Université Paris Nanterre
200 av. de la République
92001 Nanterre Cedex
www.parisnanterre.fr

École doctorale 139 :
Connaissance, langage, modélisation
Modal'X

Membre de l'université Paris Lumières

Carlos Alejandro Fernández Sanz

Estimation problems on null recurrent time series

Thèse présentée et soutenue publiquement le 08/06/2023
en vue de l'obtention du doctorat de Mathématiques appliquées et applications des mathématiques de l'Université Paris Nanterre
sous la direction de M. Patrice Bertail (Université Paris Nanterre)
et de Mme. Cécile Durot (Université Paris Nanterre)

Jury:

Rapporteur :	M. Jens-Peter Kreiss	Professeur, Technische Universität Braunschweig
Rapporteure :	Mme. Sana Louhichi	Professeur, Université de Grenoble
Membre du jury :	M. Randal Douc	Professeur, Telécom SudParis
Membre du jury :	Mme. Elisabeth Gassiat	Professeur, Université Paris-Saclay
Membre du jury :	M. Rik Lopuhaä	Professeur, Delft University of Technology
Invitée :	Mme. Anna Dudek	Professeur, AGH-University of Sciences and Technology

Acknowledgements

A mis padres, que a pesar de la distancia, y de que casi no nos hemos visto en estos 4 años, siempre los tengo presentes. Cualquier logro mío, pasado o futuro, se debe en gran medida a ustedes.

A mi familia, a quienes prácticamente no he podido ver, pero los llevo en el corazón.

To my advisor, M. Patrice Bertail, who, besides being a great mathematician, is the kindest person I know. He has treated me like a son since we met and has helped me in every possible way. I cannot thank you enough for everything you have done for me

To my advisor, Mme. Cécile Durot, for all the help provided during these years, especially during the preparation of the thesis. It has been a great pleasure to work under your guidance.

To M. Stephan Clêmençon, for all the valuable contributions provided during the research of Chapter 3 and for giving me the opportunity to do a post-doc next year.

I also extend my thanks to the reviewers who gave a lot of their time and provided valuable feedback on this thesis. In particular, I am indebted to Professors Jens-Peter Kreiss and Sana Louhichi for agreeing to evaluate this thesis and to be

part of the defense.

A mis 3 grandes amigos (no los nombro, para que no haya celos), que a pesar de no vernos casi nunca, seguimos hablando/discutiendo como el primer día.

A Ubaldo e Hiram, por toda la ayuda y comprensión que me han brindado durante estos años.

A Gabriela y Glenda, por todas las cookies que nos hemos comido juntos (y las que nos faltan).

A mis amigos de la Lenin, que 15 años después seguimos filosofando sobre los mismos temas y aún no nos podemos poner de acuerdo.

A mis compañeros del ICIMAF, que en el 2015 me permitieron volver a hacer Matemática y en especial a Jesús Eladio, por introducirme en el mundo de la estadística.

To my friends at the ATER/Doctorate office, especially Cristina, Ellie, and El Mehdi, for all the enjoyable moments we have spent together while stressing about the bureaucracy and our careers.

Y finalmente, a mi novia, quien ha estado a mi lado durante todo este tiempo y ha soportado todas mis locuras y malcriadeces, espceialmente durante los tiempos más duros del Covid. Sin ella, esta tesis hubiera sido imposible.

Abstract

In the field of Markov chain theory, β-null recurrent Markov chains represent a class of stochastic processes that exhibit challenging and peculiar properties. These nonstationary chains possess infinite invariant measures, making the estimation problems associated with them particularly intricate.

This thesis delves into several estimation problems in the context of β-null recurrent Markov chains, providing new insights and methodologies to tackle these challenges. Our first contribution is the proposal of a tail index estimator for generalized discrete Pareto distributions, which is then used to estimate the parameter β in atomic β-null recurrent Markov chains. The second contribution involves the adaptation and validation of the Regeneration-based bootstrap and Regenerative Block bootstrap methods for these types of chains. Lastly, we develop an estimator for monotone functions in nonlinear cointegrated models, where the underlying process is a Harris recurrent Markov chain (positive or β-null recurrent).

Keywords: Markov chains, Harris recurrence, null recurrence, tail-index estimation, regenerative bootstrap, monotone regression, dependent data, nonstationary data, nonlinear cointegration

Résumé

Parmi chaînes de Markov, les chaînes de Markov β-nulles récurrentes représentent une classe de processus stochastiques aux propriétés complexes et particulières. Ces chaînes non stationnaires possèdent des mesures invariantes infinies, rendant les problèmes d'estimation associés particulièrement délicats.

Cette thèse aborde plusieurs problèmes d'estimation dans le contexte des chaînes de Markov β-nulles récurrentes, en apportant de nouvelles perspectives pour relever ces défis. Notre première contribution est la proposition d'un estimateur d'indice de queue pour les distributions de Pareto discrètes généralisées, qui est ensuite utilisé pour estimer le paramètre β des chaînes de Markov β-nulles récurrentes atomiques. La deuxième contribution concerne l'adaptation et la validation des méthodes de bootstrap basées sur la régénération et le bootstrap par blocs régénératifs pour ces types de chaînes. Enfin, nous développons un estimateur pour les fonctions monotones dans les modèles de cointégration non linéaires, où le processus sous-jacent est une chaîne de Markov récurrente de Harris (récurrente positive ou β-nulle).

Mots-clés : chaînes de Markov, récurrence de Harris, récurrence nulle, estimation de l'indice de queue, bootstrap régénératif, régression monotone, données dépendantes, données non stationnaires, cointégration non linéaire.

Introduction et résumé (en Français)

En 1898, Pavel Nekrasov, alors vice-président de la Société Mathématique de Moscou, écrivit un article intitulé “Propriétés générales de nombreux événements indépendants en relation avec le calcul approximatif des fonctions de très grands nombres”. Il y prétendait que *l’indépendance est une condition nécessaire pour la loi des grands nombres*. Cette affirmation provoqua la colère d’Andrey Markov, qui était convaincu que ce n’était pas le cas.

Pour prouver que Nekrasov a tort, Markov commence d’étudier certains types de variables aléatoires dépendantes afin de relâcher l’hypothèse d’indépendance. Plus précisément, il étudie des variables $X_0, X_1, \ldots, X_n$ dont la dépendance les unes avec les autres diminue rapidement à mesure que leur distance mutuelle dans le temps augmente. Cette construction le conduit à développer un nouveau modèle mathématique, qu’il appele une *“chaîne”* - la désormais célèbre chaîne de Markov.

Selon les propres mots de Markov, les chaînes de Markov[1] est

> *une séquence infinie* $X_0, X_1, \ldots, X_n, X_{n+1}, \ldots$ *de variables liées de telle manière que* X_{n+1} *pour n’importe quel* n *est indépendant de* $X_0, \ldots, X_{n-1}$*, dans le cas où* X_n *est connu*[84]

Il définit une chaîne *homogène* lorsque les distributions conditionnelles de X_{n+1}

[1] Le terme *chaîne de Markov* a été inventé par Bernstein dans son article de 1927 “Sur l’extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes”[13]. Pour une histoire détaillée des premiers développements des chaînes de Markov, voir [11].

étant donné X_n sont indépendantes de n. Il considére également des chaînes plus complexes dans lesquelles *"chaque nombre est directement lié non pas à un seul mais à plusieurs nombres précédents"* [11].

Le première application d'une chaîne de Markov a été faite par Markov lui-même lorsqu'il l'utilise pour analyser la séquence des voyelles et des consonnes dans le poème "Eugene Onegin" d'Alexandre Pouchkine. Depuis, les chaînes de Markov ont rencontré de nombreuses applications en sciences et en ingénénie, telles que la modélisation des phénomènes naturels ou génétiques, la simulation des systèmes complexes, la génération de données aléatoires, l'optimisation des algorithmes et l'analyse des réseaux.

Une chaîne de Markov homogène $\mathbf{X} = \{X_j\}_{j \geqslant 0}$, définie dans un espace probabilisé $(E, \mathcal{E}, \mathbb{P})$ est déterminée de manière unique par une mesure initiale λ et un noyau P, au sens où

$$\mathbb{P}\left(X_n \in A\right) = \lambda P^n\left(A\right) \quad \forall A \in \mathcal{E}, n \geqslant 1.$$

Une chaîne de Markov homogène est *irréductible* s'il existe une mesure σ-finie ϕ sur $(E, \mathcal{E})$ telle que pour tout $x \in E$ et tout $A \in \mathcal{E}$ avec $\phi(A) > 0$ nous avons $P^n(x, A) > 0$ pour un certain $n \geqslant 1$. Dans ce cas, il existe une mesure d'irréductibilité maximale ψ (toutes les autres mesures d'irréductibilité sont absolument continues par rapport à ψ). Dans la suite, toutes les chaînes de Markov sont supposées être irréductibles avec une mesure d'irréductibilité maximale ψ.

Lorsqu'une mesure π vérifie $\pi P = \pi$, on dit qu'elle est invariante pour la chaîne de Markov. Lorsque la mesure invariante est finie (et peut donc être normalisée en une probabilité), la chaîne est dite *récurrente positive*, lorsqu'elle est seulement σ-finie, elle est dite *récurrente nulle*. Si la mesure initiale d'une chaîne

de Markov récurrente positive coïncide avec la probabilité invariante, la chaîne est dite *stationnaire* puisque les marginales de chaîne changent pas.

Les chaînes de Markov irréductibles ont de nombreuses propriétés et peuvent être subdivisées en plusieurs groupes (que nous décrirons en détail au chapitre 2), parmi ceux-ci, on trouve les chaînes de Markov récurrentes[2] et les chaînes de Markov récurrentes de Harris[3]. En termes simples, une chaîne de Markov est récurrente si le nombre attendu de visites à tout état accessible, quel que soit le point de départ, est infini. La récurrence de Harris est une forme renforcée de récurrence où le nombre de fois où la chaîne visite tout état accessible est infini avec probabilité 1.

Une chaîne de Markov est atomique s'il existe un ensemble accessible[4] $\boldsymbol{\alpha}$ tel que $P(x, A) = P(y, A)$ pour tous $x, y \in \boldsymbol{\alpha}$, $A \in \mathcal{E}$. En termes plus simples, un atome est un ensemble sur lequel toutes les probabilités de transition sont identiques. Ainsi, chaque fois que la chaîne de Markov atteint $\boldsymbol{\alpha}$, elle ignore son historique précédent et recommence (elle se régénère). Les chaînes atomiques récurrentes ont de nombreuses propriétés remarquables (voir section 2.2.7), les deux plus importantes étant l'existence d'une mesure invariante (unique à une constante multiplicative près) et la décomposition en blocs[5], qui permet de diviser la chaîne en une série de blocs i.i.d. Le théorème de Kac[6] indique qu'une chaîne atomique est récurrente positive si et seulement si $\mathbb{E}_{\boldsymbol{\alpha}} \tau_{\boldsymbol{\alpha}} < +\infty$, où $\tau_{\boldsymbol{\alpha}}$ désigne le temps de retour à l'atome. Le manque de moments pour les chaînes récurrentes nulles est la plus grande source de problèmes lors de leur manipulation (voir le théorème 4.8 et l'explication qui suit).

[2] Section 2.2.5
[3] Section 2.2.9
[4] Un ensemble A est accessible si $\psi(A) > 0$.
[5] Voir Théorème 2.2.17
[6] Théorème 2.2.16.

Une chaîne de Markov vérifie la condition de minorisation $M(m_0, s, \nu)$ si il existe un entier $m_0 \geqslant 1$, une fonction $0 \leqslant s(x) \leqslant 1$ et une mesure ν telle que $P^{m_0}(x, A) \geqslant s(x)\,\nu(A)$ pour tout $x \in E$ et $A \in \mathcal{E}$. Lorsqu'une chaîne de Markov vérifie cette condition de minorisation, la fonction s et la mesure ν sont appelées *petite fonction* et *petite mesure* respectivement.

Dans un article fondateur, Nummelin [89] a développé une technique d'extension (dite de "spliting") qui permet, pour toute chaîne de Markov qui vérifie la condition de minorisation $M(m_0, s, \nu)$, d'étendre l'espace des probabilités de telle manière que l'extension de $\mathbf{X}$ dans le nouvel espace soit atomique. En utilisant cette extension, il a pu démontrer que chaque chaîne de Markov récurrente de Harris admet une mesure invariante unique (à une constante multiplicative près). Cela implique que chaque chaîne récurrente de Harris est soit récurrente positive, soit récurrente nulle.

La grande majorité des résultats dans la littérature se concentrent sur les chaînes récurrentes positives et traitent de l'estimation du noyau de transition ou de la distribution stationnaire, du test de stationnarité ou de l'ordre de la chaîne de Markov. Cependant, seuls quelques articles (essentiellement de Tjøstheim et ses coauteurs) traitent des problèmes d'estimation et de tests d'hypothèses dans les cas où la chaîne est récurrente nulle [49, 65, 67, 88, 112].

Dans cette thèse, nous nous concentrons sur les chaînes récurrentes nulles, en particulier sur celles qui sont β-nulles récurrentes[7], c'est-à-dire les chaînes de Markov récurrentes de Harris telles qu'il existe une petite fonction h, une mesure initiale λ, une constante $\beta \in (0, 1)$ et une fonction à variation lente L_h telle que

$$\mathbb{E}_\lambda \left[\sum_{t=0}^{n} h(X_t) \right] \sim \frac{1}{\Gamma(1+\beta)} n^\beta L_h(n)$$

[7] Voir la section 2.2.10

lorsque n tend vers $+\infty$.

Lorsque $\mathbf{X}$ est atomique, alors $\mathbf{X}$ est β-nulle récurrente si et seulement s'il existe une constante $\beta \in (0,1)$ et une fonction à variation lente L telle que le temps de retour dans l'atome ait une queue de type Pareto

$$\mathbb{P}\left(\tau_{\boldsymbol{\alpha}} > n\right) \sim \frac{1}{n^{\beta} L\left(n\right)}.$$

Cette caractérisation implique que $\beta = \sup\left\{p \geqslant 0 : \mathbb{E}_{\boldsymbol{\alpha}}\left[\tau_{\boldsymbol{\alpha}}^{p}\right] < \infty\right\}$.

Parmi les exemples les plus connus de chaînes de Markov β-nulles récurrentes, on trouve les marches aléatoires dans $\mathbb{R}$, qui sont $1/2$-nulles récurrentes [64], les marches aléatoires de Bessel [3], [36] et certains types de processus autorégressifs à seuil (TAR) [49] et de processus autorégressifs vectoriels (VAR) [88].

0.1 Résultats principaux

Dans la section suivante, nous décrivons les motivations de nos travaux et les principaux résultats obtenus dans chaque chapitre. Sauf indication contraire, nous supposons que $\mathbf{X}$ est une chaîne de Markov récurrente de Harris de mesure invariante π.

0.1.1 Théorie générale des chaînes de Markov

Dans le Chapitre 2, nous donnons un aperçu de la théorie des chaînes de Markov qui sera utilisée tout au long de la thèse. Nous mettons un accent particulier sur les propriétés des chaînes de Markov récurrentes β-nulles.

Bien que le chapitre soit principalement une récapitulation de la théorie actuelle, nous y considérons également trois nouvelles extensions : ce sont les Théorèmes

2.2.36, 2.2.38 et 2.2.40. Le plus intéressant est le Théorème 2.2.40, qui est une généralisation fonctionnelle du Théorème 2.3 de [28]. Nous le reproduisons ci-dessous.

Theorem 0.1.1 (page 73). *Supposons que* $\boldsymbol{X}$ *soit une chaîne de Markov récurrente* β*-nulle qui satisfait la condition de minorisation* $M(1, s, \nu)$ *et soit* λ *une mesure initiale, si* $f \in L^1(E, \pi)$ *et* $\pi(f) \neq 0$*, alors si on pose* $S_{n,f}(t) = \frac{\sum_{j=0}^{\lfloor nt \rfloor} f(X_j)}{\pi(f) n^{\beta} L_s(n)}$*, on a*

$$S_{n,f} \xrightarrow{\mathscr{D}_{[0,+\infty)}} M_{\beta},$$

où M_{β} *est un processus de Mittag-Leffler de paramètre* β*.*

0.1.2 Estimation de l'indice de queue

Motivé par le besoin d'estimer le paramètre β pour une chaîne de Markov β-récurrente nulle, dans le Chapitre 3, nous abordons le problème d'estimation de l'indice de queue pour une distribution de Pareto discrète généralisée. Plus précisément, nous considérons la distribution d'une variable aléatoire S prenant des valeurs dans $\mathbb{N}^*$, et dont la fonction de survie est la suivante :

$$\mathbb{P}(S > n) = n^{-\beta} L(n) \text{ pour tout } n \geqslant 1, \tag{1}$$

où $L : \mathbb{R}_+ \to \mathbb{R}$ est une fonction à variation lente. L'inférence statistique pour les distributions discrètes à queue lourde n'a pas reçu beaucoup d'attention dans la littérature. La plupart des quelques méthodes dédiées dans la litérature traitent soit de cas très spécifiques comme dans *e.g.* [51], [85] ou [31], soit consistent à appliquer des techniques initialement conçues pour les distributions à queue lourde continues aux données discrètes après une addition préliminaire d'un bruit uniforme indépendant, voir *e.g.* [116]. La grande majorité des estimateurs d'indice de

variation régulière proposés dans la littérature, en particulier les estimateurs de Hill ou Pickand (*cf* [56], [100]), sont basés sur les statistiques d'ordre, ce qui pose des difficultés évidentes dans le cas discret en raison de la possible occurrence de nombreuses répétitions dans l'échantillon.

En revanche, l'estimateur que nous étudions dans le Chapitre 3 est basé sur l'analyse de la probabilité des événements de queue séparés exponentiellement. Il repose sur le fait que $\ln(p_k) - \ln(p_{k+1}) = \beta + \ln(L(e^k)/L(e^{k+1}))$, où $p_l = \mathbb{P}(S > e^l)$ pour tout $l \in \mathbb{N}$, et que $L(e^{k+1})/L(e^k)$ est censé être très proche de 1 pour k suffisamment grand. Une technique d'inférence naturelle (plug-in) peut alors être conçue en remplaçant les probabilités de queue p_l par leurs versions empiriques $\widehat{p}_l^{(n)} = (1/n)\sum_{i=1}^n \mathbb{I}\{S_i > e^l\}$ pour $l \in \mathbb{N}$.

$$\widehat{\beta}_n(k) = \ln\left(\widehat{p}_k^{(n)}\right) - \ln\left(\widehat{p}_{k+1}^{(n)}\right). \tag{2}$$

Nous démontrons que pour un choix approprié de l'hyperparamètre $k = k_n$ (typiquement choisi de l'ordre de $\ln n$) l'estimateur (2) est fortement consistant (Théorème 3.2.2) et asymptotiquement normal (Théorème 3.2.3) lorsque $n \to +\infty$. Des bornes de confiance supérieures non asymptotiques pour les écarts absolus entre $\widehat{\beta}_n(k)$ et β sont également établies (Proposition 3.2.1).

Pour une chaîne de Markov β-récurrente nulle atomique, soit $T(n)+1$ le nombre de fois où la chaîne visite l'atome jusqu'au temps n et soit $\tau_1, \ldots, \tau_{T(n)+1}$ les instants de ces visites. Dans le Théorème 3.3.2, nous montrons que l'estimateur $\widehat{\beta}$ reste fortement consistant lorsque nous considérons $S_i = \tau_{i+1} - \tau_i$ et où $T(n)$

joue le rôle de n. Pour $k_n = \ln n$, cet estimateur prend la forme

$$\widehat{\beta}_{T(n)}\left(\ln T\left(n\right)\right) = \ln\left(\frac{\sum_{i=1}^{T(n)} \mathbb{I}\left\{S_i > T\left(n\right)\right\}}{\sum_{i=1}^{T(n)} \mathbb{I}\left\{S_i > eT\left(n\right)\right\}}\right). \tag{3}$$

0.1.3 Bootstrap régénératif

Depuis son introduction par Efron dans [43] pour des données i.i.d., les méthodes de bootstrap ont connu d'importants développements, donnant naissance à divers schémas de bootstrap adaptés aux contextes i.i.d. et dépendants [71, 77]. Cela a conduit à leur utilisation extensive dans une multitude d'applications statistiques. Dans de nombreux cas, les techniques de bootstrap fournissent des approximations plus précises des distributions statistiques, des probabilités de couverture des intervalles de confiance et des probabilités de rejet des tests d'hypothèses par rapport à la théorie des distributions asymptotiques de premier ordre (pour une discussion détaillée, voir [57]).

Dans le cas markovien, de nombreuses approches ont été proposées et développées. Une idée originale est d'estimer la distribution marginale et la fonction de probabilité de transition en utilisant des techniques d'estimation de fonctions non paramétriques, puis de rééchantillonner à partir de ces estimations. Cette idée a été développée et étendue, entre autres, dans [6, 57, 75, 94, 97, 105]. Pour une explication détaillée de cette idée, voir la section 4 de [71].

Dans [6], une approche distincte de ce problème a été proposée. Au lieu d'utiliser des probabilités de transition estimées, les auteurs ont exploité les propriétés de régénération d'une chaîne de Markov lorsqu'un atome accessible est visité infiniment souvent. L'idée fondamentale de cette méthode est de diviser la chaîne en un

nombre aléatoire de blocs de régénération i.i.d. puis de rééchantillonner un nombre équivalent de blocs de régénération. Cette technique, connue sous le nom de *Bootstrap basé sur la régénération*, a été démontrée comme étant valide pour les chaînes atomiques à états finis dans [6] et a été étendue aux chaînes de Markov atomiques récurrentes positives générales dans [34].

En s'appuyant sur le concept d'exploitation des propriétés régénératives des chaînes de Markov, le *Bootstrap par Blocs Régénératifs* (RBB) a été introduit dans [15]. Cette méthode simule la structure de renouvellement de la chaîne en échantillonnant des blocs de données de régénération jusqu'à ce que la longueur de la série de bootstrap réassemblée dépasse la longueur de la série de données originale n (notez le contraste avec le *Bootstrap basé sur la régénération*, où le nombre de blocs échantillonnés est égal au nombre de blocs de régénération dans la chaîne d'origine).

Il a été démontré dans [15] que pour les chaînes de Markov atomiques récurrentes positives, le RBB pour l'estimation de l'intégrale d'une fonction par rapport à la probabilité invariante présente une vitesse de convergence uniforme de la distribution de l'ordre $O_{\mathbb{P}}\left(n^{-1}\right)$, qui est la même que celle dans le cas i.i.d.

Dans le Chapitre 4, nous adaptons à la fois le *Bootstrap basé sur la régénération* et le *Bootstrap par Blocs Régénératifs* pour les chaînes de Markov β-récurrentes nulles et montrons que les deux sont asymptotiquement valides (Théorèmes 4.3.1 et 4.4.1) pour l'estimation des intégrales par rapport à la mesure invariante lorsque la chaîne β-récurrente nulle possède un atome accessible.

Sous-produit important de nos recherches, dans le Lemme 4.2.1, nous présentons une généralisation du Théorème Central Limite pour des séquences indexées aléatoirement où nous remplaçons l'exigence usuelle du contrôle en probabilité de la séquence d'indexation par l'existence d'une limite de la séquence vers un processus stochas-

tique. Ce résultat a été essentiel dans la preuve du Théorème 4.3.1 qui valide l'utilisation des techniques de regénération après standadisation aléatoire adéquate.

0.1.4 Modèles cointégrés non linéaires monotones

Dans le Chapitre 5, nous étudions les modèles *cointégrés non linéaires* tels que

$$Z_t = f_0(X_t) + W_t, \tag{4}$$

où f_0 est une fonction non linéaire, X_t est une chaîne de Markov récurrente de Harris et W_t est un processus inobservé avec $E(W_t|X_t) = 0$.

Le problème d'estimation de f_0 sous l'hypothèse markovienne sur X_t a été étudié en utilisant l'estimateur de Nadaraya-Watson dans [23, 65], des estimateurs linéaires de type M dans [24, 80] et en utilisant des concepts avancés tels que le temps local et les transformations non linéaires de processus de type mouvement brownien dans [117, 118, 119]. Un résumé complet sur les dernières avancées sur ce problème se trouve dans [112].

À notre connaissance, l'estimation de f_0, lorsqu'elle est soumise à des contraintes de forme, n'a pas été étudiée dans un cadre markovien. Dans un cadre i.i.d. ces estimateurs sont fortement non linéaires et présentent des défis théoriques considérables. Ils s'adaptent mal au cadre dépendant car ils font intervenir une notion d'ordre. Dans le contexte d'observations indépendantes, des contraintes telles que la convexité, la concavité et la log-concavité sont connues pour être encore plus complexes que les contraintes de monotonie (voir [54, 109] et les références correspondantes). Dans le Chapitre 5, nous nous concentrons sur le cas monotone mais il seraient intéressant de considérer ultérieurement des extensions.

La construction de notre estimateur est la suivante:

Soit C un ensemble dont l'intérieur contient notre point d'intérêt x_0. Ayant observé $\{(X_t, Z_t)\}_{t=0}^n$ nous notons par $T_n(C)$ le nombre de fois que $\mathbf{X}$ a visité C jusqu'au temps n et par $\sigma_C(i)$ le temps de la i-ième visite. Notre estimateur $\widehat{f}_n$ est alors l'estimateur non paramétrique défini comme le minimiseur de

$$f \mapsto \sum_{i=1}^{T_n(C)} \left(Z_{\sigma_C(i)} - f\left(X_{\sigma_C(i)}\right)\right)^2 \tag{5}$$

sur l'ensemble des fonctions décroissantes. Cet estimateur peut être calculé à l'aide de simples algorithmes comme discuté dans [10]. De plus, contrairement aux estimateurs à noyau avec paramètre de lissage, C ne dépend pas de n et la vitesse de convergence de l'estimateur ne dépend pas de C.

Dans le Théorème 5.3.1, nous montrons que sous des hypothèses très générales, $\widehat{f}_n(x_0)$ est un estimateur fortement consistant de $f(x_0)$, et avec des hypothèses légèrement plus restrictives, nous montrons dans le Théorème 5.4.1 que la vitesse de convergence de $\widehat{f}_n$ est $u(n)^{-1/3}$, où $u(n) = n$ si $\mathbf{X}$ est récurrente positive et $u(n) = n^{\beta} L(n)$ si elle est récurrente nulle. Remarquons que dans le cas récurrent positif, nous obtenons le même taux, $n^{-1/3}$, que dans le cas i.i.d. [53, Chapitre 2].

L'utilisation d'un estimateur localisé est due au fait que nous devons contrôler le comportement de la chaîne autour de x_0, et, pour ce faire, nous devons estimer la "distribution" asymptotique de $\mathbf{X}$ dans un voisinage de x_0. Pour les chaînes de Markov récurrentes de Harris, le comportement à long terme de la chaîne est donné par sa mesure invariante. Dans le cas récurrent positif, la mesure invariante est finie et peut être estimée en considérant simplement la fonction de répartition empirique des X_t, cependant, dans le cas récurrent nul, la mesure invariante est seulement σ-finie, d'où la nécessité de localiser notre analyse dans un ensemble suffisamment grand pour que la chaîne le visite infiniment souvent, mais suffisam-

ment petit pour que la restriction de la mesure invariante à cet ensemble soit finie. À cet égard, deux résultats de type Glivenko-Cantelli (Lemmes 5.5.1 et 5.5.3) ont été obtenus pour les chaînes de Markov récurrentes de Harris localisées. Nous obtenons également un résultat (Lemme 5.5.2) qui permet de contrôler le nombre de recouvrements d'une classe de fonctions définies sur des blocs localisés à partir du nombre de recouvrements de la classe de départ, ce qui constitue en soit un résultat nouveau et intéressant. .

0.2 Organisation de la thèse

La thèse est organisé comme suit :

- Chapitre 2 : Nous donnons une vue d'emsemble de la théorie des chaînes de Markov, en mettant un accent particulier sur les propriétés et les particularités des chaînes de Markov β-récurrentes nulles.

- Chapitre 3 : Il est basé sur l'article [17]. Il se concentre sur l'estimation de l'indice de queue d'une distribution de Pareto discrète généralisée. Nous montrons que l"estimateur proposé est fortement consistant et asymptotiquement normal dans le cas i.i.d. Dans le cas des chaînes atomiques β-récurrentes nulles, notre estimateur de β sur lequel il est basé est démontré être fortement consistant.

- Chapitre 4 : Il est basé sur l'article [47]. Il traite du bootstrap régénératif sur les chaînes de Markov β-récurrentes nulles. Le *bootstrap basé sur la régénération* et le *bootstrap de blocs régénératifs* sont démontrés être valides pour estimer l'intégrale par rapport à la mesure invariante dans le cas atomique.

- Chapitre 5 : Il est basé sur [41]. Il aborde le problème de l'estimation d'une fonction monotone dans un modèle *cointégré non linéaire*, où X_t est une chaîne de Markov récurrente de Harris. L'estimateur est démontré être fortement consistant et nous obtenons sa vitesse de convergence dans les cas récurrent positif et nul.

Après lecture du Chapitre 2, les chapitres suivants peuvent être lus dans n'importe quel ordre, car sans dépendances de notation entre eux. Nous nous excusons par avance pour les légères redondances entre les chapitres.

Contents

List of Tables

List of Figures

Notations and symbols

$\mathbb{N}$	The natural numbers
$\overline{\mathbb{N}}$	The extended natural numbers: $\mathbb{N} \cup \{\infty\}$
$\mathbb{Z}$	The integer numbers
$\mathbb{Z}_+$	The non negative integer numbers
$\overline{\mathbb{Z}}_+$	The extended non negative integer numbers: $\mathbb{Z}_+ \cup \{\infty\}$
$\mathbb{R}$	The real numbers
$\overline{\mathbb{R}}$	The extended real numbers: $\{-\infty\} \cup \mathbb{R} \cup \{+\infty\}$
$\mathbb{R}_+$	The non-negative real numbers
$\overline{\mathbb{R}}_+$	The extended non-negative real numbers: $\mathbb{R}_+ \cup \{+\infty\}$
$\mathcal{P}(E)$	Collection of all subsets of E
$\sigma(\mathcal{A})$	σ-algebra generated by the collection of sets $\mathcal{A}$.
$\mathcal{B}(X)$	σ-algebra generated by the borelian sets of the metric space X.
$\mathscr{M}(\mathcal{E})_+$	Collection of all nonnegative measures on $(E, \mathcal{E})$
A^c	Complement of the set A
δ_x	Dirac's measure with mass at x
$L^1(E, \pi)$	Space of all measurable functions f, defined on E, such that $\int_E \lvert f \rvert \, d\pi < \infty$
$\mathscr{D}[a, b]$	Space of all real-valued cadlag functions defined on $[a, b]$ where $a < b \leqslant +\infty$.

Chapter 1

Introduction

In 1898, Pavel Nekrasov, then vice president of the Moscow Mathematical Society, wrote a paper named "General properties of numerous independent events in connection with approximate calculation of functions of very large numbers". There, he claimed that *independence is a necessary condition for the law of large numbers*. This claim infuriated Andrey Markov, who was convinced that it was not the case.

To prove Nekrasov wrong, Markov started studying certain types of dependent random variables in order to relax the independence assumption. Specifically, he studied variables $X_0, X_1, \ldots, X_n$ whose dependence on one another quickly lessens as their mutual distance increases. This observation led him to develop a novel mathematical model, which he called a *"chain"* - the now well-known Markov chain.

In Markov's own words, a Markov chain[1]

> *an infinite sequence $X_0, X_1, \ldots, X_n, X_{n+1}, \ldots$ of variables connected in such a way that X_{n+1} for any n is independent of $X_0, \ldots, X_{n-1}$, in case*

[1] The term *Markov chain* was coined by Bernstein in his 1927 paper "Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes"[13]. For a detailed history of the early developments of Markov chains see [11].

X_n *is known*[84]

He called a chain *homogeneous* if the conditional distributions of X_{n+1} given X_n were independent of n. He also considered complex chains in which *"every number is directly connected not with a single but with several preceding numbers"* [11].

The initial application of a Markov chain was done by Markov himself when he used it to analyze the sequence of vowels and consonants in the poem "Eugene Onegin" by Alexander Pushkin. After that, Markov chains have been encountered many applications in science and engineering, such as modeling natural phenomena, simulating complex systems, generating random data, optimizing algorithms and analyzing networks.

Every homogeneous Markov chain $\mathbf{X} = \{X_j\}_{j \geqslant 0}$, defined in a probability space $(E, \mathcal{E}, \mathbb{P})$ is uniquely determined by an initial measure λ and a kernel P, in the sense that

$$\mathbb{P}\left(X_n \in A\right) = \lambda P^n\left(A\right) \quad \forall A \in \mathcal{E}, n \geqslant 1.$$

An homogeneous Markov chain is *irreducible* if there exists a σ-finite measure ϕ on $(E, \mathcal{E})$ such that for all $x \in E$ and all $A \in \mathcal{E}$ with $\phi(A) > 0$ we have $P^n(x, A) > 0$ for some $n \geqslant 1$. In this case, there exists a maximal irreducibility measure ψ (all other irreducibility measures are absolutely continuous with respect to ψ). In the following, all Markov chains are supposed to be irreducible with maximal irreducibility measure ψ.

When a measure π satisfies $\pi P = \pi$, we say that it is invariant for the Markov chain. When the invariant measure is finite (and hence can be normalized into a probability), the chain is called *positive recurrent*, when is only σ-finite is called *null-recurrent*. If the initial measure of a positive recurrent Markov chain coincides

with the invariant probability, the chain is called *stationary*.

Irreducible Markov chains have many properties and can be subdivided into many groups (which we will describe in detail in Chapter 2), among those, we find the recurrent[2] Markov chains and Harris recurrent Markov chains[3]. Loosely speaking a Markov chain is recurrent if the expected number of visits to any accessible state, no matter the starting point, is infinite. Harris recurrence if a strengthened form of recurrence where the number of times the chain visits any accessible state is infinite with probability 1.

A Markov chain is atomic if there exists an accessible set[4] $\boldsymbol{\alpha}$ such that $P(x, A) = P(y, A)$ for all $x, y \in \boldsymbol{\alpha}$, $A \in \mathcal{E}$. In simpler terms, an atom is a set where all the transition probabilities are identical. Hence, whenever the Markov chain reaches $\boldsymbol{\alpha}$, it disregards its previous history and starts anew (regenerates). Recurrent atomic chains have many remarkable properties (see Section 2.2.7), the two most important being, the existence of an invariant measure (unique up to some multiplicative constant) and the block decomposition[5], which allows splitting the chain into a series of i.i.d. blocks. Kac's theorem[6] indicates that an atomic chain is positive recurrent if and only $\mathbb{E}_{\boldsymbol{\alpha}}\tau_{\boldsymbol{\alpha}} < +\infty$, where $\tau_{\boldsymbol{\alpha}}$ denotes the time of first return to the atom. This lack of moments for null recurrent chains is the biggest source of problems when working with these chains (see Theorem 4.8 and the explanation afterward).

A Markov chain satisfies the minorization condition $M(m_0, s, \nu)$ if there exists an integer $m_0 \geqslant 1$, a function $0 \leqslant s(x) \leqslant 1$ and a measure ν such that $P^{m_0}(x, A) \geqslant s(x)\nu(A)$ for all $x \in E$ and $A \in \mathcal{E}$. When a Markov chain satis-

[2] Section 2.2.5
[3] Section 2.2.9
[4] A set A is accessible if $\psi(A) > 0$.
[5] See Theorem 2.2.17
[6] Theorem 2.2.16.

fies this minorization condition, the function s and the measure ν are called *small function* and *small measure* respectively.

In a seminal paper, Nummelin [89] developed a splitting technique that allows, for any Markov chain that satisfies the minorization condition $M(m_0, s, \nu)$, to extend the probability space in such a way that the extension of $\mathbf{X}$ into the new space is atomic. Using this extension, he was able to show that every Harris recurrent Markov chain admits a unique (up to some multiplicative constant) invariant measure. This implies that every Harris recurrent chain is either positive recurrent or null recurrent.

The vast majority of results in the literature focus on positive recurrent chains, and deal with the estimation of the transition kernel or the stationary distribution or testing stationarity or the order of the Markov chain. However, only a few articles (essentially by Tjøsheim and his coauthors) deal with estimation and hypothesis testing issues in cases where the chain is null recurrent[49, 65, 67, 88, 112].

In this thesis, we put our focus on null recurrent chains, specifically in β-null recurrent ones[7], that is, Harris recurrent Markov chains such that there exists a small function h, an initial measure λ, a constant $\beta \in (0,1)$ and a slowly varying function L_h such that

$$\mathbb{E}_\lambda \left[\sum_{t=0}^{n} h(X_t) \right] \sim \frac{1}{\Gamma(1+\beta)} n^\beta L_h(n)$$

as n goes to $+\infty$.

When $\mathbf{X}$ is atomic, then $\mathbf{X}$ is β-null recurrent if and only if there is a constant

[7] See section 2.2.10

$\beta \in (0,1)$ and a slowly varying function L such that

$$\mathbb{P}(\tau_{\boldsymbol{\alpha}} > n) \sim \frac{1}{n^{\beta} L(n)}.$$

This characterization implies that $\beta = \sup\{p \geqslant 0 : \mathbb{E}_{\boldsymbol{\alpha}}[\tau_{\boldsymbol{\alpha}}^{p}] < \infty\}$.

Some of the most well-known examples of β-null recurrent Markov chain are the random walks in $\mathbb{R}$, which are $1/2$-null recurrent [64], the Bessel random walks [3], [36] and some types of threshold autoregressive (TAR) [49] and vector autoregressive processes (VAR) [88].

1.1 Main results

In this section, we describe the motivation and main results of each chapter. Unless stated otherwise, we assume that $\mathbf{X}$ is a Harris recurrent Markov chain with invariant measure π.

1.1.1 General Markov chain theory

In Chapter 2 we provide an overview of the Markov chain theory that will be used throughout the thesis. We make special emphasis on the properties of β-null recurrent Markov chains.

Although the chapter is mostly a recapitulation of the current theory, we have added three new contributions: those are Theorems 2.2.36, 2.2.38 and 2.2.40. The most interesting being Theorem 2.2.40 which is a functional generalization of Theorem 2.3 in [28], we reproduce it below.

Theorem 1.1.1 (page 73). *Assume $\boldsymbol{X}$ is a β-null recurrent Markov chain that satisfies the minorization condition $M(1, s, \nu)$ and let λ be any initial measure, if $f \in$*

$L^1(E,\pi)$ *and* $\pi(f) \neq 0$, *then*

$$S_{n,f} \xrightarrow{\mathscr{D}_{[0,+\infty)}} M_\beta,$$

where $S_{n,f}(t) = \frac{\sum_{j=0}^{\lfloor nt \rfloor} f(X_j)}{\pi(f) n^\beta L_s(n)}$, *and* M_β *is a Mittag-Leffler process with parameter* β.

1.1.2 Tail index estimation

Motivated by the need to estimate the parameter β for a β-null recurrent Markov chain, in Chapter 3, we address the problem of estimating the tail index for a generalized discrete Pareto distribution. Specifically, we consider the distribution of a random variable S taking values in $\mathbb{N}^*$, and whose survival function is as follows:

$$\mathbb{P}(S > n) = n^{-\beta} L(n) \text{ for all } n \geqslant 1, \tag{1.1}$$

where $L : \mathbb{R}_+ \to \mathbb{R}$ is a slowly varying function.

Statistical inference for discrete heavy-tailed distributions has not received much attention in the literature. Most of the very few dedicated methods documented either deal with very specific cases as in *e.g.* [51], [85] or [31] or else consists in applying techniques originally designed for continuous heavy-tailed distributions to the discrete data after a preliminary addition of an independent uniform noise, see *e.g.* [116]. The vast majority of the regular variation index estimators proposed in the literature, Hill's or Pickand's estimators in particular (*cf* [56], [100]), are based on order statistics, which causes obvious difficulties in the discrete case because of the possible occurrence of many ties.

In contrast, the estimator that we study in Chapter 3 is based on the analysis of the probability of exponentially separated tail events. It relies on the fact that

$\ln(p_k) - \ln(p_{k+1}) = \beta + \ln(L(e^k)/L(e^{k+1}))$, where $p_l = \mathbb{P}(S > e^l)$ for all $l \in \mathbb{N}$, and that $L(e^{k+1})/L(e^k)$ is expected to be very close to 1 for k sufficiently large. A natural (plug-in) inference technique can be then devised by replacing the tail probabilities p_l with their empirical versions $\widehat{p}_l^{(n)} = (1/n)\sum_{i=1}^n \mathbb{I}\{S_i > e^l\}$ for $l \in \mathbb{N}$. This yields the estimator

$$\widehat{\beta}_n(k) = \ln\left(\widehat{p}_k^{(n)}\right) - \ln\left(\widehat{p}_{k+1}^{(n)}\right). \tag{1.2}$$

We prove that for an appropriate choice of the hyperparameter $k = k_n$ (typically chosen of order $\ln(n)$), the estimator (1.2) is strongly consistent (Theorem 3.2.2) and asymptotically normal (Theorem 3.2.3) as $n \to +\infty$. Nonasymptotic upper confidence bounds for the absolute deviations between $\widehat{\beta}_n(k)$ and β are also established (Proposition 3.2.1).

For an atomic β-null recurrent Markov chain, let $T(n) + 1$ be the number of times the chain visits the atom and denote by $\tau_1, \ldots, \tau_{T(n)+1}$ the times of those visits. In Theorem 3.3.2, we show that the estimator $\widehat{\beta}$ remains strongly consistent when we consider $S_i = \tau_{i+1} - \tau_i$ and take n as $T(n)$. For $k_n = \ln n$, this estimator takes the form

$$\widehat{\beta}_{T(n)}(\ln T(n)) = \ln\left(\frac{\sum_{i=1}^{T(n)} \mathbb{I}\{S_i > T(n)\}}{\sum_{i=1}^{T(n)} \mathbb{I}\{S_i > eT(n)\}}\right). \tag{1.3}$$

1.1.3 Regenerative bootstrap

Since its inception by Efron in [43] for i.i.d. data, bootstrap methods have advanced, giving rise to various bootstrap schemes tailored for both i.i.d. and dependent contexts [71, 77]. This has led to their extensive use in a multitude of statistical

applications. In many cases, bootstrap techniques deliver more accurate approximations of statistical distributions, confidence interval coverage probabilities, and hypothesis test rejection probabilities when compared to first-order asymptotic distribution theory (for a detailed discussion, see [57]).

In the Markovian case, numerous approaches have been developed and examined. The original idea was to estimate the marginal distribution and the transition probability function using nonparametric function estimation techniques and then resample from those estimates. This idea was developed and expanded, among others, in [6, 57, 75, 94, 97, 105]. For a detailed explanation of this idea, refer to Section 4 in [71].

In [6], a distinct approach to this problem was proposed. Instead of employing estimated transition probabilities, the authors made use of the regeneration properties of a Markov chain when an accessible atom is visited infinitely often. The fundamental idea of this method is to split the chain into a random number of i.i.d. regeneration blocks and then resample an equivalent number of regeneration blocks. This technique, known as the *Regeneration based bootstrap*, was demonstrated to be valid for finite state atomic chains in [6] and was further extended to general atomic positive recurrent Markov chains in [34].

Building on the concept of exploiting the regenerative properties of Markov chains, the *Regenerative Block bootstrap* (RBB) was introduced in [15]. This method simulates the renewal structure of the chain by sampling regeneration data blocks until the length of the reassembled bootstrap series surpasses the original data series length n (note the contrast with the *Regeneration based bootstrap*, where the number of sampled blocks is equal to the number of regeneration blocks in the original chain). It was demonstrated in [15] that for atomic positive recurrent Markov chains, the RBB for estimating the integral of a function with respect to

the invariant probability exhibits a uniform rate of convergence of order $O_{\mathbb{P}}\left(n^{-1}\right)$, which is the same as in the i.i.d. case.

In Chapter 4, we adapt both *Regeneration based-bootstrap* and *Regenerative Block bootstrap* for β-null recurrent Markov chains and show that both are valid (Theorems 4.3.1 and 4.4.1) for the estimation of integrals with respect to the invariant measure when the β-null recurrent chain possesses an accessible atom.

As a byproduct of our research, in Lemma 4.2.1 we present a generalization of the Central Limit Theorem for randomly indexed sequences where we replace the requirement of the control in probability of the indexing sequence by the existence of the limit of a stochastic process defined in terms of the said sequence. This result was instrumental in the proof of Theorem 4.3.1.

1.1.4 Nonlinear monotone cointegrated models

In Chapter 5, we study *nonlinear cointegrated* models such that

$$Z_t = f_0(X_t) + W_t, \tag{1.4}$$

where f_0 is a nonlinear function, X_t is a Harris recurrent Markov chain and W_t is an unobserved process with $E(W_t|X_t) = 0$.

The problem of estimating f_0 under the Markovian assumption on X_t has been studied using Nadaraya-Watson estimator in [23, 65], linear M-type estimators in [24, 80] and using advanced concepts like local time and nonlinear transformations of Brownian motion-like processes in [117, 118, 119]. A comprehensive survey of the latest advances in this problem can be found in [112].

To our knowledge, the estimation of f_0, when it is subject to shape constraints has not been explored under Markovian assumptions. These estimators are non-

linear and present considerable theoretical challenges. In the context of independent observations, constraints such as convexity, concavity, and log-concavity are known to be even more intricate than monotonicity constraints (refer to [54, 109] and the citations therein). Consequently, in Chapter 5 we focus on the monotone case.

The construction of our estimator is as follows:

Let C be a set whose interior contains our point of interest x_0. Having observed $\{(X_t, Z_t)\}_{t=0}^n$, we denote by $T_n(C)$ the number of times that $\mathbf{X}$ visited C up to time n and by $\sigma_C(i)$ the time of the i-th visit. Our estimator $\widehat{f}_n$ is then the nonparametric LSE defined as the minimizer of

$$f \mapsto \sum_{i=1}^{T_n(C)} \left(Z_{\sigma_C(i)} - f\left(X_{\sigma_C(i)}\right)\right)^2 \tag{1.5}$$

over the set of non-increasing functions. This estimator can be computed using simple algorithms as discussed in [10]. Moreover, contrary to the bandwidth in kernel type estimators, C does not depend on n, and the rate of convergence of the estimator does not depend on C.

In Theorem 5.3.1 we show that under very general assumptions, $\widehat{f}_n(x_0)$ is a strongly consistent estimator $f(x_0)$, and with slightly more restrictive hypotheses, we show in Theorem 5.4.1 that the rate of convergence of $\widehat{f}_n$ is $u(n)^{-1/3}$, where $u(n) = n$ if $\mathbf{X}$ is positive recurrent and $u(n) = n^\beta L(n)$ if is null-recurrent. Notice that in the positive recurrent, we obtain the same rate, $n^{-1/3}$, as in the i.i.d. case [53, Chapter 2].

The use of a localized estimator is due to the fact that we need to control the behavior of the chain around x_0, and, to do this, we need to estimate the asymptotic "distribution" of $\mathbf{X}$ in a vicinity of x_0. For Harris recurrent Markov chains,

the long-term behavior of the chain is given by its invariant measure. In the positive recurrent case, the invariant measure is finite and it can be estimated by simply considering the empirical distribution function of the X_t, however, in the null recurrent case, the invariant measure is only σ-finite, hence, we need to localize our analysis in a set big enough that the chain visits it infinitely often, but small enough that the restriction of the invariant measure to it is finite. In this regard, two Glivenko-Cantelli type results (Lemmas 5.5.1 and 5.5.3) were obtained for localized Harris recurrent Markov chains, as well as a result (Lemma 5.5.2) that allows controlling the covering number of a class of functions defined over the localized blocks.

1.2 Outline

The rest of the thesis is organized as follows:

- Chapter 2: We provide a recapitulation of Markov chain theory, making special emphasis on the properties and peculiarities of β-null recurrent Markov chains.

- Chapter 3: Is based on [17]. It focuses on the estimation of the tail index of a generalized discrete Pareto distribution. The proposed estimator is shown to be strongly consistent and asymptotically normal in the i.i.d. case. In the case of atomic β-null recurrent chains, it is shown to be strongly consistent.

- Chapter 4: Is based on [47]. Deals with the regenerative bootstrap on β-null recurrent Markov chains. The *Regeneration-based bootstrap* and the *Regenerative Block bootstrap* are shown to be valid for estimating the integral with respect to the invariant measure in the atomic case.

- Chapter 5: Is based on [41]. It tackles the problem of estimating a monotone function in a *nonlinear cointegrated* model, where X_t is a Harris recurrent Markov chain. The estimator is shown to be consistent and its rate of convergence is obtained.

Upon completing Chapter 2, the remaining chapters can be read in any order, as they function independently, with no notation dependencies between them. We apologize in advance for the slight redundancies across chapters.

Chapter 2

Markov chains

In this chapter, we introduce the basic concepts that will be used throughout the thesis. We will make special emphasis on the properties of Markov chains, especially in null-recurrent ones.

Most of the definitions and results of this chapter are classic and can be found in [5, 38, 87, 90, 108]. However, in subsection 2.2.10 we have added a few new contributions in theorems 2.2.36, 2.2.38 and 2.2.40.

2.1 Kernels

Through this chapter $(E, \mathcal{E})$ denotes a measurable space where the σ-algebra $\mathcal{E}$ is countably generated[1][2]. The points of E are called *states* and $(E, \mathcal{E})$ is called the *state space*. With a slight abuse of notation, by $\mathcal{E}$ we will also denote the set of measurable functions from $(E, \mathcal{E})$ to $\left(\overline{\mathbb{R}}, \mathcal{B}\left(\overline{\mathbb{R}}\right)\right)$, and we will use $\mathcal{E}_+$ for the

[1] A σ-algebra $\mathcal{E}$ is countably generated if there exists a countable collection of subsets $\{A_n\}_{n=1}^{+\infty} \subseteq \mathcal{E}$ such that $\mathcal{E} = \sigma\left(\{A_n\}_{n=1}^{+\infty}\right)$.

[2] This assumption is used in almost all literature about Markov chain theory because it removes the possibility of extremely pathological examples known as "anormal" chains [37]. For a detailed overview of the matter, see [37, 61] and pp. 91 in [107]. An example of an "anormal" chain is provided in [22]. Furthermore, this assumption does not result in a significant loss of generality, given that most of the time, $\mathcal{E} = \mathcal{B}\left(\mathbb{R}^d\right)$, which is countably generated.

collection of measurable functions from $(E, \mathcal{E})$ to $\left(\overline{\mathbb{R}}_+, \mathcal{B}\left(\overline{\mathbb{R}}_+\right)\right)$.

Definition 2.1.1 (Kernel on $(E, \mathcal{E})$)**.** *Function $K : (E, \mathcal{E}) \to [0, +\infty]$ such that*

i) For every x, the mapping $A \to K(x, A)$ is a measure on $\mathcal{E}$.

ii) For every $A \in \mathcal{E}$, the mapping $x \to K(x, A)$ is a measurable function from $(E, \mathcal{E})$ to $\left(\overline{\mathbb{R}}_+, \mathcal{B}\left(\overline{\mathbb{R}}_+\right)\right)$.

A kernel is said to be σ-finite if there exists a $\mathcal{E} \otimes \mathcal{E}$-measurable function f that is almost everywhere positive and such that $\int_E f(x, y) K(x, dy) < +\infty \quad \forall x \in E$; it is *bounded* if $\sup_{x \in E} K(x, E) < +\infty$; *substochastic* if $K(x, E) \leqslant 1$ for all x in E and *stochastic* (or *markovian*) if $K(x, E) = 1$ for all x in E.

Example 2.1.1 (Integral kernel)**.** *If λ is a σ-finite measure on $(E, \mathcal{E})$ and k is a non-negative $\mathcal{E} \otimes \mathcal{E}$ measurable function, then, the function*

$$K(x, A) = \int_A k(x, y)\, d\lambda(y)$$

is a non-negative kernel. This kernel is often called an integral kernel with basis k and density λ.

The following example allows us to interpret any σ-finite measure λ on $(E, \mathcal{E})$ as a kernel over the same state space.

Example 2.1.2 (Measures seen as kernels)**.** *Taking $k(x, y) \equiv 1$ in Example 2.1.1 we get that*

$$K(x, A) = \lambda(A) \quad \forall x \in E, A \in \mathcal{E}$$

is a kernel on $(E, \mathcal{E})$. This kernel is stochastic if and only if λ is a probability measure.

Example 2.1.3 (Tensor product). *If f is a non-negative $\mathcal{E}$-measurable function, and λ is a σ-finite measure on $(E, \mathcal{E})$, we can define their tensor product $f \otimes \lambda$ as $f \otimes \lambda(x, A) = f(x)\lambda(A)$. Taking $k(x, y) = f(x)$ in 2.1.1 shows that $f \otimes \lambda$ is a kernel on $(E, \mathcal{E})$.*

Another interesting application of Example 2.1.1 is when E is a countable set, and $\mathcal{E} = \mathcal{P}(E)$ is the σ-algebra of all the subsets of E.

Example 2.1.4 (Kernels on countable sets). *If we take λ as the counting measure on E we obtain that every kernel K defined on $(E, \mathcal{P}(E))$ satisfies,*

$$K(x, A) = \sum_{y \in A} K(x, \{y\}).$$

This shows that every kernel K on $(E, \mathcal{P}(E))$ can be identified with the matrix $k(x, y) = K(x, \{y\})\ \forall x, y \in E$.

2.1.1 Operations with Kernels

Given a kernel K on $(E, \mathcal{E})$ and a function $f \in \mathcal{E}_+$ we can define the function $Kf : E \to \overline{\mathbb{R}}_+$ as follows:

$$Kf(x) = \int_E f(y) K(x, dy). \tag{2.1}$$

Notice that if K is stochastic, then $\|Kf\|_\infty \leqslant \|f\|_\infty$.

The following result shows that a Kf is an additive operator on the space of non-negative integrable functions over $\mathcal{E}$.

Theorem 2.1.1. [3] *Let K be a kernel on $(E, \mathcal{E})$ and define $\mathcal{K} : \mathcal{E}_+ \to \mathcal{E}_+$ as $\mathcal{K}(f) = Kf$ where Kf is as in (2.1). Then*

[3] Proposition 1.2.5 in [38].

i) The function $\mathcal{K}$ is an additive and positive homogeneous operator over $\mathcal{E}_+$.

ii) If $\{f_n\}_{n\in\mathbb{N}} \subset \mathcal{E}_+$ is an increasing sequence of functions, then $\lim_n \uparrow \mathcal{K}(f_n) = \mathcal{K}\left(\lim_n \uparrow f_n\right)$.

The following result establishes a converse

Theorem 2.1.2. [4] *Let $\mathcal{N} : \mathcal{E}_+ \to \mathcal{E}_+$ be an additive and positively homogeneous operator such that $\lim_n \uparrow \mathcal{N}(f_n) = \mathcal{N}\left(\lim_n \uparrow f_n\right)$ for every increasing sequence $\{f_n\}_{n\in N} \subset \mathcal{E}_+$. Then,*

i) The function $N(x, A) = \mathcal{N}(I_A)(x), x \in E, A \in \mathcal{E}_+$ is a kernel on $(E, \mathcal{E})$.

ii) $Nf = \mathcal{N}(f)$ for all $f \in \mathcal{E}_+$.

A kernel can also define an operator over the set of non-negative measures $\mathbb{M}(\mathcal{X})_+$ as it is shown in the following result.

Theorem 2.1.3. [5] *For every non-negative measure $\lambda \in \mathscr{M}(\mathcal{E})_+$ and every kernel K on $(E, \mathcal{E})$, the function $\lambda K : \mathcal{E} \to \overline{\mathbb{R}}_+$ defined as*

$$\lambda K(A) = \int_E K(x, A)\, d\lambda(x),$$

is a measure on $(E, \mathcal{E})$.

If K_1 and K_2 are kernels on $(E, \mathcal{E})$ we define their sum and the multiplications by positive scalars in the typical way. We define their product (or convolution) K_1K_2 as follows

$$K_1K_2(x, A) = \int_E K_2(y, A)\, K_1(x, dy) \quad \forall x \in E, A \in \mathcal{E}. \tag{2.2}$$

[4] Proposition 1.2.6 in [38]
[5] Proposition 1.2.7 in [38].

The following result shows that K_1K_2 is indeed a kernel on $(E, \mathcal{E})$

Theorem 2.1.4. [6] *[Product of kernels] If K_1 and K_2 are kernels on $(E, \mathcal{E})$, then, the function K_1K_2 defined in (2.2) is a kernel on $(E, \mathcal{E})$. Furthermore, this product is associative and for every function $f \in \mathcal{E}_+$ we have $K_1K_2(f) = K_1(K_2f)$.*

Example 2.1.5 (Identity kernel)**.** *The kernel I, defined as*

$$I(x, A) = \mathbb{I}_A(x) = \begin{cases} 0, & x \notin A \\ 1, & x \in A \end{cases}$$

satisfies that, for every kernel K on $(E, \mathcal{E})$, $f \in \mathcal{E}_+$ and $\lambda \in \mathcal{M}(\mathcal{E})$

$$\begin{aligned} IK(x, A) &= \int_E K(y, A)\, d\delta_x(y) = K(x, A) \quad \forall x \in E, A \in \mathcal{E}, \\ KI(x, A) &= \int_E I(y, A)\, K(x, dy) = K(x, A) \quad \forall x \in E, A \in \mathcal{E}, \\ If(x) &= \int_E f(y)\, d\delta_x(y) = f(x) \quad \forall x \in E, \\ \lambda I(A) &= \int_E I_A(x)\, d\lambda(x) = \lambda(A) \quad \forall A \in \mathcal{E}. \end{aligned}$$

Therefore, $IK = KI = K$, $If = f$ and $\lambda I = \lambda$. This explains why I is called the identity kernel.

The iterates K^n of a kernel K are defined by setting $K^0 = I$, and iteratively, $K^n = KK^{n-1}$. Henceforth, we'll assume that all iterates of K are σ-finite. Notice that if K is substochastic, all the iterates are substochastic.

An immediate consequence of the associativity of the product of kernels is the celebrated Chapman-Kolmogorov equation.

[6] Proposition 1.2.8 in [38]

Theorem 2.1.5 (Chapman-Kolmogorov equation)**.** *Let K be a kernel on $(E, \mathcal{E})$. For any m with $0 \leqslant m \leqslant n$, we have*

$$K^{n+m}(x, A) = \int_E K^n(x, dy)\, K^m(y, A) \tag{2.3}$$

Definition 2.1.2 (Potential kernel)**.** *If K is a kernel on $(E, \mathcal{E})$, the potential kernel of K is defined as $G = \sum_{n=0}^{+\infty} K^n$. The n-partial sum of the potential is denoted by $G^{(n)} = \sum_{j=0}^{n} K^j$.*

It can happen that G is not σ-finite, since is possible that G only admits the values 0 and $+\infty$. For example, this happens if we take $K = I$.

The following result resumes the main properties of the potential kernel.

Theorem 2.1.6. *Let K be a kernel, then, for any $n \geqslant 1$*

$$G = \sum_{j=0}^{n-1} K^j + K^n G = \sum_{j=0}^{n-1} K^j + G K^n,$$

and for any $f \in \mathcal{E}_+$

$$\lim_n \downarrow K^n G f(x) = 0, \quad \forall x \in \{y \in E : Gf(y) < \infty\}.$$

Proof. This is Proposition 2.1 in [90]. □

Closely related to the potential kernel, in the markovian scenario, are the P_ε-resolvent kernels.

Definition 2.1.3. *[P_ε-resolvent of K] Given a markovian kernel P, for every $\varepsilon \in (0, 1)$, the P_ε-resolvent kernel is defined as*

$$P_\varepsilon(x, A) = (1 - \varepsilon) \sum_{i=0}^{\infty} \varepsilon^i P^i(x, A), \quad x \in E, A \in \mathcal{E}.$$

The existence of a kernel K on $(E, \mathcal{E})$ allows us to establish a communication relation in $E \times \mathcal{E}$ as follows:

$$x \to A \Leftrightarrow \exists n \geqslant 1 : K^n(x, A) > 0. \tag{2.4}$$

When $x \to A$, we say that A is *accessible* from the point x. If B is a set and $x \to A$ for every $x \in B$, then we say that A is *accessible* from B. When $B = E$, we say that is *accessible*. In the following two sections, we study the structure induced by this relation.

2.1.2 Closed sets for a Kernel

A closed set will be defined as a set whose complement is not accessible from any point in A.

Definition 2.1.4 (Closed set)**.** *Non-empty set $A \in \mathcal{E}$ such that $K(x, A^c) = 0$ for all $x \in A$.*

We say that a set A is *absorbing* if $K(x, A) = K(x, E) = 1$ for all $x \in E$.

Lemma 2.1.1. *If K is a kernel defined in $(E, \mathcal{E})$, then*

i) A set A is closed for K if and only if $x \not\to A^c$ for all $x \in E$.

ii) An absorbing set is always closed.

Definition 2.1.5 (Indecomposable set)**.** *Set $A \in \mathcal{E}$ such that are not two disjoint closed sets $B_1, B_2 \subseteq A$.*

When $B \in \mathcal{E}$ is closed, we denote by $K|_B$ the restriction of K to $(B \cap E, \mathcal{E} \cap B)$. Not surprisingly, $K|_B$ is itself a kernel which is called the *restriction of K to the*

closed set B. The following Lemma contains the main results concerning the restriction of kernels.

Lemma 2.1.2. *Let* K *be a kernel in* $(E, \mathcal{E})$ *and* $B \in \mathcal{E}$ *a closed set for* K*, then*

i) $K|_B$ *is a kernel in* $(B \cap E, \mathcal{E} \cap B)$.

ii) $(K|_B)^n = (K^n|_B)$ *for all* n.

iii) If B *is absorbing, the kernel* $K|_B$ *is markovian.*

iv) $K|_{B^c}$ *is a kernel in* $(B^c \cap E, \mathcal{E} \cap B^c)$ *and* $(K|_{B^c})^n = (K^n|_{B^c})$ *for all* n.

2.1.3 Irreducibility

Irreducibility is the idea that all big enough parts of the space can be reached, no matter the starting point. In this section, we formalize this concept.

Let φ be a σ-finite measure on $(E, \mathcal{E})$. We say that a set $A \in \mathcal{E}$ is *φ-positive* if $\varphi(A) > 0$. Lastly, for any set B, define $B^+ = B \cup \{x \in E : x \to B\} = \{G\mathbb{I}_B > 0\}$.

Definition 2.1.6 (*φ-communicating* set for K)**.** *Set* B *such that every* φ*-positive subset* $A \subseteq B$ *is accessible from* B.

A kernel is *φ-irreducible* if the whole space is φ-communicating. In this case, the measure φ is called an *irreducible measure* for K. Notice that, if K is φ-irreducible, then $A^+ = E$ for all $A \in \mathcal{E}$ such that $\varphi(A) > 0$.

Lemma 2.1.3. [7] *Let* K *be a kernel and* B *a measurable set*

i) The set B^+ *is either closed of empty.*

ii) Every φ*-communicating set is indecomposable.*

[7] Proposition 2.3 and pp. 12-13 in [90].

iii) If B is φ-communicating then $K|_{B^+}$ is $\varphi\mathbb{I}_B$-irreducible.

iv) Any measure ψ which is absolutely continuous with respect to an irreducibility measure is itself an irreducibility measure.

The φ-irreducibility assumption eliminates several forms of reducible behavior. The definition ensures that the chain will reach "big" sets (as defined by φ) with positive probability from any starting point, preventing the chain from dividing into separate parts. However, to achieve certain objectives, it's necessary to know that "negligible" sets (B, where $\varphi(B) = 0$) are avoided with certainty from most starting points.

Definition 2.1.7 (Maximal irreducibility measure for K)**.** *Irreducibility measure ψ such that all other irreducibility measures of K are absolutely continuous with respect to ψ.*

Notice that, by definition, if a maximal irreducibility measure exists for a kernel K, it is unique up to the equivalence of measures. The following result shows that, for every φ-irreducible kernel, there exists a maximal irreducible measure ψ and that accessible sets are precisely the ψ-positive sets.

Theorem 2.1.7. [8] *Suppose that K is φ irreducible*

i) There exists a maximal irreducibility measure for K.

ii) If ψ is a maximal irreducibility measure and $\psi(B) = 0$, then $\psi(B^+) = 0$.

When K is irreducible with maximal irreducibility measure ψ, we will denote by $\mathcal{E}^+$ the subset of all non-negative measurable functions in $\mathcal{E}_+$ that are ψ-positives, i.e. $\mathcal{E}^+ = \{f \in \mathcal{E}_+ : \psi(f) > 0\}$ where $\psi(f) = \int_E f(x)\, d\psi(x)$.

[8] Proposition 2.4 in [90].

2.1.4 Small functions and small sets

Let K be a kernel with maximal irreducibility measure ψ. We say that K satisfies the *minorization condition* $M(m_0, \beta, s, \nu)$ if there exists $m_0 \in \mathbb{N}$, $\beta > 0$, $s \in \mathcal{E}^+$ and a positive non-trivial measure ν such that $K^{m_0}(x, A) \geqslant \beta s(x)\, \nu(A)$ for all $x \in E$ and $A \in \mathcal{E}$, or, using the tensor product defined in Example 2.1.3,

$$K^{m_0} \geqslant \beta s \otimes \nu. \tag{2.5}$$

Example 2.1.6. [9] *Let $f : \mathbb{R} \to \mathbb{R}$ be an integrable function with respect to the Lebesgue measure in $\mathbb{R}$, such that f is bounded on every compact set. Let $g : \mathbb{R} \to \mathbb{R}_+$ be a density function such that $0 < \inf_{x \in C} g(x)$ on every compact set C. Consider the markovian kernel P defined as,*

$$P(x, A) = \int_A g\big(y - f(x)\big)dy \quad x \in \mathbb{R}, A \in \mathcal{B}(\mathbb{R})$$

Fix a compact set C, define the function $\rho(y) = \inf_{x \in C} g\big(y - f(x)\big)$, then

$$P(x, A) \geqslant \mathbb{I}_C(x) \int_A \rho(y)\, dy, \quad \forall x \in C, A \in \mathcal{B}(\mathbb{R}).$$

Hence, the kernel P satisfies the minorization condition $M(1, 1, \mathbb{I}_C, \mu)$ where μ is the measure defined by $\mu(A) = \int_A \rho(y)\, dy$.

A function $s \in \mathcal{E}^+$ is called a *small* (for the kernel K), if there exist m_0, β and ν such that the kernel K satisfies $M(m_0, \beta, s, \nu)$. We will denote by $\mathscr{S}^+$ be the class of all small functions for K. A closely related concept is the following:

[9] Example 3.1 in [67]

Definition 2.1.8 (Small set). *Positive set $C \in \mathcal{E}$ such that its indicator function, $\mathbb{I}_C$, is small.*

A small function remains small if multiplied by a constant, hence, there is no loss of generality in assuming $\beta = 1$. Moreover, if K is a markovian kernel that satisfies $M(m_0, \beta, s, \nu)$, then ν is a finite measure and $\frac{1}{\beta\nu(E)} \geqslant s(x) \geqslant 0$, therefore, K satisfies the condition $M\left(m_0, 1, \beta\nu(E)\, s, \frac{\nu}{\nu(E)}\right)$. This means that, in the markovian case, the minorization condition (2.5) is equivalent to the existence of $m_0 \in \mathbb{N}$ such that

$$K^{m_0} \geqslant s \otimes \nu, \tag{2.6}$$

where $0 \leqslant s(x) \leqslant 1$ and ν is a probability measure. In these cases, we will say that the markovian kernel P satisfies the minorization condition $M(m_0, s, \nu)$ and if $m_0 = 1$, we will call the pair (s, ν) an *atom* for the kernel K. This condition will be crucial in section 2.2.8.

Remark 2.1.1. *In Example 2.1.6, we have that (s, ν) is an atom for the kernel P, where $s(x) = K\mathbb{I}_C$ and $\nu = \frac{\mu}{K}$ with $K = \int_E \rho(y)\, dy$.*

Notice that if a markovian kernel P satisfies the minorization condition $M(m_0, s, \nu)$, then, for all $\varepsilon \in (0,1)$, $x \in E$ and $A \in \mathcal{E}$ we have

$$P_\varepsilon(x, A) \geqslant (1-\varepsilon)\,\varepsilon^{m_0} P^{m_0}(x, A) \geqslant (1-\varepsilon)\,\varepsilon^{m_0} s(x)\,\nu(A),$$

from where the next result follows immediately.

Theorem 2.1.8. *Suppose a markovian kernel P satisfies the minorization condition $M(m_0, s, \nu)$, then for any $\varepsilon \in (0,1)$, the pair $\left((1-\varepsilon)\,\varepsilon^{m_0} s, \nu\right)$ is an atom for the P_ε-resolvent.*

From its definition, it is not evident that small functions exist. However, the following theorem shows not only that they exist for any irreducible kernel, but also that there are plenty of them.

Theorem 2.1.9. [10] *[Existence of small functions and sets] If K is an irreducible kernel with maximal irreducibility measure ψ, then*

i) $\mathscr{S}^{+} \neq \varnothing$.

ii) For every $A \in \mathcal{E}$ such that $\psi(A) > 0$, there exists $C \subseteq A$, such that C is small.

Corollary 2.1.1. *If P is an irreducible markovian kernel, then, every P_{ε}-resolvent satisfies the minorization condition $M(1, s_{\varepsilon}, \nu)$, where ν does not depend on ε.*

2.1.5 Invariant measures

By Theorem 2.1.3, a kernel K defines an operator in $\mathscr{M}(\mathcal{E})_{+}$. The fixed points for this operator, if they exist, are called *invariant measures*.

Definition 2.1.9 (Invariant measure for kernel K)**.** *σ-finite measure $\pi \in \mathscr{M}(\mathcal{E})_{+}$ such that $\pi K = \pi$.*

The following result shows that the sets of invariant measures for a markovian kernel coincide with the set of invariant measures of any of its ε-resolvents.

Theorem 2.1.10. [11] *Let P be a markovian kernel. For every $\varepsilon \in (0,1)$ a measure π is invariant for P if and only if it is invariant for the resolvent P_{ε}.*

[10] Part i is Theorem 2.1 in [90], part ii is Proposition 2.6 of the same book.
[11] Theorem 10.4.3 in [87]

2.2 Markov chains

This section is devoted to giving a general overview of the theory of Markov chains that will be used in this thesis. It is organized as follows: we will start with the concept of stochastic process, which will allow us to define a homogeneous Markov chain in subsection 2.2.2 and we will see some examples in subsection 2.2.3. In subsections 2.2.4 through 2.2.6, we will study the concepts of stopping times, recurrence, transience and aperiodicity of Markov chains and in 2.2.7 we will see the main properties of atomic chains. Subsection 2.2.8 is dedicated to the construction of the split chain, while in subsection 2.2.9 we will study the Harris recurrent Markov chains. Finally, in subsection 2.2.10 we introduce the β-null recurrent Markov chains.

2.2.1 Stochastic process

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $(E, \mathcal{E})$ a measurable space and T a set. A collection of E-valued random variables indexed by a totally ordered set T is called a E-valued *stochastic process*.

If X is an E-valued random variable, we'll denote by $\mathcal{L}(X)$ its probability distribution (or its *law*), defined as the probability measure induced by X in $(E, \mathcal{E})$ i.e. $\mathcal{L}(X)(A) = \mathbb{P}(X \in A)$ for all $A \in \mathcal{E}$.

A collection of σ-algebras $\mathcal{F}_t$ of $\mathcal{F}$ such that $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s \leqslant t$ is called a *filtration*. A *filtered probability space*, denoted by $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in T}, \mathbb{P})$, is a probability space equipped with a filtration.

Definition 2.2.1 (Stochastic process adapted to the filtration $\mathcal{F}$)**.** *Stochastic process $\{X_t\}_{t\in T}$ such that X_t is $\mathcal{F}_t$-measurable for each $t \in T$. The notation $\{X_t, \mathcal{F}_t\}_{t\in T}$ indicates that the stochastic process $\{X_t\}_{t\in T}$ is adapted to the filtration $\{\mathcal{F}_t\}_{t\in T}$.*

The σ-field $\mathcal{F}_t$ can be interpreted as the information available at time t. When a process is adapted, it means that the probability of events related to X_t can be computed using solely the information available at time t.

Every stochastic process $\{X_t\}_{t\in T}$ is trivially adapted to the filtration $\left\{\mathcal{F}_t^X\right\}_{t\in T}$, where $\mathcal{F}_t^X = \sigma\left(X_s, s \leqslant t\right)$. This filtration is named the *internal history*.

Remark 2.2.1. *When T is countable, an adapted stochastic process $\{X_t, \mathcal{F}_t\}_{t\in T}$ can be viewed as an element of $\Psi_E = \prod_{t\in T} E$, measurable with respect to $\mathcal{G} = \sigma\left(\bigcup_{t\in T} \mathcal{F}_t\right)$.*

2.2.2 Homogeneous Markov chains

Loosely speaking, a Markov Chain is a stochastic process that it is forgetful of all but its most immediate past. The formal definition is as follows:

Definition 2.2.2 (Markov Chain)**.** *Let $\left(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n\in\mathbb{N}}, \mathbb{P}\right)$ be a filtered probability space. An adapted Stochastic process $\{X_n, \mathcal{F}_n\}_{n\in\mathbb{N}}$ is a Markov Chain if for all $n \in \mathbb{N}$ and all $A \in \mathcal{E}$*

$$\mathbb{P}\left(X_{n+1} \in A | \mathcal{F}_n\right) = \mathbb{P}\left(X_{n+1} \in A | X_n\right) \quad \mathbb{P} - a.s. \tag{2.7}$$

The distribution $\mathcal{L}\left(X_0\right)$ of X_0 is called the initial distribution of the chain.

A Markov Chain is said to be *homogeneous* if the transition probabilities in (2.7) can be expressed with a Markovian kernel.

Definition 2.2.3 (Homogeneous Markov Chain)**.** *Let $\left(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n\in\mathbb{N}}, \mathbb{P}\right)$ be a filtered probability space. A Markov Chain $\{X_n, \mathcal{F}_n\}_{n\in\mathbb{N}}$ is called homogeneous with kernel P and initial measure μ if there exists a Markovian kernel P and a probability*

measure μ, both defined on $(E, \mathcal{E})$ such that for all $n \in \mathbb{N}$ and all $A \in \mathcal{E}$

$$\mathbb{P}\left(X_{n+1} \in A | \mathcal{F}_n\right) = P\left(X_n, A\right) \quad \mathbb{P} - a.s. \tag{2.8}$$

$$\mathcal{L}\left(X_0\right) = \mu. \tag{2.9}$$

Remark 2.2.2. *If $\{X_n, \mathcal{F}_n\}_{n\in\mathbb{N}}$ is a homogeneous Markov Chain, then, $\left\{X_n, \mathcal{F}_n^X\right\}_{n\in\mathbb{N}}$ is a homogeneous Markov Chain as well. From now on, we will always consider homogeneous Markov Chains adapted to its internal history, and we will write $\boldsymbol{X}$ instead of $\left\{X_n, \mathcal{F}_n^X\right\}_{n\in\mathbb{N}}$ to ease the notation.*

If the kernel of a homogeneous Markov chain is φ-irreducible we will say that the chain $\mathbf{X}$ is φ-irreducible and that φ is an irreducibility measure for $\mathbf{X}$. When we do not need to specify the irreducibility measure, we will just say that $\mathbf{X}$ is *irreducible*. Similarly, we will say that a set $A \in \mathcal{E}$ is accessible from B if A is accessible from B given the communication relation induced by the kernel (see (2.4)), and we will say that A is *accessible* if is accessible from E. When $\mathbf{X}$ is irreducible with maximal irreducibility measure ψ, accessible sets are precisely the sets $A \in \mathcal{E}$ such that $\psi\left(A\right) > 0$.

From the definition of a homogeneous Markov Chain, two questions arise: is a homogeneous Markov Chain uniquely determined by its initial distribution and its Kernel? and, given an initial probability measure and a Kernel, does there exist a homogeneous Markov Chain such that (2.8) and (2.9) holds? The answer to both questions is Yes and it is resumed in the following theorem, which combines Theorem 3.4.1 in [87] and Theorem 1.3.4 in [38].

Theorem 2.2.1. *Given a Markovian kernel P and a probability measure μ, both defined on $(E, \mathcal{E})$, a E-valued stochastic process $\{X_n\}$ is a homogeneous Markov*

Chain with kernel P and initial measure μ if and only if,

$$\mathbb{P}\left(X_0 \in A_0, X_1 \in A_1, \ldots, X_n \in A_n\right) = \int_{A_0} \mu\left(dx_0\right) \int_{A_1} P\left(x_0, dx_1\right) \ldots \int_{A_n} P\left(x_{n-1}, dx_n\right). \tag{2.10}$$

for all $n \geqslant 0$, $A_0, \ldots, A_n \in \mathcal{E}$. Moreover, for every μ and P there exists a probability space containing a homogeneous Markov Chain with kernel P and initial measure μ.

If we take $A_i = E, i = 1, \ldots, n-1$ in (2.10), we obtain,

$$\begin{aligned}\mathbb{P}\left(X_n \in A\right) &= \int_E \mu\left(dx_0\right) \int_E P\left(x_0, dx_1\right) \ldots \int_E P\left(x_{n-2}, dx_{n-1}\right) \int_A P\left(x_{n-1}, dx_n\right) \\ &= \int_E \mu\left(dx_0\right) P^n\left(x_0, A\right) \\ &= \mu P^n(A),\end{aligned} \tag{2.11}$$

for any $A \in \mathcal{E}$, therefore, $\mathcal{L}\left(X_n\right) = \mu P^n$. Similarly, taking $\mu = \delta_x$ in (2.10) for a fixed $x \in E$, we get that

$$\mathbb{P}\left(X_n \in A | X_0 = x\right) = P^n\left(x, A\right) \quad \forall n \geqslant 1, A \in \mathcal{E}, x \in E. \tag{2.12}$$

Conditioning on $X_0 = x$, we will write $\mathbb{P}_x(X_n \in A)$ instead of $P\left(X_n \in A | X_0 = x\right)$, and for a measurable function f, we will write $\mathbb{E}_x f\left(X_n\right)$ instead of $\mathbb{E}\left[f\left(X_n\right) | X_0 = x\right]$.

The equality (2.10), known as the *Markov property*[12], has the following equivalent in terms of expectations.

Theorem 2.2.2. [13] *[Markov property for expectations] If $\boldsymbol{X}$ is a homogeneous Markov Chain with initial measure μ and $f : \Psi_E \to \mathbb{R}$ is a bounded and measurable function,*

[12] See pp 62 in [87].
[13] This is Proposition 3.4.3 in [87].

then

$$\mathbb{E}\left[f\left(X_{n+1}, X_{n+2}, \ldots\right) | X_0, X_1, \ldots, X_n; X_n = x\right] = \mathbb{E}_x\left[f\left(X_1, X_2, \ldots\right)\right] \quad (2.13)$$

Applying Chapman Kolmogorov equation to (2.12) we get, for any m with $0 \leqslant m \leqslant n$,

$$\begin{aligned}\mathbb{P}_x\left(X_n \in A\right) &= \int_E P^m\left(x, dy\right) P^{n-m}\left(y, A\right) \\ &= \int_E \mathbb{P}_x\left(X_m \in dy\right) \mathbb{P}_y\left(X_{n-m} \in A\right).\end{aligned}$$

This can be understood as follows: as $\mathbf{X}$ transitions from x to A in n steps, at any intermediate step m it must take a value $y \in E$. As a Markov chain, it forgets its past at that time m and continues the remaining $n - m$ steps based on the appropriate law starting at y.

The m-step kernel is a transition kernel by itself, therefore it describes the Markov Chain $\mathbf{X}^m = \{X_{nm}\}$ with transition laws $\mathbb{P}_x\left(X_{nm} \in A\right) = P^{nm}\left(x, A\right)$. This chain $\mathbf{X}^m$ has a special name.

Definition 2.2.4 (m-skeleton of the chain $\mathbf{X}$). *Given a Markov Chain $\boldsymbol{X}$ with initial distribution μ and kernel P, its m-skeleton is the Markov Chain $\boldsymbol{X}^m$ with initial distribution μ and kernel P^m.*

Recall from Definition 2.1.3 that, for every $\varepsilon \in (0, 1)$, the P_ε-resolvent of P is the kernel

$$P_\varepsilon\left(x, A\right) = (1 - \varepsilon) \sum_{i=0}^{\infty} \varepsilon^i P^i\left(x, A\right), \quad x \in E, A \in \mathcal{E}. \quad (2.14)$$

A Markov chain with the same initial distribution as $\mathbf{X}$ and with kernel P_ε is called the *P_ε-chain of $\boldsymbol{X}$*. The following example shows how to extract a subchain

of **X** with kernel P_ε.

Example 2.2.1. *[P_ε-chain] Let **X** be a Markov chain with kernel P and $\varepsilon \in (0,1)$ arbitrary but fixed. Define an i.i.d. sequence $\{\alpha_k\}_{k\geqslant 1}$ of Bernoulli random variables, independent of **X**, such that $\mathbb{P}(\alpha_1 = 1) = 1-\varepsilon$. Lastly, consider the renewal sequence $\{\gamma(k)\}_{k\geqslant 0}$ defined as:*

$$\gamma(0) = 0\,, \quad \gamma(k) = \inf\{m > \gamma(k-1) :\ \alpha_m = 1\} \quad \text{for } k \geqslant 1. \tag{2.15}$$

The sequence $\{\gamma(k) - \gamma(k-1)\}_{k\geqslant 1}$ is i.i.d. with geometric distribution given by $\mathbb{P}(\gamma(1) = k) = (1-\varepsilon)\,\varepsilon^k$. By equation (5.9) in [35] (and pp.19 in [28]) the random sequence $\left\{X_{\gamma(n)}\right\}_{n\geqslant 0}$ is a Markov chain with kernel P_ε.

As stated in Remark 2.2.1, an adapted Markov Chain can be viewed as an element in the space $\Psi_E = \prod_{n\geqslant 0} E$, measurable with respect to $\mathcal{G} = \sigma\left(\bigcup_{n\geqslant 0}\mathcal{F}_n\right)$.

The *shift operator* $\theta : \Psi_E \to \Psi_E$ is defined as

$$\theta(\{x_0, x_1, \ldots, x_n, \ldots\}) = \{x_1, x_2, \ldots, x_n, \ldots\}$$

and its iterations are defined inductively by

$$\theta^1 = \theta, \quad \theta^{k+1} = \theta \circ \theta^k,\ k \geqslant 1.$$

When H is a random variable in $(\Psi_E, G, \mathbb{P})$, θ^k acts over H as

$$\left(\theta^k H\right)(w) = H \circ \theta^k(w)\,,$$

therefore, $X_n \circ \theta^k(w) = X_{n+k}$. Then, if $H = h(X_0, X_1, \ldots)$, where h is a measurable function on Ψ_E, we have $\theta^k H = h(X_k, X_{k+1}, \ldots)$. Because $\mathbb{E}_x H$ is a measur-

able function on E, it follows that $\mathbb{E}_{X_n} H$ is a random variable on $(\Psi_E, G, \mathbb{P})$ for any initial distribution. Then, the Markov property (2.13) can be written succinctly as

$$\mathbb{E}_\mu \left[\theta^n H | \mathcal{F}_n^X\right] = \mathbb{E}_{X_n} [H] \quad \mathbb{P}_\mu \, a.s. \tag{2.16}$$

for any bounded and measurable function h and fixed n.

2.2.3 Examples

In this section, we give some examples of time-homogeneous Markov Chains.

Example 2.2.2 (Countable space Markov Chain). [14] *Suppose that the space E is discrete and $\mathcal{E}$ is the σ-algebra of all the subsets of E. Let $\boldsymbol{X}$ be a homogeneous Markov Chain with initial distribution μ and kernel P.*

The initial probability μ satisfies $\mu(A) = \sum_{x \in A} \mu(\{x\})$ and the kernel P can be identified with the transition matrix M

$$M(x, y) \overset{def}{=} \mathbb{P}(X_{n+1} = y | X_n = x) = P(x, \{y\}) \quad x, y \in E, n \geqslant 0,$$

and satisfying $M^n(x, y) = P^n(x, \{y\})$, where M^n is the usual power of matrices.

From (2.11) and (2.12) we obtain, for all $x, y \in E, n \geqslant 0$

$$\mathbb{P}(X_n = y | X_0 = x) = M^n(x, y),$$
$$\mathbb{P}(X_n = y) = \sum_{x \in E} \mu(x) M^n(x, y).$$

For our next example, consider a scenario where a person plays a series of rounds of a game in a gambling house. On each round, a game is played and an amount is won or lost, with the successive totals of the amounts representing the

[14] Example 1.2 in [90]

fluctuations in the gambler's fortune. It is reasonable to assume, that if the same game is played each time, then the winnings Z_n at each time n are i.i.d. In this context, the total winnings or losings at time n can be represented by X_n where

$$X_{n+1} = X_n + Z_{n+1} \quad n \geqslant 1.$$

This stochastic process is called a *random walk* and is, perhaps, the most well-known instance of a Markov chain. In the following example, we give its formal definition and the form of its kernel.

Example 2.2.3. *[Random walk in $\mathbb{R}$][15] Let $\{Z_n\}_{n\in\mathbb{N}}$ be a sequence of i.i.d. random variables with common distribution Z and X_0 is a random variable, independent of $\{Z_n\}_{n\in\mathbb{N}}$ such that $\mathcal{L}(X_0) = \mu$. The process $\boldsymbol{X} = \{X_n\}_{n\geqslant 0}$ defined by*

$$X_{n+1} = X_n + Z_{n+1} \quad n \geqslant 1,$$

is a Markov chain in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with initial measure $\lambda = \mathcal{L}(X_0)$ and kernel P given by,

$$P(x, A) = \mathbb{P}(Z + x \in A) \quad \forall x \in \mathbb{R}, A \in \mathcal{B}(\mathbb{R}).$$

Example 2.2.4 (Simple symmetric random walk)**.** *Random walks can also be defined over $\mathbb{Z}$. A well-known example in this scenario is the simple symmetric random walk. Defined as in Example 2.2.3, but with*

$$P(Z_1 = 1) = \frac{1}{2} \quad , \quad P(Z_1 = -1) = \frac{1}{2}.$$

This random walk and its variation has been widely studied. See [107] for an excellent

[15] Example 2.1.1, pp. 28 in [38]

compendium of the theory.

In a queuing system, let s_n be the time of service of the n-th customer and take t_n as the arrival epoch of the n-th customer. The waiting time, X_n of the n-th customer before service is then,

$$\begin{aligned} X_1 &= 0 \\ X_{n+1} &= \max\left(X_n + s_n - (t_{n+1} - t_n), 0\right) \quad n > 1. \end{aligned}$$

If we assume that $s_1, \ldots$ are i.i.d. and $t_1, t_2 - t_1, \ldots$ are also i.i.d. and independent of $\{s_n\}_{n\geqslant 0}$, then, the sequence of random variables $\{s_{n-1} - (t_n - t_{n-1})\}_{n\geqslant 1}$ is i.i.d. and the process X_n is a Markov chain as the following example shows.

Example 2.2.5 (Random walk on a half line)**.** [16] *Let X_0 and $\{Z_n\}_{n\in\mathbb{N}}$ be as in Example 2.2.3, define X_n iteratively as follows*

$$X_{n+1} = \max\left(X_n + Z_{n+1}, 0\right) \quad n \geqslant 1.$$

In section 3.5.1 of [87], it is shown that this process is a Markov chain in $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$. Its kernel P is defined as follows: For any $A \in \mathcal{B}(\mathbb{R}_+)$ such that $A \subseteq (0, +\infty)$

$$P(x, A) = \mathbb{P}(Z + x \in A) \quad \forall x \geqslant 0,$$

whilst,

$$P(x, \{0\}) = \mathbb{P}(Z \leqslant -x).$$

*The chain **X** follows the path of a random walk but is restricted to stay at zero when-*

[16] Example RWHL1 in [87]. Also, example 1.2 d) in [90].

ever the underlying random walk drops below zero. It leaves 0 *only when the next positive value in the sequence* $\{Z_n\}$ *is encountered.*

For our last example, suppose we are interested in modeling the exchange rate X_n between two currencies. This can be represented as a function of its past several values $X_{n-1}, ..., X_{n-p}$, modified by the volatility of the market which is incorporated as a disturbance term Z_n (see pp.4 in [87]). The auto-regressive model shown below describes the essential behavior of such a system.

Example 2.2.6 (Auto-regressive process: $AR(p)$)**.** [17] *Let* $\alpha = \alpha_1, \ldots, \alpha_p$ *be real numbers,* $\{Z_n\}_{n\in\mathbb{N}}$ *a sequence of i.i.d. real-valued random variables with finite variance and* $X_0, X_{-1}, \ldots, X_{-p+1}$ *random variables independent of* $\{Z_n\}_{n\in\mathbb{N}}$*. For* $n > 0$*, define* X_n *as*

$$X_n = \alpha_1 X_{n-1} + \alpha_2 X_{n-2} + \cdots + \alpha_p X_{n-p} + Z_n.$$

Assume that the roots of the polynomial $1 - \alpha_1 x - \alpha_2 x^2 - \cdots - \alpha_p x^p$ *are all outside the unit circle. Then, the vector process* $\boldsymbol{X} = \left\{(X_n, X_{n-1}, \ldots, X_{n-p+1})^t\right\}_{n\geqslant 0}$*, known as a causal* $AR(p)$ *process, is a Markov chain in* $(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p))$*. This process can be written in matrix form as*

$$\boldsymbol{X}_n = \alpha \boldsymbol{X}_{n-1} + BZ_n,$$

where

$$\alpha = \begin{pmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_p \\ 1 & 0 & \ldots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & \ldots & 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

[17] Example 2.1.2 in [38]

The kernel of this process is

$$P(\boldsymbol{x}, A) = \mathbb{P}(\alpha \boldsymbol{x} + BZ_0 \in A) \quad \boldsymbol{x} \in \mathbb{R}^p, A \in \mathcal{B}(\mathbb{R}^p).$$

2.2.4 Hitting, stopping times and the strong Markov property

Definition 2.2.5 (Occupation time)**.** *Number of visits by* $\boldsymbol{X}$ *to the set* A*.*

$$T(A) = \sum_{n=0}^{\infty} \mathbb{I}\{X_n \in A\}. \tag{2.17}$$

We will write $T_n(A)$ for number of visits to A up to time n. This sequence is called the *occupation time sequence.*

Definition 2.2.6 (Hitting times and return times)**.** *For* $A \in \mathcal{E}$*, the time of first visit* σ_A *and the time of first return* τ_A *by the Markov Chain* $\boldsymbol{X}$ *to the set* A *are defined by*

$$\sigma_A = \inf\{n \geqslant 0 : X_n \in A\}. \tag{2.18}$$

$$\tau_A = \inf\{n \geqslant 1 : X_n \in A\}. \tag{2.19}$$

where we use the convention that $\inf \varnothing = +\infty$*. The subsequent return times* $\tau_A(k)$*,* $k \geqslant 1$ *are defined inductively as follows*

$$\begin{aligned} \tau_A(1) &= \tau_A \\ \tau_A(k) &= \min\{n > \tau_A(k-1) : X_n \in A\} \end{aligned} \tag{2.20}$$

For any set $A \in \mathcal{E}$, $T(A)$, τ_A and σ_A are measurable functions from Ω to $\overline{\mathbb{Z}}_+$.

Let $U(x,A) = \mathbb{E}_x T(A)$ and denote by G_P the potential kernel of P, by Theorem 2.1.6, we obtain,

$$U(x,A) = \mathbb{E}_x T(A) = \sum_{n=1}^{+\infty} P^n(x,A) = G_P P(x,A).$$

An analysis of the number of visits to a given set often requires consideration of the behavior after the first visit to the set, rather than the behavior after fixed times. Markov chain theory is noteworthy for its "forgetfulness" properties holding for stopping times, which are random interruptions, as well as for fixed times n.

Definition 2.2.7 (Stopping time). *A function $\zeta : \Omega \to \overline{\mathbb{Z}}_+$ is a stopping time for the Markov Chain **X**, if for any initial distribution μ, the event $\{\zeta = n\} \in \mathcal{F}_n^X$ for all $n \geqslant 0$.*

Notice that

$$\{\sigma_A = n\} = \bigcap_{i=0}^{n-1} \{X_y \notin A\} \bigcup \{X_n \in A\},$$

and $\{\tau = n\} = \bigcap_{i=1}^{n-1} \{X_y \notin A\} \bigcup \{X_n \in A\}$, therefore, for any $A \in \mathcal{E}$, both σ_A and τ_A are stopping times for **X**.

The following result expresses the distribution of τ_A and σ_A conditioned to the starting point in terms of the kernel P.

Theorem 2.2.3. *For all $x \in E$, $A \in \mathcal{E}$,*

$$\begin{aligned} \mathbb{P}_x(\tau_A = k) &= \left(P\mathbb{I}_{A^c}\right)^{k-1} P(x,A), \\ \mathbb{P}_x(\sigma_A = 0) &= \mathbb{I}_A(x). \end{aligned} \tag{2.21}$$

Moreover, if $x \in A^c$

$$\mathbb{P}_x(\sigma_A = n) = \mathbb{P}_x(\tau_A = n).$$

From (2.21) we obtain that the probability that the chain ever returns to a A starting from the state x is

$$\mathbb{P}_x\left(\mathbf{X} \text{ ever visits A}\right) = \sum_{k=1}^{\infty} \left(P\mathbb{I}_{A^c}\right)^{k-1} P\left(x, A\right).$$

The Markov Property holds for any bounded, measurable function and any fixed time n, the strong Markov Property allows us to extend it to random stopping times. Before stating this fundamental property, we need some definitions.

Let ζ be a stopping time. Then $X_\zeta^{-1}\left(A\right) = \bigcup_{n=0}^{\infty}\Big\{\left\{X_n^{-1}\left(A\right)\right\} \cap \left\{\zeta = n\right\}\Big\}$ for any $A \in \mathcal{E}$, hence, X_ζ is a random variable.

Define $\mathcal{F}_\zeta^X = \left\{A \in \mathcal{F} : \left\{\zeta = n\right\} \cap A \in \mathcal{F}_n^X \ \forall n \geqslant 0\right\}$, which represents the history of the chain until the stopping time ζ. Finally, if $H = h\left(X_0, X_1, \ldots\right)$ is a random variable, define the shift θ^ζ as $\theta^\zeta H = h\left(X_\zeta, X_{\zeta+1}, \ldots\right)$ on the set $\left\{\zeta < \infty\right\}$. Then,

Theorem 2.2.4 (Strong Markov property)**.** [18] *Let $\boldsymbol{X}$ be a Markov chain. For any initial distribution μ, any real-valued bounded measurable function h on Ψ_E and any stopping time ζ, it holds that*

$$\mathbb{E}_\mu\left[\theta^\zeta H | \mathcal{F}_\zeta^X\right] = \mathbb{E}_{X_\zeta}\left[H\right] \quad \mathbb{P}_\mu \ a.s. \tag{2.22}$$

on the set $\left\{\zeta < \infty\right\}$.

The strong Markov property tells us that the process $\left\{X_{\zeta+k}\right\}$, restricted to $\left\{\zeta < \infty\right\}$, is a Markov chain with the same kernel as the original chain and independent of the chain's history up to τ.

[18] Proposition 3.4.6 in [87].

2.2.5 Recurrence and transience

In this section, we will delve into the fundamental concepts of recurrence and transience in the context of Markov chains, which are very important in order to understand the long-term behavior of a Markov chain.

Definition 2.2.8 (Recurrence). *A set $A \in \mathcal{E}$ is said to be recurrent if $U(x, A) = \infty$ for all $x \in A$. A Markov chain is recurrent if its kernel is irreducible and every accessible set is recurrent.*

Closely related to this is the concept of transience.

Definition 2.2.9 (Uniformly transient set). *Set $A \in \mathcal{E}$ such that $\sup_{x \in A} U(x, A) < \infty$*

Definition 2.2.10 (Transient set). *Set $A \in \mathcal{E}$ such that $A = \bigcup_{n=1}^{\infty} A_n$ where A_n is uniformly transient for all n.*

Definition 2.2.11 (Transient Markov chain). *Irreducible Markov chain $\boldsymbol{X}$ such that E is transient.*

The next result shows that irreducible chains are either recurrent or transient and this property is inherited by its P_ε chains.

Theorem 2.2.5. *If $\boldsymbol{X}$ is an irreducible Markov chain, then $\boldsymbol{X}$ is recurrent or transient. Moreover, $\boldsymbol{X}$ is recurrent (transient) if and only if each P_ε-chain is recurrent (transient).*

In Theorem 2.2.26 we will see that if a chain is irreducible and recurrent it admits an invariant measure. Then, we will subdivide the class of recurrent irreducible Markov chains into two classes: the ones that admit an invariant probability and the ones that do not.

Definition 2.2.12 (Positive Markov chain)**.** *Recurrent and irreducible Markov chain that admits and invariant probability measure π.*

Definition 2.2.13 (Null recurrent Markov chains)**.** *Recurrent and irreducible Markov chain that does not admit an invariant probability measure.*

2.2.6 Aperiodicity

A key concept in Markov chains is aperiodicity, which refers to the property of a state having a finite number of steps before returning to the same state. In other words, aperiodicity describes the pattern of repeating states in a Markov chain. This section is formalize this concept.

Let **X** be a ψ-irreducible Markov chain, by Theorem 2.1.9, there exists a small set $C \in \mathcal{E}$, i.e. there exists $m_0 \in \mathbb{N}$ and a measure $\nu \in \mathscr{M}(\mathcal{E})_+$ such that $\nu(C) > 0$ and

$$P^{m_0}(x, A) \geqslant \nu(A) \quad \forall x \in C,\ A \in \mathcal{E}.$$

Then, if the chain starts in C, there is a non-zero probability that the chain will return to C at time m_0. Define E_C as the set of natural numbers m such that P satisfies the minorization condition $M(m, \mathbb{I}_C, \gamma_m \nu)$ for some $\gamma_m \in \mathbb{R}_+$, i.e.

$$E_C = \{m \in \mathbb{N} : \exists \gamma_m \in \mathbb{R}_+ \text{ such that } P^m(x, A) \geqslant \gamma_m \nu(A) \quad \forall x \in C\ A \in \mathcal{E}\}.$$

The set E_C is closed under addition and the greatest common divisor of the elements of E_C belongs to E_C. The following Theorem shows that this greatest common divisor is a property of the whole chain **X**, and is independent of the particular small set chosen.

Theorem 2.2.6. *Suppose **X** is an irreducible Markov chain with maximal irreducibil-*

ity measure ψ. Let $C \in \mathcal{E}$ be a small set and denote by d the greatest common divisor of the set E_C. Then, there exists disjoint sets $D_1, \ldots, D_{d-1} \in \mathcal{E}$ (called a d-cycle) such that

i) for $x \in D_i$, $P(x, D_{i+1}) = 1, \quad i = 0, \ldots, d-1 \ (\mod d)$.

ii) the set $N = \left(\bigcup_{i=0}^{d-1} D_i\right)^c$ is ψ-null.

Moreover, the d-cycle is maximal in the sense that for any other collection $\left\{D'_0, \ldots, D'_{d'-1}\right\}$ satisfying $i)$ and $ii)$, we have that d' divides d, and if $d = d'$, then, by reordering the indexes if necessary, $D_i = D'_i \quad \psi - a.e.$

From this, we can define the *period* of a ψ-irreducible Markov chain as the size of the largest d-cycle.

Definition 2.2.14 (Aperiodic Markov chain)**.** *Irreducible Markov chain whose period is 1.*

Most of the results that we obtain in this thesis assume that the chain is aperiodic. In practice this is not greatly restrictive, since

Theorem 2.2.7. *Let $\boldsymbol{X}$ be a ψ-irreducible Markov chain with period d and d-cycle $\{D_1, \ldots, D_d\}$. Then, each of the sets D_i is an absorbing ψ-irreducible set for the chain $\boldsymbol{X}^d$ corresponding to the transition kernel P^d. $\boldsymbol{X}^d$ on each D_i is aperiodic.*

Notice that if for a small set C there exist $\beta > 0$ and a measure $\nu \in \mathscr{M}(\mathcal{E})_+$ such that the minorization condition $M(1, \beta \mathbb{I}_C, \nu)$ is satisfied, then the chain is automatically aperiodic. These types of chains are called *strongly aperiodic.*

Theorem 2.2.8. [19] *If $\boldsymbol{X}$ is an irreducible Markov chain, then every P_ε-chain is strongly aperiodic for all $0 < \varepsilon < 1$.*

[19] This is part ii) of Proposition 5.4.5 in [87]

The final result of this section, which is a direct consequence of Theorem 2.1.9, shows that if a chain is aperiodic we can find a m-skeleton that is strongly aperiodic.

Theorem 2.2.9. *Suppose that $\boldsymbol{X}$ is ψ-irreducible, aperiodic chain, then, every skeleton of $\boldsymbol{X}$ is ψ-irreducible and aperiodic and there exists m such that the m-skeleton is strongly aperiodic.*

2.2.7 Atoms

Definition 2.2.15 (Atom for **X**). *Set $\boldsymbol{\alpha} \in \mathcal{E}$ such that*

$$P(x, A) = \nu(A), \quad x \in \boldsymbol{\alpha},$$

for some measure ν on $\mathbb{E}$. If $\boldsymbol{X}$ is φ-irreducible, and $\varphi(\boldsymbol{\alpha}) > 0$ then $\boldsymbol{\alpha}$ is called an accessible atom. A Markov chain with an accessible atom is named an atomic chain.

If $\boldsymbol{\alpha}$ is an atom for **X**, with a slight abuse of notation we will write $P(\boldsymbol{\alpha}, \bullet)$ to represent the measure ν.

A simple consequence of the existence of atoms is the irreducibility of the chain under mild conditions. In effect, by (2.3), for any $x \in E$, $A \in \mathcal{E}$ and $n \geqslant 1$, we have

$$\begin{aligned} P^{n+1}(x, A) &\geqslant \int_{\boldsymbol{\alpha}} P^n(x, dy) P(y, A) = \int_{\boldsymbol{\alpha}} P^n(x, dy) P(\boldsymbol{\alpha}, A) \\ &\geqslant P^n(x, \boldsymbol{\alpha}) P(\boldsymbol{\alpha}, A), \end{aligned}$$

then, $U(x, A) \geqslant \nu(A) U(x, \boldsymbol{\alpha})$. Hence, if $U(x, \boldsymbol{\alpha}) > 0$ for all $x \in X$, any set A such that $\nu(A) > 0$ will be accessible, which shows that **X** is ν-irreducible.

Theorem 2.2.10. *Suppose there is an atom $\boldsymbol{\alpha}$ in $\boldsymbol{X}$ such that $U(x, \boldsymbol{\alpha}) > 0$ for all*

$x \in E$. Then $\boldsymbol{X}$ is ν-irreducible, with $\nu = P(\boldsymbol{\alpha}, \bullet)$ and $\boldsymbol{\alpha}$ is an accessible atom.

Remark 2.2.3. *In section 2.1.8 we defined an atom for a kernel P as a pair (s, ν),where $s \in \mathcal{E}^+$ and ν is a probability measure, that satisfies $P \geqslant s \otimes \nu$. Notice that if $\boldsymbol{X}$ has an accessible atom $\boldsymbol{\alpha}$, then, P satisfies the minorization condition $M(1, \mathbb{I}_{\boldsymbol{\alpha}}, P(\boldsymbol{\alpha}, \bullet))$, therefore, the pair $(\mathbb{I}_{\boldsymbol{\alpha}}, P(\boldsymbol{\alpha}, \bullet))$ is an atom for the kernel P and $\boldsymbol{\alpha}$ is a small set. Moreover,*

$$(P - \mathbb{I}_{\boldsymbol{\alpha}} \otimes P(\boldsymbol{\alpha}, \bullet))(x, A) = \mathbb{I}_{\boldsymbol{\alpha}^c}(x) P(x, A) \quad \forall x \in E, A \in \mathcal{E}.$$

Theorem 2.2.11. [20] *[Atomic maximum principle] If the Markov chain $\boldsymbol{X}$ has an accessible atom $\boldsymbol{\alpha}$, then,*

$$U(x, \boldsymbol{\alpha}) = \mathbb{P}_x(\sigma_{\boldsymbol{\alpha}} < \infty) U(\boldsymbol{\alpha}, \boldsymbol{\alpha}) \quad \forall x \in E.$$

An important property of an accessible atom is that it can be used to characterize accessible sets. Essentially, a measurable set is accessible if and only if it can be accessed from the atom, in a finite time, with non-zero probability.

Theorem 2.2.12. [21] *Let $\boldsymbol{X}$ be an atomic Markov chain with accessible atom $\boldsymbol{\alpha}$ and $A \in \mathcal{E}$, then*

i) A is accessible if and only if $\mathbb{P}_{\boldsymbol{\alpha}}(\tau_A < \infty) > 0$.

ii) If A is not accessible, then A^c is accessible.

The following result indicates that atoms are either recurrent or transient. Furthermore, it shows that, in any atomic chain, accessible atoms are either all recurrent or all transient.

[20] Lemma 6.1.3 in [38]
[21] Lemma 6.1.4 in [38]

Theorem 2.2.13. [22] *Let $\boldsymbol{X}$ be a Markov chain with an atom $\boldsymbol{\alpha}$.*

i) $\boldsymbol{\alpha}$ *is recurrent if any of the following equivalent properties is satisfied*

(a) $\mathbb{P}_{\boldsymbol{\alpha}}(\tau_\alpha < \infty) = 1$,

(b) $\mathbb{P}_{\boldsymbol{\alpha}}(T(\alpha) = \infty) = 1$,

(c) $U(\boldsymbol{\alpha}, \boldsymbol{\alpha}) = \infty$

ii) $\boldsymbol{\alpha}$ *is transient if any of the following equivalent properties is satisfied*

(a) $\mathbb{P}_{\boldsymbol{\alpha}}(\tau_\alpha < \infty) < 1$,

(b) $\mathbb{P}_{\boldsymbol{\alpha}}(T(\alpha) < \infty) = 1$,

(c) $U(\boldsymbol{\alpha}, \boldsymbol{\alpha}) < \infty$

iii) If $\boldsymbol{\alpha}$ *is accessible and recurrent, any atom* $\boldsymbol{\beta}$ *satisfying* $\mathbb{P}_{\boldsymbol{\alpha}}(\tau_{\boldsymbol{\beta}} < \infty) > 0$ *is accessible and recurrent and* $\mathbb{P}_{\boldsymbol{\alpha}}(T(\boldsymbol{\beta}) = \infty) = \mathbb{P}_{\boldsymbol{\beta}}(T(\boldsymbol{\alpha}) = \infty) = 1$.

iv) If $\boldsymbol{\alpha}$ *is recurrent and there exists an accessible atom* $\boldsymbol{\beta}$*, then* $\boldsymbol{\alpha}$ *is accessible.*

From parts iii and iv of Theorem 2.2.13, we have that accessible atoms are either all recurrent or all transient. The next result shows that, in the atomic chain scenario, the recurrence or transience of an accessible atom characterizes the recurrence or transience of the chain.

Theorem 2.2.14. [23] *Suppose $\boldsymbol{X}$ is an atomic Markov chain with accessible atom $\boldsymbol{\alpha}$. Then it holds:*

i) $\boldsymbol{X}$ *is recurrent if and only* $\boldsymbol{\alpha}$ *is recurrent.*

ii) $\boldsymbol{X}$ *is transient if and only* $\boldsymbol{\alpha}$ *is transient.*

[22] Theorem 6.2.2 and Proposition 6.2.4 in [38].
[23] Theorem 6.2.7 in [38]

For an atom $\boldsymbol{\alpha}$ in $\mathbf{X}$, define a measure $\pi_{\boldsymbol{\alpha}}$ as follows:

$$\pi_{\boldsymbol{\alpha}}(A) = \mathbb{E}_{\boldsymbol{\alpha}}\left(\sum_{k=1}^{\tau_{\boldsymbol{\alpha}}} \mathbb{I}_A\{X_k\}\right). \tag{2.23}$$

The following theorem shows that $\boldsymbol{\alpha}$ is recurrent if and only if $\pi_{\boldsymbol{\alpha}}$ is invariant for the kernel.

Theorem 2.2.15. [24] *[Existence of an invariant measure] Let $\boldsymbol{X}$ be an atomic Markov chain with kernel P possessing and accessible atom $\boldsymbol{\alpha}$. Then,*

i) If $\boldsymbol{\alpha}$ is recurrent, then $\pi_{\boldsymbol{\alpha}}$ is invariant for P.

ii) If $\pi_{\boldsymbol{\alpha}}$ is invariant for P, then $\boldsymbol{\alpha}$ is recurrent.

iii) If $\boldsymbol{\alpha}$ is recurrent, then every other invariant measure π is proportional to $\pi_{\boldsymbol{\alpha}}$, satisfies $\pi(\boldsymbol{\alpha}) < \infty$ and $\pi(B) = \pi(\boldsymbol{\alpha})\pi_{\boldsymbol{\alpha}}(B) \quad \forall B \in \mathcal{E}$.

Part iii) of the previous result tells us that an atomic Markov chain has a unique, up to a multiplicative constant, invariant measure. The celebrated Kac's theorem uses this to characterize the positivity of irreducible atomic chains.

Theorem 2.2.16. *[Kac's Theorem][25] Suppose that $\boldsymbol{\alpha}$ is a recurrent atom for the atomic Markov chain $\boldsymbol{X}$ with kernel P. Then, $\boldsymbol{X}$ is positive recurrent if and only if $\mathbb{E}_{\boldsymbol{\alpha}}\tau_{\boldsymbol{\alpha}} < \infty$; and if π is the invariant probability measure for $\boldsymbol{X}$, then $\pi = \frac{\pi_{\boldsymbol{\alpha}}}{\mathbb{E}_{\boldsymbol{\alpha}}\tau_{\boldsymbol{\alpha}}}$.*

The following construction, presented in [88], allows us to construct, given a random variable Z, an atomic Markov chain where $\mathcal{L}(\tau_\alpha) = \mathcal{L}(\lfloor Z \rfloor)$.

Example 2.2.7. *Let Z_n be a sequence of i.i.d. random variables. Define X_n as*

[24] Theorem 6.4.2 in [38].
[25] Theorem 10.2.2 in [87]

$$X_n = \begin{cases} X_{n-1} - 1, & \text{if } X_{n-1} > 1, \\ Z_n, & \text{if } X_{n-1} \in [0,1]. \end{cases}$$

Then, X_n is an irreducible Markov Chain and the interval $[0,1]$ is an atom for the chain. Moreover, $P_x\left(\tau_{[0,1]} > n\right) = P\left(\lfloor Z_1 \rfloor > n\right)$.

We have saved for last what is perhaps the most important property of atomic chains: the *Block decomposition*. This property enables us to partition an atomic chain into independent blocks, which facilitates the study of this type of Markov chain, making it comparable to the i.i.d. case.

Theorem 2.2.17 (Block decomposition). [26] *Let $\boldsymbol{X}$ be an atomic Markov chain with an accessible atom $\boldsymbol{\alpha}$. Then, for any initial measure $\lambda \in \mathscr{M}(\mathcal{E})_+$ such that $\mathbb{P}_\lambda\left(\tau_{\boldsymbol{\alpha}} < \infty\right) = 1$ the following random blocks*

$$\begin{aligned} \mathcal{B}_0 &= \left(X_0, X_1, \ldots, X_{\tau_{\boldsymbol{\alpha}}(1)}\right) \\ \mathcal{B}_1 &= \left(X_{\tau_{\boldsymbol{\alpha}}(1)+1}, \ldots, X_{\tau_{\boldsymbol{\alpha}}(2)}\right) \\ &\ldots \\ \mathcal{B}_n &= \left(X_{\tau_{\boldsymbol{\alpha}}(n)+1}, \ldots, X_{\tau_{\boldsymbol{\alpha}}(n+1)}\right) \\ &\ldots \end{aligned}$$

are independent, and among them, $\{\mathcal{B}_n\}_{n\geqslant 1}$ are i.i.d. with common law $\mathcal{L}_{\mathbb{P}_{\boldsymbol{\alpha}}}\left(X_0, X_1, \ldots, X_{\tau_{\boldsymbol{\alpha}}(1)}\right)$.

As an application of Theorem 2.2.17 we will show how it can be used to study the sums of the values of a function over $\mathbf{X}$.

Let f be a function defined over E and define the random variables $f(\mathcal{B}_0) =$

[26] Corollary 2.3 in [29].

$\sum_{j=0}^{\tau_\alpha} f(X_j)$ and for $n \geqslant 1$,

$$f(\mathcal{B}_n) = \sum_{j=\tau_{\boldsymbol{\alpha}}(n)+1}^{\tau_\alpha(n+1)} f(X_j).$$

The next result is an immediate consequence of Theorem 2.2.17.

Theorem 2.2.18 (Independence of the excursions). *If $\boldsymbol{X}$ is an atomic Markov chain with a recurrent atom $\boldsymbol{\alpha}$, then, under $\mathbb{P}_{\boldsymbol{\alpha}}$, the sequence $\{f(\mathcal{B}_n)\}_{n\in\mathbb{Z}_+}$ is i.i.d. Moreover, for every initial measure $\lambda \in \mathscr{M}(\mathcal{E})_+$ such that $\mathbb{P}_\lambda(\tau_{\boldsymbol{\alpha}} < \infty) = 1$, the random variables $f(\mathcal{B}_n), n \in \mathbb{Z}_+$ are independent and for $n \geqslant 1$ they are i.i.d.*

For any measurable function f defined on E we will denote by $S_n(f)$ the partial sums of f over the chain, that is

$$S_n(f) = \sum_{k=0}^{n} f(X_k). \tag{2.24}$$

For a fixed atom, Theorem 2.2.18 allow us to express $S_n(f)$ as a sum of independent random variables as follows:

$$S_n(f) = f(\mathcal{B}_0) + \sum_{j=1}^{T(n)} f(\mathcal{B}_j) + \sum_{i=\tau_{\boldsymbol{\alpha}}(T(n)+1)+1}^{n} f(X_i), \tag{2.25}$$

where $T(n) = T_n(\boldsymbol{\alpha}) - 1$ counts the number of i.i.d. blocks up to time n. This term is called *number of regenerations up to time* n.

The following *random Law of Large Numbers* follows from (2.25), Theorem 6.8.1 in [55] and the fact that $\mathbb{E}[f(\mathcal{B}_i)] = \pi_\alpha(f)$ for $i \geqslant 1$.

Theorem 2.2.19. [27] *Let $\boldsymbol{X}$ be an atomic Markov chain with an accessible atom $\boldsymbol{\alpha}$ and let f be a $\pi_{\boldsymbol{\alpha}}$-integrable function. Then, for every initial distribution μ such that*

[27] Lemma 6.6.1 in [38]

$\mathbb{P}_\mu(\tau_{\boldsymbol{\alpha}} < \infty) = 1$,

$$\frac{S_n(f)}{T(n)} \to \pi_{\boldsymbol{\alpha}}(f) \quad \mathbb{P}_\mu - a.s.$$

This random Law of Large Numbers suggests that if we standardize by the number of complete blocks, we can derive a version of the Central Limit Theorem. In the positive recurrent scenario, the following result confirms the validity of this approach. In the null recurrent case, it is also true provided that the time of return to the atom does not have an excessively heavy tail. Before stating the theorem, we need a definition.

Definition 2.2.16 (Slowly varying functions)**.** *Measurable and positive function L, defined in $[a, +\infty)$ for some $a \geqslant 0$, that satisfies*

$$\lim_{x \to +\infty} \frac{L(xt)}{L(x)} = 1 \quad \forall t \geqslant a.$$

Two slowly varying functions are said to be equivalent if $\lim_{x\to+\infty} \frac{L(x)}{L'(x)} = 1$. A fundamental result in the theory of slowly varying functions[28] indicates that if L is a slowly varying function then, for every $\alpha > 0$, there exists an equivalent version L_α of L such that $x^\alpha L_\alpha(x)$ is strictly increasing and continuous in $[x_\alpha, +\infty]$ for some $x_\alpha \geqslant 0$. These L_α are called *normalizations of* L.

Remark 2.2.4. *Throughout this thesis, all slowly varying functions are unique up-to equivalence and, without loss of generality, we always use normalized versions.*

Theorem 2.2.20. *[CLT with random number of terms] Suppose $\boldsymbol{X}$ is an atomic, positive recurrent Markov chain with an accessible atom $\boldsymbol{\alpha}$. Let $f \in L^1(E, \pi_{\boldsymbol{\alpha}})$ such that*

[28] Proposition 1.3.4 and Theorem 1.5.5 in [21]

$\mathrm{Var}_{\boldsymbol{\alpha}}\left[f\left(\mathcal{B}_1\right)\right]$ *is finite, then,*

$$\sqrt{T\left(n\right)}\left(\frac{\sum_{i=1}^{T(n)} f\left(\mathcal{B}_i\right)}{T\left(n\right)} - \pi_\alpha\left(f\right)\right) \xrightarrow{d} N\left(0, \mathrm{Var}\left[f\left(\mathcal{B}_1\right)\right]\right). \tag{2.26}$$

and

$$\sqrt{T\left(n\right)}\left(\frac{S_n\left(f\right)}{T\left(n\right)} - \pi_\alpha\left(f\right)\right) \xrightarrow{d} N\left(0, \mathrm{Var}\left[f\left(\mathcal{B}_1\right)\right]\right). \tag{2.27}$$

Moreover, if $\boldsymbol{X}$ *is null recurrent and satisfies*

$$\mathbb{P}\left(\tau_{\boldsymbol{\alpha}} > n\right) \sim \frac{1}{n^\beta L\left(n\right)} \tag{2.28}$$

where $0 < \beta < 1$ *and* L *is slowly varying, then* (2.26) *holds, and, if in addition* $\mathbb{E}\left[\left(\left|f\right|\left(\mathcal{B}_1\right)\right)^2\right] < +\infty$ *then* (2.27) *also holds.*

The proof of Theorem 2.2.20 relies on being able to control the behavior of $T\left(n\right)$ in such a way that the convergence in distribution of $\sqrt{n}\left(\frac{\sum_{i=1}^{n} f(\mathcal{B}_i)}{n} - \pi_{\boldsymbol{\alpha}}\left(f\right)\right)$ does not change when we replace n by $T\left(n\right)$.

For positive recurrent chains, this is achieved thanks to Anscombe's Theorem,[29] using the fact that $\frac{n}{T(n)} \xrightarrow{a.s.} \mathbb{E}_{\boldsymbol{\alpha}}\tau_{\boldsymbol{\alpha}}$. The details of the proof can be found in section 6.7 of [38].[30]

In the null recurrent case, Anscombe's Theorem approach does not work because, even with the tail condition (2.28), $T\left(n\right)$ can only be controlled in distribution. Hence, in this scenario, the proof is based on the functional convergence of the stochastic processes $\sqrt{n}\left(\frac{\sum_{i=1}^{\lfloor nt\rfloor} f(\mathcal{B}_i)}{n} - \pi_{\boldsymbol{\alpha}}\left(f\right)\right)$ and $\frac{T(\lfloor nt\rfloor)}{n^\beta L(n)}$ (see Theorem 2.2.39), and a deep result by [68] that implies the functional convergence of the joint pro-

[29] The original theorem appeared in [4]. See Theorem E.4.5 in [38] for a more modern formulation.

[30] Equation (2.26) appears as an intermediate result in the proof of their Theorem 6.7.1 and (2.27) is a direct consequence of their equation (6.7.6).

cess. The full proof can be found in [9]. [31]

In the positive recurrent case, Theorem 2.2.20 can be improved to have a formulation very similar to the CLT for i.i.d. data. In effect, assume that $\mathbf{X}$ is positive recurrent with invariant probability measure π and f satisfies the conditions of Theorem 2.2.20. Define the function $\widetilde{f} = f - \pi(f)$, then

$$\pi_{\boldsymbol{\alpha}}\left(\widetilde{f}\right) = \pi_{\boldsymbol{\alpha}}(f - \pi(f)) = \pi_{\boldsymbol{\alpha}}(f) - \pi_{\boldsymbol{\alpha}}(E)\,\pi(f)$$

which equals 0 by Kac's theorem. Applying (2.27) to $\widetilde{f}$ and using Slutsky's Theorem[32] to replace $T(n)$ by n, we get the following result:

Theorem 2.2.21. [33] *[CLT for atomic positive recurrent Markov chains] If $\boldsymbol{X}$ is an atomic, positive recurrent Markov chain with an accessible atom $\boldsymbol{\alpha}$ and invariant probability measure π and $f \in L^1(E, \pi_{\boldsymbol{\alpha}})$ is such that $\mathrm{Var}_{\boldsymbol{\alpha}}[f(\mathcal{B}_1)]$ is finite, then,*

$$n^{-\frac{1}{2}} \sum_{k=1}^{n} \left(f(X_k) - \pi(f)\right) \xrightarrow{d} N\left(0, \frac{\mathrm{Var}\left[\widetilde{f}(\mathcal{B}_1)\right]}{\mathbb{E}_{\boldsymbol{\alpha}}\tau_{\boldsymbol{\alpha}}}\right), \tag{2.29}$$

where

$$\mathrm{Var}\left[\widetilde{f}(\mathcal{B}_1)\right] = E_{\boldsymbol{\alpha}}\left[\left(\sum_{k=1}^{\tau_{\boldsymbol{\alpha}}}\left(f(X_k) - \pi(f)\right)\right)^2\right].$$

Observe that the only dissimilarity between (2.29) and the CLT for i.i.d. data lies in the form of the variance.

Remark 2.2.5. *In the null-recurrent case, due to the fact that π is σ-finite but not finite, if f is such that $\pi(f) \neq 0$, then there is no deterministic recentering.*

Atomic chains are not rare, for example, when E is countable (Example 2.2.2)

[31] See page 1140 for (2.26) and their equation (6) for (2.27).
[32] Theorem 5.11.4, pp 248 in [55].
[33] Theorem 6.7.1 in [38].

every singleton (set with only one element) is an atom, and if the chain is irreducible, then every singleton is an accessible atom. In the case of the Random Walk on a half line, as described in Example 2.2.3, the set $\{0\}$ is an accessible atom if $\mathbb{P}(Z < 0) > 0$[34]. However, the reason for studying atoms is not just because of the existence of accessible singletons in some models, but rather the ability to artificially construct sets with an atomic structure by extending the probabilistic structure of the chain in the irreducible case. This permits the application of decomposition of the chain into separate, identical parts (via Theorem 2.2.17) which serve as building blocks in subsequent analysis.

2.2.8 Splitting

In this section, we will show how to "construct" atomic chains from irreducible chains. The idea is to construct a split chain $\check{\mathbf{X}}$ in a split space $\check{E} = E_0 \cup E_1$ where E_i are "copies", in a specific sense, of the original space E, in such a way that:

i) The original chain $\mathbf{X}$ is a marginal chain of $\check{\mathbf{X}}$, in the sense that for all initial distribution λ, $A \in \mathcal{E}$ and $n \geqslant 0$ we have $\lambda P^n(A) = \check{\lambda}\check{P}^n\left(\check{A}\right)$, where $\check{P}$, $\check{\lambda}$ and $\check{A}$ are respectively a kernel, an initial probability and a measurable set defined in the split space.

ii) E_1 is an accessible atom for $\check{\mathbf{X}}$.

Let us assume, for the moment, that $\mathbf{X}$ is a φ-irreducible Markov chain and its kernel P satisfies the minorization condition $M(1, s, \nu)$, that is, there exist a small function s, taking values in the interval $[0, 1]$, and a non-trivial probability measure ν such that

$$P \geqslant s \otimes \nu. \tag{2.30}$$

[34] See Example 6.1.2 in [38] for a proof of this result

The minorization condition (2.30) allows us to define the following sub-stochastic kernel in $(E, \mathcal{E})$,

$$Q(x, A) = \begin{cases} (1 - s(x))^{-1} [P(x, A) - s(x)\nu(A)] & , \quad s(x) < 1 \\ \mathbb{I}_A(x) & , \quad s(x) = 1. \end{cases} \tag{2.31}$$

By using Q, we can break down P into two parts:

$$P(x, A) = s(x)\nu(A) + [1 - s(x)] Q(x, A). \tag{2.32}$$

Remark 2.2.6. *Consider the set $C = \{x \in E : s(x) = 1\}$. If $\varphi(C) > 0$, then C is a small set*[35]*, therefore, $P(x, A) \geqslant \nu(A)$ for all $x \in C$ and $A \in \mathcal{E}$, which implies that C is an accessible atom for $\boldsymbol{X}$ and $P(x, A) = \nu(A) \quad \forall x \in C, A \in \mathcal{E}$*[36]*.*

The decomposition (2.32) indicates that a transition starting from any state x in E can be thought of as happening in two steps. First, a coin is flipped with the probability of landing on "head" equal to $s(x)$. If "head" comes up, the Markov chain moves based on the probability law of ν, otherwise, it moves based on $Q(x, \bullet)$. The most important aspect here is that getting "head" results in a transition law that is independent of the state x.

To properly formalize this heuristic, we will work on the space $\left(\check{E}, \check{\mathcal{E}}\right)$ defined as the product of the measurable spaces $(E, \mathcal{E})$ and $(\{0, 1\}, \mathcal{P}(\{0, 1\}))$. We will use the term *split space* to refer to both the set $\check{E}$ and the measurable space $\left(\check{E}, \check{\mathcal{E}}\right)$.

For all $x \in E$, $A \in \mathcal{E}$ we denote

$$x_0 = (x, 0), \qquad x_1 = (x, 1);$$

[35] Remark 2.1-iv, pp. 16 of [90]
[36] See pp. 191 of [38].

$$\check{A}_0 = A \times \{0\}, \qquad \check{A}_1 = A \times \{1\}, \qquad \check{A} = A \times \{0,1\}.$$

In the following, we identify each set $A \in \mathcal{E}$ with its corresponding set $\check{A} \in \check{\mathcal{E}}$, as defined in the previous paragraph. Thus, the σ-algebra $\mathcal{E}$ can be viewed as a subset of $\check{\mathcal{E}}$. For any measure $\lambda \in \mathscr{M}(\mathcal{E})_+$, we can define its extension $\check{\lambda}$ to $\mathscr{M}\left(\check{\mathcal{E}}\right)_+$ by setting its values on the sets $\check{A}_i$ (where $A \in \mathcal{E}$ and $i = 0, 1$) as follows:

$$\check{\lambda}\left(\check{A}_0\right) = \int_A \left(1 - s\left(x\right)\right) d\lambda\left(x\right) \quad , \quad \check{\lambda}\left(\check{A}_1\right) = \int_A s\left(x\right) d\lambda\left(x\right).$$

A $\mathcal{E}$-measurable function f, is extended to a $\check{\mathcal{E}}$-measurable function by setting $\check{f}\left(x_0\right) = \check{f}\left(x_1\right) = f\left(x\right)$. With these definitions, for every $\mathcal{E}$-measurable function f and every measure $\lambda \in \mathscr{M}(\mathcal{E})_+$ we have $\int_E f\left(x\right) d\lambda\left(x\right) = \int_{\check{E}} \check{f}\left(z\right) d\check{\lambda}\left(z\right)$.

Now the final and most subtle step in the construction of the split chain is to build a kernel $\check{P}$ in $\left(\check{E}, \check{\mathcal{E}}\right)$ that expresses the heuristic we described before. For $x \in E$ and $\check{A} \in \check{\mathcal{E}}$ define the following kernel[37] on the split space

$$\check{P}\left(x_0, \check{A}\right) = \begin{cases} \check{\nu}\left(\check{A}\right) & , \quad x \in \{s\left(x\right) = 1\}, \\ \check{Q}\left(x, \check{A}\right) & , \quad x \in \{s\left(x\right) < 1\}; \end{cases}$$

$$\check{P}\left(x_1, \check{A}\right) = \check{\nu}\left(\check{A}\right),$$

where $\check{Q}\left(x, \bullet\right)$ is defined as the extension to $\left(\check{E}, \check{\mathcal{E}}\right)$ of the measure $Q\left(x, \bullet\right)$.

Let $\check{\mathbf{X}} = \left\{\left(\check{X}_n, \check{Y}_n\right)\right\}_{n \geqslant 0}$ be a Markov chain (defined in the split space) with initial measure $\check{\lambda}$ and kernel $\check{P}$. It was shown in pp.61 of [90] that, for any $x \in$

[37] See pp.311 in [89].

$E, A \in \mathcal{E}$ the random variables $\check{X}_n$ and $\check{Y}_n$ satisfy

$$\begin{aligned}
\mathbb{P}\left(\check{X}_{n+1} \in A | \check{X}_n = x\right) &= P(x, A), \\
\mathbb{P}\left(\check{Y}_n = 1 | \check{X}_n = x\right) &= s(x), \\
\mathbb{P}\left(\check{X}_{n+1} \in A | \check{Y}_n = 1\right) &= \nu(A), \\
\mathbb{P}\left(\check{X}_{n+1} \in A | \check{Y}_n = 0, \check{X}_n = x\right) &= Q(x, A).
\end{aligned}$$

The following theorem shows that the bivariate process $\check{\mathbf{X}}$ is an atomic Markov chain and the distribution of its first component is identical to the distribution of our original chain.

Theorem 2.2.22. [38] *The split chain $\check{\boldsymbol{X}}$ is an atomic Markov chain and the set $\check{E}_1 = E \times \{1\}$ is an accessible atom. Moreover, for any probability measure λ on $(E, \mathcal{E})$, the marginal distribution of the first component $\left\{\check{X}_n\right\}$ of the split chain $\check{\boldsymbol{X}}$ and the distribution of the original chain $\boldsymbol{X}$ are identical. In particular,*

$$\check{\lambda}\check{P}^n\left(\check{A}\right) = \lambda P^n(A) \quad \forall A \in \mathcal{E}.$$

In the following, we will identify the original chain with $\check{X}_n$ and we will write X_n instead of $\check{X}_n$. We will denote by $\check{\boldsymbol{\alpha}}$ the atom $\check{E}_1$ and we will write $\mathbb{P}_{\check{\boldsymbol{\alpha}}}$ for the probability measure defined on $\sigma\left(\cup_{n\in\mathbb{N}}\mathcal{F}_n^{\check{X}}\right)$ and corresponding to the initial state $Y_0 = 1$, i.e. $\mathbb{P}_{\check{\boldsymbol{\alpha}}} = \mathcal{L}\left(X_n, \check{Y}_n, n \geqslant 1 | Y_0 = 1\right)$. Similarly, for any $x \in E$, we will denote by $\mathbb{P}_x$ the probability measure corresponding to the initial state $X_0 = x$.

Theorem 2.2.23. [39] *The split kernel $\check{P}$ satisfies*

i) $\check{P}^n(\check{\boldsymbol{\alpha}}, A) = \mathbb{P}_{\check{\boldsymbol{\alpha}}}(X_n \in A) = \nu P^{n-1}(A) \quad \forall A \in \mathcal{E}, n \geqslant 1.$

[38] This result appears as Theorems 1 and 2 in [89] and as Theorem 4.2 in [90].
[39] Equation (4.19) in [90]

ii) $\check{P}(x, \check{\boldsymbol{\alpha}}) = \mathbb{P}_x\left(\check{Y}_n = 1\right) = \mathbb{E}_x\left[s\left(X_n\right)\right] = P^n s\left(x\right) \quad \forall x \in E, n \geqslant 0.$

Theorem 2.2.24. [40] *Let $\sigma_{\check{\boldsymbol{\alpha}}}$ and $\tau_{\check{\boldsymbol{\alpha}}}$ be the hitting and return times of the atom $\check{\boldsymbol{\alpha}}$ in the split chain, then*

i) $\mathbb{P}_{\check{\boldsymbol{\alpha}}}\left(\tau_{\check{\boldsymbol{\alpha}}} = n\right) = \nu\left(P - s \otimes \nu\right)^{n-1} s$ *for* $n \geqslant 1$.

ii) $\mathbb{P}_x\left(\sigma_{\check{\boldsymbol{\alpha}}} = n\right) = \left(P - s \otimes \nu\right)^n s\left(x\right)$ *for* $x \in E$ *and* $n \geqslant 0$.

iii) $\mathbb{P}_{\check{\boldsymbol{\alpha}}}\left(X_{n+1} \in A \cap \tau_{\check{\boldsymbol{\alpha}}} \geqslant n+1\right) = \nu\left(P - s \otimes \nu\right)^n\left(A\right)$ *for all* $A \in \mathcal{E}$.

Remark 2.2.7. *If* $\mathbf{X}$ *has an atom* $\boldsymbol{\alpha}$*, remark 2.2.3 shows that it satisfies the minorization condition* $M\left(1, \mathbb{I}_{\boldsymbol{\alpha}}, P\left(\boldsymbol{\alpha}, \bullet\right)\right)$ *and* $P - \mathbb{I}_{\boldsymbol{\alpha}} \otimes P\left(\boldsymbol{\alpha}, \bullet\right) = \mathbb{I}_{\boldsymbol{\alpha}^c} P$*. In this case, the auxiliary process* $\mathbf{Y}$ *is defined as* $Y_n = \mathbb{I}_{\boldsymbol{\alpha}}\left(X_n\right)$ *and all the results in this subsection hold.*

At the beginning of this section, we assumed that the kernel P satisfied the minorization condition $M\left(1, s, \nu\right)$. This assumption is not too strict, because, by Corollary 2.1.1 and Theorem 2.1.9, we have that if $\mathbf{X}$ is φ-irreducible with kernel P, then the P_ε-chains are strongly aperiodic and there exists m such that the minorization condition $M\left(1, s, \nu\right)$ is satisfied for P^m. Therefore, we can apply Theorem 2.2.22 to the P_ε-chains (and to some m-skeleton chain) and construct a split chain. A typical proof of a property $\mathbf{P}$ for $\mathbf{X}$ using this technique is as follows:

Step 1 Prove $\mathbf{P}$ for atomic chains.

Step 2 Apply Theorem 2.2.22 to construct the split chain $\check{\mathbf{X}}$. Show that if $\mathbf{P}$ holds for $\check{\mathbf{X}}$ it also holds for $\mathbf{X}$. Then, by step 1, $\mathbf{P}$ holds for chains that satisfy the minorization condition with $m_0 = 1$.

[40] Equations (4.20), (4.21) and (4.22) in [90].

Step 3 Show that if **P** holds for the P_ε-chains (or for some m-skeleton) then it holds for the original chain. Hence, the result follows by applying step 2 to the P_ε-chains (or to the m-skeleton that satisfy the minorization condition with $m_0 = 1$).

As a demonstration, we use this technique to prove the following result, which, as a direct corollary shows that every irreducible Markov chain admits a unique (up to a multiplicative constant) invariant measure.

Theorem 2.2.25. [41] *If **X** is irreducible, aperiodic and satisfies the minorization condition $M(m_0, s, \nu)$, then, the measure π defined as*

$$\pi(A) = \nu \sum_{n=0}^{+\infty} \left(P^{m_0} - s \otimes \nu\right)^n (A) \quad A \in \mathcal{E}, \tag{2.33}$$

*is an invariant measure for the chain **X**.*

Proof. As promised, the proof will be divided in three steps:

Step 1 If **X** is atomic, the existence of the invariant measure $\pi_{\boldsymbol{\alpha}}$ and its representation is given by (2.23). By Fubini's Theorem,

$$\pi_{\boldsymbol{\alpha}}(A) = \sum_{n=1}^{+\infty} \mathbb{P}_{\boldsymbol{\alpha}}(X_n \in A \cap \tau_{\boldsymbol{\alpha}} \geqslant n)$$

By part iii of Proposition 4.4 in [90], $\mathbb{P}_{\boldsymbol{\alpha}}(X_n \in A \cap \tau_{\boldsymbol{\alpha}} \geqslant n) = \nu\left(\mathbb{I}_{\boldsymbol{\alpha}^c} P\right)^{n-1}$ where $\mathbb{I}_{\boldsymbol{\alpha}^c} P$ is the kernel defined by $\left(\mathbb{I}_{\boldsymbol{\alpha}^c} P\right)(x, A) = \mathbb{I}_{\boldsymbol{\alpha}^c}(x)\, P(x, A)$ and $\nu = \mathbb{P}(\boldsymbol{\alpha}, \bullet)$. Because **X** is atomic, remark 2.2.3 shows that it satisfies the minorization condition $M\left(1, \mathbb{I}_{\boldsymbol{\alpha}}, P(\boldsymbol{\alpha}, \bullet)\right)$ and $P - \mathbb{I}_{\boldsymbol{\alpha}} \otimes P(\boldsymbol{\alpha}, \bullet) = \mathbb{I}_{\boldsymbol{\alpha}^c} P$. Hence, the theorem holds for atomic chains.

[41] Corollary 5.2 in [90]

Step 2 Now assume that $\mathbf{X}$ satisfies the minorization condition $M(1, s, \nu)$. By Theorem 2.2.22, we can construct the split chain $\check{\mathbf{X}}$ with an accessible atom $\check{\boldsymbol{\alpha}}$. By Step 1, this chain admits an invariant measure $\pi_{\check{\boldsymbol{\alpha}}}$ and the form of this measure is given by (2.23).

When restricted to $\mathbf{X}$, this measure is invariant for $\mathbf{X}$, as was shown in (5.7) of [90]. A simple application of Fubini's Theorem and part iii of Theorem 2.2.24 shows that, for $A \in \mathcal{E}$

$$\pi(A) = \mathbb{E}_{\check{\boldsymbol{\alpha}}}\left(\sum_{k=1}^{\tau_{\check{\boldsymbol{\alpha}}}} \mathbb{I}_A\{X_k\}\right) = \nu \sum_{n=0}^{+\infty} (P - s \otimes \nu)^n (A) = \nu G_{s,\nu}(A),$$

where $G_{s,\nu} = \sum_{n=0}^{+\infty} (P - s \otimes \nu)^n$. Therefore, the result holds for chains that satisfy the minorization condition $M(1, s, \nu)$.

Step 3 Given that $\mathbf{X}$ satisfies the minorization condition $M(m_0, s, \nu)$, the m_0-skeleton chain, satisfies the same condition but with $m = 1$, therefore, by step 2 π is an invariant measure for the m_0-skeleton. By Theorem 10.4.5 in [87], π is also invariant for $\mathbf{X}$, which completes the proof. □

The following result is an extension of Theorem 2.2.25. Its proof is an example of the technique we described. It can be found as Theorem 10.4.9 in [87].

Theorem 2.2.26. *Let $\mathbf{X}$ be a ψ-irreducible recurrent chain with transition kernel P. Then, $\mathbf{X}$ admits a non-zero invariant measure π that satisfies*

i) π is unique up to a multiplicative constant.

ii) $\pi(C) < \infty$ for every small set C.

iii) For any accesible set A and $B \in \mathcal{E}$,

$$\pi(B) = \int_A \mathbb{E}_y \left(\sum_{k=1}^{\tau_A} \mathbb{I}\{X_k \in B\} \right) d\pi(y).$$

iv) π is equivalent to ψ.

The splitting technique shown in this section is one of the pillars of the modern Markov chain Theory. It was discovered almost simultaneously, but by slightly different methods, by Nummelin [89, 90] and Athreya and Ney [7]. The construction presented in this thesis follows the approach introduced by Nummelin, which is the most widely used in the literature.

2.2.9 Harris recurrent Markov chain

For atomic chains, we have seen in part b of Theorem 2.2.13, that recurrence of an atom is equivalent to the property that the number of visits to the atom is almost surely infinite when starting from the atom. In the general case, this no longer holds, as the following example shows.

Example 2.2.8. [42] *Define a Markov chain $\boldsymbol{X}$ in $\mathbb{N}$, such that if $X_n = k$, then, the chain moves to $k+1$ with probability $e^{-\frac{1}{n^2}}$ or jumps back to zero where it is absorbed. The kernel is as follows:*

$$P(0,0) = 1 \quad , \quad P(k,k+1) = e^{-\frac{1}{k^2}} \quad , \quad P(k,0) = 1 - e^{-\frac{1}{k^2}}, \quad k \geqslant 1.$$

The Markov chain in question is irreducible and contains an absorbing set comprising only the state 0. As a consequence, δ_0 is a maximal irreducibility measure and every

[42] Example 10.2.3 in [38]

accessible set must necessarily include the absorbing state 0. *For* $k \geqslant 1$, *the Markov property implies that*

$$\mathbb{P}_k\left(\tau_0 = +\infty\right) = \exp\left(-\sum_{j=k}^{+\infty} \frac{1}{j^2}\right) < 1,$$

therefore, $0 < \mathbb{P}_k\left(\tau_0 = +\infty\right) < 1$. *Consider an accessible set* A, *such that* $k \geqslant 1$ *is in* A. *Then,* $\mathbb{P}_k\left(T\left(A\right) = +\infty\right) \geqslant \mathbb{P}_k\left(\tau_0 = +\infty\right)$, *which implies that* $\mathbb{E}_k T\left(A\right) = +\infty$, *and hence the recurrence of the chain. On the other hand, if* A *is finite,* $\mathbb{P}_k\left(T\left(A\right) = +\infty\right)$ *equals* $\mathbb{P}_k\left(\tau_0 = +\infty\right) < 1$ *hence, the probability of returning to* A *infinitely often is not 1.*

In order to handle situations like the one described in the previous example, we need to define a stronger type of recurrence.

Definition 2.2.17 (Harris recurrence)**.** *An irreducible Markov Chain* $\boldsymbol{X}$ *with maximal irreducibility measure* ψ *is said to be Harris recurrent if for all* $x \in E$ *and all* $A \in \mathcal{E}$ *such that* $\psi\left(A\right) > 0$ *we have* $\mathbb{P}_x\left(\boldsymbol{X} \text{ visits } A \text{ infinitely often}\right) = 1$.

In other words, $\mathbf{X}$ is Harris recurrent if $\mathbb{P}_x\left(T\left(A\right) = \infty\right) = 1$ for all $x \in E$ and all ψ-positive sets A. The following result shows that Harris recurrence can be analyzed by looking at the P_ε-chains.

Theorem 2.2.27. [43] $\boldsymbol{X}$ *is Harris recurrent if and only if for some* ε *(and then for all) the* P_ε*-chain is Harris recurrent.*

Recall from (2.24) that if f is a function defined on E, then $S_n\left(f\right) = \sum_{k=0}^{n} f\left(X_k\right)$. The next theorem shows that, under Harris recurrence, the order of $S_n\left(f\right)$ is the same for every measurable function f such that $\int f d\pi \neq 0$, where π is an invariant

[43] Proposition 8.2.13 in [40]

measure for **X**. Moreover, it shows that this property characterizes Harris recurrence.

Theorem 2.2.28. [44] *The following propositions are equivalent when a σ-finite invariant measure π exists for **X***

1. *For every $f, g \in L^1(\pi)$ with $\int g d\pi \neq 0$*

$$\lim_{n\to\infty} \frac{S_n(f)}{S_n(g)} = \frac{\pi(f)}{\pi(g)}. \tag{2.34}$$

2. *The invariant σ-field Σ is $\mathbb{P}_x$ trivial for all $x \in E$.*

3. ***X** is Harris recurrent.*

The following result provides a similar ratio limit result as (2.34) but for expectations. It is important to notice that in the null recurrent case, it only applies to small functions.

Theorem 2.2.29. [45] *If **X** is Harris recurrent with invariant measure π, then, for every initial measures $\lambda, \mu \in \mathscr{M}(\mathcal{E})_+$ and small functions s, h such that $\pi(h) \neq 0$, we have*

$$\lim_{n\to+\infty} \frac{\mathbb{E}_\lambda\left[\sum_{t=0}^n s(X_t)\right]}{\mathbb{E}_\mu\left[\sum_{t=0}^n h(X_t)\right]} = \lim_{n\to+\infty} \frac{\lambda G^{(n)} s}{\mu G^{(n)} h} = \frac{\pi(s)}{\pi(h)}, \tag{2.35}$$

where $G^{(n)}$ is the n-partial sum of the potential kernel of P[46]*. If **X** is positive recurrent, then h, s can be taken as elements of $L^1(\pi)$.*

Equation (2.34), known as the *Ratio Limit Theorem*, allows us to obtain strong convergence results for $S_n(f)$.

[44] Theorem 17.3.2 in [87]

[45] Both statements appear in pp. 130 of [90]. It is worth pointing out that a small function is a special function.

[46] See Definition 2.1.2

Theorem 2.2.30. *Let $\boldsymbol{X}$ be a Harris recurrent Markov chain with invariant measure π and initial probability λ. Let f be a finite π integrable function.*

i) If $\boldsymbol{X}$ is positive recurrent and π is the unique invariant probability measure, then

$$\frac{S_n(f)}{n} \to \pi(f) \quad \mathbb{P}_\mu \ a.s. \tag{2.36}$$

ii) If $\boldsymbol{X}$ is null recurrent, then

$$\frac{S_n(f)}{n} \to 0 \quad \mathbb{P}_\mu \ a.s. \tag{2.37}$$

From the previous theorem, we see that when $\mathbf{X}$ is positive recurrent the behavior of $S_n(f)$ is similar to the i.i.d. case, however, (2.37) indicates that the null recurrent scenario is a whole different story. In order to properly study this case, we need a few new tools.

Let D be a small set and λ an initial measure. Define $a(t)$

$$a(t) = \pi(D)^{-1} \sum_{k=1}^{\lfloor t \rfloor} \lambda P^k(D) = \pi(D)^{-1} \lambda G^{\lfloor t \rfloor}(\mathbb{I}_D) \tag{2.38}$$

Remark 2.2.8. *This non-negative and increasing function is called the truncated Green function. By (2.35) the asymptotic order of $a(t)$ (when $t \to +\infty$) depends only on the transition kernel of the Markov chain.*

Example 2.2.9. [47] *[Truncated Green function for random walks in $\mathbb{R}$] Consider a random walk as the ones defined in example 2.2.3, with $\mathbb{E}Z_1 = 0$ and $\mathbb{E}Z_1^2 < +\infty$. It*

[47] See Section 6 in [28].

was shown in [28] that the truncated Green function of this chain satisfies

$$a(t) \sim \sqrt{\frac{2t}{\pi \mathbb{E}(Z_1^2)}}.$$

as t goes to $+\infty$.

The following two results, presented in [28], show that the behavior of $S_n(f)$ when n goes to $+\infty$ is closely related with $a(n)$.

Theorem 2.2.31. [48] *Let $\boldsymbol{X}$ be a Harris recurrent Markov chain with invariant measure π. Then, for every nonnegative function $f \in L^1(E, \pi)$ with $\pi(f) > 0$ and every initial distribution λ both the sequences*

$$\left\{\frac{S_n(f)}{a(n)}\right\}_{n\geqslant 1} \quad \text{and} \quad \left\{\left(\frac{S_n(f)}{a(n)}\right)^{-1}\right\}_{n\geqslant 1}$$

are bounded in probability, where the random variables in the second term are allowed to take the value ∞.

Theorem 2.2.32. [49] *Under the same hypothesis of Theorem 2.2.31, there exists a positive constant K such that*

$$\limsup_{n\to\infty} \frac{S_n(f)}{a\left(\frac{n}{H(a(n))}\right) H(a(n))} = K \int f(x)\, d\pi(x) \quad a.s.$$

where $H(x) = \log\log(\max\{x, e^e\})$.

In [30], a version of Theorem 2.2.32 is given for the case que when $\pi(f) = 0$

Theorem 2.2.33. [50] *Let $\boldsymbol{X}$ be a Harris recurrent Markov chain with invariant measure π and kernel P and f be a measurable function satisfying:*

[48] Theorem 2.1 in [28]
[49] Theorem 2.2 in [28]
[50] Theorem 1.1 in [30]

i) $\int f(x)\, d\pi(x) = 0 \quad , \quad \int f^2(x)\, d\pi(x) < \infty$;

ii) $\sup_{0<a<1} \int \left| \sum_{k=1}^{+\infty} a^k f(x) P^k f(x) \right| d\pi(x) < +\infty$,

then, there exists a constant $\Lambda_f \geqslant 0$ *such that*

$$\limsup_{n\to\infty} \frac{S_n(f)}{\sqrt{a\left(\frac{n}{H(a(n))}\right) H(a(n))}} = \Lambda_f \quad a.s.$$

Further, if in addition to i and ii the following two conditions also hold

ii') $\sum_{k=1}^{+\infty} f(x) \left(P^k f\right)(x) \in L^1(E, \pi)$,

iii) $\sigma_f = \int f^2(x)\, d\pi(x) + 2\sum_{k=1}^{\infty} f(x) P^k f(x)\, d\pi(x) > 0$.

Then $\Lambda_f > 0$.

2.2.10 β-null recurrent Markov chains

With the objective of finding the exact limit distribution of $\frac{S_n(f)}{a(n)}$ in the null recurrent case, the concept of *regularly varying* Markov chain was introduced in [28]. Before stating this concept, we need a few definitions.

A positive and measurable function g, defined in $[a, +\infty)$, where $a \geqslant 0$, is *regularly varying* at infinity if $\lim_{x\to+\infty} \frac{g(xt)}{g(x)}$ exists for all $t > a$. Theorem 1.4.1. in [21] shows that g is regularly varying at infinity if and only if there exists a real constant β and a slowly varying function L such that $g(t) = t^\beta L(t)$. The number β is called *index of regular variation* and it is unique for the function, i.e. if there exists β, β_1 and L, L_1 such that $g(t) = t^\beta L(t) = t^{\beta_1} L_1(t)$ then $\beta = \beta_1$ and $L = L_1$.

We will say that a Markov chain $\mathbf{X}$ is *regularly varying* if its truncated Green function is regularly varying. Thanks to the uniqueness of the index of regular

variation β we now introduce the concept of β-*regular* Markov chain.

Definition 2.2.18 (β-regular Markov chain)**.** *ψ-irreducible, Harris recurrent Markov chain $\boldsymbol{X}$ such that its truncated Green function is regularly varying with index β.*

Remark 2.2.9. *As was pointed out in [28], every irreducible, positive recurrent Markov chain is 1-regular, and if $\boldsymbol{X}$ is null recurrent and β-regular, then $0 \leqslant \beta < 1$. Therefore, for any β-regular chain, $\beta \in [0, 1]$.*

The notation used in the literature regarding β-regular chains differs a little between papers. In order to use the same notation in all of our results, define

$$u(t) = \begin{cases} a(t), & \beta \in \{0, 1\} \\ \Gamma(1+\beta)\, a(t), & 0 < \beta < 1 \end{cases},$$

where Γ is the Gamma function.

In Theorem 2.2.31, we saw that under Harris recurrence, $\frac{S_n(f)}{a(n)}$ is bounded in probability. The following result describes the asymptotic limits assuming β-regularity.

Theorem 2.2.34. [51] *Let $\boldsymbol{X}$ be a β-regular Markov chain with kernel P and invariant measure π. Then, for every non-negative function $f \in L^1(E, \pi)$, and every initial distribution λ,the sequence of random variables $\frac{S_n(f)}{u(n)}$ converges weakly (for $\beta < 1$). Moreover, the limit distributions are as follows:*

i) When $\beta = 0$, the limit distribution is the exponential distribution with parameter $\int f(x)\, d\pi(x)$.

[51] Theorem 2.3 in [28]

ii) If $0 < \beta < 1$, *the limit distribution is*

$$M_\beta(1) \int f(x)\, d\pi(x),$$

where $M_\beta(1)$ *is a Mittag-Leffler distribution with parameter* β.

iii) When $\beta = 1$,

$$\frac{S_n(f)}{u(n)} \to \int f(x)\, d\pi(x) \quad \text{in probability.}$$

Remark 2.2.10. *The Mittag-Leffler distribution with parameter* β *is a non-negative continuous distribution, whose moments are given by*

$$\mathbb{E}\left(M_\beta^m(1)\right) = \frac{m!}{\Gamma(1+m\beta)} \quad m \geqslant 0.$$

By (3.39) in [67], its Laplace transform is

$$\mathbb{E}\left[\exp\left(-sM_\beta(1)\right)\right] = \sum_{k=0}^{+\infty} \frac{(-s)^k}{\Gamma(1+k\beta)} \quad s \geqslant 0. \tag{2.39}$$

Remark 2.2.11. *In the original formulation of Theorem 2.2.34 (Theorem 2.3 in [28]), the normalization sequence is* $a(n)$ *and the limit distribution when* $0 < \beta < 1$ *is written as* $G_\beta^{-\beta} \int f(x)\, d\pi(x)$, *where* G_β *is a stable random variable with Laplace transform*

$$\mathbb{E}\left(-tG_\beta\right) = \exp\left(-\frac{t^\beta}{\Gamma(\beta+1)}\right).$$

Notice that $G_\beta = \frac{Y}{\Gamma(1+\beta)^{\frac{1}{\beta}}}$ *where* Y *has Laplace transform* $\exp\left(-s^\beta\right)$. *Let* F *be the cumulative distribution function of* Y. *Then,* $\mathbb{P}\left(G_\beta^{-\beta} \leqslant x\right) = 1 - F\left(\frac{\Gamma(1+\beta)^{\frac{1}{\beta}}}{x^{\frac{1}{\beta}}}\right)$.

By Example b, pp 453 of [46], this implies that the Laplace transform of $G_\beta^{-\beta}$ *is*

$$\sum_{k=0}^{+\infty} \frac{\left(-\Gamma\left(1+\beta\right)s\right)^k}{\Gamma\left(1+k\beta\right)} \quad s \geqslant 0,$$

which coincides with the Laplace transform of $\Gamma\left(1+\beta\right) M_\beta\left(1\right)$ *by* (2.39).

The equivalent theorem, but for the case when $\pi\left(f\right) = 0$ was presented in [30] and is as follows:

Theorem 2.2.35. *Let* $\boldsymbol{X}$ *be a* β*-regular Harris recurrent Markov chain. Under conditions i and ii' of Theorem 2.2.33 we have*

$$\frac{S_n\left(f\right)}{\sqrt{u\left(n\right)}} \xrightarrow{d} \sigma_f \sqrt{D_\beta} U \tag{2.40}$$

where U *and* D_β *are independent random variables,* $U \sim N\left(0,1\right)$*, and the distribution of* D_β *depends on* β *as follows:*

i) When $\beta = 0$*,* D_0 *is an exponential distribution with parameter 1.*

ii) When $\beta = 1$*,* $D_1 \equiv 1$*.*

iii) When $0 < \beta < 1$*,* $D_\beta = M_\beta\left(1\right)$ *with* $M_\beta\left(1\right)$ *defined as in Remark 2.2.10.*

Similarly to remark 2.2.11, when $0 < \beta < 1$, the original formulation of Theorem 2.2.35 (Theorem 2.4 in [30]) uses $a\left(n\right)$ instead of $u\left(n\right)$ and has that $D_\beta^{-\frac{1}{\beta}}$ is a stable distribution with Laplace transform

$$\mathbb{E}\left[\exp\left(-tD_\beta^{-\frac{1}{\beta}}\right)\right] = \exp\left(-\frac{t^\beta}{\Gamma\left(1+\beta\right)}\right). \tag{2.41}$$

By the same argument used in remark 2.2.11 (considering $G_\beta = D_\beta^{-\frac{1}{\beta}}$), we have that the distribution of $D_\beta^{-\frac{1}{\beta}}$ equals the distribution of $\Gamma\left(1+\beta\right) M_\beta\left(1\right)$.

With the objective of developing a non-parametric estimation theory in null-recurrent scenario, the concept of β-null recurrent Markov chain was introduced in [67].

Definition 2.2.19 (β-null recurrent Markov Chain)**.** *Markov chain $\boldsymbol{X}$ such that is ψ-irreducible and there exists a small function h, an initial measure λ, a constant $\beta \in (0,1)$ and a slowly varying function L_h such that*

$$\mathbb{E}_\lambda\left[\sum_{t=0}^{n} h\left(X_t\right)\right] \sim \frac{1}{\Gamma\left(1+\beta\right)} n^\beta L_h\left(n\right). \tag{2.42}$$

as n goes to $+\infty$.

Remark 2.2.12. *A condition similar to (2.42) was introduced as Hypothesis (C) in page 147 of [113], with the objective of finding a law of the iterated logarithm for additive functionals over $\boldsymbol{X}$ which are close to square integrable martingales with respect to the invariant measure of the chain. The main difference between both conditions is that in [113], the function h is not required to be a small function.*

The most widely known example of β-null recurrent processes are the random walks defined in Example 2.2.3. In effect, if the distribution Z is continuous, centered, and has a finite variance, then the random walk is $\frac{1}{2}$-null recurrent (see [64] and pp.8 of [88] for the form the slowly varying function).

Lemma 3.1 in [67] shows that, if $\mathbf{X}$ satisfies the minorization condition $M\left(1,s,\nu\right)$, and for some β, h, λ condition (2.42) is fulfilled, then it is fulfilled for every small function, which implies that β does not depend on the small function nor on the initial measure, and therefore is a global parameter of the chain $\mathbf{X}$. In Theorem 2.2.36 we remove the $M\left(1,s,\nu\right)$ assumption.

Theorem 2.2.36. *Assume* $\boldsymbol{X}$ *is* β*-null recurrent, and let* π *be a fixed invariant measure. Then we can find a slowly varying function* L *such that condition* (2.42) *holds with* $L_h = \pi(h) L$ *for every small function* h*.*

Before proving the theorem, we need a few preliminary results that will be useful in the sequel.

For a measure $\lambda \in \mathscr{M}(\mathcal{E})_+$ and a measurable function h define,

$$g(h, \lambda, x) = \lambda G^{(\lfloor x \rfloor)}(h) = \mathbb{E}_\lambda \left[\sum_{t=0}^{\lfloor x \rfloor} h(X_t) \right]. \tag{2.43}$$

For $\varepsilon \in (0,1)$, let $\left\{X_{\gamma(n)}\right\}_{n \geqslant 0}$ be the P_ε-chain defined in example 2.2.1, with initial measure λ and kernel P_ε. Define $g_\varepsilon(h, \lambda, x)$ as in (2.43) but for the P_ε-chain.

Lemma 2.2.1. *For any measurable function* h *and any initial measure* $\lambda \in \mathscr{M}(\mathcal{E})_+$*, when* x *goes to* $+\infty$*:*

$$g_\varepsilon(h, \lambda, x) \sim (1-\varepsilon)\, g\left(h, \lambda, \frac{x}{1-\varepsilon}\right) \quad \forall \varepsilon \in (0,1). \tag{2.44}$$

And, if X *is* β*-null recurrent, then*

$$g_\varepsilon(h, \lambda, x) \sim g(h, \lambda, x)(1-\varepsilon)^{\beta-1} \quad \forall \varepsilon \in (0,1). \tag{2.45}$$

Proof. Let $\mathbf{X}^\varepsilon$ be the P_ε-chain described in Example 2.2.1. Let $n = \lfloor x \rfloor$ and $\mathcal{F}_n^\varepsilon = \sigma\left(\{\alpha_k\}_{k=1}^{\gamma(n)}, \{\gamma(j)\}_{j=0}^n\right)$ Notice that,

$$\sum_{k=1}^{n} h\left(X_{\gamma(k)}\right) = \sum_{k=1}^{\gamma(n)} \alpha_k h(X_k). \tag{2.46}$$

The expectation of the left hand side of (2.46) is $g_\varepsilon(h, \lambda, n)$. For the right-hand

side, we have

$$\mathbb{E}_\lambda\left[\sum_{k=1}^{\gamma(n)}\alpha_k h\left(X_k\right)\right]=\mathbb{E}_\lambda\left[\mathbb{E}_\lambda\left(\sum_{k=1}^{\gamma(n)}\alpha_k h\left(X_k\right)\middle|\mathcal{F}_n^\varepsilon\right)\right]=\mathbb{E}_\lambda\left[\sum_{k=1}^{\gamma(n)}\alpha_k \mathbb{E}_\lambda h\left(X_k\right)\right],$$

Using the fact that $\gamma(n)$ is a stopping time for the sequence $\{\alpha_k\}$ and Wald's equality[52] , we get

$$\mathbb{E}_\lambda\left[\sum_{k=1}^{\gamma(n)}\alpha_k \mathbb{E}_\lambda h\left(X_k\right)\right]=(1-\varepsilon)\,\mathbb{E}_\lambda\left[\mathbb{E}_\lambda\left(\sum_{k=1}^{\gamma(n)}h\left(X_k\right)\right)\right]=(1-\varepsilon)\,\mathbb{E}_\lambda g\left(h,\lambda,\gamma\left(n\right)\right).$$

By the Law of Large Numbers, $\frac{\gamma(n)}{n}$ converges to $(1-\varepsilon)^{-1}$ almost surely and in every L-norm, then, $\mathbb{E}_\lambda g\left(h,\lambda,\gamma\left(n\right)\right)\sim\mathbb{E}_\lambda g\left(h,\lambda,\frac{n}{1-\varepsilon}\right)$, which completes the proof of (2.44).

When $\mathbf{X}$ is β-null recurrent, we have

$$g\left(h,\lambda,\frac{n}{1-\varepsilon}\right)\sim\frac{n^\beta}{(1-\varepsilon)^\beta}\frac{L\left(\frac{n}{1-\varepsilon}\right)}{\Gamma\left(1+\beta\right)}\sim\frac{g\left(h,\lambda,n\right)}{(1-\varepsilon)^\beta}.$$

where the last equivalence holds due to the slow variation of L at infinity. Equation (2.45) now follows immediately. □

Corollary 2.2.1. *A Harris recurrent Markov chain is β-null recurrent if and only if all of its P_ε-chains are β-null recurrent.*

Proof. If $\mathbf{X}$ is β-null recurrent, there is a small function h that satisfies (2.42) for some β. By Theorem 2.1.8 the function h is small for all the P_ε-resolvents, and by (2.45) it satisfies condition (2.42) for the P_ε-chains.

For the converse, suppose that for some ε the P_ε-chain is β-null recurrent. Let h be a small function for the original chain $\mathbf{X}$, by Theorem 2.1.8, h is also small

[52]Theorem 14.14.3 in [55].

for the P_ε-chain, hence, $g_\varepsilon\left(\lambda, h, \frac{n}{1-\varepsilon}\right) \sim \frac{(1-\varepsilon)^\beta n^\beta L_h(n)}{\Gamma(1+\beta)}$ and the result follows from (2.44). □

For the proof of Theorem 2.2.36 we will follow a similar approach as the one used to prove 2.2.25, with the detail that we only need to take care of Step 3, because Steps 1 and 2 were handled by Lemma 3.1 in [67].

Proof of Theorem 2.2.36. Take $\varepsilon \in (0,1)$ fixed. Let h and λ satisfy condition (2.42) for $\mathbf{X}$. By Lemma 2.2.1, h and λ satisfy the condition (2.42) for the P_ε-chain, therefore, the P_ε-chain is β-null recurrent.

By Theorem 2.2.8, the P_ε-chain satisfies the minorization condition $M(1, s, \nu)$ for some (s, ν). Let π_s be the measure defined by (2.33). This measure is invariant for both $\mathbf{X}$ and the P_ε-chain by Theorem 2.1.10. These conditions allow us to apply Lemma 3.1 in [67], obtaining that there exists a slowly varying function L_s such that, for any small function f, condition (2.42) is satisfied with $L_f = \pi_s(f) L_s$, therefore, $g(h, \lambda, n) \sim \frac{n^\beta \pi_s(h) L_s(n)}{\Gamma(1+\beta)}$. By the uniqueness of the invariant measure, there exists a constant K_s such that $\pi_s = K_s \pi$, hence $g(h, \lambda, n) \sim \frac{n^\beta \pi(h) L(n)}{\Gamma(1+\beta)}$ where $L(n) = K_s L_s(n)$. □

The next result shows under null recurrence, there is no distinction between β-regular and β-null recurrent

Theorem 2.2.37. *If $\mathbf{X}$ is null recurrent, the concepts of β-regular and β-null recurrent coincide.*

Proof. In effect, notice that

$$a(t) = \frac{\lambda G^{(\lfloor t \rfloor)}(\mathbb{I}_D)}{\pi(D)} \quad \text{and} \quad g(h, \lambda, t) = \mathbb{E}_\lambda\left[\sum_{k=0}^{n} h(X_k)\right] = \lambda G^{(\lfloor t \rfloor)}(h) \tag{2.47}$$

Then, by (2.35), both functions are of the same order. □

Theorem 3.1 in [67] characterizes β-null recurrent chains that satisfy the minorization condition $M(1, s, \nu)$ in terms of the tail of the time of return to the pseudo atom. Here, we present a small extension of that result, where we characterize β-null recurrent chains in terms of the time of return to the pseudo-atom of the P_ε-chains. This result is a direct consequence of Corollary 2.2.1 and the aforementioned Theorem 3.1 in [67].

Theorem 2.2.38. *Assume $\boldsymbol{X}$ is Harris recurrent and $\check{\boldsymbol{\alpha}}$ is an atom for the split chain of the P_ε-chain. Then, $\boldsymbol{X}$ is β-null recurrent if and only if*

$$\mathbb{P}_{\check{\boldsymbol{\alpha}}}(\tau_{\check{\boldsymbol{\alpha}}} > n) = \frac{1}{\Gamma(1-\beta)\, n^{\beta} L(n)} (1 + o(1)). \tag{2.48}$$

where L is as in Theorem 2.2.36.

Moreover, if $\boldsymbol{X}$ is β-null recurrent, then $\beta = \sup\{p \geqslant 0 : \mathbb{E}_{\check{\boldsymbol{\alpha}}}[\tau_{\check{\boldsymbol{\alpha}}}^p] < \infty\}$.

When the chain is atomic, (2.48) implies that is equivalent to the condition (2.28) that we had to impose to null recurrent chains in order to get a Central Limit Theorem. Incidentally, Theorem 2.2.38 also implies that the atomic $\mathbf{X}$ is β-null recurrent if and only if $\tau_{\boldsymbol{\alpha}}$ belongs to the domain of attraction of a stable law with index β[53].

Remark 2.2.13. *A strengthened version of (2.48) was used in [32] to obtain a strong invariance result in the null-recurrent case.*

Using this characterization of β-null recurrence and the construction outlined in Example 2.2.7, it is possible to create β-null recurrent processes for any value of β in the range of $(0, 1)$ as well as null recurrent processes that are not β-null recurrent. Aside from these constructions, there are many examples of β-null recurrent Markov chains in the literature. Bellow we show a couple of examples.

[53] See pp.3 of [28]

Example 2.2.10 (Bessel random walks). *A Bessel random walk is a stochastic process defined on $\mathbb{Z}_+ = \{0, 1, 2, \ldots\}$, reflecting at 0, with steps ± 1 and transition probabilities of the form*

$$\begin{aligned}
\mathbb{P}\left(X_{n+1} = x+1 \,|X_n = x\right) &= p_x = \frac{1}{2}\left(1 - \frac{\delta}{2x} + h\left(x\right)\right) \quad x \geqslant 1, \\
\mathbb{P}\left(X_{n+1} = x-1 \,|X_n = x\right) &= 1 - p_x \quad x \geqslant 1, \\
\mathbb{P}\left(X_{n+1} = 1 \,|X_n = 0\right) &= 1,
\end{aligned}$$

where $h\left(x\right) = o\left(\frac{1}{x}\right)$ as $x \to +\infty$.

The parameter δ is named the drift parameter. A Bessel random walk is recurrent if $\delta > -1$, positive recurrent if $\delta > 1$ and transient if $\delta = -1$; for $\delta = 1$ recurrence of transience depends on the function $h(x)$. In the null recurrent case, the chain is β-regular with $\beta = \frac{1+\delta}{2}$ and $P\left(\tau_0 \geqslant n\right) \sim n^{-\frac{1+\delta}{2}} L^*\left(n\right)$ where L^* is a slowly varying function (see Theorem 2.1 in [3]). For $\delta = 0$ and $h \equiv 0$ this process corresponds with a reflected random walk with $p = \frac{1}{2}$.

Bessel random walks are widely used in statistical physics, see for example [3], [36] and the references therein.

Example 2.2.11 (TAR model). *Let*

$$X_n = \alpha_1 X_{n-1} \mathbb{I}\left\{X_{n-1} \in S\right\} + X_{n-1} \mathbb{I}\left\{X_{n-1} \in S^C\right\} + x_n, \quad X_0 = 0$$

where α_1 is a real constant, S is a compact set in $\mathbb{R}$, S^C is its complement and x_n is an i.i.d sequence of random variables such that $\mathbb{E}x_n = 0$, $\mathbb{E}x_n^4 < \infty$, its distribution function is absolutely continuous with respect to the Lebesgue measure with density function f_0 such that $\inf_{x \in \mathbb{C}_0} f_0\left(x\right) > 0$ for all compact sets $\mathbb{C}_0$.

In Section 4.5 of [49], this model is used to study the relationship between the

logarithm of the British pound/American dollar real exchange rate and the Consumer Price Index. In that same paper, it was proven that the index of this model is $\frac{1}{2}$. This shows that having index $\frac{1}{2}$ does not characterize random walks.

Assume for the moment that $\mathbf{X}$ satisfies the minorization condition $M(1, s, \nu)$ and take $\check{\boldsymbol{\alpha}}$ as an atom in split chain $\check{\mathbf{X}}$. Define $T(n) = T_n(\check{\boldsymbol{\alpha}}) - 1$ as in (2.25) and let $u(n) = n^{\beta} L_s(n)$ where L_s is as in (2.48) and let $v(n)$ be its inverse.

Define the following stochastic processes:

$$T_n(t) = \frac{T(\lfloor nt \rfloor)}{u(n)} \quad , \quad C_n(t) = \frac{1}{v(n)} \sum_{k=1}^{\lfloor nt \rfloor} \Big(\tau_{\check{\boldsymbol{\alpha}}}(k) - \tau_{\check{\boldsymbol{\alpha}}}(k-1) \Big). \tag{2.49}$$

Consider the space of càdlàg functions defined on the interval $[0, +\infty)$, denoted by $\mathscr{D}_{[0,+\infty)}$. This space consists of the real functions that are right-continuous with left limits and defined over $[0, +\infty)$. More precisely, a function $g \in \mathscr{D}_{[0,+\infty)}$ if and only if g is right-continuous, has left limits at all points $t > 0$, and $\lim_{t \downarrow 0} g(t) = g(0)$. The space $\mathscr{D}_{[0,+\infty)}$ is equipped with the Skorokhod[54] topology, making it a completely separable metric space. We will use $\xrightarrow{\mathscr{D}_{[0,+\infty)}}$ to denote weak convergence in this space, and $\xrightarrow{\text{fd}}$ for convergence of finite-dimensional laws.

It is proven in [68] that $C_n \xrightarrow{\mathscr{D}_{[0,+\infty)}} S_\beta$ where S_β is the one-sided stable Levy process[55] defined by the marginal characteristics

$$\mathbb{E}\left[\exp\left(isC_\beta(t)\right)\right] = \exp\left(is^\beta t\right) \quad s \in \mathbb{R}, t \in [0, +\infty].$$

The Mittag-Leffler process with parameter β is defined as the inverse of S_β. It

[54] See Chapter 6 of [60] or Chapter 3 in [20] for more details about this space.

[55] A Levy process is a stochastic process with stationary, independent increments and sample paths in $\mathscr{D}_{[0,+\infty)}$.

is an increasing continuous stochastic process defined as

$$M_\beta(t) = t^\beta M_\beta(1) \quad , \quad \mathbb{E}\left(M_\beta^m(1)\right) = \frac{m!}{\Gamma(1+m\beta)} \quad m \geqslant 0.$$

Theorem 2.2.39. [56] *Assume* $\boldsymbol{X}$ *is a* β*-null recurrent Markov chain that satisfies the minorization condition* $M(1, s, \nu)$ *and let* λ *be any initial measure. Then*

$$T_n \xrightarrow{\mathscr{D}_{[0,+\infty)}} M_\beta$$

Moreover, if C *is a small set, then, the process* $T_{C,n}(t) = \frac{T_{\lfloor nt \rfloor}(C)}{\pi_s(C)u(n)}$ *also converges weakly to* M_β *in* $\mathscr{D}_{[0,+\infty)}$.

Using Theorem 2.2.39 we are able to show the following functional generalization of Theorem 2.2.34

Theorem 2.2.40. *Under the same assumptions of Theorem 2.2.39, if* $f \in L^1(E, \pi)$ *and* $\pi(f) \neq 0$*, then*

$$S_{n,f} \xrightarrow{\mathscr{D}_{[0,+\infty)}} M_\beta, \tag{2.50}$$

where $S_{n,f}(t) = \frac{S_{\lfloor nt \rfloor}(f)}{\pi(f)u(n)}$.

Proof. [57] Let C be a fixed small set. Without loss of generality, let us assume $\pi(C) = 1$. By Theorem 2.2.39, $T_{C,n}$ converges weakly to the process M_β, therefore, if we show that $D_n = S_{f,n} - T_{C,n}$ converges to the 0 process, then (2.50) will follow by Lemma 3.31 in [60] and the identity $S_{f,n} = (S_{f,n} - T_{C,n}) + T_{C,n}$.

By Lemma 3.30 in [60], we just need to show that

$$\sup_{t \leqslant M} |D_n(t)| = o_p(1) \quad \forall M > 0. \tag{2.51}$$

[56] Theorem 3.2 and Lemma 3.6 in [67]

[57] This proof follows the line of the proof of Lemma 3.6 in [67].

Take $0 < \delta < 1$ and define $n_\delta = n^{-(1-\delta)}$, then

$$\begin{aligned}\sup_{t\leqslant n_\delta}|D_n(t)| &\leqslant \frac{1}{u(n)}\sup_{t\leqslant n_\delta}\left|\frac{S_{\lfloor nt\rfloor}(f)}{\pi(f)} - T_C\lfloor nt\rfloor\right| \\ &\leqslant \frac{1}{u(n)}\left(\left|\frac{S_{n^\delta}(f)}{\pi(f)}\right| + T_C\left(n^\delta\right)\right) \\ &\leqslant \frac{S_{\lfloor n^\delta\rfloor}(|f|)}{u(n)\,|\pi(f)|} + \frac{T_C\left(n^\delta\right)}{u(n)}.\end{aligned}$$

Notice that,

$$\frac{S_{\lfloor n^\delta\rfloor}(|f|)}{u(n)} = \frac{S_{\lfloor n^\delta\rfloor}(|f|)}{u\left(n^\delta\right)}\frac{u\left(n^\delta\right)}{u(n)} = \frac{S_{\lfloor n^\delta\rfloor}(|f|)}{u\left(n^\delta\right)}\frac{L\left(n^\delta\right)}{L(n)\,n^{\beta(1-\delta)}}.$$

By Theorem 2.2.34, $\frac{S_{\lfloor n^\delta\rfloor}(|f|)}{u(n^\delta)}$ is bounded in probability. On the other hand, $\frac{L\left(n^\delta\right)}{L(n)}$ is a slowly varying function, therefore $\frac{L\left(n^\delta\right)}{L(n)n^{\beta(1-\delta)}} \to 0$ by Proposition 1.3.6-v in [21]. Hence, $\frac{S_{\lfloor n^\delta\rfloor}(|f|)}{u(n)} = o_p(1)$. The same argument proves that $\frac{T_C\left(n^\delta\right)}{u(n)} = o_p(1)$. Then, $\sup_{t\leqslant n_\delta}|D_n(t)| = o_p(1)$.

For $n_\delta \leqslant t \leqslant M$ we have

$$\begin{aligned}\sup_{n_\delta\leqslant t\leqslant M}|D_n(t)| &\leqslant \sup_{n_\delta\leqslant t\leqslant M}\frac{T_C(\lfloor nt\rfloor)}{u(n)}\left|\frac{S_{\lfloor nt\rfloor}(f)}{\pi(f)\,T_C(\lfloor nt\rfloor)} - 1\right| \\ &\leqslant T_{C,n}(M)\sup_{n_\delta\leqslant t\leqslant M}\left|\frac{S_{\lfloor nt\rfloor}(f)}{\pi(f)\,T_C(\lfloor nt\rfloor)} - 1\right| \\ &\leqslant T_{C,n}(M)\sup_{m\geqslant n^\delta}\left|\frac{S_m(f)}{\pi(f)\,T_C(m)} - 1\right|\end{aligned}$$

which is $o_p(1)$ thanks to the boundness in probability of $T_{C,n}(M)$ (by Theorem 2.2.39) and the almost sure convergence of $\frac{S_m(f)}{T_C(m)}$ to $\pi(f)$ (consequence of (2.34)).

□

The following result is a functional generalization of Theorem 2.2.35. It appears as Lemma A.2 in [49]

Theorem 2.2.41. *Suppose $\boldsymbol{X}$ is a β-null recurrent Markov chain that satisfies the minorization condition $M(1, s, \nu)$ and f is such such that $\mathbb{E}|f(\mathcal{B}_1) - \pi(f)|^{2m}$ is finite for some $m > 1$, then*

$$\left(\frac{S_{\lfloor nt \rfloor}(f) - \pi(f) T(\lfloor nt \rfloor)}{\sqrt{u(n)}\sigma_f}, T_n \right) \xrightarrow{\mathscr{D}_{[0,+\infty)}} (B \circ M_\beta, M_\beta),$$

where B is a standard Brownian motion independent of M_β and $\sigma_f = Var(f(\mathcal{B}_1))$.

Remark 2.2.14. *When $\pi(f) = 0$ and $t = 1$ the limit of the first component in Theorem 2.2.41 is $B(M_\beta(1))$. Because for each s, $\frac{B(s)}{\sqrt{s}}$ has a standard normal distribution and B and M_β are independent, then $U = \frac{B(M_\beta(1))}{\sqrt{M_\beta(1)}}$ also has standard normal distribution and is independent of $M_\beta(1)$[58]. Hence, we can write $B(M_\beta(1))$ as $\sqrt{M_\beta(1)}U$ which coincides with the form of the limit in Theorem 2.2.35.*

Under random normalization, a Central Limit Theorem was proved in Lemma A.3 of [49].

Theorem 2.2.42 (Central Limit Theorem)**.** *Under the same hypothesis of Theorem 2.2.41. For any small set C,*

$$\frac{\sqrt{T_n(C)\pi(C)}}{\sigma_f} \left(\frac{S_n(f)}{T_n(C)} - \frac{\pi(f)}{\pi(C)} \right) \xrightarrow{d} N(0, 1).$$

[58] See Theorem A.1 in [67]

Chapter 3

Tail Index Estimation for Discrete Heavy-Tailed Distributions

The content of this chapter is based in [17]. It is the result of a collaboration with Patrice Bertail[1] and Stephan Clemençon[2].

Abstract: It is the purpose of this paper to investigate the issue of estimating the regularity index $\beta > 0$ of a discrete heavy-tailed r.v. Z, namely a random value. Z valued in $\mathbb{N}^*$ such that $\mathbb{P}\{Z > n\} = L(n) \cdot n^{-\beta}$ for all $n \geqslant 1$, where $L : \mathbb{R}_+^* \to \mathbb{R}_+$ is a slowly varying function. Such discrete probability laws, referred to as generalized Zipf's laws sometimes, are commonly used to model rank-size distributions after a preliminary range segmentation in a wide variety of areas, ranging from quantitative linguistics to social sciences through information theory. As a first go, we consider the situation where inference is based on independent copies $Z_1, \ldots, Z_n$ of the generic r.v. Z. Just like the popular Hill estimator in the continuous heavy-tail situation, the estimator $\widehat{\beta}$ we propose can be derived

[1]MODAL'X, UMR 9023 - Université Paris Nanterre, CNRS, UPL, 200 ave de la République, F92000 Nanterre.

[2]LTCI, Telecom Paris, Institut Polytechnique de Paris.

by means of a suitable reformulation of the regularly varying condition, replacing Z's survivor function with its empirical counterpart. Under mild assumptions, a nonasymptotic bound for the deviation between $\widehat{\beta}$ and β is established, as well as limit results (consistency and asymptotic normality). Beyond the i.i.d. case, the inference method proposed is extended to the estimation of the regularity index of an atomic β-null recurrent Markov chain. Since the parameter β can be then viewed as the tail index of the (regularly varying) distribution of the return time of the chain $\mathbf{X}$ to any atom, the estimator is constructed from the successive regeneration times. We prove that in this case, the consistency of the estimator promoted is preserved. In addition to the theoretical analysis carried out, simulation results provide empirical evidence of the relevance of the inference technique proposed.

3.1 Introduction

This article is devoted to the study of the problem of estimating the regularity index $\beta > 0$ of a generalized discrete Pareto distribution, namely the probability distribution of a random variable S defined on a probability space $(\Omega,\ \mathcal{F},\ \mathbb{P})$, taking its values in $\mathbb{N}^*$ and such that:

$$\mathbb{P}\left(S > n\right) = n^{-\beta} L\left(n\right) \text{ for all } n \geqslant 1, \tag{3.1}$$

where $L : \mathbb{R}_+ \to \mathbb{R}$ is a slowly varying function, *i.e.* such that $L(\lambda z)/L(z) \to +1$ as $z \to +\infty$ for any $\lambda > 0$, see [21]. Such discrete power law probability distributions also referred to as generalized Zipf's laws are often used to model the distribution of discrete data exhibiting a specific rank-frequency relationship, namely when the logarithm of the frequency and that of the rank order are nearly pro-

portional. Such a phenomenon has been empirically observed in many ranking systems: in quantitative linguistics (*i.e.* when analyzing word frequency law in natural language, refer to *e.g.* [83]) in the first place, as well as in a very wide variety of situations, too numerous to be exhaustively listed here (see [1], [78] or [121] among many others). In this paper, we first consider the issue of estimating the parameter β involved in (3.1) (supposedly unknown, like the function L) in the classic (asymptotic) i.i.d. statistical setting, *i.e.* based on an increasing number $n \geqslant 1$ of independent copies $S_1, \ldots, S_n$ of the generic r.v. S. Statistical inference for discrete heavy-tailed distributions has not received much attention in the literature. Most of the very few dedicated methods documented either deal with very specific cases as in *e.g.* [51], [85] or [31] or else consists in applying techniques originally designed for continuous heavy-tailed distributions to the discrete data after a preliminary addition of an independent uniform noise, see *e.g.* [116]. The vast majority of the regular variation index estimators proposed in the literature, Hill's or Pickand's estimators in particular (*cf* [56], [100]), are based on order statistics, which causes obvious difficulties in the discrete case because of the possible occurrence of many ties. In contrast, the estimator under study here is based on the analysis of the probability of exponentially separated tail events. It simply rests on the fact that, as can be immediately deduced from (3.1), we have $\ln(p_k) - \ln(p_{k+1}) = \beta + \ln(L(e^k)/L(e^{k+1}))$, where $\ln(x)$ denotes the natural logarithm of any real number $x > 0$ and $p_l = \mathbb{P}(S > e^l)$ for all $l \in \mathbb{N}$, and that $L(e^{k+1})/L(e^k)$ is expected to be very close to 1 for $k \in \mathbb{N}$ chosen sufficiently large. A natural (plug-in) inference technique can be then devised by replacing the tail probabilities p_l with their empirical versions $\widehat{p}_l^{(n)} = (1/n)\sum_{i=1}^n \mathbb{I}\{S_i > e^l\}$ for $l \in \mathbb{N}$, where $\mathbb{I}\{\mathcal{E}\}$ denotes the indicator function of the event $\mathcal{E}$. This yields the

estimator

$$\widehat{\beta}_n(k) = \ln\left(\widehat{p}_k^{(n)}\right) - \ln\left(\widehat{p}_{k+1}^{(n)}\right), \tag{3.2}$$

provided that $\widehat{p}_{k+1}^{(n)} > 0$. We point out that it has exactly the same form as that proposed and analyzed in [25] in a different context, that of (continuous) *approximately Pareto distributions*[3] namely. In the discrete generalized Pareto framework, we prove that for an appropriate choice of the hyperparameter $k = k_n$ (typically chosen of order $\ln(n)$), the estimator (3.2) is strongly consistent and asymptotically normal as $n \to +\infty$. Nonasymptotic upper confidence bounds for the absolute deviations between $\widehat{\beta}_n(k)$ and β are also established here.

As explained in [28, 30, 67], for β-null recurrent Markov chains, the regularity index $\beta \in (0,1)$ controls the (sublinear) rate at which the number of visits to any given Harris set increases with observation time n, no matter the initial distribution. In the *regenerative* case (*i.e.* when the chain $\mathbf{X}$ possesses an *accessible atom*, a Harris set on which the transition probability is constant), the distribution of the regenerative time, the return time to the atom, is a discrete generalized Pareto (3.1) and the parameter β is its tail index. Due to the non-standard behavior of traditional estimators in this context, statistical inference for null-recurrent Markov chains is very poorly documented in the literature (see for instance [49, 66, 67, 88]) and, to the best of our knowledge, estimation of the key quantity β has not been considered besides the estimator described in [67, Remark 3.7], which is of limited practical use due to its slow convergence (see Section 3.3.3 for a more precise formulation of this statement). Hence, it is also the goal of this article to extend the use of the estimator (3.2) to the case where the S_i's are the successive durations

[3] The distribution of a real-valued r.v. X is said to be *approximately Pareto* with tail index $\beta > 0$ iff its survivor function is of the form: $\forall x > 0$, $\mathbb{P}(X > x) = L(x)x^{-\beta}$, where L is asymptotically constant at infinity, *i.e.* there exists $C \in (0, \infty)$ s.t. $L(x) \to C$ as $x \to +\infty$.

between the consecutive regeneration times up to time n. The main difficulty naturally arises from the fact that the number $T(n)$ of regeneration times (and thus the number of durations) is now random and the variables $S_1, \ldots, S_{T(n)}$ are not independent anymore (in particular, their sum is less than n by construction). We show that the strong consistency of the estimator is preserved. For illustration purposes, numerical experiments have been carried out, providing empirical evidence of the relevance of the estimation method promoted.

The paper is organized as follows. A thorough analysis of the behavior of the estimator (3.2) in the i.i.d. case is first carried out in section 3.2. The asymptotic results thus established are next extended in section 3.3.3 to the regenerative β-null recurrent Markovian setup, when the estimator is computed based on a single finite-length trajectory of the atomic chain. Illustrative numerical results are presented in section 3.4, while technical proofs are deferred to Section 3.5.

3.2 Tail Index Estimation - The Discrete Heavy-Tailed i.i.d. Case

Throughout this section, $S_1, \ldots, S_n$ are independent copies of a generic discrete generalized Pareto r.v. S, *i.e.* a random variable S with survivor function of type 3.1, where the parameter $\beta > 0$ and the slowly varying function L are supposedly unknown. As a first go, we start to investigate the behavior of the estimator (3.2) in this basic general framework and next develop the analysis in specific situations, *i.e.* for particular choices of the function L.

3.2.1 Main Results - Confidence Bounds and Limit Theorems

As explained in the Introduction section, the estimator (3.2) can be viewed as an empirical counterpart of the quantity

$$\beta(k) := \ln(p_k) - \ln(p_{k+1}) = \beta + \ln\left(\frac{L(e^k)}{L(e^{k+1})}\right), \tag{3.3}$$

see (3.1), which tends to β as $k \to \infty$ by virtue of the slow variation property of L. As previously emphasized, unless the function L is supposed to be asymptotically constant (*i.e.* there exists $C > 0$ s.t. $L(x) \to C$ as $x \to +\infty$), the discrete generalized Pareto model (3.1) is not a discrete version of the (continuous) approximately β-Pareto model considered in [25] and, consequently, the validity framework established therein does not apply here.

The proposition below provides an upper confidence bound for the absolute deviations between (3.2) and β (respectively, between (3.2) and $\beta(k)$).

Proposition 3.2.1. *Let $\delta \in (0,\ 1/2)$ and set $u_n(\delta) = \ln(2/\delta)/n$ for all $n \geqslant 1$. If $k \geqslant 1$ is such that $p_{k+1} \geqslant 16u_n(\delta)$, then, with probability at least $1 - 2\delta$, we have:*

$$\left|\widehat{\beta}_n(k) - \beta\right| \leqslant 6\sqrt{\frac{u_n(\delta)}{p_{k+1}}} + \left|\ln\left(\frac{L(e^k)}{L(e^{k+1})}\right)\right|. \tag{3.4}$$

The bound (3.4) reveals some sort of 'bias-variance' trade-off, ruled by the hyperparameter $k > 0$. The second term on the right-hand side can be viewed as the bias of the inference method, insofar as the estimator (3.2) can be seen as an empirical version of the approximant (3.3). It decays to 0 as k increases towards infinity, while the first term, whose presence is due to the random nature of the estimator, tends to $+\infty$. We point out that *second-order slow variation conditions* (see [50]) are required to bound the (vanishing) bias term in (3.4), as shall be explained

in subsection 3.2.2. The following result reveals that for an appropriate choice of $k = k_n$, the estimator (3.2) is strongly consistent.

Theorem 3.2.2 (Strong consistency). *Suppose that, as $n \to +\infty$, we have $k_n \to +\infty$ so that $(\ln n)\exp(k_n\beta)/n = o(L(\exp(k_n))$. Then, we have:*

$$\widehat{\beta}_n(k_n) \to \beta \text{ almost surely, as } n \to +\infty.$$

In particular, as stated below, strong consistency is guaranteed when k_n is of logarithmic order.

Corollary 3.2.1. *Let $0 < A < 1/\beta$. Then, we have:*

$$\widehat{\beta}_n(A\ln(n)) \to \beta \text{ almost surely, as } n \to +\infty.$$

Now the following results establish the asymptotic normality of the deviation between (3.2) and $\beta(k_n)$, when appropriately normalized.

Theorem 3.2.3 (Asymptotic normality). *Suppose that k_n satisfies the conditions of Theorem 3.2.2 and $k_n = o(n)$ as $n \to +\infty$.*

(i) Then, as $n \to +\infty$, we have the convergence in distribution:

$$\sqrt{np_{k_n}}\left(\widehat{\beta}_n(k_n) - \beta(k_n)\right) \Rightarrow \mathcal{N}\left(0,\ e^{\beta} - 1\right).$$

(ii) In addition, asymptotic normality holds true for the 'standardized' deviation:

$$\frac{\sqrt{n\widehat{p}_{k_n}^{(n)}}\left(\widehat{\beta}_n\left(k_n\right) - \beta\left(k_n\right)\right)}{\sqrt{e^{\widehat{\beta}_n(k_n)} - 1}} \Rightarrow \mathcal{N}\left(0,\ 1\right), \text{ as } n \to +\infty.$$

The asymptotic normality results above can be extended to the deviation between (3.2) and β, provided that the bias term $\beta(k_n) - \beta$ vanishes at an appropriate rate, as stated below.

Corollary 3.2.2. *Suppose that the conditions of Theorem 3.2.3 are fulfilled. In addition, assume that k_n is such that*

$$\sqrt{np_{k_n}} \left(1 - \frac{L\left(e^{k_n}\right)}{L\left(e^{k_n+1}\right)} \right) \to 0, \ \text{as } n \to +\infty. \tag{3.5}$$

(i) Then, we have the convergence in distribution

$$\sqrt{np_{k_n}} \left(\widehat{\beta}_n\left(k_n\right) - \beta \right) \Rightarrow \mathcal{N}(0, \ e^{\beta} - 1) \ \text{as } n \to +\infty.$$

(ii) In addition, the 'studentized' version is asymptotically normal:

$$\frac{\sqrt{n\widehat{p}_{k_n}^{(n)}} \left(\widehat{\beta}_n(k_n) - \beta \right)}{\sqrt{e^{\widehat{\beta}_n(k_n)} - 1}} \Rightarrow \mathcal{N}(0, 1) \ \text{as } n \to +\infty.$$

Of course, the condition (3.5) on k_n can be hardly checked in practice. This is a classic issue in tail estimation and in the statistical analysis of extreme values more generally. The choice of the hyperparameter k somehow rules the (asymptotic) bias-variance trade-off: the estimator (3.2) is expected to be of large variance when k is large and to have a large bias if k is too small. As depicted in Fig. 3.1, to choose k, one may use the same approach as that originally proposed for the Hill estimator (see *e.g.* [106]), which consists in plotting the values of (3.2) for a range of values of k and choosing k in a region where a certain degree of stability is exhibited.

Averaged versions. Rather than picking a single value for k, another natural approach would consist in averaging the estimators (3.2) over a range of values for

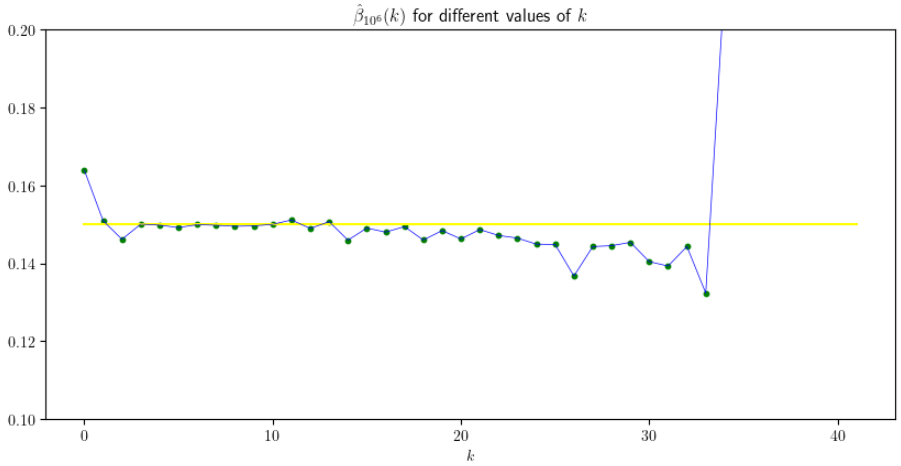

Figure 3.1: Behaviour of $\widehat{\beta}_n(k)$ for different values of k, to estimate the parameter $\beta = 0.15$ based on a dataset of 10^6 independent realizations of a Zeta distribution with parameter $\alpha = \beta - 1$ (see section 3.4 for its definition).

the hyperparameter. Let k and m such that $k > m$ and define

$$\beta(k,m) = \frac{1}{2m+1} \sum_{j=-m}^{m} \beta(k+j), \tag{3.6}$$

$$\widehat{\beta}_n(k,m) = \frac{1}{2m+1} \sum_{j=-m}^{m} \widehat{\beta}_n(k+j). \tag{3.7}$$

One may easily check that

$$\beta(k,m) = \beta + \frac{1}{2m+1} \ln\left(\frac{L\left(e^{k-m}\right)}{L\left(e^{k+m+1}\right)}\right).$$

The nonasymptotic result in Proposition 3.2.1 can be extended to (3.7), as revealed by the bound stated below, which suggests that a more favorable balance between bias and variance could be attained by means of an adequate choice of the range defined by k and m.

Proposition 3.2.4. *Let k and m such that $k > m$ and let $\delta \in (0,\ 1/(2(1+2m))$. Then, as soon as $p_{k+m+1} \geqslant 16u_n(\delta)$, we have with probability larger than $1 - 2\delta(1+$*

$2m)$:

$$\left|\widehat{\beta}_n\left(k,m\right)-\beta\right| \leqslant 6\sqrt{\frac{u_n\left(\delta\right)}{p_{k+m+1}}}+\frac{1}{2m+1}\left|\ln\left(\frac{L\left(e^{k-m}\right)}{L\left(e^{k+m+1}\right)}\right)\right|. \tag{3.8}$$

3.2.2 Refined Bias Analysis - Examples

We now consider several specific cases of distributions of type (3.1) (*i.e.* several instances of the slowly varying functions L) to explicit the asymptotic order of magnitude of the terms $1/\sqrt{np_{k+1}}$ and $|\ln(L(e^k)/L(e^{k+1}))|$ involved in the bound (3.4), when k_n is picked as in Corollary 3.2.1: $k_n = A\ln n$ with $0 < A < 1/\beta$.

• **The logarithmic case.** Suppose that $L(n) = C\ln n$, where $C > 0$. In this situation, we have $|\ln(L(e^{k_n})/L(e^{k_n+1}))| \sim 1/(A\ln n)$ as $n \to +\infty$, whereas $1/\sqrt{np_{k+1}} = O(1/\sqrt{n^{1-A\beta}\ln n})$.

• **The inversely logarithmic case.** Consider now the situation where $L(n) = C/\ln n$ with $C > 0$. Then, we still have we have $|\ln(L(e^{k_n})/L(e^{k_n+1}))| \sim 1/(A\ln n)$, while $1/\sqrt{np_{k+1}} = O(\sqrt{(\ln n)/n^{1-A\beta}})$ as $n \to +\infty$.

We point out that, in the two examples above, the conditions of Corollary 3.2.2 are not met, the bias being too big to get asymptotic normality (centered at β).

• **The asymptotically constant case.** Suppose that $L(n) = e^{C_0}(1+\varepsilon(n))$ where $C_0 > 0$ and $\varepsilon(n) \to 0$ as $n \to +\infty$. In this case, $|\ln(L(e^{k_n})/L(e^{k_n+1}))| = O(\varepsilon(n^A))$ and $1/\sqrt{np_{k+1}} = O(1/\sqrt{n^{1-A\beta}})$. Hence, if $|\varepsilon(n^A)| = O(n^{-\lambda})$ for some $\lambda > 0$, then the conditions of Corollary 3.2.2 are satisfied with $k_n = A\ln n$ such that $\max\{(1-2\lambda)/\beta,\ 0\} < A < 1/\beta$.

• **Slow variation with a remainder ($SR2$).** Consider the case where the slowly varying function satisfies the condition $SR2$ introduced in [21]: there exist two

real-valued functions k and g defined on $\mathbb{R}_+$ such that, for all $\lambda > 0$,

$$\frac{L(\lambda x)}{L(x)} - 1 \sim \kappa(\lambda) g(x), \text{ as } x \to +\infty, \tag{3.9}$$

where $\kappa(\lambda) = c\int_1^\lambda \theta^{\rho-1} d\theta$, $c > 0$ and g is regularly varying with index $\rho \leqslant 0$, *i.e.* $g(x) = x^\rho U(x)$ where U is a slowly varying function. Under the additional assumption that g has positive decrease, Corollary 3.12.3 in [21] gives the following representation:

$$L(x) = C\left(1 - c|\rho|^{-1} g(x) + o\left(g\left(x\right)\right)\right), \text{ as } x \to +\infty, \tag{3.10}$$

where C is a finite constant. The result below provides a precise control of the bias of the estimation method in this case.

Lemma 3.2.1. *Suppose that conditions* (3.9) *and* (3.10) *are fulfilled. Then, as* $n \to +\infty$, *we have:*

$$\ln\left(\frac{L\left(n^A\right)}{L\left(en^A\right)}\right) = -c|\rho|^{-1} n^{-A|\rho|}\left(U\left(n^A\right) - e^{-|\rho|} U\left(en^A\right)\right) + o\left(n^{-A|\rho|} U\left(n^A\right)\right).$$

In this situation, the bias of the method is thus of order $O(n^{-A|\rho|})$, while $1/\sqrt{np_{k+1}}$ is of order $O(n^{-(1-A\beta)/2})$. Hence, if $1/(\beta+2|\rho|) \leqslant A < 1/\beta$, satisfies the conditions of Corollary 3.2.2 are satisfied with $k_n = A \ln n$.

3.3 Regular Null-Recurrent Chains - Regularity Index Estimation

We start by setting out the notations used throughout this section and listing first the properties satisfied by the class of Markov chains under study. One may refer to [87] for an excellent account of the Markov chain theory. The concept of β-regularity for describing how fast a Harris chain returns to Harris sets is then recalled, together with related asymptotic results, invoked in the subsequent statistical analysis, for clarity's sake.

3.3.1 Harris recurrent Markov chains

Here and throughout, $\mathbf{X} = (X_n)_{n\in\mathbb{N}}$ denotes a time-homogeneous Markov chain, with state space E, equipped with a countably generated σ-field $\mathcal{E}$, and transition probability $P(x, dy)$. For any probability distribution λ on E, we denote by $\mathbb{P}_\lambda$ the probability distribution on the underlying space such that $X_0 \sim (dx)$ and by $\mathbb{E}_\lambda[.]$ the corresponding expectation. For notational convenience, we shall write $\mathbb{P}_x$ and $\mathbb{E}_x[.]$ when λ is the Dirac mass at $x \in E$.

We suppose that the chain $\mathbf{X}$ is ψ-irreducible, meaning that there exists some σ-finite measure ψ on $(E, \mathcal{E})$ such that any measurable set $B \subset E$, weighted by ψ, can be reached by the chain with positive probability in a finite number of steps, *i.e.* $\sum_{n\geqslant 1} P^n(x, B) > 0$, no matter the starting point $x \in E$, denoting by $P^n(x, dy)$ the n-th iterate of the transition probability $P(x, dy)$. An irreducibility measure is said to be *maximal* if it dominates any other irreducibility measure. A measurable set $A \in \mathcal{E}$ is called *accessible* if $\psi(A) > 0$. An irreducible chain possesses an accessible atom, and hence is called *atomic*, if there is an accessible set $\boldsymbol{\alpha} \in \mathcal{E}$ such that for

all x, y in $\boldsymbol{\alpha}$: $P(x, \bullet) = P(y, \bullet)$.

If $\mathbf{X}$ is ψ-irreducible, there is $d' \in \mathbb{N}^*$ and disjoints sets $D_1, \ldots, D_{d'}$ $D_{d'+1} = D_1$ weighted by ψ such that $\psi(E \setminus \cup_{1 \leqslant i \leqslant d'} D_i) = 0$ and $\forall x \in D_i, P(x, D_{i+1}) = 1$. The g.c.d. d of such integers is called the *period* of the chain. $\mathbf{X}$ is said to be *aperiodic* if $d = 1$.

For any set $A \in \mathcal{E}$ denote by $T(A)$ the total number of visits of $\mathbf{X}$ to A. Similarly, denote by $T_n(A)$ the number of visits $\mathbf{X}$ to A up to time n, i.e.

$$T_n(A) = \sum_{i=1}^{n} \mathbb{I}\{X_i \in A\}, \tag{3.11}$$

this sequence is called the *occupation time sequence*.

An irreducible Markov chain is Harris recurrent if it visits any accessible set infinitely often with probability one, no matter the starting point, i.e.

$$\mathbb{P}_x(T(A) = \infty) = 1 \quad \forall x \in E.$$

Every Harris recurrent chain admits a nonzero invariant measure, that is, a measure π such that $\int_{x \in E} \pi(dx) P(x, A) = \pi(A)$ for all $A \in \mathcal{E}$. This measure is unique up to a multiplicative constant and it is also a maximal irreducibility measure for $\mathbf{X}$. Measurable sets weighted by π are said to be Harris. When the measure $\pi(dx)$ is finite, the chain is said to be *positive recurrent* otherwise, is called *null recurrent*.

For Harris recurrent chains, the following strong ratio limit theorem holds: as $n \to \infty$,

$$\frac{\sum_{i=1}^{n} \mathbb{I}\{X_i \in A\}}{\sum_{i=1}^{n} \mathbb{I}\{X_i \in B\}} \to \frac{\pi(A)}{\pi(B)} \ \mathbb{P}_\lambda\text{-almost-surely}, \tag{3.12}$$

for any initial distribution λ and any measurable sets A and B such that $\pi(B) > 0$.

3.3.2 β-null recurrent Markov Chains

For a wide class of Harris Markov chains, the *regularity index* describes how fast the *occupation time* related to a Harris set C (*i.e.* the number of visits to C) increases with time n. When $\mathbf{X}$ is positive recurrent, it follows from the Strong Law of Large Numbers for Markov chains that occupation times of Harris sets (3.11) grow in a linear fashion with the observation time: as $n \to \infty$,

$$T_n(C) \sim \pi(C)n \;\; \mathbb{P}_\nu\text{-almost surely.}$$

Dealing with null recurrent chains is considerably more challenging, given that a comprehensive theory of non-parametric estimation for this type of chain does not exist. To tackle this problem, Karlsen and Tjøstheim developed in [67] the concept of a β-null recurrent Markov chain (called β-regular Markov chains in [28]), which establishes a regularity condition that makes the issue more manageable. Before formulating it, we need a few definitions.

Denote by $\mathcal{E}^+$ the class of nonnegative measurable functions with positive ψ support. A function $s \in \mathcal{E}^+$ is called *small* if there exists an integer $m_0 \geqslant 1$ and a measure $\nu \in \mathscr{M}(\mathcal{E})_+$ such that

$$P^{m_0}(x, A) \geqslant s(x)\,\nu(A) \quad \forall x \in E, A \in \mathcal{E}. \tag{3.13}$$

When a chain possesses a small function s, we say that it satisfies the *minorization inequality* $M(m_0, s, \nu)$. As pointed out in [90], there is no loss of generality in assuming that $0 \leqslant s(x) \leqslant 1$ and $\int_E s(x)d\nu(x) > 0$.

A set $A \in \mathcal{E}$ is said to be *small* if the function $\mathbb{I}_A$ is small. Similarly, a measure ν is *small* if there exist m_0, and s that satisfy (3.13). By Theorem 2.1 in [90], every

irreducible Markov chain possesses a small function and Proposition 2.6 of the same book shows that every accessible set contains a small set. Moreover, if π is an invariant measure and A is a small set, then $0 < \pi(A) < +\infty$ [90, Proposition 5.6, pp. 72].

We will say that an irreducible and Harris recurrent Markov chain $\mathbf{X}$ is *β-null recurrent* (or *β-regular*) if there exists a small function h, an initial measure λ, a constant $\beta \in (0,1)$ and a slowly varying function L_h such that

$$\mathbb{E}_\lambda \left[\sum_{t=0}^{n} h(X_t) \right] \sim \frac{1}{\Gamma(1+\beta)} n^\beta L_h(n). \tag{3.14}$$

as n goes to $+\infty$. Here $\Gamma(z) = \int_{t \geqslant 0} t^{z-1} e^{-t} dt$ denotes the Gamma function.

Lemma 3.1 in [67] shows that the parameter β is a global parameter of the chain, following [28], we will call this parameter the *regularity index* of the chain. Furthermore, the lemma shows that the function L_h in (3.14) is essentially unique up to a multiplicative constant. This means that if a small function h satisfies (3.14), then for any small function f, there exists a constant $K_{f,h}$ such that f satisfies (3.14) with $L_f = K_{f,h} L_h$.

When particularized to functions of the form $\mathbb{I}_A$ where A is a small set, (3.14) shows that under β-null recurrence, the expectation of the growth of the occupation time of small sets is sublinear

$$\mathbb{E}_\lambda T_n(C) \sim \frac{1}{\Gamma(1+\beta)} n^\beta L_C(n).$$

The following theorem, which is the particularization of Theorem 3.2 and Lemma 3.6 of [67] to the atomic case, shows that when $\mathbf{X}$ is atomic, β-null recurrence is characterized by the fact that the time of the first return to the atom has

a generalized Pareto distribution. The theorem also provides the limit distribution for the occupation time of small sets.

Theorem 3.3.1. *Assume $\boldsymbol{X}$ is a β-null recurrent atomic Markov chain with accessible atom $\boldsymbol{\alpha}$. Let $\tau_{\boldsymbol{\alpha}} = \min\{n \geqslant 1 : X_n \in \boldsymbol{\alpha}\}$. Then,*

i) There exists a slowly varying function $L_{\boldsymbol{\alpha}}$ such that

$$\mathbb{P}_{\boldsymbol{\alpha}}\left(\tau_{\boldsymbol{\alpha}} > n\right) \sim \frac{1}{\Gamma(1-\beta) n^{\beta} L_{\boldsymbol{\alpha}}(n)}. \tag{3.15}$$

ii) For any small set C, there is a constant $K_{\boldsymbol{\alpha},C} > 0$ such that

$$\frac{T_n(C)}{n^{\beta} L_{\boldsymbol{\alpha}}(n) K_{\boldsymbol{\alpha},C}} \xrightarrow{d} M_{\beta}(1) \tag{3.16}$$

where $M_{\beta}(1)$ is a Mittag-Leffler[4] distribution of index β.

The class of β-null recurrent Markov chains contains many stochastic processes widely used in probabilistic modeling.

Bessel random walks

A Bessel random walk is a stochastic process defined on $\mathbb{Z}_+ = \{0, 1, 2, \ldots\}$, reflecting at 0, with steps ± 1 and transition probabilities of the form

$$P\left(X_{n+1} = x+1 \,|X_n = x\right) = p_x = \frac{1}{2}\left(1 - \frac{\delta}{2x} + h(x)\right) \quad x \geqslant 1,$$

$$P\left(X_{n+1} = x-1 \,|X_n = x\right) = 1 - p_x \quad x \geqslant 1,$$

[4]The Mittag-Leffler distribution with index β is a non-negative continuous distribution, whose moments are given by

$$\mathbb{E}\left(M_{\beta}^{m}(1)\right) = \frac{m!}{\Gamma(1+m\beta)} \quad m \geqslant 0.$$

See (3.39) in [67] for more details.

$$P\left(X_{n+1}=1\,|X_n=0\right)=1,$$

where $h\left(x\right)=o\left(\frac{1}{x}\right)$ as $x\to+\infty$.

The parameter δ is named the *drift* parameter. A Bessel random walk is recurrent if $\delta>-1$, positive recurrent if $\delta>1$ and transient if $\delta<-1$; for $\delta=-1$ recurrence of transience depends on the function $h(x)$. In the null recurrent case, the chain is β-regular with $\beta=\frac{1+\delta}{2}$ and $P\left(\tau_0\geqslant n\right)\sim n^{-\frac{1+\delta}{2}}L^*\left(n\right)$ where L^* is a slowly varying function (see Theorem 2.1 in [3]). For $\delta=0$ and $h\equiv 0$ this process corresponds with a reflected random walk with $p=\frac{1}{2}$.

TAR model

Let

$$X_n=\alpha_1 X_{n-1}\mathbb{I}\left\{X_{n-1}\in S\right\}+X_{n-1}\mathbb{I}\left\{X_{n-1}\in S^C\right\}+x_n,\quad X_0=0$$

where α_1, is a real constant, S is a compact set in $\mathbb{R}$, S^C is its complement and x_n is an i.i.d sequence of random variables such that $\mathbb{E}x_n=0$, $\mathbb{E}x_n^4<\infty$, its distribution function is absolutely continuous with respect to the Lebesgue measure with density function f_0 such that $\inf_{x\in\mathbb{C}_0}f_0\left(x\right)>0$ for all compact sets $\mathbb{C}_0$.

In Section 4.5 of [49], this model is used to study the relationship between the logarithm of the British pound/American dollar real exchange rate and the Consumer Price Index. In that same paper, it was proven that the index of this model is $\frac{1}{2}$. This shows that having index $\frac{1}{2}$ does not characterize random walks.

Null recurrent, but not β-regular

The following construction, presented in [88], allows us to construct β-null recurrent chains for a given value of β and also null recurrent chains that are not β-regular.

Let η_n be a sequence of i.i.d. random variables. Let's define X_n as

$$X_n = \begin{cases} X_{n-1} - 1, & \text{if } X_{n-1} > 1, \\ \eta_n, & \text{if } X_{n-1} \in [0,1]. \end{cases}$$

Then, X_n is an irreducible Markov Chain and the interval $[0,1]$ is an atom for the chain. Moreover, $P_x\left(\tau_{[0,1]} > n\right) = P\left(\lfloor \eta_1 \rfloor > n\right)$, therefore, X_n is null recurrent if and only if $\mathbb{E}\lfloor \eta_1 \rfloor = \infty$ and is β-null recurrent if and only if

$$P\left(\lfloor \eta_1 \rfloor > n\right) \sim n^{-\beta} L\left(n\right)$$

for some slowly varying function $L(n)$.

3.3.3 Regularity Index of a Regular Chain - Statistical Inference

The estimation of the regularity index β has not received much attention in the literature. To our knowledge, the only consistent estimator of this parameter is $\frac{\ln T_n(C)}{\ln n}$ where C is a small set. This estimator was proposed in Remark 3.7 of [67], where it was shown to be strongly consistent for β-null recurrent chains where the minorization condition $M\left(1, s, \nu\right)$ is satisfied. It was pointed out by the authors that this estimator is of limited practical use due to its slow convergence. To make this statement more clear, consider the simplest case, where $\mathbf{X}$ is atomic and $L_{\boldsymbol{\alpha}}$

is constant. By part ii of Theorem 3.3.1, we can find a constant $K'_{\boldsymbol{\alpha},C}$ such that $\frac{T_n(C)}{n^\beta}$ converges in distribution to $K'_{\boldsymbol{\alpha},C} M_\beta(1)$. Then, by the Continuous Mapping Theorem,

$$\ln n \left(\frac{\ln T_n(C)}{\ln n} - \beta \right) \xrightarrow{d} \ln \left(K'_{\boldsymbol{\alpha},C} M_\beta(1) \right),$$

which shows that the rate of convergence, in this case, is of order $1/\ln n$.

In this section, we will show that in the atomic case, we can use the estimator proposed in Section 3.2 to consistently estimate β.

Assume that we observe the first n points of an atomic β-null recurrent homogeneous Markov Chain. Let $T_n(\boldsymbol{\alpha})$ be the number of times the chain visits the atom up to time n, and denote by $\tau_1, \ldots, \tau_{T_n(\boldsymbol{\alpha})}$ the times of those visits. By the Strong Markov property, the random variables $S_i = \tau_{i+1} - \tau_i$ are i.i.d. and by (3.15) their survival function has the form of (3.1) (with $L = \frac{1}{\Gamma(1-\beta) L_{\boldsymbol{\alpha}}}$).

The recurrence of the chain implies that $T(n) = T_n(\boldsymbol{\alpha}) - 1$ converges almost surely to $+\infty$, then, by Theorem 8.1 in page 302 of [55], we can replace n by $T(n)$ on the strong consistency results we presented on section 3.2, to obtain equivalent results for the sequence $S_1 \ldots, S_{T(n)}$.

Theorem 3.3.2. *If k_n is a sequence that satisfy the hypothesis of Theorem 3.2.2, then $\widehat{\beta}_{T(n)}\left(k_{T(n)}\right)$ converges almost surely to β.*

From this result, we get that if we chose $k_n = \ln n$, then the estimator is strongly consistent and takes the form

$$\widehat{\beta}_{T(n)}(\ln T(n)) = \ln \left(\frac{\sum_{i=1}^{T(n)} \mathbb{I}\{S_i > T(n)\}}{\sum_{i=1}^{T(n)} \mathbb{I}\{S_i > eT(n)\}} \right). \tag{3.17}$$

Theorem 3.3.3. *$\widehat{\beta}_{T(n)}(\ln T(n))$ is a strongly consistent estimator of β.*

The following result is the application of Theorem 3.5.1 to the Markovian case.

Theorem 3.3.4. *Let k_n and m_n be sequences that satisfy the hypothesis of Theorem 3.5.1, then $\widehat{\beta}_{T(n)}\left(k_{T(n)}, m_{T(n)}\right) \to \beta$ almost surely.*

Lastly, if we take $k_n = \ln n$, $m_n = \left\lfloor \frac{\ln n}{l} \right\rfloor$ and replace n by $\ln T(n)$ we get that

$$\widehat{\beta}_{T(n)}\left(\ln T(n), \left\lfloor \frac{\ln T(n)}{l} \right\rfloor\right) = \frac{1}{2\left\lfloor \frac{\ln T(n)}{l} \right\rfloor + 1} \ln \left(\frac{\sum_{j=1}^{T(n)} \mathbb{I}\left\{ S_i > e^{-\left\lfloor \frac{\ln T(n)}{l} \right\rfloor} T(n) \right\}}{\sum_{j=1}^{T(n)} \mathbb{I}\left\{ S_i > e^{\left\lfloor \frac{\ln T(n)}{l} \right\rfloor + 1} T(n) \right\}} \right) \tag{3.18}$$

is a strongly consistent estimator of β.

Theorem 3.3.5. *If l is a positive number such that $l \geqslant \frac{\beta}{1-\beta}$, then*

$$\widehat{\beta}_{T(n)}\left(\ln T(n), \left\lfloor \frac{\ln T(n)}{l} \right\rfloor\right) \to \beta \quad a.s.$$

Remark 3.3.1. *Due to the impossibility of controlling the sequence $T_n(\boldsymbol{\alpha})$ (and hence $T(n)$) by a deterministic quantity in probability (part ii of Theorem 3.3.1), we have not been able to extend to the markovian case the asymptotic normality results of Theorem 3.2.3 and Corollary 3.2.2. Heuristically, if in Theorem 3.2.3 we take $k_n = \ln T(n)$ and replace $T(n)$ by its approximate expectation $n^{\beta} L_{\boldsymbol{\alpha}}(n)$ [67, Lemma 3.3], we would get a convergence rate of order $n^{-\beta(1-\beta)/2} L_{\boldsymbol{\alpha},\beta}(n)$, where $L_{\boldsymbol{\alpha},\beta}(n)$ is the slowly varying function given by $\sqrt{\frac{L_{\boldsymbol{\alpha}}(n^{\beta} L_{\boldsymbol{\alpha}}(n))}{L_{\boldsymbol{\alpha}}(n)^{1-\beta}}}$. This suggest a convergence rate of order $n^{-\beta(1-\beta)/2}$ when $L_{\boldsymbol{\alpha}}$ is asymptotically constant. However, we have not been able to prove this result.*

3.4 Illustrative Numerical Experiments

In order to test the accuracy of the estimators proposed in this paper, we first consider the following simulation framework: for different values of β we generate 64 samples of 10^5 points from a Zeta distribution[5] with parameter $\alpha = \beta - 1$. For each sample, we estimate β. Table 3.1 contains the mean squared error for each value of β.

For comparison, we have also estimated β using the implementation of the Hill estimator provided in [116], where they add a small uniform noise to each sample in order to improve the stability of the estimator for discrete data and use a double bootstrap method [33, 104] to find the order statistic that provides the best estimator. These changes make the Hill estimator more precise, but considerably slower, especially for small values of β.

Table 3.1: Mean squared error

β	Hill	$\beta\,(\ln n)$	$\beta\left(\ln n, \frac{\ln n}{4}\right)$
0.1	1.7×10^{7}	3.47×10^{-3}	3.04×10^{-3}
0.2	103860	1.09×10^{-3}	8.71×10^{-4}
0.35	398	3.5×10^{-4}	1.63×10^{-4}
0.4	2.81×10^{-5}	5.81×10^{-4}	1.25×10^{-4}
0.5	4.09×10^{-5}	2.18×10^{-3}	4.59×10^{-4}
0.75	3.05×10^{-5}	8.48×10^{-2}	1.84×10^{-2}

[5] A discrete random variable S follows a Zeta distribution with parameter β if

$$P\left(S=k\right)=\frac{k^{-\alpha}}{\zeta\left(\alpha\right)}$$

where ζ is the Riemann zeta function. The cumulative distribution function of a Zeta distribution satisfies (see Lemma 9.1 in [120])

$$P\left(S\geqslant k\right)\sim\frac{k^{\alpha+1}}{\zeta\left(\alpha\right)\left(\alpha+1\right)}.$$

This distribution is also known as Zipf's distribution due to its relationship with Zipf's law.

Using the same methodology, we have generated different β-null recurrent Markov Chains (each one with 10^5 points) and we have estimated the value of β.

Table 3.2: Mean squared error

Markov chain	$T(n)$	Hill	$\beta(\ln n)$	$\beta\left(\ln n, \frac{\ln n}{4}\right)$
Random walk ($\beta = 0.5$)	2187	5.2×10^{-3}	3.57×10^{-2}	1.92×10^{-2}
Bessel random walk with $\beta = 0.35$	148	1.4×10^{-2}	3.78×10^{-2}	1.73×10^{-2}
Bessel random walk with $\beta = 0.4$	833	4.19×10^{-3}	1.83×10^{-2}	1.39×10^{-2}
Bessel random walk with $\beta = 0.7$	7722	8.64×10^{-4}	4.95×10^{-2}	3.26×10^{-2}

3.5 Technical Proofs

3.5.1 Proof of Proposition 3.2.1

Lemma 3.5.1. *Let $\delta > 0$ and k such that $p_{k+1} \geqslant 16u_n(\delta)$, then*

$$\left|\widehat{\beta}_n(k) - \beta(k)\right| \leqslant 6\sqrt{\frac{u_n(\delta)}{p_{k+1}}}, \tag{3.19}$$

with probability larger than $1 - 2\delta$.

Proof. In order to prove this result, we need the following lemma, proved in the supplementary material of [25].

Lemma 3.5.2. ***Bernstein's inequality for Bernoulli random variables*** *Let $X_1, \ldots, X_n$ be i.i.d. samples from a distribution F, and we define $p_k = 1 - F(e^k)$, $\widehat{p}_k^n = \frac{1}{n}\sum_{i=1}^{n} I\left\{X_i > e^k\right\}$ and $u_n(\delta) = \frac{\ln(2/\delta)}{n}$. Let $\delta > 0$ and also let n large enough so that $p_k \geqslant 4u_n(\delta)$, then with probability $1 - \delta$,*

$$|\widehat{p}_k^n - p_k| \leqslant 2\sqrt{p_k u_n(\delta)}.$$

Because $p_k \geqslant 16u_n(\delta)$ we can apply the previous lemma, then with probability greater than $1-\delta$ we have

$$-2\sqrt{p_k u_n(\delta)} \leqslant \widehat{p}_k^n - p_k \leqslant 2\sqrt{p_k u_n(\delta)}$$
$$p_k\left(1-2\sqrt{\frac{u_n(\delta)}{p_k}}\right) \leqslant \widehat{p}_k^n \leqslant p_k\left(1+2\sqrt{\frac{u_n(\delta)}{p_k}}\right),$$

taking the log in the previous equation we get

$$\begin{aligned}\ln p_k + \ln\left(1-2\sqrt{\frac{u_n(\delta)}{p_k}}\right) \leqslant \ln \widehat{p}_k^n \leqslant \ln p_k + \ln\left(1+2\sqrt{\frac{u_n(\delta)}{p_k}}\right)\\ \ln\left(1-2\sqrt{\frac{u_n(\delta)}{p_k}}\right) \leqslant \ln \widehat{p}_k^n - \ln p_k \leqslant \ln\left(1+2\sqrt{\frac{u_n(\delta)}{p_k}}\right)\\ -3\sqrt{\frac{u_n(\delta)}{p_k}} \leqslant \ln \widehat{p}_k^n - \ln p_k \leqslant 2\sqrt{\frac{u_n(\delta)}{p_k}},\end{aligned} \tag{3.20}$$

where the last pair of inequalities is obtained by using $\ln(1+x) \leqslant x$ and $\ln(1-x) \geqslant -\frac{3x}{2}$ $(x < 1/2)$.

Inequality (3.20) implies that

$$|\ln \widehat{p}_k^n - \ln p_k| \leqslant 3\sqrt{\frac{u_n(\delta)}{p_k}} \tag{3.21}$$

with probability bigger that 1-δ.

Applying (3.21) for $k+1$ we get with probability bigger that 1-δ

$$\left|\ln \widehat{p}_{k+1}^n - \ln p_{k+1}\right| \leqslant 3\sqrt{\frac{u_n(\delta)}{p_{k+1}}}. \tag{3.22}$$

Combining the triangular inequality and the equations 3.21 and 3.22 we get

with probability bigger than $1 - 2\delta$

$$\begin{aligned}\left|\widehat{\beta}_n(k) - \beta(k)\right| &\leqslant \left|\ln \widehat{p}_k^n - \ln p_k\right| + \left|\ln \widehat{p}_{k+1}^n - \ln p_{k+1}\right| \\ &\leqslant 3\sqrt{\frac{u_n(\delta)}{p_k}} + 3\sqrt{\frac{u_n(\delta)}{p_{k+1}}} \\ &\leqslant 6\sqrt{\frac{u_n(\delta)}{p_{k+1}}}.\end{aligned}$$

□

Finally, Theorem 3.2.1 follows after noticing that $\left|\widehat{\beta}_n(k) - \beta\right| \leqslant \left|\widehat{\beta}_n(k) - \beta(k)\right| + |\beta(k) - \beta|$ and applying equation (3.3) and lemma 3.5.1.

3.5.2 Proof of Theorem 3.2.2

Lemma 3.5.3.

$$\lim_{k \to +\infty} \beta(k) = \beta.$$

Proof. Because L is an slowly varying function, $\lim_{x\to\infty} \frac{L(\lambda x)}{L(x)} = 1$ (see 1.2.1 of [21]) for all $\lambda > 0$, therefore $\frac{L(e^k)}{L(e^{k+1})} = \frac{L(e^k)}{L(ee^k)} \to 1$ and the result follows by taking limits in (3.3). □

Let $\varepsilon > 0$. Because $k_n \to \infty$, Lemma 3.5.3 implies that $\beta(k_n) \to \beta$, therefore we can find $N_1 \in \mathbb{N}$ such that

$$|\beta(k_n) - \beta| \leqslant \frac{\varepsilon}{2} \quad \forall n \geqslant N_1, \tag{3.23}$$

Let's take $\delta = \frac{2}{n^2}$, then $u_n(\delta) = \frac{2 \ln n}{n}$.

Because L is slowly varying, $L\left(e^{k_n+1}\right) \sim L\left(e^{k_n}\right)$, then $e^{k_n \beta} \frac{\ln n}{n} = o\left(L\left(e^{k_n+1}\right)\right)$

and we can find $N_2 \in \mathbb{N}$ such that

$$p_{k_n+1} = \frac{L(e^{k_n+1})}{e^{(k_n+1)\beta}} \geqslant \frac{32 \ln n}{n} = 16u_n(\delta) \quad \forall n \geqslant N_2,$$

therefore, we can apply Lemma 3.5.1, obtaining that, with probability bigger than $1 - \frac{4}{n^2}$

$$\left|\widehat{\beta}_n(k_n) - \beta(k_n)\right| \leqslant 6\sqrt{\frac{2 \ln n}{n p_{k_n+1}}} \quad \forall n \geqslant N_2, \tag{3.24}$$

Combining triangular inequality with equations (3.23) and (3.24) we have that for all $n \geqslant \max(N_1, N_2)$

$$\begin{aligned}\left|\widehat{\beta}_n(k_n) - \beta\right| &\leqslant \left|\widehat{\beta}_n(k_n) - \beta(k_n)\right| + |\beta(k_n) - \beta| \\ &\leqslant 6\sqrt{\frac{2 \ln n}{n p_{k_n+1}}} + \frac{\varepsilon}{2}\end{aligned} \tag{3.25}$$

with probability bigger than $1 - \frac{4}{n^2}$.

Plugging $p_{k_n+1} = \frac{L(e^{k_n+1})}{e^{(k_n+1)\beta}}$ in the first term of right hand side of (3.25), we get

$$6\sqrt{\frac{2 \ln n}{n p_{k_n+1}}} = 6\sqrt{\frac{2 \ln n}{n} \frac{e^{(k_n+1)\beta}}{L(e^{k_n+1})}} = 6\sqrt{2e^{\beta}}\sqrt{\frac{\ln n}{n} \frac{e^{(k_n)\beta}}{L(e^{k_n+1})}}$$

The assumption that $e^{k_n \beta} \frac{\ln n}{n} = o\left(L\left(e^{k_n+1}\right)\right)$ implies the above equality converges to 0 therefore we can find $N_3 \in \mathbb{N}$ such that $\left|6\sqrt{\frac{2 \ln n}{n p_{k_n+1}}}\right| \leqslant \frac{\varepsilon}{2}$ for all $n \geqslant N_3$.

Then, for all $n \geqslant \max(N_1, N_2, N_3)$

$$\left|\widehat{\beta}_n(k_n) - \beta\right| \leqslant \varepsilon$$

with probability bigger than $1 - \frac{4}{n^2}$. Because this is valid for all $\varepsilon > 0$, it implies that $\widehat{\beta}_n(k_n)$ converges in probability to β. Moreover, because $\sum_n \frac{4}{n^2}$ converges,

Borell-Cantelli lemma implies that $\widehat{\beta}_n(k_n) \to \beta$ almost surely.

3.5.3 Proof of Corollary 3.2.1

If we take $k_n = A \ln n$, we have $e^{\beta k_n} = n^{A\beta}$ then

$$\lim_n e^{k_n \beta} \frac{\ln n}{nL(e^{k_n})} = \lim_n \frac{n^{A\beta} \ln n}{nL(n^A)} = \lim_n \frac{\ln n}{n^{1-A\beta} L(n^A)}$$
$$= \lim_n \frac{\ln n}{n^{(1-A\beta)/2}} \frac{1}{n^{(1-A\beta)/2} L(n^A)} = 0.$$

For the last limit we have used that if L is slowly varying, then $L(n^A)$ is also slowly varying and that $\lim_n n^\gamma L(n) \to +\infty$ for $\gamma > 0$ and L slowly varying [21, Proposition 1.3.6.v]. Theorem 3.2.1 now follows by Theorem 3.2.2.

3.5.4 Proof of Theorem 3.2.3 and Corollary 3.2.2

Lemma 3.5.4. *Let X_n be a sequence of positive random variables and a_n and b_n two positive sequences such that $a_n > 0$, $\frac{b_n}{a_n} \to 0$. If there exists a random variable Y with continuous distribution function F such that*

$$\frac{X_n - a_n}{b_n} \underset{d}{\to} Y,$$

then,

$$a_n \left(\frac{\ln X_n - \ln a_n}{b_n} \right) \underset{d}{\to} Y.$$

Proof. Let $x \in \mathbb{R}$ be fixed. Because $\frac{X_n - a_n}{b_n} \underset{d}{\to} Y$, we have

$$P\left(X_n \leqslant a_n + b_n x\right) \to F\left(x\right).$$

Using that $a_n + b_n x = a_n \left(1 + \frac{b_n}{a_n} x\right)$ and taking logs we get

$$P\left(\ln X_n \leqslant \ln a_n + \ln\left(1 + \frac{b_n}{a_n} x\right)\right) \to F(x).$$

$\frac{b_n}{a_n} \to 0$ implies that $\frac{b_n}{a_n} x \to 0$, therefore, for n big enough,

$$\ln\left(1 + \frac{b_n}{a_n} x\right) = \frac{b_n}{a_n} x + o\left(\frac{b_n}{a_n}\right).$$

Then,

$$P\left(\ln X_n - \ln a_n \leqslant \frac{b_n}{a_n} x + o\left(\frac{b_n}{a_n}\right)\right) \to F(x)$$
$$P\left(a_n \frac{\ln X_n - \ln a_n}{b_n} \leqslant x + o(1)\right) \to F(x)$$

and the result follows from the continuity of F. □

Lemma 3.5.5. *If k_n satisfies the hypothesis of Theorem 3.2.2, then,*

$$\frac{\widehat{p}_{k_n}^n}{\bar{F}\left(e^{k_n}\right)} \xrightarrow{a.s.} 1.$$

Proof. By Lemma 3.5.2, for any $\delta > 0$ such that $p_k \geqslant 4u_n(\delta)$ we have that, with probability bigger than $1 - \delta$,

$$|\widehat{p}_k^n - p_k| \leqslant 2\sqrt{p_k u_n(\delta)},$$

then,

$$P\left(\left|\frac{\widehat{p}_k^n}{p_k} - 1\right| \leqslant 2\sqrt{\frac{u_n(\delta)}{p_k}}\right) \geqslant 1 - \delta. \tag{3.26}$$

As in the proof of Theorem 3.2.2, let $\delta = \frac{2}{n^2}$, so $u_n(\delta) = \frac{2 \ln n}{n}$. The condition

$e^{k_n\beta}\frac{\ln n}{n} = o\left(L\left(e^{k_n+1}\right)\right) = o\left(L\left(e^{k_n}\right)\right)$ implies that we can find $N_1 \in \mathbb{N}$ such that $p_{k_n} \geqslant \frac{8\ln n}{n}$ for all $n \geqslant N_1$, therefore, by equation (3.26),

$$P\left(\left|\frac{\widehat{p}^n_{k_n}}{p_{k_n}} - 1\right| \leqslant 2\sqrt{\frac{2\ln n}{np_{k_n}}}\right) \geqslant 1 - \frac{2}{n^2} \quad \forall n \geqslant N_1.$$

Let $\varepsilon > 0$. Notice that $\frac{\ln n}{np_{k_n}} = \frac{\ln n}{ne^{-k_n\beta}L(e^{k_n})} = \frac{\ln(n)e^{k_n\beta}}{nL(e^{k_n})}$ and this goes to 0 as n goes to $+\infty$, therefore, we can find N_2 such that $2\sqrt{\frac{2\ln n}{np_{k_n}}} \leqslant \varepsilon$ for all $n \geqslant N_2$, then

$$P\left(\left|\frac{\widehat{p}^n_{k_n}}{p_{k_n}} - 1\right| \leqslant \varepsilon\right) \geqslant 1 - \frac{2}{n^2} \quad \forall n \geqslant \max\left(N_1, N_2\right),$$

and the Lemma follows by Borel-Cantelli's Lemma. □

The following lemma can be obtained using the same arguments of Example 11 on [39].

Lemma 3.5.6. *Let X_n be a sequence of i.i.d. random variables with survival function (3.1), ϕ_1 and ϕ_2 bounded functions and u_n an increasing sequence of real numbers such that $u_n \to +\infty$. Define*

$$X_{n,i} = \frac{X_i}{u_n}\mathbb{I}\left\{\frac{X_i}{u_n} > 1\right\}, \quad v_n = P\left(X_{n,i} \neq 0\right) \quad \text{and}$$

$$\widetilde{Z}_n\left(\phi_k\right) = \frac{1}{\sqrt{nv_n}}\sum_{i=1}^{n}\left(\phi_k\left(X_{n,i}\right) - E\phi_k\left(X_{n,i}\right)\right).$$

If there exists a sequence r_n such that

(A1) $r_n = o(n)$.

(A2) $r_n v_n \to 0$.

(A3) $nv_n \to +\infty$.

(A4) $E\left(\sum_{i=1}^{r_n}\phi_k\left(X_{n,i}\right)\right)^4 = O\left(r_n v_n\right) \quad k=1,2.$

(A5) $\lim_n \frac{1}{r_n v_n}\sum_{i=1}^{r_n}\sum_{j=1}^{r_n} E\left(\phi_k\left(X_{n,i}\right)\phi_l\left(X_{n,j}\right)\right) = \sigma_{kl}.$

then $\left(\widetilde{Z}_n\left(\phi_k\right)\right)_{1\leqslant k\leqslant 2}$ *converges weakly to a centered normal distribution with covariance matrix* $\left(\sigma_{kl}\right)_{1\leqslant k,l\leqslant 2}$.

Let k_n satisfy the conditions of Theorem 3.2.2, take $u_n = e^{k_n}$, $\phi_1(x) = \mathbb{I}\{x>1\}$ and $\phi_2(x) = \mathbb{I}\{x>e\}$. With this notation

$$\begin{aligned}
\phi_1\left(X_{n,i}\right) &= \mathbb{I}\left\{\left(\frac{X_i}{u_n}\mathbb{I}\left\{\frac{X_i}{u_n}>1\right\}\right)>1\right\} = \mathbb{I}\left\{\frac{X_i}{u_n}>1\right\},\\
\phi_2\left(X_{n,i}\right) &= \mathbb{I}\left\{\left(\frac{X_i}{u_n}\mathbb{I}\left\{\frac{X_i}{u_n}>1\right\}\right)>e\right\} = \mathbb{I}\left\{\frac{X_i}{u_n}>e\right\},\\
E\phi_1\left(X_{n,i}\right) &= P\left(\frac{X_i}{u_n}>1\right) = \overline{F}\left(u_n\right),\\
E\phi_2\left(X_{n,i}\right) &= P\left(\frac{X_i}{u_n}>e\right) = \overline{F}\left(eu_n\right),\\
v_n &= P\left(X_{n,i}\neq 0\right) = P\left(\frac{X_i}{u_n}>1\right) = \overline{F}\left(u_n\right).
\end{aligned}$$

Let $w_n = \overline{F}\left(eu_n\right)$, $\lambda_n = \frac{\overline{F}(u_n)}{\overline{F}(eu_n)} = \frac{v_n}{w_n}$ (notice that $\lambda_n \to e^{\beta}$) and $y_n = \sqrt{\frac{v_n}{n{w_n}^2}}$, then,

$$\begin{aligned}
\widehat{\lambda}_n &= \frac{\sum_{i=1}^{n}\mathbb{I}\{X_i>u_n\}}{\sum_{i=1}^{n}\mathbb{I}\{X_i>eu_n\}} = \frac{\sum_{i=1}^{n}\phi_1\left(X_{n,i}\right)}{\sum_{i=1}^{n}\phi_2\left(X_{n,i}\right)}\\
&= \frac{nE\phi_1\left(X_{n,i}\right)+\sum_{i=1}^{n}\left\{\phi_1\left(X_{n,i}\right)-E\phi_1\left(X_{n,i}\right)\right\}}{nE\phi_2\left(X_{n,i}\right)+\sum_{i=1}^{n}\left\{\phi_2\left(X_{n,i}\right)-E\phi_2\left(X_{n,i}\right)\right\}}\\
&= \frac{\frac{E\phi_1(X_{n,i})}{E\phi_2(X_{n,i})}+\frac{\sum_{i=1}^{n}\{\phi_1(X_{n,i})-E\phi_1(X_{n,i})\}}{nE\phi_2(X_{n,i})}}{1+\frac{\sum_{i=1}^{n}\{\phi_2(X_{n,i})-E\phi_2(X_{n,i})\}}{nE\phi_2(X_{n,i})}} = \frac{\frac{E\phi_1(X_{n,i})}{E\phi_2(X_{n,i})}+\widetilde{Z}_n\left(\phi_1\right)\sqrt{\frac{v_n}{n{w_n}^2}}}{1+\widetilde{Z}_n\left(\phi_2\right)\sqrt{\frac{v_n}{n{w_n}^2}}}
\end{aligned}$$

$$= \frac{\lambda_n + \widetilde{Z}_n(\phi_1) y_n}{1 + \widetilde{Z}_n(\phi_2) y_n}. \tag{3.27}$$

To apply Lemma 3.5.6, let's take $r_n = k_n$ and assume that $r_n = o(n)$. For condition (A2) we have:

$$\lim_n r_n v_n = \lim_n k_n \overline{F}\left(e^{k_n}\right) = \lim_n \frac{k_n L\left(e^{k_n}\right)}{e^{k_n \beta}} = \lim_n \frac{k_n}{e^{\frac{k_n \beta}{2}}} \frac{L\left(e^{k_n}\right)}{e^{\frac{k_n \beta}{2}}} = 0.$$

Because $e^{k_n \beta} \frac{\ln n}{n} = o\left(L\left(e^{k_n+1}\right)\right)$ and L is slowly varying, we can write $e^{k_n \beta} \frac{\ln n}{n} = L\left(e^{k_n}\right) \varepsilon(n)$ where $\varepsilon(n) \to 0$ and this implies (A3).

For (A4), note that $\sum_{i=1}^{r_n} \phi_1(X_{n,i})$ follows a binomial distribution with parameters r_n and v_n, then, $E\left(\sum_{i=1}^{r_n} \phi_1(X_{n,i})\right)^4 = r_n v_n \left(1 - h(r_n, v_n)\right)$, where h is a two variables polynomial of degree 3 such that the degree of r_n on each monomial is always greater or equal than the degree of v_n. Condition (A2), and the fact that $v_n \to 0$ implies $E\left(\sum_{i=1}^{r_n} \phi_1(X_{n,i})\right)^4 = O(r_n v_n)$. With a similar argument and using the fact that $\frac{w_n}{v_n} \to e^{-\beta}$, it can be shown that $E\left(\sum_{i=1}^{r_n} \phi_2(X_{n,i})\right)^4 = O(r_n v_n)$ and therefore condition (A4) is fulfilled.

For (A5), first, notice that

$$E\left(\phi_1(X_{n,i})\, \phi_1(X_{n,j})\right) = \begin{cases} v_n\ ,\ i = j \\ {v_n}^2\ ,\ i \neq j \end{cases}$$

$$E\left(\phi_1(X_{n,i})\, \phi_2(X_{n,j})\right) = \begin{cases} w_n\ ,\ i = j \\ v_n w_n\ ,\ i \neq j \end{cases}$$

$$E\left(\phi_2(X_{n,i})\, \phi_2(X_{n,j})\right) = \begin{cases} w_n\ ,\ i = j \\ {w_n}^2\ ,\ i \neq j \end{cases}$$

therefore,

$$\sigma_{11} = \lim_n \frac{1}{r_n v_n} \sum_{i=1}^{r_n} \sum_{j=1}^{r_n} E\left(\phi_1\left(X_{n,i}\right)\phi_1\left(X_{n,j}\right)\right) = \lim_n \frac{r_n v_n + r_n\left(r_n-1\right){v_n}^2}{r_n v_n} = 1$$

$$\sigma_{12} = \lim_n \frac{1}{r_n v_n} \sum_{i=1}^{r_n} \sum_{j=1}^{r_n} E\left(\phi_1\left(X_{n,i}\right)\phi_2\left(X_{n,j}\right)\right) = \lim_n \frac{r_n w_n + r_n\left(r_n-1\right)v_n w_n}{r_n v_n} = e^{-\beta}$$

$$\sigma_{22} = \lim_n \frac{1}{r_n v_n} \sum_{i=1}^{r_n} \sum_{j=1}^{r_n} E\left(\phi_2\left(X_{n,i}\right)\phi_2\left(X_{n,j}\right)\right) = \lim_n \frac{r_n w_n + r_n\left(r_n-1\right){w_n}^2}{r_n v_n} = e^{-\beta}.$$

By Lemma 3.5.6, $\left(\widetilde{Z}_n\left(\phi_k\right)\right)_{1\leqslant k\leqslant 2}$ converges to a centered normal distribution with covariance matrix $\left(\sigma_{kl}\right)_{1\leqslant k,l\leqslant 2}$. Taking into account that $y_n \sim \frac{e^\beta}{\sqrt{nv_n}}$, it follows that

$$\begin{aligned}\widehat{\lambda}_n &= \left(\lambda_n + \widetilde{Z}_n\left(\phi_1\right) y_n\right)\left(1 - \widetilde{Z}_n\left(\phi_2\right) y_n + o_P\left(\frac{1}{\sqrt{nv_n}}\right)\right)\\ &= \lambda_n + y_n\left(\widetilde{Z}_n\left(\phi_1\right) - \lambda_n \widetilde{Z}_n\left(\phi_2\right)\right) + o_P\left(\frac{1}{\sqrt{nv_n}}\right).\end{aligned}$$

Then, $\sqrt{nv_n}\left(\widehat{\lambda}_n - \lambda_n\right)$ converges weakly to a centered normal distribution with variance $e^{2\beta}\left(\sigma_{11} + e^{2\beta}\sigma_{22} - 2e^{\beta}\sigma_{12}\right) = e^{2\beta}\left(e^{\beta}-1\right)$. This can be resumed in the following lemma.

Lemma 3.5.7. *Let X_n and u_n be as in Lemma 3.5.6, if k_n satisfies the conditions of Theorem 3.2.2 and $k_n = o(n)$, then*

$$\sqrt{n\bar{F}\left(e^{k_n}\right)}\left(\frac{\sum_{i=1}^{n}\mathbb{I}\left\{X_i > e^{k_n}\right\}}{\sum_{i=1}^{n}\mathbb{I}\left\{X_i > e^{k_n+1}\right\}} - \frac{\bar{F}\left(e^{k_n}\right)}{\bar{F}\left(e^{k_n+1}\right)}\right)$$

converges weakly to a centered normal distribution with variance $e^{2\beta}\left(e^{\beta}-1\right)$.

Lemmas 3.5.3, 3.5.4 and 3.5.7 combined with equation (3.3) imply the first part

of Theorem 3.2.3, the second part follows from Lemma 3.5.5 and Slutsky's Theorem (pp. 248 in [55]). Corollary 3.2.2 follows immediately.

3.5.5 Averaged Estimators

Here we collect some remarks and results related to the averaged estimator (3.7). First, we detail how to get the expression (3.2.1) from (3.3). Let $k > 0$ be fixed, we have:

$$\beta(k+j) = \beta + \ln\left(\frac{L\left(e^{k+j}\right)}{L\left(e^{k+j+1}\right)}\right), \quad \forall j,$$

so that

$$\begin{aligned}\frac{1}{2m+1}\sum_{j=-m}^{m}\beta(k+j) &= \frac{1}{2m+1}\sum_{j=-m}^{m}\beta + \frac{1}{2m+1}\sum_{j=-m}^{m}\ln\left(\frac{L\left(e^{k+j}\right)}{L\left(e^{k+j+1}\right)}\right)\\ &= \beta + \frac{1}{2m+1}\ln\left(\prod_{j=-m}^{m}\frac{L\left(e^{k+j}\right)}{L\left(e^{k+j+1}\right)}\right)\\ &= \beta + \frac{1}{2m+1}\ln\left(\frac{L\left(e^{k-m}\right)}{L\left(e^{k+m+1}\right)}\right).\end{aligned}$$

The result below establishes the strong consistency of the estimator (3.7).

The following results show that, for well-chosen k_n and m_n, the estimator $\widehat{\beta}_n(k_n, m_n)$ is strongly consistent.

Theorem 3.5.1 (Strong consistency). *Let k_n and m_n such that*

i $k_n - m_n \to +\infty$.

ii $\sum_{n=1}^{+\infty}\frac{4}{n^2}(1-2m_n)$ *is convergent.*

iii $e^{(k_n+m_n)\beta\frac{\ln n}{n}} = o\left(L\left(e^{k_n+m_n}\right)\right)$.

then, $\widehat{\beta}_n\left(k_n, m_n\right)$ *converges almost surely to* β.

Corollary 3.5.1. *Let* A, l *be a positive numbers such that* $l > 1$ *and* $\frac{1}{\beta} > \frac{A\beta(l+1)}{l}$ *then*

$$\widehat{\beta}_n\left(A\ln n, \frac{A\ln n}{l}\right) \to \beta \quad a.s.$$

3.5.6 Proof of Theorem 3.2.4

The following lemma provides us a bound for the difference between $\widehat{\beta}(k,m) - \beta(k,m)$.

Lemma 3.5.8. *Let* $\delta > 0$ *and* k *and* m *such that* $p_{k+m+1} \geqslant 16u_n(\delta)$, *then*

$$\left|\widehat{\beta}_n(k,m) - \beta(k,m)\right| \leqslant 6\sqrt{\frac{u_n(\delta)}{p_{k+m+1}}}, \tag{3.28}$$

with probability larger than $1 - 2\delta\left(1 - 2m\right)$.

Proof. For the left hand side of equation (3.28) we have

$$\begin{aligned}\left|\widehat{\beta}_n(k,m) - \beta(k,m)\right| &= \frac{1}{2m+1}\left|\sum_{j=-m}^{m}\left(\widehat{\beta}_n\left(k+j\right) - \beta\left(k+j\right)\right)\right| \\ &= \frac{1}{2m+1}\left|\sum_{j=0}^{2m}\left(\widehat{\beta}_n\left(k-m+j\right) - \beta\left(k-m+j\right)\right)\right|,\end{aligned}$$

then,

$$\left|\widehat{\beta}_n(k,m) - \beta(k,m)\right| \leqslant \frac{1}{2m+1}\sum_{j=0}^{2m}\left|\widehat{\beta}_n\left(k-m+j\right) - \beta\left(k-m+j\right)\right|. \tag{3.29}$$

Because $p_{k+m+1} \geqslant 16u_n(\delta)$, we have that $p_{k-m+j+1} \geqslant 16u_n(\delta)$ for all j be-

tween 0 and $2m$, therefore, we can apply Lemma 3.5.1 obtaining that, for each j

$$\left|\widehat{\beta}_n\left(k-m+j\right)-\beta\left(k-m+j\right)\right| \leqslant 6\sqrt{\frac{u_n(\delta)}{p_{k-m+j+1}}}, \tag{3.30}$$

with probability bigger than $1-2\delta$. The probability that inequality (3.30) is true for all j between 0 and $2m$ is bigger than $(2m+1)(1-2\delta)-2m = 1-2\delta(1-2m)$, therefore, with at least that probability, equation (3.29) becomes,

$$\left|\widehat{\beta}_n(k,m)-\beta(k,m)\right| \leqslant \frac{1}{2m+1}\sum_{j=0}^{2m} 6\sqrt{\frac{u_n(\delta)}{p_{k-m+j+1}}} \leqslant 6\sqrt{\frac{u_n(\delta)}{p_{k+m+1}}}.$$

□

Similarly to the proof of Theorem 3.2.1, Theorem 3.2.4 now follows by triangular inequality, equation (3.6) and Lemma 3.5.8.

3.5.7 Proof of Theorem 3.5.1

Lemma 3.5.9. *Let α_n, k_n and b_n be sequences such that, $\alpha_n \to \alpha$, $k_n \to +\infty$ and $k_n - b_n \to +\infty$. Then,*

$$\frac{1}{2b_n+1}\sum_{j=-b_n}^{b_n} \alpha_{k_n+j} \to \alpha.$$

Proof. Let $\lambda_n = \frac{1}{2b_n+1}\sum_{j=-b_n}^{b_n} \alpha_{k_n+j}$, then

$$\lambda_n - \alpha = \frac{1}{2b_n+1}\sum_{j=-b_n}^{b_n}(\alpha_{k_n+j}-\alpha) = \frac{1}{2b_n+1}\sum_{j=0}^{2b_n}(\alpha_{k_n-b_n+j}-\alpha).$$

Take $\varepsilon > 0$ fixed, by the convergence of α_n there exists N_1 such that, for all $n \geqslant N_1$, $|\alpha_n - \alpha| < \varepsilon$. Because $k_n - b_n \to +\infty$, we can find N_2 such that $k_n - b_n \geqslant N_1$, for all $n \geqslant N_2$, then, $|\alpha_{k_n-b_n+j}-\alpha| \leqslant \varepsilon$ for all $n \geqslant N_2$ and all

$j \in \mathbb{N}$. This implies that for all $n \geqslant N_2$

$$|\lambda_n - \alpha| \leqslant \frac{1}{2b_n+1} \sum_{j=0}^{2b_n} |\alpha_{k_n - b_n + j} - \alpha| \leqslant \frac{1}{2b_n+1} \sum_{j=0}^{2b_n} \varepsilon = \varepsilon.$$

□

Lemma 3.5.9 shows that if $k_n - m_n \to +\infty$, then

$$\beta\left(k_n, m_n\right) \to \beta. \tag{3.31}$$

Theorem 3.5.1 follows by the same argument used to prove Theorem 3.2.2, using (3.31) instead of Lemma 3.5.3 and Lemma 3.5.8 instead of Lemma 3.5.1.

3.5.8 Proof of Corollary 3.5.1

We just need to show that sequences $k_n = A \ln n$ and $m_n = \frac{A \ln n}{l}$ satisfy conditions i, ii and iii in Theorem 3.5.1. The first two are trivially satisfied, for the third one, notice that

$$\begin{aligned}\lim_n \frac{e^{\left(A \ln n + \frac{A \ln n}{l}\right)\beta}}{L\left(e^{A \ln n + \frac{A \ln n}{l}}\right)} \frac{\ln n}{n} &= \lim_n \frac{n^{\left(1+\frac{1}{l}\right)A\beta}}{L\left(n^{\left(1+\frac{1}{l}\right)A}\right)} \frac{\ln n}{n} = \lim_n \frac{\ln n}{n^{1-\left(1+\frac{1}{l}\right)A\beta} L\left(n^{\left(1+\frac{1}{l}\right)A}\right)} \\ &= \lim_n \frac{1}{n^{\frac{1-\left(1+\frac{1}{l}\right)A\beta}{2}} L\left(n^{\left(1+\frac{1}{l}\right)A}\right)} \frac{\ln n}{n^{\frac{1-\left(1+\frac{1}{l}\right)A\beta}{2}}}.\end{aligned}$$

The condition $\frac{1}{\beta} > \frac{A(l+1)}{l}$ implies that $1 - \left(1 + \frac{1}{l}\right) A\beta > 0$, therefore,

$$\lim_n \frac{1}{n^{\frac{1-\left(1+\frac{1}{l}\right)A\beta}{2}} L\left(n^{\left(1+\frac{1}{l}\right)}\right)} = 0 \quad \text{and} \quad \lim_n \frac{\ln n}{n^{\frac{1-\left(1+\frac{1}{l}\right)A\beta}{2}}} = 0,$$

which shows that k_n and m_n satisfy condition iii in Theorem 3.5.1.

3.5.9 Proof of Lemma 3.2.1

The representation is a direct application of Lemma 3.5.10 and the fact that g is a regularly varying function of index ρ.

Lemma 3.5.10. *Assume that L satisfies $SR2$, has positive decrease and x is big enough such that representation the (3.10) holds, then*

$$\ln\left(\frac{L(x)}{L(\lambda x)}\right) = -c|\rho|^{-1}\left(g(x) - g(\lambda x)\right) + o\left(g(x)\right). \tag{3.32}$$

Proof. Let's denote $A(x) = c\rho^{-1}g(x) + o\left(g\left(x\right)\right)$. By (3.10) we have

$$\begin{aligned}\ln\left(\frac{L(x)}{L(\lambda x)}\right) &= \ln\left(\frac{C(1+A(x))}{C(1+A(\lambda x))}\right) = \ln\left(\frac{1+A(x)}{1+A(\lambda x)}\right)\\ &= \ln\left(1+A\left(x\right)\right) - \ln(1+A(\lambda x)).\end{aligned} \tag{3.33}$$

Applying the first order expansion for $\ln(1+A(x))$ we have that

$$\begin{aligned}\ln\left(1+A\left(x\right)\right) &= c\rho^{-1}g(x) + o\left(g\left(x\right)\right) + \underbrace{o\left(c\rho^{-1}g(x) + o\left(g\left(x\right)\right)\right)}_{o(g(x))}\\ &= c\rho^{-1}g(x) + o\left(g\left(x\right)\right).\end{aligned} \tag{3.34}$$

Applying (3.34) to λx we get

$$\ln\left(1+A\left(\lambda x\right)\right) = c\rho^{-1}g(\lambda x) + o\left(g\left(x\right)\right), \tag{3.35}$$

where we have used that if g is regularly varying then $o\left(g\left(\lambda x\right)\right) = o\left(g\left(x\right)\right)$.

The result now follows by plugging (3.34) and (3.35) into (3.33). □

Chapter 4

Regenerative bootstrap for β-null recurrent Markov chains

The content of this chapter is based on [47].

Abstract: Two regeneration-based bootstrap methods, namely, the *Regeneration based-bootstrap* [6, 34] and the *Regenerative Block bootstrap* [15] are shown to be valid for the problem of estimating the integral of a function with respect to the invariant measure in a β-null recurrent Markov chain with an accessible atom. An extension of the Central Limit Theorem for randomly indexed sequences is also presented.

4.1 Introduction

In [43], Bradley Efron introduced the Bootstrap as a way to overcome some limitations of classical methods that often relied on strong assumptions about the data's underlying distribution or the model's form. Since then, these techniques, first studied in the i.i.d. case, have been developed and extended to time-series

(see [77] for an extensive survey of methods) and applied to a wide range of problems in various fields such as signal processing [122, 123], soil science [111] and econometrics [58, 82]. They can handle any level of complexity in data or models from fully parametric to completely nonparametric cases. These methods are easy to implement with modern computing power and can provide more accurate and reliable inferences than traditional methods in many situations.

Although originally designed for i.i.d. sampling, there has been significant interest in adapting the bootstrap to situations where the data is dependent. Several resampling methods have been proposed for time series data: these include the autoregressive-sieve bootstrap [74], block bootstrap [76] , circular bootstrap [101], the stationary bootstrap [102], continuous-path block bootstrap [93], tapered block bootstrap [95], frequency-domain bootstrap [72, 92], and local bootstrap [96]. For detailed reviews and comparisons of these methods see [27, 48, 71, 73] and the references therein.

In the Markovian case, numerous approaches have been developed and examined. In [75], the authors proposed a block resampling scheme that consists in resampling from a nonparametric estimate of the one-step transition matrix of a finite state Markov chain. This method was extended to the countable case in [6]. Extensions of this method have been proposed for the case where the state space is Euclidean, as seen in [105], [94, 97] and [57]. The general concept is to estimate the marginal distribution and the transition probability function using a nonparametric function estimation technique and then resample from those estimates. For a detailed explanation of this approach, refer to Section 4 in [71].

A completely new approach to this problem was introduced in [6]. Instead of using estimated transition probabilities, they exploit the regeneration properties of a Markov chain when an accessible atom is visited infinitely often. The main

idea underlying this method consists in dividing the chain into a random number of i.i.d. regeneration blocks and then resampling the same number of regeneration blocks. This method, named *Regeneration based bootstrap*, was proved to be valid for finite state atomic chains in [6], and it was extended to general atomic positive recurrent Markov chains in [34].

It was pointed out in [16] that the *Regeneration based bootstrap* is not second-order correct (its rate is $O_{\mathbb{P}}(n^{-1/2})$ only). To overcome this limitation, a variation of this method, called *Regenerative Block bootstrap* (RBB), was introduced in [15]. This method consists in imitating the renewal structure of the chain by sampling regeneration data blocks, until the length of the reconstructed bootstrap series is larger than the length n of the original data series (notice the difference with the *Regeneration based bootstrap*, where the number of sampled blocks is equal to the number of regeneration blocks in the original chain). It was shown in [15] that, for atomic positive recurrent Markov chains, the RBB for estimating the integral of a function with respect to the invariant probability, has a uniform rate of convergence of order $O_{\mathbb{P}}\left(n^{-1}\right)$ (the same as in the i.i.d. case).

Despite all these efforts in the positive recurrent case, up to our knowledge, no bootstrap method has been studied in the null-recurrent scenario. Hence, our objective in this paper is to start this study and show that both *Regeneration based-bootstrap* and *Regenerative Block bootstrap* are valid schemes for estimating integrals with respect to the invariant measure when the Markov chain is β-null recurrent and possesses an accessible atom.

The paper is organized as follows: in section 4.2 we provide a brief introduction to null recurrent Markov chains, making a special emphasis on atomic ones and presenting the main results that we use throughout the paper. In subsection 4.2.3 we present an extension of the Central Limit Theorem for randomly indexed se-

quences (Lemma 4.2.1). Section 4.3 is dedicated to the *Regenerative Block bootstrap* in β-null recurrent Markov chains, while Section 4.4 is devoted to the *Regeneration based-bootstrap*. In section 4.5 we have added a few simulations to show the behavior of both algorithms in practice. The technical proofs are postponed to Section 4.6.

4.2 A short introduction to null-recurrent Markov chains

In this section, we introduce some notation and review some important concepts from Markov chain theory that will be used throughout the paper. For more details, please refer to [38, 87].

4.2.1 Notation and definitions

Consider an homogeneous Markov chain $\mathbf{X} = X_0, X_1, \ldots,$ on a countably generated state space $(E, \mathcal{E})$, with transition kernel P and initial probability distribution λ. This means that for any $B \in \mathcal{E}$ and $n \in \mathbb{N}$, we have $\mathcal{L}(X_0) = \lambda$ and

$$\mathbb{P}(X_{n+1} \in B \mid X_0, \ldots, X_n) = P(X_n, B) \quad \text{almost surely.}$$

Note that the assumption of a countably generated state space is commonly used in Markov chain theory to avoid pathological examples known as 'anormal' chains [37]. For more information on this topic, see [37], [61], and [107]. An example of an 'anormal' chain can be found in [22]. This assumption does not significantly limit the generality of our results since most of the time $\mathcal{E} = \mathcal{B}\left(\mathbb{R}^d\right)$, which is countably generated.

In the following, we use $\mathbb{P}_\lambda$ (or $\mathbb{P}_x$ for x in E) to denote the probability measure on the underlying space such that $X_0 \sim \lambda$ (or $X_0 = x$). We use $\mathbb{E}_\lambda(.)$ to represent the $\mathbb{P}_\lambda$-expectation (or $\mathbb{E}_x(.)$ to represent the $\mathbb{P}_x$-expectation), and $\mathbb{I}\{\mathcal{A}\}$ to represent the indicator function of event $\mathcal{A}$.

A homogeneous Markov chain is said to be irreducible if there exists a σ-finite measure ϕ on $(E, \mathcal{E})$ such that for all $x \in E$ and all $A \in \mathcal{E}$ with $\phi(A) > 0$, there exists some $n \geqslant 1$ such that $P^n(x, A) > 0$. In this case, there exists a maximal irreducibility measure ψ with respect to which all other irreducibility measures are absolutely continuous. If $\mathbf{X}$ is ψ-irreducible, there is $d' \in \mathbb{N}^*$ and disjoints sets $D_1, \ldots, D_{d'}$ $D_{d'+1} = D_1$ weighted by ψ such that $\psi(E \backslash \cup_{1 \leqslant i \leqslant d'} D_i) = 0$ and $\forall x \in D_i, P(x, D_{i+1}) = 1$. The the g.c.d. d of such integers is called the *period* of the chain. $\mathbf{X}$ is said to be *aperiodic* if $d = 1$.

In the following, we assume that the Markov chains under consideration are homogeneous, aperiodic, and irreducible with maximal irreducibility measure ψ.

An irreducible chain possesses an accessible atom, if there is a set $\boldsymbol{\alpha} \in \mathcal{E}$ such that for all x, y in $\boldsymbol{\alpha}$: $P(x, \bullet) = P(y, \bullet)$ and $\psi(\boldsymbol{\alpha}) > 0$. For instance, when a chain can take a countable number of values, any single point visited by the chain is an atom. Denote by $\sigma_{\boldsymbol{\alpha}}$ and $\tau_{\boldsymbol{\alpha}}$, respectively, the times of first visit and first return of the chain to $\boldsymbol{\alpha}$, i.e. $\tau_{\boldsymbol{\alpha}} = \inf\{n \geqslant 1 : X_n \in \boldsymbol{\alpha}\}$ and $\sigma_{\boldsymbol{\alpha}} = \inf\{n \geqslant 0 : X_n \in \boldsymbol{\alpha}\}$. The subsequent visit and return times $\sigma_{\boldsymbol{\alpha}}, \tau_{\boldsymbol{\alpha}}(k)$, $k \geqslant 1$ are defined inductively as follows:

$$\tau_{\boldsymbol{\alpha}}(1) = \tau_{\boldsymbol{\alpha}} \quad , \quad \tau_{\boldsymbol{\alpha}}(k) = \min\{n > \tau_{\boldsymbol{\alpha}}(k-1) : X_n \in \boldsymbol{\alpha}\}, \tag{4.1}$$

$$\sigma_{\boldsymbol{\alpha}}(1) = \sigma_{\boldsymbol{\alpha}} \quad , \quad \sigma_{\boldsymbol{\alpha}}(k) = \min\{n > \sigma_{\boldsymbol{\alpha}}(k-1) : X_n \in \boldsymbol{\alpha}\}. \tag{4.2}$$

We use $T_n(\boldsymbol{\alpha})$ to represent the random variable that counts the number of times

the chain visits the set $\boldsymbol{\alpha}$ up to time n, i.e., $T_n(\boldsymbol{\alpha}) = \sum_{t=0}^{n} \mathbb{I}\{X_t \in \boldsymbol{\alpha}\}$. Similarly, we use $T(\boldsymbol{\alpha})$ to represent the total number of visits of chain $\mathbf{X}$ to $\boldsymbol{\alpha}$. An atom $\boldsymbol{\alpha}$ is called *recurrent* if $\mathbb{E}_x T(\boldsymbol{\alpha}) = +\infty$ for all $x \in \boldsymbol{\alpha}$; otherwise, it is called *transient*. A notable property of atomic chains is that all accessible atoms are either all recurrent or all transient. Therefore, we say that an atomic chain is recurrent if one (and thus all) of its accessible atoms is recurrent.

Denote by $\mathbb{P}_{\boldsymbol{\alpha}}$ and $\mathbb{E}_{\boldsymbol{\alpha}}(.)$ the probability and the expectation conditionally to $X_0 \in \boldsymbol{\alpha}$. If $\mathbf{X}$ possesses an accessible atom and is aperiodic, the probability of returning infinitely often to the atom $\boldsymbol{\alpha}$ is equal to one, no matter the starting point, i.e.

$$\mathbb{P}_x\big(T(\boldsymbol{\alpha}) = \infty\big) = 1 \quad \forall x \in E.$$

A fundamental tool for understanding the long-term behavior of Markov chains is the existence of invariant measures, that is, a measure π such that

$$\pi(A) = \int P(x, A)\, d\pi(x) \quad \forall A \in \mathcal{E}.$$

Every irreducible and recurrent Markov chain admits a unique (up to a multiplicative constant) invariant measure [87, Theorem 10.4.9]. In the atomic case, the invariant measure is just the occupation measure over $\mathcal{B}_1 = \left(X_{\tau_{\boldsymbol{\alpha}}(1)+1}, \ldots, X_{\tau_{\boldsymbol{\alpha}}(2)}\right)$ [38, Theorem 6.4.2], i.e.

$$\pi_{\boldsymbol{\alpha}}(A) = \mathbb{E}_{\boldsymbol{\alpha}}\left(\sum_{j=1}^{\tau_{\boldsymbol{\alpha}}} \mathbb{I}\{X_j \in A\}\right), \quad \forall A \in \mathcal{E}. \tag{4.3}$$

An irreducible Markov chain is *positive recurrent* if its invariant measure is finite. When the invariant measure is just σ-finite, then the chain is called *null recurrent*. From (4.3), it is clear that an atomic Markov chain is positive recurrent if

and only if $\mathbb{E}_{\boldsymbol{\alpha}}\tau_{\boldsymbol{\alpha}} < +\infty$, and in this case, the measure defined by $\frac{\pi_{\boldsymbol{\alpha}}}{\mathbb{E}_{\boldsymbol{\alpha}}\tau_{\boldsymbol{\alpha}}}$ is an invariant probability for the chain. The existence of this invariant probability makes the theory of positive recurrent Markov chains, very similar to the i.i.d. case [87, Chapter 17].

Conversely, dealing with null recurrent chains is considerably more challenging, and a comprehensive theory of non-parametric estimation for this type of chain does not exist. To address this issue, Karlsen and Tjøstheim introduced in [67] a regularity condition for the tail behavior of the distribution of $\tau_{\boldsymbol{\alpha}}$ that renders the problem more tractable. Specifically, a chain is referred to as β-null recurrent (refer to [67, Definition 3.2 and Theorem 3.1]) if there is a constant $\beta \in (0, 1)$ and a slowly varying function[1] L such that

$$\mathbb{P}_{\boldsymbol{\alpha}}(\tau_{\boldsymbol{\alpha}} > n) \sim \frac{1}{\Gamma(1-\beta)n^{\beta}L(n)}. \tag{4.4}$$

The number β, also known as the *regularity index* (see [28, 30]) satisfies

$$\beta = \sup\{p > 0 : \mathbb{E}_{\boldsymbol{\alpha}}(\tau_{\boldsymbol{\alpha}}^{p}) < +\infty\}.$$

Some of the most well-known examples of β-null recurrent Markov chain are the random walks in $\mathbb{R}$, which are $1/2$-null recurrent [64], the Bessel random walks [3], [36] and some types of threshold autoregressive (TAR) [49] and vector autoregressive processes (VAR) [88].

[1] A measurable and positive function L is said to be *slowly varying at* $+\infty$ if it is defined in $[a, +\infty)$ for some $a \geqslant 0$, and satisfies $\lim_{x\to+\infty} \frac{L(xt)}{L(x)} = 1$ for all $t \geqslant a$. For a detailed discussion on these types of functions, refer to [21].

4.2.2 Renewal properties and Block decomposition

The *strong Markov property* implies that the sample paths of an atomic Markov chain can be partitioned into independent blocks of random length corresponding to consecutive visits to $\boldsymbol{\alpha}$, given by:

$$\begin{aligned}
\mathcal{B}_0 &= \left(X_0, X_1, \ldots, X_{\tau_{\boldsymbol{\alpha}}(1)}\right) \\
\mathcal{B}_1 &= \left(X_{\tau_{\boldsymbol{\alpha}}(1)+1}, \ldots, X_{\tau_{\boldsymbol{\alpha}}(2)}\right) \\
&\ldots \\
\mathcal{B}_n &= \left(X_{\tau_{\boldsymbol{\alpha}}(n)+1}, \ldots, X_{\tau_{\boldsymbol{\alpha}}(n+1)}\right) \\
&\ldots
\end{aligned}$$

Note that the distribution of $\mathcal{B}_0$ depends on the initial measure, and thus it does not have the same distribution as $\mathcal{B}_j$ for $j \geqslant 1$. The sequence $\{\tau_{\boldsymbol{\alpha}}(j)\}_{j\geqslant 1}$ defines successive times at which the chain forgets its past, which are called *regeneration times*. Similarly, the sequence of i.i.d. blocks $\{\mathcal{B}_j\}_{j\geqslant 1}$ is called *regeneration blocks*. As customary in the β-null recurrent Markov chain literature, we will use $T(n)$ to denote the number of complete regeneration blocks up to time n, i.e. $T(n) = \max\left(T_n(\boldsymbol{\alpha}) - 1, 0\right)$. We will denote by $\ell(\mathcal{B}_i)$ the length of the i-th block, therefore,

$$\ell(\mathcal{B}_j) = \begin{cases} \tau_{\boldsymbol{\alpha}} & , \quad j = 0 \\ \tau_{\boldsymbol{\alpha}}(j+1) - \tau_{\boldsymbol{\alpha}}(j) & , \quad j \geqslant 1 \end{cases} \tag{4.5}$$

The random variable $T(n)$, and its relationship with $\sum_{j=0}^{k} \ell(\mathcal{B}_j)$, is crucial in the theory we will develop in this paper, therefore, we will state in this section its main properties in the β-null recurrent scenario.

Assume $\mathbf{X}$ is a β-null recurrent Markov chain with an accessible atom $\boldsymbol{\alpha}$. By

(3.27) in [67], the function L in (4.4) can be normalized in such a way that

$$u(z) = z^{\beta} L(z) \tag{4.6}$$

is a continuous function that is strictly increasing in the interval $[z_0, +\infty)$ for some $z_0 \in \mathbb{R}_+$. Define $v(z)$ as

$$v(z) = u^{(-1)}(z) = \inf\{s : u(s) > z\}, \tag{4.7}$$

then, $u(v(z)) = v(u(z)) = z$ for $z \geqslant z_0$.

Consider the space of càdlàg functions defined on the interval $[0, +\infty)$, denoted by $\mathscr{D}_{[0,+\infty)}$. This space consists of the real functions that are right-continuous with left limits and defined over $[0, +\infty)$. More precisely, a function $g \in \mathscr{D}_{[0,+\infty)}$ if and only if g is right-continuous, has left limits at all points $t > 0$, and $\lim_{t \downarrow 0} g(t) = g(0)$. The space $\mathscr{D}_{[0,+\infty)}$ is equipped with the Skorokhod[2] topology, making it a completely separable metric space. We will use $\xrightarrow{\mathscr{D}_{[0,+\infty)}}$ to denote weak convergence in this space, and $\xrightarrow{\text{fd}}$ for convergence of finite-dimensional laws. Two stochastic processes Y_n, Z_n in $\mathscr{D}_{[0,+\infty)}$ are said to be *equivalent* if $Y_n - Z_n$ converges weakly to the zero process. If $Y_n \xrightarrow{\mathscr{D}_{[0,+\infty)}} Y$ and Y_n and Z_n are equivalent, then $Z_n \xrightarrow{\mathscr{D}_{[0,+\infty)}} Y$ (see Lemma 3.31 in [60]).

Define the following processes

$$T_n(t) = \frac{T(\lfloor nt \rfloor)}{u(n)}, \quad C_n(t) = \frac{1}{v(n)} \sum_{k=0}^{\lfloor nt \rfloor} \ell(\mathcal{B}_k), \tag{4.8}$$

and $C_n^{(-1)}(t) = \inf\{x : C_n(x) > t\}$. The following Theorem, proved in [67], shows that these three processes converge in $\mathscr{D}_{[0,+\infty)}$ and that T_n and $C_n^{(-1)}$ are equiva-

[2] See Chapter 6 of [60] or Chapter 3 in [20] for more details about this space.

lent.

Theorem 4.2.1. *Assume* $\boldsymbol{X}$ *is a* β*-null recurrent atomic Markov chain. Then,*

i) $C_n \xrightarrow{\mathscr{D}_{[0,+\infty)}} S_\beta$ *where* S_β *is the one-sided stable Levy process defined by the marginal characteristics*

$$\mathbb{E}\left[\exp\left(isS_\beta\left(t\right)\right)\right] = \exp\left(is^\beta t\right) \;\; s \in \mathbb{R}, t \in [0, +\infty].$$

ii) $C_n^{(-1)}$ *and* T_n *are equivalent processes and both converge in* $\mathscr{D}_{[0,+\infty)}$ *to the Mittag-Leffler process of parameter* β*.*

Remark 4.2.1. *The Mittag-Leffler process with parameter* β *is defined as the inverse of* S_β*. It is a strictly increasing continuous stochastic process defined as*

$$M_\beta\left(t\right) = t^\beta M_\beta\left(1\right) \quad , \quad \mathbb{E}\left(M_\beta^m\left(1\right)\right) = \frac{m!}{\Gamma\left(1+m\beta\right)} \;\; m \geqslant 0.$$

Theorem 4.2.1 shows a striking difference between positive and null recurrent Markov chains. While in the former the existence of moments for $\ell\left(B_j\right)$ implies that C_n and T_n (taking $u\left(n\right) = n$) converge almost surely respectively to $t\mathbb{E}_{\boldsymbol{\alpha}}\tau_{\boldsymbol{\alpha}}$ and $\frac{t}{\mathbb{E}_{\boldsymbol{\alpha}}\tau_{\boldsymbol{\alpha}}}$, and therefore, $T\left(n\right)$ can be approximated almost surely by the deterministic quantity n, in the latter, we only have weak convergence, hence $T\left(n\right)$ can only be controlled by the deterministic quantity $u(n)$ in distribution.

4.2.3 Properties of linear functionals defined on β-null recurrent chains

For a measurable function $f : E \to \mathbb{R}$, and an atomic Markov chain $\mathbf{X}$ with an accessible atom $\boldsymbol{\alpha}$, consider the problem of estimating $\pi_{\boldsymbol{\alpha}}(f) = \int f d\pi_{\boldsymbol{\alpha}}$, where $\pi_{\boldsymbol{\alpha}}$

is as in (4.3) and $\pi_{\boldsymbol{\alpha}}(f) < +\infty$. Denote by $S_n(f)$ the partial sums of f over the chain, that is

$$S_n(f) = \sum_{k=0}^{n} f(X_k). \tag{4.9}$$

The Ratio Limit Theorem for atomic chains [38, Theorem 6.6.2] shows that if g is a measurable function, then, for every invariant measure π we have

$$\frac{S_n(f)}{S_n(g)} \xrightarrow{\text{a.s.}} \frac{\pi(f)}{\pi(g)}, \tag{4.10}$$

as long as $\pi(g) \neq 0$.

Remark 4.2.2. *From (4.10) is clear that $\frac{S_n(f)}{T(n)}$ is a strongly consistent estimator of $\pi_{\boldsymbol{\alpha}}(f)$, and, in the positive recurrent case, $\frac{S_n(f)}{n} \xrightarrow{a.s.} \frac{\pi_{\boldsymbol{\alpha}}(f)}{\mathbb{E}_{\boldsymbol{\alpha}}\tau_{\boldsymbol{\alpha}}}$. In the null recurrent case, however, $\frac{S_n(f)}{n} \xrightarrow{a.s.} 0$ (see Corollary 6.6.3 in [38]) and there is no deterministic sequence $a(n)$ such that $\frac{S_n(f)}{a(n)}$ converges almost surely to a non-zero limit [28].*

Given that our interest in this paper is to apply the bootstrap method to the estimation of $\pi_{\boldsymbol{\alpha}}(f)$ we need to find a series of i.i.d. random variables whose mean strongly converges to $\pi_{\boldsymbol{\alpha}}(f)$. To do this, define the following random variables

$$f(\mathcal{B}_j) = \begin{cases} \sum_{i=0}^{\tau_{\alpha}} f(X_i) & , \quad j = 0 \\ \sum_{i=\tau_{\boldsymbol{\alpha}}(j)+1}^{\tau_{\alpha}(j+1)} f(X_i) & , \quad j \geqslant 1 \end{cases}.$$

The strong Markov property implies that under $\mathbb{P}_{\boldsymbol{\alpha}}$, the sequence $\{f(\mathcal{B}_j)\}_{j\geqslant 0}$ is i.i.d. Moreover, for every initial measure $\lambda \in \mathscr{M}(\mathcal{E})_+$ such that $\mathbb{P}_\lambda(\tau_{\boldsymbol{\alpha}} < \infty) = 1$, the random variables $f(\mathcal{B}_j), j \geqslant 0$ are independent and for $j \geqslant 1$ they are i.i.d. Therefore, $S_n(f)$ can now be written as a sum of independent random variables as follows:

$$S_n(f) = f(\mathcal{B}_0) + \sum_{j=1}^{T(n)} f(\mathcal{B}_j) + \sum_{i=\tau_{\boldsymbol{\alpha}}(T(n)+1)+1}^{n} f(X_i), \tag{4.11}$$

with the convention that the sum of an empty set is 0. As customary in the β-null recurrent literature, we will denote the last term in (4.11) by $f\left(\mathcal{B}_{(n)}\right)$.

Equation (4.3) indicates that

$$\mathbb{E}_{\boldsymbol{\alpha}} f(\mathcal{B}_j) = \pi_{\boldsymbol{\alpha}}(f), \quad j = 1, \ldots. \tag{4.12}$$

hence, if the assume that $\pi_{\boldsymbol{\alpha}}(|f|) < +\infty$, the Law of Large Numbers for randomly indexed sequences [55, Theorem 8.2, pp 302] shows that

$$\frac{1}{T(n)} \sum_{j=1}^{T(n)} f(\mathcal{B}_j) \xrightarrow{\text{a.s.}} \pi_{\boldsymbol{\alpha}}(f). \tag{4.13}$$

Remark 4.2.3. *The almost sure convergence of both* $\frac{S_n(f)}{T(n)}$ *and* $\frac{1}{T(n)}\sum_{j=1}^{T(n)} f(\mathcal{B}_j)$ *to* $\pi_{\boldsymbol{\alpha}}(f)$ *and the decomposition* (4.11) *shows that* $\frac{f(\mathcal{B}_0)}{T(n)}$ *and* $\frac{f\left(\mathcal{B}_{(n)}\right)}{T(n)}$ *both converge almost surely to 0. This allow us to only consider in our estimations the i.i.d. blocks* $f(\mathcal{B}_j), j \geqslant 1$.

If we suppose further that $f(\mathcal{B}_1)$ has finite second moment and we denote by σ^2 the variance of $f(\mathcal{B}_1)$, then

$$\frac{1}{T(n)} \sum_{j=1}^{T(n)} \left(f(\mathcal{B}_j) - \frac{1}{T(n)} \sum_{i=1}^{T(n)} f(\mathcal{B}_i) \right)^2 \xrightarrow{\text{a.s.}} \sigma^2. \tag{4.14}$$

Much of the work carried out in this investigation deals with sequences indexed by the sequence of random variables $T(n)$. As explained at the end of Section 4.2.2, this sequence, although it converges almost surely to $+\infty$, can not be deterministically approximated in probability, it only admits an approximation in distribution.

This creates huge problems, even for simple tasks, as to obtaining a CLT, because, CLTs for randomly indexed sequences (see [4] for the original formulation and Th. 17.2 in [20] for its more general form) require being able to control deterministically, at least in probability, the sequence of the number of terms. The result we present below, extends this CLT, replacing the requirement of the control in probability by the existence of the limit of a stochastic process defined in terms of the sequence of the number of terms.

Lemma 4.2.1 (CLT for randomly indexed sequences)**.** *Let $X_1, X_2 \dots$ be i.i.d. random variables such that $E(X_1) = \mu$ and $\operatorname{Var} X_1 = \sigma^2 > 0$. Let $N(n)$ be a sequence of integer-valued random variables. If there exists an unbounded increasing sequence of real numbers u_n such that the process $N_n(t) = \frac{N(\lfloor nt \rfloor)}{u_n}$ satisfy the following conditions:*

- *Exists a process S_n in $\mathscr{D}_{[0,+\infty)}$ that is non-negative and non-decreasing for each n.*
- *$S_n \xrightarrow{\mathscr{D}_{[0,+\infty)}} S$ where S is a strictly increasing non-negative process with independent increments, no fixed jumps, and $S(0) \equiv 0$.*
- *N_n is equivalent to $S_n^{(-1)}$.*

Then, N_n converges to $S^{(-1)}$,

$$\sqrt{N(n)} \left(\frac{\sum_{j=1}^{N(n)} (X_j - \mu)}{N(n)\sigma} \right), \tag{4.15}$$

converges weakly to a standard Normal distribution and this distribution is independent of $S^{(-1)}(1)$.

Corollary 4.2.1. *[Theorem 17.2 in [20]] Suppose $X_1, \ldots, X_n$ are i.i.d. with $\mathbb{E}X_1 = \mu$ and $\operatorname{Var} X_1 = \sigma^2$. If $N(n)$ is a sequence of integer-valued random variables such that*

$$\frac{N(n)}{u_n} \xrightarrow{p} \theta, \tag{4.16}$$

where θ is a positive random variable and the u_n is sequence of positive numbers going to infinity, then

$$\sqrt{N(n)} \left(\frac{\sum_{j=1}^{N(n)} (X_j - \mu)}{N(n)\sigma} \right)$$

converges in distribution to a standard normal random variable.

Using Lemma 4.2.1 and Theorem 4.2.1 we can provide a different proof of the following Central Limit Theorem for β-null recurrent atomic Markov chains, which was originally proved in [8].

Proposition 4.2.2. *Let $\boldsymbol{X}$ be a β-null recurrent Markov chain, with an accessible atom $\boldsymbol{\alpha}$. For every $\pi_{\boldsymbol{\alpha}}$- measurable function f such that $\mathbb{E}\left[(f(\mathcal{B}_1))^2\right] < +\infty$,*

$$\sqrt{T(n)} \left(\frac{\sum_{j=1}^{T(n)} f(\mathcal{B}_j)}{T(n)} - \int f d\pi_{\boldsymbol{\alpha}} \right) \tag{4.17}$$

converges in distribution to a Normal random variable with mean 0 and variance σ^2. Moreover, $\frac{T(n)}{n^\beta L(n)}$ converges to a Mittag-Leffler distribution with parameter β that is independent of the limiting distribution of (4.17).

The following corollary follows directly by Proposition 4.2.2, equation (4.14) and Slutsky's theorem.

Corollary 4.2.2. *Under the same hypothesis of Proposition 4.2.2,*

$$\frac{\sqrt{T(n)}}{s_n}\left(\frac{\sum_{j=1}^{T(n)} f(\mathcal{B}_j)}{T(n)} - \int f d\pi_{\boldsymbol{\alpha}}\right) \tag{4.18}$$

converges weakly to a Normal distribution with mean 0 *and variance* 1. *Here* $s_n^2 = \frac{1}{T(n)}\sum_{j=1}^{T(n)}\left(f(\mathcal{B}_j) - \frac{1}{T(n)}\sum_{j=1}^{T(n)} f(\mathcal{B}_j)\right)^2$.

4.3 The regenerative block-bootstrap algorithm

Let $\mathbf{X}^{(n)} = (X_0, ..., X_n)$ be observations drawn from a β-null recurrent Markov chain $\mathbf{X}$ with an *a priori* known accessible atom $\boldsymbol{\alpha}$. As in the previous section, let f be a $\pi_{\boldsymbol{\alpha}}$-integrable function such that $f(\mathcal{B}_1)$ has a finite second moment. Denote by σ^2 the variance of $f(\mathcal{B}_1)$.

The bootstrap method we study in this section was introduced in [15] for positive recurrent Markov chains. In the atomic case, it was shown to have a uniform rate of convergence of $O_p(n^{-1})$ under mild conditions.

In this section, we show that the method is also applicable in the β-null recurrent case, although, we have not been able to obtain a rate.

Proposition 3.1 in [14] shows that for positive recurrent chains, in the nonstationary case (when the initial law λ is not the invariant probability measure), the first data block $\mathcal{B}_0$ induces a bias of order $O(n^{-1})$, which cannot be estimated from a single realization $\mathbf{X}^{(n)}$ of the chain starting from λ. The last block $\mathcal{B}_{(n)}$ (which is incomplete) induces a first-order term in the bias too. This led the authors in [15] to only consider statistics based on the regenerative data blocks $\mathcal{B}_1, ..., \mathcal{B}_{T(n)}$.

In the β-null recurrent case, the lack of finite first moment for the block sizes

suggests that considering the non-regenerative blocks will incur in an even worst bias, hence, as in [15], we will only consider statistics based on the regenerative data blocks $\mathcal{B}_1, ..., \mathcal{B}_{T(n)}$.

While our asymptotic results are specifically stated for integrals with respect to the invariant measure, the algorithm can be applied to a broader range of statistics G_n that have an appropriate standardization S_n. This includes non-degenerate U-statistics and differentiable functionals.

The RBB procedure is performed in four steps as follows:

1. Count the number of visits $T_n(\boldsymbol{\alpha})$ to the atom $\boldsymbol{\alpha}$ up to time n. And divide the observed sample path $\mathbf{X}^{(n)} = (X_0, ..., X_n)$ into $T_n(\boldsymbol{\alpha}) + 1$ blocks, $\mathcal{B}_0$, $\mathcal{B}_1, ..., \mathcal{B}_{T_n(\boldsymbol{\alpha})-1}, \mathcal{B}^{(n)}_{T_n(\boldsymbol{\alpha})}$ valued in the torus $\mathbb{T} = \cup_{n=1}^{\infty} E^n$, corresponding to the pieces of the sample path between consecutive visits to the atom $\boldsymbol{\alpha}$. Drop the first and last (non-regenerative) blocks. Denote by $T(n)$ the number of remaining blocks.

2. Draw sequentially bootstrap data blocks $\mathcal{B}^*_{1,T(n)}, ..., \mathcal{B}^*_{k,T(n)}$ independently from the empirical distribution $F_n = T(n)^{-1} \sum_{j=1}^{T(n)} \delta_{\mathcal{B}_j}$ of the blocks $\{\mathcal{B}_j\}_{1 \leqslant j \leqslant T(n)}$ conditioned on $\mathbf{X}^{(n)}$, until the length $\ell^*(k) = \sum_{j=1}^{k} \ell(\mathcal{B}^*_{j,T(n)})$ of the bootstrap data series is larger than n. Let $T^*_n(\boldsymbol{\alpha}) = \inf\{k \geqslant 1, \ell^*(k) > n\}$ and $T^*(n, T(n)) = T^*_n(\boldsymbol{\alpha}) - 1$.

3. From the data blocks generated in step 2, reconstruct a pseudo-trajectory of size $l^*(T^*(n, T(n)))$ by binding the blocks together

$$X^{*(n)} = (\mathcal{B}^*_{1,T(n)}, ..., \mathcal{B}^*_{T^*(n,T(n)),T(n)}).$$

Compute the *RBB statistic* $G^*_n = G_n(X^{*(n)})$.

4. If $S_n = S(\mathcal{B}_1, ..., \mathcal{B}_{T(n)})$ is an appropriate standardization of the original statistic G_n, compute $S_n^* = S(\mathcal{B}^*_{1,T(n)}, ..., \mathcal{B}^*_{T^*(n,T(n)),T(n)})$.

The *RBB distribution* is then given by

$$H_{RBB}(x) = \mathbb{P}^*\left(S_n^{*-1}\left(G_n^* - G_n\right) \leqslant x\right)$$

where $\mathbb{P}^*(\bullet) = \mathbb{P}\left(\bullet \mid \mathbf{X}^{(n)}\right)$ denotes the conditional probability given $\mathbf{X}^{(n)}$.

Our main asymptotic result, in the case of integrals concerning the invariant measure, is the following.

Theorem 4.3.1. *Let $\boldsymbol{X}$ be a β-null recurrent Markov chain with an accessible atom $\boldsymbol{\alpha}$, and let f be a $\pi_{\boldsymbol{\alpha}}$-integrable function such that $\mathbb{E}\left[\left(f\left(\mathcal{B}_1\right)\right)^2\right] < +\infty$. Then we have,*

$$\sqrt{T^*\left(n, T\left(n\right)\right)}\left(\frac{\sum_{j=1}^{T^*(n,T(n))}\left(f(\mathcal{B}^*_{j,T(n)}) - \frac{1}{T(n)}\sum_{i=1}^{T(n)} f(\mathcal{B}_i)\right)}{T^*\left(n, T\left(n\right)\right)\sigma_{T(n)}}\right) \xrightarrow{d^*} N\left(0,1\right),$$

in probability along the data, where d^* denotes the convergence in distribution conditionally to the data and

$$\sigma^2_{T(n)} = \frac{1}{T\left(n\right)}\sum_{j=1}^{T(n)}\left(f(\mathcal{B}_j) - \frac{1}{T\left(n\right)}\sum_{i=1}^{T(n)} f(\mathcal{B}_i)\right)^2.$$

This theorem yields that the bootstrap distribution of the standardized sum has asymptotically the same distribution as the statistics $\frac{\sum_{j=1}^{T(n)} f(\mathcal{B}_j)}{T(n)}$ estimating $\int f d\pi_{\boldsymbol{\alpha}}$. The regenerative block bootstrap is thus first-order correct. In particular, this justifies the use of the quantiles of the bootstrap distribution (with or without stan-

dardizing) to obtain confidence intervals for $\int f d\pi_{\boldsymbol{\alpha}}$.

4.4 The regeneration-based bootstrap algorithm

In this section, we adapt the *Regeneration-base bootstrap* to the β-null recurrent Markov chain scenario.

Similarly to Section 4.3, consider observations $\mathbf{X}^{(n)} = (X_0, \ldots, X_n)$ drawn from a β-null recurrent Markov chain $\mathbf{X}$ that has an accessible atom $\boldsymbol{\alpha}$ known beforehand. Suppose that f is a function such $\pi_{\boldsymbol{\alpha}}(f)$ is finite and the second moment of $f(\mathcal{B}_1)$ is also finite. Let σ^2 represent the variance of $f(\mathcal{B}_1)$.

The algorithm we present in this section was introduced in [6, 34] for positive recurrent Markov chains with an accessible known atom. Similarly to the RBB, it consists on dividing the chain into $\mathcal{B}_1, \ldots, \mathcal{B}_{T(n)}$ regenerative blocks and then resampling blocks to form the empirical distribution of $\mathcal{B}_1, \ldots, \mathcal{B}_{T(n)}$. The main difference between the Regeneration-based bootstrap and the RBB is that in the former, the number of bootstrapped blocks is $T(n)$, hence, non-random conditionally to $\mathbf{X}^{(n)}$, while in the latter is random.

The full algorithm is as follows:

1. Count the number of visits $T_n(\boldsymbol{\alpha})$ to the atom $\boldsymbol{\alpha}$ up to time n. And divide the observed sample path $\mathbf{X}^{(n)} = (X_0, ..., X_n)$ into $T_n(\boldsymbol{\alpha}) + 1$ blocks, $\mathcal{B}_0$, $\mathcal{B}_1$, ..., $\mathcal{B}_{T_n(\boldsymbol{\alpha})-1}$, $\mathcal{B}^{(n)}_{T_n(\boldsymbol{\alpha})}$ valued in the torus $\mathbb{T} = \cup_{n=1}^{\infty} E^n$, corresponding to the pieces of the sample path between consecutive visits to the atom $\boldsymbol{\alpha}$. Drop the first and last (non-regenerative) blocks. Denote by $T(n)$ the number of remaining blocks.

2. Draw $T(n)$ bootstrap data blocks $\mathcal{B}^*_{1,T(n)}, ..., \mathcal{B}^*_{T(n),T(n)}$ independently from

the empirical distribution $F_n = T(n)^{-1} \sum_{j=1}^{T(n)} \delta_{\mathcal{B}_j}$ of the blocks $\{\mathcal{B}_j\}_{1 \leqslant j \leqslant T(n)}$ conditioned on $\mathbf{X}^{(n)}$.

3. From the bootstrap data blocks generated at step 2, reconstruct a trajectory by binding the blocks together, getting the reconstructed sample path

$$X^{*(n)} = (\mathcal{B}^*_{1,T(n)}, ..., \mathcal{B}^*_{T(n),T(n)}).$$

Compute the statistic $G^*_n = G_n\left(X^{*(n)}\right)$.

4. If $S_n = S(\mathcal{B}_1, \ldots, \mathcal{B}_{T(n)})$ is an appropriate standardization of the original statistic G_n, compute $S^*_n = S(\mathcal{B}^*_{1,T(n)}, \ldots, \mathcal{B}^*_{T(n),T(n)})$.

As in the RBB case, the asymptotic result stated below shows the validity of this bootstrap scheme when used in estimations of integrals with respect to the invariant measure.

Theorem 4.4.1. *Let $\boldsymbol{X}$ be a β-null recurrent Markov chain with an accessible atom $\boldsymbol{\alpha}$, and let f be a $\pi_{\boldsymbol{\alpha}}$-integrable function such that $\mathbb{E}\left[(f(\mathcal{B}_1))^2\right] < +\infty$, then*

$$\sqrt{T(n)} \left(\frac{\sum_{j=1}^{T(n)} \left(f(\mathcal{B}^*_{j,T(n)}) - \frac{1}{T(n)} \sum_{i=1}^{T(n)} f(\mathcal{B}_i) \right)}{T(n)\, \sigma_{T(n)}} \right) \xrightarrow{d^*} N(0,1),$$

almost surely along the data, where d^* denotes the convergence in distribution conditionally to the data and

$$\sigma^2_{T(n)} = \frac{1}{T(n)} \sum_{j=1}^{T(n)} \left(f(\mathcal{B}_j) - \frac{1}{T(n)} \sum_{i=1}^{T(n)} f(\mathcal{B}_i) \right)^2.$$

Remark 4.4.1. *In its original formulation for the positive recurrent case, the estimator used was* $\frac{S_n(f)}{n}$*, however, by Remark 4.2.2, this can not be done in the null recurrent case, hence, we need to use* $\frac{1}{T(n)}\sum_{j=1}^{T(n)} f(\mathcal{B}_i)$.

4.5 Simulations

To illustrate the convergence of the regenerative bootstraps method described in the previous two sections we will do the following simulation experiment.

Take **X** as the simple symmetric random walk in $\mathbb{Z}$, that is

$$X_t = \begin{cases} 0 & , \quad t = 0 \\ \sum_{k=1}^{t} Y_i & , \quad t \geqslant 1 \end{cases} \tag{4.19}$$

with $P(Y_i = 1) = P(Y_i = -1) = \frac{1}{2}$. In this random walk, the state 0 is an atom and the invariant measure is $\pi_0(i) \equiv 1$ (see pp.1143 in [8]). Consider the function $f(k) = \frac{1}{k^2}$ if $k \neq 0$ and $f(0) = 0$, then

$$\int f(x)\, d\pi(x) = 2\sum_{k=1}^{+\infty} \frac{1}{k^2} = \frac{\pi_0^2}{3}.$$

In order to show the validity of the proposed methods, we have simulated the first 10^8 points of a simple symmetric random walk (see figure 4.1). Using this data, we have applied both the RBB and the Regeneration Based-bootstrap 1000 times each and computed the values of

$$Z^*_{RBB} = \sqrt{T^*(n, T(n))}\left(\frac{\sum_{j=1}^{T^*(n,T(n))}\left(f(\mathcal{B}^*_{j,T(n)}) - \frac{1}{T(n)}\sum_{i=1}^{T(n)} f(\mathcal{B}_i)\right)}{T^*(n, T(n))\,\sigma_{T(n)}}\right),$$

$$Z^*_{RegBB} = \sqrt{T(n)} \left(\frac{\sum_{j=1}^{T(n)} \left(f(\mathcal{B}^*_{j,T(n)}) - \frac{1}{T(n)} \sum_{i=1}^{T(n)} f(\mathcal{B}_i) \right)}{T(n)\, \sigma_{T(n)}} \right).$$

Figure 4.2 shows the validity of both methods, despite the fact that we observed a huge block (52% of the whole trajectory is inside this block) and 25% of the realization is in the final incomplete block. The 95% confidence interval for $\int f d\pi_0$ using the RBB is $(3.1439, 3.3096)$ and using the regenerative based-bootstrap is $(3.1434, 3.3067)$. Notice that the true value of $\int f d\pi_0$ is 3.2899, while the estimation obtained using (4.13) is 3.2226, and the confidence interval obtained via Proposition 4.2.2 is $(3.1432, 3.302)$.

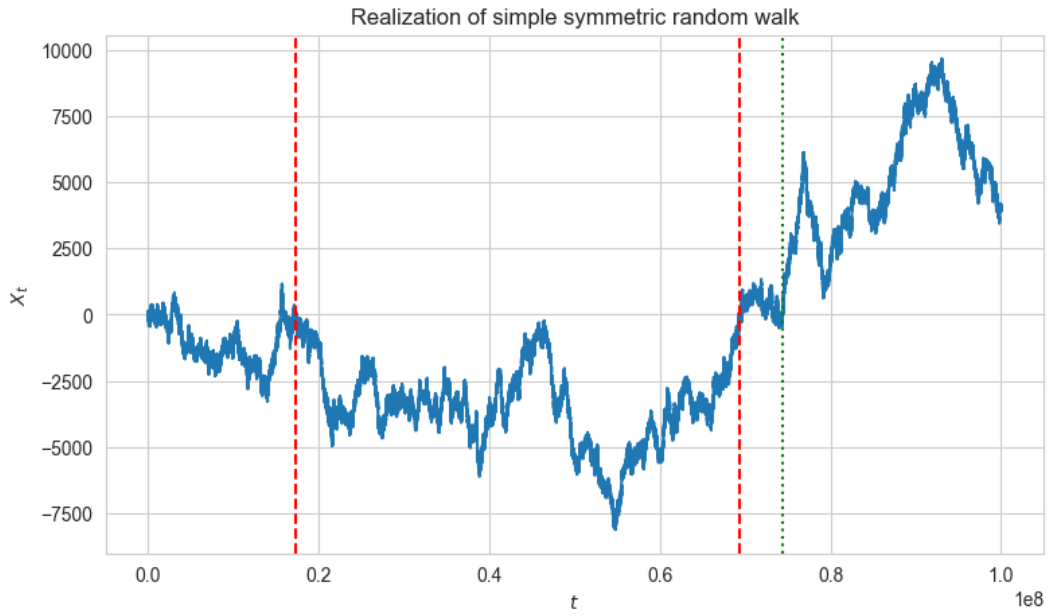

Figure 4.1: First 10^8 points of a realization of a simple symmetric random walk starting at 0. There are 9406 complete blocks in this realization. The red dashed lines delimit the largest block, while the green dotted line marks the end of the last complete block.

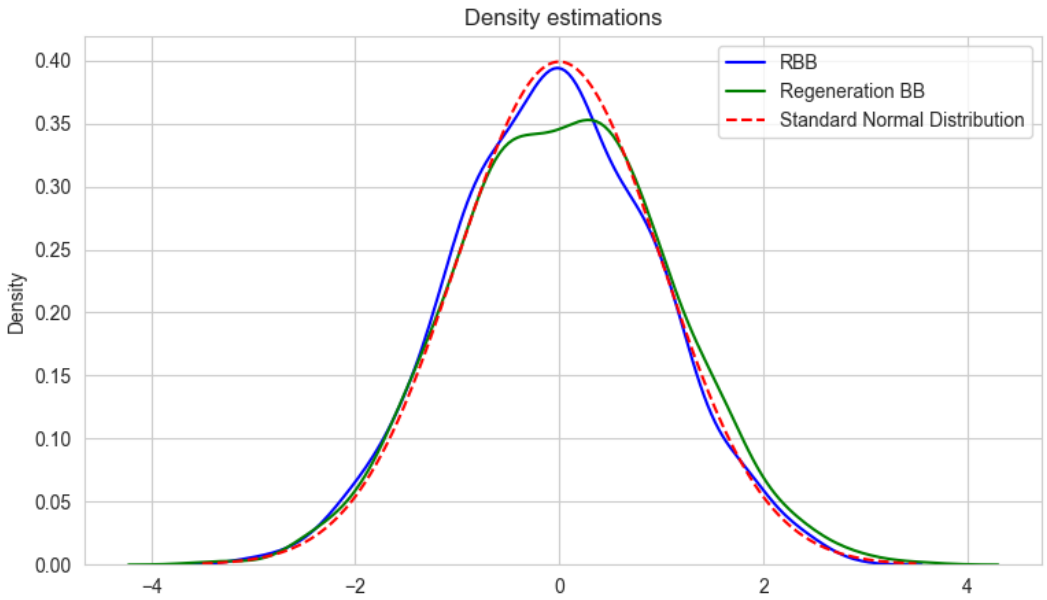

Figure 4.2: Density estimation of the bootstrap distributions Z^*_{RBB} and Z^*_{RegBB} after 10^3 simulations.

4.6 Proofs

4.6.1 Proof of Lemma 4.2.1

For the proof of Lemma 4.2.1 we need the following result, which appears as part A.3 of Theorem A.1 in [67].

Lemma 4.6.1. *Let A_n and B_n be a pair of stochastic processes which are càdlàg, where A_n is non-negative and non-decreasing. Let B denote a Brownian motion defined for $t \in \mathbb{R}$ and let A denote a strictly increasing non-negative process with independent increments, $A(0) \equiv 0$ and with no fixed jumps. Assume that $B_n \xrightarrow{\mathscr{D}_{[0,+\infty)}} B$ and $A_n \xrightarrow{\mathscr{D}_{[0,+\infty)}} A$. Then, $A_n^{(-1)} \xrightarrow{\mathscr{D}_{[0,+\infty)}} A^{(-1)}$ and*

$$\left(A_n^{(-1)}(t), \frac{B_n \circ A_n^{(-1)}(t)}{\sqrt{A_n^{(-1)}(t)}} \right) \xrightarrow{d} \left(A^{(-1)}(t), Z \right) \quad \forall t \in (0,1],$$

where Z is standard normal variable independent of $A^{(-1)}(t)$.

To prove Lemma 4.2.1, let $W_k = \sigma^{-1}\big(X_k - \mu\big)$, then $\{W_k\}_{k=1}^{\infty}$ is an i.i.d. sequence with $\mathbb{E}(W_k) = 0$ and $\operatorname{Var} W_k = 1$ for all k.

Let's define the following continuous time process for $t \geqslant 0$

$$Q_n(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^{\lfloor nt \rfloor} W_k. \tag{4.20}$$

By Theorem 23 and Example 24 in [103], $Q_n \xrightarrow{\mathscr{D}_{[0,+\infty)}} B$ and given that u_n is an unbounded increasing sequence, we also have Q_{u_n} converges weakly to B in $\mathscr{D}_{[0,+\infty)}$.

The conditions imposed to the process N_n allow us to apply Lemma 4.6.1 with $A_n = S_n$ and $B_n = Q_{u_n}$. Taking into account that N_n is equivalent to $S_n^{(-1)}$ we obtain that for all $t > 0$

$$\frac{Q_{u_n}\left(N_n\left(t\right)\right)}{\sqrt{N_n\left(t\right)}} \xrightarrow{d} N\left(0,1\right). \tag{4.21}$$

Using that $N\left(\lfloor nt \rfloor\right) = u_n N_n(t)$, we get

$$Q_{u_n}(N_n(t)) = \frac{\sigma^{-1}}{\sqrt{u_n}} \sum_{j=1}^{N(\lfloor nt \rfloor)} (X_j - \mu), \tag{4.22}$$

and Lemma 4.2.1 follows after plugging (4.22) into (4.21) and taking $t = 1$.

4.6.2 Proof of Corollary 4.2.1

We assume, at first, that θ is bounded, that is, there exists a constant K such that $0 < \theta < K$ with probability 1. Without loss of generality, assume the u_n are

integers. Define the process

$$N_n(t) = \begin{cases} \frac{tN(n)}{u_n} & , \text{ if } \frac{N(n)}{u_n} < 1 \\ t\theta & , \text{ otherwise} \end{cases}$$

As stated in pp. 147 of [20], this process converges to the process $t\theta$ and trivially satisfies the conditions of Lemma 4.2.1 (using $S_n(t) = \frac{t}{\theta}$, $S_n^{-1}(t) = t\theta$).

The case when K is unbounded can be treated by following the same argument as in pp. 148 of [20].

4.6.3 Proof of Proposition 4.2.2

Recall from Section 4.2.3 that, by the Strong Markov Property, the sequence $\{f(\mathcal{B}_j)\}_{j=1}^{+\infty}$ is i.i.d. with mean $\int f d\pi_{\boldsymbol{\alpha}}$ and variance σ^2. Consider the processes $T_n(t)$ and C_n defined in (4.8)

$$T_n(t) = \frac{T(\lfloor nt \rfloor)}{u(n)} \quad , \quad C_n(t) = \frac{1}{v(n)} \sum_{k=0}^{\lfloor nt \rfloor} \ell(\mathcal{B}_k).$$

By Theorem 4.2.1, we can apply Lemma 4.2.1 with $X_i = f(\mathcal{B}_i)$, $\mu = \int f d\pi_{\boldsymbol{\alpha}}$, $N(n) = T(n)$ and $u_n = n^{\beta} L(n)$, which completes the proof.

4.6.4 Proof of Theorem 4.3.1

Assume we have observed the chain until time n, i.e., $\mathbf{X}^{(n)} = X_0, X_1, \ldots, X_n$, and we have extracted the $T(n)$ regeneration blocks: $\mathcal{B}_1, \ldots, \mathcal{B}_{T(n)}$.

Now we start to sequentially bootstrap data blocks $\mathcal{B}^*_{1,T(n)}, \ldots, \mathcal{B}^*_{k,T(n)}$ independently from the empirical distribution $F_{T(n)} = T(n)^{-1} \sum_{j=1}^{T(n)} \delta_{\mathcal{B}_j}$ of the blocks $\{\mathcal{B}_j\}_{1 \leqslant j \leqslant T(n)}$, conditioned on $\mathbf{X}^{(n)}$, until the length $\ell^*(k) = \sum_{j=1}^{k} \ell\left(\mathcal{B}^*_{1,T(n)}\right)$ of

the bootstrap data series is larger than n.

For each m, define

$$T^{*}\left(m, T\left(n\right)\right)=\max\left\{k:\sum_{j=1}^{k}\ell\left(f(\mathcal{B}_{j,T(n)}^{*})\right)\leqslant m\right\}, \tag{4.23}$$

$$U^{*}\left(m, T\left(n\right)\right)=\sqrt{m}\left(\frac{\sum_{j=1}^{m}\left(f(\mathcal{B}_{j,T(n)}^{*})-\frac{1}{T(n)}\sum_{i=1}^{T(n)}f(\mathcal{B}_{i})\right)}{m\sigma_{T(n)}}\right). \tag{4.24}$$

Theorem 4.3.1 will be proved if we show that

$$\mathbb{P}^{*}\left(U^{*}\left(T^{*}\left(n, T\left(n\right)\right), T\left(n\right)\right)\leqslant x\right)\xrightarrow{p}\mathbb{P}\left(N\leqslant x\right)\quad\forall x\in\mathbb{R}, \tag{4.25}$$

where P is a standard normal random variable and $\mathbb{P}^{*}\left(\bullet\right)=\mathbb{P}\left(\bullet\mid\mathbf{X}^{(n)}\right)$ denotes the conditional probability given $\mathbf{X}^{(n)}$.

Given that we will bootstrap $T^{*}\left(n, T\left(n\right)\right)$ terms, which is a random quantity conditionally to the data, we will use Lemma 4.6.1 to prove (4.25). In order to do this we need, conditionally to the data:

1. Find a process $S_{n,T(n)}^{*}(t)$ that is non-negative, non-decreasing that converges in $\mathscr{D}_{[0,+\infty)}$ to a process S^{*} that is non-negative, strictly increasing, has independent increments, no fixed jumps and $S^{*}\left(0\right)\equiv 0$.

2. Show that $T_{n,T(n)}^{*}(t)=\frac{T^{*}(\lfloor nt\rfloor)}{T(n)}=\frac{T^{*}(\lfloor nt\rfloor,T(n))}{T(n)}$ is equivalent in $\mathscr{D}_{[0,+\infty)}$ to $S_{n,T(n)}^{*(-1)}$.

3. Find a process $Q_{n,T(n)}^{*}\left(t\right)$ that converges in $\mathscr{D}_{[0,+\infty)}$ to a Brownian motion

when n goes to $+\infty$. This process should satisfy, for some $t > 0$

$$U^*\left(T^*\left(n, T\left(n\right)\right), T\left(n\right)\right) = \frac{Q^*_{n,T(n)} \circ T^*_{n,T(n)}\left(t\right)}{\sqrt{T^*_{n,T(n)}\left(t\right)}}. \tag{4.26}$$

A natural choice for $Q^*_{n,T(n)}$, which satisfies (4.26) for $t = 1$, is

$$Q^*_{n,T(n)}\left(t\right) = \sqrt{T\left(n\right)} \left(\frac{\sum_{j=1}^{\lfloor T(n)t \rfloor} \left(f\left(\mathcal{B}^*_{j,T(n)}\right) - \frac{1}{T(n)} \sum_{j=1}^{T(n)} f(\mathcal{B}_i) \right)}{T\left(n\right) \sigma_{T(n)}} \right). \tag{4.27}$$

Take $S^*_{n,T(n)}(t)$ as

$$S^*_n\left(t\right) = \frac{1}{v^*\left(T\left(n\right)\right)} \sum_{i=1}^{\lfloor T(n)t \rfloor} \ell\left(\mathcal{B}^*_{i,T(n)}\right), \tag{4.28}$$

where $v^*\left(T\left(n\right)\right) = \sum_{i=0}^{T(n)} \ell\left(\mathcal{B}_i\right)$.

Following the notation of [69], let $Y_i = l\left(\mathcal{B}_i\right)$ and let $Y_{1,n} \geqslant Y_{2,n} \geqslant \ldots \geqslant Y_{n,n}$ be the order statistics of the sizes of the first n blocks, and take $Z_{k,n} = \frac{Y_{k,n}}{v(n)}$ where $v\left(n\right)$ is as in (4.7). By Theorem 1 in [69],

$$Z^{(n)} = \left(Z_{1,n}, Z_{2,n}, \ldots, Z_{n,n}, 0, \ldots, 0\right) \xrightarrow{d} \left(Z_1, Z_2, \ldots,\right) = Z, \tag{4.29}$$

where $Z_k = \left(E_1 + \cdots + E_k\right)^{-\frac{1}{\beta}}$ and E_i is a sequence of i.i.d. of exponential random variables with mean 1. By Skorokhod-Dudley-Wichura Theorem (see pp. 1171 in [69] and pp. 476 in [12]) we can choose a probability space such that, without changing the distribution of the left hand side of (4.29),

$$Z^{(n)} \xrightarrow{a.s.} Z. \tag{4.30}$$

The following Lemma shows that in that space, conditionally to the data, the process $S^*_{n,T(n)}$ converges in $\mathscr{D}_{[0,+\infty)}$.

Lemma 4.6.2. *Suppose that* (4.30) *holds, then* $\frac{T(n)}{u(n)}$ *converges almost surely to a positive random variable and*

$$S^*_{n,T(n)} \xrightarrow{\mathscr{D}_{[0,+\infty)}} S^* \quad \text{and} \quad S^{*(-1)}_{n,T(n)} \xrightarrow{\mathscr{D}_{[0,+\infty)}} S^{*(-1)} \tag{4.31}$$

almost surely along the data.

Here, $S^*(t) = KR^*(t) + t$, $R^*(t) = \sum_{j=1}^{+\infty} Z_j \left(\lambda^*_j(t) - t\right)$, $\lambda^*_j(t)$ *are independent Poisson processes with parameter* 1 *and* K *is a positive constant. Moreover, the process* S^* *is non-negative, strictly increasing, continuous, with independent increments and* $S^*(0) \equiv 0$.

Proof. When (4.30) holds, by Theorem 1 and Remark 1.3 in [79],

$$\frac{1}{v(n)} \sum_{j=1}^{n} \ell(\mathcal{B}_j) \xrightarrow{a.s.} \sum_{j=1}^{+\infty} Z_j.$$

The length of the first block, $\ell(\mathcal{B}_0)$, is finite with probability 1 and does not depend on n, hence $\frac{\ell(\mathcal{B}_0)}{v(n)}$ converges almost surely to 0. This implies that

$$\frac{1}{v(n)} \sum_{j=0}^{n} \ell(\mathcal{B}_j) \xrightarrow{a.s.} \sum_{j=1}^{+\infty} Z_j. \tag{4.32}$$

In (4.7), we defined $v(z)$ as the inverse of $u(z) = z^\beta L(z)$, then, by Proposition 1.5.15 in [21], $v(z) \sim z^{1/\beta} L_1(z)$ where L_1 is a slowly varying function, hence,

$$\frac{1}{v(n)} \sum_{j=0}^{\lfloor nt \rfloor} \ell(\mathcal{B}_j) \xrightarrow{a.s.} t^{\frac{1}{\beta}} \sum_{j=1}^{+\infty} Z_j \quad \forall t > 0. \tag{4.33}$$

For each $t > 0$, let $S_n(t) = \frac{1}{v(n)} \sum_{j=0}^{\lfloor nt \rfloor} \ell(\mathcal{B}_j)$, $S_n^{(-1)}(t) = \inf\{x > 0 : S_n(x) > t\}$ and $S(t) = t^{\frac{1}{\beta}} \sum_{j=1}^{+\infty} Z_j$, and define the three processes as 0 on $t = 0$. By (4.33) and the Continuous Mapping Theorem, $S_n^{(-1)} \xrightarrow{a.s.} S^{-1}$.

Similar to what is described on page 1141 in [8], suppose that y is such that $y < S_n^{(-1)}(1)$. Then, since $S_n(y) < 1$, it follows that $\sum_{j=0}^{\lfloor ny \rfloor} \ell(\mathcal{B}_j) < v(n)$. Consequently, we have $T(\lfloor v(n) \rfloor) \geqslant \lfloor ny \rfloor > ny - 1$. This in turn implies that $\frac{T(\lfloor v(n) \rfloor)}{n} \geqslant y - \frac{1}{n} \geqslant S_n^{(-1)}(1) - \frac{1}{n}$ for all n. In a similar way, but taking $y > S_n^{(-1)}$, we show that $\frac{T(\lfloor v(n) \rfloor)}{n} \leqslant S_n^{(-1)}(1) + \frac{1}{n}$ for all n. Then,

$$S_{u(n)^{(-1)}(1)} - \frac{1}{u(n)} \leqslant \frac{T(\lfloor v(u(n)) \rfloor)}{u(n)} \leqslant S_{u(n)^{(-1)}(1)} + \frac{1}{u(n)}. \tag{4.34}$$

The first part of the lemma now follows from (4.34), the convergence of $S_{u(n)}^{(-1)}(1)$ to $S^{-1}(1)$ and the fact that $u(v(n)) = n$ for n big enough.

To show (4.31), consider the following process, which was studied in [12],

$$Z_{m,n}^*(t) = \frac{1}{v(n)} \sum_{j=1}^{\lfloor mt \rfloor} \left(\ell\left(\mathcal{B}_{j,n}^*\right) - \frac{\sum_{i=1}^n \ell(\mathcal{B}_i)}{n} \right).$$

By Corollary 1.2 in [12] (and its proof[3]), we see that when (4.30) holds, for any m_n such that $\frac{m_n}{n} \to c$, conditionally to the data, the process $Z_{m_n,n}^*$ converges weakly in $\mathscr{D}([0,1])$ to $R^*(ct)$. Let $C > 1$, on $[0, C]$ define the process

$$W_n^*(t) = \frac{1}{v(n)} \sum_{j=1}^{\lfloor nt \rfloor} \left(\ell\left(\mathcal{B}_{j,n}^*\right) - \frac{\sum_{i=1}^n \ell(\mathcal{B}_i)}{n} \right).$$

Notice that $W_n^*(t) = Z_{nC,n}^*\left(\frac{t}{C}\right)$, hence, $W_n^* \xrightarrow{\mathscr{D}[0,C]} R^*$ as $n \to +\infty$. Because

[3] In [12], they standardize by $T_n = \max_{1 \leqslant k \leqslant n} l(\mathcal{B}_k)$ but from the proof is clear that the result remains valid if we standardize by $v(n)$ (b_n in their notation).

this convergence holds for arbitrary $C > 0$, by Lemma 1.3.ii in [68] we have that $W_n^* \xrightarrow{\mathscr{D}_{[0,+\infty)}} R^*$, and therefore, $W_{T(n)}^* \xrightarrow{\mathscr{D}_{[0,+\infty)}} R^*$.

The process $S_{n,T(n)}^*$ can be written as

$$S_{n,T(n)}^*(t) = \frac{v(T(n))}{v^*(T(n))} W_{T(n),T(n)}^*(t) + \frac{\lfloor T(n)\, t \rfloor}{T(n)} \tag{4.35}$$

Conditionally to the data, $\frac{v(T(n))}{v^*(T(n))} = \left(\frac{1}{v(T(n))} \sum_{j=0}^{T(n)} \ell(\mathcal{B}_j)\right)^{-1}$ converges to a positive constant K by equation (4.32). Equation (4.31) now follows from the convergence of $W_{T(n),T(n)}^*$ and (4.35).

The continuity of S^* was shown in pp. 466 of [12], and the rest of the properties are evident from the form of R^*. □

The next Lemma handles the equivalence of $T_{n,T(n)}^*$ and $S_{n,T(n)}^{*(-1)}$ in $\mathscr{D}_{[0,+\infty)}$.

Lemma 4.6.3. *Under the same hypothesis of Lemma 4.6.2, the processes $T_{n,T(n)}^*$ and $S_{n,T(n)}^{*(-1)}$ are equivalent in $\mathscr{D}_{[0,+\infty)}$.*

Proof. The proof of this result follows the proof of Theorem 3.2 on [67] with slight modifications.

We need to show that, for any $\varepsilon > 0$ given,

$$\mathbb{P}\left(\sup_{0<t\leqslant K} \left|T_{n,T(n)}^*(t) - S_{n,T(n)}^{*(-1)}(t)\right| > \varepsilon\right) \to 0 \quad \forall K > 0. \tag{4.36}$$

To prove this, we will show that

$$\mathbb{P}\left(\sup_{0<t\leqslant K} \left|T_{v^*(T(n)),T(n)}^*(t) - S_{n,T(n)}^{*(-1)}(t)\right| > \varepsilon\right) \to 0 \quad \forall K > 0, \tag{4.37}$$

$$\mathbb{P}\left(\sup_{0<t<K} \left|T_{v^*(T(n)),T(n)}^*(t) - T_{n,T(n)}^*(t)\right| > \varepsilon\right) \to 0 \quad \forall K > 0. \tag{4.38}$$

from where (4.36) will follow by triangular inequality.

Let $\eta > 0$

$$\begin{aligned}\left\{S_{n,T(n)}^{*(-1)}(t) < \eta\right\} &\subseteq \left\{S_{n,T(n)}^{*}(\eta) > t\right\} \\ &= \left\{\frac{1}{v^{*}(T(n))}\sum_{i=1}^{\lfloor T(n)\eta\rfloor}\ell\left(\mathcal{B}_{i,T(n)}^{*}\right) > t\right\} \\ &= \left\{\sum_{i=1}^{\lfloor T(n)\eta\rfloor}\ell\left(\mathcal{B}_{i,T(n)}^{*}\right) > tv^{*}(T(n))\right\} \\ &= \left\{\frac{T^{*}(\lfloor v^{*}(T(n))t\rfloor, T(n))}{T(n)} < \frac{\lfloor T(n)\eta\rfloor}{T(n)}\right\}. \end{aligned} \tag{4.39}$$

Because $T(v^{*}(n)) = n$, we can write,

$$T_{v^{*}(T(n)),T(n)}^{*}(t) = \frac{T^{*}(\lfloor v^{*}(T(n))\,t\rfloor, T(n))}{u^{*}(v^{*}(T(n)))} = \frac{T^{*}(\lfloor v^{*}(T(n))\,t\rfloor, T(n))}{T(n)},$$

therefore, equation (4.39) becomes

$$\left\{S_{n,T(n)}^{*(-1)}(t) < \eta\right\} \subseteq \left\{T_{v^{*}(T(n)),T(n)}^{*}(t) < \frac{\lfloor T(n)\eta\rfloor}{T(n)}\right\}. \tag{4.40}$$

Similarly, we obtain that

$$\left\{S_{n,T(n)}^{*(-1)}(t) > \eta\right\} \subseteq \left\{T_{v^{*}(T(n)),T(n)}^{*}(t) \geqslant \frac{\lfloor T(n)\eta\rfloor}{T(n)}\right\}. \tag{4.41}$$

Let $\varepsilon_1 \in (0,1)$ be fixed and take $\eta_1 < \eta_2$, then, by (4.41) and (4.40),

$$\begin{aligned}\left\{\eta_1 \leqslant S_{n,T(n)}^{*(-1)}(t) < \eta_2\right\} &\subseteq \left\{\eta_1(1-\varepsilon_1) < S_{n,T(n)}^{*(-1)}(t) < \eta_2\right\} \\ &\subseteq \left\{\frac{\lfloor T(n)\eta_1(1-\varepsilon_1)\rfloor}{T(n)} \leqslant T_{v^{*}(T(n)),T(n)}^{*}(t) < \frac{\lfloor T(n)\eta_2\rfloor}{T(n)}\right\}.\end{aligned}$$

This means, that, if $S^{*(-1)}_{n,T(n)}(t) \in [\eta_1, \eta_2)$, then

$$\frac{\lfloor T(n)\eta_1(1-\varepsilon_1)\rfloor}{T(n)} - \eta_2 < T^*_{v^*(T(n)),T(n)}(t) - S^{*(-1)}_{n,T(n)}(t) < \frac{\lfloor T(n)\eta_2\rfloor}{T(n)} - \eta_1,$$

which implies that, if $S^{*(-1)}_{n,T(n)}(t) \in [\eta_1, \eta_2)$, then

$$\left|T^*_{v^*(T(n)),T(n)}(t) - S^{*(-1)}_{n,T(n)}(t)\right| \leqslant \eta_2 - \eta_1 + \varepsilon_1\eta_1 + \frac{1}{T(n)}. \tag{4.42}$$

Let $\varepsilon > 0$ be fixed. For any s we have

$$\begin{aligned}\mathbb{P}\left(\sup_{t\leqslant K}\left|\xi^*_{n,T(n)}(t)\right| > \varepsilon\right) \leqslant\ & P\left(\sup_{t\leqslant K}\left|\xi^*_{n,T(n)}(t)\right| > \varepsilon, \sup_{t\leqslant K} S^{*(-1)}_{n,T(n)}(t) < s\right) \\ & + P\left(\sup_{t\leqslant K} S^{*(-1)}_{n,T(n)}(t) \geqslant s\right),\end{aligned}$$

where $\xi^*_{n,T(n)}(t) = T^*_{v^*(T(n)),T(n)}(t) - S^{*(-1)}_{n,T(n)}(t)$.

By (4.31),

$$\lim_{s\uparrow\infty}\lim_{n\to\infty} P\left(\sup_{t\leqslant K} S^{*(-1)}_{n,T(n)}(t) \geqslant s\right) = 0.$$

Therefore, for any $\delta > 0$ we can choose s_0 such that $\mathbb{P}\left(\sup_{t\leqslant K} S^{*(-1)}_{n,T(n)}(t) \geqslant s_0\right) \leqslant \delta$ for all n big enough. By (4.42), $\sup_{t\leqslant K} S^{*(-1)}_{n,T(n)}(t) < s_0$ implies that

$$\left|\xi^*_{n,T(n)}(t)\right| \leqslant \eta_2 - \eta_1 + \varepsilon_1\eta_1 + \frac{1}{T(n)} \quad \forall t \in [0,K] \quad , \quad \forall \varepsilon_1 \in (0,1).$$

Choose $\eta_0, \ldots, \eta_L, N_1, \varepsilon_1$ with $\eta_0 = 0 < \eta_1 < \ldots < \eta_{L-1} < \eta_L = s_0$ such that $\eta_i - \eta_{i+1} < \frac{\varepsilon}{3}$ for all i. Let $\varepsilon_1 < \frac{\varepsilon}{s_0}$ and choose N_1 such that $\frac{1}{T(N_1)} < \frac{\varepsilon}{3}$.

Notice that for all $t \in [0,K]$ there is only one $i_{n,t}$ such that $S^{*(-1)}_{n,T(n)}(t)$ belongs

to $\left[\eta_{i_{n,t}}, \eta_{i_{n,t}+1}\right)$, then, by (4.42)

$$\left|\xi^*_{n,T(n)}(t)\right| \leqslant \eta_{i_{n,t}} - \eta_{i_{n,t}+1} + \varepsilon_1\eta_1 + \frac{1}{T(n)} \leqslant \varepsilon \quad \forall t \in [0,K]\,,\ \forall n > N_1,$$

whenever $S^{*(-1)}_{n,T(n)}(t) < s_0$. This implies that

$$\mathbb{P}\left(\sup_{t\leqslant K}\left|\xi^*_{n,T(n)}(t)\right| > \varepsilon, \sup_{t\leqslant K} S^{*(-1)}_{n,T(n)}(t) < s_0\right) = 0 \quad \forall n \geqslant N_1.$$

Hence,

$$\mathbb{P}\left(\sup_{t\leqslant K}\left|\xi^*_{n,T(n)}(t)\right| > \varepsilon\right) < \delta \quad \forall n > N_1. \tag{4.43}$$

which implies (4.37).

Now we turn to the proof of (4.38).

According to the definition of v^*, $v^*(T(n)) = \sum_{i=0}^{T(n)} l(\mathcal{B}_i) \leqslant n$, therefore,

$$T^*_{v^*(T(n)),T(n)}(t) = \frac{T^*(\lfloor v^*(T(n))\,t\rfloor, T(n))}{T(n)} \leqslant \frac{T^*(\lfloor nt\rfloor, T(n))}{T(n)} = T^*_{n,T(n)}(t) \quad \forall n, t.$$

Notice that $v^*(T(n)+1) = \sum_{i=0}^{T(n)+1} l(\mathcal{B}_i) > n$, therefore,

$$T^*_{n,T(n)}(t) \leqslant T^*_{v^*(T(n)+1),T(n)}(t)\,\frac{T(n)+1}{T(n)} \quad \forall n,\ t.$$

Hence,

$$T^*_{v^*(T(n)),T(n)}(t) \leqslant T^*_n(t) \leqslant T^*_{v^*(T(n)+1)}(t)\,\frac{T(n)+1}{T(n)} \quad \forall n, t.$$

Equation (4.38) now follows from the convergence of both $T^*_{v^*(T(n)),T(n)}$ and $T^*_{v^*(T(n)+1),T(n)}$ to $S^{*(-1)}$ and the fact that $\frac{T(n)+1}{T(n)} \xrightarrow{a.s.} 1$. □

By (4.27), Lemmas 4.6.1, 4.6.2 and 4.6.3 we have that, in a space where (4.30) holds, the convergence in (4.25) holds almost surely. Therefore, in the original space we have the weakly-weakly (see pp.2550 in [26]) convergence

$$\mathbb{P}^*\left(U^*\left(T^*\left(n, T\left(n\right)\right), T\left(n\right)\right) \leqslant x\right) \xrightarrow{d} \mathbb{P}\left(N \leqslant x\right) \quad \forall x \in \mathbb{R}. \tag{4.44}$$

However, given that the right hand side of (4.44) is a constant for each x, the convergence in (4.44) can be improved to convergence in probability, which completes the proof.

4.6.5 Proof of Theorem 4.4.1

This proof follows the line of the proof of Theorem 2.1 in [19]. As in that paper, let Γ_2 be the set of distribution functions G satisfying $\int x^2 dG\left(x\right) < \infty$ and define the following notion of convergence in Γ_2

$$G_n \Rightarrow G \quad \text{iff} \quad G_n \to G \text{ weakly and } \int x^2 dG_n\left(x\right) \to \int x^2 dG\left(x\right). \tag{4.45}$$

Denote by d_2 a Mallows metric that metricizes the $\Rightarrow$ convergence in Γ_2 (see details in Section 8 of [19])

If $Y_1, \ldots, Y_n$ are i.i.d. random variables with common distribution G, denote by $G^{(m)}$ the distribution of

$$m^{-\frac{1}{2}} \sum_{j=1}^{m} \left(Y_j - \mathbb{E}Y_j\right)$$

By pp. 1198 in [19], if $G, H \in \Gamma_2$ then $G^{(m)}$ and $H^{(m)}$ are also in Γ_2 and

$$d_2\left(G^{(m)}, H^{(m)}\right) \leqslant d_2\left(G, H\right). \tag{4.46}$$

Let F be the distribution of $f(\mathcal{B}_1)$ and denote by F_n the empirical distribution function of $f(\mathcal{B}_1), \ldots, f(\mathcal{B}_n)$. By (2.1) in [19] and the fact that $T(n) \to +\infty$ a.s., $F_{T(n)} \Rightarrow F$ along almost almost all sample paths, hence, conditionally to the data

$$d_2\left(F_{T(n)}, F\right) \to 0. \tag{4.47}$$

Denote by N_σ a standard distribution with mean 0 and variance σ^2. By Proposition 4.2.2,

$$d_2\left(F^{(T(n))}, N_\sigma\right) \to 0. \tag{4.48}$$

Conditionally to the data, the distribution of

$$\sqrt{T(n)}\left(\frac{\sum_{j=1}^{T(n)}\left(f\left(\mathcal{B}^*_{j,T(n)}\right) - \frac{1}{T(n)}\sum_{i=1}^{T(n)} f\left(\mathcal{B}_i\right)\right)}{T(n)}\right)$$

is $F^{(T(n))}_{T(n)}$, then, conditionally to the data,

$$d_2\left(F^{(T(n))}_{T(n)}, N_\sigma\right) \leqslant d_2\left(F^{(T(n))}_{T(n)}, F^{(T(n))}\right) + d_2\left(F^{(T(n))}, N_\sigma\right)$$

which goes to 0 by (4.47) and (4.48). The theorem now follows by (4.45), (4.14) and Slutsky's theorem.

Chapter 5

Harris recurrent Markov chains and nonlinear monotone cointegrated models

The content of this chapter is based on [41]. It is the result of a collaboration with my advisors Cécile Durot[1] and Patrice Bertail[1].

Abstract: In this paper, we study a nonlinear cointegration-type model of the form $Z_t = f_0(X_t) + W_t$ where f_0 is a monotone function and X_t is a Harris recurrent Markov chain. We use a nonparametric Least Square Estimator to locally estimate f_0, and under mild conditions, we show its strong consistency and obtain its rate of convergence. New results (of the Glivenko-Cantelli type) for localized null recurrent Markov chains are also proved.

5.1 Introduction

The concept of *linear cointegration* refers to two time series, Z_t and X_t, that are both nonstationary and of unit root type, and where there exists a stationary

[1]MODAL'X, UMR 9023 - Université Paris Nanterre, CNRS, UPL, 200 ave de la République, F92000 Nanterre.

linear combination of X_t and Z_t. This concept was first introduced in [52] and has since been extensively studied, particularly in the field of econometrics [44, 62, 63, 98, 99].

However, the long-term relationship between the two series, Z_t and X_t, might not necessarily be linear, nor X_t be linearly generated. This has led to the study of *nonlinear cointegrated* models such as,

$$Z_t = f_0(X_t) + W_t, \tag{5.1}$$

where a nonlinear function f_0 and a nonlinearly generated input process X_t are incorporated to model the relationship between the series.

In [23, 65] a relationship like (5.1) has been studied under the assumptions that f_0 is nonlinear, X_t and W_t are independent processes, and X_t is a positive or β-null recurrent Markov chain. They have applied the Nadaraya-Watson method to estimate f_0 and established the asymptotic theory of the proposed estimator.

The problem of estimating f_0 under the Markovian assumption has also been tackled using local linear M-type estimators in [24, 80] and using advanced concepts like local time and nonlinear transformations of Brownian motion-like processes in [117, 118, 119]. A comprehensive survey of the latest advances in this problem can be found in [112].

To the best of our knowledge, the case where f_0 is subject to shape constraints has not been addressed under Markovian assumptions. Such estimators are nonlinear and therefore pose significant theoretical challenges. In the context of independent observations, constraints such as convexity, concavity, and log-concavity are known to be even more complex than monotonicity constraints (see [54, 109] and the references therein). As a result, we have chosen to initiate our study of

shape-constrained estimators in the Markovian setting by focusing on the monotone case.

In this paper, we wish to establish a nonparametric estimation theory of the nonparametric least squares estimator (LSE) for the function f_0 in the model (5.1) under the constraints that f_0 is monotone non-increasing. Here, $\{W_t\}$ is an unobserved process such that $E(W_t|X_t) = 0$ to ensure identifiability of f_0. Since a minimal condition for undertaking asymptotic analysis on $f_0(x_0)$ at a given point x_0 is that, as the number of observations on $\{X_t\}$ increases, there must be infinitely many observations in the neighborhood of x_0, the process $\{X_t\}$ will be assumed to be a Harris recurrent Markov chain (cf section 5.2).

This model is clearly very attractive in situations where monotonicity is a reasonable assumption but commonly assumed structures such as linearity or additivity are not. Indeed, this formulation, in the i.i.d. case, has found useful applications in econometrics [59], biology [81, 91], medicine [110], engineering [86] among others. However, up to our knowledge, it has not been treated under the markovian assumption on X_t.

5.1.1 The estimator

Let C be a set that its interior contains our point of interest x_0. Having observed $\{(X_t, Z_t)\}_{t=0}^n$, we denote by $T_n(C)$ the number of times that $\mathbf{X}$ visited C up to time n and by $\sigma_C(i)$ the time of the i-th visit. Then, we consider the nonparametric LSE defined as the minimizer of

$$f \mapsto \sum_{i=1}^{T_n(C)} \left(Z_{\sigma_C(i)} - f\left(X_{\sigma_C(i)}\right)\right)^2 \tag{5.2}$$

over the set of non-increasing functions f on $\mathbb{R}$. The nonparametric LSE $\widehat{f}_n$ has a well know characterization, as follows. Let m be the number of unique values of $X_{\sigma_C(1)}, \ldots, X_{\sigma_C(T_n(C))}$, and $Y_1 < \cdots < Y_m$ be the corresponding order statistics. Then, $\widehat{f}_n(Y_k)$ is the left-hand slope at $\sum_{i=1}^{T_n(C)} \mathbb{I}\{X_{\sigma_C(i)} \leqslant Y_k\}$ of the least concave majorant of the set of points

$$\left\{(0,0),\ \left(\sum_{i=1}^{T_n(C)} \mathbb{I}\{X_{\sigma_C(i)} \leqslant Y_k\},\ \sum_{i=1}^{T_n(C)} Z_{\sigma_C(i)}\mathbb{I}\{X_{\sigma_C(i)} \leqslant Y_k\}\right),\ k = 1, \ldots, m\right\}, \tag{5.3}$$

and it can be computed using simple algorithms as discussed in [10]. Thus, the constrained LSE is uniquely defined at the observation points, however, it is not uniquely defined between these points: any monotone interpolation of these values is a constrained LSE. As is customary, we consider in the sequel the piecewise-constant and left-continuous LSE that is constant on every interval $(Y_{k-1}, Y_k]$, $k = 2, \ldots, m$ and also on $(-\infty, Y_1]$ and on $[Y_m, \infty)$.

The use of a localized estimator is due to the fact that we need to control the behavior of the chain around x_0, and, to do this, we need to estimate the asymptotic "distribution" of $\mathbf{X}$ in a vecinity of x_0. For Harris recurrent Markov chains, the long-term behavior of the chain is given by its invariant measure (see Section 5.2). In the positive recurrent case, the invariant measure is finite and it can be estimated by simply considering the empirical cumulative distribution function of the X_t. However, in the null recurrent case, the invariant measure is only σ-finite, hence, we need to localize our analysis in a set big enough that the chain visits it infinitely often, but small enough that the restriction of the invariant measure to it is finite. Moreover, contrary to the bandwidth in kernel type estimators, C does not depend on n, and the rate of convergence of the estimator does not depend on C.

5.1.2 Outline

Since our paper draws quite heavily on the theory of Harris recurrent Markov chains, we have added a small introduction to the subject as well as the main results that we use throughout the paper in Section 5.2. In Section 5.3, we show that under very general assumptions, our estimator $\widehat{f}_n$ is strongly consistent, while its rate of convergence is presented in Section 5.4. In Section 5.5, we present three new results concerning Harris recurrent Markov chains that have emerged during our investigation and we believe are interesting in their own right. Section 5.6 contains an overview of the proofs of our main results, while the technical proofs are presented in Section 5.7.

5.2 Markov chain theory and notation

In this section, we present the notation and main results related to Markov chains that are needed throughout the paper. For further details, we refer the reader to [38, 87, 90].

Let $\mathbf{X} = X_0, X_1, X_2, \ldots$ be a time-homogeneous Markov Chain defined on a probability space $(E, \mathcal{E}, \mathbb{P})$ where $\mathcal{E}$ is countably generated. Let $P(x, A)$ denote its transition kernel, i.e. for $x \in E$, $A \in \mathcal{E}$ we have

$$P(x, A) = \mathbb{P}(X_{i+1} \in A \,|X_i = x), \quad i = 0, 1, \ldots$$

Let $P^n(x, A)$ denote the n-step transition probability, i.e.

$$P^n(x, A) = \mathbb{P}(X_{i+n} \in A \,|X_i = x) \quad \forall i.$$

If λ is a probability measure in $(E, \mathcal{E})$ such that $\mathcal{L}(X_0) = \lambda$, then λ is called the *initial measure* of the chain $\mathbf{X}$. A homogeneous Markov chain is uniquely identified by its kernel and initial measure.

When the initial measure of the chain is given, we will write $\mathbb{P}_\lambda$ (and $\mathbb{E}_\lambda$) for the probability (and the expectation) conditioned on $\mathcal{L}(X_0) = \lambda$. When $\lambda = \delta_x$ for some $x \in E$ we will simply write $\mathbb{P}_x$ and $\mathbb{E}_x$.

An homogeneous Markov chain is *irreducible* if there exists a σ-finite measure ϕ on $(E, \mathcal{E})$ such that for all $x \in E$ and all $A \in \mathcal{E}$ with $\phi(A) > 0$ we have $P^n(x, A) > 0$ for some $n \geqslant 1$. In this case, there exists a maximal irreducibility measure ψ (all other irreducibility measures are absolutely continuous with respect to ψ). In the following, all Markov chains are supposed to be irreducible with maximal irreducibility measure ψ.

For any set $C \in \mathcal{E}$, we will denote by σ_C and τ_C, respectively, the times of first visit and first return of the chain to the set C, i.e. $\tau_C = \inf\{n \geqslant 1 : X_n \in C\}$ and $\sigma_C = \inf\{n \geqslant 0 : X_n \in C\}$. The subsequent visit and return times $\sigma_C, \tau_C(k)$, $k \geqslant 1$ are defined inductively as follows:

$$\tau_C(1) = \tau_C \quad , \quad \tau_C(k) = \min\{n > \tau_C(k-1) : X_n \in C\}, \tag{5.4}$$

$$\sigma_C(1) = \sigma_C \quad , \quad \sigma_C(k) = \min\{n > \sigma_C(k-1) : X_n \in C\}. \tag{5.5}$$

Given that our methods will only deal with the values of $\mathbf{X}$ in a fixed set C, if A is a measurable set, we will write $\mathbb{I}_C\{X_t \in A\}$ instead of $\mathbb{I}\{X_t \in A \cap C\}$ and if $A = E$, then we will simply write $\mathbb{I}_C(X_t)$.

We will use $T_n(C)$ to denote the random variable that counts the number of times the chain has visited the set C up to time n, that is $T_n(C) = \sum_{t=0}^{n} \mathbb{I}_C(X_t)$. Similarly, we will write $T(C)$ for the total of numbers of visits the chain $\mathbf{X}$ to C.

The set C is called *recurrent* if $\mathbb{E}_x T(C) = +\infty$ for all $x \in C$ and the chain $\mathbf{X}$ is recurrent if every set $A \in \mathcal{E}$ such that $\psi(A) > 0$ is recurrent.

Although recurrent chains possess many interesting properties, a stronger type of recurrence is required in our analysis. An irreducible Markov chain is *Harris recurrent* if for all $x \in E$ and all $A \in \mathcal{E}$ with $\psi(A) > 0$ we have

$$\mathbb{P}(X_n \in A \text{ infinitely often} \,|X_0 = x) = 1.$$

An irreducible chain possesses an accessible atom, if there is a set $\boldsymbol{\alpha} \in \mathcal{E}$ such that for all x, y in $\boldsymbol{\alpha}$: $P(x, .) = P(y, .)$ and $\psi(\boldsymbol{\alpha}) > 0$. Denote by $\mathbb{P}_{\boldsymbol{\alpha}}$ and $\mathbb{E}_{\boldsymbol{\alpha}}(.)$ the probability and the expectation conditionally to $X_0 \in \boldsymbol{\alpha}$. If $\mathbf{X}$ possesses an accessible atom and is Harris recurrent, the probability of returning infinitely often to the atom $\boldsymbol{\alpha}$ is equal to one, no matter the starting point, i.e. $\forall x \in E, \mathbb{P}_x(\tau_{\boldsymbol{\alpha}} < \infty) = 1$. Moreover, it follows from the *strong Markov property* that the sample paths may be divided into independent blocks of random length corresponding to consecutive visits to $\boldsymbol{\alpha}$:

$$\begin{aligned}
\mathcal{B}_0 &= \left(X_0, X_1, \ldots, X_{\tau_{\boldsymbol{\alpha}}(1)}\right) \\
\mathcal{B}_1 &= \left(X_{\tau_{\boldsymbol{\alpha}}(1)+1}, \ldots, X_{\tau_{\boldsymbol{\alpha}}(2)}\right) \\
&\ldots \\
\mathcal{B}_n &= \left(X_{\tau_{\boldsymbol{\alpha}}(n)+1}, \ldots, X_{\tau_{\boldsymbol{\alpha}}(n+1)}\right) \\
&\ldots
\end{aligned}$$

taking their values in the torus $\mathbb{T} = \cup_{n=1}^{\infty} E^n$. Notice that the distribution of $\mathcal{B}_0$ depends on the initial measure, therefore it does not have the same distribution as $\mathcal{B}_j$ for $j \geqslant 1$. The sequence $\{\tau_{\boldsymbol{\alpha}}(j)\}_{j \geqslant 1}$ defines successive times at which the chain

forgets its past, called *regeneration times*. Similarly, the sequence of i.i.d. blocks $\{\mathcal{B}_j\}_{j\geqslant 1}$ are named *regeneration blocks*. The random variable $T(n) = T_n(\boldsymbol{\alpha}) - 1$ counts the number of i.i.d. blocks up to time n. This term is called *number of regenerations up to time* n.

Notice that for any function defined on E, we can write $\sum_{t=0}^{n} f(X_t)$ as a sum of independent random variables as follows:

$$\sum_{t=0}^{n} f(X_t) = f(\mathcal{B}_0) + \sum_{j=1}^{T(n)} f(\mathcal{B}_j) + f\left(\mathcal{B}_{(n)}\right), \tag{5.6}$$

where, $f(\mathcal{B}_0) = \sum_{t=0}^{\tau_\alpha} f(X_t)$, $f(\mathcal{B}_j) = \sum_{t=\tau_{\boldsymbol{\alpha}}(j)+1}^{\tau_\alpha(j+1)} f(X_t)$ for $j = 1, \ldots, T(n)$ and $f\left(\mathcal{B}_{(n)}\right) = \sum_{t=\tau_{\boldsymbol{\alpha}}(T(n)+1)+1}^{n} f(X_t)$.

When an accessible atom exists, the *stochastic stability* properties of $\mathbf{X}$ amount to properties concerning the speed of return time to the atom only. For instance, the measure $\pi_{\boldsymbol{\alpha}}$ given by:

$$\pi_{\boldsymbol{\alpha}}(B) = \mathbb{E}_{\boldsymbol{\alpha}}\left(\sum_{n=1}^{\tau_{\boldsymbol{\alpha}}} \mathbb{I}\{X_i \in B\}\right), \quad \forall B \in \mathcal{E} \tag{5.7}$$

is invariant, i.e.

$$\pi_{\boldsymbol{\alpha}}(B) = \int P(x, B)\, d\pi_{\boldsymbol{\alpha}}(x).$$

Denote by $\mathcal{E}^+$ the class of nonnegative measurable functions with positive ψ support. A function $s \in \mathcal{E}^+$ is called *small* if there exists an integer $m_0 \geqslant 1$ and a measure $\nu \in \mathscr{M}(\mathcal{E})_+$ such that

$$P^{m_0}(x, A) \geqslant s(x)\, \nu(A) \quad \forall x \in E, A \in \mathcal{E}. \tag{5.8}$$

When a chain possesses a small function s, we say that it satisfies the *minorization*

inequality $M(m_0, s, \nu)$. As pointed out in [90], there is no loss of generality in assuming that $0 \leqslant s(x) \leqslant 1$ and $\int_E s(x)d\nu(x) > 0$.

A set $A \in \mathcal{E}$ is said to be *small* if the function $\mathbb{I}_A$ is small. Similarly, a measure ν is *small* if there exist m_0, and s that satisfy (5.8). By Theorem 2.1 in [90], every irreducible Markov chain possesses a small function and Proposition 2.6 of the same book shows that every measurable set A with $\psi(A) > 0$ contains a small set. In practice, finding such a set consists in most cases in exhibiting an accessible set, for which the probability that the chain returns to it in m steps is uniformly bounded below. Moreover, under quite wide conditions a compact set will be small, see [45].

If $\mathbf{X}$ does not possess an atom but is Harris recurrent (and therefore satisfies a minorization inequality $M(m_0, s, \nu)$), a *splitting technique*, introduced in [89, 90], allows us to extend in some sense the probabilistic structure of $\mathbf{X}$ in order to artificially construct an atom. The general idea behind this construction is to expand the sample space so as to define a sequence $(Y_n)_{n\in\mathbb{N}}$ of Bernoulli r.v.'s and a bivariate chain $\check{\mathbf{X}} = \{(X_n, Y_n)\}_{n=0}^{+\infty}$, named *split chain*, such that the set $\check{\boldsymbol{\alpha}} = (E, 1)$ is an atom of this chain. A detailed description of this construction can be found in [90].

The whole point of this construction consists in the fact that $\check{\mathbf{X}}$ inherits all the communication and stochastic stability properties from $\mathbf{X}$ (irreducibility, Harris recurrence,...). In particular, the marginal distribution of the first coordinate process of $\check{\mathbf{X}}$ and the distribution of the original $\mathbf{X}$ are identical. Hence, the splitting method enables us to establish all the results known for atomic chains to general Harris chains, for example, the existence of an invariant measure which is unique up to multiplicative constant (see Proposition 10.4.2 in [87]).

The invariant measure is finite if and only if $\mathbb{E}_{\check{\boldsymbol{\alpha}}}\tau_{\check{\boldsymbol{\alpha}}} < +\infty$, in this case we say

the chain is *positive recurrent*, otherwise, we say the chain is *null recurrent*. A null recurrent chain is called β-null recurrent (c.f. Definition 3.2 in [67]) if there exists a small nonnegative function h, a probability measure λ, a constant $\beta \in (0,1)$ and a slowly varying function L_h such that

$$\mathbb{E}_\lambda \left(\sum_{t=0}^{n} h\left(X_t\right) \right) \sim \frac{1}{\Gamma\left(1+\beta\right)} n^\beta L_h\left(n\right) \quad \text{as } n \to \infty.$$

As argued in [67], is not a too severe restriction to assume $m_0 = 1$. Therefore, throughout this paper we assume that $\mathbf{X}$ satisfies the minorization inequality $M(1, s, \nu)$, i.e, there exist a measurable function s and a probability measure ν such that $0 \leqslant s\left(x\right) \leqslant 1$, $\int_E s(x) d\nu(x) > 0$ and

$$P\left(x, A\right) \geqslant s\left(x\right) \nu\left(A\right). \tag{5.9}$$

Remark 5.2.1. *The extensions to the case where $m_0 > 1$ of the results that will be presented in this paper can be carried out (although they involve some complicated notations/proofs) using the m-skelethon or the resolvent chains, as described in [28, 30] and Chapter 17 of [87]. However, they are not treated in this paper.*

A measurable and positive function L, defined in $[a, +\infty)$ for some $a \geqslant 0$, is called *slowly varying at* $+\infty$ if it satisfies $\lim_{x\to+\infty} \frac{L(xt)}{L(x)} = 1$ for all $t \geqslant a$. See [21] for a detailed compendium of these types of functions.

It was shown in Theorem 3.1 of [67] that if the chain satisfies the minorization condition (5.9), then it's β-null recurrent if and only if

$$\mathbb{P}\left(\tau_{\check{\boldsymbol{\alpha}}} > n\right) \sim \frac{1}{\Gamma\left(1-\beta\right) n^\beta L\left(n\right)}, \tag{5.10}$$

where L is a slowly varying function.

The following theorem is a compendium of the main properties of Harris's recurrent Markov chains that will be used throughout the paper. Among other things, it shows that the asymptotic behaviour of $T(n)$ is similar to the function $u(n)$ defined as

$$u(n) = \begin{cases} n, & \text{if } \mathbf{X} \text{ is positive recurrent} \\ n^{\beta} L(n), & \text{if } \mathbf{X} \text{ is } \beta\text{-null recurrent} \end{cases}. \tag{5.11}$$

Theorem 5.2.1. [2] *Suppose $\boldsymbol{X}$ is a Harris recurrent, irreducible Markov chain, with initial measure λ, that satisfies the minorization condition (5.9). Let $T(n)$ be the number of complete regenerations until time n of the split chain $\check{\boldsymbol{X}}$, let $C \in \mathcal{E}$ be a small set and π be an invariant measure for $\boldsymbol{X}$. Then,*

1. $0 < \pi(C) < +\infty$.
2. *For any function f, defined on E, the decomposition (5.6) holds. Moreover, there is a constant K_{π}, that only depends on π, such that if $f \in L^1(E, \pi)$, then $\mathbb{E}_{\lambda} f(\mathcal{B}_1) = K_{\pi} \int_E f d\pi$.*
3. $\frac{T(n)}{T_n(C)}$ *converges almost surely to a positive constant.*
4. $\frac{T(n)}{u(n)}$ *converges almost surely to a positive constant if $\boldsymbol{X}$ is positive recurrent and converges in distribution to a Mittag-Leffler*[3] *random variable with index β if $\boldsymbol{X}$ is β-null recurrent.*

[2] Part 1 is Proposition 5.6.ii of [90], part 2 is equation (3.23) of [67], part 3 is an application of the Ratio Limit Theorem (Theorem 17.2.1 of [87]). For the positive recurrent case, part 4 also follows by the aforementioned Ratio Limit Theorem while the claim for the null recurrent case appears as Theorem 3.2 in [67].

[3] The Mittag-Leffler distribution with index β is a non-negative continuous distribution, whose moments are given by

$$\mathbb{E}\left(M_{\beta}^{m}(1)\right) = \frac{m!}{\Gamma(1 + m\beta)} \quad m \geqslant 0.$$

See (3.39) in [67] for more details.

5.3 Consistency

The aim of the section is to show that for an arbitrary x_0 in the support of f_0, the LSE $\widehat{f}_n(x_0)$ is consistent. We make the following assumptions on the processes $\mathbf{X} = \{X_t\}$ and $\mathbf{W} = \{W_t\}$.

(A1) $\mathbf{X}$ is a Harris recurrent Markov chain whose kernel $P(x, A)$ satisfies the minorization condition (5.9).

Let $\mathcal{F}_n = \sigma(\{X_0, \ldots, X_n\})$ be sigma algebra generated by the chain $\mathbf{X}$ up to time n.

(A2) For each n, the random variables $W_1, \ldots, W_n$ are conditionally independent given $\mathcal{F}_n$, $\mathbb{E}(W_t|\mathcal{F}_n) = 0$ and $\mathrm{Var}(W_t|\mathcal{F}_n) \leqslant \sigma^2$ for some $\sigma > 0$.

It follows from Assumption (A1) that the Markov Chain $\mathbf{X}$ admits a unique (up to a multiplicative constant) σ-finite invariant measure π. Let C be a set such that $0 < \pi(C) < \infty$ and $x_0 \in C$. We denote by F_n the process defined by

$$F_n(y) = \frac{1}{T_n(C)} \sum_{i=1}^{T_n(C)} \mathbb{I}\{X_{\sigma_C(i)} \leqslant y\} = \frac{1}{T_n(C)} \sum_{t=0}^{n} \mathbb{I}_C\{X_t \leqslant y\} \tag{5.12}$$

for all $y \in \mathbb{R}$, which is a localized version of the empirical distribution function of the X_t's. It is proved in Lemma 5.5.1 that F_n converges almost surely to the distribution function F supported on C and defined by

$$F(y) = \frac{\pi(C \cap (-\infty, y])}{\pi(C)}. \tag{5.13}$$

Our next two assumptions guarantee that there is a compact C, that is a small set and contains x_0 as an interior point. Sets like this can be found under very wide conditions (cf [45]).

(A3) There is $\delta = \delta(x_0)$ such that the set $C = [x_0 - \delta, x_0 + \delta]$ is *small*.

(A4) x_0 belongs to the interior of the support of X_t.

Notice that by part 1 of Theorem 5.2.1, (A3) guarantees that $\pi(C)$ is finite and positive, and hence, F is properly defined.

In addition to the assumptions on the processes $\{X_t\}$ and $\{W_t\}$, we need smoothness assumptions on F and on f_0. In particular, we will assume that F and f_0 are continuous and strictly monotone in C. This implies that f_0 and F are invertible in C, so we can find neighborhoods of $f_0(x_0)$ and $F(x_0)$ respectively, over which the inverse functions are uniquely defined. We denote by f_0^{-1} and F^{-1} respectively the inverses of f_0 and F over such a neighborhood of $f_0(x_0)$ and $F(x_0)$ respectively. The function f_0 is assumed to be monotone on its whole support.

(A5) F is locally continuous and strictly increasing in the sense that for all x' in C, for all $\varepsilon > 0$, there exists $\gamma > 0$ such that $|F^{-1}(u) - x'| > \gamma$ for all u such that $|u - F(x')| \geqslant \varepsilon$.

(A6) f_0 is non-increasing, and f_0 is locally strictly decreasing in the sense that for all x' in C, for all $\varepsilon > 0$, there exists $\gamma > 0$ such that $|f_0(x') - f_0(y)| > \gamma$ for all y such that $|y - x'| \geqslant \varepsilon$.

(A7) f_0 continuous in C.

Assumptions (A1), (A3) and (A5) ensure that X_t visits infinite times any small enough neighborhood of x_0 with probability 1, and guarantee that x_0 is not at the boundary of the recurrent states. Assumptions (A1) and (A3) and Lemma 3.2 in [67] imply that $T_n(C) \to \infty$ almost surely.

Theorem 5.3.1. *Suppose that assumptions (A1)-(A7) are satisfied. Then, as* $n \to \infty$, *one has*

$$\widehat{f}_n(x_0) = f_0(x_0) + o_P(1), \tag{5.14}$$

and

$$\widehat{f}_n^{-1}(f_0(x_0)) = x_0 + o_P(1). \tag{5.15}$$

5.4 Rates of convergence

To compute rates of convergence, we need stronger assumptions than for consistency. We replace assumption (A1) for the following stronger version

(B1) $\{X_t\}$ is a positive or β-null recurrent, aperiodic and irreducible Markov Chain whose kernel $P(x, A)$ satisfies the minorization condition (5.9).

We replace, (A5), (A6) and (A7), for the following slightly more restrictive assumption

(B2) The function f_0 is non-increasing, the functions f_0 and F_C are differentiable in C, and the derivatives F_C' and f_0' are bounded, in absolute value, above and away from zero in C.

Let λ be the initial measure of $\mathbf{X}$. Our next hypothesis imposes some control on the behaviour of the chain in the first regenerative block.

(B3) There exists a constant K and a neighborhood V of 0, such that

$$\mathbb{E}_\lambda \left(\sum_{t=0}^{\tau_{\breve{\alpha}}} \left(\mathbb{I}_C \left\{ X_t \leqslant x_0 + \gamma \right\} - \mathbb{I}_C \left\{ X_t \leqslant x_0 - \gamma \right\} \right) \right) \leqslant K\gamma \quad \forall \gamma \in V.$$

Assumption (B3)is satisfied if we assume that the initial measure of the chain is the small measure used for the construction of the split chain (see equation 4.16c in [90]). In the positive recurrent case, taking λ equal to the unique invariant probability measure of the chain also satisfies (B3).

And finally, we need to control the number of times the chain visits C in a regeneration block.

(B4) $\ell_C(\mathcal{B}_1) = \sum_{t\in\mathcal{B}_1} \mathbb{I}_C\{X_t\}$ has finite second moment.

Theorem 5.4.1. *Assume that (A2), (A3), (A4), (B1), (B2), (B3) and (B4) hold. Then, as $n \to \infty$, one has*

$$\widehat{f}_n(x_0) = f_0(x_0) + O_P(u\,(n)^{-1/3}), \tag{5.16}$$

with $u\,(n)$ as defined in (5.11).

The rate $u\,(n)$ comes from Lemmas 5.5.3 and 5.6.7, and as it can be seen from Theorem 5.2.1, it is a deterministic approximation of $T\,(n)$. Note that in the positive recurrent case, $u\,(n) = n$, hence we obtain the same rate $n^{-1/3}$ as in the i.i.d. case [53, Chapter 2]. In the β-null recurrent case, however, the rate of convergence is $n^{\beta/3}L^{1/3}\,(n)$ which is slower than the usual rate. This is due to null recurrence of the chain because it takes longer for the process to return to a neighborhood of the point x_0 and it is these points in the neighborhood of x_0 which are used in nonparametric estimation.

5.5 Localized Markov chains

Given the localized nature of our approach, in this section, we present some results that are particularly useful in this scenario. These results are well known for positive recurrent chains but are new in the null recurrent case.

The first result can be viewed as an extension of the Glivenko-Cantelli theorem to the localized scenario.

Lemma 5.5.1. *Assume that (A1) and (A3) hold. Then, there exists a stationary σ-finite measure π, and F defined by (5.13), such that,*

$$\sup_{y\in\mathbb{R}} |F_n(y) - F(y)| \to 0 \quad a.s. \tag{5.17}$$

as $n \to \infty$. If (A5) is also satisfied, then, for all sufficiently small $\varepsilon > 0$, as $n \to \infty$ we have

$$\sup_{|p-F(x_0)|\leqslant\varepsilon} \left|F_n^{-1}(p) - F^{-1}(p)\right| \to 0 \quad a.s. \tag{5.18}$$

Our next result (Lemma 5.5.2), which is an extension of Lemma 2 in [18] to the localized β-null recurrent case, deals with the properties of classes of functions defined over the regeneration blocks. Before presenting the result, we need some machinery.

Recall that $E \subseteq \mathbb{R}$ denotes the state space of X. Define $\widehat{E} = \cup_{k=1}^{\infty} E^k$ (i.e. the set of finite subsets of E) and let the *localized occupation measure* M_C be given by

$$M_C(B, dy) = \sum_{x\in B\cap C} \delta_x(y), \qquad \text{for every } B \in \widehat{E}.$$

The function that gives the size of the localized blocks is $\ell_C : \widehat{E} \to \mathbb{N}$

$$\ell_C(B) = \int M_C(B, \mathrm{d}y), \qquad \text{for every } B \in \widehat{E}.$$

Let $\widehat{\mathcal{E}}$ denote the smallest σ-algebra formed by the elements of the σ-algebras $\mathcal{E}^k$, $k \geqslant 1$, where $\mathcal{E}^k$ stands for the classical product σ-algebra. Let $\widehat{Q}$ denote a probability measure on $(\widehat{E}, \widehat{\mathcal{E}})$. If $B(\omega)$ is a random variable with distribution Q', then

$M_C(B(\omega), \mathrm{d}y)$ is a random measure, i.e., $M_C(B(\omega), \mathrm{d}y)$ is a (counting) measure on $(E, \mathcal{E})$, almost surely, and for every $A \in \mathcal{E}$, $M_C(B(\omega), A) = \int_A M_C(B(\omega), \mathrm{d}y)$ is a measurable random variable (valued in $\mathbb{N}$). Henceforth $\ell(B(\omega)) \times \int f(y)\, M_C(B(\omega), \mathrm{d}y)$ is a random variable and, provided that $\widehat{Q}(\ell^2) < \infty$, the map Q_C, defined by

$$Q_C(A) = E_{\widehat{Q}}\left(\ell_C(B) \times \int_A M_C(B, \mathrm{d}y)\right) / E_{\widehat{Q}}(\ell_C^2), \qquad \text{for every } A \in \mathcal{E}, \quad (5.19)$$

is a probability measure on $(E, \mathcal{E})$. The notation E_{Q_C} stands for the expectation with respect to the underlying measure Q_C. Introduce the following notations: for any function $g : E \to \mathbb{R}$, let $\widehat{g}_C : \widehat{E} \to \mathbb{R}$ be given by

$$\widehat{g}_C(B) = \int g(y)\, M_C(B, dy) = \sum_{x \in B \cap C} g(x) = \sum_{x \in B} g_C(x), \qquad (5.20)$$

and for any class $\mathcal{G}$ of real-valued functions defined on E, denote the localized version of the sums on the blocks by $\widehat{G}_C = \{\widehat{g}_C : g \in \mathcal{G}\}$.

Notice that, for any function g,

$$\mathbb{E}_{Q_C}(g) = \frac{\mathbb{E}_{\widehat{Q}}\left(\ell_C(B) \times \int g(y)\, M_C(B, dy)\right)}{\mathbb{E}_{\widehat{Q}}(\ell_C^2)} = \frac{\mathbb{E}_{\widehat{Q}}(\ell_C(B)\, \widehat{g}_C(B))}{\mathbb{E}_{\widehat{Q}}(\ell_C^2)}. \qquad (5.21)$$

Lemma 5.5.2. *Let $\widehat{Q}$ be a probability measure on $(\widehat{E}, \widehat{\mathcal{E}})$ such that $0 < \|\ell_C\|_{L^2(\widehat{Q})} < \infty$ and $\mathcal{G}$ be a class of measurable real-valued functions defined on $(E, \mathcal{E})$. Then we have, for every $0 < \varepsilon < \infty$,*

$$\mathcal{N}\left(\varepsilon \|\ell_C\|_{L_2(\widehat{Q})}, \widehat{\mathcal{G}}_C, L^2\left(\widehat{Q}\right)\right) \leqslant \mathcal{N}\left(\varepsilon, \mathcal{G}, L^2(Q)\right),$$

where Q is given in (5.19). Moreover, if $\mathcal{G}$ belongs to the Vapnik–Chervonenkis (VC) class of functions with constant envelope U and characteristic $(\boldsymbol{C}, v)$, then $\widehat{\mathcal{G}}$ is VC

with envelope $U\ell_C$ *and characteristic* $(\boldsymbol{C}, v)$.

Remark 5.5.1. *For a probability measure* μ*, and a class of functions* $\mathcal{H}$*, the covering number* $\mathcal{N}\left(\varepsilon, \mathcal{H}, L^r\left(\mu\right)\right)$ *is the minimum number of* $L^r\left(\mu\right)$ ε*-balls needed to cover* $\mathcal{H}$*. For more details about this concept and the VC class of functions, see [70].*

To put into perspective Lemma 5.5.2, consider a class of bounded functions $\mathcal{G}$ that is VC with finite envelope. Lemma 5.5.2 tells us that the class of unbounded functions $\widehat{\mathcal{G}}_C$ is also VC. If we also have that (B4) holds, then Theorem 2.5 in [70] tells us that $\widehat{\mathcal{G}}_C$ is a Donsker class. A reasoning like this is used in the proof of the following result, which is a stronger version of Lemma 5.5.1 under assumptions (B1) and (B2) and has some interest on its own.

Lemma 5.5.3. *Assume that (B1), (B2), (A3), (A4) and (B4) hold. Then, for all sufficiently small* $\varepsilon > 0$ *we have,*

$$T_n(C) \sup_{|y-x_0|\leqslant\varepsilon} \left|F_n(y) - F(y)\right|^2 = O_p\left(1\right) \tag{5.22}$$

when n *goes to* $+\infty$*. If (B2) is also satisfied, as* $n \to \infty$ *we have*

$$T_n(C) \sup_{|p-F(x_0)|\leqslant\varepsilon} \left|F_n^{-1}(p) - F^{-1}(p)\right|^2 = O_p\left(1\right). \tag{5.23}$$

5.6 Proofs

In this section we give a general outline of the proofs of Theorems 5.3.1 and 5.4.1. The technical proofs can be found in sections 5.7.2 to 5.7.5.

5.6.1 Outline of the proof of Theorem 5.3.1

Recall that we consider the piecewise-constant and left-continuous LSE $\widehat{f}_n$, that is constant on every interval $(Y_{k-1}, Y_k]$, $k = 2, \dots, m$ and also on $(-\infty, Y_1]$ and on $[Y_m, \infty)$. With $\delta > 0$ fixed, we denote by $T_n(C)$ the number of times the Markov Chain $\mathbf{X}$ visits the set $C := [x_0 - \delta, x_0 + \delta]$ until time n:

$$T_n(C) = \sum_{t=0}^{n} \mathbb{I}\{X_t \in C\}. \tag{5.24}$$

Let $l_k = \sum_{t=1}^{n} \mathbb{I}_C\{X_t \leqslant Y_k\}$ for all $k \in \{1, \dots, m\}$ and $l_0 = 0$.

Our aim is to provide a characterization of $\widehat{f}_n(x_0)$. Recall from (5.12) that the localized empirical distribution function F_n is defined as

$$F_n(y) = \frac{1}{T_n(C)} \sum_{i=0}^{T_n(C)} \mathbb{I}\{X_{\sigma_C(i)} \leqslant y\} = \frac{1}{T_n(C)} \sum_{t=0}^{n} \mathbb{I}_C\{X_t \leqslant y\}$$

for $y \in \mathbb{R}$. F_n is 0 on $(-\infty, Y_1)$, so, with an arbitrary random variable $Y_0 < Y_1$ we have $F_n(y) = F_n(Y_0) = 0$ for all $y < Y_1$. Let $\mathcal{K}$ be the set

$$\mathcal{K} := \{F_n(Y_k),\ k = 0, \dots, m\} \tag{5.25}$$

and let Λ_n be the continuous piecewise-linear process on $[F_n(Y_0), F_n(Y_m)]$ with knots at the points in $\mathcal{K}$ and values

$$\Lambda_n\left(F_n(Y_k)\right) = \frac{1}{T_n(C)} \sum_{t=0}^{n} Z_t \mathbb{I}_C\{X_t \leqslant Y_k\} \tag{5.26}$$

at the knots. The characterization of $\widehat{f}_n$ in Lemma 5.6.2 involves the least concave majorant of Λ_n. Note that we use $T_n(C)$ as a normalization in the definitions of the processes F_n and Λ_n since this choice ensures that F_n and Λ_n converge to fixed

functions, see Lemma 5.5.1.

Lemma 5.6.1. *For all* $y \in [F_n(Y_0), F_n(Y_m)]$,

$$\Lambda_n(y) = L_n(y) + M_n(y),$$

where,

$$L_n(y) = \int_0^y f \circ F_n^{-1}(u)\, du, \tag{5.27}$$

and M_n *is a piece-wise linear processes with knots at* $F_n(Y_k)$ *for* $k \in \{0, \ldots, m\}$ *such that*

$$M_n(F_n(Y_k)) = \frac{1}{T_n(C)} \sum_{t=0}^{n} W_t \mathbb{I}_C\{X_t \leqslant Y_k\}.$$

Moreover, M_n *can be written as*

$$M_n(y) = \begin{cases} 0 & , \text{ if } y = 0 \\ R_n^j(y) + M_n^j & , \text{ otherwise} \end{cases} \tag{5.28}$$

where,

$$M_n^j = M_n(F_n(Y_j)) = \frac{1}{T_n(C)} \sum_{t=0}^{n} W_t \mathbb{I}_C\{X_t \leqslant Y_j\}, \tag{5.29}$$

$$R_n^j(y) = \frac{\sum_{t=0}^{n} W_t \mathbb{I}_C\{X_t = Y_{j+1}\}}{l_{j+1} - l_j} (y - F_n(Y_j)), \tag{5.30}$$

and j *is such that* $Y_{j+1} = F_n^{-1}(y)$.

In the next lemma, we give an alternative characterization of the monotone nonparametric LSE $\widehat{f}_n$ at the observation points $Y_1, \ldots, Y_m$.

Lemma 5.6.2. *Let $C = [x - \delta, x + \delta]$ for some fixed $\delta > 0$. Let $\widehat{\lambda}_n$ be the left-hand slope of the least concave majorant of Λ_n. Then,*

$$\widehat{f}_n(Y_k) = \widehat{\lambda}_n \circ F_n(Y_k), \quad \forall k \in \{1, \dots, m\}. \tag{5.31}$$

with probability 1 for n big enough.

We consider below the generalized inverse function of $\widehat{f}_n$ since it has a more tractable characterization than $\widehat{f}_n$ itself. To this end, let us define precisely the generalized inverses of all processes of interest. Since $\widehat{\lambda}_n$ is a non-increasing left-continuous step function on $(F_n(Y_0), F_n(Y_m)]$ that can have jumps only at the points $F_n(Y_k)$, $k \in \{1, \dots, m\}$, we define its generalized inverse $\widehat{U}_n(a)$, for $a \in \mathbb{R}$, as the greatest $y \in (F_n(Y_0), F_n(Y_m)]$ that satisfies $\widehat{\lambda}_n(y) \geqslant a$, with the convention that the supremum of an empty set is $F_n(Y_0)$. Then for every $a \in \mathbb{R}$ and $y \in (F_n(Y_0), F_n(Y_m)]$, one has

$$\widehat{\lambda}_n(y) \geqslant a \text{ if and only if } \widehat{U}_n(a) \geqslant y. \tag{5.32}$$

Likewise, since $\widehat{f}_n$ is a left-continuous non-increasing step function on $\mathbb{R}$ that can have jumps only at the observation times $Y_1 < \cdots < Y_m$, we define the generalized inverse $\widehat{f}_n^{-1}(a)$, for $a \in \mathbb{R}$, as the greatest $y \in [Y_0, Y_m]$ that satisfies $\widehat{f}_n(y) \geqslant a$, with the convention that the supremum of an empty set is Y_0. We then have

$$\widehat{f}_n(y) \geqslant a \text{ if and only if } \widehat{f}_n^{-1}(a) \geqslant y \tag{5.33}$$

for all $a \in \mathbb{R}$ and $y \in (Y_0, Y_m]$. On the other hand, since F_n is a right-continuous non-decreasing step function on $\mathbb{R}$ with range $[F_n(Y_0), F_n(Y_m)]$, we define the generalized inverse $F_n^{-1}(a)$, for $a \leqslant F_n(Y_m)$, as the smallest $y \in [Y_0, Y_m]$ which

satisfies $F_n(y) \geqslant a$. Note that the infimum is achieved for all $a \leqslant F_n(Y_m)$. We then have

$$F_n(y) \geqslant a \text{ if and only if } F_n^{-1}(a) \leqslant y \tag{5.34}$$

for all $a \leqslant F_n(Y_m)$ and $y \in [Y_0, Y_m]$, and thanks to Lemma 5.6.2 we have

$$\widehat{f}_n^{-1} = F_n^{-1} \circ \widehat{U}_n \tag{5.35}$$

on $\mathbb{R}$. Moreover, one can check that

$$\widehat{U}_n(a) = \underset{p \in [F_n(Y_0), F_n(Y_m)]}{\operatorname{argmax}} \{\Lambda_n(p) - ap\}, \text{ for all } a \in \mathbb{R}, \tag{5.36}$$

where argmax denotes the greatest location of maximum (which is achieved on the set $\mathcal{K}$ in (5.25)). Thus, the inverse process $\widehat{U}_n$ is a location process that is more tractable than $\widehat{f}_n$ and $\widehat{\lambda}_n$ themselves. A key idea in the following proofs is to derive properties of $\widehat{U}_n$ from its argmax characterization (5.36), then, to translate these properties to $\widehat{f}_n^{-1}$ thanks to (5.35), and finally to translate them to $\widehat{f}_n$ thanks to (5.33). The last step will be the aim of Section 5.7.3. We consider below the first two steps.

To go from $\widehat{U}_n$ to $\widehat{f}_n^{-1}$ using (5.35) requires to approximate F_n^{-1} by a fixed function. Hence, in the sequel, we are concerned by the convergence of the process F_n given in (5.12), where $\delta > 0$ is chosen sufficiently small, and by the convergence of the corresponding inverse function F_n^{-1}.

It is stated in Lemma 5.5.1 that under (A1) and (A3), F_n converges to a fixed distribution function F that depends on C, hence on δ. If, moreover, F is strictly increasing in C, then we can find a neighborhood of $F(x_0)$ over which the (usual) inverse function F^{-1} is uniquely defined, and F_n^{-1} converges to F^{-1}.

In the following lemma, we show that $F(x_0)$ belongs to the domain of Λ_n with probability that tends to one as $n \to \infty$.

Lemma 5.6.3. *Assume that (A1), (A3), (A4) and (A5) hold. Then, we can find $\varepsilon > 0$ such that the probability that $Y_1 + \varepsilon \leqslant x_0 \leqslant Y_m - \varepsilon$ tends to one as $n \to \infty$. Moreover, the probability that $F_n(Y_1) \leqslant F(x_0) \leqslant F_n(Y_m)$ tends to one as $n \to \infty$.*

We will also need to control the noise $\{W_t\}$. The following lemma shows that the noise is negligible under our assumptions.

Lemma 5.6.4. *Assume that (A1) and (A2) hold. Let $\mathcal{F}_n = \sigma(\{X_1, \ldots, X_n\})$. Then,*

$$\sum_{t=0}^{n} W_t \mathbb{I}_C\{X_t = A_n\} = o_P\left(T_n(C)\right),$$

and

$$\sup_{u > A_n} \left| \sum_{t=0}^{n} W_t \mathbb{I}_C\left\{X_t \in (A_n, u]\right\} \right| = o_P\left(T_n(C)\right).$$

for any sequence of random variables A_n, independent of the process $\{W_t\}$, that is adapted to the filtration $\{\mathcal{F}_n\}$.

With the above lemmas, we can prove convergence of $\widehat{U}_n$ given by (5.36), and then convergence of $\widehat{f}_n^{-1}$ given by (5.35), at the fixed point $f_0(x_0)$.

Lemma 5.6.5. *Suppose that assumptions (A1)-(A7) are satisfied. Then, as $n \to \infty$, one has*

$$\widehat{U}_n(f_0(x_0)) = F(x_0) + o_P(1). \tag{5.37}$$

5.6.2 Outline of the proof of Theorem 5.4.1

The proof of Theorem 5.4.1, uses similar ideas as the ones used in the proof of Theorem 5.3.1 but under stronger assumptions (and therefore using stronger lemmas).

The first intermediate result is the following stronger version of Lemma 5.6.4.

Lemma 5.6.6. *Assume that (A2), (A3), (A4), (B1), (B2) and (B3) hold. Then, there exists $K > 0$, $\gamma_0 > 0$ that do not depend on n and $N_{\gamma_0} \in \mathbb{N}$, such that for all $\gamma \in [0, \gamma_0]$ and $n \geqslant N_{\gamma_0}$ one has*

$$\mathbb{E}_\lambda \left(\sup_{|y-x_0| \leqslant \gamma} \left| \sum_{t=0}^{n} W_t \left(\mathbb{I}_C\{X_t \leqslant y\} - \mathbb{I}_C\{X_t \leqslant x_0\} \right) \right|^2 \right) \leqslant K u(n) \gamma \tag{5.38}$$

$$\mathbb{E}_\lambda \left(\sup_{|y-x_0| \leqslant \gamma} \left| \sum_{t=0}^{n} W_t \mathbb{I}_C\{X_t = y\} \right|^2 \right) \leqslant K u(n) \gamma \tag{5.39}$$

Then, we need to quantify how well we can approximate $T_n(C)$ by $u(n)$.

Lemma 5.6.7. *Assume that (B1) and (A3) hold. Then we have*

a) As $n \to \infty$ we have

$$\frac{u(n)}{T_n(C)} = O_P(1).$$

b) Let α and η be positive constants, then there exists constants N_η, $\underline{c}_\eta$ and $\overline{c}_\eta$, such that

$$\mathbb{P}\left(\left(\frac{T_n(C)}{a(n)} \right)^\alpha \in \left[\underline{c}_\eta, \overline{c}_\eta \right] \right) \geqslant 1 - \eta, \quad \forall n \geqslant N_\eta.$$

With the above lemmas (including Lemma 5.5.3 and the ones used in Section 5.6.1), we can obtain the rate of convergence of $\widehat{U}_n$ given by (5.36), and then the rate of convergence of $\widehat{f}_n^{-1}$ given by (5.35), at the fixed point $f_0(x_0)$.

Lemma 5.6.8. *Assume that (A2), (A3), (A4), (B1), (B2), (B3) and(B4) hold. Then, as $n \to \infty$, one has*

$$\widehat{U}_n(f_0(x_0)) = F(x_0) + O_P\left(u(n)^{-1/3} \right), \tag{5.40}$$

and

$$\widehat{f}_n^{-1}(f_0(x_0)) = x + O_P\left(u\,(n)^{-1/3}\right). \tag{5.41}$$

5.7 Technical proofs

5.7.1 Technical proofs for Section 5.5

Proof of Lemma 5.5.1. Equation (5.17) follows from Corollary 2 in [9] and part 2 of Theorem 5.2.1.

Now, we turn to the proof of (5.18). To do this, we adapt some of the ideas presented in the proof of Lemma 21.2 in [114].

Let V a normal random variable independent of the X_i's, and Φ its distribution function. it follows from (5.17) that conditionally on the X_t's, $F_n(V)$ converges almost surely to $F(V)$. Thus, denoting by $\mathbb{P}_X$ the conditional probability given the X_t's, it follows from (5.34) that $\Phi(F_n^{-1}(u)) = \mathbb{P}_X(F_n(V) < u)$ converges almost surely to $\mathbb{P}_X(F(V) < u) = \Phi(F^{-1}(u))$ at every u at which the limit function is continuous . Since F is strictly increasing in C, one can find $\varepsilon > 0$ such that F^{-1} is continuous on $[F(x_0) - \varepsilon, F(x_0) + \varepsilon]$, so the above limit function is continuous at every $u \in [F(x_0) - \varepsilon, F(x_0) + \varepsilon]$. By continuity of Φ^{-1} on $(0, 1)$, $F_n^{-1}(u)$ converges almost surely to $F^{-1}(u)$ for every such u. By monotonicity, the convergence is uniform, hence

$$\sup_{|p - F(x_0)| \leqslant \varepsilon} |F_n^{-1}(p) - F^{-1}(p)| = o(1) \quad a.s.$$

as $n \to \infty$.

Proof of Lemma 5.5.2. This proof is an adaptation to the localized case of the proof of Lemma 2 in [18]. Let $f'_C \in \mathcal{F}'_C$, i.e., there exists $f \in \mathcal{F}$ such that

$f'_C(B) = \int f(y)\, M_C(B, \mathrm{d}y)$. By Cauchy–Schwarz inequality,

$$\left(\int f(y) M_C(B, dy) \right)^2 \leqslant \ell_C(B) \left(\int f^2 M_C(B, dy) \right),$$

then

$$\mathbb{E}_{Q'}(f'^2_C) \leqslant \mathbb{E}_{Q'} \left(\ell_C(B) \left(\int f(y)^2\, M_C(B, dy) \right) \right) = \mathbb{E}_{Q_C}(f^2) E_{Q'}(\ell^2_C),$$

where the last equality follows from (5.21). Applying this to the function

$$f'_C(B) - f'_k(B) = \int (f(y) - f_k(y))\, M_C(B, dy),$$

when each f_k is the center of an ε-cover of the space $\mathcal{F}$ and $\|f - f_k\|_{L_2(Q_C)} \leqslant \varepsilon$ gives the first assertion of the lemma. To obtain the second assertion, note that $U'_C = U\ell_C$ is an envelope for $\mathcal{F}'_C$. In addition, we have that

$$\|U'_C\|_{L_2(Q')} = U\|\ell_C\|_{L_2(Q')}.$$

From this, we derive that, for every $0 < \varepsilon < 1$,

$$\mathcal{N}(\varepsilon\|U'_C\|_{L_2(Q')}, \mathcal{U}'_C, L_2(Q')) = \mathcal{N}(\varepsilon U\|\ell_C\|_{L_2(Q')}, \mathcal{U}', L_2(Q')).$$

Then using the first assertion of the lemma, we obtain for every $0 < \varepsilon < 1$,

$$\mathcal{N}(\varepsilon\|U'_C\|_{L_2(Q')}, \mathcal{F}'_C, L_2(Q')) \leqslant \mathcal{N}(\varepsilon U, \mathcal{F}, L_2(Q_C)),$$

which implies the second assertion of the Lemmaz whenever the class $\mathcal{F}$ is VC with envelope U.

Proof of Lemma 5.5.3. Let $B \in \widehat{E}$ and $g : E \times \mathbb{R} \to \mathbb{R}_+$. For each $y \in \mathbb{R}$ we define $g_y(x) = g(x, y)$, then, using the notation of section 5.6.2 we will have $\widehat{g}_y(B) = \sum_{x \in B \cap C} g(x, y)$. Finally, for any function $h : \mathbb{R} \to \mathbb{R}$, we define

$$\widetilde{g}_y^h(B) = \widehat{(g_y - h(y))}(B) = \sum_{x \in B \cap C} (g(x, y) - h(y)) = \widehat{g}_y(B) - \ell_C(B)\, h(y).$$

Let $g(x, y) = \mathbb{I}\{x \leqslant y\}$, and $h = F$ as defined in (5.13). Then, $\widehat{g}_y(B) = \sum_{x \in B} \mathbb{I}_C\{x \leqslant y\}$ and

$$\widetilde{g}_y^F(B) = \sum_{x \in B \cap C} (\mathbb{I}\{x \leqslant y\} - F(y)) = \widehat{g}_y(B) - \ell_C(B)\, F(y).$$

From now on, we'll remove the superindex from $\widetilde{g}_y^F$ to ease the notation.

By the definition of F_n and F ((5.12) and (5.13)), we have that

$$\begin{aligned} F_n(y) - F(y) &= \frac{1}{T_n(C)} \sum_{i=1}^{T_n(C)} \big(\mathbb{I}\{X_{\sigma_C(i)} \leqslant y\} - F(y)\big) \\ &= \frac{1}{T_n(C)} \sum_{i=0}^{n} (\mathbb{I}_C\{X_t \leqslant y\} - \mathbb{I}_C\{X_i\} F(y)) \\ &= \frac{1}{T_n(C)} \left(\widetilde{g}_y(\mathcal{B}_0) + \sum_{i=1}^{T(n)} \widetilde{g}_y(\mathcal{B}_i) + \widetilde{g}_y\big(\mathcal{B}_{(n)}\big) \right), \end{aligned}$$

therefore,

$$\sqrt{T_n(C)}\Big(F_n(y) - F(y)\Big) = \frac{\widetilde{g}_y(\mathcal{B}_0)}{\sqrt{T_n(C)}} + \frac{\sum_{i=1}^{T(n)} \widetilde{g}_y(\mathcal{B}_i)}{\sqrt{T_n(C)}} + \frac{\widetilde{g}_y\big(\mathcal{B}_{(n)}\big)}{\sqrt{T_n(C)}}.$$

Notice that $|\widetilde{g}_y(\mathcal{B}_0)| \leqslant 2\ell_C(\mathcal{B}_0) < +\infty$ and $T_n(C) \to +\infty$ almost surely, therefore, the first term in the last equation converges almost surely to 0 uniformly

in y. For the last term, we have that

$$\frac{\left|\widetilde{g}_y\left(\mathcal{B}_{(n)}\right)\right|}{\sqrt{T_n(C)}} \leqslant \frac{2\ell_C(\mathcal{B}_{T(n)})}{\sqrt{T_n(C)}} = 2\sqrt{\frac{T\left(n\right)}{T_n(C)}}\frac{\ell_C(\mathcal{B}_{T(n)})}{\sqrt{T\left(n\right)}},$$

by (B4), the expectation of $\ell_C^2(\mathcal{B}_1)$ is finite, then, Lemma 1 in [9] shows that $\frac{\ell_C^2(\mathcal{B}_n)}{n} \to 0$ a.s. which implies that $\frac{\ell_C(\mathcal{B}_n)}{\sqrt{n}}$ also converges to 0 a.s. Since $T\left(n\right) \to +\infty$ a.s., by Theorem 6.8.1 in [55] we have $\frac{\ell_C(\mathcal{B}_{T(n)})}{\sqrt{T(n)}} \to 0$ almost surely. Joining this with the almost sure convergence of $\frac{T(n)}{T_n(C)}$ to a positive constant (see Theorem 5.2.1) we obtain that $\frac{\left|\widetilde{g}_y\left(\mathcal{B}_{(n)}\right)\right|}{\sqrt{T_n(C)}}$ converges almost surely to 0 uniformly in y. Therefore,

$$\sqrt{T_n(C)}\Big(F_n(y) - F(y)\Big) = \frac{\sum_{i=1}^{T(n)} \widetilde{g}_y\left(\mathcal{B}_i\right)}{\sqrt{T(n)}} + o_P\left(1\right). \tag{5.42}$$

where we have used that $\frac{T_n(C)}{T(n)}$ converges almost surely to a positive constant to use $T(n)$ instead of $T_n(C)$.

Then, (5.22) will be proved if we show that, for ε small enough

$$\sup_{|y-x_0|\leqslant\varepsilon} \frac{\left|\sum_{i=1}^{T(n)} \widetilde{g}_y\left(\mathcal{B}_i\right)\right|}{\sqrt{T(n)}} = O_p\left(1\right). \tag{5.43}$$

Fix $\eta > 0$ arbitrarily. By Lemma 5.6.7 and Slutsky's theorem, we can find positive numbers $\underline{a}_\eta, \overline{a}_\eta$ and an integer N_η such that $P\left(\mathcal{E}_n\right) \geqslant 1 - \frac{\eta}{2}$ for all $n \geqslant N_\eta$, where

$$\mathcal{E}_n = \left\{\underline{a}_\eta u\left(n\right) \leqslant T(n) \leqslant \overline{a}_\eta u\left(n\right)\right\}. \tag{5.44}$$

Define $W_n(\varepsilon) = \sup_{|y-x_0|\leqslant\varepsilon} \left|\sum_{i=1}^{n} \widetilde{g}_y\left(\mathcal{B}_i\right)\right|$ and let M_η be a fixed positive number. Then, for all $n \geqslant N_\eta$

$$\mathbb{P}\left(\frac{1}{\sqrt{T(n)}}W_{T(n)} > M_\eta\right) < \mathbb{P}\left(\left\{\frac{1}{\sqrt{T(n)}}W_{T(n)} > M_\eta\right\} \cap \mathcal{E}_n\right) + 1 - \mathbb{P}\left(\mathcal{E}_n\right)$$

$$< \mathbb{P}\left(\left\{\frac{1}{\sqrt{T(n)}}W_{T(n)} > M_\eta\right\} \cap \mathcal{E}_n\right) + \frac{\eta}{2}. \tag{5.45}$$

On $\mathcal{E}_n$, $\underline{a}_\eta u(n) \leqslant T(n) \leqslant \overline{a}_\eta u(n)$, therefore for all $n \geqslant N_\eta$

$$\begin{aligned}\mathbb{P}\left(\left\{\frac{1}{\sqrt{T(n)}}W_{T(n)} > M_\eta\right\} \cap \mathcal{E}_n\right) &< \mathbb{P}\left(\left\{\frac{1}{\sqrt{\underline{a}_\eta u(n)}}\max_{1\leqslant k\leqslant \overline{a}_\eta u(n)} W_k > M_\eta\right\} \cap \mathcal{E}_n\right), \\ &< \mathbb{P}\left(\frac{1}{\sqrt{\underline{a}_\eta u(n)}}\max_{1\leqslant k\leqslant \overline{a}_\eta u(n)} W_k > M_\eta\right).\end{aligned} \tag{5.46}$$

The random variables $\left\{\widetilde{g}_{(\cdot)}(\mathcal{B}_k)\right\}_{k=1}^{\overline{a}_\eta u(n)}$ are i.i.d., therefore, by Montgomery-Smith's inequality (Lemma 4 in [2]), there exists a universal constant K such that for all $n \geqslant N_\eta$,

$$\begin{aligned}\mathbb{P}\left(\frac{1}{\sqrt{\underline{a}_\eta u(n)}}\max_{1\leqslant k\leqslant \overline{a}_\eta u(n)} W_k > M_\eta\right) &< K\mathbb{P}\left(\frac{1}{\sqrt{\underline{a}_\eta u(n)}}W_{\overline{a}_\eta u(n)} > \frac{M_\eta}{K}\right), \\ &< K\mathbb{P}\left(\frac{1}{\sqrt{\underline{a}_\eta u(n)}}\sup_{|y-x_0|\leqslant\varepsilon}\left|\sum_{i=1}^{\overline{a}_\eta u(n)}\widetilde{g}_y(\mathcal{B}_i)\right| > \frac{M_\eta}{K}\right).\end{aligned} \tag{5.47}$$

For an arbitrary set T, let $\ell^{+\infty}(T)$ be the space of all uniformly bounded, real functions on T, equipped with the uniform norm. Weak convergence to a tight process in this space is characterized by asymptotic tightness plus convergence of marginals (see Chapter 1.5 in [115]).

The class of functions $\mathcal{G} - F = \{g_y(\cdot) - F(y)\}_{y\in\mathbb{R}}$ is VC with constant envelope 2, hence, by Lemma 5.5.2, the class of functions $\widehat{\mathcal{G} - F}$ is also VC and has $2\ell_C$ as envelope. $\mathbb{E}\ell_C^2(\mathcal{B}_1)$ is finite (by (B4)), therefore, by Theorem 2.5 in [70],

$\widehat{\mathcal{G}-F}$ is Donsker. Then, the process $\frac{1}{\sqrt{\underline{a}_\eta u(n)}}\left|\sum_{i=1}^{\overline{a}_\eta u(n)} \widetilde{g}_y\left(\mathcal{B}_i\right)\right|$ converges weakly in $\ell^\infty\left[\widehat{\mathcal{G}-F}\right]$ to a tight process Z. The map $y \mapsto \|y\|_\infty$ from $\ell^\infty\left[\widehat{\mathcal{G}-F}\right]$ to $\mathbb{R}$ is continuous with respect to the supremum norm (cf. pp 278 of [114]), therefore, $\frac{1}{\sqrt{\underline{a}_\eta u(n)}} \sup_{|y-x_0|\leqslant\varepsilon}\left|\sum_{i=1}^{\overline{a}_\eta u(n)} \widetilde{g}_y\left(\mathcal{B}_i\right)\right|$ converges in distribution to $\sup_{|y-x_0|\leqslant\varepsilon} Z\left(y\right)$, hence, we can find V_η and N'_η such that

$$\mathbb{P}\left(\frac{1}{\sqrt{\underline{a}_\eta u\left(n\right)}} \sup_{|y-x_0|\leqslant\varepsilon}\left|\sum_{i=1}^{\overline{a}_\eta u(n)} \widetilde{g}_y\left(\mathcal{B}_i\right)\right| > V_\eta\right) < \frac{\eta}{2K}, \quad \forall n > N'_\eta. \tag{5.48}$$

Choosing $M_\eta = KV_\eta$ in 5.48 and joining (5.47), (5.46) and (5.45), completes the proof of (5.22).

Now we proceed to prove (5.23). Let η be fixed, by (5.22) and Lemma 5.6.7, we can find ε', M'_η and N'_η such that

$$\mathbb{P}\left(\sqrt{T_n(C)} \sup_{|y-x_0|\leqslant\varepsilon'} |F_n(y) - F(y)| > M'_\eta\right) < \frac{\eta}{4} \quad \forall n \geqslant N'_\eta \tag{5.49}$$

$$\mathbb{P}\left(\mathcal{D}_n\right) \geqslant 1 - \frac{\eta}{2} \quad \forall n \geqslant N'_\eta \tag{5.50}$$

where $\mathcal{D}_n = \left\{\underline{a}_\eta u\left(n\right) \leqslant T_n(C) \leqslant \overline{a}_\eta u\left(n\right)\right\}$. Define the sets

$$U_n = \left\{\sqrt{T_n(C)} \sup_{|p-F(x_0)|\leqslant\varepsilon} \left|F_n^{-1}(p) - F^{-1}(p)\right| > M_\eta\right\},$$

$$U_n^1 = \left\{\exists p \in [F(x_0)-\varepsilon, F(x_0)+\varepsilon] : F_n^{-1}(p) - F^{-1}(p) > \frac{M_\eta}{\sqrt{T_n(C)}}\right\},$$

$$U_n^2 = \left\{\exists p \in [F(x_0)-\varepsilon, F(x_0)+\varepsilon] : F^{-1}(p) - F_n^{-1}(p) > \frac{M_\eta}{\sqrt{T_n(C)}}\right\}.$$

where ε and M_η are constants that will be specified later.

On $U_n^1 \cap \mathcal{D}_n$, $F_n^{-1}(p) > \frac{M_\eta}{\sqrt{T_n(C)}} + F^{-1}(p) > \frac{M_\eta}{\sqrt{\overline{a}_\eta u(n)}} + F^{-1}(p)$, hence,

$$F_n\left(\frac{M_\eta}{\sqrt{\overline{a}_\eta u(n)}} + F^{-1}(p)\right) \leqslant F_n\left(F_n^{-1}(p)\right) \leqslant p + \frac{1}{T_n(C)} \leqslant p + \frac{1}{\underline{a}_\eta u(n)}. \tag{5.51}$$

Assumption (B2) indicates that F has bounded derivative in C, take K_1 as the maximum value of this derivative in C, then, the Mean Value Theorem implies that

$$p = F\left(F^{-1}(p)\right) = F\left(\frac{M_\eta}{\sqrt{\overline{a}_\eta u(n)}} + F^{-1}(p)\right) - \frac{F'(\theta_p) M_\eta}{\sqrt{\overline{a}_\eta u(n)}} \leqslant F\left(\frac{M_\eta}{\sqrt{\overline{a}_\eta u(n)}} + F^{-1}(p)\right) - \frac{K_1 M_\eta}{\sqrt{\overline{a}_\eta u(n)}}.$$

After plugging this into (5.51) we get

$$F\left(\frac{M_\eta}{\sqrt{\overline{a}_\eta u(n)}} + F^{-1}(p)\right) - F_n\left(\frac{M_\eta}{\sqrt{\overline{a}_\eta u(n)}} + F^{-1}(p)\right) \geqslant \frac{K_1 M_\eta}{\sqrt{\overline{a}_\eta u(n)}} - \frac{1}{\underline{a}_\eta u(n)}.$$

Because $u(n) \to +\infty$, we can find N_1 such that $\sqrt{\frac{\overline{a}_\eta}{u(n)}} \frac{1}{\underline{a}_\eta K_1} < 1$ for all $n \geqslant N_1$, taking M_η bigger than $\frac{M'_\eta}{K_1}\sqrt{\frac{\overline{a}_\eta}{\underline{a}_\eta}} + 1$ and using that $T_n(C) \leqslant \underline{a}_\eta u(n)$ on $\mathcal{D}_n$, we obtain, for all $n \geqslant N_1$

$$F\left(\frac{M_\eta}{\sqrt{\overline{a}_\eta u(n)}} + F^{-1}(p)\right) - F_n\left(\frac{M_\eta}{\sqrt{\overline{a}_\eta u(n)}} + F^{-1}(p)\right) > \frac{M'_\eta}{\sqrt{T_n(C)}}. \tag{5.52}$$

Let $N_{2,\eta}$ be such that $\frac{M_\eta}{\sqrt{\overline{a}_\eta u(n)}} < \frac{\varepsilon'}{2}$ for $n \geqslant N_{2,\eta}$. By the continuity of F^{-1} in $F(x_0)$ there exists $\varepsilon > 0$ such that $|F^{-1}(p) - x_0| \leqslant \frac{\varepsilon'}{2}$ for all p in $[F(x_0) - \varepsilon, F(x_0) + \varepsilon]$, therefore, the triangular inequality implies that $\frac{M_\eta}{\sqrt{\overline{a}_\eta u(n)}} + F^{-1}(p)$ lies in the in-

terval $[x_0 - \varepsilon', x_0 + \varepsilon']$ for all $n \geqslant N_\eta = \max(N_1, N_{2,\eta})$. This, alongside (5.52), shows that for all $n \geqslant N_\eta$

$$\begin{aligned} U_n^1 \cap \mathcal{D}_n &\subseteq \left\{ \exists y \in [x_0 - \varepsilon', x_0 + \varepsilon'] : F(y) - F_n(y) > \frac{M'_\eta}{\sqrt{T_n(C)}} \right\} \\ &\subseteq \left\{ \sqrt{T_n(C)} \sup_{|y - x_0| \leqslant \varepsilon'} |F_n(y) - F(y)| > M'_\eta \right\}. \end{aligned}$$

By a similar argument, it can be shown that

$$U_n^2 \cap \mathcal{D}_n \subseteq \left\{ \exists y \in [x_0 - \varepsilon', x_0 + \varepsilon'] : F_n(y) - F(y) > \frac{M'_\eta}{\sqrt{T_n(C)}} \right\} \quad \forall n \geqslant N_\eta.$$

Using (5.49) and $U_n = U_n^1 \cup U_n^2$ we obtain that $\mathbb{P}(U_n \cap \mathcal{D}_n) \leqslant \frac{\eta}{2}$ for all $n \geqslant N_\eta$. Equation (5.23) now follows by (5.50).

5.7.2 Technical proofs for Section 5.6.1

Proof of Lemma 5.6.1. Combining (5.26) and (5.1) yields

$$\Lambda_n(F_n(Y_k)) = \frac{1}{T_n(C)} \sum_{t=0}^{n} f_0(X_t) \mathbb{I}_C\{X_t \leqslant Y_k\} + \frac{1}{T_n(C)} \sum_{t=0}^{n} W_t \mathbb{I}_C\{X_t \leqslant Y_k\}.$$

The first term on the right hand side of the previous display can be rewritten as follows:

$$\begin{aligned} \frac{1}{T_n(C)} \sum_{t=0}^{n} f_0(X_t) \mathbb{I}_C\{X_t \leqslant Y_k\} &= \frac{1}{T_n(C)} \sum_{j=1}^{m} f_0(Y_l)(l_j - l_{j-1}) \mathbb{I}_C\{Y_j \leqslant Y_k\} \\ &= \sum_{j=1}^{k} \int_{l_{j-1}/T_n(C)}^{l_j/T_n(C)} f_0 \circ F_n^{-1}(u) du, \end{aligned}$$

using that $F_n^{-1}(u) = Y_j$ for all $u \in (l_{j-1}/T_n(C), l_j/T_n(C)]$. Hence, for all k in

$\{0, \ldots, m\}$

$$\Lambda_n(F_n(Y_k)) = \int_0^{l_k/T_n(C)} f_0 \circ F_n^{-1}(u)du + \frac{1}{T_n(C)} \sum_{t=0}^{n} W_t \mathbb{I}_C\{X_t \leqslant Y_k\}. \qquad (5.53)$$

Combining (5.53) with the piece-wise linearity of Λ_n yields

$$\Lambda_n\left(F_n\left(Y_k\right)\right) = L_n\left(F_n\left(Y_k\right)\right) + M_n\left(F_n\left(Y_k\right)\right),$$

where L_n and M_n are piece-wise linear processes with knots at $F_n(Y_k)$ for k in $\{0, \ldots, m\}$ and such that

$$L_n(F_n(Y_k)) = \int_0^{l_k/T_n(C)} f_0 \circ F_n^{-1}(u)du$$

and

$$M_n(F_n(Y_k)) = \frac{1}{T_n(C)} \sum_{t=0}^{n} W_t \mathbb{I}_C\{X_t \leqslant Y_k\}.$$

In order to ease the notation, we will write $F_n^i = F_n(Y_i)$, $L_n^i = L_n\left(F_n\left(Y_i\right)\right)$ and $M_n^i = M_n\left(F_n\left(Y_i\right)\right)$. Let $y \in \left(F_n\left(Y_0\right), F_n\left(Y_m\right)\right]$, take j such that $Y_{j+1} = F_n^{-1}\left(y\right)$, then $F_n\left(Y_j\right) < y \leqslant F_n\left(Y_{j+1}\right)$. With this notation,

$$L_n\left(y\right) = \frac{L_n^{j+1} - L_n^j}{F_n^{j+1} - F_n^j}\left(y - F_n^j\right) + L_n^j,$$
$$M_n\left(y\right) = \frac{M_n^{j+1} - M_n^j}{F_n^{j+1} - F_n^j}\left(y - F_n^j\right) + M_n^j.$$

Notice that

$$L_n^{j+1} - L_n^j = \int_{\frac{l_j}{T_n(C)}}^{\frac{l_{j+1}}{T_n(C)}} f_0 \circ F_n^{-1}\left(u\right) du = \frac{l_{j+1} - l_j}{T_n(C)} f\left(Y_{j+1}\right),$$

$$F_n^{j+1} - F_n^j = \frac{l_{j+1} - l_j}{T_n(C)},$$

therefore,

$$L_n(y) = f_0(Y_{j+1})\left(y - F_n^j\right) + L_n^j = \int_{\frac{l_j}{T_n(C)}}^{y} f_0 \circ F_n^{-1}(u)\, du + L_n^j = \int_0^y f_0 \circ F_n^{-1}(u)\, du,$$

which proves (5.27).

For M_n we have,

$$M_n^{j+1} - M_n^j = \frac{1}{T_n(C)} \sum_{t=0}^{n} W_t \mathbb{I}_C\{X_t = Y_{j+1}\},$$

then,

$$M_n(y) = \frac{\sum_{t=1}^{n} W_t \mathbb{I}_C\{X_t = Y_{j+1}\}}{l_{j+1} - l_j}\left(y - F_n^j\right) + M_n^j = R_n^j(y) + M_n^j.$$

and this completes the proof.

Proof of Lemma 5.6.2. By definition, with $l_0 = 0$, and $l_k = \sum_{t=0}^{n} \mathbb{I}_C\{X_t \leqslant Y_k\}$ for all $k \in \{1, \ldots, m\}$, we have $F_n(Y_k) = a l_k$ for all $k \in \{0, \ldots, m\}$, where $a = 1/T_n(C)$ and does not depend on k. Moreover,

$$\Lambda_n(F_n(Y_k)) = a \sum_{t=0}^{n} Z_t \mathbb{I}_C\{X_t \leqslant Y_k\}$$

Since $\widehat{f}_n(Y_k)$ is the left-hand slope at l_k of the least concave majorant of the set of points in (5.3), the equality in (5.31) follows from Lemma 2.1 in [42].

Proof of Lemma 5.6.3. The first assertion follows from Assumption (A4) and the second immediately follows from the first one by (5.17) combined with the

strict monotonicity of F in C.

Proof of Lemma 5.6.4. Let $\mathcal{F}_n = \sigma\left(\{X_0, \ldots, X_n\}\right)$ be sigma algebra generated by the chain $\{X_t\}$ up to time n. Denote by $\mathbb{P}_{\mathcal{F}_n}$ the probability conditioned to $\mathcal{F}_n$. Take $\varepsilon > 0$.

By Chebyshev's inequality,

$$\mathbb{P}_{\mathcal{F}_n}\left(\left|\frac{\sum_{t=0}^{n} W_t \mathbb{I}_C\left\{X_t = A_n\right\}}{T_n\left(C\right)}\right| > \varepsilon\right) \leqslant \frac{\sigma^2 \sum_{t=0}^{n} \mathbb{I}_C\left\{X_t = A_n\right\}}{\varepsilon^2 {T_n}^2\left(C\right)} \leqslant \frac{\sigma^2}{\varepsilon^2 T_n\left(C\right)},$$

which implies the first part of the Lemma because $T_n(C) \to \infty$ with probability 1.

For the second part, let $\gamma_n(u)$ be the number of times the chain visits $(A_n, u] \cap C$ up to time n and $A_n\left(u\right) = \left\{t \leqslant n : X_t \in (A_n, u] \cap C\right\} = \left\{a_1, \ldots, a_{\gamma_n(u)}\right\}$ the times of those visits. Using that $\gamma_n = \sup_{u > A_n} \gamma_n\left(u\right) \leqslant T_n\left(C\right)$ and Kolmogorov's inequality (Th 3.1.6, pp 122 in [55]) we obtain,

$$\begin{aligned}
\mathbb{P}_{\mathcal{F}_n}\left(\sup_{u>A_n}\left|\frac{\sum_{t=0}^{n} W_t \mathbb{I}_C\left\{X_t \in (A_n, u]\right\}}{T_n\left(C\right)}\right| > \varepsilon\right) &= \mathbb{P}_{\mathcal{F}_n}\left(\sup_{u>A_n}\left|\sum_{i=1}^{\gamma_n(u)} \frac{W_{t_{a_i}}}{T_n\left(C\right)}\right| > \varepsilon\right) \\
&\leqslant \mathbb{P}_{\mathcal{F}_n}\left(\sup_{1 \leqslant k \leqslant \gamma_n}\left|\sum_{i=1}^{k} \frac{W_{t_{a_i}}}{T_n\left(C\right)}\right| > \varepsilon\right) \\
&\leqslant \frac{\sigma^2}{\varepsilon^2 T_n\left(C\right)}.
\end{aligned}$$

which by the same argument as before, implies the second part of the Lemma.

Proof of Lemma 5.6.5. In the sequel we set $a = f_0(x_0)$. We begin with the proof of (5.37).

Fix $\varepsilon > 0$ arbitrarily, and let $\nu > 0$ and $\gamma > 0$ be such that $|F^{-1}(u) - x_0| > \nu$ for all u such that $|u - F(x_0)| \geqslant \varepsilon/2$, and $|f_0(x_0) - f_0(y)| > \gamma$ for all y such that

$|y - x_0| \geqslant \nu/2$. Note that existence of ν and γ is ensured by assumptions (A5) and (A6).

By Lemma 5.6.3, we can assume without loss of generality that $F(x_0)$ belongs to the domain $[F_n(Y_1), F_n(Y_m)]$ of Λ_n, since this occurs with probability that tends to one. Therefore, we can find $j(x_0)$ such that $Y_{j(x_0)} = {F_n}^{-1}(F(x_0))$. It follows from the characterization in (5.36) that the event $E_n^1 := \{\widehat{U}_n(a) > F(x_0) + \varepsilon\}$ is contained in the event that there exists $p \in \mathcal{K}$ such that $p > F(x_0) + \varepsilon$ and

$$\Lambda_n(p) - ap \geqslant \Lambda_n(F(x_0)) - aF(x_0),$$

where we recall that $a = f_0(x_0)$.

By Lemma 5.6.1, E_n^1 is contained in the event that there exists $p \in \mathcal{K}$ such that $p > F(x_0) + \varepsilon$ and

$$L_n(p) + M_n(p) - ap \geqslant L_n(F(x_0)) + M_n(F(x_0)) - aF(x_0) \tag{5.54}$$

Using (5.27) in (5.54) we obtain that E_n^1 is contained in the event that there exists $p \in \mathcal{K}$ such that $p > F(x_0) + \varepsilon$ and

$$\int_{t_0/T_n(C)}^{p} f_0 \circ F_n^{-1}(u)du + S_n - ap \geqslant \int_{t_0/T_n(C)}^{F(x_0)} f_0 \circ F_n^{-1}(u)du - aF(x_0),$$

where

$$S_n = \sup_{p > F(x_0)+\varepsilon,\, p \in \mathcal{K}} \{M_n(p) - M_n(F(x_0))\}.$$

Let j and k such that $Y_{j+1} = F_n^{-1}(F(x_0))$ and $p = F_n(Y_k)$. By equation (5.28) we

have $M_n(p) - M_n(F(x_0)) = M_n^k - M_n^j - R_n^j(F(x_0))$, therefore,

$$\begin{aligned}
S_n &= \sup_{\substack{p>F(x_0)+\varepsilon \\ p\in\mathcal{K}}} \left\{M_n^k - M_n^j\right\} - R_n^j\left(F\left(x_0\right)\right) \\
&\leqslant \sup_{\substack{p>F(x_0)+\varepsilon \\ p\in\mathcal{K}}} \left| \frac{1}{T_n(C)} \sum_{t=0}^{n} W_t \left(\mathbb{I}_C\{X_t \leqslant F_n^{-1}(p)\} - \mathbb{I}_C\{X_t \leqslant F_n^{-1}(F(x_0))\}\right)\right| \\
&\quad + \left|R_n^j\left(F_n\left(Y_{j+1}\right)\right)\right| \\
&\leqslant \sup_{\substack{p>F(x_0)+\varepsilon \\ p\in\mathcal{K}}} \left| \frac{1}{T_n(C)} \sum_{t=0}^{n} W_t \mathbb{I}_C\left\{X_t \in \left(F_n^{-1}\left(F\left(x_0\right)\right); F_n^{-1}(p)\right]\right\}\right| \\
&\quad + \frac{\left|\sum_{t=0}^{n} W_t \mathbb{I}_C\{X_t = F_n^{-1}\left(F\left(x_0\right)\right)\}\right|}{T_n(C)}.
\end{aligned}$$

Hence,

$$\begin{aligned}
T_n(C) S_n \leqslant \sup_{\substack{p>F(x_0)+\varepsilon \\ p\in\mathcal{K}}} &\left| \sum_{t=0}^{n} W_t \mathbb{I}_C\left\{X_t \in \left(F_n^{-1}\left(F\left(x_0\right)\right); F_n^{-1}(p)\right]\right\}\right| \\
&+ \left| \sum_{t=0}^{n} W_t \mathbb{I}_C\left\{X_t = F_n^{-1}\left(F\left(x_0\right)\right)\right\}\right|.
\end{aligned}$$

Therefore, the event E_n^1 is contained in the event that there exists $p > F(x_0) + \varepsilon$ such that

$$\int_{F(x_0)}^{p} f_0 \circ F_n^{-1}(u)du + S_n \geqslant a(p - F(x_0)).$$

Now, let E_n^2 be the event that

$$\sup_{|u-F(x_0)|\leqslant\varepsilon} |F_n^{-1}(u) - F^{-1}(u)| \leqslant \eta$$

where $\eta \in (0, \nu/4)$ is such that $|f_0(y) - f_0(x_0)| \leqslant \gamma/2$ for all y such that $|x_0 - y| \leqslant \eta$. Note that the existence of η is ensured by assumption (A7). Then, it follows from

the monotonicity of f_0 and F_n that on E_n^2,

$$\int_{F(x_0)}^{p} f_0 \circ F_n^{-1}(u)du \leqslant \int_{F(x_0)}^{F(x_0)+\varepsilon/2} f_0(F^{-1}(u)-\eta)du + \int_{F(x_0)+\varepsilon/2}^{p} f_0(F_n^{-1}(F(x_0)+\varepsilon/2))du.$$

Hence, it follows from the definitions of η, ν and γ that on E_n^2,

$$\begin{aligned}\int_{F(x_0)}^{p} f_0 \circ F_n^{-1}(u)du &\leqslant \frac{\varepsilon}{2}f_0(x_0) + \frac{\gamma\varepsilon}{4} + (p - F(x_0) - \varepsilon/2)f_0(F^{-1}(F(x_0)+\varepsilon/2) - \eta) \\ &\leqslant \frac{\varepsilon}{2}f_0(x_0) + \frac{\gamma\varepsilon}{4} + (p - F(x_0) - \varepsilon/2)f_0(x_0 + \nu/2) \\ &\leqslant \frac{\varepsilon}{2}f_0(x_0) + \frac{\gamma\varepsilon}{4} + (p - F(x_0) - \varepsilon/2)(f_0(x_0) - \gamma).\end{aligned}$$

This implies that on E_n^2,

$$\begin{aligned}\int_{F(x_0)}^{p} f_0 \circ F_n^{-1}(u)du &\leqslant a(p - F(x_0)) - (p - F(x_0) - 3\varepsilon/4)\gamma \\ &\leqslant a(p - F(x_0)) - \varepsilon\gamma/4\end{aligned}$$

for all $p > F(x_0) + \varepsilon$. Hence, the event $E_n^1 \cap E_n^2$ is contained in the event $\{S_n \geqslant \varepsilon\gamma/4\}$. Now, on E_n^2, for all $p > F(x_0) + \varepsilon$ we have

$$\begin{aligned}F_n^{-1}(p) &\geqslant F_n^{-1}(F(x_0) + \varepsilon) \\ &\geqslant F^{-1}(F(x_0) + \varepsilon) - \eta \\ &\geqslant x + \nu - \eta \\ &\geqslant F_n^{-1}(F(x_0)) + \nu - 2\eta \\ &\geqslant F_n^{-1}(F(x_0)) + \nu/2,\end{aligned}$$

since $\nu > 4\eta$. Therefore,

$$T_n(C)S_n \leqslant \sup_{u>F_n^{-1}(F(x_0))+\nu/2} \left| \sum_{t=0}^{n} W_t \mathbb{I}_C\{X_t \in (F_n^{-1}(F(x_0)), u]\} \right| + \\ + \left| \sum_{t=0}^{n} W_t \mathbb{I}_C\{X_t = F_n^{-1}(F(x_0))\} \right|.$$

Hence, it follows from Lemma 5.6.4 that S_n converges in probability to zero as $n \to \infty$, so that the probability of the event $\{S_n \geqslant \varepsilon\gamma/4\}$ tends to zero as $n \to \infty$. It follows from Lemma 5.5.1 that for ε sufficiently small, the probability of the event E_n^2 tends to one as $n \to \infty$, so we conclude that the probability of E_n^1 tends to zero as $n \to \infty$. Similarly, the probability of the event $\{\widehat{U}_n(a) < F(x_0) - \varepsilon\}$ tends to zero as $n \to \infty$, so that

$$\lim_{n\to\infty} \mathbb{P}(|\widehat{U}_n(a) - F(x_0)| > \varepsilon) = 0$$

for all $\varepsilon > 0$. This completes the proof of (5.37).

5.7.3 Proof of Theorem 5.3.1

We first prove (5.15). Fix $\varepsilon > 0$ arbitrarily small. It follows from (5.35) and (5.34) that

$$\begin{aligned} \mathbb{P}\left(\widehat{f}_n^{-1}(a) > x_0 + \varepsilon\right) &\leqslant \mathbb{P}\left(F_n^{-1} \circ \widehat{U}_n(a) > x_0 + \varepsilon\right) \\ &\leqslant \mathbb{P}\left(\widehat{U}_n(a) \geqslant F_n(x_0 + \varepsilon)\right) \\ &\leqslant \mathbb{P}\left(\widehat{U}_n(a) \geqslant F(x_0 + \varepsilon) - K_n\right), \end{aligned}$$

where

$$K_n = \sup_{|y-x_0|\leqslant\varepsilon} |F_n(y) - F(y)|.$$

With $\nu := F(x_0+\varepsilon) - F(x_0)$, we obtain

$$\mathbb{P}\left(\widehat{f}_n^{-1}(a) > x_0+\varepsilon\right) \;\leqslant\; \mathbb{P}\left(\widehat{U}_n(a) \geqslant F(x_0)+\nu-K_n\right),$$

and ν is strictly positive since F is strictly increasing in the neighborhood of x_0. Hence, it follows from (5.17) that for sufficiently small $\varepsilon > 0$ one has

$$\mathbb{P}\left(\widehat{f}_n^{-1}(a) > x_0+\varepsilon\right) \;\leqslant\; \mathbb{P}\left(\widehat{U}_n(a) \geqslant F(x_0)+\nu/2\right) + o(1),$$

so it follows from (5.37) that the probability that $\widehat{f}_n^{-1}(a) > x_0+\varepsilon$ tends to zero as $n \to \infty$. Similarly, the probability that $\widehat{f}_n^{-1}(a) < x_0-\varepsilon$ tends to zero as $n \to \infty$ so we conclude that the probability that $|\widehat{f}_n^{-1}(a) - x_0| > \varepsilon$ tends to zero as $n \to \infty$. This completes the proof of (5.15). □

To prove (5.14), fix $\varepsilon > 0$ sufficiently small so that F and f_0 are continuous and strictly increasing in the neighborhood of $x' := f_0^{-1}(f_0(x_0)+\varepsilon)$. Equation (5.15) shows that

$$\widehat{f}_n^{-1}(f_0(x_0)+\varepsilon) = f_0^{-1}(f_0(x_0)+\varepsilon) + o_P(1), \tag{5.55}$$

as $n \to \infty$. Now, it follows from the switch relation (5.32) that

$$\begin{aligned}\mathbb{P}\left(\widehat{f}_n(x_0) > f_0(x_0)+\varepsilon\right) &\leqslant \mathbb{P}\left(\widehat{f}_n^{-1}(f_0(x_0)+\varepsilon) \geqslant x\right) \\ &\leqslant \mathbb{P}\left(\widehat{f}_n^{-1}(f_0(x_0)+\varepsilon) \geqslant f_0^{-1}(f_0(x_0)+\varepsilon)+\nu\right),\end{aligned} \tag{5.56}$$

where $\nu := x - f_0^{-1}(f_0(x_0)+\varepsilon) > 0$. It follows from (5.55) that the probability on

the right-hand side tends to zero as $n \to \infty$. Hence, the probability on the left-hand side tends to zero as well as $n \to \infty$.

Similarly, the probability that $\widehat{f}_n(x_0) < f_0(x_0) - \varepsilon$ tends to zero as $n \to \infty$ so we conclude that the probability that $|\widehat{f}_n(x_0) - f_0(x_0)| > \varepsilon$ tends to zero as $n \to \infty$. This completes the proof of Theorem 5.3.1.

5.7.4 Technical proofs for Section 5.6.2

Proof of Lemma 5.6.6. Let $\mathcal{F}_n = \sigma(\{X_0, \ldots, X_n\})$ be sigma algebra generated by the chain $\mathbf{X}$ up to time n. Denote by $\mathbb{E}_{\mathcal{F}_n}$ the expected value conditioned to $\mathcal{F}_n$. Take $0 < \gamma \leqslant \delta$ and define $I_0 = [x_0 - \gamma, x_0]$, $I_1 = [x_0, x_0 + \gamma]$ and

$$S_0(\gamma) = \sup_{y \in I_0} \left| \sum_{t=0}^{n} W_t \left(\mathbb{I}\{X_t \leqslant y\} - \mathbb{I}\{X_t \leqslant x_0\} \right) \right|^2$$
$$S_1(\gamma) = \sup_{y \in I_1} \left| \sum_{t=0}^{n} W_t \left(\mathbb{I}\{X_t \leqslant y\} - \mathbb{I}\{X_t \leqslant x_0\} \right) \right|^2$$

then,

$$S(\gamma) = \sup_{|y - x_0| \leqslant \gamma} \left| \sum_{t=0}^{n} W_t \left(\mathbb{I}\{X_t \leqslant y\} - \mathbb{I}\{X_t \leqslant x_0\} \right) \right|^2 = \max\left(S_0(\gamma), S_1(\gamma) \right),$$
$$\leqslant S_0(\gamma) + S_1(\gamma) \tag{5.57}$$

Following the notation of section 5.2, let

$$\alpha_n^{(0)}(\gamma) = \sup_{y \in I_0} T_n([y, x_0]) \quad , \quad \alpha_n^{(1)}(\gamma) = \sup_{y \in I_1} T_n([x_0, y]),$$

with this notation, $S_0 = \sup_{y \in I_0} \left| \sum_{i=1}^{T_n([y,x_0])} W_{\sigma_{[y,x_0)}(i)} \right|^2$ and $S_1 = \sup_{y \in I_1} \left| \sum_{i=1}^{T_n([x_0,y])} W_{\sigma_{[x_0,y]}(i)} \right|^2$.

By Doob's maximal inequality (Th 10.9.4 in [55]), we have, for $j = 0, 1$,

$$\begin{aligned}\mathbb{E}_{\mathcal{F}_n} S_j(\gamma) &\leqslant 4\mathbb{E}_{\mathcal{F}_n}\left(\sum_{i=1}^{\alpha_n^{(j)}} W_{t_i}\right)^2 = 4\sigma^2\alpha_n^{(j)}(\gamma) \\ &\leqslant 4\sigma^2 T_n([x_0-\gamma, x_0+\gamma]) \\ &\leqslant 4\sigma^2\sum_{t=0}^{n}\left(\mathbb{I}\{X_t \leqslant x_0+\gamma\} - \mathbb{I}\{X_t < x_0-\gamma\}\right). \end{aligned} \tag{5.58}$$

Therefore, by (5.57) and (5.58)

$$\mathbb{E}_{\mathcal{F}_n} S(\gamma) \leqslant 8\sigma^2\sum_{t=0}^{n}\Big(\mathbb{I}\{X_t \leqslant x_0+\gamma\} - \mathbb{I}\{X_t < x_0-\gamma\}\Big). \tag{5.59}$$

Define,

- $h(y,\gamma) = \mathbb{I}\{y \in [x_0-\gamma, x_0+\gamma]\}$,
- $h(\mathcal{B}_j,\gamma) = \begin{cases} \sum_{t=0}^{\tau_\alpha} h(X_t,\gamma) & , \quad j=0 \\ \sum_{t=\tau_A(j)+1}^{\tau_A(j+1)} h(X_t,\gamma) & , \quad j \geqslant 1 \end{cases}$
- $Z_n(\gamma) = \sum_{t=0}^{n} h(X_t,\gamma)$
- $\ell(\mathcal{B}_j) = \begin{cases} \tau_{\boldsymbol{\alpha}} & , \quad j=0 \\ \tau_{\boldsymbol{\alpha}}(j+1) - \tau_{\boldsymbol{\alpha}}(j) & , \quad j \geqslant 1 \end{cases}$
- $\widetilde{T}(n) = \min\left\{k : \sum_{i=0}^{k}\ell(\mathcal{B}_j) \geqslant n\right\}$.
- $\mathcal{G}_k = \sigma\left(\{(h(\mathcal{B}_j,\gamma), \ell(\mathcal{B}_j))\}_{j=0}^{k}\right)$ for $k \geqslant 0$.

By the Strong Markov property, $\{(h(\mathcal{B}_j,\gamma), \ell(\mathcal{B}_j))\}_{j=1}^{+\infty}$ is an i.i.d. sequence which is independent of $(h(\mathcal{B}_0,\gamma), \ell(\mathcal{B}_0))$ (and, therefore, of the initial measure

λ). For n fixed, the random variable $\widetilde{T}(n)$ is a stopping time for the sequence $\{(h(\mathcal{B}_j,\gamma),\ell(\mathcal{B}_j))\}_{j=0}^{+\infty}$, in effect

$$\left\{\widetilde{T}(n)=0\right\}=\{\ell(\mathcal{B}_0)\geqslant n\}\in\mathcal{G}_0,$$
$$\left\{\widetilde{T}(n)=k\right\}=\bigcap_{j=0}^{k-1}\left\{\sum_{i=0}^{j}\ell(\mathcal{B}_i)<n\right\}\bigcap\left\{\sum_{i=0}^{k}\ell(\mathcal{B}_j)\geqslant n\right\}\in\mathcal{G}_k\quad\forall k\geqslant 1.$$

For each n and γ we have that

$$\begin{aligned}Z_n(\gamma)&=\sum_{t=0}^{\tau_\alpha}h(X_t,\gamma)+\sum_{j=1}^{T(n)}h(\mathcal{B}_j,\gamma)+\sum_{t=t_\alpha(T(n))+1}^{n}h(X_t,\gamma)\\&\leqslant h(\mathcal{B}_0,\gamma)+\sum_{j=1}^{\widetilde{T}(n)}h(\mathcal{B}_j,\gamma).\end{aligned}\tag{5.60}$$

where the last inequality is justified by the fact that, $T(n)\leqslant\widetilde{T}(n)$ and $h(y,\gamma)$ is a nonnegative function. Because $\ell(\mathcal{B}_j)\geqslant 1$ for all j, we have that,

$$\sum_{j=1}^{\widetilde{T}(n)}h(\mathcal{B}_j,\gamma)=\sum_{j=1}^{n}h(\mathcal{B}_j,\gamma)\,\mathbb{I}\left\{\widetilde{T}(n)\geqslant j\right\},$$

then,

$$\mathbb{E}\left(\sum_{j=1}^{\widetilde{T}(n)}h(\mathcal{B}_j,\gamma)\right)=\sum_{j=1}^{n}\mathbb{E}\left(h(\mathcal{B}_j,\gamma)\,\mathbb{I}\left\{\widetilde{T}(n)\geqslant j\right\}\right).\tag{5.61}$$

For each j we have,

$$\mathbb{E}_\lambda\left(h(\mathcal{B}_j,\gamma)\,\mathbb{I}\left\{\widetilde{T}(n)\geqslant j\right\}\right)=\mathbb{E}_\lambda\left(\mathbb{E}\left(h(\mathcal{B}_j,\gamma)\,\mathbb{I}\left\{\widetilde{T}(n)\geqslant j\right\}|\mathcal{G}_{j-1}\right)\right)$$

Notice that $\mathbb{I}\left\{\widetilde{T}(n)\geqslant j\right\}=1-\mathbb{I}\left\{\widetilde{T}(n)\leqslant j-1\right\}\in\mathcal{G}_{j-1}$ and $h(\mathcal{B}_j,\gamma)$ is inde-

pendent of $\mathcal{G}_{j-1}$, therefore,

$$\mathbb{E}_\lambda\left(h\left(\mathcal{B}_j,\gamma\right)\mathbb{I}\left\{\widetilde{T}\left(n\right)\geqslant j\right\}\right)=\mathbb{E}_\lambda\left(\mathbb{I}\left\{\widetilde{T}\left(n\right)\geqslant j\right\}\right)\mathbb{E}\left(h\left(\mathcal{B}_j,\gamma\right)\right).$$

Plugging this into equation (5.61) we get,

$$\mathbb{E}_\lambda\left(\sum_{j=1}^{\widetilde{T}(n)}h\left(\mathcal{B}_j,\gamma\right)\right)=\sum_{j=1}^{n}\mathbb{E}\left(h\left(\mathcal{B}_j,\gamma\right)\right)\mathbb{P}_\lambda\left(\widetilde{T}\left(n\right)\geqslant j\right)\leqslant\mathbb{E}\left(h\left(\mathcal{B}_1,\gamma\right)\right)\mathbb{E}_\lambda\widetilde{T}\left(n\right).$$

Then, by taking expectation in (5.60) we obtain

$$\begin{aligned}\mathbb{E}_\lambda Z_n\left(\gamma\right)&\leqslant\mathbb{E}_\lambda h\left(\mathcal{B}_0,\gamma\right)+\mathbb{E}\left(h\left(\mathcal{B}_1,\gamma\right)\right)\mathbb{E}_\lambda\widetilde{T}\left(n\right)\\&\leqslant\mathbb{E}_\lambda h\left(\mathcal{B}_0,\gamma\right)+\mathbb{E}\left(h\left(\mathcal{B}_1,\gamma\right)\right)\mathbb{E}_\lambda\left(T\left(n\right)+1\right).\end{aligned}\tag{5.62}$$

By Theorem 5.2.1 and the fact that F is Lipschitz we can find K_1 independent of γ such that,

$$\begin{aligned}\mathbb{E}\left(h\left(\mathcal{B}_1,\gamma\right)\right)&=\int h\left(t,\gamma\right)d\pi\left(t\right)=K_\pi\pi\left(C\right)\left(F\left(x_0+\gamma\right)-F\left(x_0-\gamma\right)\right)\\&\leqslant K_1\gamma.\end{aligned}\tag{5.63}$$

If $\mathbf{X}$ is positive recurrent, by Theorem 5.2.1, $\frac{T(n)}{u(n)}$ converges almost surely to a positive constant $K_2>0$. Moreover, $\frac{T(n)}{u(n)}\leqslant 1$ therefore, by the Dominated Convergence Theorem we obtain that $\mathbb{E}_\lambda T\left(n\right)\sim\frac{u(n)}{K_2}$. If $\mathbf{X}$ is β-null recurrent, by Lemma 3.3 in [67], $\mathbb{E}_\lambda T\left(n\right)\sim\frac{u(n)}{\Gamma(1+\beta)}$, hence, for both positive and β-null recurrent chains, we can find K_2 and N, both independent of γ, such that $\mathbb{E}_\lambda T\left(n\right)\leqslant K_2u\left(n\right)$ for

all $n \geqslant N$. Using this with (5.62) and (5.63) we get,

$$\frac{\mathbb{E}_\lambda Z_n(\gamma)}{u(n)\gamma} \leqslant \frac{\mathbb{E}_\lambda h(\mathcal{B}_0, \gamma)}{u(n)\gamma} + K_1 K_2 \quad \forall n \geqslant N,\ \forall \gamma \in (0, \delta]. \tag{5.64}$$

Combining (5.64) with assumption (B3) and the fact that $Z_n(0) \equiv 0$ we obtain that there exist positive constants K_3 and γ_0 such that

$$\mathbb{E}_\lambda Z_n(\gamma) \leqslant u(n)\gamma \quad \forall n \geqslant N,\ \forall \gamma \in (0, \gamma_0].$$

Equation (5.38) now follows after taking expectation in (5.59). The proof of (5.39) follows the same reasoning, but using

$$S_j(\gamma) = \sup_{y \in I_j} \left| \sum_{t=0}^{n} W_t \left(\mathbb{I}_C \{X_t = y\} \right) \right|^2.$$

Proof of Lemma 5.6.7. a) If **X** is positive recurrent, Theorem 5.2.1 implies that there exists a positive constant K such that $\frac{T_n(C)}{u(n)}$ converges almost surely to $K\pi(C)$, which is not zero by (A3).

On the other hand, if **X** is β-null recurrent, Theorem 5.2.1 and Slutsky's Theorem implies that there exists a constant $K > 0$ such that $\frac{T_n(C)}{u(n)}$ converges in distribution to $KM_\beta(1)$ where $M_\beta(1)$ denotes a Mittag-Leffler distribution with parameter β. This distribution is continuous and strictly positive with probability 1, then, by the Continuous Mapping Theorem, $\frac{u(n)}{T_n(C)}$ converges in distribution to a multiple of $\frac{1}{M_\beta}$, therefore, $\frac{u(n)}{T_n(C)}$ is bounded in probability by Theorem 2.4 in [114].

b) Let **X** be positive recurrent, then, we can find N_η such that

$$\mathbb{P}\left(\left| \left(\frac{T_n(C)}{u(n)} \right)^\alpha - K^\alpha \pi(C)^\alpha \right| \leqslant \left(\frac{K\pi(C)}{2} \right)^\alpha \right) \geqslant 1 - \eta, \quad \forall n \geqslant N_\eta.$$

hence,

$$\mathbb{P}\left(\left(\frac{T_n(C)}{u(n)}\right)^{\alpha} \in \left[\frac{K^{\alpha}\pi(C)^{\alpha}}{2}, \frac{3K^{\alpha}\pi(C)^{\alpha}}{2}\right]\right) \geqslant 1-\eta, \quad \forall n \geqslant N_{\eta}.$$

Now let $\mathbf{X}$ be β-null recurrent. Let $Z = (KM_{\beta}(1))^{\alpha}$, This random variable is continuous and positive, therefore, we can find positive constants $\underline{c}_{\eta}$ and $\overline{c}_{\eta}$ such that

$$\mathbb{P}\left(Z \in \left[\underline{c}_{\eta}, \overline{c}_{\eta}\right]\right) \geqslant 1 - \frac{\eta}{2}. \tag{5.65}$$

By the Continuous Mapping Theorem, $\left(\frac{T_n(C)}{u(n)}\right)^{\alpha}$ converges in distribution to Z, therefore, we can find $N_{\eta} \in \mathbb{N}$ such that

$$\left|\mathbb{P}\left(\left(\frac{T_n(C)}{u(n)}\right)^{\alpha} \in \left[\underline{c}_{\eta}, \overline{c}_{\eta}\right]\right) - \mathbb{P}\left(Z \in \left[\underline{c}_{\eta}, \overline{c}_{\eta}\right]\right)\right| \leqslant \frac{\eta}{2}, \quad \forall n \geqslant N_{\eta}, \tag{5.66}$$

Combining (5.65) and (5.66) we obtain that

$$\mathbb{P}\left(\left(\frac{T_n(C)}{u(n)}\right)^{\alpha} \in \left[\underline{c}_{\eta}, \overline{c}_{\eta}\right]\right) \geqslant 1-\eta, \quad \forall n \geqslant N_{\eta}. \tag{5.67}$$

Proof of Lemma 5.6.8. Fix $\varepsilon \in (0,1)$ small enough so that F' and $|f_0'|$ are bounded from above and away from zero on $[F^{-1}(F(x_0)-2\varepsilon), F^{-1}(F(x_0)+2\varepsilon)]$, see the assumption (B2). Then, the proper inverse functions of F and f_0 are well defined on $[F(x_0)-2\varepsilon, F(x_0)+2\varepsilon]$ and

$$[f_0 \circ F^{-1}(F(x_0)-2\varepsilon), f_0 \circ F^{-1}(F(x_0)+2\varepsilon)]$$

respectively. We denote the inverses on that intervals by F^{-1} and f_0^{-1} respectively.

Let

$$U_n(a) = \underset{|p-F(x_0)|\leqslant\varepsilon}{\operatorname{argmax}} \{\Lambda_n(p) - ap\} \tag{5.68}$$

where $a = f_0(x_0)$ and where the supremum is restricted to $p \in [F_n(Y_0), F_n(Y_m)]$. We will show below that

$$U_n(a) = F(x_0) + O_P(u\,(n)^{-1/3}), \tag{5.69}$$

as $n \to \infty$. Combining (5.36) to Lemma 5.6.5 ensures that $\widehat{U}_n(a)$ coincides with $U_n(a)$ with probability that tends to one as $n \to \infty$, so (5.40) follows from (5.69).

We turn to the proof of (5.69). Fix $\eta > 0$ arbitrarily and let

$$\gamma_n = K_0 u\,(n)^{-1/3} \tag{5.70}$$

for some $K_0 \geqslant 1$ sufficiently large so that

$$\gamma_n \geqslant \frac{1}{\sqrt{u\,(n)}}. \tag{5.71}$$

Then, by part ii) of Lemma 5.6.7, we can find positive constants $\underline{c}_\eta$, $\bar{c}_\eta$ and N_η such that

$$\mathbb{P}\left(T_n(C)^{2/3}\gamma_n u\,(n)^{-1/3} \in [K_0\underline{c}_\eta, K_0\bar{c}_\eta]\right) \geqslant 1 - \eta/2 \quad \forall n \geqslant N_\eta, \tag{5.72}$$

Let $\underline{c} = K_0\underline{c}_\eta$ and $\bar{c} = K_0\bar{c}_\eta$. It follows from (5.23) that for sufficiently small $\varepsilon > 0$, we can find $K_1 > 0$ such that

$$\mathbb{P}\left(T_n(C) \sup_{|p-F(x_0)|\leqslant 2\varepsilon} |F_n^{-1}(p) - F^{-1}(p)|^2 \leqslant K_1\right) \geqslant 1 - \eta/2$$

for all n. Hence for $n \geqslant \mathbb{N}_\eta$,

$$\mathbb{P}(\mathcal{E}_n) \geqslant 1 - \eta,$$

where $\mathcal{E}_n$ denotes the intersection of the events

$$T_n(C)^{2/3}\gamma_n u\,(n)^{-1/3} \in [\underline{c}, \bar{c}] \tag{5.73}$$

and

$$T_n(C) \sup_{|p-F(x_0)|\leqslant 2\varepsilon} |F_n^{-1}(p) - F^{-1}(p)|^2 \leqslant K_1. \tag{5.74}$$

Combining equations (5.73) and (5.74), we obtain that, in $\mathcal{E}_n$,

$$\sup_{|p-F(x_0)|\leqslant 2\varepsilon} \left|F_n^{-1}(p) - F^{-1}(p)\right|^2 \leqslant K_2 a(n)^{-1} \tag{5.75}$$

where $K_2 = K_1\left(\frac{K_0}{\underline{c}}\right)^{3/2}$ is independent of n and K_0.

By Lemma 5.6.3, we can assume without loss of generality that $F(x_0)$ belongs to $[F_n(Y_0), F_n(Y_m)]$, since this occurs with probability that tends to one. Hence, by (5.68), the event $\{|U_n(a) - F(x_0)| \geqslant \gamma_n\}$ is contained in the event that there exists $p \in [F_n(Y_0), F_n(Y_m)]$ with $|p - F(x_0)| \leqslant \varepsilon$, $|p - F(x_0)| \geqslant \gamma_n$ and

$$\Lambda_n(p) - ap \geqslant \Lambda_n(F(x_0)) - aF(x_0). \tag{5.76}$$

Obviously, the probability is equal to zero if $\gamma_n > \varepsilon$ so we assume in the sequel that $\gamma_n \leqslant \varepsilon$. For all $p \in [F(x_0) - \varepsilon, F(x_0) + \varepsilon]$ define

$$\Lambda(p) = \int_{F(x_0)}^{p} f_0 \circ F^{-1}(u)du.$$

Let $c > 0$ such that $|f_0'|/F' > 2c$ on the interval $[F^{-1}(F(x_0)-2\varepsilon), F^{-1}(F(x_0)+$

$2\varepsilon)]$. Since $\Lambda'(F(x_0)) = a$ and $\Lambda'' = f_0' \circ F^{-1}/F' \circ F^{-1}$, it then follows from Taylor's expansion that

$$\Lambda(p) - \Lambda(F(x_0)) \leqslant (p - F(x_0))a - c(p - F(x_0))^2$$

for all $p \in [F(x_0) - \varepsilon, F(x_0) + \varepsilon]$ and therefore, (5.76) implies that

$$\Delta_n(p) - \Delta_n(F(x_0)) - c(p - F(x_0))^2 \geqslant 0$$

for all such p's, where we set $\Delta_n := \Lambda_n - \Lambda$. Hence, for all $n \geqslant N_\eta$,

$$\begin{aligned}
&\mathbb{P}\left(|U_n(a) - F(x_0)| \geqslant \gamma_n\right) \\
&\quad \leqslant \eta + \mathbb{P}\left(\sup_{|p - F(x_0)| \in [\gamma_n, \varepsilon]} \{\Delta_n(p) - \Delta_n(F(x_0)) - c(p - F(x_0))^2\} \geqslant 0 \text{ and } \mathcal{E}_n\right) \\
&\quad \leqslant \eta + \sum_j \mathbb{P}\left(\sup_{|u| \in [\gamma_n 2^j, \gamma_n 2^{j+1}]} \{\Delta_n(F(x_0) + u) - \Delta_n(F(x_0))\} \geqslant c(\gamma_n 2^j)^2 \text{ and } \mathcal{E}_n\right) \\
&\quad \leqslant \eta + \sum_j \mathbb{P}\left(\sup_{|u| \leqslant \gamma_n 2^{j+1}} |\Delta_n(F(x_0) + u) - \Delta_n(F(x_0))| \geqslant c(\gamma_n 2^j)^2 \text{ and } \mathcal{E}_n\right) \quad (5.77)
\end{aligned}$$

where the sums are taken over all integers $j \geqslant 0$ such that $\gamma_n 2^j \leqslant \varepsilon$. Recall that we have (5.53) for all $k \in \{0, \ldots, m\}$. Since Λ_n is piecewise-linear with knots at $F_n(Y_0), \ldots, F_n(Y_m)$, by (5.27) and (5.28) we get that for every j in the above sum,

$$\begin{aligned}
&\sup_{|u| \leqslant \gamma_n 2^{j+1}} |\Delta_n(F(x_0) + u) - \Delta_n(F(x_0))| \\
&\qquad \leqslant \sup_{|u| \leqslant \gamma_n 2^{j+1}} \left| \int_{F(x_0)}^{F(x_0)+u} \left(f_0 \circ F_n^{-1}(y) - f_0 \circ F^{-1}(y)\right) dy \right| \\
&\qquad\quad + \sup_{|u| \leqslant \gamma_n 2^{j+1}} |M_n(F(x_0) + u) - M_n(F(x_0))| . \quad (5.78)
\end{aligned}$$

Moreover, $|f_0'|$ is bounded above on $[F^{-1}(F(x_0)-2\varepsilon), F^{-1}(F(x_0)+2\varepsilon)]$, so we obtain that for every j with $\gamma_n 2^j \leqslant \varepsilon$, the first term on the right-hand side of (5.78) satisfies

$$\begin{aligned}
&\sup_{|u|\leqslant\gamma_n 2^{j+1}} \left| \int_{F(x_0)}^{F(x_0)+u} \left(f_0 \circ F_n^{-1}(p) - f_0 \circ F^{-1}(p)\right) dp \right| \\
&\qquad \leqslant \int_{F(x_0)-\gamma_n 2^{j+1}}^{F(x_0)+\gamma_n 2^{j+1}} \left| f_0 \circ F_n^{-1}(p) - f_0 \circ F^{-1}(p) \right| dp \\
&\qquad \leqslant K_3 \gamma_n 2^j \sup_{|p-F(x_0)|\leqslant 2\varepsilon} \left| F_n^{-1}(p) - F^{-1}(p) \right|,
\end{aligned}$$

for some $K_3 > 0$ that does not depend on n. Hence, it follows from the previous display and (5.75) that

$$\begin{aligned}
&\mathbb{E}\left(\sup_{|u|\leqslant\gamma_n 2^{j+1}} \left| \int_{F(x_0)}^{F(x_0)+u} (f_0 \circ F_n^{-1}(p) - f_0 \circ F^{-1}(p) dp \right|^2 \mathbb{I}(\mathcal{E}_n) \right) \\
&\qquad \leqslant K_3^2 \gamma_n^2 2^{2j} \mathbb{E}\left(\sup_{|p-F(x_0)|\leqslant 2\varepsilon} |F_n^{-1}(p) - F^{-1}(p)|^2 \mathbb{I}(\mathcal{E}_n) \right) \\
&\qquad \leqslant K_3^2 \gamma_n^2 2^{2j} K_2 u\,(n)^{-1}.
\end{aligned}$$

Hence, taking $K_4 = K_3^2 K_2$ we get that for all j with $\gamma_n 2^j \leqslant \varepsilon \leqslant 1$.

$$\mathbb{E}\left(\sup_{|u|\leqslant\gamma_n 2^{j+1}} \left| \int_{F(x_0)}^{F(x_0)+u} (f_0 \circ F_n^{-1}(p) - f_0 \circ F^{-1}(p) dp \right|^2 \mathbb{I}(\mathcal{E}_n) \right) \leqslant K_4 \gamma_n 2^j u\,(n)^{-1}. \tag{5.79}$$

By equations (5.28) and (5.29) in Lemma 5.6.1, the second term on the right-hand side of (5.78) satisfies,

$$\sup_{|u|\leqslant\gamma_n 2^{j+1}} |M_n(F(x_0)+u) - M_n(F(x_0))| \leqslant I_1^{n,j} + I_2^{n,j}, \tag{5.80}$$

where $I_1^{n,j}$ and $I_2^{n,j}$ are given by

$$I_1^{n,j} = \frac{1}{T_n(C)} \sup_{|u|\leqslant\gamma_n 2^{j+1}} \left| \sum_{t=0}^{n} W_t \Big(\mathbb{I}_C\{X_t \leqslant F_n^{-1}(F(x_0)+u)\} - \mathbb{I}_C\{X_t \leqslant F_n^{-1}(F(x_0))\} \Big) \right|,$$
$$I_2^{n,j} = \frac{2}{T_n(C)} \sup_{|u|\leqslant\gamma_n 2^{j+1}} \left| \sum_{t=0}^{n} W_t \Big(\mathbb{I}_C\big\{X_t = F_n^{-1}\left(F\left(x_0\right)+u\right)\big\} \Big) \right|.$$

For $I_1^{n,j}$, it follows from the triangle inequality that

$$I_1^{n,j} \leqslant \frac{2}{T_n(C)} \sup_{|u|\leqslant\gamma_n 2^{j+1}} \left| \sum_{t=0}^{n} W_t \left(\mathbb{I}_C\{X_t \leqslant F_n^{-1}(F(x_0)+u)\} - \mathbb{I}_C\{X_t \leqslant x_0\} \right) \right|.$$

Combining (5.75) and the fact that F^{-1} is Lipschitz in $[F(x_0)-2\varepsilon, F(x_0)+2\varepsilon]$ we can find $K_5 = \max\left(\sqrt{K_2}, \sup\left(F^{-1}\right)\right)$ independent of n such that, on $\mathcal{E}_n$,

$$\sup_{|p-F(x_0)|\leqslant 2\varepsilon} |F_n^{-1}(p) - F^{-1}(p)| \leqslant \frac{K_5}{\sqrt{u\left(n\right)}}$$

and $|F^{-1}(F(x_0)+u) - x| = |F^{-1}(F(x_0)+u) - F^{-1}(F(x_0))| \leqslant K_5|u|/2$ for all u with $|u| \leqslant 2\varepsilon$. Hence, on $\mathcal{E}_n$

$$I_1^{n,j} \leqslant \frac{2}{T_n(C)} \sup_{|y-x_0|\leqslant K_5\gamma_n 2^j + K_5/\sqrt{u(n)}} \left| \sum_{t=0}^{n} W_t \left(\mathbb{I}_C\{X_t \leqslant y\} - \mathbb{I}_C\{X_t \leqslant x_0\} \right) \right|,$$
$$I_2^{n,j} \leqslant \frac{2}{T_n\left(C\right)} \sup_{|y-x_0|\leqslant K_5\gamma_n 2^j + K_5/\sqrt{u(n)}} \left| \sum_{t=0}^{n} W_t \Big(\mathbb{I}_C\{X_t = y\} \Big) \right|.$$

It follows from (5.71) that $\gamma_n 2^j \geqslant \gamma_n \geqslant 1/\sqrt{u\left(n\right)}$ for all $j \geqslant 0$, then, on $\mathcal{E}_n$

$$I_1^{n,j} \leqslant \frac{2}{T_n(C)} \sup_{|y-x_0|\leqslant 2K_5\gamma_n 2^j} \left| \sum_{t=0}^{n} W_t \left(\mathbb{I}_C\{X_t \leqslant y\} - \mathbb{I}_C\{X_t \leqslant x_0\} \right) \right|,$$
$$I_2^{n,j} \leqslant \frac{2}{T_n\left(C\right)} \sup_{|y-x_0|\leqslant 2K_5\gamma_n 2^j} \left| \sum_{t=0}^{n} W_t \Big(\mathbb{I}_C\{X_t = y\} \Big) \right|.$$

By Lemma 5.6.6, we conclude that there exists $K_6 > 0$ and N'_η such that, for $n \geqslant N'_\eta$

$$\mathbb{E}\Big(\left(I_1^{n,j} + I_2^{n,j}\right)^2 \mathbb{I}(\mathcal{E}_n)\Big) \leqslant K_6 \gamma_n 2^j u\,(n)^{-1} \tag{5.81}$$

Combining (5.78), (5.79), (5.80) and (5.81), we conclude that there exists $K_7 > 0$, independent of n and K_0, such that for all $n \geqslant N'_\eta$ and $j \geqslant 0$ where $\gamma_n 2^j \leqslant \varepsilon$, one has

$$\mathbb{E}\left(\sup_{|u| \leqslant \gamma_n 2^{j+1}} |\Delta_n(F(x_0)+u) - \Delta_n(F(x_0))|^2 \mathbb{I}(\mathcal{E}_n)\right) \leqslant K_7 \gamma_n 2^j u\,(n)^{-1}.$$

Combining this with (5.77) and the Markov inequality, we conclude that there exist $K_8 > 0$ and N''_η, that do not depend on n nor K_0, such that, for all $n \geqslant N''_\eta$,

$$\begin{aligned}\mathbb{P}\left(|U_n(a) - F(x_0)| \geqslant \gamma_n\right) &\leqslant \eta + K_8 \sum_{k \geqslant 0} \frac{\gamma_n 2^j u\,(n)^{-1}}{(\gamma_n 2^j)^4} \\ &\leqslant \eta + K_8 \gamma_n^{-3} u\,(n)^{-1} \sum_{j \geqslant 0} 2^{-3j}.\end{aligned}$$

The sum on the last line is finite, so there exists $K > 0$, independent of n and K_0, such that for n bigger than N''_η

$$\mathbb{P}\left(|U_n(a) - F(x_0)| \geqslant \gamma_n\right) \leqslant \eta + K \gamma_n^{-3} u\,(n)^{-1} = \eta + \frac{K}{K_0^3}. \tag{5.82}$$

The above probability can be made smaller than 2η by setting (5.70) for some sufficiently large K_0 independent of n. This proves (5.69) and completes the proof of (5.40).

Now, we turn to the proof of (5.41). It follows from (5.35) combined to (5.37) and

Lemma 5.5.3 that

$$\widehat{f}_n^{-1}(f_0(x_0)) = F^{-1} \circ \widehat{U}_n(f_0(x_0)) + T_n(C)^{-1/2} O_P(1).$$

Hence, by Lemma 5.6.7 we have

$$\widehat{f}_n^{-1}(f_0(x_0)) = F^{-1} \circ \widehat{U}_n(f_0(x_0)) + O_P\left(u(n)^{-1/2}\right).$$

Now, it follows from the assumption (B2) that F^{-1} has a bounded derivative in the neighborhood of $F(x_0)$, to which $\widehat{U}_n(f_0(x_0))$ belongs with probability that tends to one. Hence, it follows from Taylor's expansion that

$$\begin{aligned}\widehat{f}_n^{-1}(f_0(x_0)) &= F^{-1} \circ F(x_0) + O\left(|\widehat{U}_n(f_0(x_0)) - F(x_0)|\right) + O_P\left(u(n)^{-1/2}\right) \\ &= x + O_P(u(n)^{-1/3}) + O_P\left(u(n)^{-1/2}\right),\end{aligned}$$

where we used (5.40) for the last equality. This proves (5.41) and completes the proof of Lemma 5.6.8.

5.7.5 Proof of Theorem 5.4.1.

Inspecting the proof of Lemma 5.6.8, one can see that the convergences in (5.40) and (5.41) hold in a uniform sense in the neighborhood of x_0. More precisely, there exists $\gamma > 0$, independent on n, such that for all $\eta > 0$ we can find $K_1 > 0$ such that

$$\sup_{|a - f_0(x_0)| \leqslant \gamma} \mathbb{P}\left(\left|\widehat{U}_n(a) - F \circ f_0^{-1}(a)\right| > K_1 u(n)^{-1/3}\right) \leqslant \eta$$

and

$$\sup_{|a - f_0(x_0)| \leqslant \gamma} \mathbb{P}\left(\left|\widehat{f}_n^{-1}(a) - f_0^{-1}(a)\right| > K_1 u(n)^{-1/3}\right) \leqslant \eta.$$

Let $\varepsilon = K_1 u\,(n)^{-1/3}$ where $K_1 > 0$ does not depend on n, and recall (5.56) where $\nu = x - f_0^{-1}(f_0(x_0) + \varepsilon) > 0$. It follows from the assumption (B2) that f_0^{-1} has a derivative that is bounded in sup-norm away from zero in a neighborhood of $f_0(x_0)$. Hence, it follows from the Taylor expansion that there exists $K_2 > 0$ that depends only on f_0 such that $\nu \geqslant K_2\varepsilon$, provided that n is sufficiently large to ensure that $f_0(x_0) + \varepsilon$ belongs to this neighborhood of $f_0(x_0)$. Hence,

$$\begin{aligned}\mathbb{P}\left(\widehat{f}_n(x_0) > f_0(x_0) + \varepsilon\right) &\leqslant \mathbb{P}\left(\widehat{f}_n^{-1}(f_0(x_0) + \varepsilon) \geqslant f_0^{-1}(f_0(x_0) + \varepsilon) + K_2\varepsilon\right). \\ &\leqslant \sup_{|a - f_0(x_0)| \leqslant \gamma} \mathbb{P}\left(\left|\widehat{f}_n^{-1}(a) - f_0^{-1}(a)\right| > K_2 K_1 u\,(n)^{-1/3}\right),\end{aligned}$$

provided that n is sufficiently large to ensure that $f_0(x_0) + \varepsilon$ belongs to the above neighborhood of $f_0(x_0)$, and that $\gamma \geqslant Cu\,(n)^{-1/3}$. For fixed $\eta > 0$ one can choose $K_2 > 0$ such that the probability on the right-hand side of the previous display is smaller than or equal to η and therefore,

$$\lim_{n\to\infty} \mathbb{P}\left(\widehat{f}_n(x_0) > f_0(x_0) + K_2 u\,(n)^{-1/3}\right) \leqslant \eta.$$

Similarly, for all fixed $\eta > 0$, one can find K_3 that does not depend on n such that

$$\lim_{n\to\infty} \mathbb{P}\left(\widehat{f}_n(x_0) < f_0(x_0) - K_3 u\,(n)^{-1/3}\right) \leqslant \eta.$$

Hence, for all fixed $\eta > 0$, there exists $K > 0$ that independent of n such that

$$\lim_{n\to\infty} \mathbb{P}\left(|\widehat{f}_n(x_0) - f_0(x_0)| > K u\,(n)^{-1/3}\right) \leqslant \eta.$$

This completes the proof of Theorem 5.4.1.

References

[1] A. Sidra, H. S., and Nadeem, A. B. Zipf's law and city size distribution: A survey of the literature and future research agenda. *Physica A: Statistical Mechanics and Its Applications*, 492 (2018), 75–92.

[2] Adamczak, R. A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electronic Journal of Probability 13*, none (2008), 1000 – 1034.

[3] Alexander, K. Excursions and local limit theorems for bessel-like random walks. *Electronic Journal of Probability 16* (2011).

[4] Anscombe, F. J. Large-sample theory of sequential estimation. *Mathematical Proceedings of the Cambridge Philosophical Society 48* (10 1952).

[5] Asmussen, S. *Applied Probability and Queues*, 2nd ed. Stochastic Modelling and Applied Probability. Springer, 2010.

[6] Athreya, K., and Fuh, C. Bootstrapping markov chains: countable case. *Journal of Statistical Planning and Inference 33*, 3 (1992), 311–331.

[7] Athreya, K. B., and Ney, P. A new approach to the limit theory of recurrent markov chains. *Transactions of the American Mathematical Society 245* (1978), 493–501.

[8] Athreya, K. B., and Roy, V. Estimation of integrals with respect to infinite measures using regenerative sequences. *Journal of Applied Probability 52*, 4 (2015), 1133 – 1145.

[9] Athreya, K. B., and Roy, V. General glivenko–cantelli theorems. *Stat 5*, 1 (2016), 306–311.

[10] Barlow, R. E., Bartholomew, D. J., Bremner, J., and Brunk, H. D. *Statistical inference under order restrictions: The theory and application of isotonic regression.* Wiley New York, 1972.

[11] Basharin, G. P., Langville, A. N., and Naumov, V. A. The life and work of a.a. markov. *Linear Algebra and its Applications 386* (2004), 3–26. Special Issue on the Conference on the Numerical Solution of Markov Chains 2003.

[12] Berkes, I., Horváth, L., and Schauer, J. Non-central limit theorems for random selections. *Probability theory and related fields 147*, 3-4 (2010), 449–479.

[13] Bernstein, S. Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. *Mathematische Annalen 97* (1927), 1–59.

[14] Bertail, P., and Clémençon, S. Edgeworth expansions of suitably normalized sample mean statistics for atomic markov chains. *Probability Theory and Related Fields 130*, 3 (2004), 388–414.

[15] Bertail, P., and Clémençon, S. Regenerative block bootstrap for markov chains. *Bernoulli 12*, 4 (2006), 689–712.

[16] Bertail, P., and Clémençon, S. Second-order properties of regeneration-based bootstrap for atomic markov chains. *TEST 16*, 1 (2007), 109–122.

[17] Bertail, P., Clémençon, S., and Fernández, C. Tail index estimation for discrete heavy-tailed distributions with application to statistical inference for regular markov chains. *preprint* (2023).

[18] Bertail, P., and Portier, F. Rademacher complexity for Markov chains: Applications to kernel smoothing and Metropolis–Hastings. *Bernoulli 25*, 4B (Nov. 2019), 3912–3938.

[19] Bickel, P. J., and Freedman, D. A. Some asymptotic theory for the bootstrap. *The Annals of Statistics 9* (11 1981).

[20] Billingsley, P. *Convergence of probability measures*. Wiley, 1968.

[21] Bingham, N. H., Goldie, C. M., and Teugels, J. L. *Regular variation*. Encyclopedia of mathematics and its applications 27. Cambridge University Press, 1987.

[22] Blackwell, D. The existence of anormal chains. *Bulletin of the American Mathematical Society 51*, 6.P1 (1945), 465 – 468.

[23] Cai, B., and Tjøstheim, D. Nonparametric regression estimation for multivariate null recurrent processes. *Econometrics 3*, 2 (2015), 265–288.

[24] Cai, Z., and Ould-Saiid, E. Local m-estimator for nonparametric time series. *Statistics and Probability Letters 65*, 4 (2003), 433–449.

[25] Carpentier, A., and Kim, A. K. H. Adaptive and minimax optimal estimation of the tail coefficient. *Statistica Sinica 25*, 3 (2015), 1133–1144.

[26] Cavaliere, G., and Georgiev, I. Inference under random limit bootstrap measures. *Econometrica 88*, 6 (2020), 2547–2574.

[27] Cavaliere, G., Politis, D. N., and Rahbek, A. Recent developments in bootstrap methods for dependent data. *Journal of Time Series Analysis 2015-mar 16 vol. 36 iss. 3 36* (mar 2015).

[28] Chen, X. How often does a harris recurrent markov chain recur? *The Annals of Probability 27* (07 1999).

[29] Chen, X. *Limit Theorems for Functionals of Ergodic Markov Chains With General State Space*. Memoirs AMS 664. Amer Mathematical Society, 1999.

[30] Chen, X. On the limit laws of the second order for additive functionals of harris recurrent markov chains. *Probability Theory and Related Fields 116* (01 2000).

[31] Clauset, A., Shalizia, C. R., and Newman, M. E. J. Power-law distributions in empirical data. *SIAM Review 51* (11 2009).

[32] Csáki, E., and Csörgő, M. On additive functionals of markov chains. *Journal of Theoretical Probability 8*, 4 (1995), 905–919.

[33] Danielsson, J., de Haan, L., Peng, L., and de Vries, C. Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis 76* (2001).

[34] Datta, S., and McCormick, W. P. Regeneration-based bootstrap for markov chains. *Canadian Journal of Statistics 21* (06 1993).

[35] de Acosta, A. Large deviations for vector-valued functionals of a markov chain: Lower bounds. *The Annals of Probability 16*, 3 (1988), 925–960.

[36] De Coninck, J., Dunlop, F., and Huillet, T. Random walk versus random line. *Physica A: Statistical Mechanics and its Applications 388*, 19 (2009), 4034–4040.

[37] Doeblin, W. Éléments d'une théorie générale des chaînes simples constantes de markoff. *Annales scientifiques de l'École Normale Supérieure 57* (1940), 61–111.

[38] Douc, R., Moulines, E., Priouret, P., and Soulier, P. *Markov chains*. Springer Series in Operations Research and Financial Engineering. Springer, 2018.

[39] Drees, H., and Rootzén, H. Limit theorems for empirical processes of cluster functionals. *The Annals of Statistics 38*, 4 (2010), 2145 – 2186.

[40] Duflo, M. *Random Iterative Models*, 1 ed. Stochastic Modelling and Applied Probability 34. Springer-Verlag Berlin Heidelberg, 1997.

[41] Durot, C., Bertail, P., and Fernández, C. Harris recurrent markov chains and nonlinear monotone cointegrated models. *preprint* (2023).

[42] Durot, C., and Tocquet, A.-S. On the distance between the empirical process and its concave majorant in a monotone regression framework. In *Annales de l'Institut Henri Poincare (B) Probability and Statistics* (2003), vol. 39, Elsevier, pp. 217–240.

[43] Efron, B. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics 7*, 1 (1979), 1 – 26.

[44] Engle, R. F., and Granger, C. W. J. Co-integration and error correction: Representation, estimation, and testing. *Econometrica 55*, 2 (1987), 251–276.

[45] Feigin, P. D., and Tweedie, R. L. Random coefficient autoregressive processes:a markov chain analysis of stationarity and finiteness of moments. *Journal of Time Series Analysis 6* (jan 1985).

[46] Feller, W. *An Introduction to Probability Theory and Its Applications*, 2 ed., vol. 2. Wiley, 1971.

[47] Fernández, C. Regenerative bootstrap for β-null recurrent markov chains. *preprint* (2023).

[48] Franke, J., Kreiss, J.-P., and Mammen, E. Bootstrap of kernel smoothing in nonlinear time series. *Bernoulli 8*, 1 (2002), 1 – 37.

[49] Gao, J., Tjøstheim, D., and Yin, J. Estimation in threshold autoregressive models with a stationary and a unit root regime. *Journal of Econometrics 172*, 1 (2013), 1–13.

[50] Goldie, C., and Smith, R. Slow variation with remainder: theory and applications. *Quart. J. Math. Oxford 38*, 1 (1987), 45–71.

[51] Goldstein, M. L., Morris, S. A., and Yen, G. G. Problems with fitting to the power-law distribution. *The European Physical Journal B / Condensed Matter and Complex Systems 41* (09 2004).

[52] Granger, C. Some properties of time series data and their use in econometric model specification. *Journal of Econometrics 16*, 1 (1981), 121–130.

[53] Groeneboom, P., and Jongbloed, G. *Nonparametric Estimation under Shape Constraints: Estimators, Algorithms and Asymptotics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2014.

[54] Groeneboom, P., and Jongbloed, G. Some Developments in the Theory of Shape Constrained Inference. *Statistical Science 33*, 4 (2018), 473 – 492.

[55] Gut, A. *Probability : a graduate course*, 2nd ed ed. Springer texts in statistics. Springer, 2013.

[56] Hill, B. M. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics 3* (09 1975).

[57] Horowitz, J. L. Bootstrap methods for markov processes. *Econometrica 71*, 4 (2003), 1049–1082.

[58] Horowitz, J. L. Bootstrap methods in econometrics. *Annual Review of Economics 11*, 1 (2019), 193–224.

[59] Horowitz, J. L., and Lee, S. Nonparametric estimation and inference under shape restrictions. *Journal of Econometrics 201*, 1 (2017), 108–126.

[60] Jacod, J., and Shiryaev, A. N. *Limit Theorems for Stochastic Processes*, 2 ed. Grundlehren der mathematischen Wissenschaften 288. Springer-Verlag Berlin Heidelberg, 2003.

[61] Jain, N., and Jamison, B. Contributions to doeblin's theory of markov processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 8*, 1 (1967), 19–40.

[62] Johansen, S. Statistical analysis of cointegrating vectors. *Journal of Economic Dynamics and Control 12* (1988), 231.

[63] Johansen, S. Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica 59* (1991), 1551.

[64] Kallianpur, G., and Robbins, H. The sequence of sums of independent random variables. *Duke Mathematical Journal 21* (jun 1954).

[65] Karlsen, H. A., Myklebust, T., and Tjøstheim, D. Nonparametric estimation in a nonlinear cointegration type model. *The Annals of Statistics 35*, 1 (2007), 252–299.

[66] Karlsen, H. A., Myklebust, T., and Tjøstheim, D. Nonparametric regression estimation in a null recurrent time series. *Journal of Statistical Planning and Inference 140* (2010).

[67] Karlsen, H. A., and Tjostheim, D. Nonparametric estimation in null recurrent time series. *The Annals of Statistics 29* (04 2001).

[68] Kasahara, Y. Limit theorems for lévy processes and poisson point processes and their applications to brownian excursions. *J. Math. Kyoto Univ. 24*, 3 (1984), 521–538.

[69] Knight, K. On the Bootstrap of the Sample Mean in the Infinite Variance Case. *The Annals of Statistics 17*, 3 (1989), 1168 – 1175.

[70] Kosorok, M. R. *Introduction to empirical processes and semiparametric inference*, 1 ed. Springer Series in Statistics. Springer, 2008.

[71] Kreiss, J.-P., and Lahiri, S. N. 1 - bootstrap methods for time series. In *Time Series Analysis: Methods and Applications*, vol. 30 of *Handbook of Statistics*. Elsevier, 2012, pp. 3–26.

[72] Kreiss, J.-P., and Paparoditis, E. Autoregressive-aided periodogram bootstrap for time series. *The Annals of Statistics 31*, 6 (2003), 1923–1955.

[73] Kreiss, J.-P., and Paparoditis, E. Bootstrap methods for dependent data: A review. *Journal of the Korean Statistical Society 40*, 4 (2011), 357–378.

[74] Kreiss, J.-P., Paparoditis, E., and Politis, D. N. On the range of validity of the autoregressive sieve bootstrap. *The Annals of Statistics 39*, 4 (2011), 2103 – 2130.

[75] Kulperger, R. J., and Rao, B. L. S. P. Bootstrapping a finite state markov chain. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002) 51*, 2 (1989), 178–191.

[76] Kunsch, H. R. The Jackknife and the Bootstrap for General Stationary Observations. *The Annals of Statistics 17*, 3 (1989), 1217 – 1241.

[77] Lahiri, S. *Resampling methods for dependent data*. Springer Science & Business Media, 2003.

[78] Lazzardi, S., Valle, F., Mazzolini, A., Scialdone, A., Caselle, M., and Osella, M. Emergent statistical laws in single-cell transcriptomic data. *bioRxiv* (2021).

[79] LePage, R., Woodroofe, M., and Zinn, J. Convergence to a Stable Distribution Via Order Statistics. *The Annals of Probability 9*, 4 (1981), 624 – 632.

[80] Lin, Z., Li, D., and Chen, J. Local linear m-estimators in null recurrent times series. *Statistica Sinica 19*, 4 (2009), 1683–1703.

[81] Luss, R., Rosset, S., and Shahar, M. Efficient regularized isotonic regression with application to gene–gene interaction search. *The Annals of Applied Statistics 6*, 1 (2012), 253–283.

[82] MacKinnon, J. G. Bootstrap methods in econometrics*. *Economic Record 82*, s1 (2006), S2–S18.

[83] Manning, C., and Scütze, H. *Foundations of Statistical Natural Language Processing*. The MIT Press, 1999.

[84] Markov, A. A. Extension of the limit theorems of probability theory to a sum of variables connected in a chain, the notes of the imperial academy of sciences of st. *Petersburg VIII Series, Physio-Mathematical College 22*, 9 (1907).

[85] Matsui, M., Mikosch, T., and Tafakori, L. Estimation of the tail index for lattice-valued sequences. *Extremes 16* (feb 2013).

[86] Mehrjoo, M., Jafari Jozani, M., and Pawlak, M. Wind turbine power curve modeling for reliable power prediction using monotonic regression. *Renewable Energy 147* (2020), 214–222.

[87] Meyn, S., Tweedie, R., and Glynn, P. *Markov chains and stochastic stability*, 2 ed. Cambridge Mathematical Library. Cambridge University Press, 2009.

[88] Myklebust, T., Karlsen, H. A., and Tjøstheim, D. Null recurrent unit root process. *Econometric Theory 28* (2 2012).

[89] Nummelin, E. A splitting technique for harris recurrent markov chains. *Probability Theory and Related Fields 43* (1978).

[90] Nummelin, E. *General Irreducible Markov Chains and Non-Negative Operators*. Cambridge Tracts in Mathematics 83. Cambridge University Press, 1984.

[91] Obozinski, G., Lanckriet, G., Grant, C., Jordan, M. I., and Noble, W. S. Consistent probabilistic outputs for protein function prediction. *Genome Biology 9*, 1 (2008), 1–19.

[92] Paparoditis, E. *Frequency Domain Bootstrap for Time Series*. Birkhäuser Boston, Boston, MA, 2002, pp. 365–381.

[93] Paparoditis, E., and Politis, D. N. *The Continuous-Path Block-Bootstrap*. De Gruyter, Berlin, Boston, 2000, pp. 305–320.

[94] Paparoditis, E., and Politis, D. N. A markovian local resampling scheme for nonparametric estimators in time series analysis. *Econometric Theory 17*, 3 (2001), 540–566.

[95] Paparoditis, E., and Politis, D. N. Tapered block bootstrap. *Biometrika 88*, 4 (2001), 1105–1119.

[96] Paparoditis, E., and Politis, D. N. Local block bootstrap. *Comptes Rendus Mathematique 335*, 11 (2002), 959–962.

[97] Paparoditis, E., and Politis, D. N. The local bootstrap for markov processes. *Journal of Statistical Planning and Inference 108*, 1 (2002), 301–328. C.R. Rao 80th Birthday Felicitation Volume, Part II.

[98] Phillips, P. C. B. Optimal inference in cointegrated systems. *Econometrica 59* (1991), 283.

[99] Phillips, P. C. B., and Solo, V. Asymptotics for linear processes. *The Annals of Statistics 20* (1992), 971.

[100] Pickands, J. Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics 3*, 1 (1975), 119 – 131.

[101] Politis, D. N., and Romano, J. P. *A circular block-resampling procedure for stationary data.* Purdue University. Department of Statistics, 1991.

[102] Politis, D. N., and Romano, J. P. The stationary bootstrap. *Journal of the American Statistical Association 89*, 428 (1994), 1303–1313.

[103] Pollard, D. *Convergence of stochastic processes.* Springer, 1984.

[104] Qi, Y. Bootstrap and empirical likelihood methods in extremes. *Extremes 11* (2008), 81–97.

[105] Rajarshi, M. B. Bootstrap in markov-sequences based on estimates of transition density. *Annals of the Institute of Statistical Mathematics 42*, 2 (1990), 253–268.

[106] Resnick, S. *Heavy-tail phenomena: probabilistic and statistical modeling.* Springer Science & Business Media, 2007.

[107] Revesz, P. *Random Walk in Random and Non-Random Environments*, 2nd ed ed. World Scientific, 2005.

[108] Revuz, D. *Markov Chains*, 2nd ed., vol. 11 of *North Holland Mathematical Library.* Elsevier, 1984.

[109] Samworth, R. J., and Sen, B. Editorial: Special Issue on "Nonparametric Inference Under Shape Constraints". *Statistical Science 33*, 4 (2018), 469 – 472.

[110] Schell, M. J., and Singh, B. The reduced monotonic regression method. *Journal of the American Statistical Association 92*, 437 (1997), 128–135.

[111] Tang, X.-S., Huang, H.-B., Liu, X.-F., Li, D.-Q., and Liu, Y. Efficient bayesian method for characterizing multiple soil parameters using parametric bootstrap. *Computers and Geotechnics 156* (2023), 105296.

[112] Tjøstheim, D. Some notes on nonlinear cointegration: A partial review with some novel perspectives. *Econometric Reviews 39*, 7 (2020), 655–673.

[113] Touati, A. Loi Fonctionnelle du Logarithme Itere Pour les Processus de Markov Recurrents. *The Annals of Probability 18*, 1 (1990), 140 – 159.

[114] Van der Vaart, A. W. *Asymptotic statistics*. Cambridge university press, 2000.

[115] van der Vaart, A. W., and Wellner, J. A. *Weak Convergence and Empirical Processes*. Springer, 1996.

[116] Voitalov, I., van der Hoorn, P., van der Hofstad, R., and Krioukov, D. Scale-free networks well done. *Physical Review Research 1*, 3 (2019), 033034.

[117] Wang, Q. *Limit Theorems for Nonlinear Cointegrating Regression*. World Scientific, 2015.

[118] Wang, Q., and Phillips, P. C. B. Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory 25*, 3 (2009), 710–738.

[119] Wang, Q., and Phillips, P. C. B. Structural nonparametric cointegrating regression. *Econometrica 77*, 6 (2009), 1901–1948.

[120] Yao, D., van der Hoorn, P., and Litvak, N. Average nearest neighbor degrees in scale-free networks. internet math. 2018, 1–38 (2018), 2018.

[121] Zanette, D. Zipf's law and the creation of musical context. *Musicae Scientiae 10*, 1 (2006), 3–18.

[122] Zoubir, A., and Boashash, B. The bootstrap and its application in signal processing. *IEEE Signal Processing Magazine 15*, 1 (1998), 56–76.

[123] Zoubir, A. M., and Iskander, D. R. *Bootstrap Techniques for Signal Processing*. Cambridge University Press, 2004.