tau_A]} \mathbb{E}_A \left[\sum_{i=1}^{\tau_A} \mathbb{I}\{X_i \in B\} \right], \text{ for all } B \in \mathcal{E} \quad (6.3)$$

and its empirical counterpart based on the sequence X_1, \dots, X_n can be rewritten as: $\forall \Omega \in \mathcal{E}$,

$$\hat{\mu}_n(\Omega) = \frac{1}{n} \sum_{i=1}^{\tau_A} \mathbb{I}\{X_i \in \Omega\} + \frac{l_n - 1}{n} \left(\frac{1}{l_n - 1} \sum_{j=1}^{l_n-1} S_j(\Omega) \right) + \frac{1}{n} \sum_{i=1+\tau_A(l_n)}^n \mathbb{I}\{X_i \in \Omega\}. \quad (6.4)$$

Recall that $l_n = \sum_{i=1}^n \mathbb{I}\{X_i \in A\}$ denotes the number of visits to A (*regenerations*), the occupation time of the set Ω between the j -th and $(j+1)$ -th regeneration times is denoted by

$$S_j(\Omega) = \sum_{\tau_A(j) < i \leq \tau_A(j+1)} \mathbb{I}\{X_i \in \Omega\} \quad \text{for } j \geq 1$$

and $l_n > 1$ and with the usual convention that empty summation is equal to zero. Observe that since μ is a stationary distribution of atomic regenerative chain $\mathbb{X} = (X_1, \dots, X_n)$, the $S_j(\Omega)$'s are integrable i.i.d. r.v.'s with common mean $\mathbb{E}_A[\tau_A]\mu(A)$ and

$$l_n \sim n/\mathbb{E}_A[\tau_A] \quad \text{a.s. as } n \rightarrow +\infty$$

and the first and last terms in the equation above both asymptotically vanish with probability one. Hence, the random variables $S_j(\Omega)$ play the role of training observations in the subsequent analysis: the smaller the expected cycle length $\mathbb{E}[\tau_A]$, the larger the probability to observe a high number of training observations. However, as explained in Chapter 1, though asymptotically i.i.d., the block sums involved in (6.4) are not independent, the sum of their length being less than n . Except of the i.i.d. situation (notice that in such case, the whole state space E can be viewed as an atom), the frequency of visits to a candidate set Ω over the path X_1, \dots, X_n is not an i.i.d. average. Decomposition (6.4) is the main ingredient to control $\sup_{\Omega \in \mathcal{G}} |\hat{\mu}_n(\Omega) - \mu(\Omega)|$.

Let $p \geq 2$. In what follows we provide the conditions we impose on chain X that we will need in the subsequent analysis. Since we heavily rely on results from the previous chapter, the conditions stated here are related (or are the same) as imposed on atomic regenerative Markov chains as in Chapter 5. The moment condition provided below is the same as Assumption 1 in Chapter 5.

Condition 1. *We have: $\mathbb{E}_A[\tau_A^p] < \infty$ and $\mathbb{E}_\nu[\tau_A^p] < \infty$.*

It is noteworthy that under Condition 1, the ergodicity rate of the chain X is at least subgeometric, polynomial namely, in the sense that

$$\sup_{h: \|h\|_\infty \leq 1} |h(X_n) - \mu(h)| = O(1/n^{p-1}).$$

The above observation can be deduced from polynomial tail inequality (see Chapter 5).

The following condition is just a particular case of Assumption 3 from the previous chapter, when \mathcal{F} is a class of indicator functions. Let $r \geq 1$.

Condition 2. *The collection of indicator functions on E ,*

$$\mathcal{F} = \{\mathbb{I}\{x \in \Omega\} : \Omega \in \mathcal{G}\}$$

is a uniform Donsker class (relative to L_1) with polynomial uniform covering numbers, i.e. there exists a constant $c > 0$ s.t. $\forall \zeta > 0$,

$$N_1(\zeta, \mathcal{F}) \stackrel{\text{def}}{=} \sup_Q N(\zeta, \mathcal{F}, L_1(Q)) \leq c(1/\zeta)^r,$$

where the supremum is taken over the set of finitely discrete probability measures on (E, \mathcal{E}) .

As the following theorem shows, Condition 1 combined with Condition 2 allows to control the fluctuations of (6.4) uniformly over \mathcal{G} .

Theorem 33. *Suppose that Conditions 1-2 are fulfilled. For all*

$$\delta = \delta_1 + \delta_2 + \delta_3 + \delta_4 + \delta_5 + \delta_6 \in (0, 1),$$

we have with probability at least $1 - \delta$: $\forall n \geq 1$,

$$\sup_{\Omega \in \mathcal{G}} |\hat{\mu}_n(\Omega) - \mu(\Omega)| \leq \max \left[\left(\frac{D_1 \mathbb{E}_\nu[\tau_A^p]}{\delta_1 n^p} \right)^{1/(p+r)}, \left(\frac{D_1 \mathbb{E}_A[\tau_A^p]}{\delta_2 n^p} \right)^{1/(p+r)}, \left(\frac{D_2 \mathbb{E}_A[\tau_A^p]}{n^{p/2} \delta_3} \right)^{1/(p+r)}, \right. \\ \left. \left(\frac{D_3 \mathbb{E}_A[\tau_A^p]}{n^{3p/8} \delta_4} \right)^{1/(p+r)}, \left(\frac{D_4}{\mathbb{E}_A[\tau_A^p] \delta_5 n^{3p/8}} \right)^{1/r}, \left(\frac{D_4}{\delta_6 n^{7p/16}} \right)^{1/r} \right] \quad (6.5)$$

where D_1, D_2, D_3, D_4 are constants that can be explicitly computed and are specified in the proof.

Proof. The proof relies on Theorem 29:

$$\mathbb{P}_\nu \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n \bar{f}(B_i) \right| \geq \zeta \right] \leq N_1(\epsilon, \mathcal{F}) \left\{ \frac{3^p \mathbb{E}_\nu [|\sum_{i=1}^{\tau_A} F(X_i)|]^p}{n^p (\zeta - 2\epsilon)^p} + \frac{3^p \mathbb{E}_A [F(B_1)]^p}{n^p (\zeta - 2\epsilon)^p} \right. \\ \left. + \frac{6^p C_p \mathbb{E}_A |F(B_1)|^p}{n^{p/2} (\zeta - 2\epsilon)^p} + \frac{6^p C_p \mathbb{E}_A |F(B_1)|^p N^p}{n^{3p/4} (\zeta - 2\epsilon)^p} + \mathbb{P}_\nu \left(n^{1/2} \left(\frac{l_n}{n} - \frac{1}{\mathbb{E}_A[\tau_A]} \right) \geq N \right) \right\},$$

when \mathcal{F} is a class of indicator functions. In what follows we set $\epsilon = \zeta/4$ and $N = n^{3/8}$. Note that under Condition 2 we get

$$\begin{aligned} & \mathbb{P}_\nu \left\{ \sup_{\Omega \in \mathcal{G}} |\hat{\mu}_n(\Omega) - \mu(\Omega)| \geq \zeta \right\} \\ & \leq \left(\frac{1}{\zeta} \right)^r \left\{ \frac{D_1 \mathbb{E}_\nu[\tau_A^p]}{\zeta^p n^p} + \frac{D_1 \mathbb{E}_A[\tau_A^p]}{\zeta^p n^p} + \frac{D_2 \mathbb{E}_A[\tau_A^p]}{n^{p/2} \zeta^p} \right. \\ & \quad \left. + \frac{D_3 \mathbb{E}_A[\tau_A^p]}{n^{3p/8} \zeta^p} + \frac{D_4}{\mathbb{E}_A[\tau_A^p] n^{3p/8} \zeta^p} + \frac{D_4}{\zeta^p n^{7p/16}} \right\}, \end{aligned} \quad (6.6)$$

where $D_1 = c 4^r 12^p$, $D_2 = c 4^r 2^p C_p$, $D_3 = c 4^r 2^p C_p$, $D_4 = c 4^{r+p} (2^p + 1)$ and C_p is the same constant as in Corollary 1. Constant c comes from Condition 2.

Next, we solve the following equations for ζ :

$$\begin{aligned} \delta_1 &= \frac{D_1 \mathbb{E}_\nu[\tau_A^p]}{\zeta^p n^p}, \\ \delta_2 &= \frac{D_1 \mathbb{E}_A[\tau_A^p]}{\zeta^p n^p} \\ \delta_3 &= \frac{D_2 \mathbb{E}_A[\tau_A^p]}{n^{p/2} \zeta^p} \\ \delta_4 &= \frac{D_3 \mathbb{E}_A[\tau_A^p]}{n^{3p/8} \zeta^p} \\ \delta_5 &= \frac{D_4}{\mathbb{E}_A[\tau_A^p] n^{3p/8} \zeta^p} \\ \delta_6 &= \frac{D_4}{\zeta^p n^{7p/16}}. \end{aligned}$$

Standard calculations show the desired bound. □

A direct application of Theorem 33 to the MV -set estimation problem yields the following result.

Theorem 34. *Suppose that assumptions of Theorem 33 are fulfilled. Then, for all*

$$\delta = \delta_1 + \delta_2 + \delta_3 + \delta_4 + \delta_5 + \delta_6 \in (0, 1),$$

and for n large enough, any solution $\hat{\Omega}_n$ of (6.2) with

$$\begin{aligned} \psi_n(\delta) \stackrel{def}{=} \max & \left[\left(\frac{D_1 \mathbb{E}_\nu[\tau_A^p]}{\delta_1 n^p} \right)^{1/(p+r)}, \left(\frac{D_1 \mathbb{E}_A[\tau_A^p]}{\delta_2 n^p} \right)^{1/(p+r)}, \left(\frac{D_2 \mathbb{E}_A[\tau_A^p]}{n^{p/2} \delta_3} \right)^{1/(p+r)}, \right. \\ & \left. \left(\frac{D_3 \mathbb{E}_A[\tau_A^p]}{n^{3p/8} \delta_4} \right)^{1/(p+r)}, \left(\frac{D_4}{\mathbb{E}_A[\tau_A^p] n^{3p/8} \delta_5} \right)^{1/r}, \left(\frac{D_4}{\delta_6 n^{7p/16}} \right)^{1/r} \right] \end{aligned}$$

satisfies, with probability at least $1 - \delta$,

$$\lambda(\widehat{\Omega}_n) \leq \lambda(\Omega_\alpha^*) + \left\{ \inf_{\Omega \in \mathcal{G}: \mu(\Omega) \geq \alpha} \lambda(\Omega) - \lambda(\Omega_\alpha^*) \right\}$$

and

$$\mu(\widehat{\Omega}_n) \geq \alpha - 2\psi_n(\delta).$$

Constants D_1, D_2, D_3, D_4 are given in the proof of Theorem 33.

Remark 31. Note that Theorem 34 works for n of the magnitude of $\psi_n(\delta)$ in order to ensure that

$$\alpha - 2\psi_n(\delta) > 0.$$

Proof. The proof is analogous to that of Theorem 11 in [126]. Let $K = (X_1, \dots, X_n)$ be a Markovian sample. Define

$$\begin{aligned} \Theta_\mu &= \{K : \mu(\widehat{\Omega}_n) < \alpha - 2\psi_n(K, \delta)\}, \\ \Theta_\lambda &= \{\Omega \in \mathcal{G} : \lambda(\Omega_n) > \lambda(\Omega)\}, \\ \Gamma_\mu &= \{K : \sup_{\Omega \in \mathcal{G}} |\widehat{\mu}_n(\Omega) - \mu(\Omega)| - \psi_n(K, \delta) > 0\}. \end{aligned}$$

Now, one can deduce the same way as in [126], page 670 (see also Appendix A therein) that

$$\Theta_\mu \cup \Theta_\lambda \subset \Gamma_\mu$$

which yields the result. □

The following remark is an observation made in [126], page 670.

Remark 32. Note that

$$\Theta_\mu \cup \Theta_\lambda \subset \Gamma_\mu,$$

where Θ_μ , Θ_λ and Γ_μ are defined in the proof of Theorem 34; can be seen as an analogy to the result in classification, namely

$$R(\widehat{h}) - \inf_{h \in \mathcal{H}} R(h) \leq 2 \sup_{h \in \mathcal{H}} |R(h) - R_n(h)|,$$

where \widehat{h} is the empirical risk minimizer and \mathcal{H} is a set of classifiers.

Remark 33. One can establish analogous bound in sub-exponential case by using Bernstein or Hoeffding type maximal inequalities for Markov chains established in Chapter 5. In consequence, one can get sharper control of tolerance parameter $\psi_n(\delta)$. Moreover, one can assume that exponential moments of return time to the atom A exist for chain X and use Lemma 1 in order to make the terms involving parameter N asymptotically negligible.

3.1 Harris recurrent case

It is straightforward to extend the preceding results into a Harris recurrent case when the regeneration properties for Harris chains can be recovered via the Nummelin's splitting technique. The following condition is the same as Assumption 4 from Chapter 5. We state it below again in order to make this chapter self-contained.

Condition 3. *We have:*

$$\sup_{x \in S} \mathbb{E}_x[\tau_S^p] < \infty, \text{ and } \mathbb{E}_\nu[\tau_S^p] < \infty.$$

It is noteworthy that the hypothesis above is independent from the small set chosen. Recall also that this condition can be replaced by Foster-Lyapunov drift conditions that are much more tractable in practice, see *e.g.* Chapter 11 in [106]. The following theorem gives a generalization bound for MV -set estimation problem in a Harris recurrent case.

Theorem 35. *Suppose that Conditions 2-3 are fulfilled. Then, for all*

$$\delta = \delta_1 + \delta_2 + \delta_3 + \delta_4 + \delta_5 + \delta_6 \in (0, 1),$$

and for n large enough, any solution $\hat{\Omega}_n$ of (6.2) with

$$\psi_n(\delta) \stackrel{\text{def}}{=} \max \left[\left(\frac{D_1 \mathbb{E}_\nu[\tau_S^p]}{\delta_1 n^p} \right)^{1/(p+r)}, \left(\frac{D_1 \sup_{y \in S} \mathbb{E}_y[\tau_S^p]}{\delta_2 n^p} \right)^{1/(p+r)}, \left(\frac{D_2 \sup_{y \in S} \mathbb{E}_y[\tau_S^p]}{n^{p/2} \delta_3} \right)^{1/(p+r)}, \right. \\ \left. \left(\frac{D_3 \sup_{y \in S} \mathbb{E}_y[\tau_S^p]}{n^{3p/8} \delta_4} \right)^{1/(p+r)}, \left(\frac{D_4}{n^{3p/8} \sup_{y \in S} \mathbb{E}_y[\tau_S^p] \delta_5} \right)^{1/r}, \left(\frac{D_4}{\delta_6 n^{7p/16}} \right)^{1/r} \right]$$

satisfies, with probability at least $1 - \delta$,

$$\lambda(\hat{\Omega}_n) \leq \lambda(\Omega_\alpha^*) + \left\{ \inf_{\Omega \in \mathcal{G}: \mu(\Omega) \geq \alpha} \lambda(\Omega) - \lambda(\Omega_\alpha^*) \right\}$$

and

$$\mu(\hat{\Omega}_n) \geq \alpha - 2\psi_n(\delta).$$

The constants D_1, D_2, D_3 and D_4 can be explicitly computed and are analogous to the constants given in the proof of Theorem 34.

Proof. The proof is analogous to the proof of Theorem 34. The difference boils down to using polynomial tail inequality for Harris recurrent Markov chains stated in Theorem 30. We obtain different constants as a result that now we work under Condition 3. \square

4 Simulations

In this section we carry out some numerical experiments in order to illustrate the performance of MV-set method when the data are Markovian. Note that our goal is not to compare our method with other algorithms but rather show that our estimators' behaviour agree with the theory. We will provide a short discussion of the behaviour of the MV-set method when considering different penalties. Below, we present numerical results where approximate solutions of (6.2) are built by means of a dyadic recursive partitioning scheme (we used the mex/matlab code from <http://web.eecs.umich.edu/~cscott>). We refer to [127] for details of dyadic trees construction and to [126] for details concerning applications to minimum volume set estimation. We illustrate the performance of the *MV*-set estimation method with dyadic trees approach with $M/G/2$ bivariate queuing process. $M/G/2$ queue is a sub-geometric regenerative Markov chain with atom $A = (0, 0)$ (see for instance [57] for more details). The $M/G/2$ queue we consider is of the following form

$$\begin{bmatrix} x_{1,t+1} \\ x_{2,t+2} \end{bmatrix} = \max \left\{ \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix} - \begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix}, 0 \right\},$$

where the process $U_t = [u_{1,t}, u_{2,t}]$ has bivariate exponential distribution with parameters (λ_1, λ_2) and correlation ρ_u . Process $W_t = [w_{1,t}, w_{2,t}]$ has bivariate Pareto distribution with a density given by

$$f(w_1, w_2) = \frac{p(p+1)(\theta_1\theta_2)^{p+1}}{(\theta_2w_1 + \theta_1w_2 - \theta_1\theta_2)^{p+2}}$$

and correlation $Cor(w, v) = 1/p$ for $p > 2$. We draw from bivariate Pareto distribution using inverse sampling which requires the cumulative distribution functions of both marginal and conditional distributions. The marginal and conditional density functions are given in [102] and are of the form

$$f(w_2) = \frac{p\theta_2}{w_2^{p+1}} \text{ and } f(w_1|w_2) = \frac{(p+1) \left(\frac{\theta_1}{\theta_2}w_2\right)^{p+1}}{\left(w_1 + \frac{\theta_1}{\theta_2}w_2 - \theta_1\right)^{p+2}}.$$

Figure 6.1 show 10000 samples from $M/G/2$ such that (we denote this process as Process 1)

$$\lambda_1 = \lambda_2 = 0.4 \text{ and } \rho_u = 0.4.$$

The parameters of Pareto distributions are:

$$\theta_1 = 3, \theta_2 = 3 \text{ and } \rho_w = 1/3.$$

The estimated mean inter-renewal time is $\widehat{\mathbb{E}}_A[\tau_A] = 2.06$ (computed by Monte-Carlo simulation). We set level $\alpha = 0.9$ for the true minimum volume set. We consider constant penalty of one per leaf, Rademacher upper bound, exact Rademacher and Rademacher upper bound with volume term respectively (see [126] for more details for the choice of penalization). Figure 6.2 show 10000 realizations of $M/G/2$ such that (we denote this process as Process 2)

$$\lambda_1 = \lambda_2 = 0.3 \text{ and } \rho_u = 0.6.$$

The parameters of Pareto distributions are:

$$\theta_1 = 3, \theta_2 = 3 \text{ and } \rho_w = 1/6.$$

The estimated mean inter-renewal time is $\widehat{\mathbb{E}}_A[\tau_A] = 26.27$. We set level $\alpha = 0.90$ for the true minimum volume set. Penalties are chosen as in the preceding case. Figure 6.3 demonstrate the behaviour of the minimum volume set when we generate 10000 samples from $M/G/2$ model with parameters (we denote this process as Process 3):

$$\lambda_1 = \lambda_2 = 0.3, \rho_u = 0.7, \theta_1 = \theta_2 = 3 \text{ and } \rho_w = 1/8.$$

The estimated mean inter-renewal time is $\widehat{\mathbb{E}}_A[\tau_A] = 82.86$. Visual inspection of the resulting minimum volume set estimates allows to detect some of the characteristics of the different penalties and their behaviour as a function of the sample size. Reconstructions are based on con = constant penalty, rad = Rademacher penalty, radub = Rademacher upper bound (which is tight up to a factor $1/\sqrt{2}$, radvol = Rademacher upper bound with volume term (adds a volume to empirical mass in the radub penalty). We refer to [126] and description of the code in <http://web.eecs.umich.edu/~cscott> for the definitions of the aforementioned penalties and more comments.

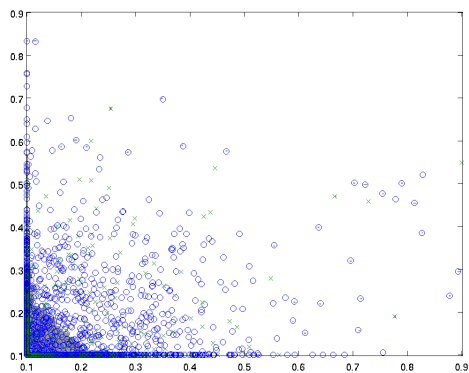
We observe that Rademacher upper bound with volume term allows isolated leafs into the final set estimate (especially for Process 2 and Process 3), which is somewhat unappealing. We see that applying Rademacher upper bound significantly eliminates the problem. The exact Rademacher penalty results in giving very reasonable estimates. Applying the constant penalty of one per leaf provides equally satisfactory results as in exact Rademacher case.

Finally, Figure 6.4 shows the error rate as a function of sample size. We see that the decrease of error rate to zero depends on $\widehat{\mathbb{E}}_A[\tau_A]$ and is the fastest for the process with $\widehat{\mathbb{E}}_A[\tau_A] = 2.06$ and the slowest for process with $\widehat{\mathbb{E}}_A[\tau_A] = 82.86$. This agrees with the theory, since $S_j(\Omega)/\mathbb{E}_A[\tau_A]$ play the role of training observations: the smaller the expected cycle length $\mathbb{E}[\tau_A]$, the larger the probability to observe a high number of training observations.

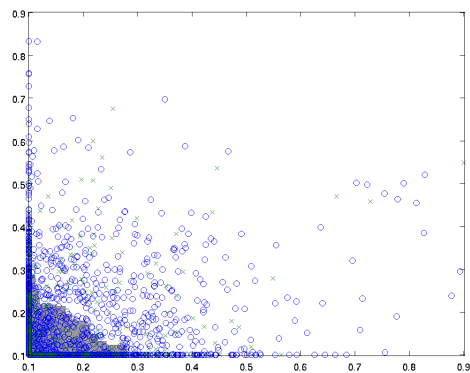
5 Conclusion

In this chapter we have shown how uniform tail bounds for suprema of empirical processes can be used to obtain generalization bounds for statistical learning algorithm when the data are Markovian. In this work we considered minimum volume set estimation algorithm which can be applied to outlier/anomaly detection, determining highest posterior density or multivariate confidence regions or clustering, see [126] for more details. More specifically, we have shown how to estimate minimum volume sets via empirical risk minimization and structural risk minimization paradigm. Finally, we illustrated our theoretical considerations by simulations.

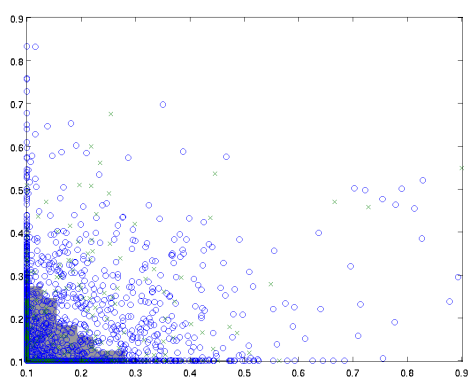
It is noteworthy that our generalization bounds depend on parameters that can be explicitly computed and work for all $n \geq 1$. However, when dealing with sub-geometric case, we struggle with full recovery of the i.i.d. rates. Interestingly, in sub-exponential case, the i.i.d. rates can be recovered (we can apply Bernstein/ Hoeffding inequality for Markov chains, see Chapter 5). The sub-geometric case requires further investigation.



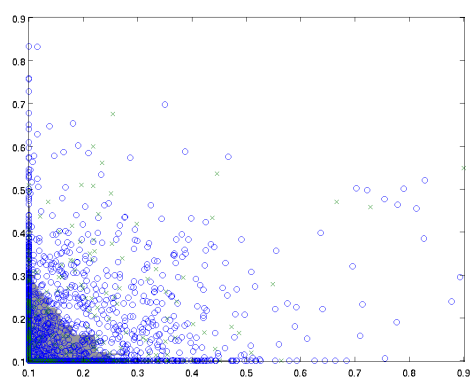
(a) con



(b) ar

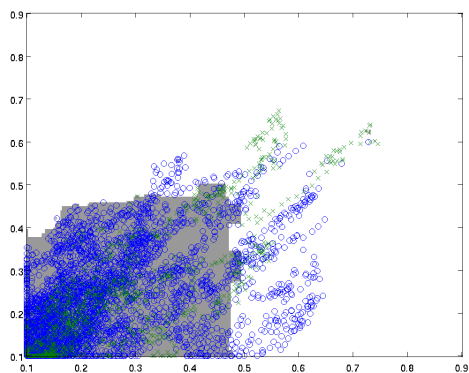


(c) radub

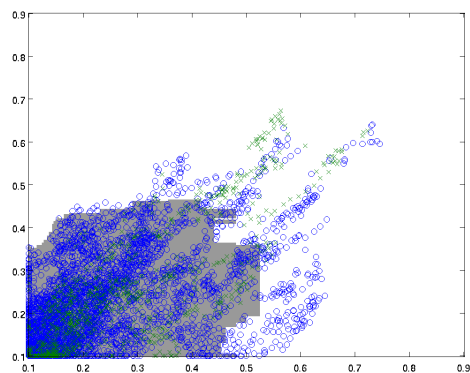


(d) radvol

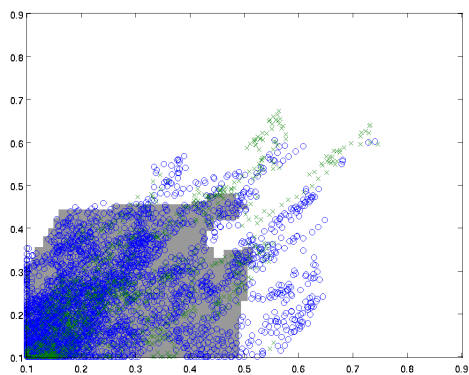
Figure 6.1: Minimum volume set estimates based on dyadic quadrees for Process 1, $\alpha = 0.9$, $n = 10000$.



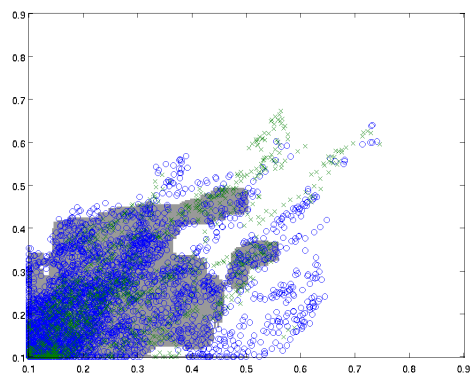
(a) con



(b) rad

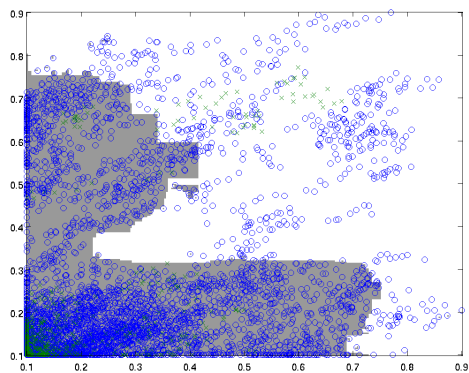


(c) radub

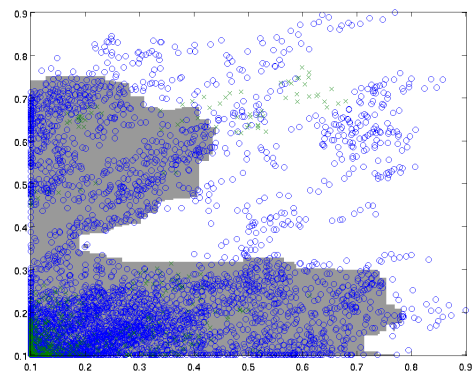


(d) radvol

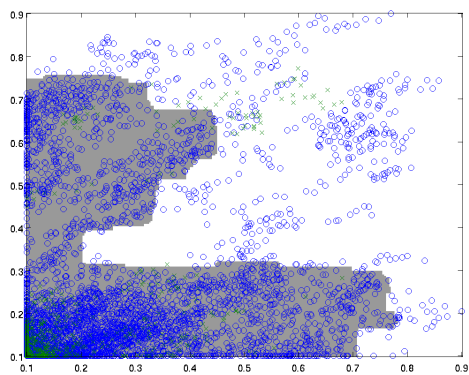
Figure 6.2: Minimum volume set estimates based on dyadic quadrees for Process 2, $\alpha = 0.9$, $n = 10000$.



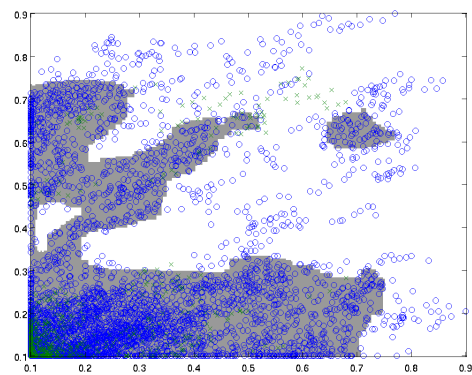
(a) con



(b) rad



(c) radub



(d) radvol

Figure 6.3: Minimum volume set estimates based on dyadic quadtree for Process 3, $\alpha = 0.90$, $n = 10000$.

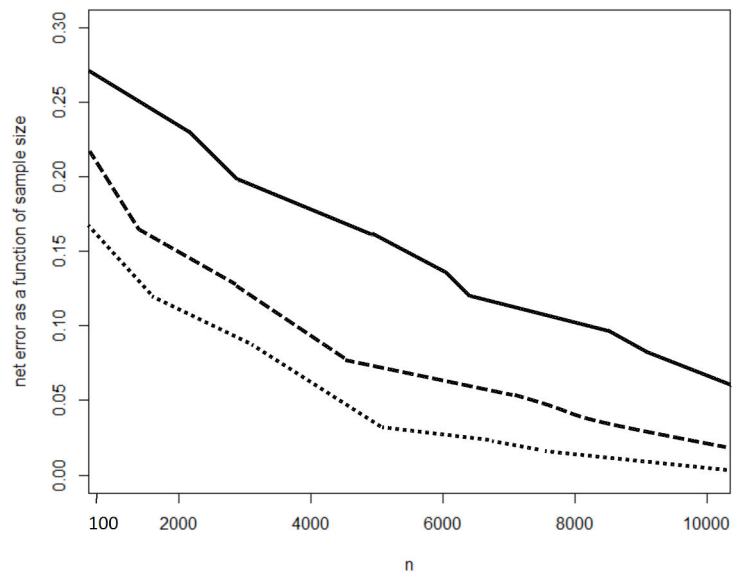


Figure 6.4: The error $\mathbb{E}[\widehat{\Omega}_n]$ as a function of sample size for Process 1 (dotted line), Process 2 (dashed line) and Process 3 (solid line).

Chapter 7

Résumé substantiel

Cette thèse se concentre sur quelques extensions de la théorie des processus empiriques lorsque les données sont Markoviennes. Nous nous intéressons dans cette thèse à quelques extensions en matière de bootstrap, de robustesse et d'apprentissage statistique dans le cadre des chaînes de Markov atomiques régénératives et Harris récurrentes.

1 Motivation

La théorie des processus empiriques joue un rôle important dans les statistiques modernes. Elle fournit des outils qui permettent de résoudre de nombreux problèmes statistiques dans différents domaines, tels l'analyse spectrale, la théorie des valeurs extrêmes, le bootstrap et l'apprentissage statistique. La théorie des processus empiriques dans le cas de données i.i.d. a déjà été étudiée en détails, voir par exemple [87], [119] et [141].

Cependant, l'hypothèse de variables i.i.d. est souvent irréaliste en pratique. Les données qui proviennent de domaines tels que la climatologie, la génétique, la finance, la géologie ou les télécommunications sont intrinsèquement temporelles par nature et par conséquent constituent des processus non i.i.d.. D'où la volonté des chercheurs d'étendre les concepts de la théorie des processus empiriques au cas dépendant (voir par exemple [54] pour une analyse détaillée des résultats obtenus pour les suites stationnaires et les données dépendantes).

Les chaînes de Markov atomiques régénératives et Harris récurrentes présentent un intérêt à la fois théorique et applicatif. Tout d'abord, la classe des chaînes de Markov (y compris les chaînes à mémoire infinie) est très générale et contient de nombreuses séries temporelles (y compris des processus non stationnaires); elles sont utilisées dans de nombreux modèles économétriques utilisant des données dépendantes. De plus, la structure régénérative des chaînes de Markov (voir par exemple [112], [12] et [111]) en fait un outil idéal pour étendre

certain résultats du cas i.i.d. au cas dépendant. En effet, [111] a prouvé qu'il était possible de découper (théoriquement) les chaînes Harris récurrentes en blocs indépendants en utilisant une extension probabiliste de la chaîne. La théorie des chaînes de Markov Harris récurrentes est bien étudiée, le lecteur pourra se reporter à [106], [136] et [111]. En particulier, on trouvera dans [98] et [44] un théorème central limite pour de tels processus, des développements bootstrap dans [26], [122], [75], des inégalités de concentration dans [2], [28], [22] et des applications à l'apprentissage statistique dans [23], [24].

De nombreux exemples de la vie réelle présentent les propriétés des chaînes de Markov régénératives atomiques et Harris récurrentes : les systèmes de stockage et les files d'attente, beaucoup de modèles dans les domaines de la finance, de l'assurance ou de l'évaluation des risques alimentaires (voir [106], [27], [31] et [25]). De nombreux résultats pour les chaînes de Markov ont été établis avec des propriétés de mélange qui sont difficiles à vérifier en pratique.

La première partie de la thèse se concentre sur des aspects relatifs au bootstrap. Le bootstrap naïf a été proposé par Efron [59]. Étant données n observations X_1, \dots, X_n i.i.d. ayant une loi inconnue F , on souhaite estimer la loi d'échantillonnage de certaines fonctionnelles $R_n(X_1, \dots, X_n, F)$. Dans cette situation, la méthode bootstrap simple non paramétrique est facile à implémenter. Tout d'abord, nous construisons la version empirique de F , c'est-à-dire $F_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$. L'étape suivante consiste à tirer n observations selon F_n : $X_i^*, i = 1, \dots, n$, où les X_i^* sont i.i.d. conditionnellement à F_n . On approxime la loi d'échantillonnage d'une fonctionnelle d'intérêt $R_n(X_1, \dots, X_n, F)$ par la loi de $R_n(X_1^*, \dots, X_n^*, F_n)$ conditionnellement à F_n . On peut se demander pourquoi ne pas simplement utiliser le théorème central limite pour étudier $R_n(X_1, \dots, X_n, F)$. Cette approche n'est pas toujours possible car la forme explicite de la loi limite peut être difficile à obtenir (voir par exemple [121] pour plus de détails et de références). Il est assez courant que la loi du processus de la limite dépend de certains paramètres inconnus. En résumé, les méthodes bootstrap permettent de remédier aux problèmes susmentionnés (voir le Chapitre 2 pour plus de détails).

Les méthodes de bootstrap non paramétrique pour les données i.i.d. ont progressivement évolué et de nouveaux types de bootstrap dans le cas i.i.d. et le cas dépendant ont été établis (voir [76], [93] et [43], [42] pour les résultats en grande dimension). Cela a conduit à énormément d'applications dans la plupart tous les domaines des statistiques (voir par exemple [80] et [76] pour plus de détails). Suite à l'intérêt accru porté à l'inférence statistique dans un cadre dépendant, de nouvelles procédures de bootstrap ont été développées. La plupart des modèles dans le cadre dépendant reposent sur des techniques exploitant une structure par blocs. Ces approches se résument essentiellement au rééchantillonnage de blocs d'observations de sorte que la structure de dépendance soit capturée. Il existe de nombreuses

variantes aux méthodes de block bootstrap pour les données dépendantes telles que le *moving block bootstrap* (MBB), *non-overlapping block bootstrap* (NBB) ou *circular bootstrap* (CBB) pour n'en nommer que quelques-unes (voir par exemple [93] pour un aperçu exhaustif des procédures susmentionnées).

Malheureusement, comme indiqué par de nombreux auteurs (voir par exemple [46] et [93]), ces procédures se heurtent à de nombreux problèmes. Le gros inconvénient est que les méthodes de bloc bootstrap sont très sensibles au choix de la longueur des blocs. En effet, la longueur optimale des blocs dépend fortement de la taille de l'échantillon et du processus de génération de données. De plus, la méthode populaire MBB nécessite la propriété de stationnarité pour les observations ce qui entraîne généralement l'échec de cette méthode en mode non stationnaire (voir [93] pour plus de détails).

Nous nous sommes attachés à rappeler les principales méthodes de bootstrap par blocs pour des données dépendantes. En effet, les procédures de bootstrap que nous considérons dans cette thèse reposent également sur la segmentation des données en blocs. Il existe cependant de nombreux autres méthodes du bootstrap que l'on peut utiliser dans le cas des données sont dépendantes, tels que bootstrap résiduel ou le wild bootstrap. Nous renvoyons le lecteur à [89] et [93] pour plus de détails sur la mise en pratique de ces procédures et leurs limitations.

Compte tenu des limitations des méthodes de bootstrap par blocs, nous avons décidé de nous concentrer sur les techniques de régénération des chaînes de Markov atomiques et Harris récurrentes. Dans [26], la procédure du bloc bootstrap régénératif (RBB) et la méthode approximative du bloc bootstrap régénératif (ARBB) sont présentées. Ces procédures ne requièrent pas le choix de la longueur des blocs contrairement à les méthodes de bootstrap par blocs. Dans le cas atomique, la division des blocs de données est entièrement déterminée par les données. Comme indiqué dans [26], le théorème central limite de type bootstrap pour la moyenne dans le cadre Markovien est valide. Nous avons développé cette théorie plus loin en établissant des théorèmes de la limite centrale du bootstrap uniformes sur des classes de fonctions non nécessairement bornées. Les théorèmes fonctionnels de la limite centrale de type bootstrap sont utiles pour prouver la validité des procédures de bootstrap (voir [122], [44] et [71]) et peuvent être utilisés dans de nombreuses applications statistiques, afin par exemple d'obtenir des versions bootstrap des résultats dans [110], [68] et [69].

Ensuite, nous utilisons les résultats susmentionnés pour obtenir un théorème de la limite centrale pour des fonctionnelles Fréchet différentiables dans un cadre Markovien. La différentiabilité de Fréchet est un concept essentiel de la statistique robuste, car elle garantit l'existence d'une fonction d'influence qui permet de détecter les valeurs aberrantes dans les

données (voir [140] pour plus de détails). Plus spécifiquement, on peut détecter des valeurs aberrantes et construire des estimateurs de plug-in robustes en éliminant les blocs ayant soit une contribution trop importante aux statistiques d'intérêt ou ayant une longueur trop grande, ce qui entraîne un biais important dans les statistiques (au lieu de prendre en compte l'impact d'une seule observation sur une statistique donnée).

Notre deuxième direction concernant les développements bootstrap dans un cadre dépendant s'attache à étudier des processus autorégressifs périodiques (PAR) qui sont un exemple de chaînes de Markov Harris récurrentes. Nous proposons des méthodes d'échantillonnage résiduel et de wild bootstrap et prouvons leur validité. Ces méthodes sont uniquement basées sur les données et ne nécessitent donc aucun étalonnage de la longueur de bloc, un aspect qui peut être fort attrayant en pratique.

La deuxième partie de la thèse se concentre sur les applications de l'apprentissage statistique à la théorie des processus empiriques. Sans surprise, la théorie de l'apprentissage statistique est principalement étudiée dans le cas i.i.d. (voir [38], [142], [144] et [64]). Cependant, il existe une forte demande motivée par le domaine du Big Data afin d'étendre les résultats au cas dépendant. La théorie de l'apprentissage automatique des processus dépendants a fait l'objet de recherches approfondies au cours des dernières années ; voir par exemple [3], [5], [132] ou [78] pour certains résultats présentés dans un cadre très général.

Dans la théorie de l'apprentissage statistique, de nombreux travaux ont établi des limites non asymptotiques évaluant la capacité de généralisation de minimiseurs empiriques du risque sous une grande variété d'hypothèses de complexité pour la classe de règles de décision sur laquelle l'optimisation est effectuée via le contrôle de la distance entre les moyennes des variables aléatoires et leurs espérance (et est géré par les inégalités de concentration qui donnent une borne supérieure sur la probabilité de taille pour suprema de processus empiriques). Ces résultats, à première vue très théoriques, sont des outils cruciaux pour étudier la capacité d'apprentissage des algorithmes d'apprentissage statistique. Il existe de nombreux résultats de concentration pour les données dépendantes. Les résultats dans le cas Markovien doivent être mentionnés [28], [2] et [1]. En dépit de divers résultats de concentration, nous établissons des inégalités de concentration de type maximales sur mesure pour nos applications, c'est-à-dire qu'elles sont établies pour des processus de Markov non stationnaires et pour des classes de fonctions non bornées \mathcal{F} .

Dans cette thèse, nous souhaiterons établir des bornes de généralisation pour les algorithmes d'apprentissage statistique lorsque les données sont Markoviennes. Nous obtenons de tels résultats via une approche empirique de minimisation des risques. Notre stratégie se résume essentiellement à 3 étapes.

- Nous obtenons des inégalités de concentration pour des classes de fonctions bornées et non bornées des chaînes de Markov (voir [22], [23] et [24]). Les inégalités exponentielles (par exemple Bernstein et Hoeffding) et les inégalités de moments polynomiales sont un outil essentiel lorsque l'on veut contrôler une minimisation empirique des risques ou calculer la vitesse de convergence d'un algorithme d'apprentissage statistique.
- Nous étudions la performance des algorithmes d'apprentissage via une minimisation empirique du risque. Il est à noter que l'analyse de l'algorithme ERM et des propriétés des algorithmes d'apprentissage statistique sont des problèmes très importants et urgents à résoudre.
- Nous étudions les propriétés de généralisation d'un algorithme d'apprentissage statistique sélectionné (dans le cadre régénératif atomique et Harris Markovien).

Dans cette thèse, nous présentons les bornes de généralisation pour l'estimation d'ensemble à volume minimal problème (minimum volume (MV) set en anglais). Le concept d'estimation MV-set a été introduit pour la première fois dans [120] et étend le concept de quantile pour les distributions de probabilité multivariées (voir [120] pour plus de détails). Cette méthode offre un cadre non paramétrique pour la detection (non supervisée) des anomalies / nouveautés.

Comme observé dans [126], l'estimation de MV-set peut être transposée dans un cadre d'apprentissage très semblable à la minimisation empirique du risque dans la classification (supervisée). Des bornes de généralisation du problème d'estimation de l'ensemble MV ont été établies dans [126] dans le cas i.i.d.. Certains résultats ont obtenus dans le cas dépendant dans [55]. Nous avons étendu les résultats précédents aux cas atomique et Harris récurrent dans [23] et [24].

2 Propriétés fondamentales des chaînes de Markov régénératives atomiques et Harris récurrentes

Dans cette section, nous introduisons quelques notations et rappelons les concepts clés de la théorie des chaînes de Markov (se référer à [106] et [136], [27] et [112] pour des rappels exhaustifs). Les résultats fournis dans cette section ont un caractère purement informatif. Le lecteur intéressé trouvera dans la littérature susmentionnée les preuves des théorèmes énoncés pour des chaînes de Markov à espace d'état général. Dans toute cette section, \mathbb{I}_A est la fonction indicatrice de l'événement A .

On suppose que $X = (X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov homogène sur un espace d'état général (E, \mathcal{E}) avec probabilité de transition Π et loi initiale ν . Le lecteur pourra noter que

pour tout $B \in \mathcal{E}$, et $n \in \mathbb{N}$, nous avons

$$X_0 \sim \nu \text{ et } \mathbb{P}(X_{n+1} \in B | X_0, \dots, X_n) = \Pi(X_n, B) \text{ p.s.}$$

Dans ce qui suit, \mathbb{P}_x (respectivement \mathbb{P}_ν) est la mesure de probabilité telle que $X_0 = x$ et $X_0 \in E$ (respectivement $X_0 \sim \nu$), et nous écrivons $\mathbb{E}_x(\cdot)$ pour la \mathbb{P}_x -espérance (respectivement $\mathbb{E}_\nu(\cdot)$ la \mathbb{P}_ν -espérance).

Dans ce qui suit, on suppose que X est ψ -irréductible et apériodique.

2.1 Chaînes de Markov régénératives atomiques

Dans cette thèse, notre intérêt particulier s'est porté sur la structure atomique des chaînes de Markov en raison de sa capacité à étendre la théorie des processus empiriques (essentielle aux développements dans le domaine du bootstrap et de l'apprentissage statistique) à partir du cas i.i.d. au cadre Markovien.

Supposons que X soit apériodique et ψ -irréductible. Un ensemble $A \in E$ est un atome accessible si pour tout $x, y \in A$ nous avons $\Pi(x, \cdot) = \Pi(y, \cdot)$ et $\psi(A) > 0$. Dans ce cas, nous appelons X atomique. Intuitivement, l'atome est un ensemble à partir duquel toutes les probabilités de transition sont les mêmes. Par conséquent, chaque fois que X arrive à A , il oublie son passé et recommence (se régénère).

La propriété de Markov forte (voir [106] pour une justification rigoureuse) garantit que, étant donné la loi initiale ν , les trajectoires peuvent être découpées en blocs i.i.d. correspondant aux visites consécutives de l'atome A par la chaîne. Les blocs sont de la forme :

$$B_j = (X_{1+\tau_A(j)}, \dots, X_{\tau_A(j+1)}), \quad j \geq 1$$

et prennent des valeurs dans le tore $\mathbb{T} = \bigcup_{k=1}^{\infty} E^k$.

Sachant que la chaîne possède la structure atomique, nous définissons la séquence des temps de régénération $(\tau_A(j))_{j \geq 1}$. La séquence est constituée des temps successifs auxquels la chaîne oublie son passé. Soit

$$\tau_A = \tau_A(1) = \inf\{n \geq 1 : X_n \in A\}$$

la première fois où la chaîne arrive à l'atome A et

$$\tau_A(j) = \inf\{n > \tau_A(j-1), X_n \in A\} \text{ for } j \geq 2.$$

Nous introduisons quelques notations supplémentaires : tout au long de la thèse, nous écrivons

$$l_n = \sum_{i=1}^n \mathbb{I}\{X_i \in A\}$$

pour le nombre total de visites consécutives de l'ensemble atomique A par la chaîne, donc nous avons $l_n + 1$ blocs. Nous faisons la convention $B_{l_n}^{(n)} = \emptyset$ quand $\tau_A(l_n) = n$. De plus, on note

$$l(B_j) = \tau_A(j+1) - \tau_A(j), \quad j \geq 1,$$

la longueur des blocs de régénération.

Ci-dessous, nous fournissons le schéma de construction des blocs. L'étape 3 n'est pertinente que dans le cadre de résultats non-asymptotiques. Nous supposons que nous observons la trajectoire $\mathbb{X}_n = (X_1, \dots, X_n)$.

Construction de blocs de régénération

Étape 1 Compter le nombre total de visites $l_n = \sum_{i=1}^n \mathbb{I}\{X_i \in A\}$ de l'atome A jusqu'à temps n .

Étape 2 Diviser les données \mathbb{X}_n en $l_n + 1$ blocs en fonction des visites consécutives de l'atome A :

$$\begin{aligned} B_0 &= (X_1, \dots, X_{\tau_A(1)}), \dots, B_j = (X_{\tau_A(j)+1}, \dots, X_{\tau_A(j+1)}), \dots, \\ B_{l_n-1} &= (X_{\tau_A(l_n-1)+1}, \dots, X_{\tau_A(l_n)}), B_{l_n}^{(n)} = (X_{\tau_A(l_n)+1}, \dots, X_n). \end{aligned}$$

Étape 3 Supprimer le premier bloc B_0 et le dernier $B_{l_n}^{(n)}$ si $\tau_A(l_n) < n$.

Dans le cadre de notre étude, nous nous intéressons également au comportement asymptotique des chaînes de Harris Markov positives et récurrentes. Supposons que X est une chaîne de Markov ψ -irréductible. La chaîne X est *Harris récurrente* si, à partir de tout point $x \in E$ et de tout ensemble tel que $\psi(A) > 0$, nous avons

$$\mathbb{P}_x(\tau_A < +\infty) = 1.$$

La propriété de Harris récurrence garantit que X visite tout ensemble A infiniment souvent p.s..

Dans notre cadre, nous nous intéressons à l'analyse en régime permanent des chaînes de Markov. Plus spécifiquement, quand une mesure invariante est finie, on peut la normaliser en une mesure de probabilité stationnaire. D'après le théorème de Kac, la chaîne de Markov X est récurrente positive si et seulement si $\mathbb{E}_A(\tau_A) < \infty$. L'unique loi de probabilité unique invariante μ est la mesure d'occupation de Pitman donnée par

$$\mu(B) = \frac{1}{\mathbb{E}_A(\tau_A)} \mathbb{E}_A \left(\sum_{i=1}^{\tau_A} \mathbb{I}\{X_i \in B\} \right), \quad \forall B \in \mathcal{E}.$$

Nous avons par le théorème de Kac :

$$\mathbb{E}(l(B_j)) = \mathbb{E}_A(\tau_A) = \frac{1}{\mu(A)}.$$

On considère la fonction μ -intégrable $f : E \rightarrow \mathbb{R}$. Par

$$u_n(f) = \frac{1}{\tau_A(l_n) - \tau_A(1)} \sum_{i=1+\tau_A(1)}^{\tau_A(l_n)} f(X_i),$$

on désigne l'estimateur de la moyenne asymptotique inconnue $\mathbb{E}_\mu(f(X_1))$.

2.2 Chaînes de Markov Harris récurrentes et scission de Nummelin

Dans cette section, nous expliquons comment les techniques de régénération peuvent être étendues du cas atomique à un cadre Harris récurrent grâce aux travaux de [12] et [111]. Nummelin [111] a notamment proposé la technique dite de *scission* qui permet, en termes simples, d'étendre la structure probabiliste de toute chaîne de Harris afin de construire artificiellement un ensemble de régénération. Dans cette section, X est une chaîne de Markov Harris récurrente avec noyau de transition Π . Dans cette section, nous adoptons les notations introduites dans [27] et invitons le lecteur à s'y référer pour un exposé détaillé de la théorie présentée ci-dessous.

Un ensemble $S \in \mathcal{E}$ est dit petit s'il existe un paramètre $\delta > 0$, une mesure de probabilité positive Φ supportée par S et un entier $m \in \mathbb{N}^*$ tels que

$$\forall x \in S, A \in \mathcal{E} \quad \Pi^m(x, A) \geq \delta \Phi(A), \quad (7.1)$$

où Π^m désigne le m -ième itéré de la probabilité de transition Π .

Comme indiqué dans [82], les petits ensembles existent pour tout noyau irréductible sous l'hypothèse que l'espace d'état est généré de manière dénombrable.

Dans ce qui suit, nous élargissons l'espace échantillon afin de définir une suite $(Y_n)_{n \in \mathbb{N}}$ de variables aléatoires indépendantes de paramètre δ . Soit $\mathbb{P}_{\nu, \mathcal{M}}$ la loi conjointe de $X^{\mathcal{M}} = (X_n, Y_n)_{n \in \mathbb{N}}$. La construction est basée sur la représentation de Π , définie sur S , à savoir

$$\Pi(x, A) = \delta \Phi(A) + (1 - \delta) \frac{\Pi(x, A) - \delta \Phi(A)}{1 - \delta}.$$

Il peut être obtenu par la randomisation suivante de la transition probabilité Π chaque fois que la chaîne X visite l'ensemble S . Si $X_n \in S$ et

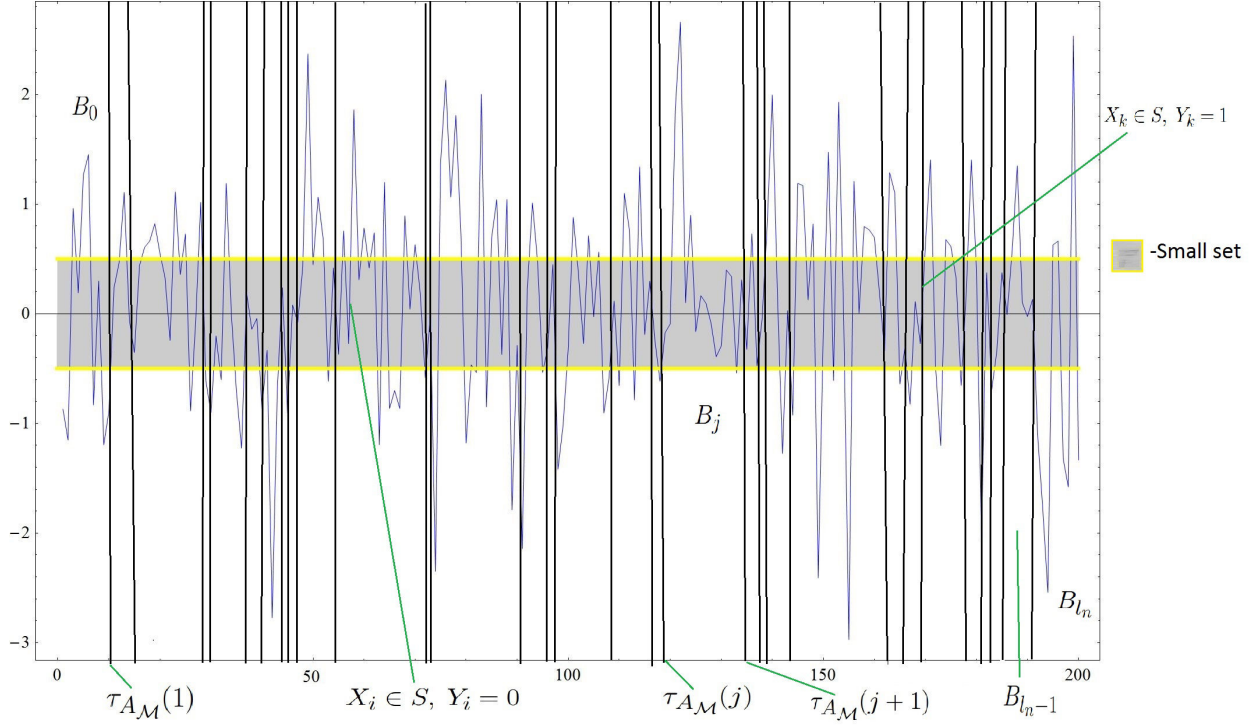


Figure 7.1 : Construction de bloc de régénération pour le modèle AR (1).

- si $Y_n = 1$ (cela arrive avec probabilité $\delta \in]0,1[$), alors X_{n+1} est distribué selon la mesure de probabilité Φ ,
- si $Y_n = 0$ (cela arrive avec probabilité $1 - \delta$), alors X_{n+1} est distribué selon la mesure de probabilité $(1 - \delta)^{-1}(\Pi(X_n, \cdot) - \delta\Phi(\cdot))$.

Cette chaîne de Markov bivariée $X^{\mathcal{M}}$ est appelée *chaîne scindée*. Elle prend ses valeurs dans $E \times \{0,1\}$ et possède un atome $S \times \{1\}$. Cette chaîne scindée $X^{\mathcal{M}}$ hérite de toutes les propriétés de stabilité et de communication de la chaîne X (voir [106] et [112] pour une présentation rigoureuse). La Figure 2.2 illustre la technique de scission pour une trajectoire de processus AR(1). Il est connu que le processus autorégressif d'ordre 1 est une chaîne de Markov Harris récurrente (voir par exemple [106] pour plus de détails).

Il convient de noter que les blocs créés par la technique de fractionnement sont i.i.d. dans le cas où $m = 1$ sous condition de minoration (7.1). Si la chaîne X satisfait $\mathcal{M}(m, S, \delta, \Phi)$ pour $m > 1$, alors les blocs de données sont 1-dépendants. Cependant, dans de nombreux cas, il est facile d'adapter la théorie du cas $m = 1$ en considérant des sommes de blocs pairs et impairs afin de traiter la dépendance entre les B_j (voir par exemple [98] ou [95] pour plus de détails) ou en vectorisant la chaîne (voir [106]).

Par souci de simplicité, dans la suite de ce manuscrit, la condition de minoration \mathcal{M} est satisfaite avec $m = 1$, sauf indication contraire.

2.3 Blocs régénératifs pour les familles dominées

Nous supposons que la famille de lois conditionnelles $\{\Pi(x, dy)\}_{x \in E}$ et la loi initiale ν sont dominées par une σ mesure finie λ de référence de sorte que

$$\nu(dy) = f(y)\lambda(dy) \quad \text{et} \quad \Pi(x, dy) = p(x, y)\lambda(dy),$$

pour tout $x \in E$. On suppose également que Φ est absolument continue par rapport à λ de telle sorte que

$$p(x, y) \geq \delta \phi(y), \quad \lambda(dy) \quad \text{p.s. pour tout } x \in S$$

avec $\Phi(dy) = \phi(y)dy$. Dans ce qui suit, soit Y une suite aléatoire binaire obtenue via la technique de Nummelin à partir des paramètres donnés par la condition \mathcal{M} . La loi de $Y^{(n)} = (Y_1, \dots, Y_n)$ conditionnellement à $X^{(n+1)} = (x_1, \dots, x_{n+1})$ est le produit tensoriel des lois de Bernoulli données par : pour tous $\beta^{(n)} = (\beta_1, \dots, \beta_n) \in \{0, 1\}^n$, $x^{(n+1)} = (x_1, \dots, x_{n+1}) \in E^{n+1}$,

$$\mathbb{P}_\nu(Y^{(n)} = \beta^{(n)} \mid X^{(n+1)} = x^{(n+1)}) = \prod_{i=1}^n \mathbb{P}_\nu(Y_i = \beta_i \mid X_i = x_i, X_{i+1} = x_{i+1}),$$

avec, pour $1 \leq i \leq n$,

- si $x_i \notin S$, $\mathbb{P}_\nu(Y_i = 1 \mid X_i = x_i, X_{i+1} = x_{i+1}) = \delta$,
- si $x_i \in S$, $\mathbb{P}_\nu(Y_i = 1 \mid X_i = x_i, X_{i+1} = x_{i+1}) = \delta \phi(x_{i+1})/p(x_i, x_{i+1})$.

Étant donné $X^{(n+1)}$, de $i = 1, \dots, n$, Y_i est distribué suivant la loi de Bernoulli de paramètre δ , à l'exception du cas où X frappe le petit ensemble S au temps i : alors, Y_i est tiré selon la loi de Bernoulli de paramètre $\delta \phi(X_{i+1})/p(X_i, X_{i+1})$. On note $\mathcal{L}^{(n)}(p, S, \delta, \phi, x^{(n+1)})$ cette loi de probabilité. Si nous pouvions générer Y_1, \dots, Y_n , de sorte que

$$X^{\mathcal{M}(n)} = ((X_1, Y_1), \dots, (X_n, Y_n))$$

soit une réalisation de la chaîne scindée $X^{\mathcal{M}}$, alors nous pourrions procéder à une décomposition en blocs de la trajectoire de l'échantillon $X^{\mathcal{M}(n)}$ conduisant à des blocs asymptotiquement i.i.d.. Le lecteur pourra noter que cette procédure nécessite de connaître la densité de transition $p(x, y)$ afin de générer des variables aléatoires (Y_1, \dots, Y_n) . Cependant, en pratique, la densité de transition est inconnue et doit être estimée. Par conséquent, nous ne pouvons pas

utiliser directement la procédure indiquée ci-dessus et devons appliquer la version approchée proposée dans [26]. La construction se decompose en deux étapes, premièrement, construire un estimateur $p_n(x, y)$ de $p(x, y)$ basé sur $X^{(n+1)}$, de sorte que $p_n(x, y)$ satisfait

$$p_n(x, y) \geq \delta\phi(y), \lambda(dy) - p.s. \text{ et } p_n(x, y) > 0, 1 \leq i \leq n. \quad (7.2)$$

Dans la deuxième étape, générer un vecteur aléatoire $\hat{\mathbb{Y}}_n = (\hat{Y}_1, \dots, \hat{Y}_n)$ conditionnellement à $X^{(n+1)}$ de loi $\mathcal{L}^{(n)}(p_n, S, \delta, \gamma, X^{(n+1)})$ qui est une approximation de la loi conditionnelle $\mathcal{L}^{(n)}(p, S, \delta, \gamma, X^{(n+1)})$ de (Y_1, \dots, Y_n) sachant $X^{(n+1)}$. La validité de cette approximation a été prouvée dans [26].

Dans ce cadre, nous définissons les temps de visite successifs de $A_{\mathcal{M}} = S \times \{1\}$ par

$$\hat{\tau}_{A_{\mathcal{M}}}(i), i = 1, \dots, \hat{l}_n, \text{ où } \hat{l}_n = \sum_{i=1}^n \mathbb{I}\{X_i \in S, \hat{Y}_i = 1\}$$

est le nombre total de visites de la chaîne scindée en $A_{\mathcal{M}}$ jusqu'à n . Ci-dessous, nous fournissons un schéma de construction approximatif. Soit $\mathbb{X}_{n+1} = (X_1, X_2, \dots, X_{n+1})$ un échantillon aléatoire tiré selon une chaîne Harris récurrente X . On suppose que X vérifie les hypothèses énoncées précédemment dans cette section. L'étape 5 n'est pertinente que dans le cadre de résultats non-asymptotiques. Pour ce qui est des résultats asymptotiques, elle pourra être omise.

Construction approximative de blocs de régénération

Étape 1 Construire un estimateur $p_n(x, y)$ de la densité de transition à l'aide de l'échantillon \mathbb{X}_{n+1} . L'estimateur $p_n(x, y)$ doit satisfaire les conditions (7.2).

Étape 2 Conditionnellement à \mathbb{X}_{n+1} , tirer $(\hat{Y}_1, \dots, \hat{Y}_n)$ à partir de $\mathcal{L}^{(n)}(p_n, S, \delta, \gamma, \mathbb{X}_{n+1})$. En pratique, \hat{Y} est tiré seulement lorsque $X_i \in S$ (voir [27] pour plus de détails). A l'instant i , tel que $X_i \in S$, tirer \hat{Y}_i selon la loi de Bernoulli de $\delta\gamma(X_{i+1}) \backslash p_n(X_i, X_{i+1})$.

Étape 3 Compter le nombre de visites

$$\hat{l}_n = \sum_{i=1}^n \mathbb{I}\{X_i \in S, \hat{Y}_i = 1\}$$

à l'atome $S_1 = S \times \{1\}$ jusqu'à n .

Étape 4 Couper la trajectoire \mathbb{X}_{n+1} en $\hat{l}_n + 1$ blocs de régénération approximatifs qui correspondent à visites réussies consécutives de (X, \hat{Y}) à S_1 . Les blocs approximatifs sont de

la forme

$$\begin{aligned}\widehat{B}_0 &= (X_1, \dots, X_{\widehat{\tau}_{A_{\mathcal{M}}}(1)}), \dots, \widehat{B}_j = (X_{\widehat{\tau}_{A_{\mathcal{M}}}(j)+1}, \dots, X_{\widehat{\tau}_{A_{\mathcal{M}}}(j+1)}), \dots, \\ \widehat{B}_{\widehat{l}_n-1} &= (X_{\widehat{\tau}_{A_{\mathcal{M}}}(\widehat{l}_n-1)+1}, \dots, X_{\widehat{\tau}_{A_{\mathcal{M}}}(\widehat{l}_n)}), \widehat{B}_{\widehat{l}_n}^{(n)} = (X_{\widehat{\tau}_{A_{\mathcal{M}}}(\widehat{l}_n)+1}, \dots, X_{n+1}).\end{aligned}$$

Étape 5 Supprimer le premier bloc \widehat{B}_0 et le dernier $\widehat{B}_{\widehat{l}_n}^{(n)}$ si $\widehat{\tau}_{S_1}(\widehat{l}_n) < n$.

Dans ce qui suit, nous désignons par

$$\widehat{n}_{A_{\mathcal{M}}} = \widehat{\tau}_{A_{\mathcal{M}}}(\widehat{l}_n) - \widehat{\tau}_{A_{\mathcal{M}}}(1) = \sum_{i=1}^{\widehat{l}_n-1} l(\widehat{B}_i)$$

le nombre total d'observations après le premier et avant les derniers temps de pseudo-régénération. Soit

$$\sigma_f^2 = \frac{1}{\mathbb{E}_{A_{\mathcal{M}}}(\tau_{A_{\mathcal{M}}})} \mathbb{E}_{A_{\mathcal{M}}} \left(\sum_{i=1}^{\tau_{A_{\mathcal{M}}}} \{f(X_i) - \mu(f)\} \right)^2$$

la variance asymptotique. En outre, nous définissons

$$\widehat{\mu}_n(f) = \frac{1}{\widehat{n}_{A_{\mathcal{M}}}} \sum_{i=1}^{\widehat{l}_n-1} f(\widehat{B}_i), \quad \text{où} \quad f(\widehat{B}_j) = \sum_{i=1+\widehat{\tau}_{A_{\mathcal{M}}}(j)}^{\widehat{\tau}_{A_{\mathcal{M}}}(j+1)} f(X_i)$$

et

$$\widehat{\sigma}_n^2(f) = \frac{1}{\widehat{n}_{A_{\mathcal{M}}}} \sum_{i=1}^{\widehat{l}_n-1} \left\{ f(\widehat{B}_i) - \widehat{\mu}_n(f) l(\widehat{B}_i) \right\}^2.$$

Comme indiqué dans [27], la technique de Nummelin repose fortement sur la condition de minoration et un petit ensemble bien choisis. Le choix du petit ensemble ainsi que la précision du minorant uniforme sur la densité de transition de X (dans la condition de minoration) joue un rôle crucial pour obtenir suffisamment de blocs. L'observation suivante provient de [27] (voir l'article susmentionné pour des exemples de choix d'un petit ensemble pour différentes séries temporelles).

Si la taille d'un petit ensemble augmente, le minorant uniforme pour la densité de transition de X diminue. Inversement : pour une réalisation donnée de la trajectoire, augmenter la taille du petit ensemble S fait croître le nombre de points de la trajectoire qui sont candidats à la détermination d'un bloc. Il convient donc de trouver un équilibre entre taille de S et $\delta = \delta(S)$.

Enfin, nous mentionnons brièvement qu'il existe une relation entre les coefficients de α -mélange et les temps de régénération pour les chaînes de Markov Harris récurrentes. Pour ce

faire, nous faisons l'hypothèse de conditions de moment imposées sur τ_A et de conditions de moment des blocs plutôt qu'utiliser des propriétés de mélange principalement en raison du fait qu'il est difficile, en pratique, de vérifier les conditions de mélange. Cependant, compte tenu du nombre considérable de travaux exprimant la dépendance en termes de conditions de mélange, nous fournissons quelques commentaires ci-dessous.

Soit \mathcal{F}_a^b la σ -algèbre engendrée par X_a, \dots, X_b . Le coefficient d' α -mélange entre deux σ -algèbres \mathcal{A} et \mathcal{B} est défini comme

$$\alpha(\mathcal{A}, \mathcal{B}) := \sup_{(A, B) \in \mathcal{A} \times \mathcal{B}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|.$$

Les coefficients de mélange liés à une suite de variables aléatoires sont définis par

$$\alpha(k) = \sup_n \sup_{A \in \mathcal{F}_{-\infty}^n} \sup_{B \in \mathcal{F}_{n+k}^{+\infty}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|.$$

Le Théorème 2 de [36] indique que pour les chaînes de Harris stationnaires pour un choix de $\lambda \geq 0$ tel que la somme $\sum_m m^\lambda \alpha(m) < \infty$, alors pour tout $B \in \mathcal{E}$ tel que $\mu(B) > 0$ nous avons

$$\mathbb{E}_\mu(\tau_B^{1+\lambda}) < \infty, \quad \text{où} \quad \tau_B = \inf\{n \geq 1 : X_n \in B\}.$$

Ce résultat garantit une décroissance polynomiale des coefficients de mélange forts. C'est une condition faible, car on suppose généralement que le taux de décroissance soit exponentiel.

3 Un aperçu des contributions et des perspectives futures

Comme indiqué dans la Section 1, cette thèse porte sur les développements dans le domaine du bootstrap et de l'apprentissage statistique lorsque les données sont Markoviennes. Nous commençons par les théorèmes central limite bootstrap uniformes pour les chaînes de Markov Harris récurrentes. Ensuite, nous généralisons les procédures d'échantillonnage résiduel et wild bootstrap pour les processus autorégressifs d'ordre p en processus autorégressifs périodiques (PAR). Les suites de PAR peuvent être écrites sous une forme markovienne et constituent un exemple de chaînes de Markov Harris récurrentes. La deuxième partie de la thèse porte sur les applications de la théorie des processus empiriques à l'apprentissage statistique. Nous obtenons des inégalités maximales de type exponentiel et polynomiales. Les inégalités de concentration sont un outil crucial pour étudier la capacité de généralisation des algorithmes d'apprentissage statistique via une approche empirique de minimisation des

risques (nous renvoyons à [142], [144] et [38] pour un exposé détaillé). Dans cette thèse, nous utilisons cette stratégie afin d'obtenir des bornes de généralisation pour un problème d'estimation d'ensembles de volumes minimaux.

Théorèmes de la limite centrale fonctionnel de type bootstrap pour des chaînes de Markov Harris récurrentes

La première contribution concerne les résultats asymptotiques pour les chaînes de Markov Harris récurrentes et a été publiée dans [44] et [25]. Nous étendons le théorème de la limite centrale bootstrap pour la moyenne établie dans [26]. Nous montrons que le théorème central limite (TCL) de type bootstrap uniforme est valable pour les classes de fonctions \mathcal{F} uniformément bornées, ainsi que dans les cas non bornés lorsque nous n'exigeons que des conditions de second ordre imposées à l'enveloppe F de \mathcal{F} .

Dans ce cadre, nous mesurons la complexité de \mathcal{F} par le *nombre de recouvrements* $N_p(\epsilon, Q, \mathcal{F})$, i.e. le nombre minimum de boules de rayon ϵ nécessaires pour couvrir \mathcal{F} au sens de la norme $L_p(Q)$, où Q est une mesure sur E dont le support est fini. Dans ce qui suit, nous imposons que *intégrale d'entropie uniforme* de \mathcal{F} soit finie, c'est-à-dire :

$$\int_0^\infty \sqrt{\log N_2(\epsilon, \mathcal{F})} d\epsilon < \infty, \quad \text{où } N_2(\epsilon, \mathcal{F}) = \sup_Q N_2(\epsilon, Q, \mathcal{F}).$$

Plus spécifiquement, nous montrons que sous certaines conditions techniques (spécifiées dans [44] et le deuxième chapitre de cette thèse) imposées à X , f , à la classe \mathcal{F} et au petit ensemble S et sous l'hypothèse

$$\int_0^\infty \sqrt{\log N_2(\epsilon, \mathcal{F})} d\epsilon < \infty$$

on obtient

$$\mathbb{Z}_n^* = n_{A_{\mathcal{M}}}^{*1/2} \left[\frac{1}{n_{A_{\mathcal{M}}}^*} \sum_{i=1}^{l_n^*-1} f(B_i^*) - \frac{1}{\hat{n}_{A_{\mathcal{M}}}} \sum_{i=1}^{\hat{l}_n-1} f(\hat{B}_i) \right] \quad (7.3)$$

converge en probabilité sous \mathbb{P}_ν vers un processus Gaussien G indexé par \mathcal{F} dont les trajectoires sont bornées et uniformément continus. Pour comprendre l'équation (7.3), nous mentionnons brièvement que $n_{A_{\mathcal{M}}}^*$ et l_n^* sont des équivalents bootstrap des quantités $n_{A_{\mathcal{M}}}$ et l_n introduites dans la sous-section 2 et B_i^* sont des blocs bootstrap, c'est-à-dire des suites i.i.d. de blocs échantillonnés avec remise à partir de la fonction de répartition empirique, basée sur les blocs régénératifs ou approximatifs.

Nos théorèmes sont des versions bootstrap des théorèmes des limites centrales uniformes pour les chaînes de Harris pour les classes de fonctions bornées présentées dans [98] et dans

le cas non borné (voir [137]). De plus, les techniques mises en oeuvre dans nos preuves permettent d'appliquer notre raisonnement à des cas régénératifs simplifiant ainsi la preuve du théorème uniforme de la limite centrale bootstrap pour les chaînes de Markov régénératives établies dans [122]. Les théorèmes de la limite centrale uniformes (et leurs versions de type bootstrap) sont un outil utile pour l'estimation non paramétrique, la théorie du maximum de vraisemblance, l'estimation de densité par noyaux ou l'estimation de la densité en ondelettes (voir par exemple [110], [68] et [69] pour plus de détails).

Nous utilisons les résultats susmentionnés pour établir des théorèmes de limite centrale uniformes pour les fonctionnelles de Fréchet différentiables des chaînes de Harris Markov. Sous certaines conditions techniques (condition d'entropie uniforme, hypothèses sur un petit ensemble S , etc.) nous avons que, pour une fonctionnelle Fréchet différentiable en μ (pour une métrique indexée par une classe de fonctions, voir [17] et le Chapitre 2 pour plus de détails) le bootstrap est valide asymptotiquement.

Estimation robuste pour les chaînes de Markov avec applications aux PDMP

Dans la contribution suivante, nous proposons une méthode pour construire des estimateurs robustes pour les chaînes de Markov régénératives atomiques et Harris récurrentes, en mettant l'accent sur les processus de Markov déterministes (PDMP). Dans ce cadre, nous nous appuyons sur la théorie du renouvellement des chaînes de Markov et sur le développement ultérieur de la méthode du bootstrap par blocs régénératifs approximatif. L'idée principale est d'éliminer les blocs ayant soit trop de contribution aux statistiques d'intérêt, soit ayant une longueur trop grande.

Il est connu (voir [29]) que certains concepts classiques de statistiques robustes peuvent être étendus naturellement à un cadre Markovien. Par exemple, une fonction d'influence peut être définie sur le tore $\mathbb{T} = \cup_{k=1}^{\infty} E^k$. En effet, prenons $\mathcal{P}_{\mathbb{T}}$ l'ensemble des mesures de probabilité sur le tore \mathbb{T} et pour tout $b \in \mathbb{T}$, définissons

$$L(b) = k \text{ si } b \in E^k, k \geq 1.$$

La fonction d'influence sur le tore peut alors être définie comme suit. Soient $(\mathcal{V}, \|\cdot\|)$ un espace de Banach séparable et $T : \mathcal{P}_{\mathbb{T}} \rightarrow \mathcal{V}$ une fonctionnelle sur $\mathcal{P}_{\mathbb{T}}$. Pour \mathcal{L} dans $\mathcal{P}_{\mathbb{T}}$, si

$$t^{-1}(T((1-t)\mathcal{L} + t\delta_b) - T(\mathcal{L}))$$

existe lorsque $t \rightarrow 0$ pour tout $b \in \mathbb{T}$, la fonction d'influence $T^{(1)} : \mathcal{P}_{\mathbb{T}} \rightarrow \mathcal{V}$ de la fonctionnelle

T en \mathcal{L} et pour tout b dans \mathbb{T} est définie par

$$T^{(1)}(b, \mathcal{L}) = \lim_{t \rightarrow 0} \frac{T((1-t)\mathcal{L} + t\delta_b) - T(\mathcal{L})}{t}. \quad (7.4)$$

La fonctionnelle $T : \mathcal{P}_{\mathbb{T}} \rightarrow \mathbb{R}$ est dite Fréchet différentiable en $\mathcal{L}_A \in \mathcal{P}_{\mathbb{T}}$ pour une métrique d , s'il existe un opérateur linéaire continu $DT_{\mathcal{L}_A}$ (de l'ensemble des mesures signées de la forme $\mathcal{L} - \mathcal{L}_A$ dans $(\mathbb{R}, \|\cdot\|)$) et une fonction

$$\epsilon^{(1)}(\cdot, \mathcal{L}_A) : \mathbb{R} \rightarrow (\mathbb{R}, \|\cdot\|),$$

qui est continue en 0 avec $\epsilon^{(1)}(0, \mathcal{L}_A) = 0$ telle que

$$\forall \mathcal{L} \in \mathcal{P}_{\mathbb{T}}, \quad T(\mathcal{L}) - T(\mathcal{L}_A) = DT_{\mathcal{L}_A}(\mathcal{L} - \mathcal{L}_A) + R^{(1)}(\mathcal{L}, \mathcal{L}_A),$$

où

$$R^{(1)}(\mathcal{L}, \mathcal{L}_A) = d(\mathcal{L}_A, \mathcal{L})\epsilon^{(1)}(d(\mathcal{L}_A, \mathcal{L}), \mathcal{L}_A).$$

De plus, T admet la représentation suivante,

$$\forall \mathcal{L}_A \in \mathcal{P}_{\mathbb{T}}, \quad DT_{\mathcal{L}_A}(\mathcal{L} - \mathcal{L}_A) = \int T^{(1)}(b, \mathcal{L}_A) \mathcal{L}(db),$$

où $T^{(1)}(b, \mathcal{L}_A)$ est la fonction d'influence en \mathcal{L}_A .

La différentiabilité de Fréchet est un outil standard pour obtenir des théorèmes de la limite centrale pour les estimateurs plug-in. En effet, supposons que $T : \mathcal{P}_{\mathbb{T}} \rightarrow \mathbb{R}$ est Fréchet différentiable en \mathcal{L}_A pour certaines métriques $d_{\mathcal{F}}$ (voir le Chapitre 3 pour plus de détails), où \mathcal{F} est une classe admissible de fonctions avec une enveloppe F , satisfaisant la condition d'entropie uniforme

$$\int_0^\infty \sqrt{\sup_Q \log N_2(\epsilon, Q, \mathcal{F})} d\epsilon < \infty.$$

Supposons en outre, dans le cas atomique régénératif

$$\mathbb{E}_A \left[\sum_{1 \leq j \leq \tau_A} F(X_j) \right]^2 < \infty, \quad \mathbb{E}_\nu(\tau_A) < \infty, \quad \mathbb{E}_A(\tau_A^2) < \infty.$$

Puis,

$$n^{1/2}(T(\mathcal{L}_n) - T(\mathcal{L}_A)) \rightarrow \mathcal{N}\left(0, \frac{\text{Var}(T^{(1)}(B_i, \mathcal{L}_A))}{E_A(\tau_A)}\right).$$

Nous établissons des résultats similaires pour les chaînes de Markov Harris récurrentes.

Dans ce cadre, nous considérons en particulier la construction d'estimateurs robustes pour la chaîne de Markov immergée associée au PDMP (processus dont le comportement est déterminé par des sauts aléatoires à des points du temps et dont l'évolution est régie de manière déterministe par une équation différentielle ordinaire entre ces moments).

Nous considérons des estimateurs robustes de plusieurs indicateurs de risque tels que la probabilité de ruine, le déficit attendu et l'indice extrémal de deux PDMP : le modèle Cramér-Lundberg et le modèle dynamique d'exposition à un risque alimentaire (brièvement KDEM -*Kinetic Dietary Exposure Model* en anglais) utilisé dans la modélisation de la pharmacocinétique des contaminants (voir [31] par exemple).

Les procédures d'échantillonnage résiduel et le wild bootstrap pour les processus autorégressifs périodiques (PAR)

Cette contribution est publiée dans [46]. Nous considérons un processus périodiquement autorégressif (PAR) de la forme

$$X_{nT+v} = \sum_{k=1}^p \phi_k(v) X_{nT+v-k} + \epsilon_{nT+v}, \quad (7.5)$$

où

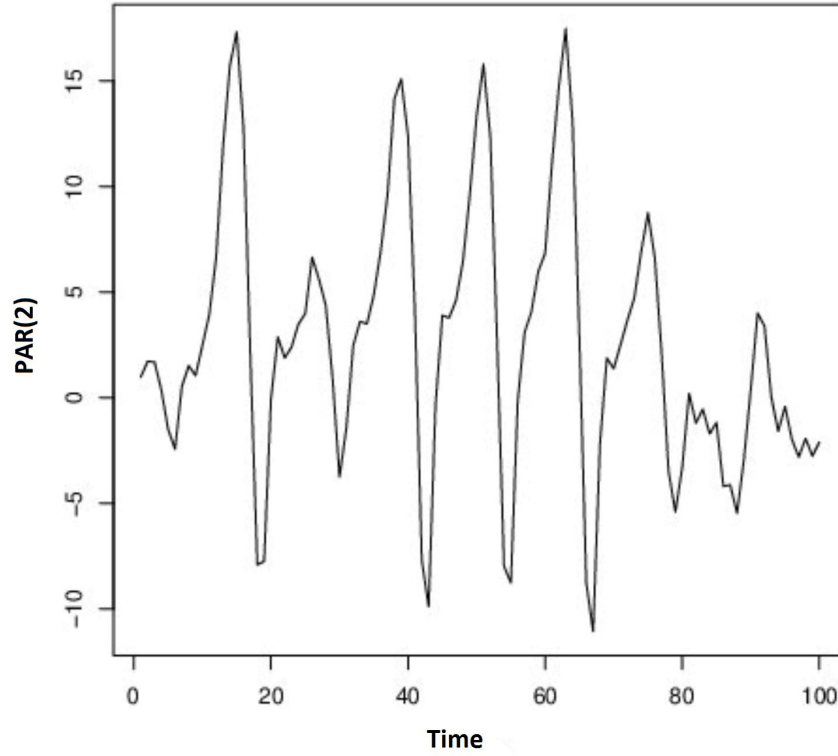
$$\Phi' = [\phi_1(1), \phi_2(1), \dots, \phi_p(1), \phi_1(2), \dots, \phi_p(2), \dots, \phi_1(T), \dots, \phi_p(T)]$$

désigne le vecteur de paramètres et $'$ est la transposée. $\{X_{nT+v}\}$ dénote la série au cours du n -ième cycle ($0 \leq n \leq N-1$) et v -ème saison ($1 \leq v \leq T$). Le $\{\epsilon_{nT+v}\}$ est le bruit blanc nul moyen avec une variance de la forme $\text{Var}(\epsilon_{nT+v}) = \sigma_v^2 > 0$ pour toutes les saisons v . Le processus dans (7.5) peut être écrit sous une forme Markovienne en utilisant une astuce de vectorisation analogue à celle de l'exemple 2 puisque le processus PAR peut être écrit sous la forme d'un modèle autorégressif (AR) T -varié (voir [20] pour plus de détails). Nous avons obtenu les estimateurs des moindres carrés des paramètres du modèle

$$\hat{\Phi}' = [\hat{\phi}_1(1), \hat{\phi}_2(1), \dots, \hat{\phi}_p(1), \hat{\phi}_1(2), \dots, \hat{\phi}_p(2), \dots, \hat{\phi}_1(T), \dots, \hat{\phi}_p(T)]$$

afin de générer leurs équivalents bootstrap.

Figure 7.2 : Trajectoire du processus PAR(2) avec 12 saisons et 100 observations.



Nous proposons des procédures wild bootstrap et résiduel les basées sur les données (car elles ne nécessitent pas de choix de la longueur des blocs bootstrap), elles peuvent donc être intéressantes pour une utilisation pratique. La procédure d'échantillonnage résiduel pour les processus autorégressifs périodiques comprend les étapes suivantes.

La procédure d'échantillonnage résiduel pour les processus autorégressifs périodiques

Étape 1 Calculer l'estimateur des moindres carrés ordinaires $\hat{\Phi}$ de Φ .

Étape 2 Calculer les résidus du modèle estimé

$$\hat{\epsilon}_{nT+v} = X_{nT+v} - \sum_{k=1}^p \hat{\phi}_k(v) X_{nT+v-k},$$

où $1 \leq v \leq T$, $0 \leq n \leq N - 1$.

Étape 3 Calculer les résidus centrés

$$\bar{\eta}_{nT+v} = \frac{\hat{\epsilon}_{nT+v}}{\sigma_v} - \frac{1}{NT} \sum_{n=0}^{N-1} \sum_{v=1}^T \frac{\hat{\epsilon}_{nT+v}}{\sigma_v},$$

où NT est le nombre total d'observations dans le modèle.

Étape 4 Générer des variables η_{nT+v}^* en tirant avec remise parmi

$$\{\bar{\eta}_1, \dots, \bar{\eta}_{NT}\}.$$

Étape 5 Générer la version bootstrap du modèle (7.5)

$$X_{nT+v}^* = \sum_{k=1}^p \hat{\phi}_k(v) X_{nT+v-k} + \sigma_v \eta_{nT+v}^*, \quad 1 \leq v \leq T.$$

Étape 6 Calculer les estimateurs bootstrap des paramètres pour chaque saison v , $1 \leq v \leq T$
 $\hat{\Phi}^*(v)$, où

$$z^*(v) = [X_v^*, \dots, X_{(N-1)T+v}^*]', \quad 1 \leq v \leq T.$$

La deuxième méthode que nous proposons est la procédure de wild bootstrap pour les processus PAR.

La procédure wild bootstrap pour les processus autorégressifs périodiques

Étape 1 Calculer l'estimateur des moindres carrés ordinaire $\hat{\Phi}$ de Φ .

Étape 2 Calculer les résidus du modèle estimé

$$\hat{\epsilon}_{nT+v} = X_{nT+v} - \sum_{k=1}^p \hat{\phi}_k(v) X_{nT+v-k},$$

où $1 \leq v \leq T$, $0 \leq n \leq N-1$.

Étape 3 Générer le processus de bootstrap X_{nT+v}^\dagger pour chaque saison v , $1 \leq v \leq T$

$$X_{nT+v}^\dagger = \sum_{k=1}^p \hat{\phi}_k(v) X_{nT+v-k} + \epsilon_{nT+v}^\dagger$$

et

$$\epsilon_{nT+v}^\dagger = \hat{\epsilon}_{nT+v} \eta_{nT+v}^\dagger,$$

où $\eta_{nT+v}^\dagger \sim \mathcal{N}(0, 1)$ et $(\eta_{nT+v}^\dagger)_{nT+v \in \mathbb{R}}$ est indépendant de $\hat{\epsilon}_{nT+v}$.

Étape 4 Calculer l'estimateur bootstrap des paramètres $\widehat{\Phi}^\dagger(v)$ pour chaque saison v , $1 \leq v \leq T$,
où

$$z^\dagger(v) = \left[X_v^\dagger, \dots, X_{(N-1)T+v}^\dagger \right]'.$$

Nous obtenons une consistance faible pour les deux méthodes. Plus précisément, nous avons montré pour une série autorégressive périodique causale X_{nT+v} définie dans (7.5) avec un quatrième moment fini que, la procédure de bootstrap résiduel que nous avons proposée est faiblement cohérente, c'est-à-dire, quand $N \rightarrow \infty$,

$$\sqrt{N} \left(\widehat{\Phi}^* - \widehat{\Phi} \right) \xrightarrow{P^*} \mathcal{N} \left(0, R^{-1} \right),$$

où

$$\widehat{\Phi}^{*'} = \left[\widehat{\phi}_1^*(1), \widehat{\phi}_2^*(1), \dots, \widehat{\phi}_p^*(1), \widehat{\phi}_1^*(2), \dots, \widehat{\phi}_p^*(2), \dots, \widehat{\phi}_1^*(T), \dots, \widehat{\phi}_p^*(T) \right]$$

est un vecteur d'estimateurs de bootstrap des paramètres pour chaque saison v obtenue à partir de l'algorithme bootstrap résiduel et R est spécifié dans le Chapitre 4. Des résultats de cohérence similaires sont obtenus pour la procédure du wild bootstrap (voir le Chapitre 4).

Enfin, nous illustrons nos considérations théoriques par des simulations.

Inégalités maximales de type exponentiel et de type polynomial pour les chaînes Harris récurrentes

Motivés par les applications en apprentissage statistique, nous établissons des bornes pour la probabilité de dépassement des processus empiriques dans un cadre Markovien. Ces contributions sont présentées dans [22], [23] et [24].

Soit $f : E \rightarrow \mathbb{R}$ une fonction mesurable. Puisque nos inégalités sont de type maximales, nous contrôlons la classe de fonctions via un nombre d'entropie uniforme. Sous l'existence des moments exponentiels sur τ_A et $f(B_j)$, nous établissons des inégalités maximales de type Bernstein et Hoeffding en fonction du nombre d'entropie uniforme et des moments des temps de retours et des blocs. Une des principales difficultés lors de la détermination de telles bornes est que même si nous supposons que f est bornée, $f(B_j)$ peut être non bornée sur tout un bloc d'observations. Pour établir les inégalités, nous nous appuyons beaucoup sur l'inégalité de Montgomery-Smith et les techniques de symétrisation de [119].

Nous montrons aussi que dans des conditions plus faibles imposées sur les temps de retours et les $f(B_j)$, les bornes polynomiales peuvent être établies. Les conditions imposées sont satisfaites par des chaînes de Markov sous-géométriquement ergodiques, auxquelles notre inégalité de type polynomiale peut être appliquée.

De plus, nous établissons des bornes pour l'espérance du suprémum du processus empirique dans un contexte Markovien, car elles s'avèrent particulièrement utiles lorsque l'on veut sélectionner un modèle via un critère de pénalisation avec un terme de pénalité dépendant de la complexité de l'ensemble des modèles.

Nous détaillons les résultats susmentionnés dans le Chapitre 5. Cependant, pour donner au lecteur un aperçu général des résultats obtenus, nous fournissons les bornes sous une forme générale (et un peu simplifiée) ci-dessous. Les conditions détaillées imposées sur la chaîne X sont omises ici et énoncées dans d'autres sections. Par souci de simplicité, nous fournissons les résultats uniquement dans le cas de la régénération atomique (nous formulons les inégalités pour les chaînes de Markov Harris récurrentes dans Chapitre 5). Soit

$$\sigma_m^2 = \max_{f \in \mathcal{F}} \sigma^2(f) > \eta > 0.$$

- **Une inégalité de concentration de type Bernstein** On suppose que $N_1(\epsilon, \mathcal{F}) < \infty$. Ensuite, dans des conditions de moments exponentiels des blocs et de temps de retour à l'atome A , nous avons pour tout $x > 0$, $0 < \epsilon < x/2$ et pour tout $n \geq 1$

$$\mathbb{P}_\nu \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(X_i) - \mu(f) \right| \geq x \right] \leq N_1(\epsilon, \mathcal{F}) K_1 \left\{ \exp \left[\frac{-n(x - 2\epsilon)^2}{K_2(\sigma_m^2 + K_3(x - 2\epsilon))} \right] \right\},$$

où K_1 , K_2 et K_3 sont des paramètres positifs spécifiés au le Chapitre 5.

- **Une inégalité de concentration de type Hoeffding** On suppose que $N_1(\epsilon, \mathcal{F}) < \infty$. On suppose en outre que la classe de fonctions \mathcal{F} est uniformément bornée. Sous des conditions de moments exponentiels des blocs et du temps de retour à l'atome A , nous avons pour tout $x > 0$, $0 < \epsilon < x/2$ et pour tout $n \geq 1$

$$\mathbb{P}_\nu \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(X_i) - \mu(f) \right| \geq x \right] \leq N_1(\epsilon, \mathcal{F}) L_1 \left\{ \exp \left[-\frac{n(x - 2\epsilon)^2}{L_2 D^2} \right] \right\},$$

où D est une constante telle que $\forall f \in \mathcal{F} |f| < D$ et L_1 et L_2 sont des paramètres positifs spécifiés au Chapitre 5.

- **Une inégalité de moments polynomiales** On suppose que $N_1(\epsilon, \mathcal{F}) < \infty$. Supposons de plus que le moment du bloc p et les moments des temps de retour à l'atome A soient finis. Nous avons pour tout $x > 0$, $0 < \epsilon < x/2$ et pour tout $n \geq 1$

$$\mathbb{P}_\nu \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(X_i) - \mu(f) \right| \geq x \right] \leq C_1 \frac{N_1(\epsilon, \mathcal{F})}{(x - 2\epsilon)^p n^{p/2}}$$

et C_1 est un paramètre positif spécifié au Chapitre 5.

- **Une borne pour l'espérance de supremum de processus empiriques** On suppose que

$$\mathbb{E}_A[l(B_1)]^2 < \infty \text{ et } \mathbb{E}_A[F(B_1)]^2 < \infty,$$

où F est une enveloppe pour \mathcal{F} . De plus, supposons que le nombre d'entropie uniforme $N_1\left(\frac{\epsilon}{R_1}, \mathcal{F}\right) < \infty$. Pour tout $\epsilon > 0$ nous avons

$$\begin{aligned} & \mathbb{E}_A \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^{l_n} (f(B_i) - \mu(f(B_1))) \right| \right] \\ & \leq R_2 \left[\epsilon + \mathcal{N}\left(\frac{\epsilon}{R_1}, \mathcal{F}\right) \times \mathbb{E}_A[F(B_1)^2]^{1/2} \sqrt{\frac{2 \log N_1\left(\frac{\epsilon}{R_1}, \mathcal{F}\right)}{n}} \right], \end{aligned}$$

où R_1 et R_2 sont des constantes positives pouvant être calculées explicitement.

Les résultats ci-dessus peuvent facilement être généralisés au cas Harris récurrent.

Bornes de généralisation pour le problème d'estimation d'ensembles de volume minimum

Les dernières contributions présentées dans cette thèse sont des bornes de généralisation pour l'estimation d'ensemble de volume minimum (ensemble MV) pour des chaînes de Markov régénératives et Harris récurrentes. Les résultats ont été présentés dans [23] et [24]. Le problème d'estimation de MV-set a été proposé pour la première fois dans [120] dans le cadre i.i.d..

Soit μ une loi de probabilité sur un espace mesurable (E, \mathcal{E}) . Soient $\alpha \in (0, 1)$ et λ une mesure de référence σ -finie sur (E, \mathcal{E}) , toute solution du problème de minimisation (7.6)

$$\min_{\Omega \in \mathcal{E}} \lambda(\Omega) \text{ sous contrainte } \mu(\Omega) \geq \alpha \quad (7.6)$$

est appelée un ensemble *MV* de niveau α . La loi μ est supposée être absolument continue par rapport à la mesure de Lebesgue λ et on note

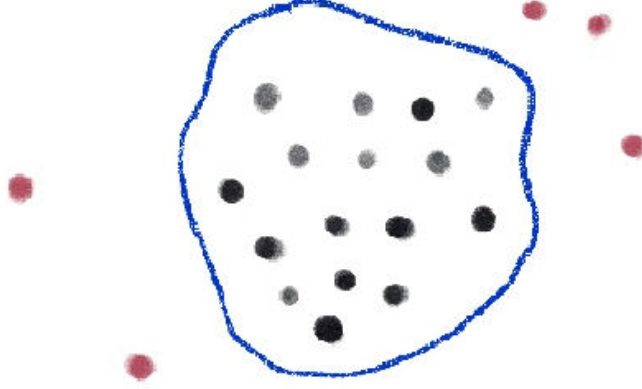
$$f(x) = (d\mu/d\lambda)(x)$$

la densité associée.

Sous certaines hypothèses techniques sur f , pour tout $\alpha \in (0, 1)$, il a été montré dans [120] que l'ensemble

$$\Omega_\alpha^* = \{x \in E : f(x) > \gamma_\alpha\},$$

Figure 7.3 : Pour les grandes valeurs de α , les points de l'échantillon qui sont dans d'ensemble de volumes minimum seront considérés comme normaux; et les autres seront considérés comme des anomalies.



où γ_α est le nombre unique tel que

$$\int_{f(x) > \gamma_\alpha} f(x) d\lambda(x) = \alpha$$

est la solution unique au problème d'estimation d'ensemble MV (7.6).

Les ensembles de volumes minimaux peuvent être interprétés comme suit : pour les petites valeurs du niveau de masse α , les ensembles *MV* permettent de récupérer les modes de la loi, tandis que les ensembles complémentaires correspondent aux *observations rares* lorsque α est grand.

En pratique, la loi μ est inconnue et est remplacée par sa version empirique μ_n . Trouver un volume minimal de niveau α revient à résoudre le problème de minimisation suivant

$$\min_{\Omega \in \mathcal{E}} \lambda(\Omega) \text{ sous contrainte } \hat{\mu}_n(\Omega) \geq \alpha - \psi_n, \quad (7.7)$$

ψ_n étant un paramètre de tolérance (voir le Chapitre 6 pour plus de détails).

La technique d'estimation d'ensembles de volume minimaux peut être utilisée en tant qu'algorithme de détection d'anomalie non supervisée puisque, dans le cas de données non étiquetées, nous considérons l'anomalie comme un événement rare.

Scott et Nowak [126] ont établi des bornes de généralisation pour le problème d'estimation d'ensembles *MV* dans le cadre i.i.d.. Afin d'établir des bornées de généralisation dans un

cadre Markovien, nous nous restreignons au cadre de minimisation du risque empirique et, par conséquent, nous nous appuyons fortement sur les inégalités de concentration, ce qui nous permet de contrôler le supremum des processus empiriques impliqués. Notre approche se résume à la décomposition de la fonction de loi empirique d'intérêt en : $\forall \Omega \in \mathcal{E}$,

$$\hat{\mu}_n(\Omega) = \frac{1}{n} \sum_{i=1}^{\tau_A} \mathbb{I}\{X_i \in \Omega\} + \frac{l_n - 1}{n} \left(\frac{1}{l_n - 1} \sum_{j=1}^{l_n - 1} S_j(\Omega) \right) + \frac{1}{n} \sum_{i=1+\tau_A(l_n)}^n \mathbb{I}\{X_i \in \Omega\}, \quad (7.8)$$

où $l_n = \sum_{i=1}^n \mathbb{I}\{X_i \in A\}$ indique le nombre de visites de l'atome A (*régénérations*), les temps d'occupation du Ω entre les j -ème et $(j+1)$ -ème temps de régénération est indiqué par

$$S_j(\Omega) = \sum_{\tau_A(j) < i \leq \tau_A(j+1)} \mathbb{I}\{X_i \in \Omega\}.$$

La décomposition (7.8) ainsi que l'application de l'inégalité maximale de type polynomiale pour les chaînes de Markov nous permettent d'étendre le résultat de [126] au cas régénératif atomique et au cas Harris récurrent.

Soit $p \geq 2$. Afin d'établir des bornes de généralisation pour le problème d'estimation d'ensembles de volumes minimaux, nous supposons que

$$\mathbb{E}_A[\tau_A^p] < \infty \text{ et } \mathbb{E}_\nu[\tau_A^p] < \infty.$$

Soit $r \geq 1$. La collection de fonctions indicatrices sur E ,

$$\mathcal{F} = \{\mathbb{I}\{x \in \Omega\} : \Omega \in \mathcal{G}\}$$

est une classe uniforme de Donsker (par rapport à L_1) avec des nombres de couverture uniformes polynomiaux tels qu'il existe une constante $c > 0$ telle que $\forall \zeta > 0$,

$$N_1(\zeta, \mathcal{F}) \stackrel{\text{def}}{=} \sup_Q N(\zeta, \mathcal{F}, L_1(Q)) \leq c(1/\zeta)^r,$$

où le supremum est pris sur l'ensemble des mesures de probabilité discrètes sur (E, \mathcal{E}) .

Sous les hypothèses précédentes sur les moments de retour de X à l'atome A et sur la complexité de la classe \mathcal{F} et pour tout $\delta \in (0, 1)$, et si ψ_n (nous spécifions ψ_n au le Chapitre 6) est une pénalité bien choisie pour la classe \mathcal{G} , alors, avec probabilité au moins $1 - \delta$,

$$\lambda(\hat{\Omega}_n) \leq \lambda(\Omega_\alpha^*) + \left\{ \inf_{\Omega \in \mathcal{G}: \mu(\Omega) \geq \alpha} \lambda(\Omega) - \lambda(\Omega_\alpha^*) \right\}$$

et

$$\mu(\hat{\Omega}_n) \geq \alpha - 2\psi_n(\delta).$$

4 Liste des travaux

Publications dans des revues à comité de lecture internationales

G. Ciolek. Bootstrap uniform central limit theorems for Harris recurrent Markov chains. *Electronic Journal of Statistics*, 10 :2157–2178, 2016.

G. Ciolek and P. Potorski. Bootstrapping periodically autoregressive models. *ESAIM : Probability and Statistics*, 21 :2157–2178, 2017.

P. Bertail and G. Ciolek. New Bernstein and Hoeffding type inequalities for regenerative Markov chains. Accepted for publication in *Latin American Journal of Probability and Mathematical Statistics*, 2018.

Publications dans des ouvrages

P. Bertail, G. Ciolek, and C. Tillier. Robust estimation for Markov chains with application to PDMP. In *Statistical Inference for Piecewise-Deterministic Markov Processes*. Wiley, 2018.

G. Ciolek. Bootstrapping Harris recurrent Markov chains. In *Nonparametric Statistics*, 3rd ISNPS, Avignon, France, June 2016, Springer Proceedings in Mathematics and Statistics. Due to December, 2018.

Publications dans des conférences à comité de lecture internationales

P. Bertail, G. Ciolek, and S. Cléménçon. Generalization bounds for minimum volume set estimation based on Markovian data. *ISAIM, International Symposium on Artificial Intelligence and Mathematics Proceedings Book*, pages 1–7, 2018.

Pré-publications

P. Bertail, G. Ciolek, and S. Cléménçon. Statistical learning based on Markovian data. Maximal deviation inequalities and learning rates. Submitted, 2018.

5 Liste des communications orales

Juillet 2018, Présentation à la 12th International Vilnius Conference on Probability Theory and Mathematical Statistics and 2018 IMS Annual Meeting on Probability and Statistics, Vilnius (Lituanie), *Rosenthal type inequality for regenerative Markov chains*,

Juin 2018, Présentation au Meeting IMS Asia Pacific Rim, Singapour, *Robust estimation for Piecewise Deterministic Markov Processes*,

Juin 2018, Présentation à la Conference of the international Society for Nonparametric Statistics, Salerno, Italie, *Learning minimum volume sets from regenerative Markov chains*,

Juin 2018, Présentation à la Conference Statistical Learning and Data Science, New York, États-Unis, *Generalization bounds for ERM algorithm with atomic regenerative Markov chain samples*,

Avril 2018, Présentation au séminaire de la Faculty of Economics, University of Tokyo, Tokyo, Japon, *Bootstrap uniform theorems for Harris recurrent Markov chains*,

Janvier 2018, Présentation à la Conference Artificial Intelligence and Mathematics, Fort Lauderdale, États-Unis, *Generalization bounds for minimum volume set estimation problem for Markovian data*,

Juillet 2017, Présentation à la Stochastic Processes and Applications Conference, Moscou, Russie, *Exponential inequalities for regenerative Markov chains*,

Mai 2017, Présentation à la conférence Les probabilités de demain, Paris, France, *Sharp Bernstein and Hoeffding type inequalities for regenerative Markov chains*,

Mai 2017, Présentation au séminaire généraliste de l'équipe de Probabilités et Statistiques, Université de Lorraine, Nancy, France, *Concentration inequalities for regenerative and Harris recurrent Markov chains with applications to statistical learning*,

Mai 2017, Présentation à la Journée annuelle de la Chaire Machine Learning for Big Data, Paris, France, *Bootstrapping periodically autoregressive models*,

Janvier 2017, Présentation au Séminaire-AgroParisTech, Paris, France, *Bootstrap uniform central limit theorems for Harris recurrent Markov chains*,

Juin 2016, Présentation à la Conference International Society for NonParametric Statistics, Avignon, France, Prix Jeune Chercheur *Bootstrap uniform central limit theorems for Harris recurrent Markov chains*,

Mai 2016, Présentation au Workshop Dependence, Stability, and Extremes, Toronto, Canada, *Bootstrapping Harris recurrent Markov chains*,

Fevrier 2016, Présentation au Workshop COMputational Statistics and MOlecular Simulation, Paris, France, *Bootstrap uniform central limit theorems for Harris recurrent Markov chains*,

Avril 2016, Présentation au Seminaire-MODAL'X, Université Paris X, Nanterre, France, *Bootstrap uniform central limit theorems for Harris recurrent Markov chains*.

Bibliography

- [1] R. Adamczak. A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electronic Journal of Probability*, 34:1000–10034, 2008.
- [2] R. Adamczak and W. Bednorz. Exponential concentration inequalities for additive functionals of Markov chains. *ESAIM: Probability and Statistics*, 19:440–481, 2015.
- [3] T.M. Adams and A.B. Nobel. Uniform convergence of Vapnik-Chervonenkis classes under ergodic sampling. *Annals of Probability*, 38:1345–1367, 2010.
- [4] A. Agarwal and J. Duchi. The generalization ability of online algorithms for dependent data. *IEEE Transactions on Information Theory*, 59:573–587, 2013.
- [5] P. Alquier and O. Wintenberger. Model selection for weakly dependent time series forecasting. *Bernoulli*, 18:883–913, 2012.
- [6] G. Alsmeyer. On the Harris recurrence and iterated random Lipschitz functions and related convergence rate results. *Journal of Theoretical Probability*, 16:217–247, 2003.
- [7] R. Arora, W. Sethares, and J. Bucklew. Latent periodicities in genome sequences. *IEEE Journal of Selected Topics in Signal Processing*, 2:332–342, 2008.
- [8] S. Asmussen and H. Albrecher. *Ruin Probabilities*. Applications of Mathematics (New York). World Scientific Publishing C0. Pte. Ltd, New York, second edition, 2010. Stochastic Modelling and Applied Probability.
- [9] K. Athreya and G. Atuncar. Kernel estimation for real-valued Markov chains. *Sankhyā: The Indian Journal of Statistics, Series A*, 60:1–17, 1998.
- [10] K.B. Athreya and C.D. Fuh. Bootstrapping Markov chains: countable case. *Journal of Statistical Planning and Inference*, 33:311–331, 1992.

- [11] K.B. Athreya and C.D. Fuh. Central limit theorem for a double array of Harris chains. *Sankhyā: The Indian Journal of Statistics, Series A*, 55:1–11, 1993.
- [12] K.B. Athreya and P. Ney. A new approach to the limit theory of recurrent Markov chains. *Transactions of the American Mathematical Society*, 245:493–501, 1978.
- [13] K.B. Athreya and S.G. Pantula. Mixing properties of Harris chains and autoregressive processes. *Journal of Applied Probability*, 23:880–892, 1986.
- [14] R. Azaïs and F. Bouguet. *Statistical Inference for Piecewise-deterministic Markov Processes*. Wiley, 2018.
- [15] Z. Bai and Z. Lin. *Probability Inequalities*. Springer, Science Press, Beijing, 2009.
- [16] Y. Baraud. Bounding the expectation of the supremum of an empirical process over a (weak) VC-major class. *Electronic Journal of Statistics*, 10:1709–1728, 2016.
- [17] P. Barbe and P. Bertail. *The Weighted Bootstrap. Lecture Notes in Statistics*. Springer Verlag, New-York, 1995.
- [18] P. Barbe, A. Fougères, and C. Genest. On the tail behavior of sums of dependent risks. *Astin Bulletin*, 36:361–373, 2006.
- [19] I.V. Basawa and R.B. Lund. Recursive prediction and likelihood evaluation for periodic ARMA models. *Journal of Time Series Analysis*, 21:75–93, 2000.
- [20] I.V. Basawa and R.B. Lund. Large sample properties of parameter for periodic ARMA models. *Journal of Time Series Analysis*, 21:75–93, 2001.
- [21] J. Berkowitz and L. Kilian. Recent developments in bootstrapping time series. *Econometric Reviews*, 19:1–54, 2000.
- [22] P. Bertail and G. Ciolek. New Bernstein and Hoeffding type inequalities for regenerative Markov chains. *Accepted in Latin American Journal of Probability and Mathematical Statistics*, 2018.
- [23] P. Bertail, G. Ciolek, and S. Cléménçon. Generalization bounds for minimum volume set estimation based on Markovian data. *ISAIM, International Symposium on Artificial Intelligence and Mathematics Proceedings Book*, pages 1–7, 2018.
- [24] P. Bertail, G. Ciolek, and S. Cléménçon. Statistical learning based on Markovian data. Maximal deviation inequalities and learning rates. *Submitted*, 2018.

- [25] P. Bertail, G. Ciolek, and C. Tillier. Robust estimation for Markov chains with application to PDMP. *In Statistical Inference for Piecewise-Deterministic Markov Processes. Wiley*, 2018.
- [26] P. Bertail and S. Cl  men  on. Regenerative block bootstrap for Markov chains. *Bernoulli*, 12:689–712, 2005.
- [27] P. Bertail and S. Cl  men  on. Regeneration-based statistics for Harris recurrent Markov chains. *Dependence in Probability and Statistics, Lecture Notes in Statistics. Springer*, 2006.
- [28] P. Bertail and S. Cl  men  on. Sharp bounds for the tails of functionals of Markov chains. *Theory of Probability and Its Applications*, 54:505–515, 2010.
- [29] P. Bertail and S. Cl  men  on. Bootstrapping robust statistics for Markovian data. Applications to regenerative \mathcal{R} - and \mathcal{L} -statistics. *Journal of Time Series Analysis*, 36:462–480, 2015.
- [30] P. Bertail, S. Cl  men  on, and C. Tillier. Extreme values statistics for Markov chains with applications to finance and insurance. *Extreme Events in Finance: A Handbook of Extreme Value Theory and its Applications. Wiley*, 2016.
- [31] P. Bertail, S. Cl  men  on, and J. Tressou. A storage model for modelling exposure to food contaminants. *Mathematical Biosciences and Engineering*, 5:35–60, 2006.
- [32] P. Bertail, S. Cl  men  on, and J. Tressou. Extreme value statistics for Markov chains via the (pseudo-) regenerative method. *Extremes*, 12:327–360, 2009.
- [33] P. Bertail, S. Cl  men  on, and J. Tressou. Regenerative block-bootstrap confidence intervals for the tail and extremal indexes. *Electronic Journal of Statistics*, 7:1224–1248, 2013.
- [34] A. Bibi and I. Lescheb. A note on integrated periodic garch processes. *Statistics and Probability Letters*, 87:121–124, 2014.
- [35] P. Bickel and J. Kwon. Inference for semiparametric models: Some current frontiers. *Statistica Sinica*, 11:863–960, 2001.
- [36] E. Bolthausen. The Berry-Essen theorem for strongly mixing Harris recurrent Markov chains. *Zeitschrift f  r Wahrscheinlichkeitstheorie und verwandte Gebiete*, 60:283–289, 1982.

- [37] S. Boucheron, G. Lugosi, and P. Massart. *Inequalities. A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.
- [38] O. Bousquet, S. Boucheron, and G. Lugosi. Introduction to statistical learning theory. *In Advanced Lectures on Machine Learning*. Springer, 2004.
- [39] P.J. Brockwell and R.A. Davis. *Introduction to Time Series and Forecasting*. New York, Springer, 2002.
- [40] E. Broszkiewicz-Suwaj, A. Makagon, R. Weron, and A. Wyłomańska. On detecting and modeling periodic correlation in financial data. *Physica A: Statistical Mechanics and its Applications*, 336:196–205, 2004.
- [41] J.R. Chazottes and F. Redig. Concentration inequalities for Markov processes via coupling. *Electronic Journal of Probability*, 14:1162–1180, 2009.
- [42] V. Chernozhukov, D. Chetverikov, and K. Kato. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, 41:2786—2819, 2013.
- [43] V. Chernozhukov, D. Chetverikov, and K. Kato. Central limit theorems and bootstrap in high dimensions. *The Annals of Probability*, 45:2309–2352, 2017.
- [44] G. Ciolek. Bootstrap uniform central limit theorems for Harris recurrent Markov chains. *Electronic Journal of Statistics*, 10:2157–2178, 2016.
- [45] G. Ciolek. Bootstrapping Harris recurrent Markov chains. *In Nonparametric Statistics, 3rd ISNPS, Avignon, France, June 2016, Springer Proceedings in Mathematics and Statistics*, 2018.
- [46] G. Ciolek and P. Potorski. Bootstrapping periodically autoregressive models. *ESAIM: Probability and Statistics*, 21:2157–2178, 2017.
- [47] S. Cléménçon. Moment and probability inequalities for sums of bounded additive functionals of a regular Markov chains via the Nummelin splitting technique. *Statistics and Probability Letters*, 55:227–238, 2001.
- [48] S. Cléménçon, P. Bertail, and G. Papa. Learning from survey training samples: rate bounds for Horvitz-Thompson risk minimizers. *In Proceedings of ACML’16*, 2016.

- [49] F. Cucker and S. Smale. On the mathematical foundations of learning. *Bulletin of the American Mathematical Society*, 39:1–49, 2002.
- [50] S. Datta and W. McCormick. Some continuous Edgeworth expansions for Markov chains with applications to bootstrap. *Journal of Multivariate Analysis*, 52:83–106, 1995.
- [51] M.H.A. Davis. Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models. *Journal of the Royal Statistical Society. Series B*, 46:353–388, 1984.
- [52] V. de la Pena and E. Giné. *Decoupling: from Dependence to Independence*. Springer, 1999.
- [53] J. Dedecker and S. Gouëzel. Subgaussian concentration inequalities for geometrically ergodic Markov chains. *Electronic Communications in Probability*, 20:1–12, 2015.
- [54] H. Dehling, T. Mikosch, and M. Sørensen. *Empirical Process Techniques for Dependent Data*. Springer Science & Business Media, 2002.
- [55] J. Di and E. Kolaczyk. Complexity-penalized estimation of minimum volume sets for dependent data. *Journal of Multivariate Analysis*, 101:1910–1926, 2004.
- [56] J. Dowell, S. Weiss, D. Hill, and D. Infield. A cyclo-stationary complex multichannel Wiener filter for the prediction of wind speed and direction. *Proceedings of the 21st European Signal Processing Conference*, 2013.
- [57] R. Duc, E. Moulines, and P. Soulier. Computable convergence rates for sub-geometric ergodic Markov chains. *Bernoulli*, 13:831–848, 2007.
- [58] R.M. Dudley. Nonlinear functionals of empirical measures and the bootstrap. *Probability in Banach Spaces. Birkhäuser, Boston*, 7:63–82, 1990.
- [59] B. Efron. Bootstrap methods: another look at the jackknife. *Annals of Statistics*, 7:1–26, 1979.
- [60] B. Efron and R. Tibshirani. Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy. *Statistical Science*, 1:54–77, 1986.
- [61] J.H.J. Einmahl and D.M. Mason. Generalized quantile process. *The Annals of Statistics*, 20:1062–1078, 1992.

- [62] P. Embrechts, C. Kluppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, 1997.
- [63] V. Feuillard, A. Gramfort, and A. Thomas. Calibration of one-class SVM for MV set estimation. *IEEE International Conference on Data Science and Advanced Analytics*, 1998.
- [64] J. Friedman, T. Hastie, and R. Tibshirani. *The Elements of Statistical Learning*. Springer, 2009.
- [65] F. Gao and L. Wu. Bernstein type’s concentration inequalities for symmetric Markov processes. *Theory of Probability and its Applications*, 58:358–382, 2014.
- [66] C. Gaucherel. Analysis of enso interannual oscillations using non-stationary quasi-periodic statistics. *International Journal of Climatology*, 30:926–934, 2010.
- [67] F. Ghaderi, K. Nazarpour, J.G. McWhirter, and S. Sanei. Removal of ballistocardiogram artifacts using the cyclostationary source extraction method. *IEEE Transactions on Biomedical Engineering*, 57:2667–2676, 2010.
- [68] E. Giné and R. Nickl. Uniform central limit theorems for kernel density estimators. *Probability Theory and Related Fields*, 141:333–387, 2008.
- [69] E. Giné and R. Nickl. Uniform limit theorems for wavelet density estimators. *The Annals of Probability*, 37:1605–1646, 2009.
- [70] E. Giné and V. Koltchinskii. Concentration inequalities and asymptotic results for ratio type empirical processes. *The Annals of Probability*, 34:1143–1216, 2006.
- [71] E. Giné and J. Zinn. Some limit theorems for empirical processes. *The Annals of Probability*, 12:929–989, 1984.
- [72] E. Giné and J. Zinn. Bootstrapping general empirical measures. *The Annals of Probability*, 18:851–869, 1990.
- [73] E.G. Gladyshev. Periodically correlated random sequences. *Soviet Mathematics*, 2:385–388, 1961.
- [74] P.W. Glynn and D. Ormoneit. Hoeffding’s inequality for uniformly ergodic Markov chains. *Statistics and Probability Letters*, 56:143–146, 2002.

- [75] A. Gorst-Rasmussen and M. Bøgsted. Asymptotic inference for waiting times and patiences in queues with abandonment. *Communication in Statistics- Simulation and Computation*, 38:318–334, 2009.
- [76] P. Hall. *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, 1992.
- [77] J.D. Hamilton. *Time Series Analysis*. New Jersey, Princeton University Press, 1994.
- [78] S. Hanneke. Learning whenever learning is possible: Universal learning under general stochastic processes. *arXiv:1706.01418*, 2017.
- [79] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.
- [80] J.L. Horowitz. The bootstrap. *Handbook of Econometrics*. Elsevier, 5:3159–3228, 2001.
- [81] B. Iqelan. *Periodically Correlated Time Series: Models and Examples*. Lambert Academic Publishing, 2011.
- [82] J. Jain and B. Jamison. Contributions to Doeblin’s theory of Markov processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 8(9):19–40, 1964.
- [83] R.H. Jones and W.M. Brelsford. Time series with periodic structure. *Biometrika*, 54:403–408, 1967.
- [84] K.Y. Kim, B. Hamilton, and H. Na. Theoretical foundation of cyclostationary EOF analysis for geophysical and climatic variables. *Concepts and Examples, Earth-Science Reviews*, 150:201–218, 2015.
- [85] A. Klenke. *Probability Theory*. Universitext, Springer,, London, German edition, 2014. A comprehensive course.
- [86] I. Kontoyiannis and S.P. Meyn. Geometric ergodicity and the spectral gap of non-reversible Markov chains. *Probability Theory and Related Fields*, 154:327–339, 2012.
- [87] M.R. Kosorok. *Introduction to Empirical Processes and Semiparametric Inference*. Springer, 2008.
- [88] J. Kreiss and S. Lahiri. Bootstrap methods for time series. *Time Series Analysis: Methods and Applications*, 30:3–23, 2012.

- [89] J. Kreiss and E. Paparoditis. Bootstrap methods for dependent data: A review. *Journal of the Korean Statistical Society*, 40:357–378, 2011.
- [90] H. Kunsch. Infinitesimal robustness for autoregressive processes. *The Annals of Statistics*, 12:843–863, 1984.
- [91] V. Kuznetsov and M. Mohri. Generalization bounds for time series prediction with non-stationary processes. In *Proceedings of ALT'14*, 2014.
- [92] A. Kyprianou and S. Gerber. *Risk Theory*. EAA Series Springer International Publishing, 2013.
- [93] S. Lahiri. *Resampling methods for Dependent Data*. Springer-Verlag, 2003.
- [94] S.N. Lahiri. Theoretical comparisons of block bootstrap methods. *The Annals of Statistics*, 27:386–404, 1999.
- [95] K. Latuszyński. *Regeneration and Fixed-Width Analysis of Markov Chain Monte Carlo Algorithms*. PhD Thesis, University of Warsaw, 2008.
- [96] M. R. Leadbetter and H. Rootzen. Extremal theory for stochastic processes. *The Annals of Probability*, 16:431–478, 1988.
- [97] J. Lederer and S. van der Geer. New concentration inequalities for suprema of empirical processes. *Bernoulli*, 20:2020–2038, 2014.
- [98] S. Levental. Uniform limit theorems for Harris recurrent Markov chains. *Probability Theory and Related Fields*, 80:101–118, 1988.
- [99] P. Lezaud. Chernoff and Berry-Esseen inequalities for Markov processes. *ESAIM: Probability and Statistics*, 5:183–201, 2001.
- [100] R. Liu. Bootstrap procedures under some non-i.i.d. models. *The Annals of Statistics*, 16:1696–1708, 2008.
- [101] S. Maiz, El. Badaoui, F. Bonnardot, and C. Serviere. New second order cyclostationary analysis and application to the detection and characterization of a runner’s fatigue. *Signal Processing*, 102:188–200, 2014.
- [102] K.V. Mardia. Multivariate Pareto distributions. *Annals of Mathematical Statistics*, 33:1008–1015, 1962.

- [103] R.D. Martin and V.J. Yohai. Influence functionals for time series. *The Annals of Statistics*, 14:781–818, 1986.
- [104] P. Massart. Some applications of concentration inequalities to statistics. *Annales de la faculté des sciences de Toulouse*, 9:245–303, 2000.
- [105] F. Merlevède and M. Peligrad. Rosenthal-type inequalities for the maximum of partial sums of stationary processes and examples. 41:914–960, 2013.
- [106] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, 1993.
- [107] T. Mikosch. *Non-life Insurance Mathematics*. Springer-Verlag, Berlin, Heidelberg, second edition, 2010.
- [108] S. J. Montgomery-Smith. Comparison of sums of independent identically distributed random vectors. *Probability and Mathematical Statistics*, 14(2):281–285, 1993.
- [109] U. Muller, A. Schick, and W. Wefelmeyer. Improved estimators for constrained Markov chain models. *Statistics and Probability Letters*, 54:427–435, 2001.
- [110] R. Nickl. Donsker-type theorems for nonparametric maximum likelihood estimators. *Probability Theory and Related Fields*, 138:411–449, 2007.
- [111] E. Nummelin. A splitting technique for Harris recurrent chains. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 43:309–318, 1978.
- [112] E. Nummelin. *General irreducible Markov chains and non-negative operators*. Cambridge University Press, 1984.
- [113] M. Pagano. On periodic and multiple autoregressions. *The Annals of Statistics*, 6:1310–1317, 1978.
- [114] D. Paulin. Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electronic Journal of Probability*, 20:1–32, 2015.
- [115] M. Peligrad. *The r -quick version of the strong law for stationary ϕ -mixing sequences. In Almost Everywhere Convergence (Columbus, OH, 1988)*. Academic Press, Boston, MA, 1989.

- [116] J.C. Perea and E. Ursu. Application of periodic autoregressive process to the modeling of the garonne river flows. *Stochastic Environmental Research and Risk Assessment*, 30:1785–1795, 2016.
- [117] R. Perfekt. Extremal behaviour of stationary Markov chains with applications. *The Annals of Applied Probability*, 4:529–548, 1994.
- [118] V.V. Petrov. *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*. Oxford studies in probability. Clarendon Press, 1995.
- [119] D. Pollard. *Convergence of Stochastic Processes*. Springer Series in Statistics, 1984.
- [120] W. Polonik. Minimum volume sets and generalized quantile processes. *Stochastic Processes and their Applications*, 69:1–24, 1997.
- [121] D. Radulović. On the bootstrap and empirical processes for dependent sequences. in: *H. Dehling, T. Mikosch, M. Sørensen (Eds.), Empirical Process Techniques for Dependent Data, Birkhäuser*, pages 345–364, 2002.
- [122] D. Radulović. Renewal type bootstrap for Markov chains. *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research*, 13:147–192, 2004.
- [123] S.I. Resnick. *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag, New York, first edition, 1987.
- [124] H. Rieder. *Robust Asymptotic Statistics*. Springer-Verlag Berlin and Heidelberg, 1994.
- [125] H. Rootzen. Maxima and exceedances of stationary Markov chains. *Advances in Applied Probability*, 20:371–390, 1988.
- [126] C. Scott and R. Nowak. Learning minimum volume sets. *Journal of Machine Learning Research*, 7:665–704, 2006.
- [127] C. Scott and R. Nowak. Minimax-optimal classification with dyadic decision trees. *IEEE Transactions on Information Theory*, 52:1335–1353, 2006.
- [128] B. Sen. *A Study of Bootstrap and Likelihood Based Methods in Non-Standard Problems*. PhD Thesis. University of Michigan, 2008.
- [129] Q. Shao. Maximal inequalities for partial sums of ρ -mixing sequences. *The Annals of Probability*, 23:948–965, 1995.

- [130] Q. Shao and P.P. Ni. Least-squares estimation and ANOVA for periodic autoregressive time series. *Statistics and Probability Letters*, 69:287–297, 2004.
- [131] K. Shimizu. *Boostrapping Stationary ARMA-GARCH Models*. Vieweg+Teubner Research, 2009.
- [132] I. Steinwart and A. Christmann. Fast learning from non-i.i.d. observations. *NIPS*, pages 1768–1776, 2009.
- [133] I. Steinwart, D. Hush, and C. Scovel. Learning from dependent observations. *Journal of Multivariate Analysis*, 100(1):175–194, 2009.
- [134] R.A. Stine. Estimating properties of autoregressive forecasts. *Journal of the American Statistical Association*, 82:1072–1078, 1987.
- [135] G.B. Thomas and D.F. Fiering. *Mathematical synthesis of stream flow sequences for the analysis of river basins by simulation*. In Design of Water Resources Systems. Harvard University Press, Cambridge, 1962.
- [136] H. Thorisson. *Coupling, Stationarity and Regeneration*. Springer, 2000.
- [137] T.H. Tsai. *The Uniform CLT and LIL for Markov Chains*. Ph.D. Thesis. University of Wisconsin, 1998.
- [138] E. Ursu and P. Duchesne. On modelling and diagnostic checking of vector periodic autoregressive time series models. *Journal of Time Series Analysis*, 30:70–96, 2009.
- [139] S.A. Utev. Sums of random variables with ϕ -mixing. *Siberian Advances in Mathematics*, 1:124–155, 1991.
- [140] A. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.
- [141] A.W. van der Vaart and J.A. Wellner. *Weak Convergence and Empirical Processes*. Springer, 1996.
- [142] V. Vapnik. *Statistical Learning Theory*. Wiley, New York, 1998.
- [143] V. N. Vapnik. *The Nature of Statistical Learning Theory*. 1999.
- [144] V.N. Vapnik. An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10:988–999, 1999.

- [145] A. Vecchia. Maximum likelihood estimation for periodic autoregressive moving average models. *Technometrics*, 27:375–384, 1985.
- [146] G. Viennet. Inequalities for absolutely regular sequences: Application to density estimation. *Probability Theory and Related Fields*, 107:467–492, 1997.
- [147] T. De Wet and J.W.J. van Wyk. Bootstrap confidence intervals for regression coefficients when the residuals are dependent. *Journal of Statistical Computation and Simulation*, 23:317–327, 1986.
- [148] O. Wintenberger. Exponential inequalities for unbounded functions of geometrically ergodic Markov chains. Applications to quantitative error bounds for regenerative Metropolis algorithms. *A Journal of Theoretical and Applied Statistics*, 51:222–234, 2017.
- [149] B. Zou, H. Zhang, and Z.B. Xu. Learning from uniformly ergodic Markov chain samples. *Journal of Complexity*, 25:188–200, 2009.