

**AN INVESTIGATION OF THE STATISTICAL PROPERTIES OF
CERTAIN CHAOTIC DYNAMICAL SYSTEMS THROUGH
EXTREMES AND RECURRENCE: A THEORETICAL AND
APPLIED APPROACH**

A Dissertation Presented to
the Faculty of the Department of Mathematics
University of Houston

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

By
Meagan Elizabeth Carney

May 2019

**AN INVESTIGATION OF THE STATISTICAL PROPERTIES OF
CERTAIN CHAOTIC DYNAMICAL SYSTEMS THROUGH
EXTREMES AND RECURRENCE: A THEORETICAL AND
APPLIED APPROACH**

Meagan Elizabeth Carney

APPROVED:

Dr. Matthew Nicol, Advisor
Department of Mathematics, University of Houston

Dr. Robert Azencott
Department of Mathematics, University of Houston

Dr. Vaughn Climenhaga
Department of Mathematics, University of Houston

Dr. Gemunu Gunaratne
Department of Physics, University of Houston

Dean, College of Natural Sciences and Mathematics

Acknowledgements

To my daughter, Elsa Lane Carney

My sincerest appreciation goes to my thesis advisor, Prof. Matthew Nicol, for guiding me through all of the most challenging moments in my graduate career with patience, understanding, and integrity. The immense support he has given to me and my family has allowed me to achieve goals which seemed quite improbable at times and advance my career in ways I could not have imagined. Without him this thesis would have, undoubtedly, not been possible. Thank you.

My special thanks goes to Dr. Robert Azencott for his invaluable insight on the applied section of this thesis and for showing me a way to approach future interesting problems. I am also very grateful for the encouraging discussions he had with me about my professional goals which helped shape my future in a most positive way.

My thanks goes out to Dr. Hong-Kun Zhang who originally encouraged me to join the program and has provided valuable knowledge and work on the billiards section of this thesis. I would like to thank Dr. Vaughn Climenhaga for taking valuable time to explain many dynamical concepts to me throughout my entire graduate career and for reading and providing feedback on this thesis. I would also like to thank Dr. Gemunu Gunaratne for taking valuable time to read and provide feedback on this thesis.

Finally, I would like to thank my husband who has been present for the amount of hours this has taken and has provided encouragement and equal support of our daughter throughout. I would also like to thank my father for his words of encouragement.

AN INVESTIGATION OF THE STATISTICAL PROPERTIES OF
CERTAIN CHAOTIC DYNAMICAL SYSTEMS THROUGH
EXTREMES AND RECURRENCE: A THEORETICAL AND
APPLIED APPROACH

An Abstract of a Dissertation
Presented to
the Faculty of the Department of Mathematics
University of Houston

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

By
Meagan Elizabeth Carney

May 2019

Abstract

Motivated by proofs in extreme value theory, we investigate the statistical properties of certain chaotic dynamical systems, including the well-known dispersing billiard model. In particular, we prove the existence of a maximal probability distribution and rare event point process in the setting of two-dimensional hyperbolic systems with singularities. We also obtain bounds on the growth rates of Birkhoff sums with non-integrable observables, where the Birkhoff ergodic theorem fails, by using the recurrence properties of the system to a point of maximization. We end with an analysis of extreme temperatures across Texas where we find compelling evidence that the probability of observing higher summer temperature extremes has increased.

Contents

1	Introduction	1
1.1	Ergodic Theory	3
1.2	Borel-Cantelli Lemmas	11
1.3	Extreme Value Theory	13
1.3.1	Extreme value theory for i.i.d. processes	14
1.3.2	Extreme value theory for dependent processes	18
1.4	Rare Event Point Processes	27
1.5	Statistical Methods for Modeling Extremes	31
1.5.1	Generalized Extreme Value distribution approximation and block maxima approach	32
1.5.2	Generalized Pareto distribution and peaks over threshold approach	36
1.6	Statistical Methods for Clustering Large Data	39
1.6.1	Mutual information and data compression	40
1.6.2	Spectral clustering methods	44
2	Growth Rates for Birkhoff Sums of Non-integrable Observables¹	50
2.1	Background	50
2.2	Main Results	53
2.2.1	Non-integrable observations.	54

2.2.2	Non-integrable observables on a class of intermittent type maps.	59
2.2.3	Dynamical systems with L^p , ($p > 1$), densities and exponential decay of correlations.	65
2.3	Numerical Results	71
3	REPP Distribution for 2-D Hyperbolic Systems with Singularities²	75
3.1	Background	75
3.2	Main Assumptions	80
3.3	Main Results	85
4	Modeling Real-World Temperature Extremes	94
4.1	Preliminary results on Houston, Texas, summer temperature extremes	97
4.2	Analysis of summer temperature extremes throughout the Texas region	100
4.2.1	Data	102
4.2.2	Results	105
4.2.3	Conclusions	120
4.3	A brief discussion of winter temperature extremes throughout the Texas region	124
4.3.1	Data, Clusters and Preliminaries	124
4.3.2	Generalized Extreme Value Distribution	124
4.3.3	Generalized Pareto Distribution	128
4.3.4	Conclusions	129
4.4	Concluding Remarks	132
A	Appendix	134
A.1	Supporting Dynamical Definitions and Proofs	134
A.1.1	Descriptions of relevant dynamical maps	134
A.1.2	Dynamical Definitions	137
A.1.3	Proof of $\mathcal{D}_q(u_n)$ for 2-D hyperbolic maps with singularities . .	139

A.2	Supporting Statistical Information	141
A.2.1	Relevant derivations	141
A.2.2	Description of statistical tests	142
	Bibliography	145

List of Figures

1.1	(a) Example orbit of an irrational rotation ($\alpha = 1/\pi$) and (b) a rational rotation ($\alpha = 1/5$).	8
1.2	Generalized extreme value probability distribution functions (PDF) for three different values of the shape parameter corresponding to the three different types with $\mu = 0$ and $\sigma^2 = 1$	17
1.3	The doubling map with a nonperiodic orbit of a point (black) and a periodic orbit of a point (red).	24
1.4	Undirected graph for five random variables where a line represents a strictly positive similarity (a) before clustering red lines indicate which edges will be removed and after clustering into two groups. (b) The corresponding random walk for the graph different colors indicate clusters black lines indicate removed edges.	44
1.5	(a) K -means clustering result. (b) Spectral clustering result. (c) Space where K -means clustering is performed in the spectral clustering algorithm.	49
2.1	Iterations of shrinking balls (<i>shrinking targets</i>) of (a) radius $\frac{1}{\sqrt{n}}$ and (b) radius $\frac{1}{n}$. (a) The balls about point q shrink slow enough so that almost every point under the map will return to the shrinking ball infinitely often. (b) The balls shrink too quickly so that at a large enough time step almost every under the map will not return to the shrinking ball.	57
2.2	The Liverani-Saussol-Vaianti (LSV) map for (a) $\alpha = 0.2$ and (b) $\alpha = 0.9$	59

2.3	Shrinking target about $q = 0$ for the first branch of the LSV map with (a) $\alpha = 0.2$ and (b) $\alpha = 0.9$. Points indicate iterations of a point under the map. Note that it takes longer for a point to escape the shrinking target for larger α values.	62
2.4	(a) Shrinking balls for two nested sets of radius $\frac{1}{n^{\eta_1}}$ (blue) and $\frac{1}{n \log^{1+\delta} n}$ (green). Orbits of points may enter the blue sequence infinitely many times while only enter the green sequence finitely many times. Red dots indicate the maximum of the orbit after the point leaves green but continues to remain in blue. (b) Pink indicates a shrinking set of radius $\frac{1}{n^{\eta_2}}$ where maximum values of the orbit after the point leaves blue but remains in pink are marked by red dots.	65
2.5	Numerical results for the Birikhoff sum with fixed point $q = 0.25$ and $q = 0$ on (a) the doubling map and (b) an Anosov diffeomorphism. Predicted value of the convergence is marked in a dotted line.	72
2.6	Numerical results for the Birikhoff sum with fixed point $q = 0.25$ and $q = 0$ on (a) the tent map and (b) the unimodal map. Predicted value of the convergence is marked with a dotted line.	73
2.7	Numerical results for the Birikhoff sum with fixed point $q = 0.25$ and $q = 0$ on the LSV map. (a) $\alpha = 0.5$ and (c) $\alpha = 0.9$ with initial value near zero and (b) $\alpha = 0.5$ and (d) $\alpha = 0.9$ with initial value near 1. Predicted value of the convergence is marked with a dotted line.	74
3.1	Billiard flow with infinite horizon on the torus with circular barriers. Collisions are represented by points.	76
3.2	(a) Billiard flow on the torus with circular barriers. Collisions are represented by points. (b) Illustration of the reduction to the billiard map for a single collision point.	77
3.3	(a) Trace of the convex front on M showing an increase in the angle of collision and (b) the corresponding billiard map for (r, θ) depicting the expansion of the unstable curve.	79
3.4	The sets $A_n^q := \{X_0 > u_n, X_q < u_n\}$ in (a) the usual Euclidean metric and (b) the adapted metric.	86
3.5	Expansion of the set A_n^q out of Ω_n under the map T	88

3.6	(a) Expansion of the set A_n^q under the map T before it hits a singularity point. (b) Fragmentation of the unstable direction of A_n^q under the map T after it hits a singularity point. This represents the "worst-case scenario" where estimates are needed to ensure the fragmented pieces do not remain in the intersection for "too long".	90
3.7	A periodic point p for the billiard flow on the torus with circular barriers.	92
4.1	Generalized extreme value distribution for the Houston station for 10-day maximum temperatures for 1941-2017 with MLE parameters (a) cumulative density function, (b) probability density function, and (c) qq plot. Anderson-Darling p -value= $4.9e - 3$	98
4.2	Generalized extreme value distribution for the Houston station with MLE parameters for 10-day maximum temperatures for 1941-1981 (a) cumulative density function, (b) probability density function, and (c) qq plot. Anderson-Darling p -value= 0.25.	100
4.3	Generalized extreme value distribution for the Houston station with MLE parameters for 10-day maximum temperatures for 1982-2017 (a) cumulative density function, (b) probability density function, and (c) qq plot. Anderson-Darling p -value= $4.9e - 4$	101
4.4	Map of all listed stations.	104
4.5	Calculated maximized mutual information between stations.	106
4.6	Calculated error on mutual information between stations after maximization.	107
4.7	Graph of nodes and edges where nodes represent stations and an edge represents a strictly positive mutual information. (a) The graph of full dimension where edges to be removed after spectral clustering are highlighted in red. (b) The reduced system after spectral clustering.	108
4.8	Ratio of eigenvalues and chosen cutoff.	108
4.9	Minimized cost for different values of K	109
4.10	Minimized distances for each iteration of K -means $K = 4$	110
4.11	The $J = 3$ dimensional projection of the nodes into the space of spectral clustering. Colors represent different clusters.	110

4.12	Similarity matrix for stations after clustering. Lighter colors indicate higher similarity.	111
4.13	Regional map with clustered stations where symbols represent clusters.	112
4.14	Chi-square independence p -values for (a) summer and (b) winter time series. The Houston station p -values referenced in the preliminaries is included.	114
4.15	histogram of Anderson-Darling p -values obtained by varying the estimated parameters within the 95% confidence interval of the MLE parameters.	115
4.16	Non-stationary generalized extreme value pdfs.	123
4.17	Chi-square independence p -values for winter time series.	125
4.18	QQ plots for the true block maximum and the maximum likelihood estimated generalized extreme value distribution for winter temperature extremes for 1941-2017.	126
4.19	Mean residual life plots for each cluster. Chosen threshold is marked with a red dashed line.	130
4.20	Generalized extreme value distribution for the winter with MLE parameters for 10-day maximum temperatures for 1941-2017 probability density function.	131

List of Tables

4.1	Table of Mann-Kendall p -values for the 3-year mean and variance of the Houston station.	102
4.2	Table of all stations used in this analysis with location and type. . . .	105
4.3	Table of minimum Davies-Bouldin index clusters.	111
4.4	Table of Anderson-Darling p -values associated to stationary GEV model of MLE parameters.	113
4.5	Table of mean change points for the yearly mean and variance of each cluster.	116
4.6	Mann-Kendall, Theil-Sen and Linear Regression results for yearly mean of the cluster.	117
4.7	Mann-Kendall, Theil-Sen and Linear Regression results for yearly variance of the cluster.	118
4.8	Anderson-Darling goodness of fit p -values before and after non-stationary fitting of the location and scale parameters.	120
4.9	Location and scale comparisons for 1941 and 2017 GEV models. . . .	122
4.10	Probability distribution comparisons for 1941 and 2017 GEV models.	122
4.11	Ratios of probability comparisons for 1941 and 2017 GEV models. . .	122
4.12	Table of minimum Davies-Bouldin index clusters.	124
4.13	Table of Anderson-Darling p -values associated to stationary GEV model of MLE parameters.	125
4.14	A-D goodness of fit p values before and after non-stationary modeling.	127

4.15	Log likelihood p values for comparing stationary and non-stationary models.	127
4.16	Mann-Kendall and Theil-Sen results for yearly mean and variance of the cluster.	128
4.17	Table of Anderson-Darling p -values associated to stationary GPD model of MLE parameters.	129

Chapter 1

Introduction

The term *dynamical system* is used to describe a process that changes over time. Dynamics play a critical role in understanding the systems that govern our everyday lives: from stock market exchanges in viewing price fluctuations; to weather patterns in observing changes in temperature or precipitation; to biological agent interactions, such as the movement of cells in a petridish or the flocking of birds.

In theory, a dynamical system is often viewed as a map (in the discrete-time case) or a flow (in the continuous-time case) on a space. While in simple systems knowing the trajectory of a single point under iterations of the map is sufficient for modeling and prediction, more complex dynamical systems often exhibit chaotic behavior, where the future is too sensitive to initial conditions to be understood in this way. As a result, many physically-relevant problems in dynamics are reduced to studying the long-term behavior of chaotic systems.

This observed randomness provides motivation for studying the statistical and asymptotic properties of our system rather than viewing the trajectories individually. We may investigate whether classical results from probability theory, such as, central limit theorems, strong laws of large numbers, recurrence properties (such as Borel-Cantelli lemmas), extreme value theory, and the like, enables us to make predictions about the system's behavior.

Mathematically, a dynamical system is defined by the pair (X, T) where T is a transformation on a state space X , and for some point $x \in X$, the sequence $x, Tx, T^2x, \dots, T^n x$ is the time-evolution of the point x .

Definition 1.0.1. A family \mathcal{B} of subsets of X is called a σ -**algebra** if the following are satisfied:

1. $X \in \mathcal{B}$;
2. for any $A \in \mathcal{B}, A^c \in \mathcal{B}$, where $A^c = \{x \in X : x \notin A\}$;
3. if $A_n \in \mathcal{B}$, for $n = 1, 2, \dots$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{B}$.

Definition 1.0.2. If X is a set and $\mathcal{B} \subset P(X)$ is a σ -algebra, (X, \mathcal{B}) is called a **measurable space** and the sets in \mathcal{B} are called **measurable sets**. $P(X)$ is called the **power set** of X and is the family of all subsets of a set X .

Definition 1.0.3. A set function $\mu: \mathcal{B} \rightarrow [0, \infty)$ is called a **measure** on \mathcal{B} if it satisfies:

1. $\mu(\emptyset) = 0$;

2. for any sequence $\{A_n\}$ of disjoint measurable sets, $A_n \in \mathcal{B}$, $n = 1, 2, \dots$,

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$$

(X, \mathcal{B}, μ) is called a measure space and if $\mu(X) = 1$, it is called a normalized measure space or *probability* space. If X can be written as a countable union of measurable sets with finite measure then μ is a σ -finite measure and (X, \mathcal{B}, μ) is called a σ -finite measure space.

The transformation T is called a measure-preserving transformation, or interchangeably (X, \mathcal{B}, T, μ) is a measure-preserving dynamical system, provided $\forall A \in \mathcal{B}$, $\mu(T^{-1}(A)) = \mu(A)$. Within this framework, μ is referred to as a T -invariant, or invariant measure.

Definition 1.0.4. If μ and ν are two measures on a σ -algebra \mathcal{B} of subsets of X , we say that ν is **absolutely continuous** with respect to μ if $\nu(A) = 0$ for any $A \in \mathcal{B}$ such that $\mu(A) = 0$. This is commonly denoted by $\nu \ll \mu$.

From now on we will restrict ourselves to measure-preserving transformations defined on probability spaces.

1.1 Ergodic Theory

In physics and thermodynamics, the ergodic theorem originates from the work of Boltzmann and states that, over long periods of time, the time spent by a system in some region of the phase space is proportional to the volume of this region. It is

then understandable that ergodic theory is a branch of mathematics concerned with the asymptotic behavior of dynamical systems with invariant measures.

If we consider a dynamical system (X, \mathcal{B}, T, μ) equipped with $T : X \rightarrow X$ a measure-preserving transformation then $T^n(x) = T \circ T \cdots \circ T(x)$ represents the trajectory of a point $x \in X$. In a probabilistic sense, \mathcal{B} represents the set of observable events and μ their probability of occurrence. Taking $n \rightarrow \infty$ allows us to consider the long-term behavior of the system, including convergence and recurrence properties. One important recurrence theorem comes from Poincaré,

Theorem 1.1.1. (*Poincaré Recurrence Theorem*) [83, Theorem 8.1]. *Let T be a measure-preserving transformation on a probability space (X, \mathcal{B}, μ) . Let $E \in \mathcal{B}$ such that $\mu(E) > 0$ then almost all points of E return infinitely often to E under iterations of T .*

Before introducing stronger versions of recurrence and other statistical properties, we need to discuss ergodicity in terms of dynamics.

Definition 1.1.2. A measure-preserving transformation T on a dynamical system (X, \mathcal{B}, T, μ) is called **ergodic** if for any $A \in \mathcal{B}$, such that $T^{-1}A = A$, then $\mu(A) = 0$ or $\mu(A^c) = 0$.

Under the ergodic assumption a stronger version of the Poincaré Recurrence Theorem holds. Let A be a measurable set with $\mu(A) > 0$ and define, for $x \in A$, the first hitting time of x to a set A .

$$n_A(x) = \min\{k \geq 1 : T^k(x) \in A\}$$

Theorem 1.1.3. (*Kac's Lemma*) [8, Theorem 3.2.4]. Let T be an ergodic measure-preserving transformation of the probability space (X, \mathcal{B}, μ) . Let $A \in \mathcal{B}$ be such that $\mu(A) > 0$. Then

$$\int_A n_A(x) d\mu(x) = 1. \quad (1)$$

If we define $\mu_A(B) = \frac{\mu(A \cap B)}{\mu(A)}$ as the conditional measure then we can rewrite (1) as,

$$\int_A n_A(x) d\mu_A(x) = \frac{1}{\mu(A)}.$$

Thus, Kac's theorem tells us that, given a set A of measure $\mu(A)$, the orbit of almost every point of A returns eventually to A and the expected time of the first recurrence is $\frac{1}{\mu(A)}$.

In dynamics it is common to consider an observable taken on the map T , which represents a physical quantity that can be measured. The time-series of a measurable function $\varphi : X \rightarrow \mathbb{R}$ given by $\varphi(x), \varphi(Tx), \varphi(T^2x), \dots$ may be thought of as the time-evolution of some physically interesting variable within our dynamical system. Moreover, if the map T taken on the measure space (X, μ) is such that μ is T -invariant, then the time-series is stationary. Since we are interested in the long-term statistical behavior of a system, it is beneficial to consider the long-term time average given by

$$\frac{1}{n} \sum_{i=0}^{n-1} \varphi \circ T^i(x).$$

Convergence in the case when $\varphi \circ T^i$ is independent and identically distributed (i.i.d) was proved by Borel (Borel's Strong Law of Large Numbers) [81]; however the independence assumption is too strong for most applications. Almost everywhere

(a.e.) convergence under more general dependence conditions is given by Birkhoff [81, Theorem 2.3]. We will consider a consequence of this theorem stated in the following corollary.

Corollary 1.1.4. [81, Theorem 4.4] *Let (X, \mathcal{B}, μ) be a probability space, $T : X \rightarrow X$ an ergodic measure-preserving transformation and $\varphi \in L^1(X, \mathcal{B}, \mu)$ then*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \varphi \circ T^i(x) = \int \varphi d\mu,$$

for μ -a.e. $x \in X$ where $\frac{1}{n} \sum_{i=0}^{n-1} \varphi \circ T^i(x)$ is often referred to as the time average and $\int \varphi d\mu$ as the space average.

In probability theory, if we define (X_i) to be an i.i.d. sequence of random variables, the property that $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} X_i = E(X_0)$ is called the **Strong Law of Large Numbers (SLLN)**. The SLLN holds provided $E(X_0)$ is finite and fails otherwise. If (X_i) are i.i.d and $E(X_0) = \infty$ then it has been shown [35] that for any sequence $b(n) > 0$, if $\lim_{n \rightarrow \infty} \frac{b(n)}{n} = \infty$ then either $\limsup \frac{S_n}{b(n)} = \infty$ a.e. or $\liminf \frac{S_n}{b(n)} = 0$ a.e. In fact, the conditions on $b(n)$ can be relaxed so that for any sequence of constants $b(n)$ either $\limsup \frac{S_n}{b(n)} = \infty$ a.e. or $\liminf \frac{S_n}{b(n)} = 0$ a.e. [25]. Hence, there is no strong law of large numbers for which the $E(X_0) = \infty$.

If we let (X, T, μ) be an ergodic, measure-preserving transformation and define $X_i = \varphi \circ T^i(x)$ then our dynamical system can be seen as a stochastic process. In this setting the Birkhoff ergodic theorem serves as a dynamical variant of the SLLN provided the observable satisfies $\int \varphi d\mu < \infty$. Note that $\int \varphi d\mu$ plays the role of $E(X_0)$ since in the dynamical setting $E(X_0) = E(\varphi \circ T^0(x)) = E(\varphi(x))$ by definition. If the observable is taken to be non-integrable, that is $\int \varphi d\mu = \infty$, a

failure of the Birkhoff ergodic theorem occurs. In fact, since $\varphi \circ T^i(x)$ is stationary by definition, Aaronson [1] showed that for any sequence $b(n) > 0$, if $\lim_{n \rightarrow \infty} \frac{b(n)}{n} = \infty$ then either $\limsup \frac{S_n}{b(n)} = \infty$ a.e. or $\liminf \frac{S_n}{b(n)} = 0$ a.e. Hence for ergodic dynamical systems there is no strong law of large numbers for non-integrable observables. When no convergence can be established, a natural question is at what rate the Birkhoff sum grows. We address this issue in detail in Chapter 2.

The property that $\frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} X_i \rightarrow \mathcal{N}(m, \sigma^2)$ (that is, converges in *distribution* to the Normal with mean m and variance σ^2) is called the **Central Limit Theorem**.

For the rest of this chapter we will only discuss processes for which the observable is integrable.

Definition 1.1.5. Let T be a measure-preserving transformation and (X, \mathcal{B}, μ) a probability space.

1. We say that T is weak-mixing if, for all $A, B \in \mathcal{B}$, we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\mu(T^{-i}A \cap B) - \mu(A)\mu(B)| = 0$$

2. We say that T is strong-mixing if, for all $A, B \in \mathcal{B}$, we have

$$\lim_{n \rightarrow \infty} \mu(T^{-n}A \cap B) = \mu(A)\mu(B)$$

That is, given two sets $A, B \in \mathcal{B}$, the sets $T^{-n}A$ and B become independent in some sense in the limit. This property and its relationship to ergodicity is well-defined. Strong-mixing implies weak-mixing and weak-mixing implies ergodicity [81, Proposition 5.2].

An example of a dynamical system which is ergodic but not mixing is the **irrational rotation** $R_\alpha = x + \alpha$. In this system α provides the angle of rotation around the circle created by "gluing" the end pieces of the $[0,1]$ interval together.

Dynamics of the Rotation Map

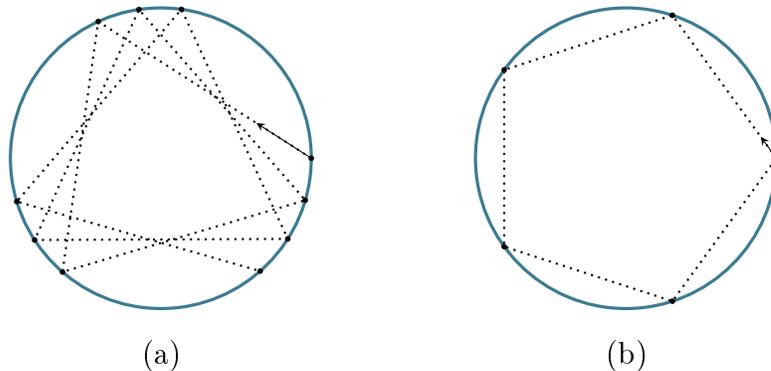


Figure 1.1: (a) Example orbit of an irrational rotation ($\alpha = 1/\pi$) and (b) a rational rotation ($\alpha = 1/5$).

Example 1.1.6. Let $A = \{e^{2\pi i x} : x \in [0, 1/4]\}$ since $\alpha \notin \mathbb{Q}$ there are infinitely many n such that $R_\alpha^{-n}(A) \cap A = \emptyset$ and thus $\lim_{n \rightarrow \infty} \mu(R_\alpha^{-n}(A) \cap A) = 0 \neq (1/4)^2$. On the other hand R_α is ergodic. To see this let $\varphi = \sum_{n \in \mathbb{Z}} a_n e^{2\pi i n x}$ then the invariance of μ implies $\mu(\varphi \circ R_\alpha(x)) = \mu(\varphi(x))$ so that $a_n e^{2\pi i n \alpha} = a_n \forall n$ and hence $\varphi(x)$ is constant on R_α .

Examples of dynamical maps that are weak-mixing but not strong-mixing are less obvious and left out of this discussion. An explanation for how to construct transformations with this property on the unit interval equipped with Lebesgue measure can be found in [14].

After convergence of the Birkhoff sum is established, a common question is to

ask *how fast* this convergence occurs. Answering this question requires an understanding of the strength of dependence in the system or how quickly the iterates $T^n x$ become independent in some sense. Mathematically, this is described by **decay of correlations**.

In general, **correlation** between integrable functions ψ_1 and ψ_2 is given by,

$$C_{\psi_1, \psi_2}(n) = \left| \int \psi_1 \cdot \psi_2 \circ T^n d\mu - \int \psi_1 d\mu \int \psi_2 d\mu \right|.$$

When $\psi_1 = \psi_2$ this is called **autocorrelation** and we write this as $C_\psi(n)$. Moreover, T is mixing if and only if correlations decay, that is, for every pair ψ_1 and $\psi_2 \in L^2$,

$$C_{\psi_1, \psi_2}(n) \rightarrow 0 \text{ as } n \rightarrow \infty$$

For simplicity let $S_n = \frac{1}{n} \sum_{i=0}^{n-1} \varphi \circ T^i(x)$ and $\mu_\varphi = \int \varphi d\mu$. Recall that under the Birkhoff ergodic theorem we have,

$$\frac{S_n - n\mu_\varphi}{n} \rightarrow 0, \quad \text{i.e.} \quad S_n = n\mu_\varphi + o(n)$$

In other words, $S_n - n\mu_\varphi$ is ultimately smaller than n . For *fast enough* decay of correlations the remainder term $o(n)$ is actually $o(\sqrt{n})$; consequently, the time average converges at a faster rate. The following is taken from [19] and describes under what conditions this is observed.

As a measure of the order of magnitude of the difference $S_n - n\mu_\varphi$ we may estimate the root-mean-square value $\sqrt{m([S_n - n\mu_\varphi]^2)}$ where,

$$m([S_n - n\mu_\varphi]^2) = nC_\varphi(0) + 2(n-1)C_\varphi(1) + 2(n-2)C_\varphi(2) + \cdots + 2C_\varphi(n-1).$$

Now suppose decay of correlations is fast enough so that (at least)

$$\sum_{n=0}^{\infty} |C_{\varphi}(n)| < \infty \quad (2)$$

Then the following sum is always positive:

$$\sigma^2 = \sum_{n=-\infty}^{\infty} C_{\varphi}(n) = C_{\varphi}(0) + 2 \sum_{n=1}^{\infty} C_{\varphi}(n).$$

and the mean square of the difference grows as

$$m([S_n - n\mu_{\varphi}]^2) = n\sigma^2 + o(n)$$

This means that on average the values $S_n - n\mu_{\varphi}$ grow as $\sigma\sqrt{n}$; they are of order \sqrt{n} .

Hence,

$$S_n = n\mu_{\varphi} + O(\sqrt{n})$$

Sometimes our correlation results can be improved to decay in normed spaces with stronger conditions on the observables [5, Definition 2.2]. Let B_1, B_2 denote Banach spaces of real valued measurable functions ψ_1 and ψ_2 defined on a probability space X equipped with invariant measure μ . Correlation in this setting is defined by,

$$C_{\psi_1, \psi_2}(n) = \frac{1}{\|\psi_1\|_{B_1} \|\psi_2\|_{B_2}} \left| \int \psi_1 \cdot \psi_2 \circ T^n d\mu - \int \psi_1 d\mu \int \psi_2 d\mu \right|.$$

We say that we have decay of correlations with respect to measure μ , for observables in B_1 *versus* observables in B_2 if, for every $\psi_1 \in B_1$ and every $\psi_2 \in B_2$ we have $C_{\psi_1, \psi_2}(n) \rightarrow 0$. For example, a common assumption occurs in the setting of decay of correlations for any real valued measurable functions $\psi_1 \in B_1$ a Banach space and $\psi_2 \in L^1(\mu)$. Decay of correlations versus $L^1(\mu)$ observables is a very strong property.

In fact, one can show that such a system has exponential decay of correlations of Hölder observables regardless of the rate of decay in $L^1(\mu)$ provided the correlation given in (2) is summable [4].

1.2 Borel-Cantelli Lemmas

In order to study the limiting behavior of a system we need to know something about recurrence in the system. In classical probability theory the question of whether an event will occur with probability one or probability zero in the limit is addressed by the Borel-Cantelli lemmas.

For a probability space $(X, \mathcal{B}, \mathbb{P})$ the classical Borel-Cantelli lemmas [34] state,

Classical Borel-Cantelli Lemmas

1. **(BC1)**: If $(A_n)_{n=0}^\infty$ is a sequence of measurable sets in X and $\sum_{n=0}^\infty \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(x \in A_n \text{ i.o.}) = 0$. (i.o. means "infinitely often")
2. **(BC2)**: If $(A_n)_{n=0}^\infty$ is a sequence of independent sets in X and $\sum_{n=0}^\infty \mathbb{P}(A_n) = \infty$, then $\mathbb{P}(x \in A_n \text{ i.o.}) = 1$.
3. **(BC3)**: If $(A_n)_{n=0}^\infty$ is a sequence of pairwise independent sets in X and $\sum_{n=0}^\infty \mathbb{P}(A_n) = \infty$, then for \mathbb{P} a.e. $x \in X$,

$$\frac{S_n(x)}{E_n} \rightarrow 1$$

where $S_n(x) = \sum_{i=0}^{n-1} \mathbb{1}_{A_i}(x)$ and $E_n = \sum_{i=0}^{n-1} \mathbb{P}(A_i)$.

Given a measure-preserving transformation $T : X \rightarrow X$ and a probability space (X, \mathcal{B}, μ) and a sequence of sets $(A_n)_{n=0}^\infty$ in \mathcal{B} it is natural to ask whether $T^i x$ enters A_i infinitely often and at what rate this reentry occurs. We may reformulate the classical Borel-Cantelli lemmas in the dynamical framework [58, Section 3.1].

Dynamical Borel-Cantelli Lemmas

1. **Borel-Cantelli (BC)**: A sequence of sets (A_n) is said to be a Borel-Cantelli (BC) sequence if $\mu(x : T^n x \in A_n \text{ i.o.}) = 1$.
2. **strong Borel-Cantelli (SBC)**: A sequence of sets (A_n) is said to be a strong Borel-Cantelli (SBC) sequence if for μ a.e. $x \in X$ we have,

$$\frac{S_n(x)}{E_n} \rightarrow 1$$

as $n \rightarrow \infty$ where $S_n(x) = \sum_{i=0}^{n-1} \mathbb{1}_{A_i} \circ T^i(x)$.

Consider $B_i = T^{-i} A_i$ then if $\sum_{i=0}^\infty \mu(B_i) = \infty$ and the events B_i are independent we have $\frac{S_n(x)}{E_n} \rightarrow 1$ almost surely as $n \rightarrow \infty$. Moreover, the independence requirement can be relaxed to pairwise independence, e.g. $\mu(B_i \cap B_j) = \mu(B_i)\mu(B_j)$ for $i \neq j$. In many cases the sequence $B_i = T^{-i} A_i$ is not pairwise independent; however, it is reasonable to consider some sufficiently fast decay of correlations so that for large enough i the sequence is *almost* pairwise independent.

Note that if a sequence is strong Borel-Cantelli then $S_n = E_n + o(E_n)$. A stronger version of the SBC property is given in terms of a quantitative estimate on the error rate and is typically used in applications.

3. **quantitative strong Borel-Cantelli (QSBC)**: A sequence is said to be quantitative strong Borel-Cantelli (QSBC) [21] if for μ -a.e. $x \in X$ and $\epsilon > 0$ we have,

$$S_n = E_n + O(E_n^{1/2} \log^{3/2+\epsilon} E_n)$$

Consider the quantity,

$$R_{ij} = \mu(B_i \cap B_j) - \mu(B_i)\mu(B_j) = \mu(T^{-i}A_i \cap T^{-j}A_j) - \mu(A_i)\mu(A_j)$$

then the assumption that,

$$\exists C > 0 : \sum_{i,j=N}^n R_{ij} \leq C \cdot \sum_{i=N}^n \mu(A_i)$$

for all $n \geq N \geq 1$ is known as the **(SP)** property (Sprindzuk Property) [93]. If a sequence $(A_n)_{n=0}^\infty$ satisfies the SP property then it is a QSBC sequence [21].

Remark 1.2.1. By definition, independent sequences satisfy the (SP) property so that the QSBC property holds; however, it is possible to obtain better convergence rates under certain boundedness assumptions on the observable.

1.3 Extreme Value Theory

Extreme value theory deals with modeling and establishing statistical properties of maxima within a system and attracts interest in multiple fields of study, including mathematics, physics, climate-science, and finance. Since extremes occur in the tail of probability distributions, they relate to predictions of rare events. An analysis of recurrence times and other statistics of these events allows us to estimate risks.

Extreme value laws have been well-established for i.i.d random variables. A main goal when studying extremes in the context of dynamics is to investigate the statistical properties of the maxima when dependence or nonstationarity is observed. The next few sections discuss classical results from extreme value theory for i.i.d random variables, extend these results under more relaxed conditions and finally discuss applications to dynamical systems viewed as a stochastic processes.

1.3.1 Extreme value theory for i.i.d. processes

Given a sequence of i.i.d. random variables X_0, \dots, X_{n-1} define F as the cumulative distribution function for X_0 so that $\mathbb{P}(X_0 \leq x) = F(x)$.

Definition 1.3.1. If a random variable has the property that the Lebesgue-Stieltjes measure associated to F , μ_F , is absolutely continuous with respect to Lebesgue measure m we may define the **probability density function (p.d.f)** as the Radon-Nikodym derivative function $f = d\mu_F/dm$. Moreover, if f is Riemann integrable and continuous at x we have $f(x) = dF(x)/dx$.

Generally, a rare event or extreme event at time n corresponds to the occurrence of an exceedance of a threshold u denoted by,

$$U(u) := \{X_n > u\}.$$

When X is essentially bounded we define,

$$u_F = \sup\{x : F(x) < 1\}.$$

Extreme values of a random variable in the i.i.d. case relate directly to the behavior of the tail of the of its distribution function (d.f.). Let $\bar{F} = 1 - F$ be the complementary d.f. then the speed at which \bar{F} approaches 0 as $u \rightarrow u_F$ establishes the type of tail. Informally, F is said to have *heavy* tails if $u_F = \infty$ and \bar{F} vanishes polynomially fast and *light* tails if $u_F < \infty$ or \bar{F} vanishes exponentially fast.

Given our sequence of random variables define $M_n = \max\{X_0, \dots, X_{n-1}\}$. Classical statistical analysis is mostly concerned with establishing almost-sure and distributional convergence results for the mean, $\frac{1}{n} \sum_{i=0}^{n-1} X_i$ and $\frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} X_i$. In the case of i.i.d random variables these results are known as the **Laws of Large Numbers** and **Central Limit Theorem** respectively. For extremes of i.i.d random variables we have that M_n converges almost surely to u_F [73]. The question of general distributional convergence for M_n is addressed in the following definition [73, Definition 2.2.1].

Definition 1.3.2. We say that we have an **extreme value law (EVL)** for M_n if there is a non-degenerate d.f. $H : \mathbb{R} \rightarrow [0, 1]$ with $H(0) = 0$ and, for every $\tau > 0$, there exists a sequence of levels $u_n = u_n(\tau)$, $n = 1, 2, \dots$, such that

$$n\mathbb{P}(X_0 > u_n(\tau)) \rightarrow \tau, \quad \text{as } n \rightarrow \infty \quad (3)$$

and for which the following holds:

$$\mathbb{P}(M_n \leq u_n(\tau)) \rightarrow \bar{H}(\tau), \quad \text{as } n \rightarrow \infty$$

where the convergence is meant to hold at the continuity points of $H(\tau)$.

Motivation for the chosen normalizing sequences satisfying (3) comes once again from the classical i.i.d case where $\mathbb{P}(M_n \leq u) = (F(u))^n$, where F is the d.f. of X_0

and hence,

$$\mathbb{P}(M_n \leq u_n(\tau)) = (1 - \mathbb{P}(X_0 > u_n(\tau)))^n \sim \left(1 - \frac{\tau}{n}\right)^n \rightarrow e^{-\tau},$$

as $n \rightarrow \infty$.

In classical theory, the sequences of real numbers $u_n = u_n(\tau)$, $n = 1, 2, \dots$ are usually taken to be one parameter linear families like,

$$u_n(\tau) = y(\tau)/a_n + b_n,$$

where $y(\tau) \in \mathbb{R}$ and $a_n > 0$, for all $n \in \mathbb{N}$ which brings us to the main classical result of extreme value theory.

Theorem 1.3.3. [73, Theorem 3.1.1] *If X_0, X_1, \dots is a sequence of i.i.d random variables and there exist linear normalizing sequences (a_n) and (b_n) , with $a_n > 0$ for all n , such that,*

$$\mathbb{P}(a_n(M_n - b_n) \leq y) \rightarrow G(y),$$

where the convergence occurs at the continuity points of G , and G is the nondegenerate (that is, there is no $y_0 \in \mathbb{R}$ such that $G(y_0) = 1$ and $G(y) = 0$ for all $y < y_0$) function $\bar{H}(\tau)$ under this scaling, then $G(y) = e^{-\tau(y)}$ where $G(y)$ can be one of the following types,

- Type 1 (Gumbel): $G(y) = e^{-e^{-y}}$
- Type 2 (Fréchet): $G(y) = \begin{cases} 0 & x \leq 0 \\ e^{-y^{-\alpha}} & \alpha > 0, y > 0 \end{cases}$

- *Type 3 (Weibull):* $G(y) = \begin{cases} e^{-(-y)^\alpha} & \alpha > 0, y \leq 0 \\ 1 & y > 0 \end{cases}$

Definition 1.3.4. We may combine all three extreme value distributions into the following d.f. known as the **Generalized Extreme Value Distribution (GEV)**,

$$G(y) = \begin{cases} e^{-(1+\xi y)^{-1/\xi}}, & 1 + \xi y > 0, \text{ if } \xi \neq 0 \\ e^{-e^{-y}}, & y \in \mathbb{R}, \text{ if } \xi = 0 \end{cases}$$

In this context the shape parameter or ξ determines the type of distribution by its tail. When $\xi = 0$ the distribution corresponds to the Gumbel type; when $\xi > 0$ it corresponds to a Fréchet; and when $\xi < 0$ it corresponds to a Weibull. Generally, an *exponential tail* is displayed in Gumbel type distributions, a *heavy tail* in the Fréchet type and the Weibull type display an *upper bound*.

Three Types of GEV Distributions

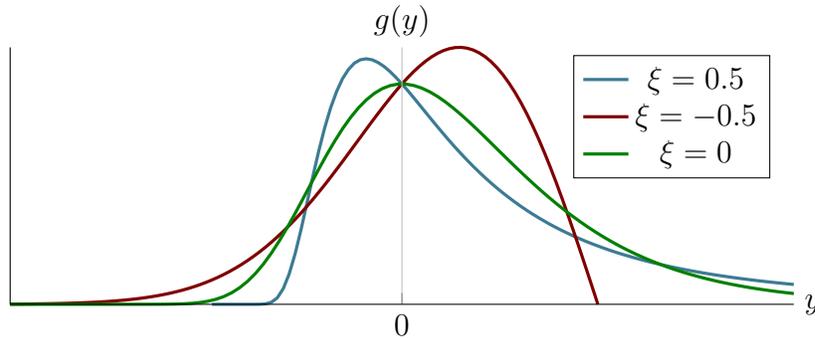


Figure 1.2: Generalized extreme value probability distribution functions (PDF) for three different values of the shape parameter corresponding to the three different types with $\mu = 0$ and $\sigma^2 = 1$.

A reformulation of the GEV commonly used in numerical settings is given in

terms of the location μ , scale σ and shape ξ parameters and has the following form.

$$\mathbb{P}(M_n \leq y) \rightarrow G(y) = \exp \left\{ - \left[1 + \xi \left(\frac{y - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}$$

defined on the set $\{y : 1 + \xi(y - \mu)/\sigma > 0\}$, where $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$. The subset of the GEV family with $\xi = 0$ is interpreted as the limit $\xi \rightarrow 0$ which leads to

$$G(y) = \exp \left[- \exp \left\{ - \left(\frac{y - \mu}{\sigma} \right) \right\} \right]$$

1.3.2 Extreme value theory for dependent processes

The requirement of i.i.d on the sequence of random variables is commonly not satisfied in applications. It is interesting to consider whether an extreme value law still holds when this assumption is modified. The general idea is to determine under which assumptions the dependent process is quantifiably close to an independent process.

1.3.2.1 Stationary sequences and dependence conditions

A sequence (X_n) is called stationary if the distributions of $(X_{j_1}, \dots, X_{j_n})$ and $(X_{j_1+k}, \dots, X_{j_n+k})$ are identical for any choice of n and j_1, \dots, j_n . Leadbetter proposed that an extreme value law can still be obtained provided the sequence of random variables is stationary and the following two dependence conditions hold [68].

Condition $D(u_n)$ (mixing condition): Given the sequence X_1, \dots, X_n , for any

integers $i_1 < \dots < i_p$ and j_1, \dots, j_k for which $j_1 - j_p > t$, and any large $n \in \mathbb{N}$,

$$|F_{i_1, \dots, i_p, j_1, \dots, j_k}(u_n) - F_{i_1, \dots, i_p}(u_n)F_{j_1, \dots, j_k}(u_n)| \leq \alpha(n, t)$$

uniformly for every $p, k \in \mathbb{N}$, where F_{i_1, \dots, i_p} denotes the joint distribution function of X_{i_1}, \dots, X_{i_p} for any $t_n = o(n)$ such that $\alpha(n, t_n) \rightarrow 0$ as $n \rightarrow \infty$.

We can think of the $D(u_n)$ condition as a mixing condition in which if two large blocks are sufficiently apart then the joint probabilities of the two blocks is approximately the product of the probability of the individual blocks.

Condition $D'(u_n)$ (recurrence condition): Given the sequence X_1, \dots, X_n there exists a sequence k_n such that $k_n \rightarrow \infty$, $\lim_{n \rightarrow \infty} k_n \alpha(n, t_n) = 0$ and $k_n t_n = o(n)$ and,

$$\lim_{n \rightarrow \infty} \sum_{j=1}^{\lfloor n/k_n \rfloor} \mathbb{P}(X_0 > u_n, X_j > u_n) = 0$$

Under the requirement $\lim_{n \rightarrow \infty} n\mathbb{P}(X_0 > u_n(\tau)) = \tau$ there are approximately τ exceedances of u_n among X_1, \dots, X_n , and thus τ/k_n among $X_1, \dots, X_{\lfloor n/k_n \rfloor}$. Condition $D'(u_n)$ bounds the probability of more than one exceedance of u_n among $X_1, \dots, X_{\lfloor n/k_n \rfloor}$ and thus determines a simple Poisson limit for this point process [73, Section 3.3.2]. Rare event point processes are discussed in more detail in Chapter 1, Section 1.4.

Theorem 1.3.5. [73, Theorem 3.2.1] *Let X_0, X_1, \dots be a stationary stochastic process and $(u_n(\tau))_{n \in \mathbb{N}}$ a sequence satisfying $\lim_{n \rightarrow \infty} n\mathbb{P}(X_0 > u_n(\tau)) = \tau$, for some $\tau > 0$. If $D(u_n)$ and $D'(u_n)$ hold, then $\bar{H}(\tau) = e^{-\tau}$.*

The core argument for this theorem comes from breaking the n observations into k_n blocks of size $\lfloor n/k_n \rfloor$ and separating them by adding a sequence of t_n observations

between each block. Condition $D(u_n)$ ensures independence between each block, condition $D'(u_n)$ decorrelates the information within each block so that the process is quantifiably close to an independent process where the difference converges to zero as $n \rightarrow \infty$. This is known as the *blocking argument*.

In the previous result, the recurrence condition $D'(u_n)$ ensures that the limiting distribution for M_n exists and is given as $\bar{H}(\tau) = e^{-\tau}$ (equivalent to the i.i.d case). When $D'(u_n)$ does not hold but $D(u_n)$ does, although existence may not be assumed a priori, another limiting distribution for M_n is defined.

Theorem 1.3.6. *[73, Theorem 3.2.3] Let X_0, X_1, \dots be a stationary stochastic process and $(u_n)_{n \in \mathbb{N}}$ a sequence satisfying $\lim_{n \rightarrow \infty} n\mathbb{P}(X_0 > u_n(\tau)) = \tau$. Suppose $D(u_n)$ holds for each choice of τ . If the limit of $\mathbb{P}(M_n \leq u_n)$ exists then there exists $0 \leq \theta \leq 1$ such that $\bar{H}(\tau) = e^{-\theta\tau}$ for all $\tau > 0$.*

The consecutive occurrences of an exceedance of a given threshold can cause $D'(u_n)$ to fail. This is often referred to as *clusters* of exceedances and is related to the memory properties of the underlying process. It has been shown that this setting essentially produces the same type of EVL with a parameter $0 \leq \theta \leq 1$ so that $\bar{H}(\tau) = e^{-\theta\tau}$ where θ is known as the *extremal index* and $\frac{1}{\theta}$ roughly measures the number of clustered exceedances.

1.3.2.2 General dependence conditions motivated by dynamics

Consider a stationary stochastic process, X_0, X_1, \dots , a measure space $(X, \mathcal{B}, \mathbb{P})$ and a map T defined by

$$X_{i-1} \circ T = X_i$$

for all $i \in \mathbb{N}$ where T^i denotes the i th composition of the map T and T^0 is the identity map on X . In this setting, stationarity is equivalent to the statement that \mathbb{P} is T -invariant. Following the same notation described in previous sections and recalling $U(u) = \{X_0 > u\}$ we define,

$$A^{(q)}(u) := U(u) \cap \bigcap_{i=1}^q T^{-i}(U(u)^c) = \{X_0 > u, X_1 \leq u, \dots, X_q \leq u\}.$$

In other words, $A^{(q)}(u)$ corresponds to the case when an exceedance occurs at time zero and does not occur again up to time $t = q$. For convenience take $A^{(0)}(u) := U(u)$, $U_n := U(u_n)$ and $A_n^{(q)} := A_n^{(q)}(u_n)$, for all $n \in \mathbb{N}$ and $q = \mathbb{N} \cup \{0\}$. Let,

$$\theta = \lim_{n \rightarrow \infty} \theta_n := \frac{\mathbb{P}(A_n^{(q)})}{\mathbb{P}(U_n)}$$

Now let $B \in \mathcal{B}$ be an event. For some $s \geq 0$ and $l \geq 0$ define,

$$\mathcal{W}_{s,l}(B) = \bigcap_{i=s}^{s+\max\{l-1,0\}} T^{-i}(B^c).$$

The disadvantage of the dependence condition $D(u_n)$ is that it cannot be verified easily in the dynamical setting by using information about mixing rates or decay in various norms of the system. To develop theory with more practical utility in this context [53] proposed a new condition, $D_2(u_n)$ for general stochastic processes to replace $D(u_n)$ and, together with $D'(u_n)$, establishes an EVL for the system.

Periodicity poses another issue since it may result in short returns of exceedances resulting in a failure of the recurrence assumption $D'(u_n)$. Further adjustments motivated by the relationship between periodicity and the presence of cluster of exceedances [42] resulted in a new set of conditions $D_p(u_n)$ and $D'_p(u_n)$ for systems of periodic behavior of period p . Finally, the following conditions were proposed which combine all scenarios with no periodic behavior, simple periodic behavior or multiple types of behavior.

Condition $\mathbb{D}_q(u_n)$: We say that $\mathbb{D}_q(u_n)$ holds for the sequence X_0, X_1, \dots if for every $\ell, t, n \in \mathbb{N}$,

$$|\mathbb{P}(A_n^{(q)} \cap \mathcal{W}_{t,\ell}(A_n^{(q)})) - \mathbb{P}(A_n^{(q)})\mathbb{P}(\mathcal{W}_{0,\ell}(A_n^{(q)}))| \leq \gamma(q, n, t)$$

where $\gamma(q, n, t)$ is decreasing in t for each n and, there exists a sequence $(t_n)_{n \in \mathbb{N}}$ such that $t_n = o(n)$ and $n\gamma(q, n, t_n) \rightarrow 0$ as $n \rightarrow \infty$.

Condition $\mathbb{D}'_q(u_n)$: We say that $\mathbb{D}'_q(u_n)$ holds for the sequence X_0, X_1, \dots if there exists a sequence $(k_n)_{n \in \mathbb{N}}$ such that $k_n \rightarrow \infty$ and $k_n t_n = o(n)$ and

$$\lim_{n \rightarrow \infty} n \sum_{j=q+1}^{\lfloor n/k_n \rfloor - 1} \mathbb{P}(A_n^{(q)} \cap T^{-j}(A_n^{(q)})) = 0$$

As stated, $\mathbb{D}_q(u_n)$ and $\mathbb{D}'_q(u_n)$ can be used to prove the existence of an EVL through the following corollary.

Corollary 1.3.7. [73, Corollary 4.1.7] *Let X_0, X_1, \dots be a stationary stochastic process and $(u_n)_{n \in \mathbb{N}}$ a sequence satisfying $\lim_{n \rightarrow \infty} n\mathbb{P}(X_0 > u_n(\tau)) = \tau$, for some $\tau > 0$. Assume that conditions $\mathbb{D}_q(u_n)$ and $\mathbb{D}'_q(u_n)$ hold for some $q \in \mathbb{N} \cup \{0\}$ and $(t_n)_{n \in \mathbb{N}}$ and $(k_n)_{n \in \mathbb{N}}$ are the sequences in those conditions. Moreover, assume that*

the limit $\theta = \lim_{n \rightarrow \infty} \theta_n := \frac{\mathbb{P}(A_n^{(g)})}{\mathbb{P}(U_n)}$ exists. Then,

$$\lim_{n \rightarrow \infty} \mathbb{P}(M_n \leq u_n) = e^{-\theta\tau}.$$

Given our dynamical system (X, T, μ) we consider an observable φ taken on the map which represents some physically relevant variable which can be measured so that our stochastic process is given by $X_0 = \varphi(x)$ and $X_n = \varphi \circ T^n(x)$ where $x \in X$. In the classical extreme value setting the observable is typically taken as a function of the distance to some fixed point, $\varphi(x) = g(\text{dist}(x, p))$ where $p \in X$. It has been shown that the choice of g determines the type of tail (extremal distribution) $G(y)$.

- Type 1 (Gumbel) law for $\varphi(x) = -\log d(x, p)$
- Type 2 (Fréchet) law for $\varphi(x) = d(x, p)^{-k}$, $k > 0$
- Type 3 (Weibull) law for $\varphi(x) = C - d(x, p)^k$, $k > 0$ and C constant

The following examples outline how each distribution type is determined for the **doubling map**.

Example 1.3.8. Consider the observable $\varphi(x) = -\log(|x - 1/3|)$ taken on $([0, 1], T, m)$ with m taken as Lebesgue measure. First we solve $F(u) = m(\{X_0 \leq u\})$ by solving $-\log(|x - 1/3|) \leq u$ for u . This gives $\{x : \varphi(x) \leq u\} = [0, 1/3 - e^{-u}] \cup [1/3 + e^{-u}, 1]$ so that $F(u) = 1 - 2e^{-u}$. The series of thresholds (u_n) defined by $\lim_{n \rightarrow \infty} n\mathbb{P}(X_0 > u_n(\tau)) = \tau$ are assumed to hold. Moreover, $n(1 - F(u_n)) = \tau$ so that $u_n = -\log(\tau) + \log(2n)$. Solving for parameter $\tau = e^{-y}$ and we conclude that the underlying distribution is of the Gumbel type.

Dynamics of the Doubling Map

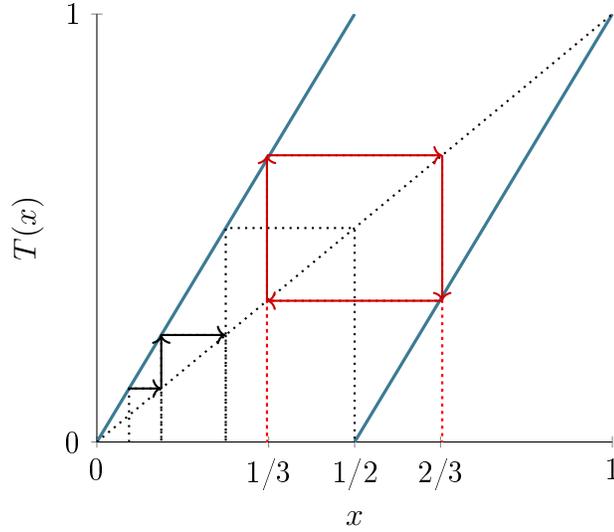


Figure 1.3: The doubling map with a nonperiodic orbit of a point (black) and a periodic orbit of a point (red).

Example 1.3.9. Consider the observable $\varphi(x) = (|x - 1/3|)^{-k}$ so that $\{x : \varphi(x) \leq u\} = [0, 1/3 - u^{-1/k}] \cup [1/3 + u^{-1/k}, 1]$ and $F(u) = 1 - 2u^{-1/k}$. By a similar calculation to the previous example we have $u_n = 2n^{-1/k}\tau^{-1/k}$. Solving for our parameter $\tau = y^{-k}$ and we conclude that the underlying distribution is of the Fréchet type.

Example 1.3.10. Consider the observable $\varphi(x) = 1 - (|x - 1/3|)^{1/k}$ so that $\{x : \varphi(x) \leq u\} = [0, 1/3 - (1 - u)^k] \cup [1/3 + (1 - u)^k, 1]$ and $F(u) = 1 - 2(1 - u)^k$. By a similar calculation to the previous example we have $u_n = 1 - \frac{\tau^{1/k}}{2n^{1/k}}$. Solving for our parameter $\tau = -y^k$ and we conclude that the underlying distribution is of the Weibull type.

Recall that, for those dependent sequences of random variables which satisfy conditions $D(u_n)$ and $D'(u_n)$, the Generalized Extreme Value distribution (GEV)

can be fit to the extremes. Typically a numerical GEV fit is performed by some version of likelihood estimation and the the value of the extremal index (which is a result of the dependence in the system) is buried within the estimated parameters. It is interesting to ask whether we are able to recover the value of θ numerically. The following theorem relates the standard GEV model to the GEV model with an extremal index [27, Theorem 5.2].

Theorem 1.3.11. *[27, Theorem 5.2] Let X_1, X_2, \dots be a stationary process and X_1^*, X_2^*, \dots be a sequence of independent variables with the same marginal distribution. Define $M_n = \max\{X_1, \dots, X_n\}$ and $M_n^* = \max\{X_1^*, \dots, X_n^*\}$. Under suitable regularity conditions ($D(u_n)$ and $D'(u_n)$ or variations thereof),*

$$\mathbb{P}\{(M_n^* - b_n)/a_n \leq z\} \rightarrow G_1(z)$$

as $n \rightarrow \infty$ for normalizing sequences $\{a_n > 0\}$ and $\{b_n\}$, where G_1 is a non-degenerate distribution function, if and only if

$$\mathbb{P}\{(M_n - b_n)/a_n \leq z\} \rightarrow G_2(z)$$

where

$$G_2(z) = G_1^\theta(z)$$

for a constant θ such that $0 < \theta \leq 1$.

Moreover, if G_1 corresponds to a GEV distribution with parameters (μ, σ, ξ) and $\xi \neq 0$ then

$$\begin{aligned} G_1^\theta(z) &= \exp\left\{-\left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right]^{-1/\xi}\right\}^\theta \\ &= \exp\left\{-\left[1 + \xi\left(\frac{z - \mu^*}{\sigma^*}\right)\right]^{-1/\xi}\right\}^\theta \end{aligned}$$

where

$$\mu^* = \mu - \frac{\sigma}{\xi}(1 - \theta^{-\xi}) \quad \text{and} \quad \sigma^* = \sigma\theta^\xi$$

If $\xi = 0$ for G_1 then it is also for G_1^θ and

$$\mu^* = \mu + \sigma \log \theta \quad \text{and} \quad \sigma^* = \sigma$$

Dynamics on frequently investigated transformations, such as the doubling map, are defined precisely so that we may consider a point which is non-periodic, compare the resulting parameters of the distribution to that of the periodic point, and recover the value of θ in a straight-forward way. It is more interesting, however, to consider time series where the underlying dynamics are relatively unknown and where this procedure is certain to fail. In this setting, results on the numerical estimation of θ use the informal definition of the extremal index (a representation of the number of clusters of exceedances) to obtain approximations on this parameter. If we let a general form for the estimator of θ be given by [73],

$$\hat{\theta} := \frac{C(u_n)}{N(u_n)},$$

where $N(u_n)$ is the number of exceedances of a high threshold u_n and $C(u_n)$ the number of clusters then the literature gives two general ways of estimating the extreme index.

Runs estimator Given $1 < \ell < n$,

$$\hat{\theta}_R := \frac{1}{N(u_n)} \sum_{i=1}^{n-\ell} \mathbb{1}_{\{X_i > u_n\}} \mathbb{1}_{\{X_{i+1} \leq u_n\}} \cdots \mathbb{1}_{\{X_{i+\ell} \leq u_n\}}.$$

In other words, $\hat{\theta}_R$ only defines a cluster if there are at least ℓ consecutive observations below the threshold between exceedances of the threshold.

Blocks estimator Given a sequence of n observations, we divide the a sample into k blocks of length m so that $n \approx km$. If an observation inside one of the k blocks exceeds (u_n) we call that block a *cluster*. We sum up all the clusters and denote this sum by $C(u_n)$. We define,

$$\hat{\theta}_B = \frac{C(u_n)}{N(u_n)}$$

where $N(u_n)$ is the total number of exceedances of the threshold.

1.4 Rare Event Point Processes

When studying rare events, it is interesting to consider records of the exceedances of the high thresholds u_n . These records are often referred to as rare event point processes (REPP) and are defined by the following,

$$\mathcal{N}_{u_n}(A) := \sum_{i \in A \cap \mathbb{N}} \mathbb{1}_{(X_i > u_n)}.$$

In other words, the REPP counts the number of exceedances of u_n so that $\mathcal{N}_{u_n}([0, n])$ represents the count of these exceedances among the first n observations of the process X_0, X_1, \dots, X_{n-1} . As is common, we consider the i.i.d case for $\mathcal{N}_{u_n}([0, n])$ which by definition is binomial with parameters $(n, \mathbb{P}(X_0 > u_n))$. By $\lim_{n \rightarrow \infty} n\mathbb{P}(X_0 > u_n(\tau)) = \tau$ we have that the average number of successes converges to $\tau \geq 0$ which implies that this average is nearly constant and $\mathcal{N}_{u_n}([0, n])$ is asymptotically Poisson. Hence, it is natural to ask whether we obtain a Poisson process in the limit for the dependent setting. For stationary processes where dependence is observed we need to define the REPP over a rescaled time period so that the average number of exceedances can be kept stabilized (converging to 1 on the unit interval $[0, 1]$). Note that

in the independent case the rescaling factor would be taken as $v_n := 1/\mathbb{P}(X_0 > u_n)$ following Kac's theorem.

Let \mathcal{S} denote the semi-ring of subsets of \mathbb{R}_0^+ whose elements are intervals of the type $[a, b)$, for $a, b \in \mathbb{R}_0^+$. Let \mathcal{R} denote the ring generated by \mathcal{S} . Recall that for every $\mathcal{J} \in \mathcal{R}$ there are $k \in \mathbb{N}$ and k intervals $I_1, \dots, I_k \in \mathcal{S}$ such that $\mathcal{J} = \cup_{j=1}^k I_j$. Let $a_j, b_j \in \mathbb{R}_0^+$ be such that $I_j = [a_j, b_j) \in \mathcal{S}$. For $I = [a, b) \in \mathcal{S}$ and $\alpha \in \mathbb{R}$, denote $\alpha I := [\alpha a, \alpha b)$ and $I + \alpha := [a + \alpha, b + \alpha)$. Similarly, for $\mathcal{J} \in \mathcal{R}$ define $\alpha \mathcal{J} := \alpha I_1 \cup \dots \cup \alpha I_k$ and $\mathcal{J} + \alpha := (I_1 + \alpha) \cup \dots \cup (I_k + \alpha)$.

Definition 1.4.1. A REPP is defined by counting the number of exceedances during the (rescaled) time period $v_n \mathcal{J} \in \mathcal{R}$, where $\mathcal{J} \in \mathcal{R}$. That is, for every $\mathcal{J} \in \mathcal{R}$ set

$$N_n(\mathcal{J}) := \mathcal{N}_{u_n}(v_n \mathcal{J}) = \sum_{j \in v_n \mathcal{J} \cap \mathbb{N} \cup \{0\}} \mathbb{1}_{X_j > u_n}$$

Next we will define a slightly stronger mixing condition [73, Section 3.3.3].

Condition $(\Delta(u_n))$ For $0 \leq i \leq j \leq n$, let $\mathcal{F}_i^j(u_n)$ be the σ -algebra generated by the events $\{X_s \leq u_n\}$, for $i \leq s \leq j$, where $(u_n)_{n \in \mathbb{N}}$ is a sequence of levels satisfying $\lim_{n \rightarrow \infty} n \mathbb{P}(X_0 > u_n(\tau)) = \tau$. Let

$$\alpha_{n,t} = \max\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{F}_1^p(u_n), B \in \mathcal{F}_{p+t}^k(u_n), 0 \leq p \leq n-1-t\}.$$

Then $\Delta(u_n)$ holds for the sequence X_0, X_1, \dots if $\alpha(n, t_n) \rightarrow_{n \rightarrow \infty} 0$, for some sequence $t_n = o(n)$.

Let $(k_n)_{n \in \mathbb{N}}$ be a sequence satisfying $D(u_n)$ then the cluster size distribution for each $j \in \mathbb{N}$

$$\pi_n(j) = \mathbb{P}(\mathcal{N}([0, n/k_n]))$$

The following main theorem of this section relates the REPP to the existence of an extreme value law (EVL).

Theorem 1.4.2. *[73, Theorem 3.3.6] Assume that X_0, X_1, \dots is a stationary stochastic process satisfying condition $\Delta(u_n)$ and $\lim_{n \rightarrow \infty} \mathbb{P}(M_n \leq u_n) = e^{-\eta}$ for some $\eta > 0$ and a sequence $(u_n)_{n \in \mathbb{N}}$ as in $\lim_{n \rightarrow \infty} n\mathbb{P}(X_0 > u_n(\tau)) = \tau$. Suppose that there exists a probability distribution π defined on \mathbb{N} such that $\pi(j) = \lim_{n \rightarrow \infty} \pi_n(j)$ for every $j \in \mathbb{N}$. Then the REPP N_n converges in distribution to a compound Poisson process N , with intensity $\theta = \eta/\tau$ and multiplicity distribution π .*

In dynamical applications the condition $\Delta(u_n)$ is typically difficult to prove, so adaptations of this condition have been introduced.

Absence of clustering. When condition $\mathbb{D}'_0(u_n) = D'(u_n)$ holds there is no clustering. In this setting the REPP converges in distribution to a standard Poisson process (since the EI $\theta = 1$) for simple point processes without multiple events and is shown to hold for those systems which satisfy the following adapted condition [13].

Condition ($D_3(u_n)$) Let $A \in \mathcal{R}$ and $t \in \mathbb{N}$. We say that $D_3(u_n)$ holds for the sequence X_0, X_1, \dots if

$$|\mathbb{P}(\{X_0 > u_n\} \cap \{\mathcal{N}_{u_n}(A+t) = 0\}) - \mathbb{P}(\{X_0 > u_n\})\mathbb{P}(\mathcal{N}_{u_n}(A) = 0)| \leq \gamma(n, t)$$

where $\gamma(n, t)$ is nonincreasing in t for each n and $n\gamma(n, t_n) \rightarrow 0$ as $n \rightarrow \infty$ for some sequence $t_n = o(n)$, that is $t_n/n \rightarrow 0$ as $n \rightarrow \infty$. The proof of condition $D_3(u_n)$ follows, after minor adjustments, the proof scheme of $\mathbb{D}_q(u_n)$.

Presence of clustering. Since condition $\mathbb{D}'_0(u_n)$ prevents the existence of clusters of exceedances the associated EVL is standard exponential $\bar{H}(\tau) = e^{-\tau}$. When $\mathbb{D}'_0(u_n)$

fails a new parameter $\theta \in (0, 1)$ is introduced so that $\bar{H}(\tau) = e^{-\theta\tau}$. In this setting distributional convergence of the REPP exists as a compound Poisson process with parameter θ has been shown to hold provided condition $\bar{D}'(u_n)$ and the following adapted condition $D_q(u_n)^*$ hold for the system.

Let p be a periodic point of prime period q . Recall $U_n = (X_0 > u_n)$ and define the sequence $(U^{(k)}(u_n))$ of nested sets centered at p given by

$$U^{(0)}(u_n) = U_n \quad \text{and} \quad U^{(k)}(u_n) = T^{-q}(U^{(k-1)}(u_n)) \cap U_n, \quad \text{for all } k \in \mathbb{N}.$$

For $i, k, l, s \in \mathbb{N} \cup \{0\}$, we define the following sets:

$$Q_{q,i}^k(u_n) := T^{-i}(U^{(k)}(u_n) - U^{(k+1)}(u_n)).$$

Note that $Q_{q,0}^0(u_n) = A_n^q$. Furthermore, $U_n = \cup_{k=0}^{\infty} Q_{q,0}^k(u_n)$.

Condition ($D_q^*(u_n)$) We say that $D_q^*(u_n)$ holds for the sequence X_0, X_1, \dots if for any integers t, k_1, \dots, k_q n and any $J = \cup_{j=2}^q I_j \in \mathbb{R}$ with $\inf\{x : x \in J\} \geq t$,

$$|\mathbb{P}(Q_{q,0}^{k_1}(u_n) \cap (\cap_{j=2}^q N_n(I_j) = k_j)) - \mathbb{P}(Q_{q,0}^{k_1}(u_n))\mathbb{P}(\cap_{j=2}^q N_n(I_j) = k_j)| \leq \gamma(n, t)$$

where for each n we have that $\gamma(n, t)$ is nonincreasing in t and $n\gamma(n, t_n) \rightarrow 0$ as $n \rightarrow \infty$, for some sequence $t_n = o(n)$.

The advantage of these new conditions is that they follow from sufficiently fast decay of correlations of the dynamical system.

Proposition 1.4.3. [73, Proposition 4.4.1] *Assume that for a system (X, T, μ) we have decay of correlations for all $\psi_1 \in BV$ and all $\psi_2 \in L^\infty$ so that there exist C , independent of ψ_1, ψ_2 and a rate function $\rho : \mathbb{N} \rightarrow \mathbb{R}$ such that,*

$$\left| \psi_1 \cdot (\psi_2 \circ T^i) d\mu - \int \psi_1 d\mu \int \psi_2 d\mu \right| \leq C \|\psi_1\|_{BV} \|\psi_2\|_{\infty} \rho(i)$$

for all $i \in \mathbb{N}$ and $n\rho(i_n) \rightarrow 0$, as $n \rightarrow \infty$ for some $i_n = o(n)$. Then conditions $\mathcal{D}_q(u_n)$, $D_3(u_n)$, and $D_q^*(u_n)$ hold.

Remark 1.4.4. In most applications it can be shown that the clustering π follows a geometric distribution of parameter $\theta \in (0, 1]$, that is, $\pi_k = \theta(1 - \theta)^{k-1}$, for every $k \in \mathbb{N}$. This means that, as in [41, 13], $N([0, t])$ follows a Pólya-Aeppli distribution

$$\mathbb{P}(N([0, t])) = k = e^{-\theta t} \sum_{j=1}^k \theta^j (1 - \theta)^{k-j} \frac{(\theta t)^j}{j!} \binom{k-1}{j-1},$$

for all $k \in \mathbb{N}$, and that $\mathbb{P}(N([0, t]) = 0) = e^{-\theta t}$.

This behavior has been observed in a variety of one-dimensional expanding maps [40, 36, 61] as well as in two-dimensional hyperbolic linear toral automorphisms [57, 30, 13].

1.5 Statistical Methods for Modeling Extremes

It is certainly of interest to use the established theory on extreme values in real-world application through statistical modeling of rare events based on recorded data. This section will introduce two commonly used techniques for modeling extremes of general data: the block maximum [27, Chapter 3] and the peaks over threshold approach [27, Chapter 4]. Recall that for an extreme value law to be established the system must be stationary. In all the dynamical systems we have described the presence of an invariant measure implies the stationary assumption; however, real-world data may not behave so "neatly". In fact, nonstationarity in the data is quite common and, consequently, adaptations on the extreme value model have

been introduced. We describe these adaptations below. Complementary, though not unique to extreme modeling, machine learning methods are described towards the end of this section. The statistical methods outlined are used in the analysis of real-world data located in Chapter 4.

1.5.1 Generalized Extreme Value distribution approximation and block maxima approach

Recall from Chapter 1, Section 1.3, the definition of the Generalized Extreme Value distribution (GEV). The unification of the extremal distribution types provides a way of simplifying statistical implementation. Since there is no longer a need to make an apriori judgement on the tail behavior of the system we can look at the asymptotic behavior of the extreme values to determine the value of ξ and hence which family they belong to. But before we can discuss fitting the GEV, we need to establish under what conditions this can be done.

Recall that the assumptions for the existence of the limit require the sequence of random variables X_0, X_1, \dots to be i.i.d. In the case of modeling extremes taken from data it is enough to require stationarity of the sequence and independence over blocks of length m given by,

$$\{X_0, \dots, X_{m-1}\}, \dots, \{X_{n-m}, \dots, X_{n-1}\}.$$

The choice of block size can be critical and amounts to a trade-off between bias and accuracy of the fit. Blocks which are too small can lead to sequences which are not quite independent and thus do not follow the required assumptions on the model. In

contrast, blocks which are too large result in too few maxima and a wide variance in parameter estimates.

To estimate the distribution of extremes in this context we define the GEV over the set of **block maxima** given by $M_N = \max\{X_N, \dots, X_{N+m}\}$. A common technique is to estimate by **maximum likelihood estimation (MLE)** the location μ , scale σ and shape ξ parameters for the block maxima M_N under which the assumptions for the existence of a GEV distribution hold.

Modeling under stationary assumptions

Under the assumption that X_0, X_1, \dots are stationary and M_1, \dots, M_N are independent variables having the GEV distribution $G(y)$, the log-likelihood for the GEV parameters when $\xi \neq 0$ is,

$$\ell(\mu, \sigma, \xi) = -N \log \sigma - (1 + 1/\xi) \sum_{i=1}^N \log \left[1 + \xi \left(\frac{y_i - \mu}{\sigma} \right) \right] - \sum_{i=1}^N \left[1 + \xi \left(\frac{y_i - \mu}{\sigma} \right) \right]^{-1/\xi},$$

for

$$1 + \xi \left(\frac{y_i - \mu}{\sigma} \right) > 0$$

and

$$\ell(\mu, \sigma) = -N \log \sigma - \sum_{i=1}^N \left(\frac{y_i - \mu}{\sigma} \right) - \sum_{i=1}^N \left\{ - \left(\frac{y_i - \mu}{\sigma} \right) \right\}$$

when $\xi = 0$.

A potential difficulty when using MLE comes from the regularity requirements violated by the GEV model: the GEV model end-points are functions of the parameter so that $\mu - \sigma/\xi$ is the upper end-point when $\xi < 0$ but the lower end-point when $\xi > 0$. This issue has been partially overcome in [92] which state under what conditions of ξ the asymptotic likelihood results are valid.

- when $\xi > -0.5$, maximum likelihood estimators are regular, in the sense of having the usual asymptotic properties;
- when $-1 < \xi \leq -0.5$, maximum likelihood estimators are generally obtainable, but do not have the standard asymptotic properties;
- when $\xi \leq -1$, maximum likelihood estimators are unlikely to be obtainable

Since the case $\xi < -0.5$ corresponds to distributions with a very short bounded upper tail it is rarely the case that applications of extreme value modeling encounter the obstacles in likelihood estimation.

Modeling under non-stationary assumptions

In many instances the sequence of variables $X_0, X_1 \dots$ (and thus M_1, \dots, M_N) is non-stationary and apparent trends in the data raise doubts about model fit. In this case a practical approach to extreme modeling is adopted by considering the location μ , scale σ and shape ξ parameters as functions of time.

A simple linear trend, for example $\mu(t) = \beta_0 + \beta_1 t$ is typically checked in initial analyses. If appropriate, more complex changes in the parameters such as a quadratic

$$\mu(t) = \beta_0 + \beta_1 t + \beta_2 t^2$$

or a change-point model

$$\mu(t) = \begin{cases} \mu_1 & \text{for } t \leq t_0, \\ \mu_2 & \text{for } t > t_0. \end{cases}$$

are considered. This idea can be extended further by looking at the parameters as functions of variables which change over time. In all the examples presented the

extreme value parameters can be written in the form,

$$\theta(t) = h(X^T \beta)$$

where θ represents either μ , σ or ξ , h is a specified function, β is a vector of parameters, and X is a model vector. Under this representation the model can be fit using optimization techniques to minimize error; however, this type of estimation may provide too broad a fit for representing models with more than one time-dependent parameter.

The advantage of maximum likelihood estimation over other estimation techniques is its adaptability to changes in the model since all parameters are considered at each time step. Under the independence assumption for M_1, \dots, M_N if we have $t = 1, \dots, N$ with,

$$M_t \sim \text{GEV}(\mu(t), \sigma(t), \xi(t))$$

then the log-likelihood function for the GEV is,

$$\begin{aligned} \ell(\mu, \sigma, \xi) = - \sum_{t=1}^m \log \sigma(t) + (1 + 1/\xi(t)) \log \left[1 + \xi(t) \left(\frac{y_t - \mu(t)}{\sigma(t)} \right) \right] \\ + \left[1 + \xi(t) \left(\frac{y_t - \mu(t)}{\sigma(t)} \right) \right]^{-1/\xi(t)} \end{aligned}$$

provided that

$$1 + \xi(t) \left(\frac{y_t - \mu(t)}{\sigma(t)} \right) > 0, \quad \text{for } t = 1, \dots, N$$

It is important to discuss under what conditions a more complex model should be considered since overfitting the data can lead a model which poorly describes future or generated data. The following theorem for comparing two models by their maximum likelihood estimators is commonly used to make this determination.

Theorem 1.5.1. [27, Theorem 2.6] *Let x_1, \dots, x_n be independent realizations from a distribution within a parametric family \mathcal{F} , and let $\hat{\theta}_0$ denote the maximum likelihood estimator of the d -dimensional model parameter $\theta_0 = (\theta^1, \theta^2)$, where θ^1 is a k -dimensional subset of θ_0 and θ^2 corresponds to the remaining $(d - k)$ components of θ_0 . Then, under suitable regularity conditions, for large n*

$$D_p(\theta^1) = 2\{\ell(\hat{\theta}_0) - \ell_p(\theta^1)\} \sim \chi_k^2.$$

Suppose \mathcal{M}_1 is a model with parameter vector θ and model \mathcal{M}_0 is the subset of model \mathcal{M}_1 obtained by constraining k of the components of $\theta = (\theta^1, \theta^2)$. Let $\ell_1(\mathcal{M}_1)$ be the maximized log-likelihood for model \mathcal{M}_1 , similarly $\ell_0(\mathcal{M}_0)$ for \mathcal{M}_0 and define the deviance statistic

$$D = 2\{\ell_1(\mathcal{M}_1) - \ell_0(\mathcal{M}_0)\}.$$

Applying the theorem we have that $D \sim \chi_k^2$ so that large values of D indicate that model \mathcal{M}_1 explains significantly (specifically at the α significance level) more of the variation in the data than \mathcal{M}_0 whereas small values of D suggest that the increase in model complexity does not provide significant improvements. This test is known as the **likelihood ratio test**.

1.5.2 Generalized Pareto distribution and peaks over threshold approach

In some cases a different approach to modeling the extreme values is necessary, particularly those for which the block does not have many realizations. For example,

if we consider a time series with a single yearly weather recording we can reasonably assume independence; however we cannot calculate block maximum in this setting since it will result in too few points to fit a GEV model. In this setting we can look to the **peaks over threshold (POT)** method of gathering maximum values.

Given a sequence of independent and identically distributed random variables, having marginal distribution function F it is natural to consider an extreme event one which exceeds some high threshold u . If we denote an arbitrary X_i by X , we can rewrite this description in terms of the following conditional probability,

$$\mathbb{P}(X > u + y | X > u) = \frac{1 - F(u + y)}{1 - F(u)}, \quad y > 0. \quad (4)$$

In practical application it is uncommon to know the parent function F . The following theorem allows us to approximate such a distribution while giving some relation to the generalized extreme value distribution. A detailed proof can be found in [68].

Theorem 1.5.2. *[27, Theorem 4.1] Let X_1, X_2, \dots be a sequence of independent random variables with common distribution function and let*

$$M_n = \max\{X_1, \dots, X_n\}$$

so that for large n ,

$$\mathbb{P}(M_n \leq z) \approx G(z),$$

where

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$

for some $\mu, \sigma > 0$ and ξ . Then, for large enough u , the distribution function given by (4) is approximately

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi}$$

defined on $\{y : y > 0, (1 + \xi y/\tilde{\sigma}) > 0\}$ where,

$$\tilde{\sigma} = \sigma + \xi(u - \mu).$$

The set of points

$$\left\{ \left(u, \frac{1}{n_u} \sum_{i=1}^{n_u} (x_{(i)} - u) \right) : u < x_{\max} \right\},$$

where $x_{(1)}, \dots, x_{(n_u)}$ consist of the n_u observations that exceed u , and x_{\max} is the largest value is termed the **mean residual life plot** or the mean excess function. A threshold u is chosen so that this relationship is roughly linear since the mean excess is linear if and only if the parent function is a generalized Pareto distribution [Section 4.3.1][27]. A point M_N is called a maximum if $M_N > u$. Increasing the value of u gives a better approximation of the generalized Pareto distribution since the mean residual is more easily approximated linearly while decreasing the value of u provides a larger pool of data at the expense of a lower fit quality.

After an appropriate threshold is chosen the GDP is fit to the maxima by approximation by maximum likelihood estimation of the log-likelihood function given by,

$$\ell(\sigma, \xi) = -n \log \sigma - (1 + 1/\xi) \sum_{i=1}^n \log(1 + \xi y_i/\sigma)$$

where y_i are the n excesses of a threshold u and $(1 + \sigma^{-1}\xi y_i) > 0$ for $i = 1, \dots, n$; otherwise $\ell(\sigma, \xi) = -\infty$.

1.6 Statistical Methods for Clustering Large Data

Statistical learning has critically impacted our society, as seen in: national security in the form of fast-processing facial recognition software; in medical research providing a way of mapping the brain; even in weather modeling giving global views to climate variability and extreme weather events. The elements of statistical learning can be generally split into two categories: supervised and unsupervised algorithms.

Some of the most well-known statistical methods for analyzing data come in the form of supervised learning where the values of one or more outputs or response variables are modeled on a set of predictor variables. Predictions are based on a subset of previously recorded data called the training sample where the joint values of all predictor and response variables are known. The model created from the predictor variables provides a value of the response variable possibly with some associated error. This error is typically characterized by some loss function, for example the squared difference between the actual and predicted values, where the main goal is to minimize such loss.

In unsupervised learning the goal is to conjecture about the underlying probability distribution of a variable, call it X , without the help of a set of predictor variables providing some measurement for the degree-of-error of each observation. Since X now must represent all response variables under consideration the dimension of X is often much higher than in supervised learning. In low-dimensional problems ($X \in \mathbb{R}^{n \leq 3}$) there are a variety of numerical methods for estimating the probability density of X and representing it graphically; however, these methods fail in higher dimensions [90].

Moreover, in unsupervised learning there is no such direct measure of success as observed in the supervised case. This makes it difficult to prove the validity of results and opens the effectiveness of an unsupervised algorithm up to a matter of a opinion. Nevertheless in recent years unsupervised learning techniques have been commonly put into practice with favorable results.

In this section we discuss a standard, unsupervised clustering technique called **spectral clustering**. Cluster analysis has a variety of goals all relating to grouping or segmenting a collection of objects into subsets or "clusters". The objects in each cluster are more closely related to other objects within the same cluster than objects in different clusters. An object is often described as a set of measurements, or by a similarity to other objects. The latter is typically represented by some positive-definite similarity matrix containing vectors which represent each object in the system and their similarities to all other objects. Recall that in unsupervised learning this matrix can carry (possibly unnecessarily) high dimension. Spectral clustering consists of two parts, the first to minimize the dimension without much loss of information and the second to run a clustering algorithm on the system based on similarities between objects.

1.6.1 Mutual information and data compression

Fundamentally, similarity can be seen as a measure of how much information one random variable can tell us about the other. It is natural to extend this idea in a probabilistic way to consider a measure which takes into account the joint probability

distribution or conditional probabilities of the two random variables.

Let Z^1 and Z^2 be random variables with realizations z_i^1 and z_j^2 respectively for $i = 1, \dots, m$ and $j = 1, \dots, n$ then the **entropy** of Z^1 is given by,

$$H(Z^1) = - \sum_{i=1}^m p(z_i^1) \log p(z_i^1)$$

and the **joint entropy** is given by,

$$H(Z^1, Z^2) = - \sum_{i=1}^m \sum_{j=1}^n p(z_i^1, z_j^2) \log p(z_i^1, z_j^2)$$

where $p(z_i^1) = \mathbb{P}(Z^1 = z_i^1)$ and $p(z_i^1, z_j^2) = \mathbb{P}(Z^1 = z_i^1, Z^2 = z_j^2)$. In general, entropy is a function which attempts to assign a value to the unpredictability of a random variable. In a similar way joint entropy assigns a value to the multi-valued random variable. This measure allows us to extend a precise mathematical definition to the information in a system and assigns a value to the similarity of two random variables [69].

Definition 1.6.1. The **mutual information** between two random variables Z^1 and Z^2 is given as

$$I(Z^1, Z^2) = H(Z^1) + H(Z^2) - H(Z^1, Z^2).$$

where the mutual information between Z^1 and Z^2 is 0 if and only if Z^1 and Z^2 are statistically independent.

Remark 1.6.2. Mutual information allows us to assign a positive similarity between two time series without the requirement that their relationship be linear as in the case of correlation.

Since the distribution of a random variable in practice is apriori unknown, a method for calculating the mutual information between two continuous random variables is to consider the transformation into discrete time. Given the data processing inequality [7],

$$I(T(X), R(Y)) \leq I(T(X), Y) \leq I(X, Y)$$

with equality only when $T(\cdot)$ and $R(\cdot)$ are invertible, our goal is to find transformations (though not invertible by definition) which compress our random variables while preserving the maximum possible mutual information between each pair of random variables.

We seek values a_{\max}^ℓ where $\ell = 1, \dots, m$ and $v_\ell \in \mathbb{Z}$ is the value the compressed vector can take such that the mutual information between two continuous random variables $X^1 = (x_1^1, x_2^1 \dots x_l^1)$ and $X^2 = (x_1^2, x_2^2 \dots x_p^2)$ is maximized. Compression is given by

$$\text{If } a^\ell < x_k^{1,2} < a^{\ell+1} \text{ then } z_k^{1,2} = v_\ell$$

$$\text{If } x_k^{1,2} > a^m \text{ then } z_k^{1,2} = v_m$$

$$\text{If } x_k^{1,2} < a^1 \text{ then } z_k^{1,2} = v_1$$

Since *entropy* is maximized when the random variable is uniformly distributed, it is reasonable to choose starting points of gradient ascent a_0^ℓ at the $1/m$ quantile of their respective time series. The gradient value of mutual information (now as a function of the interval endpoint) is estimated as,

$$\Delta I_t = \frac{I(a_t) - I(a_{t-1})}{a_t - a_{t-1}}. \quad (5)$$

Here a_t is the left-hand (right-hand) value of the interval with fixed right-hand (left-hand) at the current time step, and our updated a_{t+1} is given by

$$a_{t+1} = a_t + \gamma \Delta I_t,$$

where γ is the specified step size multiplier. Gradient ascent is then performed on (5) by varying the endpoints. Mutual information as a measure of similarity has been used in other applications of clustering real world data [6, 55, 56].

Since our transformation for compression is noninvertible, there will be an error on the calculated mutual information for the compressed data and the true mutual information between two time series. The total error is estimated by the difference between the true and calculated joint entropy. This derivation can be found in A.2.1.

The similarity matrix of mutual information is positive definite and can be represented as an undirected graph where nodes correspond to a single random variable and an edge represents a strictly positive mutual information between two random variables. If the graph is disconnected, the connected component containing the random variable of interest is chosen for the remaining analysis. From now on, we will use the word *system* and *connected component* to represent the model of relationships between random variables and their graph-matrix representation respectively.

Remark 1.6.3. The choice and estimation of the similarity measurement and compression of the data are modifications on standard spectral clustering described in the next section.

Graph and Random Walk Representations of a System with Similarities

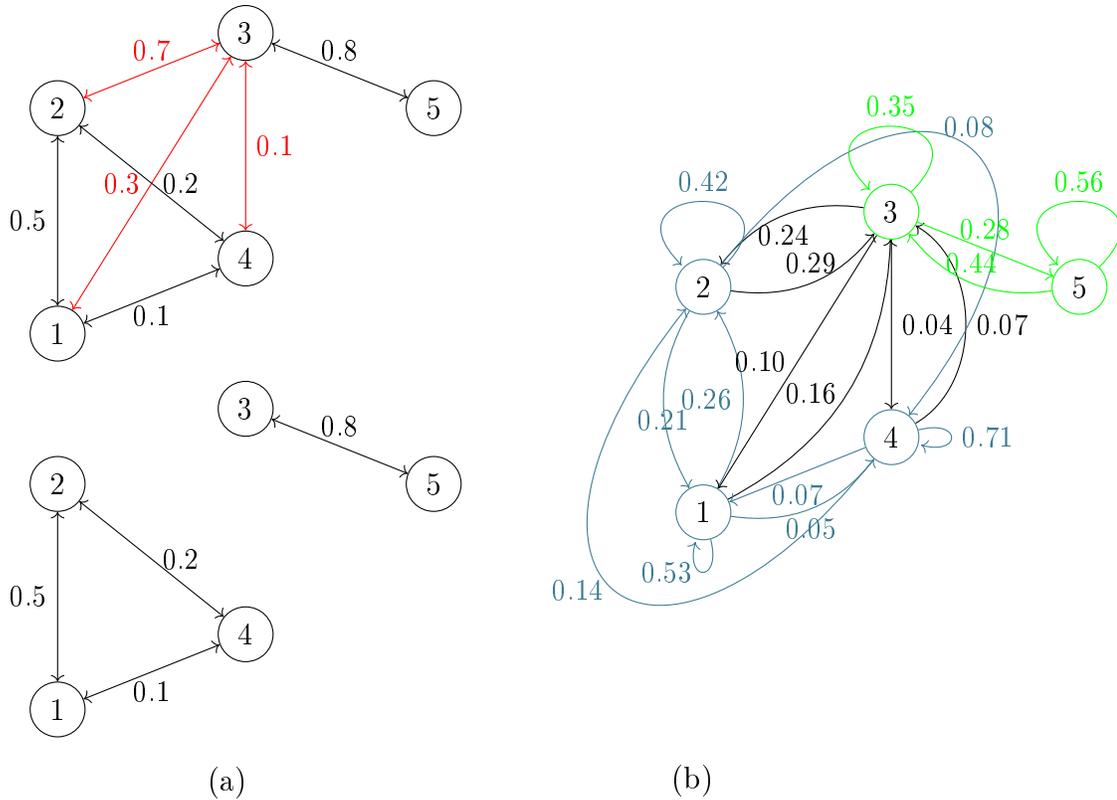


Figure 1.4: Undirected graph for five random variables where a line represents a strictly positive similarity (a) before clustering red lines indicate which edges will be removed and after clustering into two groups. (b) The corresponding random walk for the graph different colors indicate clusters black lines indicate removed edges.

1.6.2 Spectral clustering methods

The goal of spectral clustering is to remove the edges within (or *cut*) the graph to create clusters of disconnected nodes while still maintaining as much information as possible. The simplest way to approach this problem is to choose a partition of the graph which minimizes the sum of all the weights of the cuts; however, this is not reasonable in practice. Since an unbalanced graph (that is, one where the number

of edges for each node varies) can result in clusters containing one or a few nodes. Several solutions to this problem were proposed. Each is an optimization problem which minimizes the weights between each cluster while simultaneously maximizing the similarity within [47, 89, 31]. The normalized-cut problem [89] is given by,

$$\text{NCut}(C_1, \dots, C_k) = \frac{1}{2} \sum_{i=1}^k \frac{S(C_i, \bar{C}_i)}{\text{vol}(C_i)}$$

where C_1, \dots, C_k is a partition of the graph, \bar{C}_i is the compliment of C_i , $S(C_i, C_j) = \sum_{i \in C_i, j \in C_j} S_{i,j}$ is the similarity matrix and $\text{vol}(C_i)$ is the sum of the weights of all edges attached to vertices in C_i . A relaxation of this optimization problem yields the following spectral clustering algorithm.

Spectral decomposition and reducing the dimension of the system

Spectral decomposition of the system is performed by calculated the normalized Laplacian [52, 14.5.3] for the connected component given by

$$L = I - D^{-1/2} S D^{-1/2}$$

where I is the identity matrix, $S = S_{i,j}$ is the symmetric matrix of mutual information between random variables within the connected component and D is the diagonal matrix with entries $D_{i,i} = \sum_{j=1}^h S_{i,j}$. Every node in the connected component is represented by an h -dimensional vector in S where L takes this h -dimensional vector in S to an associated h -dimensional vector in \mathbb{R}^h .

In keeping with the theme, we will discuss the intuition behind spectral clustering from a probabilistic point of view. Let $P = D^{-1}S$ then P can be seen as a probability transition matrix for a random walk on the associated graph of similarities. The

largest eigenvalues of P indicate which eigenvectors carry the most information of the system. Since an eigenvalue λ of P with eigenvector v implies $1 - \lambda$ is an eigenvalue of $L = I - P$ ¹ with eigenvector v we may look at a reduction of the system by projecting each node into the J -dimensional subspace of eigenvectors of L corresponding to the J lowest eigenvalues. This provides a way of reducing the dimension of the system while maintaining as much information as possible. Clustering is then performed on the reduced system where the goal is to remove edges which carry the lowest weight and result in the highest weight within clusters. In the setting of a random walk, this amounts to partitioning the graph in places where there is a low probability of transitioning between clusters.

A special property of the normalized Laplacian is that the projections of each h -dimensional node correspond to the row vectors of the reduced $[h \times J]$ eigenvector matrix of L . These vector projections serve as inputs into the K -means algorithm.

K -means clustering

The K -means clustering algorithm [52, 14.3.6] seeks to minimize cost by performing the following until a steady state on the objective function is reached:

$$\min \sum_{K=1}^N \sum_{l=1}^M \|n(l) - C(K)\|_{\mathbb{R}^J}$$

where M is the number of nodes n in the cluster K , $C(K)$ is the centroid of cluster K and N is the number of clusters.

¹ This formulation of the Laplacian is used for simplification and does not change the results since $L = I - P$ is similar to $L = I - D^{-1/2}SD^{-1/2}$.

assignment step: Given an initial set of K random centroids, assign each observation to the cluster whose centroid has the least squared euclidean distance.

update step: Calculate the new means of the cluster to be the centroids of the observations in the new clusters.

K -means provides a simple and relatively easily computed algorithm for data clustering though it is not without limitations. Primarily these limitations exist because J -dimensional data cannot be visualized. Since K means only considers separations of clusters by hyper-planes it supplies a method of "flat" clustering.

By definition the K -means algorithm monotonically decreases the objective; however, this convergence is dependent upon the original choice of centroids. To overcome this difficulty, many analyses consider multiple runs of the K means algorithm where the *best* clustering can be loosely defined by the cluster with the lowest minimized distance.

In the case when data is dispersed for one cluster and tight for another the minimized distance may not provide the most accurate clustering solution. Determining the accuracy of resulting clusters is a wide topic of discussion in unsupervised learning. For the purposes of this paper we will consider a good cluster one which has (1) a low dispersion of points and (2) a large separation between centroids.

The Davies-Bouldin index [28], given by

$$\frac{1}{N} \sum_1^N \max \frac{MSE_j + MSE_i}{\|C(j) - C(i)\|}$$

where $i, j = 1, \dots, N$, N is the number of clusters and,

$$MSE_j = \frac{1}{M} \sum_{n(l) \in C(j)} d(n(l), C(j))^2$$

is often used to determine how separated a cluster is. Note that small values of the Davies-Bouldin index indicate larger separation of clusters.

Remark 1.6.4. If the data cannot be separated appropriately by a hyper-plane, for example in the case of complicated clusters such as the Donut and Ball, Horse Shoe or Spiral, more accurate clustering can be obtained by implementing the Kernel K -means algorithm. Kernel K -means considers functions ϕ on a Hilbert space where the distance $\|\phi(n(l)) - \phi(C(K))\|_{\mathbb{R}^J}$ is minimized. Implementing Kernel K -means may provide a more accurate cluster at the expense of computation efficiency.

The number of clusters K presents another issue since a large number of clusters will decrease the objective in an obvious way but give no usable information while small numbers make the objective unable to decrease appropriately. Typically K is chosen by running the algorithm several times for different values of K , taking the minimum value of the objective and comparing this across all K .

It is natural to ask why we perform spectral clustering rather than running K -means explicitly on the data. Projection onto the full space \mathbb{R}^h quite often result in clusters which are unnecessarily complicated and unable to be easily separated. Spectral decomposition takes clusters which are not separable by a hyperplane in \mathbb{R}^h into the lower dimensional space \mathbb{R}^J where they may be separated. This produces more reliable results with very little computational expense.

A Toy Example for Clustering Comparison

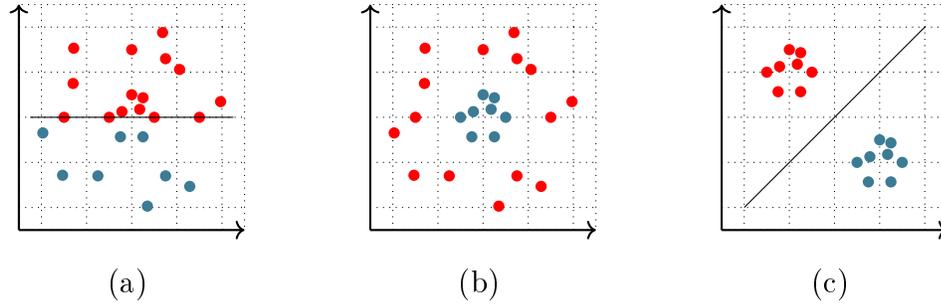


Figure 1.5: (a) K -means clustering result. (b) Spectral clustering result. (c) Space where K -means clustering is performed in the spectral clustering algorithm.

For completion of exposition, the adapted spectral clustering algorithm described is listed below.

- Compute the normalized Laplacian given by $L = I - D^{-1}S$ where I is the identity, S is the matrix of similarity and D is a diagonal matrix containing the sum of all the row values of S .
- Compute the k smallest eigenvalues of L and their corresponding eigenvectors by choosing a cut-off point.
- Form a matrix $U \in \mathbb{R}^{n \times k}$ with the eigenvectors as columns.
- Let $y_i \in \mathbb{R}^k$ be the row-vectors of U . Cluster the set of points y_i using the K -means algorithm.

Chapter 2

Growth Rates for Birkhoff Sums of Non-integrable Observables¹

2.1 Background

Let (X, T, μ) be an ergodic, measure-preserving transformation, μ a probability measure absolutely continuous with respect to Lebesgue and φ a non-integrable observable taken on the map. Since convergence for the Birkhoff sum cannot be established in this setting (see Chapter 1, Section 1.1), an interesting question to ask is about the growth rates of the Birkhoff sum. A useful result due to Aaronson [1, Proposition 2.3.1] states,

¹ **This chapter contains published work from** Carney, M., and Nicol, M. (2017) Dynamical Borel-Cantelli lemmas and rates of growth of Birkhoff sums of non-integrable observables on chaotic dynamical systems. *Nonlinearity*, 30(7), 2854-2870.

Proposition 2.1.1. *Suppose that $\varphi : X \rightarrow \mathbb{R}$ is a non-integrable measurable function. If $a(x)$ is increasing, $\lim_{x \rightarrow \infty} \frac{a(x)}{x} = 0$ and*

$$\int a(|\varphi(x)|)d\mu < \infty$$

then for μ a.e. x

$$\lim_{n \rightarrow \infty} \frac{a(|S_n|)}{n} = 0.$$

Despite the generality of its assumptions, the above gives close to optimal upper bounds on the $\limsup S_n$ in many dynamical settings, as demonstrated later. Since the Birkhoff sum is greater than or equal to the sum of its maximum values, it is natural to assume that optimal lower bounds on the $\liminf S_n$ can be obtained by counting the maximum contributions.

In [58] dynamical Borel Cantelli lemmas were used to give information on the behavior of the maxima $M_n := \max\{\varphi(x), \varphi(Tx), \varphi(T^2x), \dots, \varphi(T^n x)\}$ of a time series on a variety of chaotic dynamical systems (T, X, μ) with observables of the form $\varphi(x) = -\log d(x, q)$ and $\varphi(x) = d(x, q)^{-k}$ for a point $q \in X$, where $d(.,.)$ was a Riemannian metric on the space X , a Riemannian manifold. Recall from Chapter 1, Section 1.3, that observables of this type are motivated by applications in extreme value theory. For the integrable observable $\varphi(x) = -\log d(x, q)$ under relatively mixing conditions [58, Theorem 2.2] on the dynamical system a sequence of scaling constants $a(n)$ exists such that $\lim_{n \rightarrow \infty} \frac{M_n}{a(n)} = C > 0$ almost surely for some constant C , however, for observables of the form $\varphi(x) = d(x, q)^{-k}$, $k > 0$ there is no almost sure limit for $\frac{M_n}{a(n)}$ even if k is such that φ is integrable. The precise result for $\varphi(x) = d(x, q)^{-k}$ is stated in the following version of [58, Theorem 2.7],

Proposition 2.1.2. *Suppose that (T, X, μ) is a probability measure-preserving system with ergodic measure μ which is absolutely continuous with respect to Lebesgue measure m . Suppose for a point $q \in X$ there exists $\delta > 0$, $C > 0$ and $r_0 > 0$ such that for all $\epsilon < r < r_0$:*

$$|\mu(B(q, r + \epsilon)) - \mu(B(q, r))| \leq C\epsilon^\delta.$$

and $0 < \frac{d\mu}{dm}(q) < \infty$ where $B(q, r)$ denotes the ball of radius r centered at q . Moreover suppose that we have exponential decay of correlations in bounded variation norm (BV) versus L^1 (refer to Chapter 1, Section 1.1, and the Appendix A.1.2 for details) in the sense that there exists $C > 0$ and $0 < \theta < 1$ such that for all φ_1 of bounded variation and all $\varphi_2 \in L^1(m)$ we have:

$$\left| \int \varphi_1 \cdot \varphi_2 \circ T^j d\mu - \int \varphi_1 d\mu \int \varphi_2 d\mu \right| \leq C\theta^j \|\varphi_1\|_{BV} \|\varphi_2\|_{L^1(m)},$$

Then if $\varphi(x) = d(x, q)^{-k}$ for some $k > 0$ for any sequence $u(n) \rightarrow \infty$:

$$\mu\left(\limsup_{n \rightarrow \infty} \frac{M_n(x)}{u(n)} = 0\right) = 1, \text{ or } \mu\left(\liminf_{n \rightarrow \infty} \frac{M_n(x)}{u(n)} = \infty\right) = 1$$

The relation between Birkhoff sums and extreme values, such as the maxima, is investigated in the topic of trimmed Birkhoff sums [3, 62, 88]. In this approach the time series $\{\varphi(x), \varphi(Tx), \varphi(T^2x), \dots, \varphi(T^n x)\}$ is rearranged into increasing order $\{\varphi(T^{i_0}x) \leq \varphi(T^{i_1}x) \leq \varphi(T^{i_2}x) \dots \leq \varphi(T^{i_n}x)\} = \{M_0^n(x), M_1^n(x), \dots, M_n^n(x)\}$ so that $\varphi(T^{i_n}x) = M_n^n(x) = M_n(x)$. Almost sure limit theorems for trimmed sums consider two sequences of constants $a(n)$, $b(n)$ so that the scaled truncated sum $\frac{1}{a(n)} \sum_{j=0}^{n-b(n)} M_j^n$ satisfies a strong law of large numbers with [3, 88] containing precise limiting behavior and choice of constants $a(n)$ and $b(n)$ for certain dynamical

systems. The rate of growth of $\sum_{j=n-b(n)+1}^n M_j^n$ remains an open question; however, the trimmed sum limit has been shown to provide good estimates on the lower bound of the rate of growth of S_n particularly in the case of piecewise uniformly expanding interval maps [88, Theorem 1.8] where our techniques present less optimal bounds.

In this chapter we consider the observable $\varphi(x) = d(x, q)^{-k}$ over chaotic dynamical systems (T, X, μ) on probability spaces $\mu(X) = 1$ for ergodic, measure-preserving transformations and for values of k which ensure $\int \varphi d\mu = \infty$. Most of the results generalize to a wider class of functions of the form $\mu(\varphi > t) = \frac{L(t)}{t^\gamma}$ where $0 < \gamma < 1$ and $L(t)$ is a slowly varying function, as long as the sets $(\varphi > t)$ correspond to sets for which the dynamical Borel-Cantelli lemmas hold for large t . In a similar way we may generalize further by considering observables φ with a finite set of singularities $\{q_1, \dots, q_m\}$ such that for all t there exist constants $C_1, C_2, r > 0$ such that $0 < C_1 < \frac{\varphi(y)}{d(y, q_i)^{-k}} < C_2$ for all $y \in B(q_i, r)$ and with integrable negative part i.e. if $\varphi^- := \max\{0, -\varphi\}$ then $\int \varphi^- d\mu < \infty$. The case when φ^- is non-integrable is interesting but the techniques outlined are not immediately applicable in this setting. For interesting recent results on trimmed symmetric Birkhoff sums in the setting of infinite ergodic theory (when the underlying probability space has infinite measure) refer to [2, 64].

2.2 Main Results

We assume that (T, X, μ) is an ergodic dynamical system and X is a probability measure space and a Riemannian manifold with a Riemannian metric d . Let m denote

the Lebesgue measure on X and assume $\mu \ll m$. Let $B(q, r) := \{x : d(q, x) < r\}$ denote the ball of radius r about a point q with respect to the given metric d . The following results utilize dynamical Borel Cantelli properties, in particular the QSBC property, for shrinking balls (B_j) about a point $q \in X$. Please refer to Chapter 1, Section 1.2, for detailed information on these topics.

Examples of systems for which the QSBC property has been proved for balls nested at points q in phase space include Axiom A diffeomorphisms [21], uniformly partially hyperbolic systems preserving a volume measure with exponential decay of correlations [32], uniformly expanding C^2 maps of the interval [82], and Gibbs-Markov type maps of the interval [63] (All listed maps are defined precisely in the Appendix A.1.1). For intermittent type maps with an absolutely continuous invariant probability measure the work of Kim [63] and Gouëzel [45] gives a fairly complete picture: the Borel Cantelli property holds for nested balls except those based at the indifferent fixed point. Other results on non-uniformly expanding systems include one-dimensional maps modeled by Young towers with exponential decay of correlations [46], the general framework of [54] and other hyperbolic settings [43, 59, 74, 75]. See Appendix A.1.2 for more details on Young towers.

2.2.1 Non-integrable observations.

Let $\varphi(x) = d(x, q)^{-k}$ for some distinguished point $q \in X$ where $\dim(X) = D$ and $k > D$ and define $S_n = \sum_{i=0}^{n-1} \varphi \circ T^i(x)$.

Theorem 2.2.1. *Suppose that (T, X, μ) is an ergodic dynamical system where X*

is a Riemannian manifold with Riemannian metric d . Suppose $\dim(X) = D$. Let $\varphi(x) = d(x, q)^{-k}$ for some distinguished point q . Suppose there exist constants C_1, C_2 such that $0 < C_1 < \frac{d\mu}{dm}(q) < C_2$ and that the SBC property holds for nested balls about q .

If $k > D$ then for μ a.e. x and any $\epsilon > 0$

$$(a) \limsup_{n \rightarrow \infty} \frac{S_n}{n^{k/D} (\log n)^{k/D+\epsilon}} = 0$$

and for any $\epsilon > 0$

$$(b) \liminf_{n \rightarrow \infty} \frac{S_n}{n^{k/D-\epsilon}} = \infty$$

while

$$(c) S_n \geq n^{k/D} \log^{k/D} n \text{ infinitely often}$$

If moreover the QSBC property holds for nested balls about q then for any $\epsilon > 0$

$$(d) \liminf_{n \rightarrow \infty} \frac{S_n}{n^{k/D} (e^{-(\log n)^{1/2+\epsilon}})^{k/D}} = \infty$$

Proof. We assume first $k > D$. It is known from Aaronson [1, Proposition 2.3.1] that if $a(x)$ is increasing, $\lim_{x \rightarrow \infty} \frac{a(x)}{x} = 0$ and

$$\int a(\varphi(x)) d\mu < \infty$$

then for μ a.e. x

$$\lim_{n \rightarrow \infty} \frac{a(S_n)}{n} = 0$$

Our assumptions imply that $\mu(B(q, r)) \sim r^D$. In fact using spherical coordinates our assumption on the density implies that for any integral $f: X \rightarrow \mathbb{R}$ such

that $\int f d\mu = \int f h(x) dx = \int f(\theta_1, \dots, \theta_{D-1}, r) K(\theta_1, \dots, \theta_{D-1}) r^{D-1} h(r) dr d\theta_1 \dots d\theta_{D-1}$ where $0 < c_1 < K(\theta_1, \dots, \theta_{D-1}) < c_2$ for some constants c_1, c_2 .

By the Borel Cantelli lemma $\mu(T^n x \in B(q, \frac{1}{n^{1/D+\delta}}) \text{ i. o.}) = 0$ for any $\delta > 0$. Hence given $\delta > 0$ for μ a.e. $x \in X$ there exists a time $N(x)$ such that $T^i x \notin B(q, \frac{1}{n^{1/D+\delta}})$ for all $i > N(x)$. This implies that $\varphi \circ T^j \leq n^{k(1/D+\delta)}$ for all $j \geq N(x)$. Thus $S_n \leq C(x)n^{1+k(1/D+\delta)}$ for large n where $C(x)$ is a constant. Hence $\log(S_n) \leq c(x) \log(n)$ for some constant $c(x) > 0$. Choosing $a(x) = \frac{x^{D/k}}{\log(x)^{1+\eta}}$ for $\eta > 0$ then

$$a(S_n) = \frac{(S_n)^{D/k}}{\log(S_n)^{1+\eta}} \geq \frac{(S_n)^{D/k}}{[c(x) \log(n)]^{1+\eta}}$$

Hence for any $\epsilon > 0$

$$\limsup \frac{S_n}{n^{k/D} [\log(n)]^{k/D+\epsilon}} = 0$$

Assume now that the SBC property holds for nested balls about q . First note that if $r_n = (n)^{-1/D}$ then $T^n x \in B(q, r_n)$ i.o. Let $B_j := B(q, \frac{1}{j^{1/D}})$. From the SBC property $\sum_{j=1}^n 1_{B_j} \circ T^j(x) \sim \log(n)$.

If we define $n_l := \max\{0 < j \leq n\}$ such that $T^j x \in B(q, r_j)$ (the notation "l" in n_l suggests the "last time") then for μ a.e. $x \in X$, for any $M > 0$, $\lim_{n \rightarrow \infty} \frac{n_l}{n^{1-\delta}} > M$ for any $\delta > 0$. To see this, for a generic $x \in X$, $\lim_{n \rightarrow \infty} \frac{S_n}{\log n} = 1$. By definition of $n_l(x)$, $S_{n_l} = S_n$ and hence $\lim_{n \rightarrow \infty} \frac{S_{n_l}}{\log n} = 1$. As $\lim_{n \rightarrow \infty} \frac{S_{n_l}}{\log n_l} = 1$ we see $\lim_{n \rightarrow \infty} \frac{\log n_l}{\log n} = 1$, which implies the result.

Since $S_n > M_{n_l}$, $\liminf \frac{S_n}{n^{k/D-\epsilon}} = \infty$ for any $\epsilon > 0$.

Suppose now that we have a quantitative error estimate in the form of the QSBC

Illustration of Shrinking Targets



Figure 2.1: Iterations of shrinking balls (*shrinking targets*) of (a) radius $\frac{1}{\sqrt{n}}$ and (b) radius $\frac{1}{n}$. (a) The balls about point q shrink slow enough so that almost every point under the map will return to the shrinking ball infinitely often. (b) The balls shrink too quickly so that at a large enough time step almost every under the map will not return to the shrinking ball.

property,

$$\sum_{j=1}^n 1_{B_j} \circ T^j(x) = E_n + O(E_n^{1/2+\delta})$$

Remark 2.2.2. Since the general QSBC property does not provide much more optimal results a simpler version is used in the proofs of this chapter to avoid clutter of variables.

Then

$$S_n = E_n + O(E_n^{1/2+\delta})$$

$$S_{n_l} = E_{n_l} + O(E_{n_l}^{1/2+\delta})$$

By definition of n_l , $S_{n_l} = S_n$ and hence

$$E_n - E_{n_l} = O(E_n^{1/2+\delta})$$

We obtain

$$\log n - \log n_l = O(E_n^{1/2+\delta})$$

which implies that

$$n_l \geq ne^{-(\log n)^{\frac{1}{2}+\delta}}$$

for any $\delta > 0$.

$$\text{Hence } \liminf \frac{S_n}{n^{k/D} (e^{-(\log n)^{\frac{1}{2}+\epsilon}})^{k/D}} = \infty \text{ for any } \epsilon > 0.$$

The proofs of (a) and (c) in the case $k = D$ are unchanged, and estimates (b) and (d) are immediate consequences of the ergodic theorem. \square

Remark 2.2.3. The assumptions of Theorem 2.2.1 are satisfied by Anosov diffeomorphisms [17], uniformly expanding C^2 maps of the interval [82] and Gibbs-Markov type maps of the interval [63]. Kim also shows that for all $q \in (0, 1]$ in a class of intermittent maps of the unit interval preserving an absolutely continuous probability measure the conditions hold, except at the indifferent fixed point $x = 0$. Recent work of Tanja Schindler [88, Theorem 1.8] on trimmed Birkhoff sums has shown that for Gibbs-Markov maps the limit infimum estimate (d) can be improved to $\liminf \frac{S_n(\log \log n)^{k-1+\epsilon}}{n^k} = \infty$ for any $\epsilon > 0$.

2.2.2 Non-integrable observables on a class of intermittent type maps.

A simple model of intermittency, a form of Manneville-Pommeau map, is the class of maps T_α introduced by Liverani, Saussol and Vaienti in [71]

$$T_\alpha(x) = \begin{cases} x + 2^\alpha x^{1+\alpha}, & 0 \leq x \leq 1/2 \\ 2x - 1, & 1/2 \leq x \leq 1 \end{cases} \quad 0 \leq \alpha < 1 \quad (2.1)$$

The LSV Map

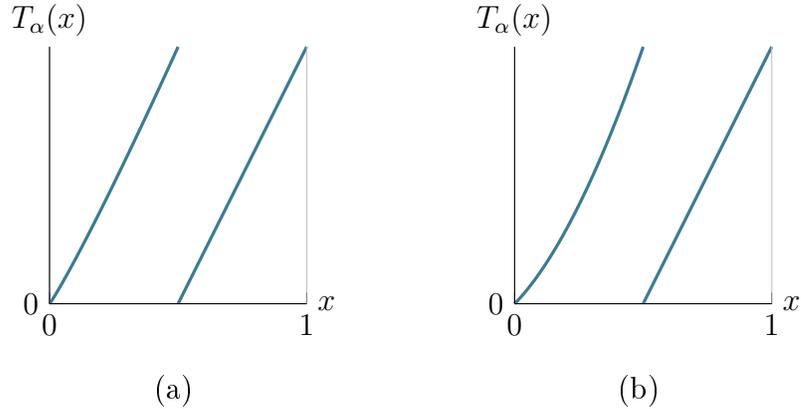


Figure 2.2: The Liverani-Saussol-Vaienti (LSV) map for (a) $\alpha = 0.2$ and (b) $\alpha = 0.9$.

The map T_α has a unique absolutely continuous probability measure μ_α if $0 \leq \alpha < 1$. We will only consider the case of a probability measure, rather than an infinite measure-preserving system. The density $h_\alpha(x)$ is Lipschitz and strictly positive on any interval of form $[a, 1]$, $a > 0$ but is unbounded at $x = 0$, where $h_\alpha(x) \sim x^{-\alpha}$, $\alpha > 0$.

Kim [63, Proposition 4.1] has shown that if $q \neq 0$ then any nested sequence of balls about q has the SBC property. In this section we improve this result to obtain

the quantitative (QSBC) property and establish the following.

Theorem 2.2.4. *Suppose $(T_\alpha, [0, 1], \mu_\alpha)$ is a Liverani-Saussol-Vaianti map with $0 \leq \alpha < 1$. Let $q \in [0, 1]$ and $\varphi(x) = d(x, q)^{-k}$ with $k \geq 1$. Define $S_n = \sum_{j=1}^n \varphi \circ T_\alpha^j$. Then if $q \neq 0$, for μ_α a.e. x and any $\epsilon > 0$*

$$\liminf_{n \rightarrow \infty} \frac{S_n}{n^k (e^{-(\log n)^{\frac{1}{2} + \epsilon}})^k} = \infty$$

and

$$\limsup_{n \rightarrow \infty} \frac{S_n}{n^k [\log(n)]^{k+\epsilon}} = 0$$

In particular

$$\lim_{n \rightarrow \infty} \frac{\log S_n}{\log n} = k$$

If $q = 0$ then for any $\epsilon > 0$

$$\liminf_{n \rightarrow \infty} \frac{S_n}{n^{k+\alpha-\epsilon}} = \infty$$

and

$$\limsup_{n \rightarrow \infty} \frac{S_n}{n^{k+\alpha+\epsilon}} = 0$$

In particular

$$\lim_{n \rightarrow \infty} \frac{\log S_n}{\log n} = k + \alpha$$

Remark 2.2.5. For generalized Manneville-Pommeau maps Dedecker, Gouëzel and Merlevede [29] proved that a strong law of large numbers with good error rates can be obtained for a large class of unbounded, but integrable observables.

Proof.

Case 1: $q \neq 0$

Proposition 2.2.6. [46] Let X be a compact interval and let \mathcal{P} be a countable partition of X into subintervals. Suppose that (T, X, μ, \mathcal{P}) is a Gibbs-Markov system. Let (B_n) be a sequence of intervals in X for which there exists $C > 0$ such that $\mu(B_j) \leq C\mu(B_i)$ for all $j \geq i \geq 0$. If $\sum_{n=0}^{\infty} \mu(B_n) = \infty$, then denoting $E_n = \sum_{j=1}^n \mu(B_j)$ for any $\epsilon > 0$,

$$\sum_{j=1}^n 1_{B_j} \circ T^j(x) = E_n + O(E_n^{1/2+\epsilon})$$

for μ a.e. $x \in X$.

A first return time Young Tower (see Appendix A.1.2) (F, ν, Δ) may be constructed for this class of intermittent maps with base $\Delta = [1/2, 1]$ [96]. Every point $q \neq 0$ has a unique representation in such a first return time Tower, in the sense that there is a unique mint t for each q such that $F^{-t}(q) \in \Delta$. Hence Proposition 2.2.6 shows that if $q \neq 0$ and (B_j) is a sequence of nested sequence of balls based about q then

$$\sum_{j=1}^n 1_{B_j} \circ T_{\alpha}^j(x) = E_n + O(E_n^{1/2+\epsilon})$$

for μ_{α} a.e. $x \in X$.

Hence by the proof of Theorem 2.2.1 for μ_{α} a.e. x

$$\liminf \frac{S_n}{(ne^{-(\log n)^{\frac{1}{2}+\epsilon}})^k} = \infty$$

for any $\epsilon > 0$, and as a consequence of Aaronson [1, Proposition 2.3.1] for any $\epsilon > 0$

$$\limsup \frac{S_n}{n^k [\log(n)]^{k+\epsilon}} = 0$$

Case 2: $q = 0$

Dynamics of the LSV Map

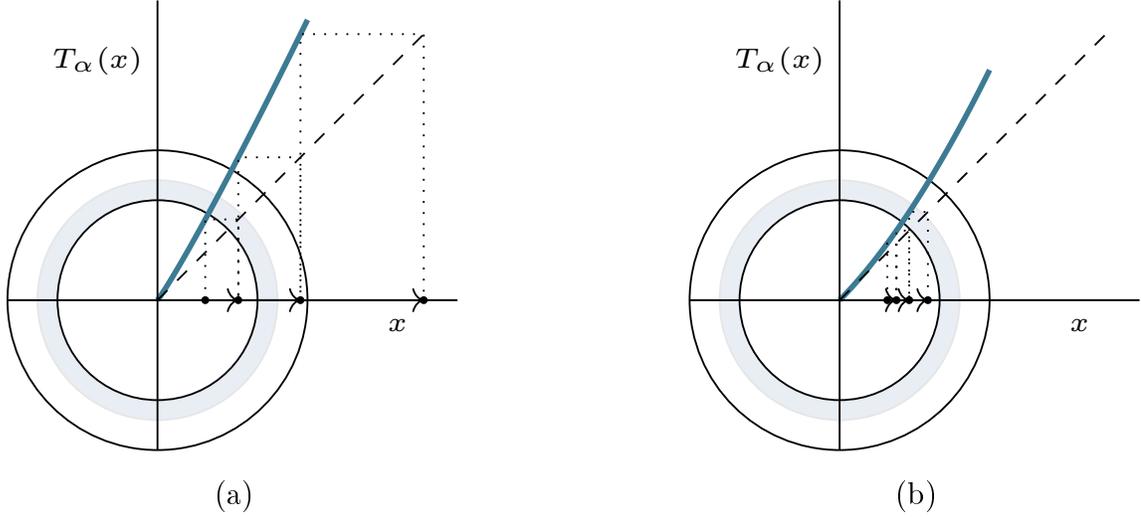


Figure 2.3: Shrinking target about $q = 0$ for the first branch of the LSV map with (a) $\alpha = 0.2$ and (b) $\alpha = 0.9$. Points indicate iterations of a point under the map. Note that it takes longer for a point to escape the shrinking target for larger α values.

Now we consider the case $q = 0$. For nested intervals based at $q = 0$ an interesting failure of the dynamical Borel-Cantelli lemma occurs, described in [63].

To understand this phenomenon let T_1 and T_2 be the two branches of the map T_α , with domains $[0, 1/2]$ and $[1/2, 1]$ respectively. Consider the sequence of balls $B_j = [0, \frac{1}{j^\gamma})$ for any $1 < \gamma \leq \frac{1}{1-\alpha}$. Kim notes that $\sum_n \mu_\alpha(B_j)$ diverges (due to $h_\alpha(x) \sim x^{-\alpha}$) while $\sum_n m(B_j) < \infty$. Note that $T_1^{-1}(B_j) \subset B_j$. Hence the only way that $T_\alpha^j(x)$ can enter B_j for infinitely many j is that $T_\alpha^{j-1}(x) \in T_2^{-1}(B_j)$ for infinitely many j . However the density $h_\alpha(x)$ is strictly positive and Lipschitz on any interval $[a, 1]$ for $a > 0$ and so $\sum_j \mu_\alpha(T_2^{-1}(B_j)) \sim \sum_j m(T_2^{-1}(B_j)) < \infty$ and the sequence (B_j) is not Borel-Cantelli.

We now consider the case of $q = 0$ and $\varphi(x) = d(x, 0)^{-k}$. In this setting using

Aaronson [1, Proposition 2.3.1] we solve $\int a(\varphi) \frac{1}{x^\alpha} dx < \infty$ which gives an upper bound roughly of form $\limsup \frac{S_n}{n^{k/(1-\alpha)}} = 0$, which is not optimal (being too large as we will see).

To get a better estimate we will consider the dynamics near the indifferent fixed point. The following local analysis of a large class of Manneville-Pommeau maps (of which the Liverani-Saussol-Vaianti map is a subclass) is taken from [97].

Fix $\epsilon_0 > 0$, let $x_0 \in (0, \epsilon_0]$ and define the sequence x_n by $x_{n-1} = T_\alpha x_n$. Young shows that $x_n \sim \frac{1}{n^\beta}$ where $\beta = \frac{1}{\alpha}$. In fact there is a uniform bound on the number of intervals $[\frac{1}{(m+1)^\beta}, \frac{1}{m^\beta}]$ that meet each $[x_{n+1}, x_n]$ and vice-versa.

This implies that if $x = \frac{1}{2} + \frac{1}{2m^\gamma}$ then $T_\alpha x = \frac{1}{m^\gamma}$. Writing $\frac{1}{m^\gamma} = x_n$ for some sequence as described above we have $\frac{1}{m^\gamma} = \frac{1}{n^\beta}$, hence it takes $n \sim m^{\gamma/\beta} = m^{\gamma\alpha}$ iterates j for $T_\alpha^{j+1}x$ to escape the region $[0, \epsilon_0]$ i.e. $T_\alpha^{j+1}x < \epsilon_0$ for $j < m^{\gamma\alpha}$. Note that $\sum_{j=1}^n \varphi(x_j) \geq \sum_{j=1}^n j^{k\beta}$ as $x_j \sim \frac{1}{j^\beta}$ and hence $S_n \geq n^{k\beta+1}$. Hence if

$$x = \frac{1}{2} + \frac{1}{2m^\gamma}$$

then

$$\sum_{j=1}^n \varphi \circ T_\alpha^j x \geq m^{(\gamma/\beta)(k\beta+1)} = m^{\gamma(k+\alpha)} \quad (\diamond)$$

This gives a lower bound on $\liminf S_n$ since if we define $n_l(x) = \max\{1 \leq j \leq n\}$ such that $T_\alpha^j(x) \in [1/2, 1/2 + \frac{1}{j}]$ then for any $\epsilon > 0$, $\liminf \frac{n_l}{n^{1-\epsilon}} \geq 1$ by the arguments of the previous section (we use the weaker SBC estimate as the stronger QSBC estimate does not help in this argument). Furthermore once $T_\alpha^n x$ enters $[1/2, 1/2 + \frac{1}{n}]$ it spends $\sim n^\alpha$ iterates in the region $(0, \epsilon_0)$ whence $S_{n+n^\alpha} \geq n^{k+\alpha-\epsilon}$. As $\alpha < 1$ this implies that $\liminf \frac{S_n}{n^{k+\alpha-\epsilon}} = \infty$ for any $\epsilon > 0$.

We will now show $\limsup \frac{S_n}{n^{k+\alpha+\epsilon}} = 0$ for any ϵ , hence $\lim_{n \rightarrow \infty} \frac{\log S_n}{\log n} = k + \alpha$.

We first sketch our argument. Let $0 < \eta < 1$ and $q \in (0, 1]$. Then $\sum_{j=1}^n 1_{B(q, \frac{1}{j^\eta})} \circ T_\alpha^j(x) \sim n^{1-\eta}$ for μ_α a.e. x . Note that if $\delta > 0$ then by Borel-Cantelli μ_α a.e. $x \in X$ has the property that $T_\alpha^n x \in B(q, (n \log^{1+\delta} n)^{-1})$ for only finitely many n . Asymptotically almost every x has the property that $T_\alpha^j x \in B(q, \frac{1}{j^\eta})$ for $\sim n^{1-\eta}$ iterates j in the interval $1 \leq j \leq n$, after a certain $L(x)$, i.e. for $j \geq L(x)$, the maximum value that $\varphi \circ T_\alpha^{j+1} x$ attains if $T_\alpha^j(x)$ enters $B(q, \frac{1}{n^\eta})$ is $n^k \log^{k(1+\delta)} n$. We break up S_n for large n into the times j that $T_\alpha^j(x)$ enters $B(q, \frac{1}{n^\eta})$, roughly $n^{1-\eta}$ times where the value $\varphi \circ T_\alpha^{j+1}(x)$ is bounded by $n^k \log^{k(1+\delta)} n$ which thus contributes at most $n^{1-\eta} n^{k+\alpha} \log^{k(1+\delta)} n$ to S_n and the times j that $T_\alpha^j(x)$ enters $B^c(q, \frac{1}{n^\eta})$, which contributes at most $n \cdot n^{\eta(k+\alpha)} = n^{1+\eta(k+\alpha)}$ to the sum S_n (using the estimate of line (\diamond)). Incorporating the log term into the exponent, by choosing $\eta = \frac{k+\alpha}{k+\alpha+1}$ we obtain $\limsup S_n \leq n^{k+\frac{1}{k+1}+\alpha}$.

We will iterate this procedure. Choose $1 > \eta_1 > \eta_2 > \dots \eta_m > 0$ and for simplicity of notation let $B_{\eta_i} = B(q, \frac{1}{n^{\eta_i}})$.

The contribution of the iterates j that enter B_{η_1} we bound by the product of the maximum value they may attain, namely the value $n^{k+\alpha} \log^{k(1+\delta)} n$ and the number of times the point enters this sequence of sets $n^{1-\eta_1}$ to arrive at $n^{k+\alpha+1-\eta_1}$ (incorporating the log term into the exponent). This accounts for those iterates that enter $B_{\eta_1} \subset B_{\eta_2}$ and we bound the contribution of those that enter $B_{\eta_2} \setminus B_{\eta_1}$ by $n^{1-\eta_2} \cdot n^{\eta_1(k+\alpha)} = n^{1-\eta_2+\eta_1(k+\alpha)}$. We bound the contribution of those that enter $B_{\eta_3} \setminus B_{\eta_2}$ by $n^{1-\eta_3} n^{\eta_2(k+\alpha)} = n^{1-\eta_3+\eta_2(k+\alpha)}$. Continuing in this way we have a sum of contributions of form $n^{1-\eta_{j+1}+\eta_j(k+\alpha)}$ terminating with the last contribution, those

Iterations of Shrinking Balls for Optimal Upper Bound Estimation

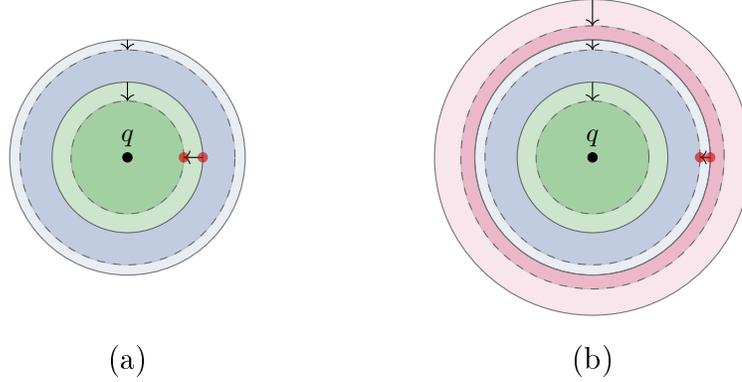


Figure 2.4: (a) Shrinking balls for two nested sets of radius $\frac{1}{n^{\eta_1}}$ (blue) and $\frac{1}{n \log^{1+\delta} n}$ (green). Orbits of points may enter the blue sequence infinitely many times while only enter the green sequence finitely many times. Red dots indicate the maximum of the orbit after the point leaves green but continues to remain in blue. (b) Pink indicates a shrinking set of radius $\frac{1}{n^{\eta_2}}$ where maximum values of the orbit after the point leaves blue but remains in pink are marked by red dots.

iterates j that lie in $B_{\eta_m}^c$ whose contribution we bound by $n^{\eta_m(k+\alpha)} \cdot n = n^{1+\eta_m(k+\alpha)}$.

If $k \geq 1$, choosing $\epsilon = \frac{1}{(k+\alpha)^m}$ and $\eta_i = 1 - (k+\alpha)^{i-1}\epsilon$ for $i = 1, \dots, m$ the leading term is $n^{k+\alpha+\epsilon}$ corresponding to $n^{k+\alpha+1-\eta_1}$, thus $\limsup S_n \leq mn^{k+\alpha+\frac{1}{(k+\alpha)^m}}$ which implies the result as m was arbitrary.

□

2.2.3 Dynamical systems with L^p , ($p > 1$), densities and exponential decay of correlations.

In this section, we consider dynamical systems with exponential decay of correlations, which possess absolutely continuous invariant measures (with respect to Lebesgue

measure m) with densities $\frac{d\mu}{dm}$ in $L^p(m)$, $p > 1$.

Suppose (T, X, μ) is an ergodic measure-preserving map of a probability space X which is a Riemannian manifold with Riemannian metric d . We assume:

(A) For all Lipschitz functions φ, ψ on X we have exponential decay of correlations in the sense that there exist constants $C, 0 < \theta < 1$ (independent of φ, ψ) such that

$$|E(\varphi \psi \circ T^k) - E(\varphi)E(\psi)| < C\theta^k \|\varphi\|_{\text{Lip}} \|\psi\|_{\text{Lip}}.$$

(B) There exist $r_0 > 0, 0 < \delta < 1$ such that for $C > 0$, all $q \in X$ and all $0 < \epsilon < r \leq r_0$

$$\mu\{x : r < d(x, q) < r + \epsilon\} < C\epsilon^\delta.$$

Under assumptions (A) and (B) Haydn, Nicol, Persson and Vaienti [54] showed:

Proposition 2.2.7. *Assume (T, X, μ) satisfies assumptions (A) and (B). Suppose $\mu(B_i) \geq C \frac{\log^\beta i}{i}$ for some $\beta > 0$, then if $E_n = \sum_{j=1}^n \mu(B_j)$ for μ a.e. $x \in X$.*

$$\sum_{j=1}^n 1_{B_j} \circ T^j(x) = E_n + O(E_n^{1/2+\epsilon})$$

for any $\epsilon > 0$.

In fact the density assumption $h := \frac{d\mu}{dm} \in L^p(m)$, $p > 1$ implies assumption (B).

Lemma 2.2.8. *Suppose m is Lebesgue measure on a D -dimensional manifold X and $h := \frac{d\mu}{dm} \in L^p(m)$, $p > 1$. Then for all $0 < r < r_0$*

$$\mu\{x : r < d(x, q) < r + \epsilon\} < \epsilon^\delta$$

for some $\delta > 0$

Proof. Let p be the conjugate of q , so that $\frac{1}{q} + \frac{1}{p} = 1$. Then $\int_{B_{r+\epsilon}(q)/B_r(q)} d\mu = \int_{B_{r+\epsilon}(q)/B_r(q)} h dx \leq \|h\|_q m(x : r < d(x, q) < r + \epsilon)^{\frac{1}{p}}$ which implies the result. \square

Remark 2.2.9. Any exponentially mixing volume preserving system satisfies (A) and (B), for example Sinai dispersing billiard maps with finite and infinite horizon [97, 17]. Furthermore for a volume preserving dynamical system the density $h(x) = \frac{d\mu}{dm}$ of the invariant measure is bounded above and is strictly positive. We consider the consequences of this in the next theorem.

Theorem 2.2.10. *Suppose a dynamical system (T, X, μ) satisfies (A) and $q \in X$ has density $h = \frac{d\mu}{dm}$ satisfying $0 < C_1 < h(q) < C_2$ for some constants C_1, C_2 . Suppose also $\dim(X) = D$. Then if $\varphi(x) = d(x, q)^{-k}$, $k > D$,*

$$\limsup_{n \rightarrow \infty} \frac{S_n}{n^{k/D} [\log(n)]^{\frac{k}{D} + \epsilon}} = 0$$

and

$$\liminf_{n \rightarrow \infty} \frac{S_n}{n^{k/D} (e^{-[\log(n)]^{\frac{1}{2} + \epsilon}})^{\frac{k}{D}}} = \infty$$

for μ a.e. x and any $\epsilon > 0$.

Remark 2.2.11. By ergodicity in the case $k = D$

$$\liminf_{n \rightarrow \infty} \frac{S_n}{n} = \infty$$

Proof. Note that if $B(q, r)$ is a ball of small radius $r > 0$ nested at q then $\mu(B(q, r)) \sim Cr^D$. Suppose $k > D$. Let $\varphi(x) = d(x, q)^{-k}$ and $a(x) = |x|^{\frac{D}{k}} / (|\log|x||)^{1+\eta}$. Then

$\int a(\varphi(x))dx < \infty$. If we define $S_n = \sum_{j=1}^n \varphi \circ T^j$, then by [1, Proposition 2.3.1]

$$\frac{a(S_n)}{n} \rightarrow 0$$

for μ a.e. $x \in X$. Hence for any $\epsilon > 0$, for μ a.e. $x \in X$

$$\limsup \frac{S_n}{n^{k/D} [\log(n)]^{\frac{k}{D} + \epsilon}} = 0$$

To obtain a limit infimum estimate we modify our previous argument. Let B_j be balls of μ (hence m) measure $\sim \frac{\log^\beta n}{n}$ nested about q . Let $E_n := \sum_{j=1}^n \mu(B_j)$

Define $n_l := \max\{0 \leq j \leq n\}$ such that $T^j(x) \in B_j$ as before we have

$$\begin{aligned} \sum_{j=1}^n 1_{B_j} \circ T^j &= E_n + O(E_n^{1/2+\delta}) \\ \sum_{j=1}^{n_l} 1_{B_j} \circ T^j &= E_{n_l} + O(E_{n_l}^{1/2+\delta}) \end{aligned}$$

By definition of n_l , $\sum_{j=1}^{n_l} 1_{B_j} \circ T^j = \sum_{j=1}^n 1_{B_j} \circ T^j$ and hence

$$E_n - E_{n_l} = O(E_n^{1/2+\delta})$$

We obtain

$$\log^{1+\beta} n - \log^{1+\beta} n_l = O(\log^{1/2+\gamma}(n))$$

where $\gamma = \delta + \frac{\beta}{2}$. As $x - y \leq x^{1+\beta} - y^{1+\beta}$ for large y and $x > y$ we see that

$$n_l \geq n e^{-(\log n)^{\frac{1}{2} + \epsilon}}$$

for any $\epsilon > 0$. Note that balls of radius r based at q satisfy $\mu(B(q, r)) \sim Cr^D$, and so we are able to bound S_n below by $M_{n_l} \geq (n e^{-[\log(n)]^{\frac{1}{2} + \epsilon}})^{\frac{k}{D}}$. Hence

$$\liminf \frac{S_n}{(n e^{-[\log(n)]^{\frac{1}{2} + \epsilon}})^{\frac{k}{D}}} = \infty$$

for any $\epsilon > 0$. □

The following recent result [86, Corollary B] extends a class of interval maps to the L^p case.

Proposition 2.2.12. *Let T be a non-degenerate smooth interval map having an exponentially mixing absolutely continuous invariant probability measure μ . Then there is $p > 1$ such that the density h of μ with respect to Lebesgue measure m is in $L^p(m)$. Moreover, μ can be obtained through a Young tower with an exponential tail estimate.*

For maps which satisfy these properties if the invariant density at q is $h(x) \sim Cd(q, x)^{-\alpha}$, $\alpha > 0$, then the following holds,

Theorem 2.2.13. *Suppose a dynamical system (T, X, μ) satisfies (A) and $q \in X$ has density satisfying $h(x) \sim Cd(q, x)^{-\alpha}$, $\alpha > 0$ for x near q . Suppose also $\dim(X) = D$. Then if $\varphi(x) = d(x, q)^{-k}$, $k \geq D - \alpha$,*

$$\limsup_{n \rightarrow \infty} \frac{S_n}{n^{k/(D-\alpha)} [\log(n)]^{k+\epsilon}} = 0$$

and

$$\liminf_{n \rightarrow \infty} \frac{S_n}{n^{k/(D-\alpha)} (e^{-[\log(n)]^{\frac{1}{2}+\epsilon}})^{\frac{k}{D-\alpha}}} = \infty$$

for μ a.e. x any $\epsilon > 0$. Hence

$$\lim_{n \rightarrow \infty} \frac{\log S_n}{\log n} = \frac{k}{D - \alpha}$$

Proof. The proof is an obvious modification of the proof of the previous theorem.

Let $\tilde{D} = D - \alpha$ and define $a(x) = \frac{|x|^{\tilde{D}/k}}{(\log |x|)^{1+n}}$. Then $\int a(\varphi(x)) dx < \infty$ and by [1, Proposition 2.3.1] $\frac{a(S_n)}{n} \rightarrow 0$ and hence

$$\limsup_{n \rightarrow \infty} \frac{S_n}{n^{k/(\tilde{D})} [\log(n)]^{k+\epsilon}} = 0$$

We now obtain our limit infimum estimate.

Let B_j be balls of μ measure $\sim \frac{\log^\beta n}{n}$ nested about q . Define $n_l := \max\{0 \leq j \leq n\}$ where $T^j(x) \in B_j$ as before we have

$$\sum_{j=1}^n 1_{B_j} \circ T^j = E_n + O(E_n^{1/2+\delta})$$

$$\sum_{j=1}^{n_l} 1_{B_j} \circ T^j = E_{n_l} + O(E_{n_l}^{1/2+\delta})$$

and hence

$$E_n - E_{n_l} = O(E_n^{1/2+\delta})$$

We have

$$\log^{1+\beta} n - \log^{1+\beta} n_l = O(\log^{1/2+\gamma}(n))$$

where $\gamma = \delta + \frac{\beta}{2}$. As $x - y \leq x^{1+\beta} - y^{1+\beta}$ for large y large and $x > y$ we see that as in the previous theorem

$$n_l \geq ne^{-(\log n)^{\frac{1}{2}+\epsilon}}$$

for any $\epsilon > 0$. Note that balls of radius r based at q satisfy $\mu(B(q, r)) \sim Cr^{\tilde{D}}$ we see that $S_n \geq M_{n_l}$ implies

$$\liminf \frac{S_n}{(ne^{-[\log(n)]^{\frac{1}{2}+\epsilon}})^{\frac{k}{D-\alpha}}} = \infty$$

for any $\epsilon > 0$. □

Corollary 2.2.14. *Suppose $T(x) = 4x(1-x)$ is a unimodal map of the interval $[0, 1]$. Let $\varphi(x) = d(x, q)^{-k}$, then if $q = 0$ or $q = 1$*

$$\lim_{n \rightarrow \infty} \frac{\log(S_n)}{\log n} = 2k$$

while if $q \in (0, 1)$

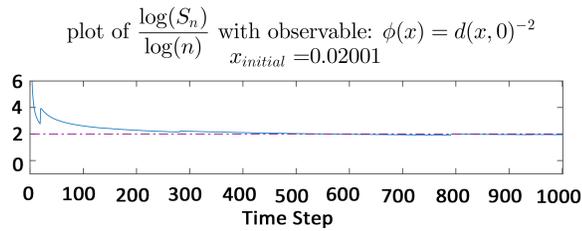
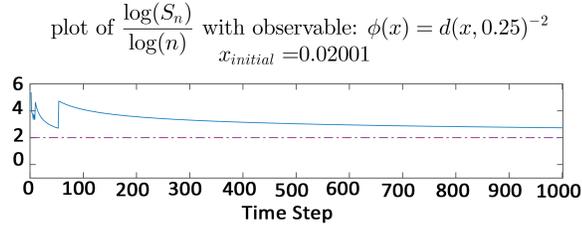
$$\lim_{n \rightarrow \infty} \frac{\log(S_n)}{\log n} = k$$

Proof. This map has invariant density $h(x) = \frac{1}{\sqrt{\pi x(1-x)}}$. First note that the unimodal map has density $h(x) \sim \frac{1}{\sqrt{x}}$ for $q = 0$ and $q = 1$ which implies the result. \square

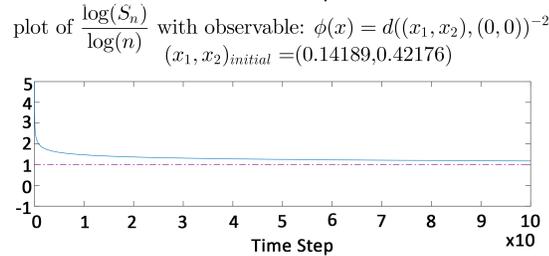
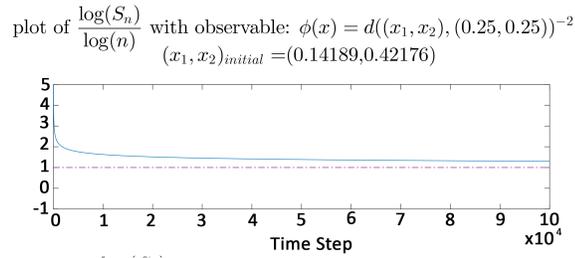
2.3 Numerical Results

Numerical simulations to illustrate Birkhoff sum growth rates under the Theorems described in this chapter are given for a few example dynamical maps.

Doubling Map and Hyperbolic Toral Automorphism. [Theorem 2.2.1]



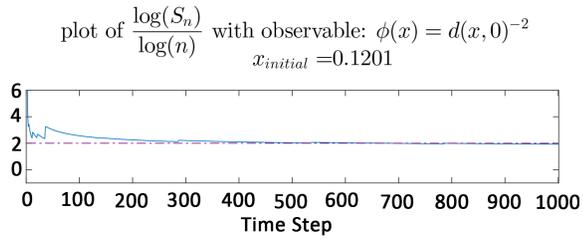
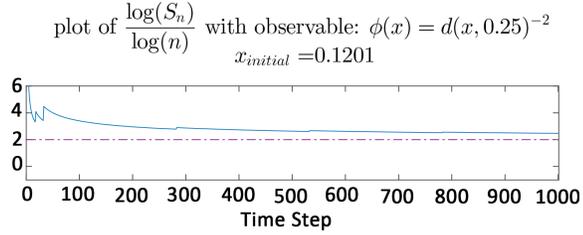
(a)



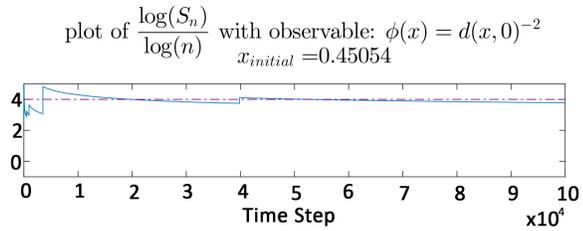
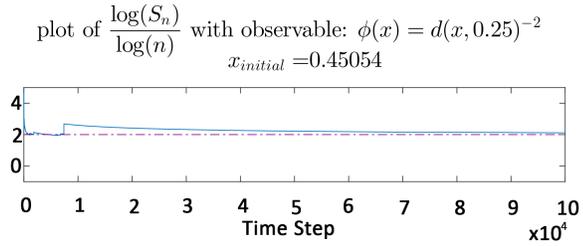
(b)

Numerical results for the Birkhoff sum with fixed point $q = 0.25$ and $q = 0$ on (a) the doubling map and (b) an Anosov diffeomorphism. Predicted value of the convergence is marked in a dotted line.

Tent Map. [Theorem 2.2.10] and Unimodal Map. [Corollary 2.2.14]



(a)

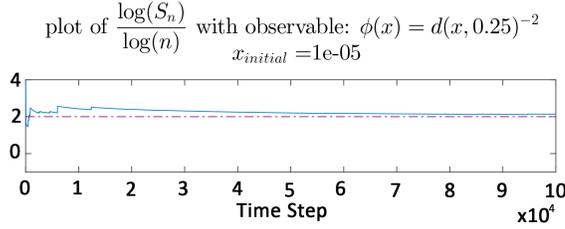


(b)

Figure 2.6: Numerical results for the Birkhoff sum with fixed point $q = 0.25$ and $q = 0$ on (a) the tent map and (b) the unimodal map. Predicted value of the convergence is marked with a dotted line.

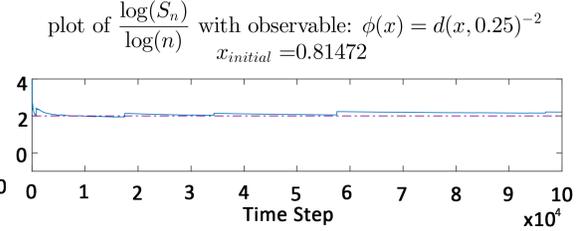
Liverani-Saussol-Vaianti Map. [Theorem 2.2.4]

INTERMITTENT MAP $\alpha = 0.5$



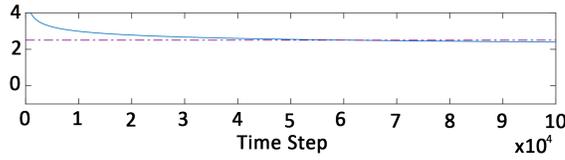
(a)

INTERMITTENT MAP $\alpha = 0.5$

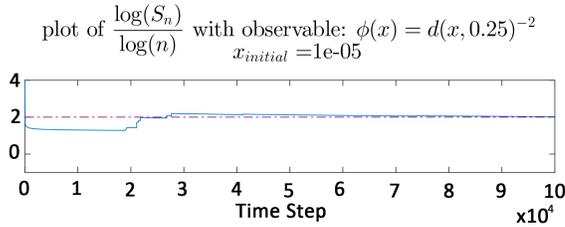


(b)

plot of $\frac{\log(S_n)}{\log(n)}$ with observable: $\phi(x) = d(x, 0)^{-2}$
 $x_{initial} = 1e-05$

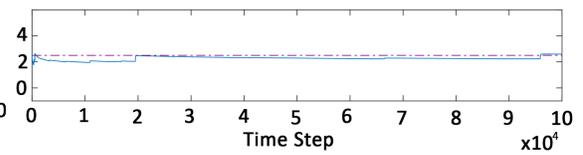


INTERMITTENT MAP $\alpha = 0.9$

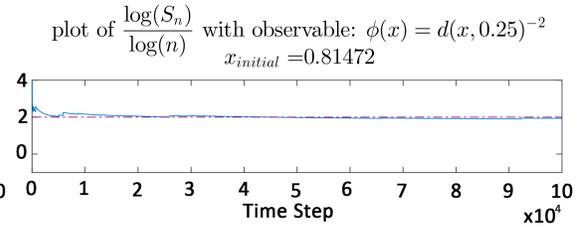


(c)

plot of $\frac{\log(S_n)}{\log(n)}$ with observable: $\phi(x) = d(x, 0)^{-2}$
 $x_{initial} = 0.81472$



INTERMITTENT MAP $\alpha = 0.9$



(d)

Numerical results for the Birkhoff sum with fixed point $q = 0.25$ and $q = 0$ on the LSV map. (a) $\alpha = 0.5$ and (c) $\alpha = 0.9$ with initial value near zero and (b) $\alpha = 0.5$ and (d) $\alpha = 0.9$ with initial value near 1. Predicted value of the convergence is marked with a dotted line.

Chapter 3

REPP Distribution for 2-D

Hyperbolic Systems with

Singularities¹

3.1 Background

Much of the work done in this chapter was motivated by a desire to understand mathematical billiards with chaotic behavior. What has come to be known as the *dispersing billiard model* was introduced by Yakov Sinai (1970) [91] and has been widely studied [11, 12, 96, 97, 23] for its physical relevance. The dispersing billiard

¹ **This chapter contains published work from** Carney, M., Nicol, M., and Zhang, H.K. (2017) Compound Poisson Law for Hitting Times to Periodic Orbits in Two-Dimensional Hyperbolic Systems with Singularities. *Journal of Statistical Physics*, 169(4), 804-823.

It is common to reduce this system to its corresponding billiard map $T : \partial Q \rightarrow \partial Q$, derive the statistical properties in this setting and then deduce the properties of the flow. The billiard map only considers points of collisions of the particle so that if r is a one-dimensional coordinatization of Γ given by arc-length and $n(r)$ the outward normal then for each $r \in \Gamma$ the tangent space at r consists of two unit vectors $(n(r), v) \geq 0$ where v is identified with an angle $\theta \in [-\pi/2, \pi/2]$. Finally, the phase space $M := \partial Q = \Gamma \times [-\pi/2, \pi/2]$ and consists of points (r, θ) . $T : M \rightarrow M$ is known as the Poincaré map that gives the position $T(r, \theta) = (r_1, \theta_2)$ after a point flows under B_t and collides with Γ .

Reducing the Billiard Flow to the Billiard Map

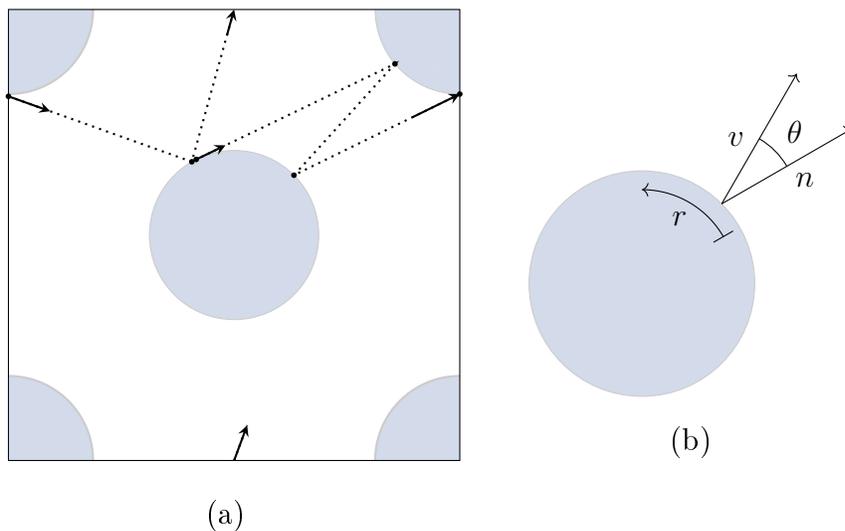


Figure 3.2: (a) Billiard flow on the torus with circular barriers. Collisions are represented by points. (b) Illustration of the reduction to the billiard map for a single collision point.

The billiard map, T , was proved to be ergodic [91] and preserves a measure $d\mu = c_M \cos \theta dr d\theta$ which is equivalent to two-dimensional Lebesgue measure $dm = dr, d\theta$ with density $\rho(x)$ where $x = (r, \theta)$. Exponential decay of correlations for

Hölder continuous observables was shown to hold by Young (1998) [96] and again by Chernov and Dolgopyat (2006) [22]. Bunimovich and Sinai (1981) [10] proved that T satisfies the central limit theorem. Finally, a critical lemma known as the **growth lemma** was established for unstable curves generated by T [16].

Background

Informally, a map T is said to be hyperbolic if the phase space can be separated into stable and unstable manifolds where expansion happens along the unstable direction in the tangent space under iterations of the map (similarly, contraction along the stable direction) [60]. The space M can be broken up into set of these stable and unstable curves where a curve is a branch of the manifold itself. This is referred to as a *foliation* of the space. The unstable curve W^u is defined in the billiard map by the trace of a convex front on M . W^u is characterized by an increase in the (r, θ) coordinates and is invariant and expanding under T . In particular, $\exists \Lambda > 1$ such that $d(Tx, Ty) \geq \Lambda d(x, y)$, $\forall W^u$ and $\forall x, y \in W^u$ where d is some distance metric. Moreover, it has been shown that "most" unstable curves are "long" where the growth lemma provides a quantifiable estimate on this.

In a general way we consider (X, T, μ) a two-dimensional hyperbolic system with singularities and an observable $\varphi(x) = -\log d(x, p)$ where d is a metric defined in terms of the stable and unstable foliation. Let $X_n = \varphi \circ T^n$ for $n \geq 0$ and consider the stochastic process generated by the time series $\{X_n\}$. Let the maxima of the time series $M_n := \max\{X_0, \dots, X_{n-1}\}$ and u_n be defined by $\lim_{n \rightarrow \infty} n\mathbb{P}(X_0 > u_n(\tau)) = \tau$.

Unstable Curves for the Billiard Map

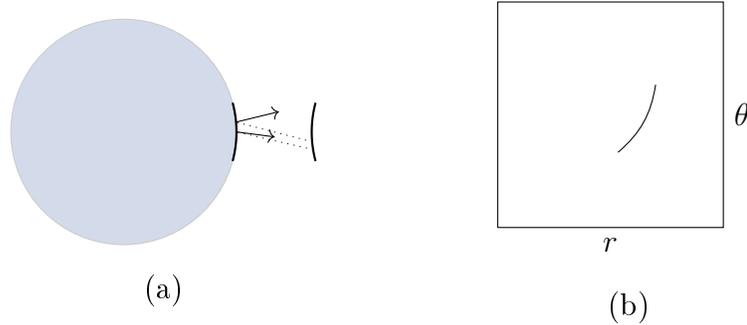


Figure 3.3: (a) Trace of the convex front on M showing an increase in the angle of collision and (b) the corresponding billiard map for (r, θ) depicting the expansion of the unstable curve.

We say that an extreme value law exists for $\{M_n\}$ provided,

$$\lim_{n \rightarrow \infty} \mu(M_n \leq u_n) = e^{-\theta \tau}$$

where $\theta \in [0, 1]$ is called the extremal index. Refer to Chapter 1, Section 1.3, for a detailed description. For certain one-dimensional uniformly expanding maps and Anosov toral automorphisms (refer to the Appendix A.1.1 for a detailed description of these systems) a strict dichotomy has been observed: $\theta = 1$ if p is not periodic and $\theta < 1$ otherwise [36, 40, 61]. In this chapter we investigate the extremal index when p is a periodic point in the setting of two-dimensional hyperbolic systems with singularities. Poisson return time statistics for generic points in a variety of billiard systems (both polynomially and exponentially mixing) were established in [53]. Related results on Poisson return time statistics were obtained for Young Towers with polynomial tails in [80]. However the result of both [53, 80] were limited to a full measure set of generic points, which excluded periodic orbits.

Our goal is to establish an extreme value law in the setting of two-dimensional

hyperbolic systems with singularities (of which the Sinai dispersing billiard model is a subset) when $\varphi(x) = -\log d(x, p)$ is assumed to be uniquely maximized at a periodic point p . We will extend the results of [53] in a similar way to that of [13] to prove the existence of a rare event point process model. We show that the resulting REPP follows a compound Poisson distribution known as the Pólya-Aeppli distribution of index θ . The techniques used here are outlined in Chapter 1, Section 1.4. In particular, we prove the dynamical variant dependence conditions $D_q^*(u_n)$ and $\bar{D}'_q(u_n)$ hold in this setting. The main difficulty in proving this comes from the presence of singularities in the system which result in a fragmentation of phase space and slow down the global expansion of unstable manifolds. This "slow down" affects the recurrence assumption $\bar{D}'_q(u_n)$. We are able to use the growth lemma and a property of the system called **one-step expansion** [20] to overcome these difficulties. See Figure 3.6 for an illustration of this phenomenon.

3.2 Main Assumptions

Let X be a two-dimensional compact Riemannian manifold, possibly with boundary. Let $\Omega \subset X$ be an open subset and let $T : \Omega \rightarrow X$ be a $C^{1+\gamma}$ diffeomorphism of Ω onto $T(\Omega)$ (here $\gamma \in (0, 1]$). We assume that $S_1 = X \setminus \Omega$ is a finite or countable union of smooth compact curves. Similarly, $S_{-1} = M \setminus T(\Omega)$ is a finite or countable union of smooth compact curves. If X has boundary ∂X , it must be a subset of both S_1 and S_{-1} . Call S_1 and S_{-1} **singularity sets** for the maps T and T^{-1} , respectively. Let Ω_i , $i \geq 1$ be the connected components of Ω ; then $T(\Omega_i)$ are the connected

components of $T(\Omega)$. Assume that $T|_\Omega$ is time-reversible, and the restriction of the map T to any component Ω_i can be extended by continuity to its boundary $\partial\Omega_i$, through extensions $\partial\Omega_i \cap \partial\Omega_j$ for $i \neq j$ need not agree. Similarly, for each i the restriction of T^{-1} to any connected component $T(\Omega_i)$ can be extended by continuity to its boundary $\partial T(\Omega_i)$.

Assume the map T is hyperbolic as defined by Katok and Streclyn [60] such that T preserves a probability measure μ . The formal definition of hyperbolicity can be found in the Appendix A.1.2. Moreover, μ a.e. $x \in X$ has two Lyapunov exponents: one positive and one negative. Also, the first and second derivatives of T and T^{-1} do not grow too rapidly near their singularity sets S_1 and S_{-1} respectively and the ϵ -neighborhood of the singularity set has measure $\mathcal{O}(\epsilon)$; this is to ensure the existence and absolute continuity of stable and unstable manifolds at μ -a.e. $x \in X$. Let $\mathcal{W}^u = \bigcap_{n \geq 0} T^n(X \setminus S_1)$. \mathcal{W}^u is (mod 0) the union of all unstable manifolds, and we assume that the partition \mathcal{W}^u of X into unstable manifolds is measurable, so that μ induces conditional distributions on μ -almost all unstable manifolds (see the definition and basic properties of conditional measures in [23, Appendix A]). Most importantly, we assume that the conditional distributions of μ on unstable manifolds $W \subset \mathcal{W}^u$ are absolutely continuous with respect to the Lebesgue measure on W . This means that μ is the so called Sinai-Ruelle-Bowen (SRB) measure. We also assume that our SRB measure μ is ergodic and mixing. In physics terms, μ is an equilibrium state for the potential $-\log DT|_{\mathcal{W}^u}$.

For $n \geq 1$, let

$$S_n = \bigcup_{i=0}^{n-1} T^{-i} S_1 \text{ and } S_{-n} = \bigcup_{i=0}^{n-1} T^i S_{-1}.$$

Then the map $T^n : X \setminus S_n \rightarrow X \setminus S_{-n}$ is a $C^{1+\gamma_0}$ diffeomorphism.

The following specific assumptions on the system (X, T, μ) provide sufficient conditions for exponential mixing rates and the coupling lemma. These assumptions have been made in other analyses [18, 22, 23, 24].

(h1) Hyperbolic cones for T^1 . There exist two families of cones C_x^u (unstable) and C_x^s (stable) in the tangent spaces $\mathcal{T}_x X$, for all $x \in X \setminus S_1$, and there exists a constant $\Lambda > 1$, with the following properties:

- (1) $D_x T(C_x^u) \subset C_{T_x}^u$ and $D_x T(C_x^s) \supset C_{T_x}^s$, wherever $D_x T$ exists.
- (2) $\|D_x T(v)\| \geq \Lambda \|v\|$, $\forall v \in C_x^u$ and $\|D_x T^{-1}(v)\| \geq \Lambda \|v\|$, $\forall v \in C_x^s$.
- (3) These families of cones are continuous on X and the angle between C_x^u and C_x^s is uniformly bounded away from zero.

We say that a smooth curve $W \subset X$ is an unstable (stable) *curve* if at every point $x \in W$ the tangent line $\mathcal{T}_x W$ belongs in the unstable (stable) cone C_x^u (C_x^s). Furthermore, a curve $W \subset X$ is an unstable (stable) *manifold* if $T^{-n}(W)$ is an unstable (stable) curve for all $n \geq 0$.

(h2) Singularities. The boundary ∂X is transversal to both stable and unstable cones. Every other smooth curve $W \subset S_1 \setminus \partial X$ (resp. $W \subset S_{-1} \setminus \partial X$) is a stable (resp. unstable) curve. Every curve in S_1 terminates either inside another curve of S_1 or on the boundary ∂X . A similar assumption is made for S_{-1} . Moreover, there exists $C > 0$ such that for any $x \in X \setminus S_1$

$$\|D_x T\| \leq C \text{dist}(x, S_1)^{-1}, \quad (1)$$

and for any $\epsilon > 0$,

$$\mu(x \in X : \text{dist}(x, S_1) < \epsilon) < C\epsilon. \quad (2)$$

Note that (2) implies that for μ -a.e. $x \in X$, there exists a stable and unstable manifold $W^{u/s}(x)$, such that $T^n W^s(x)$ and $T^{-n} W^u(x)$ does not intersect S_1 for any $n \geq 0$.

Definition 3.2.1. For every $x, y \in X$, define $s_+(x, y)$, the **forward separation time** of x, y to be the smallest integer $n \geq 0$ such that x and y belong to distinct elements of $X \setminus S_n$. Fix $\beta \in (0, 1)$, then $d(x, y) = \beta^{s_+(x, y)}$ defines a metric on X . Similarly we defined **backwards separation time** $s_-(x, y)$.

(h3) Regularity of stable/unstable curves. We assume that the following families of stable/unstable curves, denoted by $\mathcal{W}_T^{s,u}$ are invariant under T^{-1} (resp., T) and include all stable/unstable manifolds:

1. **Bounded curvature.** There exist $B > 0$ and $c_M > 0$, such that the curvature of any $W \in \mathcal{W}_T^{s,u}$ is uniformly bounded from above by B , and the length of the curve $|W| < c_M$.
2. **Distortion bounds.** There exist $\gamma_0 \in (0, 1)$ and $C_r > 1$ such that for any unstable curve $W \in \mathcal{W}_T^u$ and any $x, y \in W$,

$$|\log \mathcal{J}_W(T^{-1}x) - \log \mathcal{J}_W(T^{-1}y)| \leq C_r \text{dist}(x, y)^{\gamma_0} \quad (3)$$

3. **Absolute continuity.** Let $W_1, W_2 \in \mathcal{W}_T^u$ be two unstable curves close to each other. Denote

$$W'_i = \{x \in W_i : W^s(x) \cap W_{3-i} \neq \emptyset\}, \quad i = 1, 2.$$

The map $h : W'_1 \rightarrow W'_2$ defined by sliding along stable manifolds is called the *holonomy* map. We assume $h_*\mu_{W'_1} \prec \mu_{W'_2}$. and furthermore, there exist uniform constants $C_r > 0$ and $\vartheta_0 \in (0, 1)$, such that the Jacobian of h satisfies

$$|\log \mathcal{J}h(y) - \log \mathcal{J}h(x)| \leq C_r \vartheta_0^{s_+(x,y)}, \quad \forall x, y \in W'_1 \quad (4)$$

Similarly, for any $n \geq 1$ we can define the holonomy map

$$h_n = T^n \circ h \circ T^{-n} : T^n W_1 \rightarrow T^n W_2,$$

and then (4) and the uniform hyperbolicity **(h1)** imply

$$\log \mathcal{J}h_n(T^n x) \leq C_r \vartheta_0^n \quad (5)$$

(h4) One-step expansion. We have,

$$\liminf_{\delta \rightarrow 0} \sup_{W:|W|<\delta} \sum_n \left(\frac{|W|}{|V_n|} \right)^\alpha \cdot \frac{|T^{-1}V_n|}{|W|} < 1$$

where the supremum is taken over regular unstable curves $W \subset X$, $|W|$ denotes the length of W , and V_n , $n \geq 1$, the smooth components of $T(W)$, $\alpha \in (0, 1]$.

Lemma 3.2.2. Growth Lemma. *Let W be a small unstable curve and m_W the Lebesgue measure on W . Define G_ϵ as the set of points in W that are at most ϵ from the boundary,*

$$G_\epsilon = \{x \in W : d(x, \partial W) \leq \epsilon\}$$

and H_ϵ as the set of points in W that will be at most ϵ from the boundary,

$$H_\epsilon = \{x \in W : d(Tx, \partial(TW)) \leq \epsilon\}.$$

There exists a constant $\lambda < \Lambda$, independent of W , such that

$$m_W(H_\varepsilon) \leq \lambda m_W(G_{\varepsilon/\Lambda})$$

where Λ is the rate of expansion of W .

Remark 3.2.3. The assumptions **h1-h4** are satisfied by the billiard map associated to Sinai dispersing billiards with finite and infinite horizon.

Remark 3.2.4. The assumptions **h1-h4** along with the Growth Lemma imply the existence of a Young Tower with exponential tails [24, Lemma 17].

3.3 Main Results

Fix a hyperbolic point $p \in X$ with prime period $q > 1$, and let $\varphi : X \rightarrow \mathbb{R} \cup \{+\infty\}$ be given by,

$$\varphi(x) = -\log(d(x, p))$$

where d is a metric on X specified later. We assume that iterates of p do not lie on the singularity sets $S_1 \cup S_{-1}$. Our metric d will be adapted to the chart given by the stable and unstable manifolds of p , denoted as $W^s(p)$ and $W^u(p)$. If the stable manifold of x denoted as $W^s(x)$ intersects $W^u(p)$, say at a point \tilde{x} , then we define $x^s := \text{dist}_{W^u(p)}(\tilde{x}, p)$ i.e. the distance of \tilde{x} and p measured along the unstable manifold $W^u(p)$. Note that if $W^u(x)$ is very short and does not reach $W^s(p)$, then we may extend $W^u(x)$ as an unstable curve, thus x^u can be similarly defined. If the unstable manifold (or extension to an unstable curve) of x , denoted as $W^u(x)$ intersects $W^s(p)$ at \tilde{x} , we define $x^u := \text{dist}_{W^s(p)}(\tilde{x}, p)$. The foliation of stable and

unstable curves of a sufficiently small neighborhood of p will be Hölder continuous. Moreover, if both x, y lie in the same local chart determined by stable and unstable manifolds (or stable and unstable curves) so that $x = (x^u; x^s)$, $y = (y^u; y^s)$, we define

$$d(x, y) = \max\{|x^u - y^u|, |x^s - y^s|\}$$

In this dynamically adapted metric [13] has shown that for q not periodic the REPP follows a compound Poisson process with parameter given by the extremal index θ of the EVL so that $\theta = 1 - \frac{1}{|DT_u^q(p)|}$ where $DT_u^q(p)$ is the derivative of T^q in the unstable direction at p .

Comparing the Euclidean and Adapted Metric

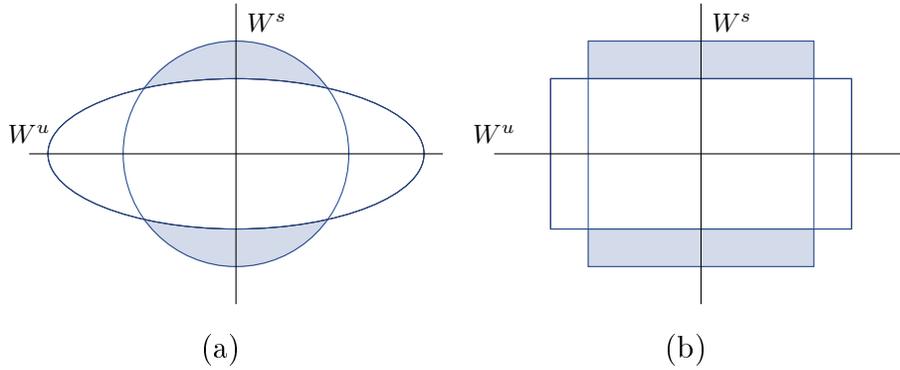


Figure 3.4: The sets $A_n^q := \{X_0 > u_n, X_q < u_n\}$ in (a) the usual Euclidean metric and (b) the adapted metric.

Theorem 3.3.1. *Let p be a periodic orbit of period q , $p \notin S_1 \cup S_{-1}$, and define*

$$\varphi(x) = -\log(d(x, p)),$$

where d is the metric adapted to the chart given by the stable and unstable foliation.

Define $M_n := \max\{\varphi, \dots, \varphi \circ T^{n-1}\}$. Then

$$\lim_{n \rightarrow \infty} \mu(M_n \leq u_n) = e^{-\theta u_n},$$

where $\theta = 1 - \frac{1}{|DT_u^q(p)|}$.

To establish Theorem 3.3.1 it suffices to prove conditions $\mathcal{D}_q(u_n)$ and $\mathcal{D}'_q(u_n)$ stated in Chapter 1, Section 1.4, since if both conditions hold then the limit exists and

$$\lim_{n \rightarrow \infty} \mu(M_n \leq u_n) = e^{-\theta\tau}$$

The proof of condition $\mathcal{D}_q(u_n)$ is somewhat standard [?, 13] and is not stated in this section. For completeness of exposition it has been assigned to the Appendix A. The novelty of this analysis comes the proof of condition $\mathcal{D}'_q(u_n)$ which uses the growth lemma of [16] applied to unstable curves to overcome issues with singularities.

Proof. Condition $\mathcal{D}'_q(u_n)$ Before checking $\mathcal{D}'_q(u_n)$, we note that we need only to consider the sum up to time $(\log n)^{1+\delta}$, for $\delta > 0$ since by exponential decay of correlations, the remaining sum goes to 0.

We assume that $p \in X$ does not lie on a discontinuity. Let Ω_n be the square of side length $1/\sqrt{n}$ centered at p corresponding to the set $\{\varphi > u_n\}$ so that $\mu(\Omega_n) = O(\frac{1}{n})$. We will say that a point $x \in \Omega_n$ *fully crosses* Ω_n if $W^u(x)$ is a smooth connected unstable manifold touching both stable manifolds defining the boundary $\partial\Omega_n$.

Let $B_n = \{x \in \Omega_n : W^u(x) \text{ does not fully cross } \Omega_n\}$ and $G_n = \{x \in \Omega_n \cap B_n^c\}$.

Short returns.

We consider points which leave our neighborhood after the first iterate and return at the j th iterate. For simplification we consider $F = T^q$, so that p is a fixed point for F . Clearly for large n , $A_n^q = \{X_0 > u_n, X_0 \circ F < u_n\}$ since (u_n) increase as

$n \rightarrow \infty$.

Recall that $DF_p^u \sim \lambda > 1$ is the expansion in the unstable direction at the fixed point p . We only consider points on the unstable manifold since by the growth lemma we have that any x not on an unstable manifold will reach an unstable manifold a finite number of iterates. If $W^u(x)$ is an unstable manifold for $x \in \Omega_n \cap G_n$, the set $A_u(x) := A_n^q \cap W^u(x)$ has two connected components (right and left subintervals) of $W^u(x) \cap \Omega_n$ which are roughly a distance $\frac{1}{\lambda\sqrt{n}}$ from p . In this case we only consider points contained in $\Omega_n \cap G_n$ since any point $x \in \Omega_n \cap B_n$ will take longer to leave Ω_n .

Expansion of A_n^q Under the Map T

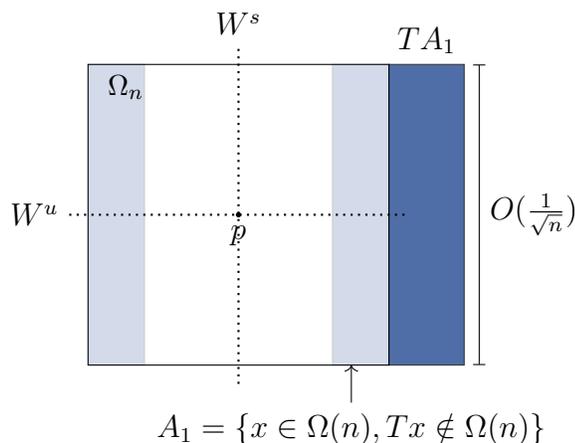


Figure 3.5: Expansion of the set A_n^q out of Ω_n under the map T .

For large n , on Ω_n the map F acts on $W^u(p) \cap \Omega_n$ as an expansion in the unstable direction for a number of iterates. We see that $\lambda^j \frac{1}{n} < 1$ and solving for j implies that for all $0 < j < C \log n$ the unstable manifold is not long enough to fully wrap around the torus and hence, if m_u is Lebesgue measure on $W^u(p)$, then $m_u((A_u(x) \cap W^u(p)) \cap F^j(A_u(x) \cap W^u(p))) = 0$ for all $0 < j < C \log n$ for some $C > 0$. We will

estimate an upper bound for C for sufficiently large n by solving $\lambda^k \frac{1}{\sqrt{n}} \sim n^{-\beta}$ for a fixed $0 < \beta < \frac{1}{2}$. We see that taking $C \sim \frac{1/2-\beta}{\log \lambda}$ will do. Given that we are using an adapted metric this implies that $\mu((A_u(x)) \cap F^j(A_u(x))) = 0$ for all $0 < j < C \log n$.

Thus

$$\lim_{n \rightarrow \infty} n \sum_{j=1}^{C \log(n)} \mu((A_n^q) \cap T^{-j}(A_n^q)) = 0$$

Intermediate returns.

The argument here is based on that of [13] in which a corresponding result was proven for linear toral automorphism. The main idea is that if an unstable manifold $W^u(x) \cap A_n^q$ of A_n^q is to intersect A_n^q for an intermediate iterate j , so that $T^j(W^u(x) \cap A_n^q) \cap A_n^q \neq \emptyset$, then $T^j(W^u(x) \cap A_n^q)$ must be quite long (and perhaps have 'wrapped around' the torus several times) and so the length of $[T^j(W^u(x) \cap A_n^q) \cap A_n^q]$ as a fraction of the length of $T^j(W^u(x) \cap A_n^q)$ is small (as $\Omega(n)$ itself has diameter only $\frac{1}{\sqrt{n}}$). And by bounded distortion (which poses no problem in the case of a linear toral automorphism) the fractional measure of $T^{-j}[T^j(W^u(x) \cap A_n^q) \cap A_n^q] \cap [W^u(x) \cap A_n^q]$ is small also, say of order $\frac{1}{\sqrt{n}}|W^u(x) \cap A_n^q|$. This implies, by a Fubini type argument, that $\mu(T^{-j}(A_n^q) \cap A_n^q) \leq C \frac{1}{n^{3/2}}$, and hence

$$\lim_{n \rightarrow \infty} n \sum_{j=C \log n}^{(\log n)^{1+\delta}} \mu((A_n^q) \cap T^{-j}(A_n^q)) \leq \frac{(\log n)^{1+\delta}}{\sqrt{n}}$$

which limits to zero as $n \rightarrow \infty$.

We will replicate this argument in the setting of hyperbolic billiards. But there are several difficulties to overcome, as the presence of singularities causes the unstable manifolds to fragment under iteration, the expansion is not uniform and care is needed with bounded distortion arguments.

Two Outcomes of Expansion of A_n^q Under T

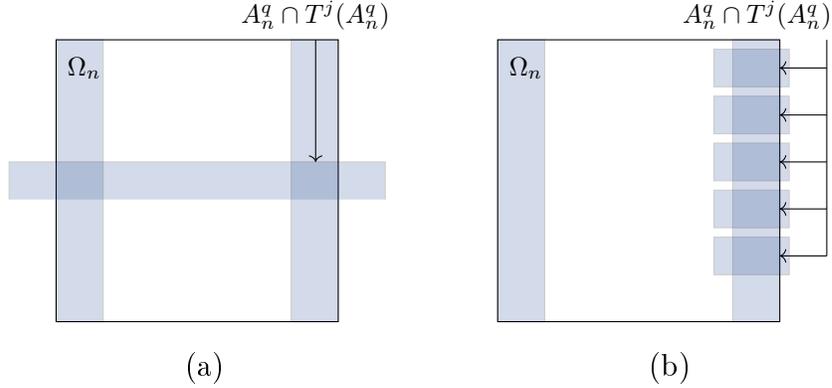


Figure 3.6: (a) Expansion of the set A_n^q under the map T before it hits a singularity point. (b) Fragmentation of the unstable direction of A_n^q under the map T after it hits a singularity point. This represents the "worst-case scenario" where estimates are needed to ensure the fragmented pieces do not remain in the intersection for "too long".

First we note that the components $W^u(x) \cap A_n^q$ do not hit a singularity set in the first $C \log n$ iterates. To see this recall that the point p is a fixed distance $\epsilon > 0$ away from the singularity set. F acts as a diffeomorphism on a sufficiently small neighborhood of p , which contains Ω_n . By expansion we have that $\lambda^k \frac{1}{n} \sim n^{-\beta}$ for $0 < \beta < 1$ so that for large n , $n^{-\beta} < \epsilon$ and points in Ω_n do not hit that singularity set in $C \log n$ iterates.

Now we consider $\mu(A_n \cap T^{-j}(A_n))$ for $C \log n \leq j \leq (\log n)^{1+\delta}$. Suppose $x \in G_n$ then $F^j(W^u(x) \cap A_n)$ consists of two connected components for $j \leq C \log n$. For simplicity we will focus on the 'right hand' component of $W^u(x) \cap A_n^q$ (the argument for the left hand is the same). We define $\gamma(x)$ to be the right hand component of $F^{[C \log n]}(W^u(x) \cap A_n^q)$. By expansion $|\gamma(x)| \sim n^{-\beta}$. If we iterate this component further such that $F^j \gamma(x)$, $j > 0$ hits a singularity then we may decompose $F^j \gamma(x)$

into smooth connected components V_n and their preimages U_n so that T^j maps U_n onto V_n diffeomorphically and with uniformly bounded distortion. Applying one-step expansion for $p \in \gamma(x)$ gives,

$$\left(\frac{|\gamma(x)|}{|V_n(p)|} \right)^\alpha \left| \frac{U_n(p)}{\gamma(x)} \right| < 1$$

Fix $\gamma(x)$ and for every point $y \in \gamma(x)$ let $d\mu_\gamma(y) = \frac{U_n(y)}{\gamma(x)}$ be the density of a probability measure $\mu_\gamma(y)$ on $\gamma(x)$ and $f(x) = \left(\frac{|\gamma(x)|}{|V_n(y)|} \right)^\alpha$ a function on this probability space. Then $\{y \in \gamma : V_n(y) < n^{-(1+\varepsilon)\beta}\} \subset \{y \in \gamma(x) : f(y) > n^{\beta\alpha\varepsilon}\}$ and by Markov's inequality $m_\gamma\{y \in \gamma : V_n(y) < n^{-(1+\varepsilon)\beta}\} \leq n^{-\beta\alpha\varepsilon}$. We choose ε sufficiently small so that $\rho_1 := \frac{1}{2} - (1 + \varepsilon)\beta > 0$ (recall $\beta < \frac{1}{2}$) and define $\rho := \min\{\frac{1}{2} - (1 + \varepsilon)\beta, \beta\alpha\varepsilon\}$. With our choice of ε if $|V_n| \geq n^{-(1+\varepsilon)\beta}$ then $\frac{|V_n \cap A_n^q|}{|V_n|} \leq C_1 n^{-\rho}$.

By bounded distortion of the map $F^{[C \log n]} : W^u(x) \cap A_n^q \rightarrow \gamma(x)$,

$$\frac{|U_n \cap T^{-j}(A_n^q)|}{|U_n|} \leq C_2 n^{-\rho}$$

and by bounded distortion again, if $i > q[C \log n]$ then

$$\frac{|W^u(x) \cap A_n^q \cap T^{-i}(A_n^q)|}{|W^u(x) \cap A_n^q|} \leq C_3 n^{-\rho}.$$

This provides a bound on the length of the intersection of a single unstable manifold. We may now use the fact that μ decomposes as a product measure on U_n so that if we consider all manifolds of $x \in G_n$ we have for $i > q[C \log n]$,

$$\mu((G_n \cap A_n) \cap T^{-i}(A_n)) \leq C_4 n^{-1-\rho}.$$

Finally, to deal with the set B_n of x for which the unstable manifold does not fully cross Ω_n we may assume by the growth lemma that $\frac{\mu(B_n)}{\mu(\Omega_n)} \leq (\log n)^{-(1+\delta)}$ and hence,

$$\mu((B_n \cap A_n^q) \cap T^{-i}(A_n^q)) \leq n^{-1} (\log n)^{-(1+\delta)}.$$

Orbit of a Periodic Point Under the Billiard Flow

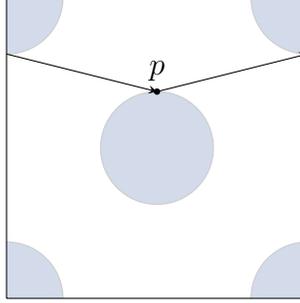


Figure 3.7: A periodic point p for the billiard flow on the torus with circular barriers.

Putting these results together implies,

$$\lim_{n \rightarrow \infty} n \sum_{j=C \log(n)}^{\log(n)^{1+\delta}} \mu((A_n^q) \cap T^{-j}(A_n^q)) = 0.$$

Condition $\mathcal{A}'_q(u_n)$ follows. □

Remark 3.3.2. The value of the EI θ is computed similarly to that of [13, Section 6]. These calculations are rather complicated and left out of this discussion for simplicity.

Theorem 3.3.3. *Let p be a periodic orbit of period q , $p \notin S_1 \cup S_{-1}$, and*

$$\varphi(x) = -\log(d(x, p))$$

where d is the metric adapted to the chart given by the stable and unstable foliation. Define $X_n = \varphi \circ T^n$ and $M_n := \max\{X_0, \dots, X_{n-1}\}$. Let $(u_n)_{n \in \mathbb{N}}$ be a sequence satisfying $\lim_{n \rightarrow \infty} n\mathbb{P}(X_0 > u_n(\tau)) = \tau$ for some $\tau \geq 0$ and $(v_n)_{n \in \mathbb{N}}$ be given by $v_n = 1/\mu(X_0 > u_n)$. Consider the REPP N_n as in Chapter 1, Section 1.4. Then the REPP N_n converges in distribution to a compound Poisson process with intensity

$\theta = 1 - \frac{1}{|DT_u^q(p)|}$ and geometric multiplicity d.f. G^* given by $G^*(k) = \theta(1 - \theta)^{k-1}$, for all $k \in \mathbb{N}$.

Proof. To establish Theorem 3.3.3 it suffices to prove condition $\mathbb{D}'_q(u_n)$ and a condition related to $\mathbb{D}_q(u_n)$, which we call Condition $D_q^*(u_n)$ [41] (see also the discussion in [13, Section 8.1]). The proof of condition $D_q^*(u_n)$ follows from $\mathbb{D}_q(u_n)$ by [Proposition 4.4.1][73]. □

Chapter 4

Modeling Real-World Temperature Extremes

Extreme temperature events can make considerable impacts on society, most obviously on human health and power consumption. As a result there has been a growing interest in examining temperature extremes in relation to climate change [44, 37, 65, 85, 95, 33, 38]. In particular in the last decade the Texas Gulf Coast region has experienced high profile heat waves, such as the summer of 2011, and momentous summer rainfall and flooding [50]. In this paper we address the question of whether the probability of extreme summer temperatures in this region has increased by using clustering techniques and fitting generalized extreme value [27] (GEV) models to the data which allow for forecast modeling of the maxima by estimating the probability of high temperatures. Although the classical GEV model requires the mean and variance of the maxima to be stationary (not change in time) it is possible to adapt this

probability model to allow for prediction in non-stationary scenarios[51, 15, 44, 9]. We find compelling evidence that the probability of extreme temperatures during summer has increased.

It is necessary to have comprehensive information on extreme weather events to make reasonable conclusions from the data. Clustering techniques allow enlargement of the time series data to permit better modeling. This is especially useful when determining macroscopic weather changes since natural variability, such as human influence over the environment [26], causes nonuniform changes [94, 70, 49, 98]. We apply clustering techniques to Texas wide weather stations to provide a larger pool of data for GEV fits of the distribution of maximum temperatures. Our approach is particularly effective when looking at large scale climate data and global environmental zones [100, 99]. We use a combination of clustering methods, including *K*-means [52], with mutual information as a measure of similarity between weather stations. In this way we give a comprehensive extreme value analysis of summer extreme temperatures throughout Texas.

Recent extreme weather events in Houston, Texas, provide motivation for our station choice in this paper. Anecdotally, higher extreme temperature patterns have been recorded in Houston for the July-August months when compared to records 40 years prior. The National Oceanic and Atmospheric Administration (NOAA) reports the longest stretch of record high temperatures August 1st - August 24th 2011 with the highest temperature ever recorded in Houston occurring on August 27, 2011 and yearly record highs near this value occurring in the summer of 2013 and 2015 [67]. Preliminary results on the temperature vector from the Houston station suggest a

significant change in the mean and variance of temperature maxima after year 1981. In particular a stationary GEV model was shown to fit the summer temperature maxima prior to 1981 at the $\alpha = 0.05$ significance level with poor fits after year 1981. Moreover, an analysis of the QQ plots indicate poor fits in the tail of the GEV distribution after 1981 suggesting a non-stationary scale parameter. Nonparametric trend tests confirmed this hypothesis. Other studies have also recorded more incidences of higher temperature outliers after 1980 [48, 39, 76, 84] with claims that this increase in temperature extremes can be attributed to global warming due to human-made greenhouse gases. This provides motivation in the following analysis for breaking the time series into time windows 1941-1981, 1982-2017 and 1941-2017. We test these periods for stationarity and find across clusters the period 1941-1981 is stationary but this is not the case for the periods 1982-2017 and 1941-2017. Based on these results we fit non-stationary GEV models to estimate the probability of current temperature extremes in our clusters. We do not consider rainfall patterns or the interaction between summer temperatures and rainfall in this paper.

We end with a brief discussion on extreme temperatures for the December-January period where no evidence of change in probabilities of extremes across Texas is found. We estimate generalized extreme value (GEV) and generalized Pareto (GP) distributions [27] for the winter temperature maxima. Even after accounting for dependence and nonstationarity, goodness of fit tests return poor results for both models for winter data. We hypothesize that these poor fits may be due to Canadian airflow temporarily lowering extremes throughout Texas and leading to a mixed extreme distribution.

4.1 Preliminary results on Houston, Texas, summer temperature extremes

The block maxima method described in Chapter 1, Section 1.5, was performed on the vector of temperature measurements where each block contains hourly records over 10 days, with the ultimate goal of fitting a GEV model to the block maxima of the time series. In order to test for independence the time series was compressed by the following: if the next time step increased, a value of 1 was assigned to the compressed time series, if it decreased a value of -1 was assigned and value 0 if it remained the same. A count matrix was created across each 10 day block and the chi-square test of independence. (See Appendix A.2.2 for more details on this test) was performed on this matrix. All blocks reported independence at the $\alpha = 0.05$ significance level. (See Figure 4.14 for the histogram of p values.) The maximum is taken over each 10 day block of the uncompressed time series.

Remark 4.1.1. It is important to note that the compression performed for chi-square independence testing is separate from compression for mutual information calculation in the next section. Since daily cycles are typically the cause of temperature dependence we test increases and decreases across 10 day cycles to determine whether the one 10 day temperature cycle determines another. The compression performed in the following section seeks to maximize mutual information (maintain as much information as possible) between whole time series.

Maximum likelihood estimation (MLE) is performed iteratively using the Nelder-Mead simplex method [77] to estimate the value of the three parameters for the GEV

function over all 76 years of block maxima ($76 \times 60/10 = 456$ maximum temperature values). Anderson-Darling goodness of fit tests were performed on the binned data with MLE parameters. (See Appendix A.2.2 for more information on the test statistic.)

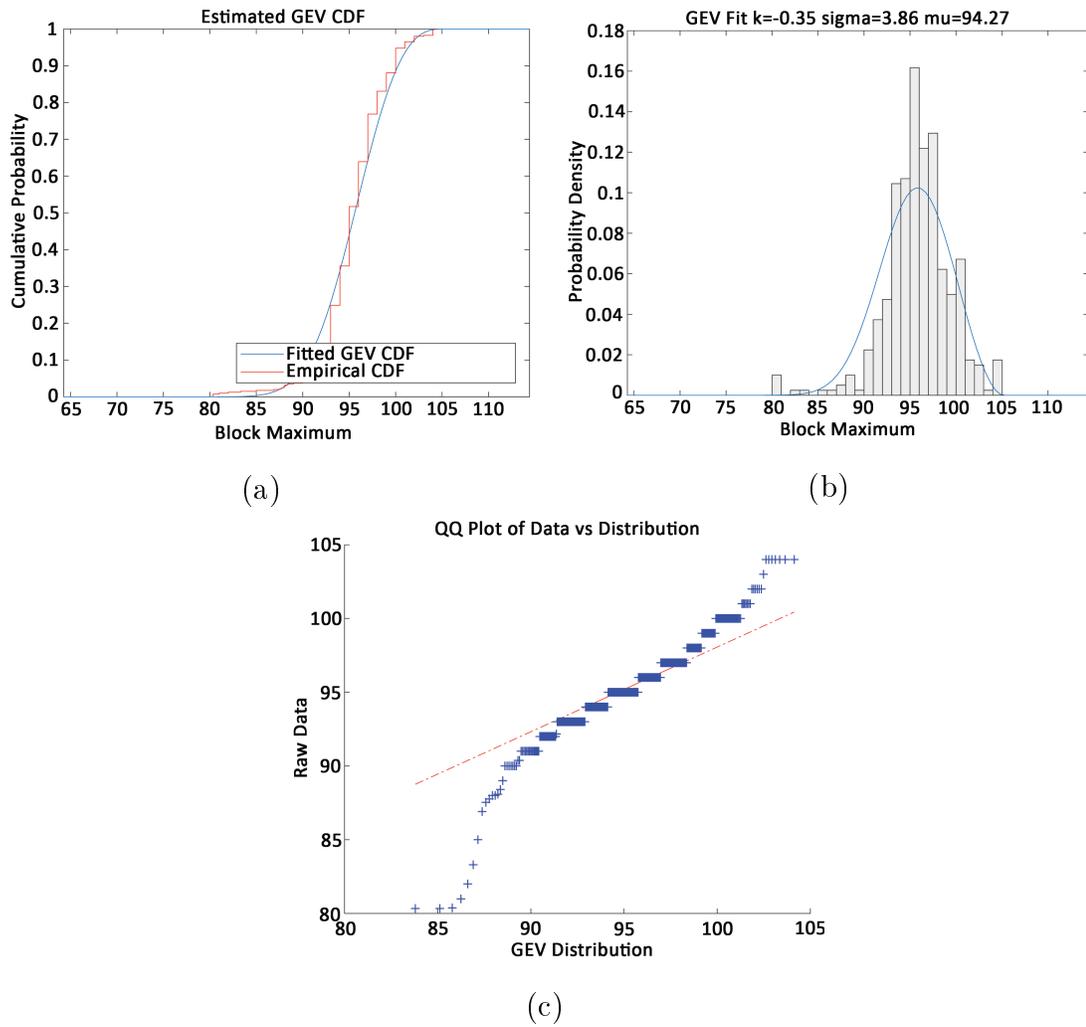
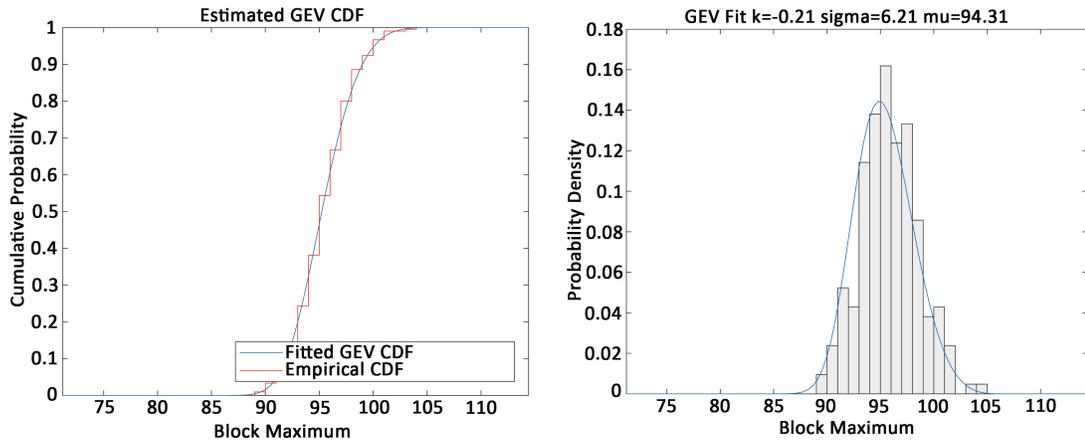


Figure 4.1: Generalized extreme value distribution for the Houston station for 10-day maximum temperatures for 1941-2017 with MLE parameters (a) cumulative density function, (b) probability density function, and (c) qq plot. Anderson-Darling p -value = $4.9e - 3$.

Comparing the Anderson-Darling goodness of fit results across Figures, in particular the QQ plot, note that poor fits for the GEV model of maximum values occur in the tails of the distribution and point to possible nonstationarity (varying in time) of the location μ and scale σ parameters. It is interesting to ask whether this nonstationarity is consistent. The time series is broken roughly in half and a GEV model is fit to the block maxima over the first 40 years (240 maximum values) and over the last 36 years (216 maximum values) with the following results.

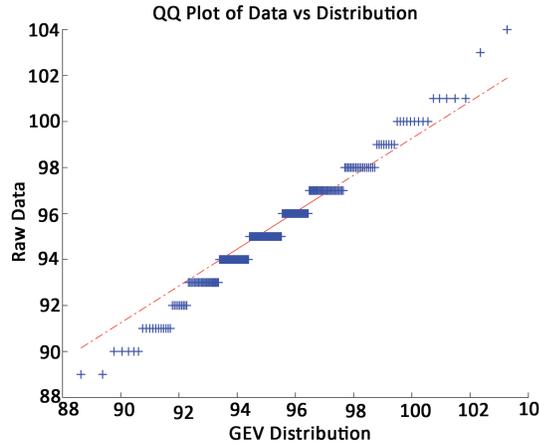
Interestingly, a good fit for the GEV model at the $\alpha = 0.05$ significance level is found over maximum values for the first 40 years followed by poor fits over the last 36 years. These results suggest nonstationarity of the GEV model parameters occur after 1981 and support further investigation of summer temperature extremes on a larger geographic scale. Other studies have also recorded more incidences of higher temperature outliers after 1980 [48, 39, 76, 84]. All this provides motivation for breaking the time series into time windows 1941-1981, 1982-2017 and 1941-2017 in the following analysis.

Since this preliminary analysis only considers only one time series for the Houston station, we calculate the 3-year mean and variance values (a single variance value calculated from 18 maximum data points). When clustering is performed in the next section, we calculate the 1-year mean and variance values since more data is available. The nonparametric Mann-Kendal p -values are given for each time window. These results support the hypothesis that the scale parameter is non-stationary after 1981 for the Houston station.



(a)

(b)



(c)

Figure 4.2: Generalized extreme value distribution for the Houston station with MLE parameters for 10-day maximum temperatures for 1941-1981 (a) cumulative density function, (b) probability density function, and (c) qq plot. Anderson-Darling p -value= 0.25.

4.2 Analysis of summer temperature extremes throughout the Texas region

In this section, spectral clustering is applied to Texas wide weather stations to provide a larger pool of data for GEV fits of the distribution of maximum temperatures.

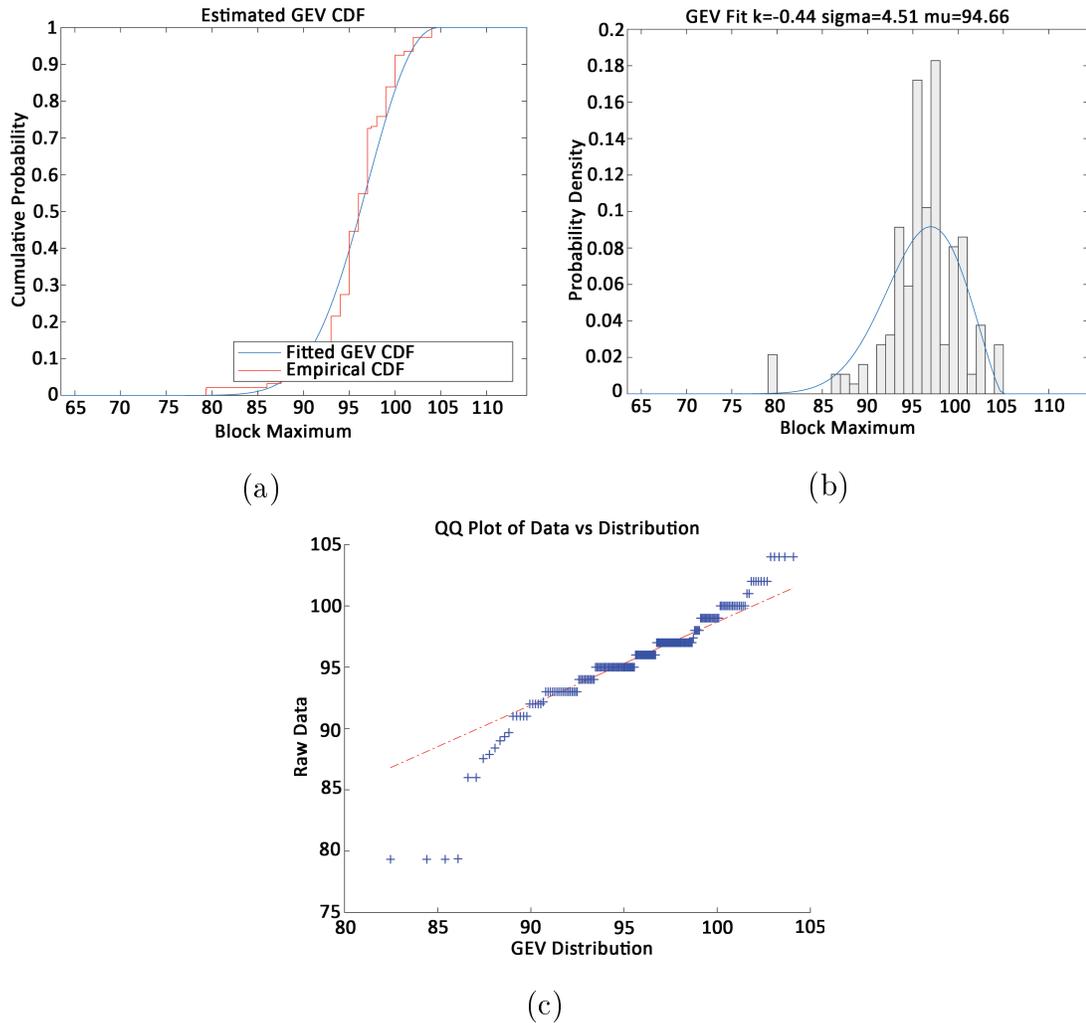


Figure 4.3: Generalized extreme value distribution for the Houston station with MLE parameters for 10-day maximum temperatures for 1982-2017 (a) cumulative density function, (b) probability density function, and (c) qq plot. Anderson-Darling p -value= $4.9e - 4$.

This is especially useful when determining macroscopic weather changes since natural variability, such as human influence over the environment, causes nonuniform changes. The increased effectiveness of this clustering analysis comes from the modifications made on standard spectral clustering techniques including choice of mutual

Mann-Kendall Values for Houston

parameter	μ			σ^2		
	1941-1981	1981-2017	1941-2017	1941-1981	1981-2017	1941-2017
years						
M-K p -value	0.12	0.06	0.90	0.48	4.4e-3	0.23

Table 4.1: Table of Mann-Kendall p -values for the 3-year mean and variance of the Houston station.

information as a similarity measurement, estimation of mutual information by gradient ascent and consideration of time-dependent clusters.

4.2.1 Data

Hourly temperature measurements (in Celsius) were obtained from the National Oceanic and Atmospheric Administration (NOAA) Integrated Surface Data (ISD) lite. The ISD lite data set comprises hourly weather observations from over 100 sources providing global climate information [72]. A subset containing the hourly temperature measurements for stations across Texas and New Orleans, Louisiana, for the July-August months from 1941-2017 was created. New Orleans, Louisiana was included because of recent similar extreme events to that of Houston. The 31st of each month was not considered for 10 day block divisibility. Though many stations were listed by the NOAA across Texas, the following stations (see Table 4.2 and Figure 4.4 on page 105) were chosen for this analysis because they contain complete records aside from possible one to two year gaps. For clarification a *time series* in this analysis will be defined as the hourly temperature vector for July-August, 1941-2017 for a single station. We convert the time series from Celsius to Fahrenheit for pragmatic reasons.

Quality control at the integration level was performed by the NOAA on the data in two phases. We list the most relevant for our purposes. Phase one ran a check that the temperature values indeed came from the same time and location as listed. This was done by comparing hourly coincident data over each day to obtain a percent score for the day. The criteria of 1-degree variance in Celsius was used as a pass/fail limit for each, with an overall 70% score for the day required for integration into the data set. Phase two performed a "two-sided" continuity validation on each temperature value per day in the following way: an increase by at least 8 degrees Celsius in one data point followed by an increase of at least 8 degrees Celsius in the next point was assumed to indicate a possible erroneous value and was replaced by a missing value (similarly for downward spikes).

Observations for the ISD lite data set were selected from the available data sets for a single location by defining a capture window to extract temperature measurements closest to the top of the hour. In many cases there was more than one element at the same observation time. In these situations, a single element was chosen according to a ranking schema which took into account data quality of the set and origin [78]. Despite this ranking system, the observation time for each element can differ by as much as ten minutes. This does not pose a problem for our purposes since we are interested in investigation the long-term behavior of the time series.

Some of the stations chosen for this analysis have up to one to two year intervals where data is missing. Common forms of missing data replacement including moving average or linear interpolation are reasonably accurate in a small time window but break down when these large gaps occur. A more accurate model to replace this

data may exist, however, dependence and trend reports for these time periods could not be trusted to represent the true fluctuations. For this reason, missing data was excluded from this analysis.

Station Map

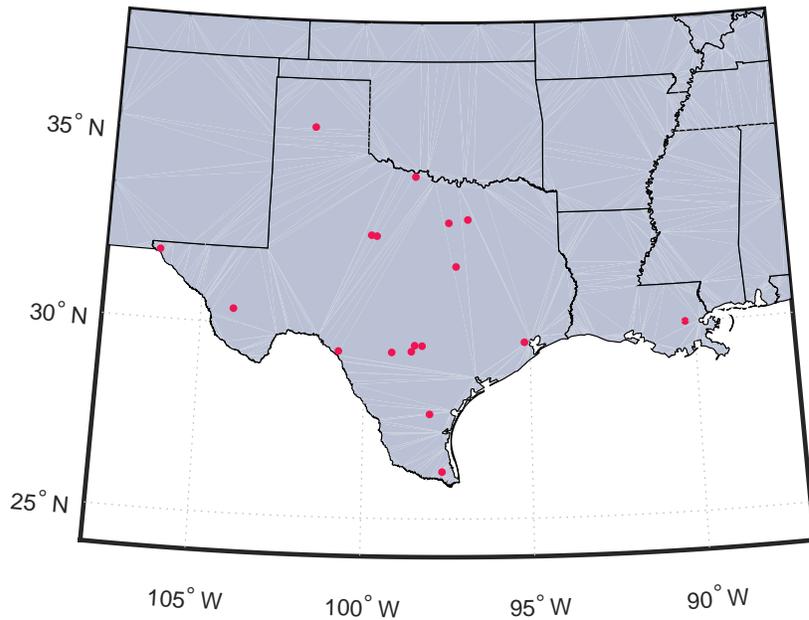


Figure 4.4: Map of all listed stations.

This data set does not account for the urban heat island [79] which may contribute to an observed increase in mean temperature due to human activity; however, the stations chosen for analysis are far enough away from cities that this should not pose an issue. Although urbanization of the surrounding area can cause bias in the long-term trend behavior of the temperature, we argue that the number of stations in which this occurs is small and contributes a minor amount to the cluster as a whole.

Station List

Station	ID	Location	Type
1	690190	Abilene, TX	Airport
2	722310	New Orleans, LA	Airport
3	722436	Houston, TX	Airport
4	722505	Harlingen, TX	Airport
5	722517	Alice, TX	Airport
6	722530	San Antonio, TX	Airport
7	722533	Hondo, TX	Airport
8	722535	San Antonio, TX	Lackland AFB
9	722536	San Antonio, TX	Randolph AFB
10	722560	Waco, TX	Airport
11	722580	Dallas, TX	Airport
12	722595	Fort Worth, TX	Naval Air Station
13	722615	Del Rio, TX	Laughlin AFB
14	722640	Marfa, TX	Airport
15	722660	Abilene, TX	Airport
16	722700	El Paso, TX	Airport
17	723510	Wichita, Falls, TX	Sheppard AFB
18	723630	Amarillo, TX	Airport

Table 4.2: Table of all stations used in this analysis with location and type.

4.2.2 Results

4.2.2.1 Clustering Results

We briefly describe the techniques used for clustering in this section. For a more detailed explanation refer to Chapter 1, Section 1.6. The continuous time series of temperature measurements was compressed by giving values -1 , 0 , or 1 to measurements which fall into the intervals $[-\infty, a]$, $[a, b]$, and $[b, \infty]$ respectively. Values for a and b were chosen by gradient ascent of the mutual information between two time series. Mutual information is taken as a measure of similarity. In this analysis mutual information between two time series was only calculated over non-missing years for *both* time series. See Figure 4.5 for histogram of maximized mutual information

and Figure 4.6 for the error on the maximized mutual information.

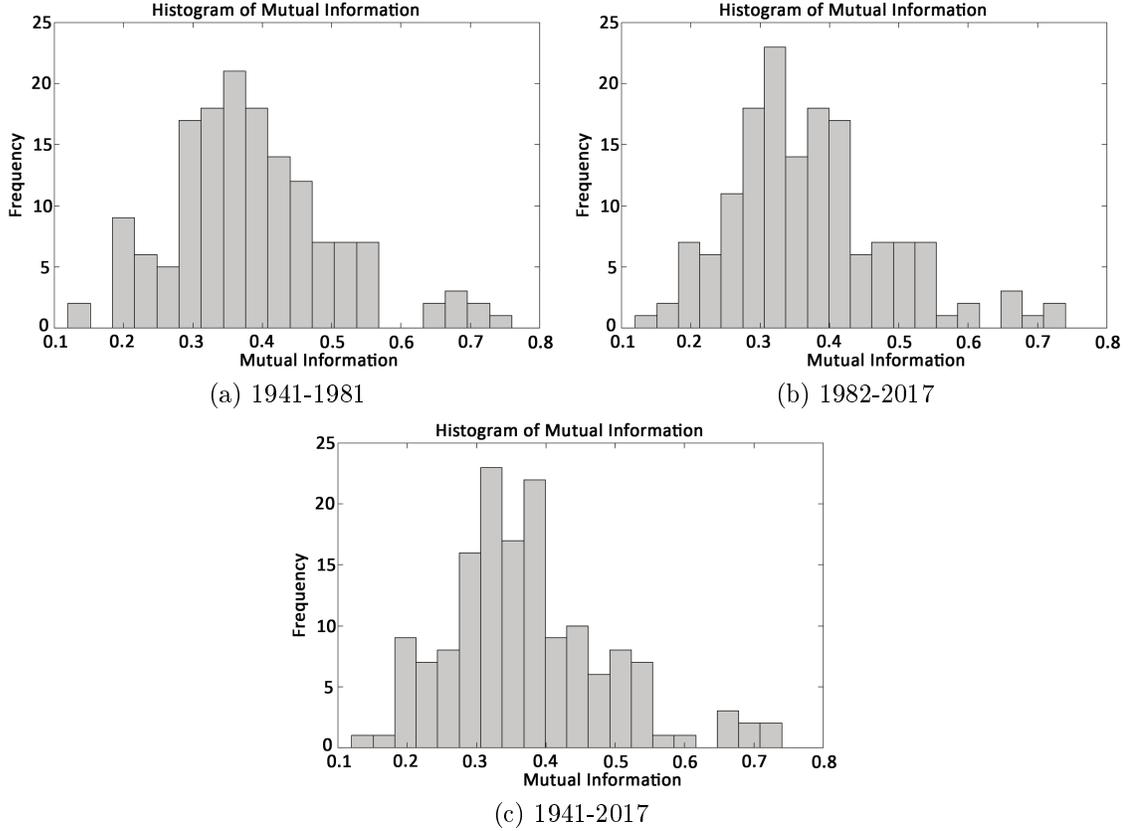


Figure 4.5: Calculated maximized mutual information between stations.

An undirected graph was created from the resulting similarity matrix where nodes correspond to stations and an edge represents a strictly positive mutual information between stations. If the mutual information between two stations was less than 0.1 it was assumed to be zero.

The normalized Laplacian and its eigenvalues and corresponding eigenvectors were calculated and sorted in ascending order. The sum of the eigenvalues $s(k) = \sum_{i=1}^k \lambda_k$ and the ratio $R(k) = \frac{s(k)}{s(h)}$ for $k = 1 \dots h$ were calculated and a cut off

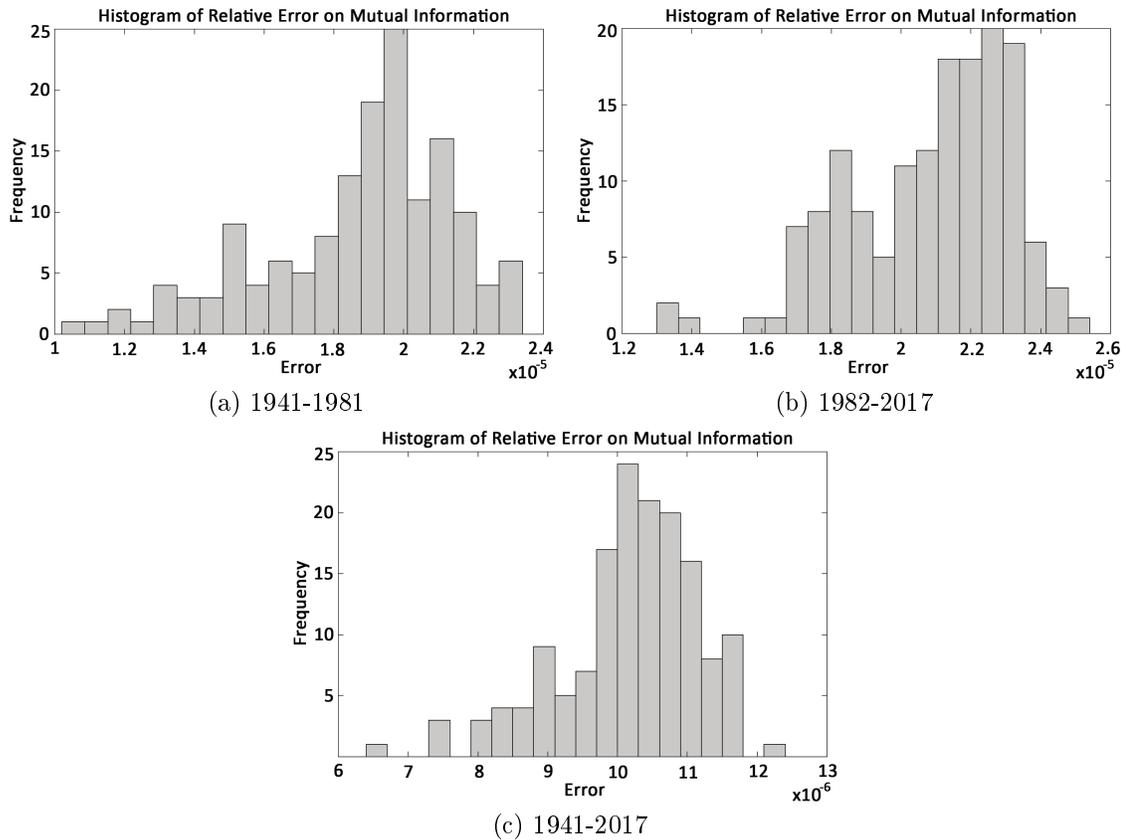
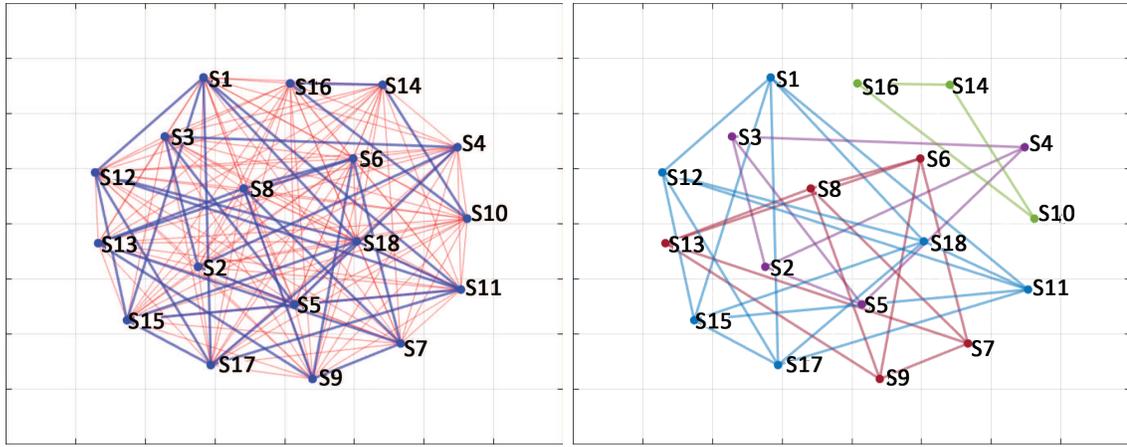


Figure 4.6: Calculated error on mutual information between stations after maximization.

point was chosen such that the number of nodes before this point are approximated $0.15 * h$ where h is the total number of nodes in the connected component. This resulted in a projection of each node (row vector in the normalized Laplacian) onto the $J = 3$ dimensional subspace of eigenvectors corresponding to the 3 smallest eigenvalues. These projections were used as the inputs into the K -means algorithm. See Figure 4.8 for eigenvalues and ratios. See Chapter 1, Section 1.6, for details on spectral clustering

Graph Representation of Station Similarity



(a)

(b)

Figure 4.7: Graph of nodes and edges where nodes represent stations and an edge represents a strictly positive mutual information. (a) The graph of full dimension where edges to be removed after spectral clustering are highlighted in red. (b) The reduced system after spectral clustering.

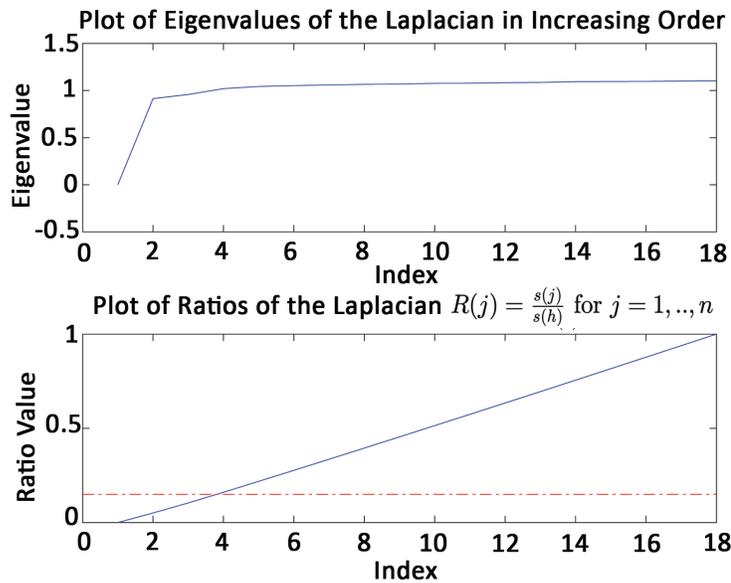


Figure 4.8: Ratio of eigenvalues and chosen cutoff.

The minimized cost values were collected for different values of K in the K -means algorithm (see Figures 4.9 and 4.10) and the value $K = 4$ was chosen as the appropriate number of clusters. The K -means algorithm was run 1,000 times on the reduced set of vectors and the Davies-Bouldin index was calculated for each run of the K -means algorithm. The set of clusters associated to the minimum Davies-Bouldin index was used in the following extreme value analysis. For more information see Chapter 1, Section 1.6.

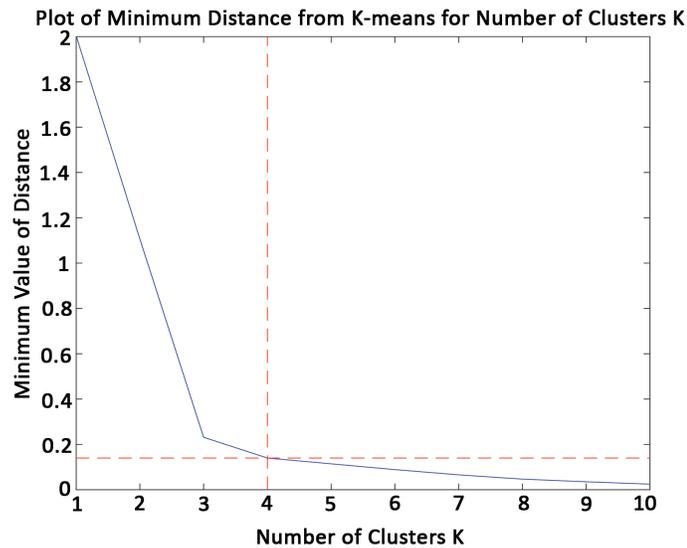


Figure 4.9: Minimized cost for different values of K .

Table 4.3 and Figure 4.13 reflect results of clustered stations associated to the minimum Davies-Bouldin index. The similarity matrix was created as an additional support of cluster reliability. Note that after reordering by cluster higher similarities should occur near the diagonal of the matrix with 4 squares along the diagonal representing each cluster (see Figures 4.12 and 4.11). Clusters are stable for each time window. Geographical locations of clusters are given by (1) Coastal (2) Southern

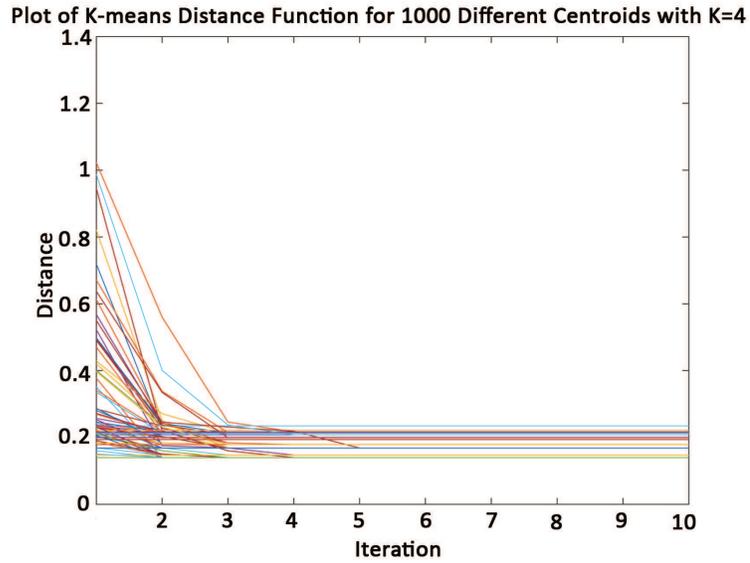


Figure 4.10: Minimized distances for each iteration of K -means $K = 4$.

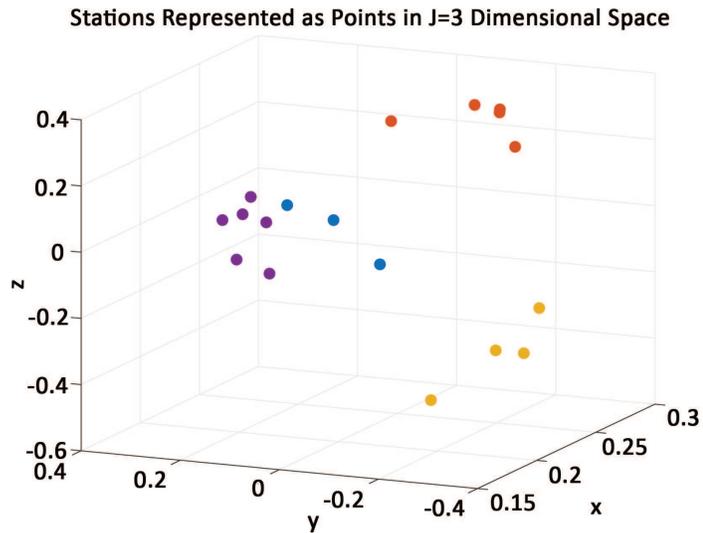


Figure 4.11: The $J = 3$ dimensional projection of the nodes into the space of spectral clustering. Colors represent different clusters.

Texas (3) Northern Texas (4) Along the Central Band.

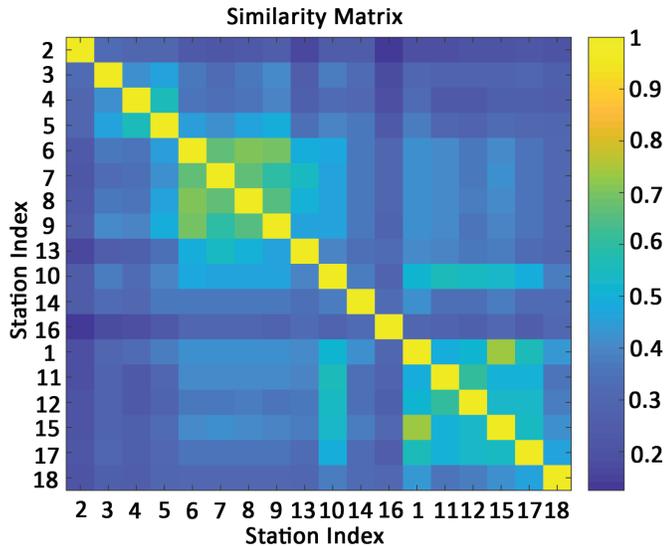


Figure 4.12: Similarity matrix for stations after clustering. Lighter colors indicate higher similarity.

Clusters for Each Time Period

time window	1941-1981	1982-2017	1941-2017
cluster (1)	2,3,4,5	"	"
cluster (2)	6,7,8,9,13	"	"
cluster (3)	1,11,12,15,17,18	"	"
cluster (4)	10,14,16	"	"
Davies-Bouldin index	7.75e-4	8.94e-4	6.34e-4

Table 4.3: Table of minimum Davies-Bouldin index clusters.

4.2.2.2 The Stationary Model

Just as in preliminary testing a chi-square test of independence was performed over blocks of 10 days with the result that every 10 day block reflects independence at the $\alpha = 0.05$ confidence level for each time series. (See Figure 4.14 for the histogram of p values.) The cluster of time series of maximum temperature is fit by maximum likelihood estimation to a generalized extreme value distribution with location parameter μ , scale parameter σ and shape parameter k . For clarification

Clustered Station Map

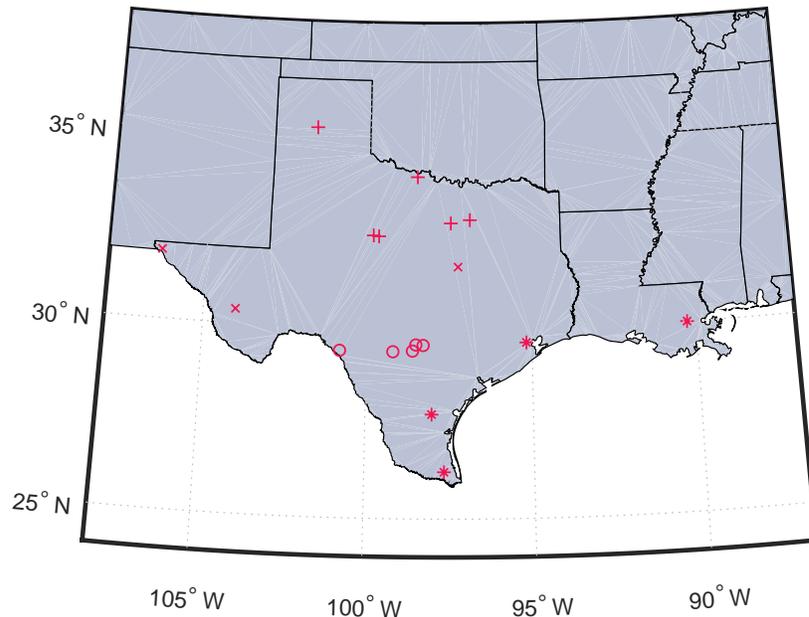


Figure 4.13: Regional map with clustered stations where symbols represent clusters.

we note that a single maximum is defined at the maximum over a 10-day block so that there are ~ 456 maxima for each time series and our distribution represents the *combined* maxima of all the time series in the cluster which is now ~ 456 times the number of time series in the cluster. MLE distributional parameters were used to create a theoretical GEV distribution. The Anderson-Darling (A-D) goodness of fit test was performed on the binned maxima with null hypothesis that the data fit the theoretical GEV distribution. Ten points were chosen uniformly within the 95% confidence intervals associated to each of the 3 MLE parameters and a total of 1,000 permutations of these 3 parameters were tested for goodness of fit. The histogram of 1,000 p values associated to each run of the A-D test for 1941-1981, 1982-2017

and 1941-2017 was used to determine whether a difference in GEV fit exists between each time interval.

Results from the Anderson-Darling goodness of fit test suggest a better fit for 1941-1981 and 1982-2017 when compared to 1941-2017 for all clusters (See Table 4.4). The histogram of p -values (See Figure 4.15) for clusters (1) and (2) show higher likelihoods of being below the $\alpha = 0.05$ confidence limit (e.g. conclude the data do not come from the specified distribution) for groups 1982-2017 and 1941-2017. Clusters (3) and (4) show higher likelihoods of being below $\alpha = 0.05$ confidence limit for 1941-2017.

For all clusters the GEV model with maximum likelihood estimates does not fit the whole of 1941-2017. Moreover, it is reasonable to conclude from A-D goodness of fit results that a change in parameters occurs between time intervals 1941-1981 and 1982-2017.

Goodness of Fit for Stationary Model

years	cluster(1)	cluster(2)	cluster(3)	cluster(4)
1941-1981	0.23	0.05	0.06	0.04
1982-2017	0.03	0.01	0.08	0.03
1941-2017	0.03	7.17e-4	0.01	6.21e-4

Table 4.4: Table of Anderson-Darling p -values associated to stationary GEV model of MLE parameters.

4.2.2.3 Location and Scale Trend Results

In this analysis the non-parametric Mann-Kendall and Theil-Sen and parametric linear regression tests were performed on the yearly mean and variance of the *combined* cluster of time series. A mean change-point analysis was also performed on the yearly

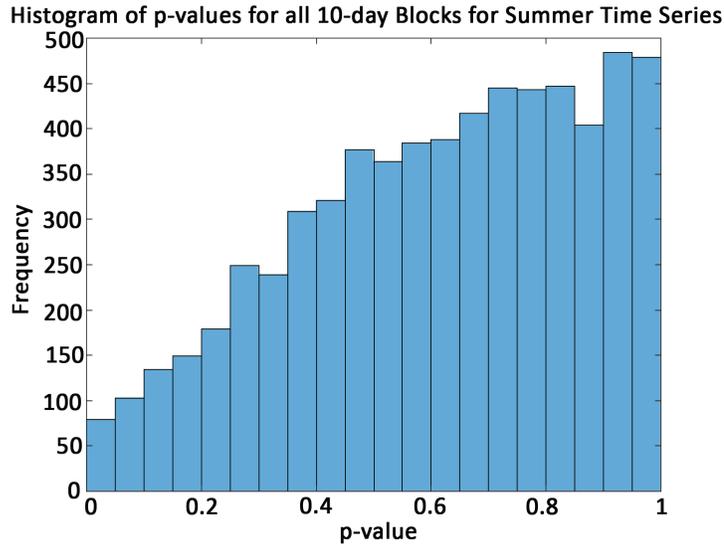


Figure 4.14: Chi-square independence p -values for (a) summer and (b) winter time series. The Houston station p -values referenced in the preliminaries is included.

mean and variance of the combined cluster of time series (See Table 4.5). These results suggest a change in the mean and variance parameters near 1981 for clusters (1), (2) and (4). It is important to note that the change-point analysis assumes at least one change point exists apriori which may not be the case. Moreover, it is unclear what constitutes a high or low residual error for this test. For more information on the statistics of these tests please refer to the Appendix A.2.2.

Cluster (4) was seen to have anomalous readings which was the result of an unusually large variance recorded from station 14. Since we could not trust the veracity of the temperature vector from this station it was removed from cluster (4) before analysis.

The Mann-Kendall test returns no monotonic trend in the variance over 1941-1981 and 1982-2017 for all clusters and a monotonic trend (e.g. a rejection of the null) for

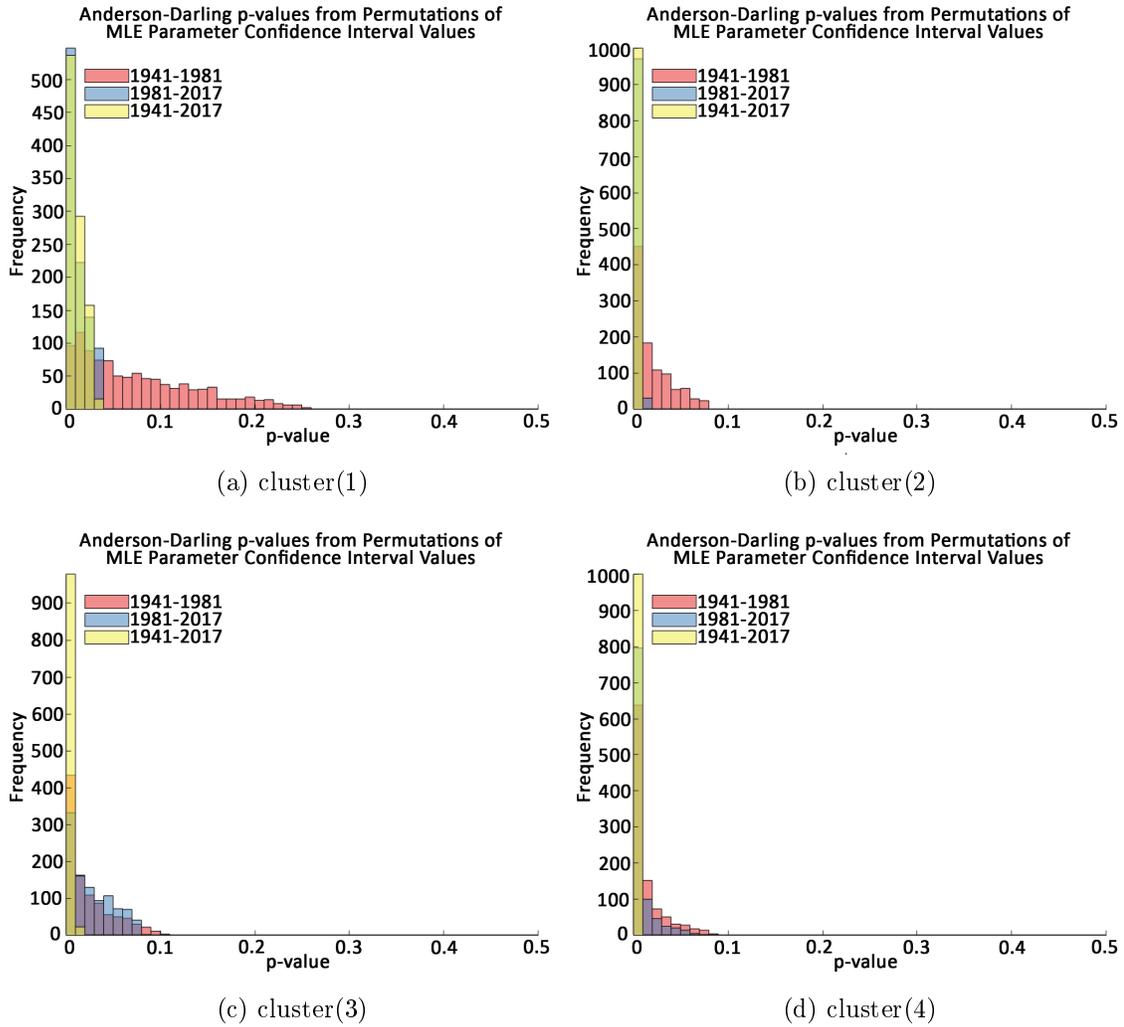


Figure 4.15: histogram of Anderson-Darling p -values obtained by varying the estimated parameters within the 95% confidence interval of the MLE parameters.

clusters (1) (2) and (4) over the time period 1941-2017 at the $\alpha = 0.05$ confidence level (See Table 4.7 and refer to A.2.2 in the Appendix for more information on this test.). Though cluster (3) does not report a monotonic trend for 1941-2017 it does return a lower p -value when compared to 1941-1981 and 1981-2017. Such a result suggests a significant difference in variance between 1941-1981 and 1982-2017.

Higher correlation coefficients returned by linear regression for 1941-2017 support conclusions from the Mann-Kendall test. Theil-Sen and linear regression models have similar values for slopes and intercepts. See Table 4.6. Results from all three tests provide motivation for fitting the variance parameters for 1941-2017 to a linear model.

Regression results for cluster (1) report sign changes between 1941-1981 and 1982-2017 which suggest a possible quadratic trend exists for the variance. Our justification for keeping a linear model for cluster (1) comes from the existence of a monotonic trend over 1941-2017 and all other clusters favoring a linear model.

The mean change point analysis checks at what point in time the mean of the yearly mean and the mean of the variance reflect a significant change. Change-point results for the yearly mean of clusters (1) (2) and (4) occurs between 1980 and 1990 and around 1960 for cluster (3). Arguably more interesting are the mean change-point results for the yearly variance of each cluster which report a change between 1970-1980 for clusters (1) (3) and (4) and around 1993 for cluster (2).

Mean Change Points

cluster	μ		σ^2	
	year	r	year	r
(1)	1983	147.40	1975	2032.69
(2)	1992	280.78	1993	1313.54
(3)	1957	321.86	1970	2418.32
(4)	1982	267.90	1971	17640.78

Table 4.5: Table of mean change points for the yearly mean and variance of each cluster.

Location Trend Results

cluster	(1)			(2)		
	1941-1981	1981-2017	1941-2017	1941-1981	1981-2017	1941-2017
years	1941-1981	1981-2017	1941-2017	1941-1981	1981-2017	1941-2017
M-K p -value	2.20e-3	1.4e-3	6.82e-5	0.01	0.13	0.55
T-S slope	-0.08	0.05	0.04	-0.08	0.05	0.01
REG slope	-0.07	0.06	0.04	-0.09	0.04	0.01
REG r	-0.46	0.44	0.45	-0.48	0.24	0.10

cluster	(3)			(4)		
	1941-1981	1981-2017	1941-2017	1941-1981	1981-2017	1941-2017
years	1941-1981	1981-2017	1941-2017	1941-1981	1981-2017	1941-2017
M-K p -value	0.03	0.32	0.74	0.29	0.92	3.61e-4
T-S slope	-0.09	0.04	4.3e-3	-0.03	3.3e-3	-0.04
REG slope	-0.07	0.05	3.82e-4	3.9e-3	0.01	-0.03
REG r	-0.35	0.26	3.9e-3	0.02	0.09	-0.32

Table 4.6: Mann-Kendall, Theil-Sen and Linear Regression results for yearly mean of the cluster.

Scale Trend Results

cluster	(1)			(2)		
	1941-1981	1981-2017	1941-2017	1941-1981	1981-2017	1941-2017
years						
M-K p -value	0.66	0.54	0.02	0.21	0.15	1.74e-4
T-S slope	0.03	-0.04	0.06	0.05	0.13	0.09
REG slope	0.01	-0.03	0.06	0.06	0.12	0.09
REG r	0.02	-0.05	0.25	0.18	0.22	0.43

cluster	(3)			(4)		
	1941-1981	1981-2017	1941-2017	1941-1981	1981-2017	1941-2017
years						
M-K p -value	0.55	0.71	0.36	0.14	0.75	1.26e-6
T-S slope	0.06	0.03	0.03	0.17	0.04	0.27
REG slope	0.09	0.05	0.03	0.27	-0.07	0.34
REG r	0.16	0.10	0.11	0.32	-0.04	0.41

Table 4.7: Mann-Kendall, Theil-Sen and Linear Regression results for yearly variance of the cluster.

4.2.2.4 The non-stationary Model

The non-stationary GEV model chosen is based on trend results and considers a quadratic location parameter $\mu(t) = \beta_0 + \beta_1 t + \beta_2 t^2$ and linear scale parameter $\sigma(t) = \alpha_0 + \alpha_1 t$ for each cluster of time series. Since the cluster of time series with $\mu(t)$ and $\sigma(t)$ would be under-determined for likelihood estimation, the estimation of parameters $\beta_0, \beta_1, \beta_2, \alpha_0, \alpha_1, k$ is given by MLE of the log-likelihood function $\ell(\mu(t), \sigma(t), k)$ for each station in the cluster. The location and scale parameters for the cluster are estimated by the following. For a distribution created from n samples of m distributions the mean of the distribution is given as,

$$\mu = \frac{\sum_{i=1}^m n \bar{x}_i}{mn}$$

where \bar{x}_i is the mean of the sample coming from one of the m distributions. The variance of the combined distribution is given as,

$$\sigma^2 = \frac{\sum_{i=1}^m n(s_i^2 + d_i^2)}{mn}$$

where s_i^2 is the variance of the sample and $d_i = \bar{x}_i - \mu$. From this observation the time-dependent location parameter for the cluster is taken to be,

$$\hat{\mu}(t) = \frac{\sum_{i=1}^m n\bar{x}(t)_i}{mn}$$

and the scale parameter

$$\hat{\sigma}(t) = \sqrt{\frac{\sum_{i=1}^m n(s_i^2(t) + d_i^2(t))}{mn}}$$

where n is the number of block maxima for each station and m is the number of stations in the cluster, $\bar{x}_i = \beta_0 + \beta_1 t + \beta_2 t^2$ is the location parameter of the i th station in the cluster with MLE parameters $\beta_{0,1,2}$ and $s_i = \alpha_0 + \alpha_1 t$ is the scale parameter of the i th station in the cluster with MLE parameters $\alpha_{0,1}$. Normalizing the block maxima for the cluster by subtracting the maxima for each station by $\hat{\mu}(t)$ and scaling by $\hat{\sigma}(t)$ allowed us to perform maximum likelihood estimation on the normalized block maxima and goodness of fit tests for the non-stationary model.

Anderson-Darling goodness of fit p -values before and after non-stationary modeling of the combined GEV for each cluster are given. See Table 4.8. Confidence intervals for the normalized location and scale parameters were checked for model credibility. Normalized GEV distributions with non-stationary location (quadratic) and scale (linear) parameters report significantly better fits from the stationary GEV distribution for all clusters. Moreover, clusters (1) and (3) conclude the normalized

block maxima come from the GEV distribution with MLE parameters at the $\alpha = 0.05$ significance level over all time intervals.

Goodness of Fit for Non-stationary Model

cluster	(1)		(2)		(3)		(4)	
	before	after	before	after	before	after	before	after
1941-1981	0.23	0.52	0.05	0.14	0.06	0.18	0.04	0.34
1982-2017	0.03	0.19	0.01	0.01	0.08	0.70	0.03	0.50
1941-2017	0.03	0.26	7.16e-4	4.60e-3	0.01	0.17	6.21e-4	0.50

Table 4.8: Anderson-Darling goodness of fit p -values before and after non-stationary fitting of the location and scale parameters.

4.2.3 Conclusions

These results demonstrate that temperature extremes for the years 1941 to 2017 in Texas are not stationary and that a non-stationary GEV model for the maximum of temperature values for each cluster of stations is needed. Particularly interesting are the differences in GEV fit between 1941-1981 and 1982-2017 which suggest significant differences in the mean and variance parameters. A comparison of GEV distributional fits for 1941 and 2017 reflect an increase in the probability for right-hand (higher) temperature extremes for all clusters particularly for temperature values greater than 100 degrees Fahrenheit. In fact, we see that the probability of occurrence in 2017 is double¹ that of 1941. See Tables 4.9, 4.10 and 4.11.

The GEV distributions generated for 1941 and 2017 based on parameter modeling are provided. Since the shape parameter of the distributions is assumed to be

¹ Where this is seen in the tails depends on the location of the pdf.

stationary the shape parameter estimated by MLE for 1941-2017 was used to generate the GEV PDFs for 1941 and 2017. (See Figure 4.16). Note that, in general, if variance parameters of the cluster change a significant amount only a small amount of change is seen in the mean (see clusters (2) and (4)). Conversely, if a small amount of change is seen in the variance a significant amount of change is seen in the mean (see clusters (1) and (3)). We observe that the most prominent example of monotonic trend in variance occurs in cluster (2) while the most prominent example of monotonic trend in mean occurs in cluster (1).

It would be interesting and useful to adapt cluster techniques to analyze temperature extreme phenomena such as heatwaves or successive days over a threshold temperature, as in [66][87]. A related investigation would be the probability of excessive precipitation and how it relates to temperature extremes (as considered in [84]). This question is very timely given the unprecedented flooding seen in Houston, Texas, in August 2017. These projects are the subject of future work.

Location and Scale Comparison

values	μ		σ	
cluster	1941	2017	1941	2017
(1)	95.18	96.70	13.76	13.28
(2)	97.58	97.81	10.13	18.13
(3)	97.98	99.09	18.07	18.84
(4)	97.95	98.45	12.66	13.71

Table 4.9: Location and scale comparisons for 1941 and 2017 GEV models.

Extreme Probability Distribution Comparisons

values	$P(X \geq 100)$		$P(X \geq 105)$		$P(X \geq 107)$		$P(X \geq 110)$	
cluster	1941	2017	1941	2017	1941	2017	1941	2017
(1)	0.07	0.09	0.01	0.02	2.20e-3	6.80e-3	0.0 1.00e-4	7.00e-4
(2)	0.11	0.09	0.02	0.04	4.40e-3	0.021	0	4.80e-3
(3)	0.09	0.09	0.04	0.06	0.02	0.03	5.50e-3	1.13e-2
(4)	0.10	0.10	0.03	0.04	0.01	0.02	1.00e-3	2.90e-3

Table 4.10: Probability distribution comparisons for 1941 and 2017 GEV models.

Extreme Probability Comparisons

2017/1941	$P(X \geq 100)$	$P(X \geq 105)$	$P(X \geq 107)$
(1)	1.32	1.89	3.09
(2)	0.80	1.88	4.86
(3)	0.99	1.30	1.57
(4)	0.99	1.26	1.51

Table 4.11: Ratios of probability comparisons for 1941 and 2017 GEV models.

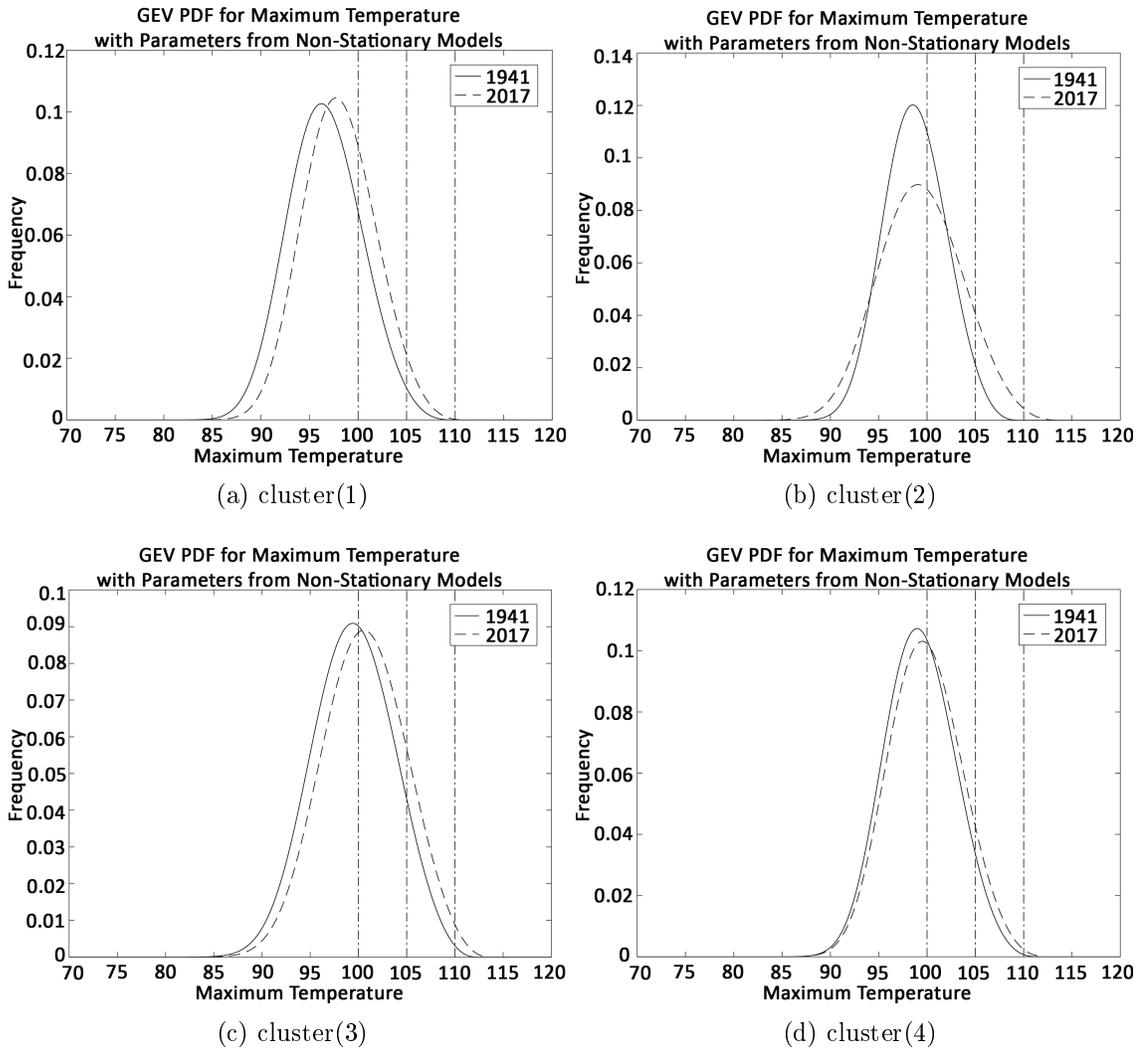


Figure 4.16: Non-stationary generalized extreme value pdfs.

4.3 A brief discussion of winter temperature extremes throughout the Texas region

4.3.1 Data, Clusters and Preliminaries

In this section a *time series* is similarly defined to that in section 4.2 with the replacement of December-January measurements.

The previously outlined clustering analysis was performed on the winter time series from each station. Results for the winter differ only by the movement of station 10 from cluster (4) to cluster (3) 4.12. Clusters are stable over all time window.

Clusters for Each Time Period for Winter Data

time window	1941-1981	1982-2017	1941-2017
cluster (1)	2,3,4,5	"	"
cluster (2)	6,7,8,9,13	"	"
cluster (3)	1,10,11,12,15,17,18	"	"
cluster (4)	14,16	"	"
Davies-Bouldin index	7.75e-4	8.94e-4	6.34e-4

Table 4.12: Table of minimum Davies-Bouldin index clusters.

4.3.2 Generalized Extreme Value Distribution

All time series reported independence across 10 day blocks from the chi-square test of independence at the $\alpha = 0.05$ significance level (see Figure 4.17). (See Table 4.13 for the histogram of p values.) The 10 day block maximum was calculated. Maximum likelihood estimation to fit a generalized extreme value distribution was performed

on the resulting time series of maximum values for each cluster. Anderson-Darling goodness of fit tests report poor fits over all clusters for time window 1941-2017 with poor fits over time windows 1941-1981 and 1981-2017 for clusters (1), (2) and (3). Comparing QQ plots (Figure 4.20) suggest a poor fit for all clusters in the tails of the distributions, in particular at the lower end of the probability distribution where block maximum values report unusually low records (see Figure 4.20).

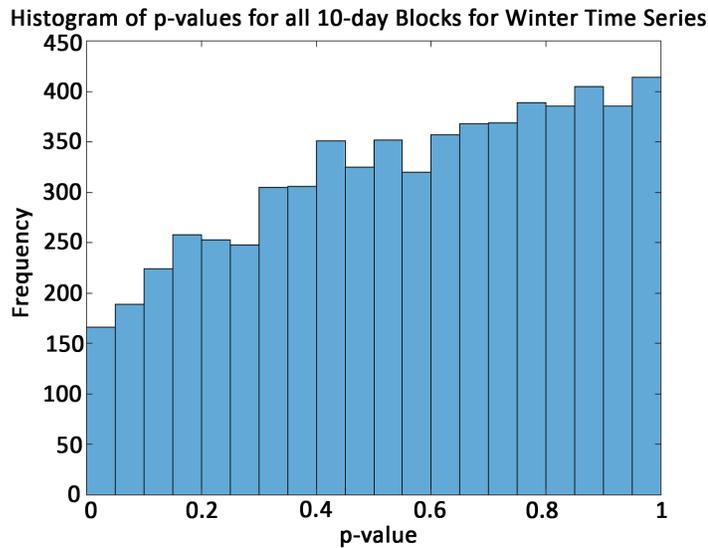


Figure 4.17: Chi-square independence p -values for winter time series.

Goodness of Fit for the Stationary Model

years	cluster(1)	cluster(2)	cluster(3)	cluster(4)
1941-1981	1.11e-4	1.48e-3	6.66e-7	0.75
1982-2017	5.59e-3	1.11e-4	7.09e-4	0.03
1941-2017	1.03e-6	3.10e-7	2.29e-7	0.02

Table 4.13: Table of Anderson-Darling p -values associated to stationary GEV model of MLE parameters.

Nonstationarity in the parameters is assumed to be the primary cause of poor distributional fits. This hypothesis is tested against the trend tests used in analysis

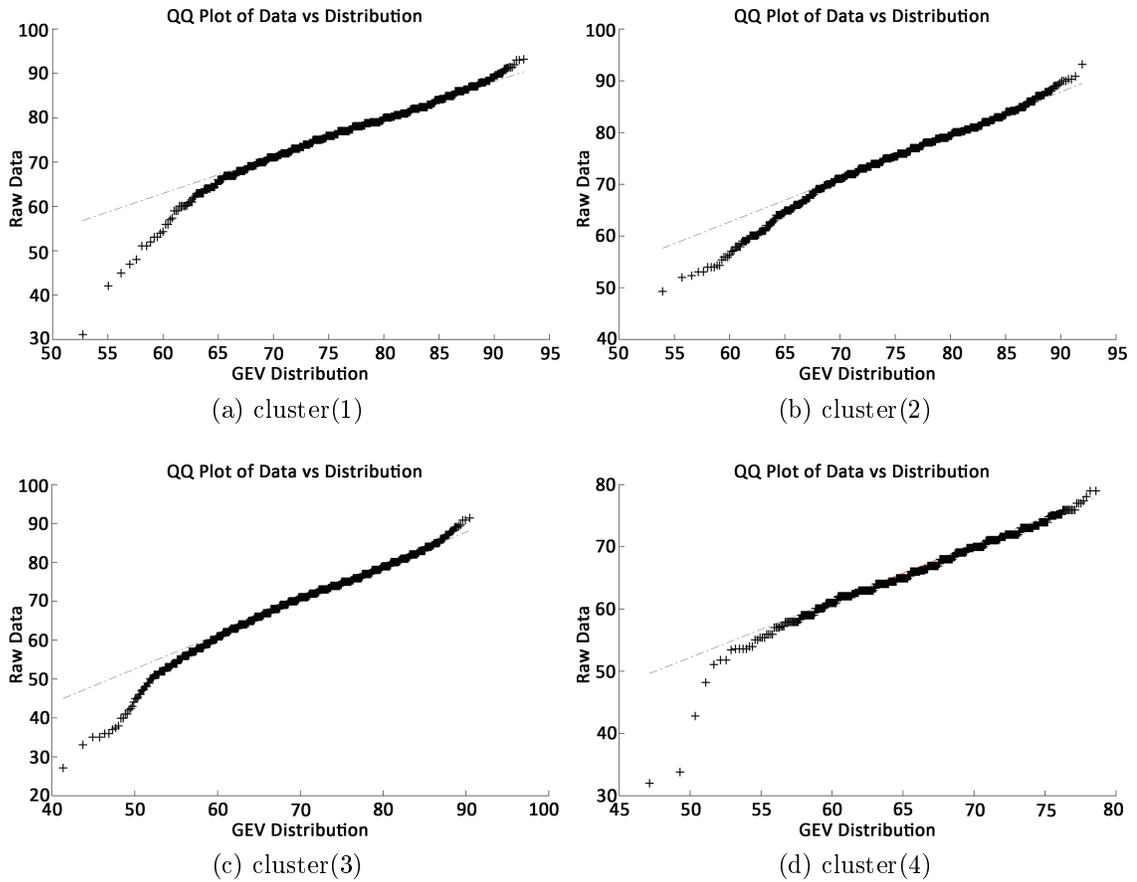


Figure 4.18: QQ plots for the true block maximum and the maximum likelihood estimated generalized extreme value distribution for winter temperature extremes for 1941-2017.

for summer data. See Table 4.16. Non-parametric Mann-Kendall and Theil-Sen were performed on the yearly mean and variance of the cluster of time series for each time window. Mann-Kendall results suggest no monotonic trend exists in the mean for the whole of 1941-2017; however a monotonic trend is observed over the time windows 1941-1980 and 1981-2017 for all aside from cluster (4). Theil-Sen values for the mean over time windows 1941-1980 and 1981-2017 change signs and point towards a quadratic trend in the mean. All tests confirm no trend in the yearly

variance for every cluster over every time window.

Based on trend results for the yearly mean and variance the non-stationary GEV model was calculated using maximum likelihood estimation for each cluster with stationary variance and quadratic mean models. Likelihood ratio results suggest no significant difference between the stationary and non-stationary quadratic mean model for most stations within the clusters. (See Table 4.15.) Anderson-Darling results confirm this analysis with poor fits for the non-stationary GEV model. (See Table 4.14.)

Goodness of Fit for the Winter Non-Stationary Model

cluster	(1)		(2)		(3)		(4)	
year	before	after	before	after	before	after	before	after
1941-2017	1.03e-6	1.26e-6	3.10e-7	3.10e-7	2.29e-7	2.29e-7	2.43e-2	2.47e-2

Table 4.14: A-D goodness of fit p values before and after non-stationary modeling.

Log-likelihood Comparisons

cluster	(1)				(2)				
station	2	3	4	5	6	7	8	9	13
p value	0.05	1.36e-4	0.06	0.06	0.09	0.07	0.58	0.86	2.95e-5

cluster	(3)							(4)	
station	1	10	11	12	15	17	18	14	16
p value	2.9e-3	0.01	0.93	0.87	0.61	0.11	0.53	0.06	0.06

Table 4.15: Log likelihood p values for comparing stationary and non-stationary models.

Winter Location and Scale Trend Results

cluster	(1)					
parameter	μ			σ		
years	1941-1981	1981-2017	1941-2017	1941-1981	1981-2017	1941-2017
M-K p -value	2.36e-4	6.58e-5	0.02	0.45	0.04	0.31
T-S slope	-0.15	0.12	0.03	0.19	-0.51	0.08

cluster	(2)					
parameter	μ			σ		
years	1941-1981	1981-2017	1941-2017	1941-1981	1981-2017	1941-2017
M-K p -value	0.19	6.2e-3	0.28	0.45	0.12	0.61
T-S slope	-0.06	0.13	0.02	0.10	-0.37	-0.04

cluster	(3)					
parameter	μ			σ		
years	1941-1981	1981-2017	1941-2017	1941-1981	1981-2017	1941-2017
M-K p -value	0.53	1.6e-4	0.40	0.50	0.82	0.85
T-S slope	-0.04	0.22	0.01	0.27	-0.07	0.03

cluster	(4)					
parameter	μ			σ		
years	1941-1981	1981-2017	1941-2017	1941-1981	1981-2017	1941-2017
M-K p -value	0.73	0.26	0.15	0.49	0.54	0.21
T-S slope	-0.02	0.04	0.02	0.14	0.11	0.08

Table 4.16: Mann-Kendall and Theil-Sen results for yearly mean and variance of the cluster.

4.3.3 Generalized Pareto Distribution

For completion of exposition we consider the peaks over threshold (POT) method where the maximum values are fit by a generalized Pareto distribution (GP) with scale parameter σ and shape parameter k . The POT approach allows nonstationarity of the location parameter to be ruled out in extreme value analysis [?].

For each cluster a threshold was chosen by looking at the mean residual life plots (see Figure 4.19) so that a linear trend is observed at the start of the threshold. A

maximum value of the time series is defined as any value which exceeds the set threshold. The generalized Pareto distribution is fit by maximum likelihood estimation of the log-likelihood function,

$$l_{\sigma,k} = -n \log \sigma - (1 + 1/k) \sum_{i=1}^n \log(1 + k \frac{x_i}{\sigma})$$

for the scale σ and shape k parameters to the maxima of the whole cluster of time series.

Although the scale parameter can be assumed to be stationary from previous results, Anderson-Darling tests continue to reflect poor fits for the GPD. (See Table 4.17.)

Goodness of Fit for the Winter Stationary Pareto Model

years	cluster(1)	cluster(2)	cluster(3)	cluster(4)
1941-2017	1.36e-6	1.54e-6	7.41e-6	1.63e-6

Table 4.17: Table of Anderson-Darling p -values associated to stationary GPD model of MLE parameters.

4.3.4 Conclusions

Even after accounting for dependence and nonstationarity, both extreme value distribution models reflect poor fit results for the maxima of each cluster of time series. In view of the unusually low maximum values observed in each cluster (see Figure 4.20) we note that it is possible that the temperature throughout Texas is affected by Canadian winds during winter which may lead to a mixed distribution for the winter extremes of the time series.

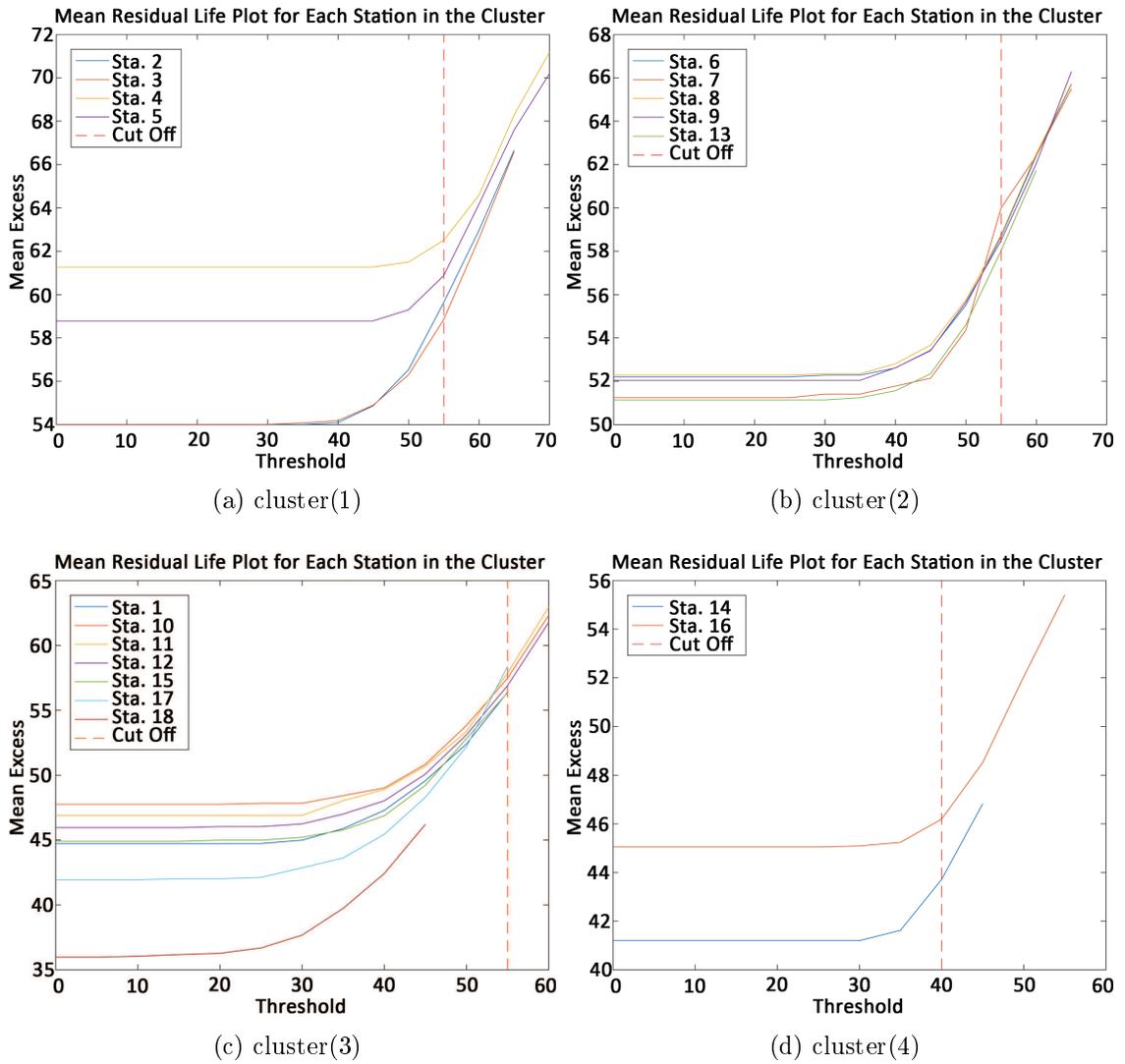


Figure 4.19: Mean residual life plots for each cluster. Chosen threshold is marked with a red dashed line.

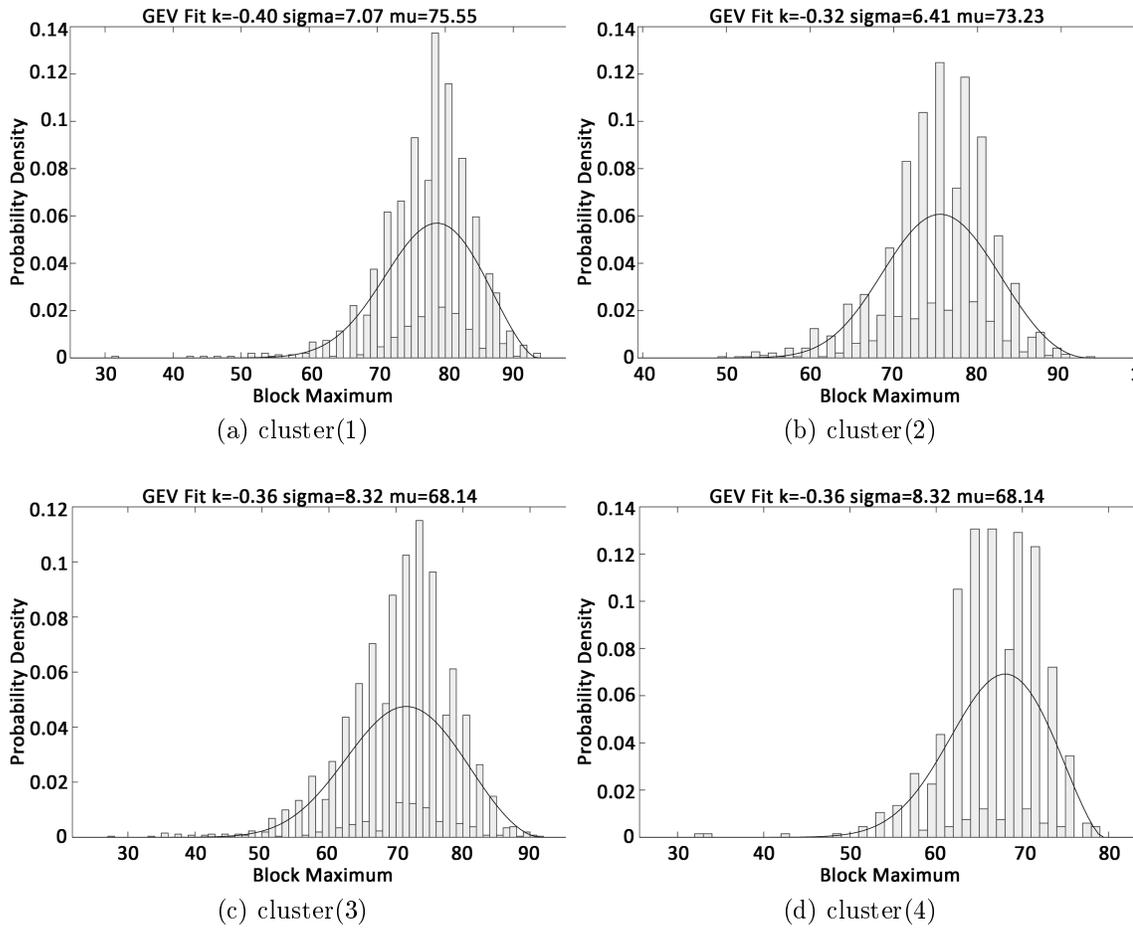


Figure 4.20: Generalized extreme value distribution for the winter with MLE parameters for 10-day maximum temperatures for 1941-2017 probability density function.

4.4 Concluding Remarks

We have established statistical properties for certain chaotic dynamical systems using information about their extreme values and recurrence probabilities. We have shown, under certain ergodic and regularity assumptions, growth rates on the Birkhoff sum can be obtained where the Birkhoff ergodic theorem fails. In a probabilistic setting, this is comparable to obtaining rates on the growth of the time average for systems with no strong law of large numbers. These results hold for a large number of chaotic systems. By using example chaotic maps, we included numerical estimations as support for the established theorems. These numerical estimations were obtained through simulations created in code.

We have extended previous results to more complex, physically relevant systems such as the Sinai dispersing billiard model. One of the theorems we have proven shows the existence of a maximal probability distribution. When used in practice, this result provides a way of estimating the probability that a maximum over a certain threshold will occur in this system. We have also proven the existence of a rare event point process, which provides the probabilities of returns of these maximums. Motivated by local climate concern, this discussion also included an analysis on summer and winter temperature extremes throughout Texas from 1941-2017. This project required the implementation of relatively new and sophisticated machine learning techniques. These techniques provided a way of clustering stations throughout Texas so that regional conclusions could be made about the data. We were then able to use results from extreme value theory (with data-specific modifications) to model

the maximum temperatures for the cluster. We found compelling evidence that the probability of observing higher summer temperature maximums has increased in the last 40 years. Modeling these changes in extremes allows us to better predict and take action against a changing climate.

We used a combination of theoretical, numerical, and applied approaches of mathematics to investigate the statistical properties of extremes in a variety of chaotic systems. The theoretical work performed here will allow us to build more sophisticated analytical tools for the future; while the applied work allows us to make conclusions about what is happening in the present. With continued effort, using combined approaches to problems, we can help shape the future in a more positive way for the next generation.

Appendix A

Appendix

The appendix contains supplementary material.

A.1 Supporting Dynamical Definitions and Proofs

A.1.1 Descriptions of relevant dynamical maps

Anosov diffeomorphisms

Let M be a compact Riemannian manifold without boundary, with metric d on M derived from the Riemannian metric. For $x \in M$ let $T_x M$ denote the tangent space at x and let TM denote the tangent bundle. Let $T : M \rightarrow M$ be a C^1 diffeomorphism of M and let $D_x T : T_x M \rightarrow T_{T(x)} M$ denote the derivative of T . We say that $T : M \rightarrow M$ is an *Anosov diffeomorphism* if the tangent bundle TM has a continuous

splitting into a direct sum of two DT -invariant sub-bundles $TM = E^s \oplus E^u$ such that there exists constant $C > 0$ and $\lambda \in (0, 1)$ such that for all $n \geq 0$,

$$\|D_x T^n v\| \leq C\lambda^n \|v\|, \quad \forall v \in E_x^s,$$

$$\|D_x T^{-n} v\| \leq C\lambda^n \|v\|, \quad \forall v \in E_x^u.$$

We call E^s and E^u the stable and unstable sub-bundles, respectively.

Uniformly expanding C^2 maps of the interval

Let $([0, 1], d)$ be a metric space and $f \in C^2([0, 1])$ then $f : [0, 1] \rightarrow [0, 1]$ is called expanding if there exists $\lambda > 1$ such that for any $x, y \in [0, 1]$,

$$d(f(x), f(y)) \geq \lambda d(x, y)$$

Gibbs-Markov maps

Let (X, \mathcal{B}, m) be a Lebesgue probability space with $X \subset \mathbb{R}$. Let \mathcal{P} be a countable measurable partition of X such that $m(\alpha) > 0$ for all $\alpha \in \mathcal{P}$.

A measure-preserving map $T : X \rightarrow X$ is said to be a *Markov map* if the following are satisfied.

1. \mathcal{P} generates \mathcal{B} under T
2. (Markov property) For all $\alpha, \beta \in \mathcal{P}$, if $m(T(\alpha) \cap \beta) > 0$ then $\beta \subset T(\alpha)$.
3. (local invertibility) For all $\alpha \in \mathcal{P}$, $T|_\alpha$ is invertible.

For integer $n \geq 0$, let \mathcal{P}_n be the partition of X defined by

$$\mathcal{P}_n = \bigvee_{i=0}^{n-1} T^{-i}(\mathcal{P})$$

Let $J_T = \frac{d(m \circ T)}{dm}$.

The quintet $(X, \mathcal{B}, m, T, \mathcal{P})$ is said to be a *Gibbs-Markov system* if T is a Markov map and the following properties also hold.

1. (full branches) For all $\alpha \in \mathcal{P}$, $T(\alpha) = X$, mod m .
2. (uniform expansion) There exists $K_1 > 0$ and $\gamma_1 \in (0, 1)$ such that $m(\alpha) \leq K_1 \gamma_1^n$ for all $n \geq 0$ and $\alpha \in \mathcal{P}_n$.
3. (distortion control) There exists $K_2 > 0$ and $\gamma_2 \in (0, 1)$ such that for all $n \geq 0$ and $\alpha \in \mathcal{P}_n$, we have

$$\left| \log \left(\frac{J_{T^n}(x)}{J_{T^n}(y)} \right) \right| \leq K_2 \gamma_2^n \tag{A.1}$$

for all $x, y \in \alpha$.

Remark A.1.1. Some authors weaken the full-branch condition in the definition of Gibbs-Markov systems by requiring merely that $m(T(\alpha)) > K > 0$ for some K independent of α .

Intermittent type maps

The family of intermittent maps can be described as maps which are expanding everywhere except at a neutral fixed point, where hyperbolicity is lost; where the local behavior of the map differs.

A.1.2 Dynamical Definitions

Bounded Variation

Let $I \subset \mathbb{R}$ be an interval and consider the collection Π of ordered points $a_1 < a_2 < \dots < a_{N+1} \in I$, where $N \in \mathbb{N}$. The total variation of f is given by,

$$TV(f) := \sup \left\{ \sum_{i=1}^N |f(a_{i+1}) - f(a_i)| : (a_1, \dots, a_{N+1}) \in \Pi \right\}$$

Let $I \subset \mathbb{R}$ be an interval. A function $f : I \rightarrow \mathbb{R}$ is said to have bounded variation if its total variation TV is bounded.

Young towers

Young Tower with exponential return time tails for a map $T : M \rightarrow M$ of a Riemannian manifold M equipped with Lebesgue measure m .

A Young Tower has a base set Δ_0 with a hyperbolic product structure as in Young [96] with an $\mathcal{L}^1(m)$ return time function $R : \Delta_0 \rightarrow \mathbb{N}$. There is a countable partition $Y_{0,i}$ of Δ_0 so that R is constant on each partition element $Y_{0,i}$. We denote $R|_{Y_{0,i}}$ by R_i . The Young Tower is defined by

$$\Delta = \bigcup_{i \in \mathbb{N}, 0 \leq l \leq R_i - 1} \{(x, l) : x \in \Lambda_{0,i}\}$$

equipped with a tower map $F : \Delta \rightarrow \Delta$ given by

$$F(x, l) = \begin{cases} (x, l + 1) & \text{if } x \in Y_{0,i}, l < R_i - 1 \\ (T^{R_i}x, 0) & \text{if } x \in Y_{0,i}, l = R_i - 1 \end{cases}.$$

We will refer to $\Delta_0 := \cup_i(Y_{0,i}, 0)$ as the base of the tower Δ and denote $Y_i := Y_{0,i}$. Similarly we call $\Delta_l = \{(x, l) : l < R(x)\}$, the l th level of the tower. A key role is played by the return map $f = T^R : \Delta_0 \rightarrow \Delta_0$ by $f(x) = T^{R(x)}(x)$, which is uniformly expanding Gibbs-Markov.

Definition of hyperbolic

Let M be a compact Riemannian manifold (perhaps, with boundary and corners), $N \subset M$ an open and dense subset and $F: N \rightarrow M$ a C^r (with $r \geq 2$) diffeomorphism of N onto $F(N)$. Note that all the iterations of F are de

ned on the set

$$\tilde{N} = \bigcap_{n=-\infty}^{\infty} F^n(N).$$

Assume that F preserves a probability measure μ on M and $\mu(\tilde{N}) = 1$.

Theorem A.1.2. [23, Theorem 3.1] *Suppose*

$$\int_M \log^+ \|D_x F\| d\mu(x) < \infty \quad \text{and} \quad \int_M \log^+ \|D_x F^{-1}\| d\mu(x) < \infty,$$

where $\log^+ s = \max\{\log s, 0\}$. Then there exists an F -invariant set $H \subset \tilde{N}$, $\mu(H) = 1$ such that for all $x \in H$ there is a DF -invariant decomposition of the tangent space

$$\mathcal{T}_x M = E_x^{(1)} \oplus \dots \oplus E_x^{(m)}$$

with some $m = m(x)$, such that for all nonzero vectors $v \in E_x^{(i)}$

$$\lim_{n \rightarrow \pm\infty} \frac{1}{n} \log \|D_x F^n v\| = \lambda_x^{(i)}$$

where $\lambda_x^{(1)} > \dots > \lambda_x^{(m)}$.

The values of $\lambda_x^{(i)}$ are called *Lyapunov exponents* of the map F at the point x .

Definition A.1.3. A point $x \in M$ is said to be **hyperbolic** if Lyapunov exponents exists at x and none of them equals zero.

A.1.3 Proof of $\mathbb{D}_q(u_n)$ for 2-D hyperbolic maps with singularities

The scheme of this proof comes from [13]. Recall the definition of a young tower A.1.2. We may form a quotiented tower (see [96] for details) by introducing an equivalence relation for points on the same stable manifold. This operation helps in our decay of correlations estimates, as it allows decay rates for the indicator function of complicated sets to be estimated in the L^∞ norm. We now list the features of the Tower that we will use.

There exists an invariant measure m_0 for $f : \Delta_0 \rightarrow \Delta_0$ which has absolutely continuous conditional measures on local unstable manifolds in Δ_0 , with density bounded uniformly from above and below.

There exists an F -invariant measure ν on Δ which is given by $\nu(B) = \frac{m_0(F^{-l}B)}{\int_{Y_0} R dm_0}$ for a measurable $B \subset Y_l$, and extended to the entire tower Δ in the obvious way. There is a projection $\pi : \Delta \rightarrow M$ given by $\pi(x, l) = T^l(x)$ which semi-conjugates F and T , so that $\pi \circ F = T \circ \pi$. The invariant measure μ , which is an SRB measure for $T : M \rightarrow M$, is given by $\mu = \pi_* \nu$. Denote by $W^s(x)$ the local stable manifold through x and by $W^u(x)$ the local unstable manifold.

Let $B(x, r)$ denote the ball of radius r centered at the point x .

We lift a function $\phi : M \rightarrow \mathbb{R}$ to Δ by defining $\phi(x, l) = \phi(T^l x)$ (we keep the same symbol for ϕ on Δ and ϕ on M).

Under the assumption of exponential tails, that is if $m(R > n) = \mathcal{O}(\gamma_1^n)$ for some $0 < \gamma_1 < 1$ then Young shows [96] there exists $0 < \Lambda_1 < 1$ such that for all Lipschitz ϕ, ψ we have

$$\left| \int \phi \cdot \psi \circ T^n d\mu - \int \phi d\mu \int \psi d\mu \right| \leq C \Lambda_1^n \|\phi\|_{Lip} \|\psi\|_{Lip} \quad (\text{A.2})$$

for some constant C . Moreover, if the lift of ψ is constant on local stable leaves of the Young Tower, then

$$\left| \int \phi \cdot \psi \circ T^n d\mu - \int \phi d\mu \int \psi d\mu \right| \leq C \Lambda_2^n \|\phi\|_{Lip} \|\psi\|_\infty. \quad (\text{A.3})$$

Let D be a set whose boundary is piecewise smooth and finite length, and define

$$H_k(D) = \{x \in D : T^k(W^s(x)) \cap \partial D \neq \emptyset\}.$$

Proposition A.1.4. *There exist constants $C > 0$ and $0 < \tau_1 < 1$ such that, for all k ,*

$$\mu(H_k(D)) \leq C \tau_1^k. \quad (\text{A.4})$$

Proof. As a consequence of the uniform contraction of local stable manifolds, there exists $0 < \tau_1 < 1$ and $C_1 > 0$ such that $d(T^n(x), T^n(y)) \leq C_1 \tau_1^n$ for all $y \in W^s(x)$. In particular, this implies that $|T^k(W^s(x))| \leq C_1 \tau_1^k$. Therefore, for every $x \in H_k(D)$,

the stable manifold $T^k(W^s(x))$ lies in an "annulus" of width $2C_1\tau_1^k$ around ∂D . By the invariance of μ the result follows.

□

A.2 Supporting Statistical Information

A.2.1 Relevant derivations

Total Error on the Mutual Information

The number of values z_k in the k th state of our newly compressed time series can be seen as binomally distribution with $E(z_k) = Np_k$ and variance $V(z_k) = Np_k(1 - p_k)$ where the true probability of being in the k th state is given by p_k and N is the length of the time series. Then the error $\tilde{\epsilon}_k$ on the estimated proportion $\hat{p}_k = z_k/N$ has normal distribution with $E(\tilde{\epsilon}_k) = 0$ and $V(\tilde{\epsilon}_k) = \hat{p}(1 - \hat{p})/N$ for large enough N . The relative error size ϵ_k on $\hat{p}_k \log \hat{p}_k$ is taken as

$$\epsilon_k = \sum_k \frac{\delta(\hat{p}_k \log \hat{p}_k)}{\delta \hat{p}_k} \tilde{\epsilon}_k$$

but $E(\epsilon_k) = 0$ so,

$$V(\epsilon_k) = \sum_k (1 + \log \hat{p}_k)^2 \frac{\hat{p}_k(1 - \hat{p}_k)}{N}$$

Note that by choosing threshold values resulting in a uniform distribution for z_k , we have that the $V(\epsilon_k) \approx 0$ (entropy is maximized) for each time series so that we take the relative error to be the $V(\epsilon_k)$ taken over the approximate joint entropy.

A.2.2 Description of statistical tests

Mean change-point analysis

The algorithm employed by the MATLAB function

```
findchangepts.m
```

performs the following steps to detect the mean change-point:

- Chooses a point and divides the signal (time series) into two sections.
- Computes an empirical estimate of the mean for each section.
- At each point within a section, measures how much the property deviates from the empirical estimate. Adds the deviation for all points.
- Adds the deviations section-to-section to find the total residual error.
- Varies the location of the division point until the total residual error attains a minimum.

Chi-square test of independence

The chi-square test of independence tests two blocks of categorical variables with null hypothesis that the two data blocks are independent and alternative that the two data blocks are dependent. The test statistic is given by,

$$\chi^2 = \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}$$

where

$$E_{i,j} = \frac{\sum_{i=1}^c O_{i,j} \sum_{k=1}^r O_{k,j}}{N}$$

is the expected value, O is a matrix where the first column corresponds to the first variable and the second column to the other. For purposes of the analysis laid out in Chapter 4, the first block of data corresponds to the compressed block of 10 days (240 data points) the second to another compressed block of 10 days. The blocks are compared within a single station and across all stations to determine independence.

Anderson-Darling test for goodness of fit

The Anderson-Darling has a test statistic given by,

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n (2i - 1) \cdot [\log F(x_i) + \log(1 - F(x_{n-i+1}))]$$

where F is value of the empirical cumulative distribution function (CDF) taken at the value of the ordered data x . The Anderson-Darling test calculates the weighted sum of the distance between the empirical and theoretical CDF multiplied by the weight function $[F(x)(1 - F(x))]$ so that the greatest value occurs in the tails of the distribution. Anderson-Darling tests the null hypothesis is that the data follow the specified distribution against the alternative that the data does not follow the specified distribution.

Non-parametric Mann-Kendall test of trend

The Mann-Kendall tests a null hypothesis of no monotonic trend and an alternative that there exists a monotonic trend. Cyclic changes in the data will not be detected

in this test. The test statistic is given by,

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{sgn}(x_j - x_k)$$

with

$$\text{sgn}(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases}$$

where in our case, x is the variance or mean vector calculated per year. As such, the results of the mann-kendall test are not impacted by the magnitude of extreme values. (Note that each data point is compared with all preceding data points so that there are $n(n - 1)/2$ comparisons.)

Non-parametric Theil-Sen test of trend

The Theil-Sen estimator m is determined by taking the median of the slopes between all pairs of sample points (x_i, y_i) . The regression line is created by setting the setting the intercept b to be the median of the values $y_i - mx_i$.

Parametric regression test of trend

Results for the slope m , intercept b and correlation coefficient r are given by linear least squares estimation of each point where $-1 \leq r \leq 1$.

Bibliography

- [1] J. Aaronson. On the ergodic theory of non-integrable functions and infinite measure spaces. *Israel J. Math*, 27(2):163–173, 1977.
- [2] J. Aaronson, Z. Kosloff, and B. Weiss. Symmetric birkhoff sums in infinite ergodic theory. *Ergodic Theory and Dynamical Systems*, 37(8):2394–2416, 2016.
- [3] J. Aaronson and H. Nakada. Trimmed sums for non-negative, mixing stationary processes. *Stochastic Process. Appl.*, 104(2):173–192, 2003.
- [4] J.F. Alves, J. Freitas, S. Luzzatto, and S. Vaienti. From rates of mixing to recurrence times via large deviations. *Adv. Math.*, 228(2):1203–1236, 2011.
- [5] H. Aytac, J. Freitas, and S. Vaienti. Laws of rare events for deterministic and random dynamical systems. *Tans. Amer. Math. Soc.*, 367:8229–8278, 2015.
- [6] R. Azencott, V. Muravina, R. Hekmati, W. Zhang, and M. Paldino. Automatic clustering in large sets of time series. in contributions to partial differential equations and applications. *Springer Contr. Part. Diff. Eq. and App.*, 47:65–75, 2019.
- [7] Normand Beaudry. An intuitive proof of the data processing inequality. *Quantum Information and Computation*, 12(5-6):432–441, 2012.
- [8] A. Boyarsky and P. Gora. *Laws of Chaos*. Springer, 1997.
- [9] S.J. Brown, J. Caesar, and Ferro C.A.T. Global changes in extreme daily temperature since 1950. *Journal of Geophysical Research*, 113:D05115, 2008.
- [10] L.A. Bunimovich and Y. Sinai. Statistical properties of lorentz gas with periodic configuration of scatterers. *Commun. Math. Phys.*, 78:479–497, 1981.
- [11] L.A. Bunimovich, Y. Sinai, and N. Chernov. Markov partitions for two-dimensional billiards. *Russ. Math. Surv.*, 45:105–152, 1990.

- [12] L.A. Bunimovich, Y. Sinai, and N. Chernov. Statistical properties of two-dimensional hyperbolic billiards. *Russ. Math. Surv.*, 46:47–106, 1991.
- [13] M. Carvalho, A. Freitas, J. Freitas, M. Holland, and M. Nicol. Extremal dichotomy for hyperbolic toral automorphisms. *Dynamical Systems: An International Journal*, 30(4):383–403, 2015.
- [14] R.V. Chacon. Weakly mixing transformations which are not strongly mixing. *Proceedings of the American Mathematical Society*, 22(3):559–562, 1969.
- [15] L. Cheng, A. AghaKouchak, E. Gilleland, and R.W. Katz. Non-stationary extreme value analysis in a changing climate. *Climatic Change*, 2:127, 2014.
- [16] N. Chernov. Decay of correlations and dispersing billiards. *J. Stat. Phys.*, 94:513–556, 1999.
- [17] N. Chernov. Decay of correlations in dispersing billiards. *Journal of Statistical Physics*, 94:513–556, 1999.
- [18] N. Chernov. Statistical properties of piecewise smooth hyperbolic systems in high dimensions. *Discr. Cont. Dynam.*, 5:425–448, 1999.
- [19] N. Chernov. Decay of correlations. *Scholarpedia*, 3(4):4862, 2008.
- [20] N. Chernov. *On statistical properties of dynamical systems*. Preprint ESI 2074, 2008.
- [21] N. Chernov and Kleinbock D. Dynamical borel-cantelli lemmas for gibbs measures. *Israel J. Math*, 122:1–27, 2001.
- [22] N. Chernov and D. Dolgopyat. Hyperbolic billiards and statistical physics. *Proceedings of International Congress of Mathematicians*, 2006.
- [23] N. Chernov and R. Markarian. Chaotic billiards. *Math. Surv. Monographs*, 127:316, 2006.
- [24] N. Chernov and H.K. Zhang. On statistical properties of hyperbolic systems with singularities. *J. Stat. Phys.*, 136:615–642, 2009.
- [25] Y.S. Chow and H. Robbins. On sums of independent random variables with infinte moments, and "fair" games. *Proc. Nat. Acad. Sci*, 47:330–335, 1961.
- [26] N. Christidis, P. Stott, and S. Brown. The role of human activity in the recent warming of extreme warm daytime temperatures. *J. Climate*, 24:1922–1930, 2011.

- [27] S. Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer, 2001.
- [28] D. Davies and D. Bouldin. A cluster separation measure. *IEEE Trans. Pat. An. and Mach. Learn.*, 2:224–227, 1979.
- [29] J. Dedecker, S. Gouëzel, and F. Merlevede. Some almost sure results for unbounded functions of intermittent maps and their associated markov chains. *Ann. Inst. Henri Poincaré Probab. Stat.*, 46(3):796–821, 2010.
- [30] M. Denker, M. Gordin, and A. Sharova. A poisson limit theorem for toral automorphisms. *Illinois J. Math*, 48(1):1–20, 2004.
- [31] C.H.Q Ding, X. He, H. Zha, G Ming, and H.D. Simon. A min-max cut algorithm for graph partitioning and data clustering. *Proceedings IEEE Int. Conf. on Data Mining*, pages 107–114, 2001.
- [32] D. Dolgopyat. Limit theorems for partially hyperbolic systems. *Trans. AMS*, 356:1637–1689, 2004.
- [33] P.B. Duffy and C. Tebaldi. Increasing prevalence of extreme summer temperatures in the u.s. *Climactic Change*, 111(2):487–495, 2012.
- [34] R. Durrett. *Probability: theory and examples*. Cambridge university press, 2010.
- [35] W. Feller. A limit theorem for random variables with infinite moments. *American Journal of Mathematics*, 68(2):257–262, 1946.
- [36] A. Ferguson and M. Pollicott. Escape rates for gibbs measures. *Ergodic Theory and Dynamical Systems*, 32(3):961–988, 2012.
- [37] J. Finkel and J.I. Katz. Changing us extreme temperature statistics. *Int. J. Climatol*, 37(13):4749–4755, 2017.
- [38] J. Finkel and J.I. Katz. Changing world extreme temperature statistics. *Int. J. Climatol*, 38(5):2613–2617, 2018.
- [39] NOAA National Centers for Environmental Information. State of the climate: Global climate report for july 2017 mean temp anomalies. *online retrieved on September 19, 2018 from <https://www.ncdc.noaa.gov/sotc/global/201703>*, 2017.

- [40] A.C.M Freitas, J. Freitas, and M. Todd. Extremal index, hitting time statistics and periodicity. *Adv. Math.*, 231(5):2626–2665, 2012.
- [41] A.C.M Freitas, J. Freitas, and M. Todd. The compound poisson limit ruling periodic extreme behavior of non-uniformly hyperbolic dynamics. *Comm. Math. Phys.*, 321(2):483–527, 2013.
- [42] J. Freitas, A. Freitas, and M. Todd. Hitting times and extreme value theory. *Probab. Theory Related Fields*, 147(3):675–710, 2010.
- [43] S. Galatolo and D.H. Kim. The dynamical borel-cantelli lemma and the waiting time problems. *Indag. Math. (N.S.)*, 18(3):421–434, 2007.
- [44] M. Gao and H. Zheng. Nonstationary extreme value analysis of temperature extremes in china. *Stochastic Environmental Research and Risk Assessment*, 32(5):1299–1315, 2018.
- [45] S. Gouëzel. A borel-cantelli lemma for intermittent interval maps. *Nonlinearity*, 20(6):1491–1497, 2007.
- [46] C. Gupta, M. Nicol, and W. Ott. A borel-cantelli lemma for non-uniformly expanding dynamical systems. *Nonlinearity*, 23(8):1991–2008, 2010.
- [47] L. Hagen and A.B. Kahng. New spectral methods for ratio cut partitioning and clustering. *IEEE Trans. Comp. Des. Integr. Circ. Syst.*, 11(9):1074–1085, 1992.
- [48] J. Hansen, M. Sato, and R. Ruedy. Perception of climate change. *PNAS*, 109(37):2415–2423, 2012.
- [49] J.E. Hansen, R. Ruedy, M. Sato, M. Imhoff, W. Lawrence, D. Easterling, T. Peterson, and T. Karl. A closer look at united states and global surface temperature change. *J. Geophys. Res.*, 106:23947–23963, 2001.
- [50] G. Hartfield, J. Blunder, and D.S. Arndt. A look at 2017: Takeaway points from the state of the climate supplement. *Bulletin of American Meteorological Society*, 99(8):1527–1539, 2017.
- [51] H. Hasan, N. Radi, and S. Kassim. Modeling of extreme temperature using generalized extreme value (gev) distribution: A case study of penang. *Proceedings of the World Congress on Engineering*, 1, 2012.
- [52] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer, 2017.

- [53] N. Haydn, J. Freitas, and M. Nicol. Convergence of rare events point processes to the poisson for billiards. *Nonlinearity*, 27:1669–1687, 2014.
- [54] N. Haydn, M. Nicol, T. Persson, and S. Vaienti. A note on borel-cantelli lemmas for non-uniformly hyperbolic dynamical systems. *Ergodic Theory Dynam. Systems*, 33(2):475–498, 2013.
- [55] R. Hekmati, R. Azencott, W. Zhang, Z. Chu, and M. Paldino. Localization of epileptic seizure focus by computerized analysis of fmri recordings. *arXiv preprint*, 2018.
- [56] R. Hekmati, R. Azencott, W. Zhang, and M. Paldino. Machine learning to evaluate fmri recordings of brain activity in epileptic patients. *Q-Bio Conference*, 2018.
- [57] M. Hirata. Poisson limit law for axiom a diffeomorphisms. *Erg. Th. Dyn. Syst.*, 13(3):533–556, 1993.
- [58] M.P. Holland, M. Nicol, and A. Török. Almost sure convergence of maxima for chaotic dynamical systems. *Stochastic Process. Appl.*, 126(10):3145–3170, 2016.
- [59] J. Jaerisch, M. Kesseböhmer, and B. Stratmann. A fréchet law and an erdős-philipp law for maximal cuspidal windings. *Ergodic Theory Dynam. Systems*, 33(4):1008–1028, 2013.
- [60] A. Katok and J.M. Strelcyn. *Invariant Manifolds, Entropy and Billiards; Smooth Maps with Singularities. Lect. Notes Math. (with the collaboration of Ledrappier, F. and Przytycki, F.)*, volume 1222. Springer, 1986.
- [61] G. Keller. Rare events, exponential hitting times and extremal indices via spectral perturbation. *Dyn. Syst.*, 27(1):11–27, 2012.
- [62] H. Kesten and R.A. Maller. The effect of trimming on the strong law of large numbers. *Proceedings of the London Mathematical Society*, 3(2):441–480, 1995.
- [63] D. Kim. The dynamical borel-cantelli lemma for interval maps. *Discrete Contin. Dyn. Syst.*, 17(4):891–900, 2007.
- [64] Z. Kosloff. A universal divergence rate for symmetric birkhoff sums in infinite ergodic theory. *Trans. Amer. Math. Soc.*, 369:6373–6388, 2017.

- [65] J. Kysely. Probability estimates of extreme temperature events: Stochastic modelling approach vs. extreme value distributions. *Studia Geophysica et Geodaetica*, 46(1):93–112, 2002.
- [66] J. Kysely. Temporal fluctuations in heat waves at prague-klementinum, the czech republic, from 1901-1997, and their relationships to atmospheric circulation. *Int. J. Climatol*, 22:33–50, 2002.
- [67] J. Lawrimore. Global summary of the year, version 1.0., noaa national centers for environmental information. *online retrieved on 2019 from <https://www.currentresults.com/Yearly-Weather/USA/TX/Houston/extreme-annual-houstonhigh-temperature.php>*, 2016.
- [68] M.R. Leadbetter, G. Lindgreen, , and H. Rootzen. *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, New York, 1983.
- [69] E.G. Learned-Miller. *Entropy and Mutual Information*. *online at: <https://people.cs.umass.edu/~elm/Teaching/Docs/mutInf.pdf>*, 2013, accessed March 2019.
- [70] E.M. Leibensperger, L.J. Mickley, D.J. Jacob, W.T. Chen, J.H. Seinfeld, A. Nenes, P.J. Adams, D.G. Streets, N. Kumar, and D. Rind. Climatic effects of 1950-2050 changes in us anthropogenic aerosols– part 2: Climate response. *Atmospheric Chemistry and Physics*, 12:3349–3362, 2012.
- [71] C. Liverani, B. Saussol, and S. Vaienti. A probabilistic approach to intermittency. *Ergodic theory and dynamical systems*, 19:671–685, 1999.
- [72] N. Lott. The quality control of the integrated surface hourly database. *online retrieved on March 18, 2018 from <https://www1.ncdc.noaa.gov/pub/data/inventories/ish-qc.pdf>*, 2003.
- [73] V. Lucarini, D. Faranda, A. Freitas, J. Freitas, M. Holland, T. Kuna, M. Nicol, M. Todd, and S. Vaienti. *Extremes and Recurrence in Dynamical Systems*. Wiley, 2016.
- [74] N. Luzia. Borel-cantelli lemma and its applications. *Trans. Amer. Math. Soc.*, 366(1):547–560, 2014.
- [75] F. Maucourant. Dynamical borel-cantelli lemma for hyperbolic spaces. *Israel J. Math*, 152:143–155, 2006.
- [76] G. Meehl, C. Tebaldi, and D. Adams-Smith. Us daily temperature records past, present and future. *PNAS*, 113(49):13977–13982, 2016.

- [77] J.A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, 1965.
- [78] National Oceanic and Atmospheric Administration. Isd-lite technical document. *online retrieved on March 18, 2019 from <https://www1.ncdc.noaa.gov/pub/data/noaa/isd-lite/isd-lite-technical-document.txt>*, 2006.
- [79] T.R. Oke. The energetic basis of the urban heat island. *Journal of the Royal Meteorological Society*, 108(455):1–24, 1982.
- [80] F. Pène and B. Saussol. Poisson law for some nonuniformly hyperbolic dynamical systems with polynomial rate of mixing. *Ergod. Th. Dyn. Syst.*, 36(8):2602–2626, 2016.
- [81] K. Petersen. *Ergodic Theory*. Cambridge University Press, 1983.
- [82] W. Phillipp. Some metrical theorems in number theory. *Pacific J. Math*, 20:109–127, 1967.
- [83] H. Poincaré. On the problem of the three bodies and the equations of dynamics. *Acta Mathematica*, 13(1-2):1–270, 1890.
- [84] R. Portmann, S. Solomon, and G. Hegerl. Spatial and seasonal patterns in climate change, temperatures, and precipitation across the united states. *PNAS*, 106(18):7324–7329, 2009.
- [85] S. Rahmstorf and D. Coumou. Increase of extreme events in a warming world. *PNAS*, 108(44):17905–17909, 2011.
- [86] J. Rivera-Letelier. Asymptotic expansion of smooth interval maps. *arxiv:1204.3071v2*.
- [87] C. Schär, P. Vidale, D. Lüthi, C. Frei, C. Häberli, M. Liniger, and C. Appenzeller. The role of increasing temperature variability in european summer heatwaves. *Nature*, 427:332–336, 2004.
- [88] T. Schindler. Generalized strong laws of large numbers for intermediately trimmed sums for non-negative stationary processes. *Thesis, Mathematik and Informatik, Universität Bremen*, 2015.
- [89] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Trans. Patt. An. and Mach. Intel.*, 22(8):888–905, 2000.

- [90] B. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, 1986.
- [91] Y. Sinai. Dynamical systems with elastic reflections. ergodic properties of dispersing billiards (russian). *Uspehi Mat. Nauk*, 25(2):141–192, 1970.
- [92] R.L. Smith. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72:67–90, 1985.
- [93] V. Sprindzuk. *Metric theory of Diophantine approximations*. V.H. Winston and Sons, Washington D.C. Translated from the Russian and edited by Richard A. Silverman, With a foreword by Donald J. Newman, Scripta Series in Mathematics, 1979.
- [94] Y. Wen, X. Liu, and G. Du. Nonuniform time-lag effects of asymmetric warming on net primary productivity across global terrestrial biomes. *Earth Interact*, 22:1–26, 2018.
- [95] G. Wergen and J. Krug. Record-breaking temperatures reveal a warming climate. *Europhysics Letters*, 92:3, 2010.
- [96] L.S. Young. Statistical properties of dynamical systems with some hyperbolicity. *Ann. of Math.*, 147:585–650, 1998.
- [97] L.S. Young. Recurrence times and rates of mixing. *Israel J. Math*, 110:153–188, 1999.
- [98] L. Yu, S. Zhong, W. Heilman, and X. Bian. Trends in seasonal warm anomalies across the contiguous united states: Contributions from natural climate variability. *Scientific Reports*, 8:3435, 2018.
- [99] X. Zhang and X. Yan. Spatiotemporal change in geographical distribution of global climate types in the context of climate warming. *Climate Dynamics*, 43(3-4):595–605, 2014.
- [100] J. Zscheischler, M. Mahecha, and S. Harmeling. Climate classifications: the value of unsupervised clustering. *Procedia Computer Science*, 9:897–906, 2012.