STATIONARITY AND RANDOM LOCATIONS

A Dissertation
Presented to the Faculty of the Graduate School
of Cornell University
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by
Yi Shen
May 2013

STATIONARITY AND RANDOM LOCATIONS

Yi Shen, Ph.D.

Cornell University 2013

We introduce a family of random locations called "intrinsic location functionals", which include most of the random locations that one may encounter in many cases, *e.g.*, the location of the path supremum/infimum over an interval, the first/last hitting times, *etc*. It is proved that the distributions of these locations must satisfy certain properties, such as the absolute continuity in the interior of the interval, and a group of constraints on the total variation of the density function. It is further shown that the list of properties that we obtained for the distributions is actually equivalent to the stationarity of the process, in the sense that a process is stationary if and only if the distributions of all intrinsic location functionals satisfy the list of properties. In this way we get an alternative characterization of stationarity from the perspective of random locations. Moreover, we develop alternative equivalent descriptions for intrinsic location functionals in terms of partially ordered random point sets and piecewise linear functions. The main results can be extended in many directions, for instance, stationary increment processes, stationary random fields and isotropic random fields.

BIOGRAPHICAL SKETCH

Yi Shen was born in Beijing, China on July 15, 1983. He did his undergraduate studies in the Academic Talent Program (also named “Class of Fundamental Sciences”) at Tsinghua University. From 2005 to 2008 he was in Ecole Polytechnique, France, where he received his Ingénieur degree. He continued to pursue a doctoral degree in the School of Operations Research and Information Engineering at Cornell University since then.

He has broad research interest in various areas in probability. His Ph.D. works, supervised by Professor Gennady Samorodnitsky, focus on the relation between stationary processes and the distributions of random locations. He is also interested in stochastic algebraic topology, extreme value theory, financial mathematics, *etc*.

Yi Shen will join the Department of Statistics and Actuarial Science at the University of Waterloo as an assistant professor upon graduation.

To my parents: Changning Shen and Liqian He, and my wife: Xuyang Ma,
for their love, encouragement and endless support.

ACKNOWLEDGEMENTS

My deepest gratitude goes first and foremost to my Ph.D. advisor, Professor Gennady Samorodnitsky, for his guidance and encouragement. He has been a perfect advisor in all aspects, and a great mentor in work and life. I have learned enormously from him, from finding the good topic to nailing down all the details, from submitting papers to participating in academic activities. I am also deeply impressed and influenced by his hard working and rigorous scholarship. None of my academic achievements during my Ph.D. study will be possible without his help.

I also wish to thank my committee members, Professor Robert Jarrow and Professor Michael Nussbaum, for their support in my study and career development. I owe a lot to Professor Sidney Resnick and Professor Mark Lewis. I have been fortunate to have them as my teachers first, then to serve as a teaching assistant for them. I am also very grateful to Professor Jim Dai, for his enlightening comments on both research and life.

Next I thank the professors in France who have taught me and helped me during my study in Ecole Polytechnique. Professor Nicole El Karoui gave me warmest encouragement to pursue a research career in probability, and to go to conferences from a very early stage. Professor Luciano Campi nicely supervised my final project in Ecole Polytechnique. It was their support that allowed me to go further, from France to North America.

My gratitude goes out to my undergraduate advisor, Professor Feng-Yu Wang, who brought me into the world of probability, and has always been very supportive during these years. I want to thank Professor Zenghu Li as well, for his valuable pieces of advice and help. Special thanks go to Professor Shouyong Pei, who has been an excellent teacher and generously offered direction to my first scientific research project.

I would also like to thank my Ph.D. fellow and friends at Cornell. Just to name a few, Chao Ding, Anjie Guo, Juan Li, Cao Ni, Chao Shi, Tia Sondjaja, Yuemeng Sun, Tuohua Wu, Shanshan Zhang, Fan Zhu. You made my life in Ithaca so pleasant and memorable.

Last but important, I am indebted to my family, to which this dissertation is dedicated. My parents always give me their unconditional support and encouragement. The education they gave me is unmatchable. Thank my wife, Xuyang, for her love and companionship, which helped me pass the highs and lows in life.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

This thesis is about how to understand and characterize stationarity and related notions of invariance, from the perspective of random locations. Playing an essential role in probability theory and its applications, stationarity has been intensively studied. A rich stream of literature can be found dealing with various classical aspects of stationary processes and random fields, such as spectral representation, sample path properties, level crossings, Palm theory, extreme values, etc. While the lasting popularity of the text by H. Cramér and M. R. Leadbetter[9] testifies to the historical importance of this area, the new book by G. Lindgren[16], just published, shows that stationarity is still attracting attention of researchers from different fields, and new results keep being published. In a larger picture, stationarity is one of the most important probabilistic symmetries, along with exchangeability, isotropy and contractability, see O. Kallenberg[12].

Although much is known about stationarity, there remain properties whose relation to stationarity may be intrinsic yet not easy to perceive. This dissertation is therefore dedicated to asking and answering the following questions under very general conditions:

-What does the stationarity mean, indeed, regarding the distributions of certain random objects, in particular, the random locations such as the location of the supremum over an interval, or the first hitting time over an interval, among many others?

-Conversely, is it possible to characterize the stationarity by these distributional properties?

-How can we extend the results, originally developed for strictly stationary processes, to different versions of the notion of stationarity, and to other probabilistic symmetries?

For example, consider the location of the supremum of a stationary process over an interval. On one hand, it may not be uniformly distributed in spite of what may appear in a naïve guess; on the other hand, the stationarity does have an influence on its distribution, in a more indirect and complicated way. This influence now can be clarified by our results, applied to the special case of the location of the path supremum.

In this work, the author introduced a notion called “intrinsic location functional” in order to carry out a general study of the impact of stationarity on the distributions of random locations. This is a large class of random locations including, but being much broader than, the location of the path supremum/infimum and the hitting times. It turns out that, despite the different origins and nature of different locations in this family, their common structure already allows us to derive interesting and important properties of their distributions under stationarity. Basic results such as the absolute continuity of the distribution in the interior of the interval and the uniform upper bounds of the density functions are established. The most important among these properties are the so-called “total variation constraints”, which is a group of conditions controlling the total variation of a density function by the sum of its values at certain points.

Surprisingly, the intrinsic location functionals actually provide a whole new framework to characterize stationarity, under which some deep aspects of stationarity can be revealed and studied in a general setting. It turns out that the total variation constraints are not merely a consequence of stationarity; they are actually the stationarity itself viewed from a different angle. More precisely, we have proved that a stochastic process with continuous paths is stationary if and only if all its intrinsic location functionals satisfy the total variation constraints for all intervals. Notice that the total variation constraints do not explicitly involve any shift invariance, as one may expect to see. Thus the new framework differs from the classical one in the sense that it does not directly deal with the finite dimensional distributions, but focuses on the distributions of random locations.

The richness of the concept of stationarity is matched by the various extensions in different directions of the notion of intrinsic location functional. Stationarity actually consists of a family of related notions, with differences in strength (strict stationarity versus wide-sense stationarity) and in object (the stationarity of the state versus the stationarity of the increment). Corresponding to this richness, our results can be naturally extended to higher dimensional cases of stationary random fields, and to the case of stationary increment processes, where a subclass of intrinsic location functionals, called doubly intrinsic location functionals, is identified. We can also restrict the definition to the intervals with a single fixed length to define "local intrinsic location functional", which are proved to have similar properties as intrinsic location functional. The impacts of some other probabilistic symmetries on the distributions of random locations, such as isotropy, have been investigated; others, such as exchangeability, are in the research plan.

The results in this dissertation have wide applications and numerous potential links to different areas. For example, stochastic algebraic topology is an emerging research area attracting more and more attention from both probabilists and statisticians. Statistically, it aims at getting information from the topological structures of data, such as the connectedness, the presence of holes and "handles" in higher dimensional cases, etc. This new approach finds extensive applications in various areas, for instance, brain imaging[8] and signal processing[10]. A review of its deterministic counterpart can be found in[7].

Analyzing data using topological properties requires new probabilistic results to derive the distributional properties of the related quantities. One of the most beautiful and powerful results in this direction was obtained by J. E. Taylor and R. J. Adler, where the expectation of the Euler characteristic of the excursion sets for Gaussian random fields on a manifold was calculated using the Lipschitz-Killing curvatures of the manifold[2],[21].

The mathematical tool used here is Morse theory, which relates Euler characteristic and more generally, the alternating sums of Betti numbers, to the number of critical points, with different degrees, of a "nice" function in the domain of interest. It is in this sense that the results presented in this dissertation have natural links to stochastic algebraic topology, since all critical points are intrinsic location functionals or a generalized version of it. Therefore by providing useful information about the locations of the critical points of a stationary random field, our results can be helpful in determining the topological structure of the excursion sets through Morse theory.

Besides stochastic algebraic topology, one important application of the results in this dissertation is the possibility of constructing a new class of statistical tests for stationarity of stochastic processes. Now we can tell whether a stochastic process is stationary or not by looking at the distributions of the random locations, which was impossible before the needed results are derived in this thesis. This approach can be elaborated into statistical tests once the corresponding functional central limit theorems are provided. Compared to existing tests, the new family of tests has the advantage of not requiring the whole observation of the process, but only the occurrence location/time of certain events.

Another remarkable application is the link between the setting of intrinsic location functionals and queueing systems with deadlines. Queueing systems with deadlines are queueing systems in which each customer only has a fixed amount of time to spend in the system. When the amount of time is used up, she/he must leave the system immediately. It turns out that there exists a correspondence between our works on random locations and certain queueing models with deadlines. The link is made through the ordered set representation of intrinsic location functional discussed in Chapter 8. Under this correspondence, our results actually answer the question about the distribution of the amount of time that the current customer has stayed in the system. It will be interesting to see what else we can get from this relation: on one hand, whether the technique that we used to study random locations can lead to more results in queueing system; on the other hand, whether it is possible to adapt the existing methods in queueing theory to obtain information about the random locations.

This dissertation consists of nine chapters. Except for the current chapter of introduction and the chapter of preliminaries, the other seven chapters are based on published, accepted or submitted papers and unpublished manuscripts:

Chapter 3 is based on the paper “Is the location of the supremum of a stationary process nearly uniformly distributed?”[19], accepted for publication in *The Annals of Probability*. It focuses on a specific random location–the location of the path supremum over a compact interval, and looks at the impact of stationarity on its distribution. A group of conditions for the distribution of this random location, including the absolute continuity of the distribution and the total variation constraints, appear for the first time. We also introduce two mild assumptions on the behavior of the process, under which further conclusions can be derived about the distribution of the location of the path supremum.

Chapter 4 is based on the paper "Distribution of the supremum location of stationary processes" [20], published in Electronic Journal of Probability. It continues the work in Chapter 3 considering the location of the path supremum. It takes, however, a different direction, to show that the group of conditions obtained in Chapter 3 is complete, in the sense that any probabilistic distribution satisfying this group of conditions can be the distribution of the location of the path supremum for some stationary process. Moreover, we show that under certain mixing condition, the renormalized distribution of the supremum location in a larger and larger interval will converge to uniform distribution, returning to the intuition that all the points in a stationary process are “similar”, if the influence of the boundaries of the interval can be omitted.

Chapter 5 is based on the paper "Intrinsic location functionals of stationary processes"[18], submitted for publication. It takes a significant step forward to expand the results in the previous two chapters to a general family of random locations, called “intrinsic location functionals”. This is a very large family of random locations including the random locations that one may encounter in many cases, *e.g.*, the location of the path supremum/infimum over an interval, the first/last hitting time of the process to certain level over an interval, the starting point of the largest shortfall during a short period, *etc.* It turns out that for any random location in this family, it always satisfies the same group of conditions as the location of the path supremum. On the other hand, if this group of conditions is satisfied for all the random locations in the family of intrinsic location functional, then the stochastic process must be stationary. In this sense, we actually find a way to characterize stationarity from the perspective of random locations. Furthermore, the structure of the sets of all possible distributions for various random locations under stationarity are discussed, and optimal bounds are derived for the expectation of functions of random locations.

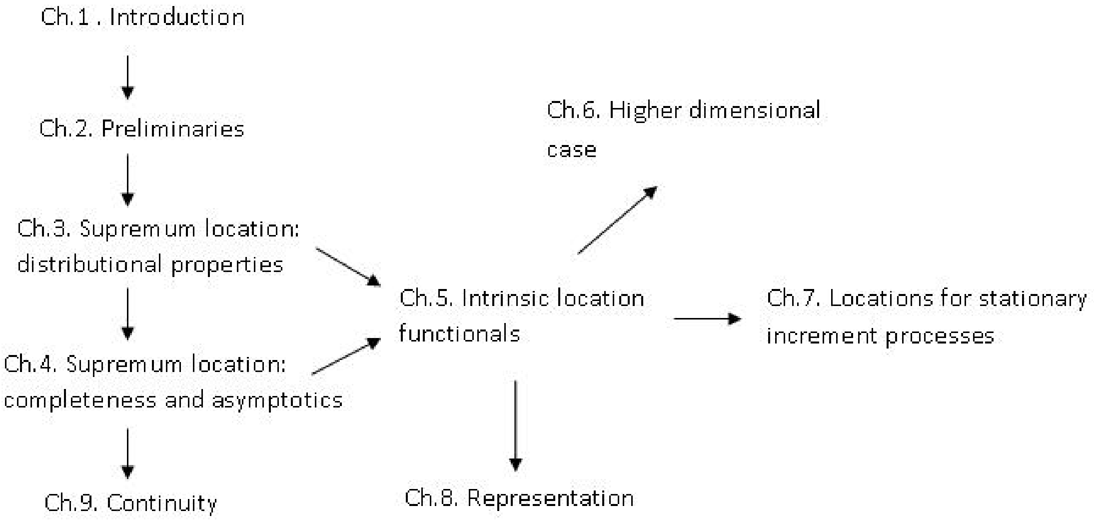
Chapter 6 discusses the higher dimensional extension of the results in Chapter 5, where stationary processes are replaced naturally by stationary random fields, and intervals become hypercubes. We show that most of the results in Chapter 5 can be migrated easily to higher dimensional setting, which now involves faces and densities in different dimensions. Meanwhile, the higher dimensional setting also allows us to study the influence of another probabilistic symmetry, isotropy, on the distributions of intrinsic location functionals. A new group of conditions, called "angular total variation constraints", is proved to hold in addition to the ordinary total variation constraints under isotropy.

Chapter 7 deals with another important generalization of the main results in Chapter 5, where the class of stationary processes is now expanded to the class of stationary increment processes, which is a much larger family of stochastic processes. Consequently, in order to keep similar properties for the distribution, a smaller class of random location, called “doubly intrinsic location functional”, is introduced and studied. Briefly speaking, an intrinsic location functional is doubly intrinsic, if its value remains unchanged under vertical shift of the path. The results in Chapter 5, including the absolute continuity, the total variation constraints and the equivalence between the total variation constraints and stationarity, all extend to stationary increment processes and doubly intrinsic location functionals.

Chapter 8 helps us to get a better understanding of the object of intrinsic location functional by establishing two representation results. The first result shows that for any intrinsic location functional, it can always be regarded as taking the location of the maximal element of a random set determined by the path, according to some (random) order on the random set. The second result characterizes intrinsic location functional by looking at how its value changes when the interval of interest moves. Along with these two representation results, a generalization of intrinsic location functional, called locally intrinsic location functional, is naturally introduced, and its relationship with intrinsic location functional is studied.

The final chapter, Chapter 9, addresses to the question of continuity and approximation. We show that for specific random locations such as the location of the path supremum and for specific stochastic processes such as stationary Gaussian processes satisfying some mild spectral conditions, the convergence of the covariance functions or the spectral measures imply the convergence of the distributions of the random location. This lays a theoretical basis for numerical approximation of the distribution of random locations.

At the end of this introduction, we present the following roadmap of the thesis. Although the chapters in this book are logically related, the author tried to make them self-containing to the largest extent possible, so the readers interested in particular parts can skip the other chapters, especially when the proofs are not the focus.



CHAPTER 2

PRELIMINARIES

2.1 Definition of stationarity

Stationarity is, indeed, a family of related notions instead of one single notion. Intuitively, it says that the distribution of a stochastic process or a random field defined in a space exhibits an invariance under the operation of translation. The specific type of the invariance determines the specific version of stationarity.

Two different versions of stationarity are most commonly used. The first one, which is also "the" stationarity used in almost all the places in this dissertation, is also called "strict stationarity". It requires that any finite dimensional distribution of the process/random field will not change under translation. This is to say, in d-dimensional Euclidean space, for any $n = 1, 2, ...$, $t_1, ..., t_n \in \mathbb{R}^d$ and $\Delta \in \mathbb{R}^d$,

$$(\mathbf{X}(t_1), ..., \mathbf{X}(t_n)) \overset{d}{=} (\mathbf{X}(t_1 + \Delta), ..., \mathbf{X}(t_n + \Delta)),$$

where $\mathbf{X}$ is the random field, or stochastic process in dimension 1, and "$\overset{d}{=}$" means that the equality holds in the sense of distribution.

In time series analysis and many other occasions, the criterion for strict stationarity may be too strong to hold. Therefore another notion called "wide sense stationarity" or "weak stationarity", which only requires the shift invariance of the first two moments, namely, expectation and covariance, is also used. More precisely, a random field $\mathbf{X}$ is called wide sense stationary, if both $E(\mathbf{X}(t))$ and $Cov(\mathbf{X}(t), \mathbf{X}(t + \Delta))$ are constants in t.

One of the most commonly used (strictly) stationary process is the Ornstein-

Uhlenbeck (OU) process, which can be defined as the solution to the following stochastic differential equation:

$$dX_t = \theta(\mu - X_t)dt + \sigma dW_t.$$

The drift term negatively correlated to the value of the process makes the process mean-reverting. If in addition, the process starts at its stationary distribution, then the OU process becomes strictly stationary. Notice that since this is a one-dimensional semi-martingale driven by a Brownian motion, it has almost surely continuous but nowhere smooth paths, and therefore possesses infinitely many local maxima and minima in any interval.

The notion of stationarity does not only apply to the stochastic process itself; it can also be used to characterize the increments of a process, leading to the notion of stationary increment processes. Again, in this dissertation we are mostly focused on strict stationary increment processes, which means that its increment process $Y(\cdot) := X(\cdot + s) - X(\cdot)$ is a strictly stationary process with any parameter $s \in \mathbb{R}$. It is easy to see that stationary increment process is a strictly larger family of stochastic processes compared to stationary processes, since each stationary process automatically has stationary increments, but the converse is not true. Many of the stochastic processes that we often meet, such as Brownian motion, or more broadly, Lévy processes, are examples of stationary increment processes.

2.2 Stationary Gaussian processes

A random vector $(N_1, ..., N_n)$ is a Gaussian random vector, if all the linear combinations of its components follow a Gaussian distribution. A continuous time stochastic process/random field $\mathbf{X}$ is Gaussian, if all its finite dimensional distributions are Gaussian.

For any Gaussian process $\mathbf{X}$, after centering, the distribution of the whole process can be characterized by the covariance function $r(s,t) := Cov(X(s), X(t))$. As an immediate result of this definition, we have $r(s,t) = r(t,s)$ for all $s,t \in \mathbb{R}$, and moreover, for any $n = 1, 2, ...$, $t_1, ..., t_n \in \mathbb{R}$ and $z_1, ..., z_n \in \mathbb{R}$, $\sum_{i,j=1}^{n} r(t_i, t_j) z_i z_j = Var(\sum_{i=1}^{n} z_i X(t_i)) \geq 0$. We call the functions with this property "nonnegative definite", and it turns out that this is the only property needed for a covariance function. That is, for any nonnegative definite function, there always exists a Gaussian process, such that the function is the covariance function for the Gaussian process. The proof of this result is given in page 80 of [9].

A special family of Gaussian processes is the stationary Gaussian processes. As its name suggests, it is the intersection of the family of Gaussian processes and the family of stationary processes. Consequently, it inherits the properties from both families. In particular, the covariance function is now a function of a single variable: $r(t) := r(0,t) = r(s, s+t)$ for any real number s. In this case, the condition for r to be nonnegative definite becomes naturally $\sum_{i,j=1}^{n} r(t_j - t_i) z_i z_j \geq 0$ for all $n = 1, 2, ...$, $t_1, ..., t_n \in \mathbb{R}$ and $z_1, ..., z_n \in \mathbb{R}$.

2.3 Spectral representation

The famous Bochner theorem shows that a necessary and sufficient condition for a (possibly complex valued) continuous function to be nonnegative definite is that it has the following spectral representation:

Theorem 2.3.1. *A continuous function $r(t)$ is nonnegative definite, if only if there exists a finite measure F, such that*

$$r(t) = \int_{-\infty}^{\infty} e^{it\lambda} dF(\lambda).$$

The measure F is called the “spectral measure”, and its moments are called “spectral moments”. When we only consider the real valued processes, the covariance function r is also real valued, in which case the spectral representation becomes

$$r(t) = \int_{0}^{\infty} cos(\lambda t) dG(\lambda)$$

for some measure G on $[0, \infty)$.

It should be pointed out that the spectral representation also has more powerful version which directly deals with the process rather than the covariance function, using stochastic integral and orthogonal increment processes. However, we will skip it here since it will not be used anywhere in this dissertation. Interested reader can read chapter 7 in [9] for details.

CHAPTER 3

DISTRIBUTION OF THE LOCATION OF PATH SUPREMUM: GENERAL PROPERTIES

3.1 Introduction

The extremes of stationary processes, especially of Gaussian processes, have attracted significant interest for a long time. Many of results are described in the books [2] and [4], with shorter versions in [1] and [3]. Roughly speaking, these results can be categorized as follows: the exact distributions of the suprema have been calculated for several particular processes: bounds on the supremum distribution have been obtained for a large number of processes; the asymptotic behavior of the level crossing probability has been studied for a large number of processes. Almost without exception, however, these results deal with the value of the supremum, while very little is known about the random location of the supremum.

The work in chapter 3 arises from an obvious attempt to understand the effect of stationarity of the process on the distribution of the location of the supremum. Therefore in this chapter, we look at stationary stochastic processes in continuous, one-dimensional, time, and we will consider the location of its global supremum over a compact interval. It turns out that answering even this, apparently simple question, leads to unexpected insights.

We now discuss out setup more formally. Let $\mathbf{X} = (X(t),\, t \in \mathbb{R})$ be a stationary process. If the sample paths of the process are upper semi-continuous, then the process is bounded from above on any compact interval $[0, T]$, and its supremum over that interval is attained. We are interested in the location of that supremum within the interval $[0, T]$.

It is, of course, entirely possible that the supremum of the process in the interval $[0, T]$ is not unique (i.e. that it is achieved at more than one point). In that case one could be more specific and take, for example, the left-most point in which the largest value over the interval is achieved, as the location of the supremum. In this and the following chapter we will sometimes deal with the situation in which, on an event of probability 1, the supremum is achieved at a single point. In either case it is easy to check that the location of the supremum is a well defined random variable.

Will the stationarity of the process guarantee a uniform distribution of the location of the supremum over the interval? The answer is negative. The examples in Section 9.4 of [15] show that even in the case of Gaussian processes with a uniquely attained supremum (thus eliminating a possible bias resulting from taking the leftmost supremum location), the supremum can still be located, with a positive probability, at one of the endpoints of the interval and, furthermore, the remaining mass in the interior of the interval does not have to be uniformly distributed there.

It is, of course, the endpoints of the interval that are responsible for the lack of uniformity. In a sense, the points near the ends of the interval have "fewer local competitors" for being the supremum than the points further from the endpoints do. But exactly how far from having the uniform distribution can the location of the supremum be? In this chapter we give a very detailed answer to this question by showing that this distribution is absolutely continuous in the interior of the interval and describing very specific conditions its density must satisfy. This is done in Section 3.3. Our results turn out to be quite complete. In fact, we will show in next chapter that, for a very broad class of stationary processes with a uniquely achieved supremum, our description actually gives all possible distributions of its location. In the present chapter we start with treating a general upper semi-continuous stationary process and (with one exception) allowing the process to have multiple supremum locations within an interval. We proceed with establishing extra conditions the density has to satisfy if the process satisfies certain assumptions. In Section 3.4 we provide the sharpest possible universal upper bounds on the density both in the general case and in the case of time-reversible stationary processes.

3.2 Notation and assumptions on the stationary process

For the remainder of this chapter $\mathbf{X} = (X(t),\, t \in \mathbb{R})$ is a stationary process with upper semi-continuous sample paths, defined on some probability space $\big(\Omega, \mathcal{F}, P\big)$. For a compact interval $[a, b]$, we will denote by

$$\tau_{\mathbf{X},[a,b]} = \min\big\{t \in [a, b] :\; X(t) = \sup_{a \le s \le b} X(s)\big\}.$$

That is, $\tau_{\mathbf{X},[a,b]}$ is the first time the overall supremum in the interval $[a, b]$ is achieved. It is elementary to check that $\tau_{\mathbf{X},([a,b])}$ is a well defined random variable. If $a = 0$, we will use the single variable notation $\tau_{\mathbf{X},b}$.

We denote by $F_{\mathbf{X},[a,b]}$ the law of $\tau_{\mathbf{X},[a,b]}$; it is a probability measure on the interval $[a, b]$. If $a = 0$, we have the corresponding single variable notation $F_{\mathbf{X},b}$. The following statements are obvious.

Lemma 3.2.1. (i) *For any* $\Delta \in \mathbb{R}$,

$$F_{\mathbf{X},[\Delta,T+\Delta]}(\cdot) = F_{\mathbf{X},T}(\cdot - \Delta)\,.$$

(ii) *For any intervals* $[c,d] \subseteq [a,b]$,

$$F_{\mathbf{X},[a,b]}(B) \le F_{\mathbf{X},[c,d]}(B) \text{ for any Borel set } B \subset [c,d].$$

The discussion of the leftmost supremum location $\tau_{\mathbf{X},[a,b]}$ in the sequel applies equally well to the rightmost supremum location, for instance, by considering the time-reversed stationary process $(X(-t),\, t \in \mathbb{R})$. In some cases we will find it convenient to assume that the supremum is achieved at a unique location. Formally, for $T > 0$ we denote by $X_*(T) = \sup_{0\le t\le T} X(t)$ the largest value of the process in the interval $[0,T]$, and consider the set

$$\Omega_T = \big\{\omega \in \Omega :\ X(t_i) = X_*(T) \text{ for at least two different } t_1, t_2 \in [0,T]\big\}.$$

It is easy to see that Ω_T is a measurable set. The following assumption says that, on a set of probability 1, the supremum over interval $[0,T]$ is uniquely achieved.

Assumption U_T: $P(\Omega_T) = 0$.

In our previous notation, under Assumption U_T, $\tau_{\mathbf{X},[a,b]}$ is the unique point at which the supremum over the interval $[0,T]$ is achieved, and $F_{\mathbf{X},T}$ is the law of that point.

Even though many of our results do not require it, the most complete description of the distribution of the location of the supremum that we have requires the following, additional, assumption.

Assumption L:

$$K := \lim_{\varepsilon \downarrow 0} \frac{P\bigl(\mathbf{X} \text{ has a local maximum in } (0,\varepsilon)\bigr)}{\varepsilon} < \infty\,.$$

It is easy to check that the limit in Assumption L exists. If, for example, the process $\mathbf{X}$ has differentiable sample paths, then a sufficient condition for Assumption L is that the expected number of times the process $Y(t) = X'(t),\ t \in \mathbb{R}$ crosses zero in a unit time interval is finite; the latter can be checked using, for instance, Theorem 7.2.4 in [15].

Assumption L rules out existence of "too frequent" local extrema of the sample paths. For sample continuous processes this also rules out rapid oscillation of the sample paths possessed, for instance, by the Gaussian Ornstein-Uhlenbeck process of Example 3.3.7 below. In fact, we will presently see that, at least for sample continuous processes, under Assumption L the process has, with probability 1, sample paths of locally bounded variation.

Lemma 3.2.2. *Let $\mathbf{X} = (X(t),\ t \in \mathbb{R})$ be a stationary sample upper semi-continuous process satisfying Assumption L. Then, for any $T > 0$, on an event of probability 1 the process has finitely many local maxima and minima in the interval $(0,T)$. In particular, if the process is sample continuous, then its sample paths are, on event of probability 1, of locally bounded variation.*

Proof. For notational simplicity we take $T = 1$. For $n = 1, 2, \ldots$ let

$$N_n = \sum_{i=1}^{2^n} \mathbf{1}\left(\text{a point in } \left[\frac{i-1}{2^n}, \frac{i}{2^n}\right) \text{ is a local maximum of } \mathbf{X}\right).$$

Clearly, the sequence N_n is nondecreasing, and $N_n \to N_\infty$, where N_∞ is the total number of local maxima of $\mathbf{X}$ in the interval $[0,1)$. By the monotone convergence theorem,

$$\begin{aligned} EN_\infty &= \lim_{n\to\infty} EN_n \\ &\le \limsup_{n\to\infty} 2^n P\big(\mathbf{X} \text{ has a local maximum in } (0,2^{-n})\big) \le K. \end{aligned}$$

Therefore, $N_\infty < \infty$ a.s. Since between any two distinct local minima there is a local maximum, the number of local minima in $[0,1)$ is a.s. finite as well. Since a sample continuous process must have a monotone path between any two consecutive local extrema, the lemma has been proved. $\square$

3.3 Description of the possible distributions of the location of the supremum

We start with a result showing existence of a density in the interior of the interval $[0,T]$ of the leftmost location of the supremum in that interval for any upper semi-continuous stationary process, as well as conditions this density has to satisfy. Only one of the statements of the theorem requires Assumption U_T, in which case the statement applies to the unique location of the supremum. See Remark 3.3.2 in the sequel.

Theorem 3.3.1. *Let $\mathbf{X} = (X(t),\, t \in \mathbb{R})$ be a stationary sample upper semi-continuous process. Then the restriction of the law $F_{\mathbf{X},T}$ to the interior $(0,T)$ of the interval is absolutely continuous. The density, denoted by $f_{\mathbf{X},T}$, can be taken to be equal to the right derivative of the cdf $F_{\mathbf{X},T}$, which exists at every point in the interval $(0,T)$. In this case the density is right continuous, has left limits, and has the following properties.*

(a) The limits

$$f_{\mathbf{X},T}(0+) = \lim_{t\to 0} f_{\mathbf{X},T}(t) \text{ and } f_{\mathbf{X},T}(T-) = \lim_{t\to T} f_{\mathbf{X},T}(t)$$

exist.

(b) The density has a universal upper bound given by

$$f_{\mathbf{X},T}(t) \le \max\left(\frac{1}{t}, \frac{1}{T-t}\right),\ 0 < t < T\,. \tag{3.1}$$

(c) Assume that the process satisfies Assumption U_T. Then the density is bounded away from zero:

$$\inf_{0<t<T} f_{\mathbf{X},T}(t) > 0\,. \tag{3.2}$$

(d) The density has a bounded variation away from the endpoints of the interval. Furthermore, for every $0 < t_1 < t_2 < T$,

$$TV_{(t_1,t_2)}(f_{\mathbf{X},T}) \le \min\bigl(f_{\mathbf{X},T}(t_1), f_{\mathbf{X},T}(t_1-)\bigr) + \min\bigl(f_{\mathbf{X},T}(t_2),\ f_{\mathbf{X},T}(t_2-)\bigr)\,, \tag{3.3}$$

where

$$TV_{(t_1,t_2)}(f_{\mathbf{X},T}) = \sup \sum_{i=1}^{n-1} \bigl| f_{\mathbf{X},T}(s_{i+1}) - f_{\mathbf{X},T}(s_i) \bigr|$$

is the total variation of $f_{\mathbf{X},T}$ on the interval (t_1, t_2), and the supremum is taken over all choices of $t_1 < s_1 < \ldots < s_n < t_2$.

(e) The density has a bounded positive variation at the left endpoint and a bounded negative variation at the right endpoint. Furthermore, for every $0 < \varepsilon < T$,

$$TV^{+}_{(0,\varepsilon)}(f_{\mathbf{X},T}) \le \min\bigl(f_{\mathbf{X},T}(\varepsilon), f_{\mathbf{X},T}(\varepsilon-)\bigr) \tag{3.4}$$

and

$$TV^{-}_{(T-\varepsilon,T)}(f_{\mathbf{X},T}) \le \min\bigl(f_{\mathbf{X},T}(T-\varepsilon), f_{\mathbf{X},T}(T-\varepsilon-)\bigr)\,, \tag{3.5}$$

where for any interval $0 \le a < b \le T$,

$$TV^{\pm}_{(a,b)}(f_{\mathbf{X},T}) = \sup \sum_{i=1}^{n-1} \bigl(f_{\mathbf{X},T}(s_{i+1}) - f_{\mathbf{X},T}(s_i)\bigr)_{\pm}$$

is the positive (negative) variation of $f_{\mathbf{X},T}$ *on the interval* (a,b)*, and the supremum is taken over all choices of* $a < s_1 < \ldots < s_n < b$.

(f) The limit $f_{\mathbf{X},T}(0+) < \infty$ *if and only if* $TV_{(0,\varepsilon)}(f_{\mathbf{X},T}) < \infty$ *for some (equivalently, any)* $0 < \varepsilon < T$*, in which case*

$$TV_{(0,\varepsilon)}(f_{\mathbf{X},T}) \leq f_{\mathbf{X},T}(0+) + \min\bigl(f_{\mathbf{X},T}(\varepsilon), f_{\mathbf{X},T}(\varepsilon-)\bigr) . \tag{3.6}$$

Similarly, $f_{\mathbf{X},T}(T-) < \infty$ *if and only if* $TV_{(T-\varepsilon,T)}(f_{\mathbf{X},T}) < \infty$ *for some (equivalently, any)* $0 < \varepsilon < T$*, in which case*

$$TV_{(T-\varepsilon,T)}(f_{\mathbf{X},T}) \leq \min\bigl(f_{\mathbf{X},T}(T-\varepsilon), f_{\mathbf{X},T}(T-\varepsilon-)\bigr) + f_{\mathbf{X},T}(T-) . \tag{3.7}$$

Proof. Choose $0 < \delta < T/2$. We claim that for every $\delta \leq t \leq T-\delta$, for every $\rho > 0$ and every $0 < \varepsilon < \delta\rho/(1+\rho)$

$$P\bigl(t < \tau_{\mathbf{X},T} \leq t+\varepsilon\bigr) \leq \varepsilon(1+\rho) \max\left(\frac{1}{t}, \frac{1}{T-t}\right) . \tag{3.8}$$

This statement, once proved, will imply absolute continuity of $F_{\mathbf{X},T}$ on the interval $(\delta, T-\delta)$ and, since $\delta > 0$ can be taken to be arbitrarily small, also on $(0,T)$. Further, (3.8) will imply that the version of the density given by

$$f_{\mathbf{X},T}(t) = \limsup_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} P\bigl(t < \tau_{\mathbf{X},T} \leq t+\varepsilon\bigr),\ 0 < t < T ,$$

satisfies the bound (3.1).

We proceed to prove (3.8). Suppose that, to the contrary, (3.8) fails for some $\delta \leq t \leq T-\delta$ and $0 < \varepsilon < \delta\rho/(1+\rho)$. Choose

$$\varepsilon < \theta < \frac{\rho}{1+\rho}\delta$$

and $0 < a < t < b < T$ such that

$$\min\bigl(t, T-t\bigr) - \theta < b-a < \min\bigl(t, T-t\bigr) - \varepsilon .$$

For $a \le s \le b$, by stationarity, we have

$$P\big(s < \tau_{\mathbf{X},[s-t,s-t+T]} \le s+\varepsilon\big) > \varepsilon(1+\rho)\max\left(\frac{1}{t}, \frac{1}{T-t}\right). \tag{3.9}$$

Further, let $a \le s_1 < s_1+\varepsilon \le s_2 \le b$. We check next that

$$\big\{s_j < \tau_{\mathbf{X},[s_j-t,s_j-t+T]} \le s_j+\varepsilon,\ j=1,2\big\} = \emptyset\,. \tag{3.10}$$

Indeed, let Ω_{s_1,s_2} be the event in (3.10). Note that the intervals $(s_1, s_1+\varepsilon)$ and $(s_2, s_2+\varepsilon)$ are disjoint and, by the choice of the parameters a and b, each of these two intervals is a subinterval of both $[s_1-t, s_1-t+T]$ and $[s_2-t, s_2-t+T]$. Therefore, on the event Ω_{s_1,s_2} we cannot have

$$X\big(\tau_{\mathbf{X},[s_1-t,s_1-t+T]}\big) < X\big(\tau_{\mathbf{X},[s_2-t,s_2-t+T]}\big)\,,$$

for otherwise $\tau_{\mathbf{X},[s_1-t,s_1-t+T]}$ would fail to be a location of the maximum over the interval $[s_1-t, s_1-t+T]$. For the same reason on the event Ω_{s_1,s_2} we cannot have

$$X\big(\tau_{\mathbf{X},[s_1-t,s_1-t+T]}\big) > X\big(\tau_{\mathbf{X},[s_2-t,s_2-t+T]}\big)\,.$$

Finally, on the event Ω_{s_1,s_2} we cannot have

$$X\big(\tau_{\mathbf{X},[s_1-t,s_1-t+T]}\big) = X\big(\tau_{\mathbf{X},[s_2-t,s_2-t+T]}\big)\,,$$

for otherwise $\tau_{\mathbf{X},[s_2-t,s_2-t+T]}$ would fail to be the leftmost location of the maximum over the interval $[s_2-t, s_2-t+T]$. This establishes (3.10).

We now apply (3.9) and (3.10) to the points $s_i = a + i\varepsilon,\ i = 0, 1, \ldots, \lceil (b-a)/\varepsilon\rceil - 1$. We have

$$1 \ge P\left(\bigcup_{i=0}^{\lceil (b-a)/\varepsilon\rceil-1} \big\{s_i < \tau_{\mathbf{X},[s_i-t,s_i-t+T]} \le s_i+\varepsilon\big\}\right)$$

$$= \sum_{i=0}^{\lceil (b-a)/\varepsilon\rceil-1} P\big(s_i < \tau_{\mathbf{X},[s_i-t,s_i-t+T]} \le s_i+\varepsilon\big) > \frac{b-a}{\varepsilon}\varepsilon(1+\rho)\max\left(\frac{1}{t}, \frac{1}{T-t}\right)$$

$$
> \Big(\min\big(t, T-t\big) - \theta\Big)(1+\rho)\max\left(\frac{1}{t}, \frac{1}{T-t}\right)
$$

$$
> \left(1 - \frac{\delta}{\min(t, T-t)}\,\frac{\rho}{1+\rho}\right)(1+\rho) \geq \left(1 - \frac{\rho}{1+\rho}\right)(1+\rho) = 1
$$

by the choice of θ. This contradiction proves (3.8).

Before proceeding with the proof of Theorem 3.3.1, we pause to prove the following important lemma.

Lemma 3.3.1. *Let* $0 \leq \Delta < T$. *Then for every* $0 \leq \delta \leq \Delta$, $f_{\mathbf{X},T-\Delta}(t) \geq f_{\mathbf{X},T}(t+\delta)$ *almost everywhere in* $(0, T-\Delta)$. *Furthermore, for every such* δ *and every* $\varepsilon_1, \varepsilon_2 \geq 0$, *such that* $\varepsilon_1 + \varepsilon_2 < T - \Delta$,

$$
\int_{\varepsilon_1}^{T-\Delta-\varepsilon_2} \big(f_{\mathbf{X},T-\Delta}(t) - f_{\mathbf{X},T}(t+\delta)\big)\,dt \tag{3.11}
$$

$$
\leq \int_{\varepsilon_1}^{\varepsilon_1+\delta} f_{\mathbf{X},T}(t)\,dt + \int_{T-\Delta-\varepsilon_2+\delta}^{T-\varepsilon_2} f_{\mathbf{X},T}(t)\,dt\,.
$$

Proof. We simply use Lemma 3.2.1. For any Borel set $B \subseteq (0, T-\Delta)$ we have

$$
\int_B f_{\mathbf{X},T-\Delta}(t)\,dt = P\big(\tau_{\mathbf{X},T-\Delta} \in B\big) \geq P\big(\tau_{\mathbf{X},[-\delta,T-\delta]} \in B\big)
$$

$$
= \int_B f_{\mathbf{X},[-\delta,T-\delta]}(t)\,dt = \int_B f_{\mathbf{X},T}(t+\delta)\,dt\,,
$$

which shows that $f_{\mathbf{X},T-\Delta}(t) \geq f_{\mathbf{X},T}(t+\delta)$ almost everywhere in $(0, T-\Delta)$.

For (3.11), notice that by Lemma 3.2.1,

$$
\int_{\varepsilon_1}^{T-\Delta-\varepsilon_2} \big(f_{\mathbf{X},T-\Delta}(t) - f_{\mathbf{X},T}(t+\delta)\big)\,dt
$$

$$
= P\big(\tau_{\mathbf{X},T-\Delta} \in (\varepsilon_1, T-\Delta-\varepsilon_2)\big) - P\big(\tau_{\mathbf{X},T} \in (\varepsilon_1+\delta, T-\Delta-\varepsilon_2+\delta)\big)
$$

$$
= P\big(\tau_{\mathbf{X},T} \notin (\varepsilon_1+\delta, T-\Delta-\varepsilon_2+\delta)\big) - P\big(\tau_{\mathbf{X},T-\Delta} \notin (\varepsilon_1, T-\Delta-\varepsilon_2)\big)
$$

$$
= P\big(\tau_{\mathbf{X},T} \in [0, \varepsilon_1+\delta)\big) + P\big(\tau_{\mathbf{X},T} \in (T-\Delta-\varepsilon_2+\delta, T]\big)
$$

$$-P\big(\tau_{\mathbf{X},T-\Delta} \in [0,\varepsilon_1)\big) - P\big(\tau_{\mathbf{X},T-\Delta} \in (T-\Delta-\varepsilon_2, T-\Delta]\big)$$

$$= P\big(\tau_{\mathbf{X},T} \in (\varepsilon_1, \varepsilon_1+\delta)\big) + \Big(P\big(\tau_{\mathbf{X},T} \in [0,\varepsilon_1)\big) - P\big(\tau_{\mathbf{X},T-\Delta} \in [0,\varepsilon_1)\big)\Big)$$

$$+P\big(\tau_{\mathbf{X},T} \in (T-\Delta-\varepsilon_2+\delta, T-\varepsilon_2)\big)$$

$$+\Big(P\big(\tau_{\mathbf{X},T} \in (T-\varepsilon_2, T]\big) - P\big(\tau_{\mathbf{X},[\Delta,T]} \in (T-\varepsilon_2, T]\big)\Big)$$

$$\le P\big(\tau_{\mathbf{X},T} \in (\varepsilon_1, \varepsilon_1+\delta)\big) + P\big(\tau_{\mathbf{X},T} \in (T-\Delta-\varepsilon_2+\delta, T-\varepsilon_2)\big)$$

$$= \int_{\varepsilon_1}^{\varepsilon_1+\delta} f_{\mathbf{X},T}(t)\,dt + \int_{T-\Delta-\varepsilon_2+\delta}^{T-\varepsilon_2} f_{\mathbf{X},T}(t)\,dt\,,$$

as required. □

We return now to the proof of Theorem 3.3.1. Our next goal is to prove that the cdf $F_{\mathbf{X},T}$ is right differentiable at every point in the interval $(0,T)$. Since we already know that $F_{\mathbf{X},T}$ is absolutely continuous on $(0,T)$, the set

$$A = \big\{t \in (0,T):\ F_{\mathbf{X},T} \text{ is not right differentiable at } t\big\} \tag{3.12}$$

has Lebesgue measure zero. Define next

$$B = \big\{t \in A^c:\ f_{\mathbf{X},T} \text{ restricted to } A^c \text{ does not have a right limit at } t\big\}. \tag{3.13}$$

We claim that the set B is at most countable. To see this, we define for $t \in A^c$

$$L(t) = \limsup_{s\downarrow t,\, s\in A^c} f_{\mathbf{X},T}(s), \quad l(t) = \liminf_{s\downarrow t,\, s\in A^c} f_{\mathbf{X},T}(s)\,.$$

Our claim about set B will follow once we check that for any $0 < \varepsilon < T/2$ and $\theta > 0$, the set

$$B_{\varepsilon,\theta} = \big\{t \in A^c \cap (\varepsilon, T-\varepsilon):\ L(t) - l(t) > \theta\big\}$$

is finite. In fact, we will show that the cardinality of $B_{\varepsilon,\theta}$ cannot be larger than $4/(\varepsilon\theta)$. If not, let $N > 4/(\varepsilon\theta)$ and find points $\varepsilon < t_1 < t_2 < \ldots < t_N < T-\varepsilon$. Choose $\delta > 0$ so small that $\delta < \varepsilon/2$ and

$$0 < \delta < \frac{1}{2}\min\Big(t_1-\varepsilon, t_2-t_1, \ldots, t_N - t_{N_1}, T-\varepsilon-t_N\Big)\,.$$

Let now $i = 1, \ldots, N$ and choose a sequence $s_n \downarrow t_i$, $s_n \in A^c$, such that $f_{\mathbf{X},T}(s_n) \to L(t_i)$. Consider n so large that $s_n - t_i < \delta/3$, and let

$$j \geq \frac{3}{\delta - (s_n - t_i)}$$

be an integer. We have

$$P\big(\tau_{\mathbf{X},T-\delta} \in (t_i - \delta, t_i)\big) \geq \sum_{k=0}^{\lfloor j(\delta-(s_n-t_i))\rfloor - 1} P\big(\tau_{\mathbf{X},T-\delta} \in (t_i - (k+1)/j, t_i - k/j)\big)\,,$$

and for each k as in the sum,

$$h_k := s_n - t_i + \frac{k+1}{j} \in (0, \delta]\,.$$

Therefore, by Lemma 3.2.1

$$\begin{aligned} &P\big(\tau_{\mathbf{X},T-\delta} \in (t_i - \delta, t_i)\big) \\ &\geq \sum_{k=0}^{\lfloor j(\delta-(s_n-t_i))\rfloor - 1} P\big(\tau_{\mathbf{X},T} \in (t_i - (k+1)/j + h_k, t_i - k/j + h_k)\big) \\ &= \lfloor j(\delta - (s_n - t_i))\rfloor P\big(\tau_{\mathbf{X},T} \in (s_n, s_n + 1/j)\big) \to (\delta - (s_n - t_i)) f_{\mathbf{X},T}(s_n) \end{aligned}$$

as $j \to \infty$. Letting $n \to \infty$, we conclude that

$$P\big(\tau_{\mathbf{X},T-\delta} \in (t_i - \delta, t_i)\big) \geq \delta L(t_i),\ i = 1, \ldots, N\,. \tag{3.14}$$

Similarly, for $i = 1, \ldots, N$ choose a sequence $w_n \downarrow t_i$, $w_n \in A^c$, such that $f_{\mathbf{X},T}(w_n) \to l(t_i)$. For large n and j we have

$$\begin{aligned} P\big(\tau_{\mathbf{X},T+\delta} \in (t_i, t_i + \delta)\big) &= P\big(\tau_{\mathbf{X},T+\delta} \in (t_i, w_n)\big) + P\big(\tau_{\mathbf{X},T+\delta} \in (w_n, w_n + \delta)\big) \\ &\leq P\big(\tau_{\mathbf{X},T+\delta} \in (t_i, w_n)\big) + \sum_{k=0}^{\lceil \delta j\rceil - 1} P\big(\tau_{\mathbf{X},T+\delta} \in (w_n + k/j, w_n + (k+1)/j)\big)\,. \end{aligned}$$

For each k as in the sum above,

$$h_k := \frac{k}{j} \in [0, \delta]\,.$$

Therefore, by Lemma 3.2.1,

$$P\big(\tau_{\mathbf{X},T+\delta} \in (t_i, t_i+\delta)\big)$$

$$\le P\big(\tau_{\mathbf{X},T+\delta} \in (t_i, w_n)\big) + \lceil \delta j \rceil P\big(\tau_{\mathbf{X},T} \in (w_n, w_n+1/j)\big).$$

Letting, once again, first $j \to \infty$ and then $n \to \infty$, we conclude that

$$P\big(\tau_{\mathbf{X},T+\delta} \in (t_i, t_i+\delta)\big) \le \delta l(t_i),\ i=1,\dots,N\,. \tag{3.15}$$

Now we use the estimate in Lemma 3.3.1 as follows. By the definition of the point t_i and the smallness of δ,

$$N\delta\theta \le P\left(\tau_{\mathbf{X},T-\delta} \in \bigcup_{i=1}^{N}(t_i-\delta, t_i)\right) - P\left(\tau_{\mathbf{X},T+\delta} \in \bigcup_{i=1}^{N}(t_i, t_i+\delta)\right)$$

$$= \int_{\cup_{i=1}^{N}(t_i-\delta,t_i)} \big(f_{\mathbf{X},T-\delta}(t) - f_{\mathbf{X},T+\delta}(t+\delta)\big)\,.$$

Using the fact that

$$\bigcup_{i=1}^{N}(t_i-\delta, t_i) \subset (\varepsilon-\delta, T-\varepsilon)\,,$$

and that, by Lemma 3.3.1, the integrand above is a.e. nonnegative, we have by the estimate in that lemma that the integral above does not exceed

$$\int_{\varepsilon-\delta}^{T-\varepsilon} \big(f_{\mathbf{X},T-\delta}(t) - f_{\mathbf{X},T+\delta}(t+\delta)\big)\,dt$$

$$\le \int_{\varepsilon-\delta}^{\varepsilon} f_{\mathbf{X},T+\delta}(t)\,dt + \int_{T-\varepsilon+\delta}^{T-\varepsilon+2\delta} f_{\mathbf{X},T+\delta}(t)\,dt\,.$$

Applying the already proved (3.1), we conclude that

$$N\delta\theta \le 2\frac{\delta}{\varepsilon-\delta} \le \frac{4\delta}{\varepsilon}\,,$$

and this contradicts the assumption that we can choose $N > 4/(\varepsilon\theta)$. This proves that the set B in (3.13) is at most countable. We notice, further, that

$$f_{\mathbf{X},T}(t) = \lim_{s\downarrow t}\frac{1}{s-t}P\big(t < \tau_{\mathbf{X},T} \le s\big) \tag{3.16}$$

$$= \lim_{s\downarrow t} \frac{1}{s-t}\int_t^s f_{\mathbf{X},T}(w)\,dw \;=\; \lim_{w\downarrow t,\, w\in A^c\setminus B} f_{\mathbf{X},T}(w)$$

for every $t \in A^c \setminus B$ (recall the set A is defined in (3.12)).

Now we are ready to prove that the right derivative of the cdf $F_{\mathbf{X},T}$ exists at every point in the interval $(0,T)$. Suppose, to the contrary, that this is not so. Then there is $t \in (0,T)$ and real numbers $a < b$ such that

$$\liminf_{\varepsilon\downarrow 0} \frac{F_{\mathbf{X},T}(t+\varepsilon) - F_{\mathbf{X},T}(t)}{\varepsilon} < a < b < \limsup_{\varepsilon\downarrow 0} \frac{F_{\mathbf{X},T}(t+\varepsilon) - F_{\mathbf{X},T}(t)}{\varepsilon}.$$

This implies that there is a sequence $t_n \downarrow t$ with $t_n \in A^c \setminus B$ for each n such that

$$f_{\mathbf{X},T}(t_{2n-1}) > b,\; f_{\mathbf{X},T}(t_{2n}) < a \text{ for all } n = 1,2,\ldots.$$

We can and will choose t_1 so close to t that $t_1 < (T+t)/2$.

Notice that by (3.16), for every $n = 1,2,\ldots$ there is $\delta_n > 0$ such that

$$f_{\mathbf{X},T}(w) > b \text{ a.e. in } (t_{2n-1}, t_{2n-1} + \delta_{2n-1})\,,$$

$$f_{\mathbf{X},T}(w) < a \text{ a.e. in } (t_{2n}, t_{2n} + \delta_{2n})$$

for $n = 1,2,\ldots$.

Let now $m \geq 1$, and consider $s > 0$ so small that both $s < \min_{n=1,\ldots,2m} \delta_n$ and $t_1 < (T+t)/2 - s$. Observe that

$$\int_t^{(T+t)/2} \big(f_{\mathbf{X},T}(w+s) - f_{\mathbf{X},T}(w)\big)_+\, dw$$

$$\geq \int_t^{t+s} \sum_{i=0}^{\lfloor (T-t)/2s\rfloor - 1} \big(f_{\mathbf{X},T}(w+(i+1)s) - f_{\mathbf{X},T}(w+is)\big)_+\, dw\,,$$

and for every point $w \in (t, t+s)$, each one of the intervals $(t_n, t_n+\delta_n)$, $n = 1,\ldots,2m$, contains at least one of the points in the finite sequence $w + is$, $i = 0,1,\ldots,\lfloor (T-t)/2s\rfloor - 1$. By construction, apart from a set of points $w \in (t, t+s)$

of measure zero, those points of the kind $w + is$ that fall in the odd-numbered intervals satisfy $f_{\mathbf{X},T}(w+is) > b$, and those points that fall in the even-numbered intervals satisfy $f_{\mathbf{X},T}(w+is) < a$. We conclude that

$$\sum_{i=0}^{\lfloor (T-t)/2s \rfloor - 1} \big(f_{\mathbf{X},T}(w+(i+1)s) - f_{\mathbf{X},T}(w+is)\big)_+ \geq m(b-a)$$

a.e. in $(t, t+s)$. Therefore, for all $s > 0$ small enough,

$$\int_t^{(T+t)/2} \big(f_{\mathbf{X},T}(w+s) - f_{\mathbf{X},T}(w)\big)_+ \, dw \geq sm(b-a)$$

and, since m can be taken arbitrarily large, we conclude that

$$\lim_{s \downarrow 0} \frac{1}{s} \int_t^{(T+t)/2} \big(f_{\mathbf{X},T}(w+s) - f_{\mathbf{X},T}(w)\big)_+ \, dw = \infty \, . \tag{3.17}$$

We will see that this is, however, impossible, and the resulting contradiction will prove that the right derivative of the cdf $F_{\mathbf{X},T}$ exists at every point in the interval $(0, T)$.

Indeed, recall that by Lemma 3.3.1, for all $s > 0$ small enough,

$$f_{\mathbf{X},T-2s}(w-s) \geq f_{\mathbf{X},T}(w+s) \text{ a.e. on } (s, T-s) \supset (t, (T+t)/2).$$

Therefore, for such s,

$$\begin{aligned} &\int_t^{(T+t)/2} \big(f_{\mathbf{X},T}(w+s) - f_{\mathbf{X},T}(w)\big)_+ \, dw \\ &\leq \int_t^{(T+t)/2} \big(f_{\mathbf{X},T-2s}(w-s) - f_{\mathbf{X},T}(w)\big)_+ \, dw \\ &\leq \int_{t-s}^{(T+t)/2-s} \big(f_{\mathbf{X},T-2s}(w) - f_{\mathbf{X},T}(w+s)\big) \, dw \end{aligned}$$

since, by another application of Lemma 3.3.1, the integrand is a.e. nonnegative over the range of integration. Applying (3.11), we see that

$$\int_t^{(T+t)/2} \big(f_{\mathbf{X},T}(w+s) - f_{\mathbf{X},T}(w)\big)_+ \, dw$$

$$\leq \int_{t-s}^{t} f_{\mathbf{X},T}(w)\, dw + \int_{(T+t)/2}^{(T+t)/2+s} f_{\mathbf{X},T}(w)\, dw \, .$$

However, we already know that the density $f_{\mathbf{X},T}$ is bounded on any subinterval of $(0,T)$ that is bounded away from both endpoints. Therefore, the upper bound obtained above shows that (3.17) is impossible. Hence the existence of the right derivative everywhere, which then coincides with the version of the density $f_{\mathbf{X},T}$ chosen above.

Next we check that this version of the density is right continuous. To this end we recall that we already know that the set A in (3.12) is empty. Next, we rule out existence of a point $t \in (0,T)$ such the limit of $f_{\mathbf{X},T}(s)$ as $s \downarrow t$ over $s \in B^c$ does not exist. Suppose that, to the contrary, that such t exists. This means that there are real numbers $a < b$ and a sequence $t_n \downarrow t$ with $t_n \in B^c$ for each n such that

$$f_{\mathbf{X},T}(t_{2n-1}) > b, \; f_{\mathbf{X},T}(t_{2n}) < a \text{ for all } n = 1, 2, \ldots .$$

However, we have already established that such a sequence cannot exist.

As in (3.16), we see that for every $t \in (0,T)$

$$f_{\mathbf{X},T}(t) = \lim_{s\downarrow t,\, s\in B^c} f_{\mathbf{X},T}(s)$$

and, since the set B is at most countable, the restriction to $s \in B^c$ in the above limit statement can be removed. This proves right continuity of the version of the density density given by the right derivative of $F_{\mathbf{X},T}$. The proof of existence of left limits is similar.

Next, we address the variation of the version of the density we are working with away from the endpoints of the interval $(0,T)$. Let $0 < t_1 < t_2 < T$. We start with a preliminary calculation. Let $0 < r_n < T - t_2$. Introduce the notation

$$C_+ = \big\{ t \in (t_1, t_2) : \; f_{\mathbf{X},T}(t + r_n) \geq f_{\mathbf{X},T}(t) \big\} \, ,$$

$$C_- = \big\{t \in (t_1, t_2) :\ f_{\mathbf{X},T}(t + r_n) < f_{\mathbf{X},T}(t)\big\}\,,$$

so that

$$\begin{aligned}&\int_{t_1}^{t_2} \big|f_{\mathbf{X},T}(t + r_n) - f_{\mathbf{X},T}(t)\big|\,dt \\ &= \int_{C_+} \big(f_{\mathbf{X},T}(t + r_n) - f_{\mathbf{X},T}(t)\big)\,dt + \int_{C_-} \big(f_{\mathbf{X},T}(t) - f_{\mathbf{X},T}(t + r_n)\big)\,dt\,.\end{aligned}$$

To estimate the two terms we will once again use Lemma 3.3.1. Since

$$f_{\mathbf{X},T-r_n}(t) \geq f_{\mathbf{X},T}(r_n + t) \text{ a.e. on } (0, T - r_n) \supset (t_1, t_2)$$

for n large enough, for such n, we have the upper bound

$$\begin{aligned}\int_{C_+} \big(f_{\mathbf{X},T}(t + r_n) - f_{\mathbf{X},T}(t)\big)\,dt &\leq \int_{C_+} \big(f_{\mathbf{X},T-r_n}(t) - f_{\mathbf{X},T}(t)\big)\,dt \\ &\leq \int_{t_1}^{t_2} \big(f_{\mathbf{X},T-r_n}(t) - f_{\mathbf{X},T}(t)\big)\,dt\,.\end{aligned}$$

We now once again use (3.11) to conclude that for all n large, we have

$$\int_{C_+} \big(f_{\mathbf{X},T}(t + r_n) - f_{\mathbf{X},T}(t)\big)\,dt \leq \int_{t_2}^{t_2+r_n} f_{\mathbf{X},T}(t)\,dt$$

so that

$$\limsup_{n\to\infty} \frac{1}{r_n} \int_{C_+} \big(f_{\mathbf{X},T}(t + r_n) - f_{\mathbf{X},T}(t)\big)\,dt \leq f_{\mathbf{X},T}(t_2)\,.$$

Similarly, by Lemma 3.3.1,

$$f_{\mathbf{X},T}(t + r_n) \geq f_{\mathbf{X},T+r_n}(t + r_n) \text{ a.e. on } (0, T - r_n) \supset (t_1, t_2)$$

for n large enough, and we obtain, for such n, using (3.11)

$$\begin{aligned}\int_{C_-} \big(f_{\mathbf{X},T}(t) - f_{\mathbf{X},T}(t + r_n)\big)\,dt &\leq \int_{C_-} \big(f_{\mathbf{X},T}(t) - f_{\mathbf{X},T+r_n}(t + r_n)\big)\,dt \\ &\leq \int_{t_1}^{t_2} \big(f_{\mathbf{X},T}(t) - f_{\mathbf{X},T+r_n}(t + r_n)\big)\,dt \leq \int_{t_1}^{t_1+r_n} f_{\mathbf{X},T+r_n}(t)\,dt\,.\end{aligned}$$

This can, in turn, be bounded from above both by

$$\int_{t_1}^{t_1+r_n} f_{\mathbf{X},T}(t)\,dt$$

and by

$$\int_{t_1}^{t_1+r_n} f_{\mathbf{X},T}(t-r_n)\,dt = \int_{t_1-r_n}^{t_1} f_{\mathbf{X},T}(t)\,dt\,.$$

Therefore,

$$\limsup_{n\to\infty} \frac{1}{r_n} \int_{C_-} \bigl(f_{\mathbf{X},T}(t) - f_{\mathbf{X},T}(t+r_n)\bigr)\,dt \le \min\bigl(f_{\mathbf{X},T}(t_1), f_{\mathbf{X},T}(t_1-)\bigr)\,.$$

Overall, we have proved that

$$\begin{aligned} &\limsup_{n\to\infty} \frac{1}{r_n} \int_{t_1}^{t_2} \bigl|f_{\mathbf{X},T}(t+r_n) - f_{\mathbf{X},T}(t)\bigr|\,dt \\ &\le \min\bigl(f_{\mathbf{X},T}(t_1), f_{\mathbf{X},T}(t_1-)\bigr) + f_{\mathbf{X},T}(t_2)\,. \end{aligned} \tag{3.18}$$

To relate (3.18) to the total variation of the density $f_{\mathbf{X},T}$ over the interval (t_1, t_2), we notice first that by the right continuity of the density, it is enough to consider the regularly spaced points $s_i = t_1 + ir_n,\ i = 1, \ldots, n$, where $r_n = (t_2 - t_1)/(n+1)$ for some $n = 1, 2, \ldots$. Write

$$\int_{t_1}^{t_2} \bigl|f_{\mathbf{X},T}(t+r_n) - f_{\mathbf{X},T}(t)\bigr|\,dt = \int_{t_1}^{t_1+r_n} \sum_{i=0}^{n} \bigl|f_{\mathbf{X},T}(t+(i+1)r_n) - f_{\mathbf{X},T}(t+ir_n)\bigr|\,dt$$

and observe that

$$\lim_{n\to\infty} \sum_{i=0}^{n} \bigl|f_{\mathbf{X},T}(t+(i+1)r_n) - f_{\mathbf{X},T}(t+ir_n)\bigr| \ge TV_{(t_1,t_2)}(f_{\mathbf{X},T})$$

uniformly in $t \in (t_1, t_2)$. Therefore, by (3.18)

$$\min\bigl(f_{\mathbf{X},T}(t_1), f_{\mathbf{X},T}(t_1-)\bigr) + f_{\mathbf{X},T}(t_2) \ge \limsup_{n\to\infty} \frac{1}{r_n} \int_{t_1}^{t_2} \bigl|f_{\mathbf{X},T}(t+r_n) - f_{\mathbf{X},T}(t)\bigr|\,dt$$

$$\ge \limsup_{n\to\infty} \frac{1}{r_n} \int_{t_1}^{t_1+r_n} \sum_{i=0}^{n} \bigl|f_{\mathbf{X},T}(t+(i+1)r_n) - f_{\mathbf{X},T}(t+ir_n)\bigr|\,dt \ge TV_{(t_1,t_2)}(f_{\mathbf{X},T})\,.$$

Now the bound (3.3) follows from the obvious fact that

$$TV_{(t_1,t_2)}(f_{\mathbf{X},T}) = \lim_{\varepsilon\downarrow 0} TV_{(t_1,t_2-\varepsilon)}(f_{\mathbf{X},T}),\,.$$

Furthermore, the proof of (3.4) and (3.5) is the same as the proof of (3.3), with each one using one side of the two-sided calculation performed above for (3.3).

Next, the boundedness of the positive variation of the density at zero, clearly, implies that the limit $f_{\mathbf{X},T}(0+) = \lim_{t\downarrow 0} f_{\mathbf{X},T}(t)$ exists, while the boundedness of the negative variation of the density at T implies that the limit $f_{\mathbf{X},T}(T-) = \lim_{t\uparrow T} f_{\mathbf{X},T}(t)$ exists as well. If $TV_{(0,\varepsilon)}(f_{\mathbf{X},T}) < \infty$ for some $0 < \varepsilon < T$, then, trivially, $f_{\mathbf{X},T}(0+) < \infty$. On the other hand, if $f_{\mathbf{X},T}(0+) < \infty$, then the same argument as we used in proving (3.3), shows that for any $0 < \varepsilon < T$,

$$TV^{-}_{(0,\varepsilon)}(f_{\mathbf{X},T}) \leq f_{\mathbf{X},T}(0+)\,,$$

which, together with (3.4), both shows that $TV_{(0,\varepsilon)}(f_{\mathbf{X},T}) < \infty$ and proves (3.6). One can prove the statement of part (f) of the theorem concerning the behaviour of the density at the right end point of the interval in the same way.

It only remains to prove part (c) of the theorem, namely the fact that the version of the density given by the right derivative of the cdf $F_{\mathbf{X},T}$ is bounded away from zero. Recall that Assumption U_T is in effect here.

Suppose, to the contrary, that (3.2) fails and introduce the notation

$$t_1 = \inf\big\{s \in (0,T) : \inf_{0<t<s} f_{\mathbf{X},T}(t) = 0\big\},$$

$$t_2 = \sup\big\{s \in (0,T) : \inf_{s<t<T} f_{\mathbf{X},T}(t) = 0\big\}\,.$$

Clearly, $0 \leq t_1 \leq t_2 \leq T$. We claim that,

$$\text{if } t_1 < t_2, \text{ then } f_{\mathbf{X},T}(t) = 0 \text{ for all } t_1 < t < t_2. \tag{3.19}$$

We start with the case $0 < t_1 < t_2 < T$. Notice that, in this case,

$$\min\big(f_{\mathbf{X},T}(t_1), f_{\mathbf{X},T}(t_1-)\big) = \min\big(f_{\mathbf{X},T}(t_2), f_{\mathbf{X},T}(t_2-)\big) = 0\,.$$

By (3.3) the density is constant on the interval (t_1, t_2). If $f_{\mathbf{X},T}(t_1) = 0$ then, by the right continuity of the density the constant must be equal to zero, so (3.19) is immediate. If $f_{\mathbf{X},T}(t_1-) = 0$ then, given $\varepsilon > 0$, choose $0 < s < t_1$ such that $f_{\mathbf{X},T}(s) \leq \varepsilon$. By (3.3) we know that $TV_{(s,t_2)}(f_{\mathbf{X},T}) \leq \varepsilon$, which implies that $f(t) \leq 2\varepsilon$ on (s, t_2), hence also on (t_1, t_2). Letting $\varepsilon \to 0$ proves (3.19). If either $t_1 = 0$ and/or $t_2 = T$, then (3.19) can be proved using a similar argument and the continuity of the density at 0 and at T shown in part (a) of the theorem. Furthermore, we also have

$$\text{if } t_1 = t_2, \text{ then } \min\bigl(f_{\mathbf{X},T}(t_1),\, f_{\mathbf{X},T}(t_1-)\bigr) = 0\,, \tag{3.20}$$

with the obvious conventions in the case $t_1 = t_2$ coincide with one of the endpoints of the interval.

It follows from (3.19), (3.20) and Lemma 3.3.1 that for any $\Delta > 0$,

$$f_{\mathbf{X},T+\Delta}(t) = 0 \text{ for } t_1 < t < t_2 + \Delta. \tag{3.21}$$

Furthermore, we know by Lemma 3.2.1 that

$$F_{\mathbf{X},T+\Delta}([0, t_1]) \leq F_{\mathbf{X},T}([0, t_1]) \tag{3.22}$$

and

$$F_{\mathbf{X},T+\Delta}([t_2 + \Delta, T + \Delta]) \leq F_{\mathbf{X},T}([t_2, T])\,. \tag{3.23}$$

Note that for $\Delta > 0$ all the quantities in the above equations refer to the leftmost location $\tau_{\mathbf{X},T+\Delta}$ of the supremum, which is no longer assumed to be unique.

Since the distributions $F_{\mathbf{X},T}$ and $F_{\mathbf{X},T+\Delta}$ have equal total masses (equal to one), it follows from (3.21), (3.22) and (3.23) that the latter two inequalities must hold as equalities for all relevant sets. We concentrate on the resulting equation

$$F_{\mathbf{X},T+\Delta}([t_2 + \Delta, T + \Delta]) = F_{\mathbf{X},T}([t_2, T])\,. \tag{3.24}$$

Since we are working with the leftmost supremum location on a larger interval, we can write for $\Delta > 0$

$$P\big(\tau_{\mathbf{X},T} \in [t_2, T]\big) = P\big(\tau_{\mathbf{X},[-\Delta,T]} \in [t_2, T]\big)$$

$$+P\big(\tau_{\mathbf{X},T} \in [t_2, T],\, \tau_{\mathbf{X},[-\Delta,T]} \in [-\Delta, 0)\big).$$

Using Lemma 3.2.1 and (3.24) we see that

$$P\big(\tau_{\mathbf{X},T} \in [t_2, T],\, \tau_{\mathbf{X},[-\Delta,T]} \in [-\Delta, 0)\big) = 0\,,$$

which implies that, if $\Delta > T - t_2$, then

$$P\big(\tau_{\mathbf{X},T} \in [t_2, T],\, \sup_{-\Delta \le t \le -\Delta + T - t_2} X(t) \ge \sup_{t_2 \le t \le T} X(t)\big) = 0\,. \tag{3.25}$$

Pick $\delta > T$. Using (3.25) with $\Delta = n\delta - t_2,\ n = 1, 2, \ldots$, we see that

$$Y_n < Y_0 \text{ a.e. on } \{\tau_{\mathbf{X},T} \in [t_2, T]\} \text{ for } n = 1, 2, \ldots,$$

where $Y_n = \sup_{t_2 - n\delta \le t \le T - n\delta} X(t),\ n = 0, 1, 2, \ldots$. Note, however, that the sequence $(Y_n,\ n = 0, 1, 2, \ldots)$ is stationary, and for a stationary sequence it is impossible that, on a set of positive probability, $Y_0 > Y_n$ for $n = 1, 2, \ldots$ (this is clear for an ergodic sequence; in general one can use the ergodic decomposition). We conclude that

$$P\big(\tau_{\mathbf{X},T} \in [t_2, T]\big) = 0\,. \tag{3.26}$$

Reversing the direction of time (or, equivalently, switching to the rightmost supremum location on a larger interval) and using Assumption U_T, we also have

$$P\big(\tau_{\mathbf{X},T} \in [0, t_1]\big) = 0\,. \tag{3.27}$$

However, (3.19), (3.26) and (3.27) rule out any possible mass of the distribution $F_{\mathbf{X},T}$. This contradiction shows that, under Assumption U_T, the version of the density given by the right derivative of the cdf $F_{\mathbf{X},T}$ is bounded away from zero. This completes the proof of the theorem. $\square$

Remark 3.3.2. The following example shows that the statement of part (c) of Theorem 3.3.1 may fail without Assumption U$_T$.

Let $(x(t),\, t \in \mathbb{R})$ be a continuous periodic function with period 1, for which $t = 0$ is a global maximum. Let U be a standard uniform random variable. Then $(X(t) = x(t+U),\, t \in \mathbb{R})$ is a continuous stationary process, that always attains its global maximum in the interval $[0,1]$. Therefore, with $T > 1$, we have $f_{\mathbf{X},T}(t) = 0$ for $1 \le t < T$.

Next we describe what extra restrictions on the distribution of the location of the supremum, in addition to the statements of Theorem 3.3.1, Assumption L of Section 3.2 imposes. Again, one of the statements of the theorem requires Assumption U$_T$. See Remark 3.3.6 for a discussion.

Theorem 3.3.3. *Let $\mathbf{X} = (X(t),\, t \in \mathbb{R})$ be a stationary sample upper semi-continuous process, satisfying Assumption L. Then the version of the density $f_{\mathbf{X},T}$ of the leftmost location of the supremum in the interval $[0,T]$ described in Theorem 3.3.1 has the following additional properties.*

(a) $f_{\mathbf{X},T}(0+) < \infty$, $f_{\mathbf{X},T}(T-) < \infty$ and $TV_{(0,T)}(f_{\mathbf{X},T}) \le f_{\mathbf{X},T}(0+) + f_{\mathbf{X},T}(T-)$. In particular, the density has a bounded variation on the entire interval $(0,T)$.

(b) Assume additionally that the process is sample continuous and satisfies Assumption U$_T$. Then either $f_{\mathbf{X},T}(t) = 1/T$ for all $0 < t < T$, or $\int_0^T f_{\mathbf{X},T}(t)\, dt < 1$.

Remark 3.3.4. Theorem 3.3.3 provides a list of specific conditions that the distribution of the supremum location has to satisfy (under Assumptions U$_T$ and L).

The list turns out to be complete. That is, for any function f satisfying the conditions described in the theorem, there is a sample continuous stationary process satisfying Assumption U_T and Assumption L, for which f is the density of the supremum location. Thus we have obtained a full characterization of the set of all possible densities. In order to decide whether a candidate function can be the density of the supremum location for some stationary process, we only need to check the list of conditions given in the theorem. This is, of course a much easier task than trying to construct an appropriate process. We refer the reader to the next chapter for details and proofs.

Remark 3.3.5. Note that part (b) of Theorem 3.3.3 says that, unless the location of the supremum is uniformly distributed in the interval $(0,T)$, the supremum is achieved, with a positive probability, at an endpoint of the interval. The proof of this part, exhibited in the following pages, actually implies more. It shows that the uniform distribution occurs only when the suprema of the process appear periodically with period equal to T:

$$P\Big(X\big(\tau_{\mathbf{X},[T,2T]}\big) = X\big(\tau_{\mathbf{X},T}\big),\ \tau_{\mathbf{X},[T,2T]} - \tau_{\mathbf{X},T} = T\Big) = 1.$$

Proof of Theorem 3.3.3. Assumption L and stationarity imply that for any $0 < t < T$,

$$\begin{aligned} f_{\mathbf{X},T}(t) &= \lim_{\varepsilon\downarrow 0} \frac{P(\tau_{\mathbf{X},T} \in (t, t+\varepsilon))}{\varepsilon} \\ &\le \limsup_{\varepsilon\downarrow 0} \frac{P(\mathbf{X} \text{ has a local maximum in } (t, t+\varepsilon))}{\varepsilon} \\ &= \limsup_{\varepsilon\downarrow 0} \frac{P(\mathbf{X} \text{ has a local maximum in } (0, \varepsilon))}{\varepsilon} \le K\,. \end{aligned}$$

This proves finiteness of $f_{\mathbf{X},T}(0+) < \infty$ and $f_{\mathbf{X},T}(T-)$. The rest of the statement in part (a) follows from (3.6) by letting $\varepsilon \uparrow T$.

We now prove part (b). Assume that $P(\tau_{\mathbf{X},T} = 0 \text{ or } T) = 0$. By stationarity this implies that $\tau_{\mathbf{X},[T,2T]} \in (T,2T)$ with probability 1. We first prove that

$$P\Big(X\big(\tau_{\mathbf{X},[T,2T]}\big) \neq X\big(\tau_{\mathbf{X},T}\big)\Big) = 0\,. \tag{3.28}$$

By symmetry, it is enough to prove the one-sided claim

$$P\Big(X\big(\tau_{\mathbf{X},[T,2T]}\big) < X\big(\tau_{\mathbf{X},T}\big)\Big) = 0\,. \tag{3.29}$$

Indeed, suppose, to the contrary, that the probability in (3.29) is positive. Under Assumption U_T we can use the continuity from below of measures to see that there is $\varepsilon > 0$ such that

$$p := P\Big(X\big(\tau_{\mathbf{X},T}\big) > X\big(\tau_{\mathbf{X},[T,2T]}\big) + \varepsilon,\, X\big(\tau_{\mathbf{X},T}\big) > \max_{t\in L_T,\, t\neq \tau_{\mathbf{X},T}} X(t) + \varepsilon\Big) > 0\,.$$

Here L_T is the (a.s. finite) set of the local maxima of $\mathbf{X}$ in the interval $(0,T)$.

Next, by the uniform continuity of the process $\mathbf{X}$ on $[0,T]$, there is $n \geq 1$ such that

$$P\Big(\sup_{0\leq s<t\leq T,\, t-s\leq T/n} \big|X(t) - X(s)\big| > \varepsilon/2\Big) \leq p/2\,.$$

We immediately conclude by the law of total probability that there is $i = 1, \ldots, n$ such that $P(A_i) > 0$, where

$$A_i = \Big\{X\big(\tau_{\mathbf{X},T}\big) > X\big(\tau_{\mathbf{X},[T,2T]}\big) + \varepsilon,\, X\big(\tau_{\mathbf{X},T}\big) > \max_{t\in L_T,\, t\neq \tau_{\mathbf{X},T}} X(t) + \varepsilon,$$
$$(i-1)T/n < \tau_{\mathbf{X},T} < iT/n,\, \sup_{(i-1)T/n\leq s,t\leq iT/n} \big|X(t) - X(s)\big| \leq \varepsilon/2\Big\}.$$

However, on the event A_i, $X(iT/n) = \sup_{iT/n\leq t\leq 2T} X(t)$, implying that $\tau_{\mathbf{X},[iT/n,iT/n+T]} = iT/n$. By stationarity, this contradicts the assumption $P(\tau_{\mathbf{X},T} = 0) = 0$. This contradiction proves (3.29) and, hence, also (3.28).

Next, we check that

$$P\Big(X\big(\tau_{\mathbf{X},[T,2T]}\big) = X\big(\tau_{\mathbf{X},T}\big),\, \tau_{\mathbf{X},[T,2T]} - \tau_{\mathbf{X},T} < T\Big) = 0\,. \tag{3.30}$$

Indeed, suppose that, to the contrary, the probability above is positive. By the continuity from below of measures, there is $\varepsilon > 0$ such that

$$P\Big(X\big(\tau_{\mathbf{X},[T,2T]}\big) = X\big(\tau_{\mathbf{X},T}\big),\ \tau_{\mathbf{X},[T,2T]} - \tau_{\mathbf{X},T} < T - \varepsilon\Big) > 0\,.$$

Take $n > 2T/\varepsilon$. By the law of total probability there are $i_1, i_2 = 1, \ldots, n$ such that $P(A_{i_1,i_2}) > 0$, where

$$A_{i_1,i_2} = \Big\{X\big(\tau_{\mathbf{X},[T,2T]}\big) = X\big(\tau_{\mathbf{X},T}\big),\ \tau_{\mathbf{X},[T,2T]} - \tau_{\mathbf{X},T} < T - \varepsilon,$$
$$(i_1 - 1)T/n < \tau_{\mathbf{X},T} < i_1 T/n,\ T + (i_2 - 1)T/n < \tau_{\mathbf{X},[T,2T]} < T + i_2 T/n\Big\}.$$

By the choice of n, $T + i_2T/n - (i_1 - 1)T/n < T$, so that, on the event A_{i_1,i_2}, the process $\mathbf{X}$ has at least two points, $\tau_{\mathbf{X},T}$ and $\tau_{\mathbf{X},[T,2T]}$, at which the supremum over the interval $[(i_1 - 1)T/n,\ (i_1 - 1)T/n + T]$ is achieved. By stationarity, this contradicts Assumption U_T. This contradiction proves (3.30).

Finally, we check that

$$P\Big(X\big(\tau_{\mathbf{X},[T,2T]}\big) = X\big(\tau_{\mathbf{X},T}\big),\ \tau_{\mathbf{X},[T,2T]} - \tau_{\mathbf{X},T} > T\Big) = 0\,. \tag{3.31}$$

The proof is similar to the proof of (3.29), so we only sketch the argument. Suppose that, to the contrary, the probability in (3.31) is positive. Use the continuity of measures to see that the probability remains positive if we require that $\tau_{\mathbf{X},[T,2T]} - \tau_{\mathbf{X},T} > T + \varepsilon$ for some $\varepsilon > 0$. Next, use Assumption U_T to separate the value of $X\big(\tau_{\mathbf{X},T}\big)$ from the values of $\mathbf{X}$ at other local maxima in $(0, T)$ and, finally, use the uniform continuity of the process $\mathbf{X}$ to show that there is a point $T < b < 2T$ and an event of positive probability on which $\tau_{\mathbf{X},[b-T,b]} = b$. By stationarity, this contradicts the assumption $P(\tau_{\mathbf{X},T} = T) = 0$.

Combining (3.28), (3.30) and (3.31), we see that the assumption $P(\tau_{\mathbf{X},T} = 0 \text{ or } T) = 0$ implies that

$$P\Big(X\big(\tau_{\mathbf{X},[T,2T]}\big) = X\big(\tau_{\mathbf{X},T}\big),\ \tau_{\mathbf{X},[T,2T]} - \tau_{\mathbf{X},T} = T\Big) = 1 \tag{3.32}$$

Let $0 < a < b < T$. We have by stationarity,

$$P\big(\tau_{\mathbf{X},T} \in (0, b-a)\big) = P\big(\tau_{\mathbf{X},[a,a+T]} \in (a,b)\big)$$

$$= P\big(\tau_{\mathbf{X},[a,a+T]} \in (a,b),\, \tau_{\mathbf{X},T} \in (0,a)\big) + P\big(\tau_{\mathbf{X},[a,a+T]} \in (a,b),\, \tau_{\mathbf{X},T} \in (a,T)\big)\,.$$

By (3.32), if $\tau_{\mathbf{X},T} \in (0,a)$, then $\tau_{\mathbf{X},[T,2T]} \in (T, T+a)$ and $X\big(\tau_{\mathbf{X},[T,2T]}\big) > \sup_{t\in[a,b]} X(t)$. Therefore, the first term in the right hand side above vanishes. Similarly, by (3.32), if $\tau_{\mathbf{X},T} \in (a,T)$ then $\tau_{\mathbf{X},[T,2T]} \in (T+a, 2T)$, and $X\big(\tau_{\mathbf{X},T}\big) > \sup_{t\in[T,T+a]} X(t)$. Therefore,

$$P\big(\tau_{\mathbf{X},T} \in (0, b-a)\big) = P\big(\tau_{\mathbf{X},T} \in (a,b)\big)$$

for any $0 < a < b < T$, which proves the uniformity of the distribution of $\tau_{\mathbf{X},T}$. $\square$

Remark 3.3.6. A simple special case of the process in Remark 3.3.2 shows that the statement of part (b) of Theorem 3.3.3 may fail without Assumption U_T.

We take, for clarity, a specific function x. Let $x(t) = 1 - 2|t|$ for $|t| \le 1/2$ and extend x to a periodic function with period 1. Then for any $T > 1$, the leftmost location of the supremum in the interval $[0,T]$ of the process $(X(t) = x(t+U),\, t \in \mathbb{R})$ is in the interval $(0,1)$ with probability 1, and (as we already know) this location is not uniformly distributed between 0 and T.

None of the statement of Theorem 3.3.3 holds, in general, without Assumption L, as the following example shows.

Example 3.3.7. Let $X(t) = e^{-t/2}B(e^t)$, $t \ge 0$, where $(B(t))$ is the standard Brownian motion. Then $\mathbf{X}$ is a stationary Gaussian process, the Ornstein-Uhlenbeck process. It is, clearly, sample continuous, and the strong Markov property of the

Brownian motion shows that, for any $T > 0$, it satisfies Assumption U_T. It is clear that Assumption L fails for the Ornstein-Uhlenbeck process.

By the law of iterated logarithm for the Brownian motion we see that, on a set of probability 1, in any interval $(0, \varepsilon)$ with $\varepsilon > 0$ there is a point t such that $X(t) > X(0)$. Therefore, $P(\tau_{\mathbf{X},T} = 0) = 0$ and, similarly, $P(\tau_{\mathbf{X},T} = T) = 0$ for any $T > 0$.

It is also easy to show, using the basic properties of the Brownian motion, that the density $f_{\mathbf{X},T}$ is not bounded near each of the two endpoints of the interval $[0, T]$, so that both statements of Theorem 3.3.3 fail for this process.

It is worthwhile to point out that although the dichotomy result in part (b) of Theorem 3.3.3 no longer holds without assumption L, it can be modified into the following trichotomy:

Proposition 3.3.8. *Let* $\mathbf{X}$ *be as defined before. Under assumption* U_T, $\tau_{\mathbf{X},[0,T]}$ *is uniformly distributed on* $[0, T]$ *if and only if*

$$P\Big(X\big(\tau_{\mathbf{X},[T,2T]}\big) = X\big(\tau_{\mathbf{X},T}\big),\ \tau_{\mathbf{X},[T,2T]} - \tau_{\mathbf{X},T} = T\Big) = 1.$$

If $\tau_{\mathbf{X},[0,T]}$ *is not uniformly distributed, then either*

$$\int_0^T f_{\mathbf{X},T}(t)dt < 1,$$

or

$$f_{\mathbf{X},T}(0+) + f_{\mathbf{X},T}(T-) = \infty.$$

Proposition 3.3.8 basically says that the location of the supremum is uniformly distributed over the interval if and only if the extremes of the process appear exactly with period T. If this is not the case, then at least for one boundary of the interval, either there is a point mass on the boundary, or the density of the supremum location explodes when approaching to that boundary. Thus the dichotomy under assumption L now becomes a trichotomy, with the new possibility of an exploding density. A intuitive explanation is that when the process can oscillate infinitely fast, the probability mass which would be located on the boundary with smooth pathes will be "spread" by the small oscillations to its neighborhood, resulting in an exploding density as an alternative to the point mass.

Proof. Let M be the set of all local maxima of process $\mathbf{X}$. We define the following two subsets in M:

$$A = \{t \in M : \exists \delta_t > 0, X(t-\delta_t) > \sup_{s \in [t-T, t-\delta_t) \cap M} X(s) \vee X(t-T),$$

$$\text{and } X \text{ is strictly increasing on } (t-\delta_t, t)\},$$

and

$$B = \{t \in M : \exists \delta_t > 0, X(t) > \sup_{s \in [t-T, t-\delta_t)} X(s)\} \backslash A$$

Intuitively, A and B are both sets of maxima which "beat" all the other local maxima to their left within distance T, and also beat the points exactly of distance T to their left. They are further distinguished by whether there is a monotone neighborhood to the left of the maximum point, or the maximum is achieved via infinitely frequent oscillation. Symmetrically we also define the sets A' and B', which are the counterparts of A and B where the direction left is replaced by the direction right.

There are then three possible cases:

Case 1: $P(A \cup A' \neq \phi) > 0$;

Case 2: $P(B \cup B' \neq \phi) > 0$;

Case 3: $P(A \cup A' \cup B \cup B' = \phi) = 1$.

We will prove that these three cases correspond, respectively, to the point mass on the boundaries, exploding density near the boundaries, and uniform distribution of the location of the supremum.

Case 1. Without loss of generality, assume $P(A \neq \phi) > 0$. By localization and stationarity, case 1 can be rewritten as $P(A \cap (0, T] \neq \phi) = 0$. Then by the continuity from above of probability and the continuity of the path, there exists $\delta > 0$, such that the set

$$A_\delta := \{t \in A \cap (0, T] : \delta_t > \delta, X(t - \delta) > \sup_{s \in [t-T-\delta, t-\delta) \cap M} X(s) \vee X(t - T - \delta)\}$$

is nonempty with positive probability. For each t in A_δ, consider the interval $(t - \delta, t]$ with fixed length δ. Then since such a point t exists on $(0, T]$ with positive probability, it is easy to see that there must exist certain point $u \in (0, T]$, such that

$$P\left(u \in (t - \delta, t] \text{ for some } t \in A_\delta\right) > 0.$$

However, by definition of A_δ, $t \in A_\delta$ implies that $\tau_{\mathbf{X},[u-T,u]} = u$. Thus $P(\tau_{\mathbf{X},[u-T,u]} = u) > 0$: there exists a point mass at the boundary u. By stationarity, this is of course equivalent to the existence of a point mass at T for interval $[0, T]$.

Case 2. Similar to case 1, assume that $P(B \neq \phi) > 0$. The localization procedure, along with the continuity from above of the probability and the continuity of the path, guarantees that there exists $\delta > 0$, such that

$$B_\delta := \{t \in B \cap (0, T] : \delta_t > \delta, X(t) > \sup_{s \in [t-T-\delta, t-\delta)} X(s)\}$$

is nonempty with positive probability, say, p. For any point $t \in M$, define $d(t) := \inf\{d > 0 : X(t+d) \geq X(t)\}$ be the distance between the point t and the next time when the process goes back to the same level. It is then not hard to check that a point $t \in (0, T]$ is in B_δ, if and only if $\delta_t > \delta, X(t) > \sup_{s \in [t-T-\delta, t-\delta)} X(s)$, and there exists a sequence $\{t_i\}_{i \in \mathbb{N}} \subseteq [t - \delta_t, t) \cap M$, such that $t_i \uparrow t$, $d(t_i) > 0, \forall i \in \mathbb{N}$, and $X(t_i) > \sup_{s \in [t_i - T, t_i - \delta_t)} X(s) \vee X(t_i - T), \forall i \in \mathbb{N}$. Fix an integer n and some $\delta > 0$. For each path, take ϵ small enough such that at least n terms among $\{d(t_i)\}_{i=1,2,\ldots}$ are greater than ϵ. Thus there exists a number $\epsilon_{n,\delta} > 0$, such that

$$p - \delta \leq P(\epsilon \geq \epsilon_{n,\delta}) \leq P(\text{at least } n \text{ terms in } \{d(t_i)\} \text{ is greater than } \epsilon_{n,\delta}).$$

Define interval $I_i = (t_i, t_i + \epsilon_{n,\delta})$. Then by construction $\{I_i\}$ is a family of at least n (random) disjoint intervals, each with length $\epsilon_{n,\delta}$. Thus there exists a point $u \in [0, T]$, such that

$$P(u \in I_i \text{ for some } i) \geq \frac{n\epsilon_{n,\delta}(p - \delta)}{T}.$$

Notice that $u \in I_i$ for some i implies that $\tau_{\mathbf{X},[u-T,u]} \in [u - \epsilon_{n,\delta}, u]$. So

$$P(\tau_{\mathbf{X},[u-T,u]} \in [u - \epsilon_{n,\delta}, u]) \geq \frac{n\epsilon_{n,\delta}(p - \delta)}{T},$$

which clearly implies that there exists a point $v_n \in [u - \epsilon_{n,\delta}, u]$, for which the density $f_{\mathbf{X},[u-T,u]}(v) \geq \frac{n(p-\delta)}{T}$. Since n can be arbitrarily large, this shows that $f_{\mathbf{X},[u-T,u]}(u-) = \infty$, and therefore also $f_{\mathbf{X},T}(T-) = \infty$ by stationarity. Symmetric results hold for the case where $P(B' \neq \phi) > 0$, but then the derivation is in almost sure sense in stead of pathwise, since we will need to use assumption U_T in this situation.

Case 3. If none of case 1 or case 2 happens, then with probability 1, there is no point of maxima which “beats” all the other local maxima to its left within distance T and also beats the point exactly of distance T to its left. Symmetric result also holds for the direction right. Then consider the locations and values of the global maxima for two consecutive intervals $[0, T]$ and $[T, 2T]$. It turns out that $X(\tau_{\mathbf{X},T}) = X(\tau_{\mathbf{X},[T,2T]})$ almost surely. Suppose this is not true. Then without loss of generality assume $P(X(\tau_{\mathbf{X},T}) < X(\tau_{\mathbf{X},[T,2T]})) > 0$. It then implies that $\tau_{\mathbf{X},[T,2T]} \in A \cup B$, thus contradicting with $P(A \cup B) = 0$. Similar reasoning further guarantees that

$$P(X(\tau_{\mathbf{X},T}) = X(\tau_{\mathbf{X},[T,2T]}), \tau_{\mathbf{X},[T,2T]} - \tau_{\mathbf{X},T} > T) = 0.$$

On the other hand, assumption U_T requires

$$P(X(\tau_{\mathbf{X},T}) = X(\tau_{\mathbf{X},[T,2T]}), \tau_{\mathbf{X},[T,2T]} - \tau_{\mathbf{X},T} < T) = 0.$$

Combining these conditions, we are left with only one possibility:

$$P(X(\tau_{\mathbf{X},T}) = X(\tau_{\mathbf{X},[T,2T]}), \tau_{\mathbf{X},[T,2T]} - \tau_{\mathbf{X},T} = T) = 1,$$

which, clearly, leads to the uniform distribution of $\tau_{\mathbf{X},T}$ over $[0, T]$.

□

Indeed, proposition 3.3.8 does not only gives out the trichotomy, but also tells us under which condition each scenario will happen. Briefly, point mass exists on at least one boundary if and only if $A \cup A'$ is nonempty with positive probability; the density explodes near to the boundary if and only if $B \cup B'$ is nonempty with positive probability. Even more precisely, which ones between A and A', B and B' are nonempty determine to which boundaries do the corresponding behaviors appear. Finally, uniform distribution can be predicted by checking whether the global maxima of the process occur with an exact period.

3.4 Universal upper bounds on the density

The upper bounds in part (b) of Theorem 3.3.1 turn out to be the best possible pointwise, as is shown in the following result.

Proposition 3.4.1. *For each $0 < t < T$ and any number smaller than the upper bound given in* (3.1)*, there is a sample continuous stationary process satisfying Assumption U_T and Assumption L for which the right continuous version of the density $f_{\mathbf{X},T}(t)$ of the supremum location at time t exceeds that number.*

Proof. By symmetry, it is enough to show that for any $0 < t < T$ and any number smaller than $1/t$ there is a stationary process of the required type for which $f_{\mathbf{X},T}(t)$ exceeds that number.

To this end, let $\tau > t$ and let $k \geq 1$ be an integer. We define a periodic function $(x(s),\ s \in \mathbb{R})$ with period $k\tau + 2T$ by defining its values on the interval $[0, k\tau + 2T]$. We set $x(i\tau) = k - i$ for $i = 0, 1, \ldots, k$ and $x(k\tau + 2T) = k$. We set, further, for $i = 0, 1, \ldots, k-1$, $x((i + 1/2)\tau) = -R$ and also $x\big(k\tau + T\big) = -R$ for a large positive R we describe in a moment. We complete the definition of the function by connecting linearly the values in neighboring points where the function has already been defined. Fix $t < r < \tau$, and choose now R so large that the condition

$$x\big(i\tau\big) > x\big(i\tau - r\big) \tag{3.33}$$

holds for all $i = 1, \ldots, k$. Now define a stationary process by $X(s) = x(s - U)$, $s \in \mathbb{R}$, where U is uniformly distributed between 0 and $k\tau + 2T$. By construction, the process is sample continuous and satisfies Assumption U_T and Assumption L.

If, for $i = 1, \dots, k$, we have $i\tau - r < U < i\tau$, then the local maximum at $s = i\tau$ of the function $\mathbf{x}$ becomes the global maximum of the process $\mathbf{X}$ over the interval $[0, T]$, and is located in the interval $(0, r)$. This contributes $1/(k\tau + 2T)$ to the value of the density $f_{\mathbf{X},T}$ at each point of the interval $(0, r)$. In particular, since $t \in (0, r)$,

$$f_{\mathbf{X},T}(t) \geq \frac{k}{k\tau + 2T} .$$

Since we can take k arbitrarily large, the value of the density can be arbitrarily close to $1/\tau$ and, since τ can be taken arbitrarily close to t, the value of the density can be arbitrarily close to $1/t$. □

Suppose now that the stationary process $\mathbf{X}$ is time reversible, i.e. if $(X(-t),\, t \in \mathbb{R}) \overset{d}{=} (X(t),\, t \in \mathbb{R})$. That would, obviously, be the case for stationary Gaussian processes. If the process satisfies also Assumption U_T, then the distribution of the unique supremum location $\tau_{\mathbf{X},T}$ is symmetric in the interval $[0, T]$, meaning that $\tau_{\mathbf{X},T} \overset{d}{=} T - \tau_{\mathbf{X},T}$. Therefore, the density $f_{\mathbf{X},T}$ satisfies

$$f_{\mathbf{X},T}(t) = f_{\mathbf{X},T}(T - t) \tag{3.34}$$

for all $0 < t < T/2$ that are continuity points of $f_{\mathbf{X},T}$. Even though the upper bound given in part (b) of Theorem 3.3.1 is symmetric around the middle of the interval $[0, T]$, it turns out that the bounded variation property in part (d) of Theorem 3.3.1 provides a better bound in this symmetric case. This bound and its optimality, even within the class of stationary Gaussian processes, is presented in the following result.

Proposition 3.4.2. *Let* $\mathbf{X} = (X(t),\, t \in \mathbb{R})$ *be a time reversible stationary sample upper semi-continuous process satisfying Assumption* U_T*. Then the density* $f_{\mathbf{X},T}$ *of the* *unique location of the supremum in the interval* $[0,T]$ *satisfies*

$$f_{\mathbf{X},T}(t) \le \begin{cases} \frac{1}{2t} & \text{if } 0 < t \le \frac{T}{3} \\ \frac{1}{T-t} & \text{if } \frac{T}{3} < t \le \frac{T}{2} \\ \frac{1}{t} & \text{if } \frac{T}{2} < t \le \frac{2T}{3} \\ \frac{1}{2(T-t)} & \text{if } \frac{2T}{3} < t < T \end{cases} . \tag{3.35}$$

Furthermore, for each $0 < t < T$ *and any number smaller than the upper bound given in* (3.35)*, there is a sample continuous Gaussian process satisfying Assumption* U_T *and Assumption L for which the density* $f_{\mathbf{X},T}(t)$ *exceeds that number.*

Proof. Since the density $f_{\mathbf{X},T}$ is right continuous, it is enough to consider only continuity points of the density and, by (3.34), it is enough to consider $0 < t < T/2$. Then $T-t$ is also a continuity point of the density. Denote $a = \inf_{0<s\le t} f_{\mathbf{X},T}(s)$, $b = \inf_{t<s<T/2} f_{\mathbf{X},T}(s)$. Note that, given $\varepsilon > 0$, there is a continuity point of the density $u \in (0,t]$ such that $f_{\mathbf{X},T}(u) \le a+\varepsilon$, and there is a continuity point of the density $v \in [t, T/2]$ such that $f_{\mathbf{X},T}(v) \le b+\varepsilon$. Observe also that

$$at + b(T/2 - t) \le \int_0^{T/2} f_{\mathbf{X},T}(s)\, ds \le \frac{1}{2} . \tag{3.36}$$

Furthermore, applying the total variation bound (3.3) to the interval $[u, T-u]$ gives us

$$\begin{aligned} 2(a+\varepsilon) &\ge f_{\mathbf{X},T}(u) + f_{\mathbf{X},T}(T-u) \\ &\ge \big|f_{\mathbf{X},T}(t) - f_{\mathbf{X},T}(u)\big| + \big|f_{\mathbf{X},T}(v) - f_{\mathbf{X},T}(t)\big| \\ &+ \big|f_{\mathbf{X},T}(T-v) - f_{\mathbf{X},T}(v)\big| + \big|f_{\mathbf{X},T}(T-t) - f_{\mathbf{X},T}(T-v)\big| \\ &+ \big|f_{\mathbf{X},T}(T-u) - f_{\mathbf{X},T}(T-t)\big| \\ &\ge 2\big(f_{\mathbf{X},T}(t) - a - \varepsilon\big)_+ + 2\big(f_{\mathbf{X},T}(t) - b - \varepsilon\big)_+ . \end{aligned}$$

Letting $\varepsilon \to 0$ and recalling that $a \le f_{\mathbf{X},T}(t)$ and $b \le f_{\mathbf{X},T}(t)$, we obtain

$$f_{\mathbf{X},T}(t) \le a + b/2\,. \tag{3.37}$$

Since $b \le f_{\mathbf{X},T}(t)$, this implies that

$$b \le 2a\,. \tag{3.38}$$

If $0 < t \le T/3$, then the largest value of the right hand side of (3.37) under the constraint (3.36) requires taking a as large as possible and b as small as possible. Taking $a = 1/2t$ and $b = 0$ in (3.37) results in the upper bound given in (3.35) in this range. If $T/3 < t \le T/2$, then the largest value of the right hand side of (3.37) under the constraint (3.36) requires taking a as small as possible and b as large as possible. By (3.38), we have to take $a = 1/2(T-t)$, $b = 1/(T-t)$ in (3.37), which results in the upper bound given in (3.35) in this case.

It remains to prove the optimality part of the statement of the corollary. By symmetry it is enough to consider $0 < t \le T/2$. Fix such t. Let $\varepsilon > 0$ be a small number and $h > 0$ be a large number, rationally independent of $t+\varepsilon$. Consider a stationary Gaussian process given by

$$X(s) = G_1 \cos\left(\frac{2\pi}{t+\varepsilon}s\right) + G_2 \sin\left(\frac{2\pi}{t+\varepsilon}s\right)$$
$$+G_3 \cos\left(\frac{2\pi}{h}s\right) + G_4 \sin\left(\frac{2\pi}{h}s\right),\ s \in \mathbb{R}\,,$$

where $G_1, \ldots, G_4$ are i.i.d. standard normal random variables. The process is, clearly, sample continuous, and it satisfies Assumption L. Furthermore, rational independence of $t+\varepsilon$ and h implies that, on a set of probability 1, the process $\mathbf{X}$ has different values at all of its local maxima, hence Assumption U_T is satisfied for any $T > 0$. Note that we can write

$$X(s) = A_1 \cos\left(\frac{2\pi}{t+\varepsilon}s + U_1\right) + A_2 \cos\left(\frac{2\pi}{h}s + U_2\right) := X_1(s) + X_2(s),\ s \in \mathbb{R}\,,$$

where A_1 and A_2 have the density $xe^{-x^2/2}$ on $(0, \infty)$, and U_1 and U_2 are uniformly distributed between 0 and 2π, with all 4 random variables being independent. Clearly, the leftmost location of the supremum of the process $\mathbf{X}_1$ is at

$$\tau_1 = (t + \varepsilon)\frac{2\pi - U_1}{2\pi},$$

which is uniformly distributed between 0 and $t + \varepsilon$. On the event $E = \{0 < U_2 < \pi - 2\pi T/h\}$ the process $\mathbf{X}_2$ is decreasing on $[0, T]$, so the value of the sum $\mathbf{X}$ at the leftmost supremum of $\mathbf{X}_1$ exceeds the value of the sum at all the other locations of the supremum of $\mathbf{X}_1$ in the interval $[0, T]$. If the supremum of the sum remained at τ_1, the density of that unique supremum would be at least $P(E)/(t + \varepsilon)$ at each point of the interval $(0, t + \varepsilon)$. Since $P(E) \to 1/2$ as $h \to \infty$, the value of the density at t would exceed any value smaller than $1/2t$ after taking h large and ε small. The location of the supremum of the sum does not remain at τ_1 but, instead, moves to $\tau_2 = \tau_2(A_1, A_2, U_1, U_2)$ defined by

$$\tau_2 = \sup\left\{s \le \tau_1 : \frac{A_1}{t + \varepsilon}\sin\left(\frac{2\pi}{t + \varepsilon}s + U_1\right) + \frac{A_2}{h}\sin\left(\frac{2\pi}{h}s + U_2\right) = 0\right\}.$$

For large h, τ_2 is nearly identical to τ_1, and straightforward but somewhat tedious calculus based on the implicit function theorem shows that the above statement remains true for τ_2: the contribution of the event E to the density of the unique supremum of the process $\mathbf{X}$ would exceed any value smaller than $1/2t$ at any point of the interval $(0, t + \varepsilon)$ after taking h large and ε small. We omit the details.

We have shown the optimality of the upper bound given in (3.35) in the case $0 < t \le T/3$. It remains to consider the case $T/3 < t \le T/2$. We will use again a two-wave stationary Gaussian process, but with a slightly different twist. Let $\varepsilon > 0$ be a small number, $h > 0$ a large number and $r > 0$ a fixed number that is rationally independent of $T - t + \varepsilon$. Consider a stationary Gaussian process given by

$$X(s) = A_1 \cos\left(\frac{2\pi}{T-t+\varepsilon}s + U_1\right) + \frac{1}{h}A_2 \cos\left(\frac{2\pi}{r}s + U_2\right)$$

$$:= X_1(s) + X_2(s),\ s \in \mathbb{R},$$

where A_1, A_2, U_1 and U_2 are as above. As above, $\mathbf{X}$ is a sample continuous Gaussian process satisfying Assumption L and Assumption U_T. Now the leftmost location of the supremum of the process $\mathbf{X}_1$ is at

$$\tau_1 = (T-t+\varepsilon)\frac{2\pi - U_1}{2\pi},$$

which is uniformly distributed between 0 and $T-t+\varepsilon$. Further, if $\tau_1 > t - \varepsilon/2$, then τ_1 is the unique supremum of $\mathbf{X}_1$ in the interval $[0, T]$. If the supremum of the sum $\mathbf{X}$ remained at τ_1, then the density of the supremum location at the point t would be at least $1/(T-t+\varepsilon)$, which would then exceed any value smaller than $1/(T-t)$ after taking ε small. The location of the supremum of $\mathbf{X}$ does not remain at τ_1, but instead moves to the unique for large h point $\tau_2 = \tau_2(A_1, A_2, U_1, U_2)$ in $[0, T]$ satisfying

$$\frac{A_1}{T-t+\varepsilon}\sin\left(\frac{2\pi}{T-t+\varepsilon}\tau_2 + U_1\right) + \frac{A_2}{hr}\sin\left(\frac{2\pi}{r}\tau_2 + U_2\right) = 0.$$

For large h, τ_2 is nearly identical to τ_1 and, as above, using the implicit value theorem allows us to conclude that, for any value smaller than $1/(T-t)$, the value of the density of τ_2 in the interval $(t - \varepsilon/2, T-t+\varepsilon)$ exceeds that value after taking ε small and h large. This proves the optimality of the upper bound given in (3.35) in all cases. $\square$

CHAPTER 4

DISTRIBUTION OF THE LOCATION OF PATH SUPREMUM: CHARACTERIZATION AND ASYMPTOTIC BEHAVIOR

4.1 Introduction

Let $\mathbf{X} = (X(t),\, t \in \mathbb{R})$ be a sample continuous stationary process. Even if, on an event of probability 1, the supremum of the process over a compact interval $[0, T]$ is attained at a unique point, this point does not have to be uniformly distributed over that interval, as is known since [15]. However, its distribution still has to be absolutely continuous in the interior of the interval, and the density has to satisfy very specific general constraints, as was shown in Chapter 3.

In this chapter we give a complete description of the family of possible densities of the supremum location for a large class of sample continuous stationary processes. The necessary conditions on these densities follow by combining certain general results cited above, and for every function satisfying these necessary conditions we construct a stationary process of the required type for which this function is the density of the supremum location. This is done in Section 4.3, which is preceded by Section 4.2 in which we describe the class of stationary processes we are considering and quote the results from the previous chapter we need in the present chapter. Next, we show that for a large class of stationary processes, under a certain strong mixing assumption, the distribution of the supremum location does converge to the uniformity for very long intervals, and it does it in a strong sense. This is shown in Section 4.4.

4.2 Preliminaries

For most of this chapter $\mathbf{X} = (X(t),\, t \in \mathbb{R})$ is a stationary process with continuous sample paths, defined on a probability space $\big(\Omega, \mathcal{F}, P\big)$, but in Section 4.4 we will allow upper semi-continuous sample paths. In most of the chapter (but not in Section 4.4) we will also impose two assumptions on the process, which we now state.

For $T > 0$ we denote by $X_*(T) = \sup_{0 \le t \le T} X(t)$, the largest value of the process in the interval $[0, T]$.

Assumption U_T:

$$P\Big(X(t_i) = X_*(T),\ i = 1, 2, \text{ for two different } t_1, t_2 \in [0, T]\Big) = 0.$$

Many processes satisfy Assumption U_T. In particular, a beautiful proof in [13] shows that any continuous Gaussian process, such that $X(s) \neq X(t)$ a.s. for any two points $s \neq t$, satisfies this assumption.

The second assumption on a stationary process deals with the fluctuations of its sample paths.

Assumption L:

$$K := \lim_{\varepsilon \downarrow 0} \frac{P\big(\mathbf{X} \text{ has a local maximum in } (0, \varepsilon)\big)}{\varepsilon} < \infty\,,$$

with the limit easily shown to exist. Under Assumption L the process $\mathbf{X}$ has sample paths of locally bounded variation; see Lemma 3.2.2.

For a compact interval $[a, b]$, we will denote by

$$\tau_{\mathbf{X},[a,b]} = \inf\big\{t \in [a, b] :\ X(t) = \max_{a \le s \le b} X(s)\big\}$$

the leftmost location of the supremum in the interval; it is a well defined random variable. If the supremum is unique, the adjective “leftmost” is, clearly, redundant. For $a = 0$, we will abbreviate $\tau_{\mathbf{X},[0,b]}$ to $\tau_{\mathbf{X},b}$, and use the same abbreviation in similar situations in the sequel.

We denote by $F_{\mathbf{X},[a,b]}$ the law of $\tau_{\mathbf{X},[a,b]}$; it is a probability measure on the interval $[a, b]$. It was proved in Chapter 3 that for any $T > 0$ the probability measure $F_{\mathbf{X},T}$ is absolutely continuous in the interior of the interval $[0, T]$, and density can be chosen to be right continuous and have left limits; we call this version of the density $f_{\mathbf{X},[a,b]}$. This version of the density satisfies a universal upper bound

$$f_{\mathbf{X},T}(t) \leq \max\left(\frac{1}{t}, \frac{1}{T-t}\right), \ 0 < t < T\,. \tag{4.1}$$

We will also use the following result from the previous chapter.

Lemma 4.2.1 (Lemma 3.3.1). *Let $0 \leq \Delta < T$. Then for every $0 \leq \delta \leq \Delta$, $f_{\mathbf{X},T-\Delta}(t) \geq f_{\mathbf{X},T}(t+\delta)$ almost everywhere in $(0, T-\Delta)$. Furthermore, for every such δ and every $\varepsilon_1, \varepsilon_2 \geq 0$, such that $\varepsilon_1 + \varepsilon_2 < T - \Delta$,*

$$\begin{aligned} &\int_{\varepsilon_1}^{T-\Delta-\varepsilon_2} \big(f_{\mathbf{X},T-\Delta}(t) - f_{\mathbf{X},T}(t+\delta)\big)\, dt \qquad (4.2) \\ &\leq \int_{\varepsilon_1}^{\varepsilon_1+\delta} f_{\mathbf{X},T}(t)\, dt + \int_{T-\Delta-\varepsilon_2+\delta}^{T-\varepsilon_2} f_{\mathbf{X},T}(t)\, dt\,. \end{aligned}$$

4.3 Processes satisfying Assumption L

In this section we prove our main theorem of this chapter, giving a full description of possible càdlàg densities $f_{\mathbf{X},T}$ for continuous stationary processes satisfying Assumption U_T and Assumption L.

For a function f of a real argument whose domain contains an interval (t_1, t_2), its total variation over the interval is defined by

$$TV_{(t_1,t_2)}(f) := \sup \sum_{i=1}^{n-1} \big| f(s_{i+1}) - f(s_i) \big| \,,$$

where the supremum is taken over all choices of $t_1 < s_1 < \ldots < s_n < t_2$.

Theorem 4.3.1. *Let $\mathbf{X} = (X(t),\, t \in \mathbb{R})$ be a stationary sample continuous process, satisfying Assumption U_T and Assumption L. Then the restriction of the law $F_{\mathbf{X},T}$ of the unique location of the supremum of the process in $[0,T]$ to the interior $(0,T)$ of the interval is absolutely continuous. The density $f_{\mathbf{X},T}$ has a càdlàg version with the following properties:*

(a) The density has a bounded variation on $(0,T)$, hence the limits

$$f_{\mathbf{X},T}(0+) = \lim_{t \to 0} f_{\mathbf{X},T}(t) \text{ and } f_{\mathbf{X},T}(T-) = \lim_{t \to T} f_{\mathbf{X},T}(t)$$

exist and are finite. Furthermore,

$$TV_{(0,T)}(f_{\mathbf{X},T}) \le f_{\mathbf{X},T}(0+) + f_{\mathbf{X},T}(T-) \,. \tag{4.3}$$

(b) The density is bounded away from zero. That is,

$$\inf_{0<t<T} f_{\mathbf{X},T}(t) > 0 \,. \tag{4.4}$$

(c) Either $f_{\mathbf{X},T}(t) = 1/T$ for all $0 < t < T$, or $\int_0^T f_{\mathbf{X},T}(t)\, dt < 1$.

Moreover, if f is a nonnegative càdlàg function satisfying (a)-(c) above, then there is a stationary sample continuous process $\mathbf{X}$, satisfying Assumption U_T and Assumption L, such that f is the density in the interior $(0,T)$ of the unique location of the supremum of the process in $[0,T]$.

Proof. The existence of a càdlàg density with properties (a)-(c) in the statement of the theorem is an immediate consequence of the statements of Theorems 3.3.1 and 3.3.3 in Chapter 3. We proceed to show the converse part of the theorem. If $f_{\mathbf{X},T}(t) = 1/T$ for all $0 < t < T$, then a required example is provided by a single wave periodic stationary Gaussian process with period T, so we need only to consider the second possibility in property (c). We start with the case where the candidate density f is a piecewise constant function of a special form.

We call a finite collection $(u_i, v_i),\ i = 1, \ldots, m$ of nonempty open subintervals of $(0, T)$ *a proper collection of blocks* if for any $i, j = 1, \ldots, m$ there are only 3 possibilities: either $(u_i, v_i) \subseteq (u_j, v_j)$, or $(u_j, v_j) \subseteq (u_i, v_i)$, or $[u_i, v_i] \cap [u_j, v_j] = \emptyset$. If $u_i = 0,\ v_i = T$, we call (u_i, v_i) a base block. If $u_i = 0,\ v_i < T$, we call (u_i, v_i) a left block. If $u_i > 0,\ v_i = T$, we call (u_i, v_i) a right block. If $u_i > 0,\ v_i < T$, we call (u_i, v_i) a central block. We start with constructing a stationary process as required in the theorem when the candidate density f satisfies requirements (a)-(c) of the theorem and has the form

$$f(t) = \frac{1}{HT} \sum_{i=1}^{m} \mathbf{1}_{[u_i, v_i)}(t),\ 0 < t < T \tag{4.5}$$

for some proper collection of blocks, with the obvious convention at the endpoints 0 and T, for some $H > 1$. Observe that for functions of the type (4.5), requirement (b) of the theorem is equivalent to requiring that there is at least one base block, and requirement (a) is equivalent to requiring that the number of the central blocks does not exceed the number of the base blocks. Finally, (the second case of) property (c) is equivalent to requiring that

$$d = \frac{1}{m} \left(HT - \sum_{i=1}^{m} (v_i - u_i) \right) > 0\,. \tag{4.6}$$

We will construct a stationary process by a uniform shift of a periodic deterministic function over its period. Now, however, the period will be equal to

$HT > T$. We start, therefore, by defining a deterministic continuous function $(x(t),\, 0 \le t \le HT)$ with $x(0) = x(HT)$, which we then extend by periodicity to the entire $\mathbb{R}$. Let $B \ge 1$ be the number of the base blocks in the collection. We partition the entire collection of blocks into B subcollection which we call *components* by assigning each base block to one component, assigning to each component at most one central block, and assigning the left and right blocks to components in an arbitrary way. For $j = 1, \ldots, B$ we denote by

$$L_j = d\big(\text{the number of blocks in the } j\text{th component}\big) \tag{4.7}$$

$$+ \text{ the total length of the blocks in the } j\text{th component}\,.$$

We set $x(0) = 2$. Using the blocks of the first component we will define the function x on the interval $(0, L_1]$ in such a way that $x(L_1) = 2$. Next, using the blocks of the second component we will define the function x on the interval $(L_1, L_1 + L_2]$ in such a way that $x(L_1 + L_2) = 2$, etc. Since

$$\sum_{j=1}^{B} L_j = dm + \sum_{i=1}^{m} (v_i - u_i) = HT\,,$$

this construction will terminate with a function x constructed on the entire interval $[0, HT]$ with $x(HT) = 2 = x(0)$, as desired.

We proceed, therefore, with defining the function x on an interval of length L_j using the blocks of the jth component. For notational simplicity we will take $j = 1$ and define x on the interval $[0, L_1]$ using the blocks of the first component. The construction is slightly different depending on whether or not the component has a central block, whether or not it has any left blocks, and whether or not it has any right blocks. If the component has $l \ge 1$ left blocks, we will denote them by $(0, v_j)$, $j = 1, \ldots, l$. If the component has $r \ge 1$ right blocks, we will denote them by (u_j, T), $j = 1, \ldots, r$. If the component has a central block, we will denote it by (u, v). We will construct the function x by defining it first on a finite number of special points and then filling in the gaps in a piecewise linear manner.

Suppose first that the component has a central block, some left blocks and some right blocks. In this case we proceed as follows.

Step 1 Recall that $x(0) = 2$ and set

$$x\left(jd + \sum_{i=1}^{j-1} v_i\right) = x\left(jd + \sum_{i=1}^{j} v_i\right) = 2 - 2^{j-l},\ j = 1, \dots, l\,.$$

Note that the last point obtained in this step is $x\big(ld + \sum_{i=1}^{l} v_i\big) = 1$.

Step 2 Set

$$x\left((l+1)d + \sum_{i=1}^{l} v_i\right) = x\left((l+1)d + \sum_{i=1}^{l} v_i + v\right)$$
$$= x\left((l+1)d + \sum_{i=1}^{l} v_i + v + T - u\right) = \frac{1}{2}\,.$$

Step 3 Set

$$x\left((l+j+1)d + \sum_{i=1}^{l} v_i + v + T - u + \sum_{i=1}^{j-1}(T - u_j)\right)$$
$$= x\left((l+j+1)d + \sum_{i=1}^{l} v_i + v + T - u + \sum_{i=1}^{j}(T - u_j)\right)$$
$$= 2 - 2^{-(j-1)},\ j = 1, \dots, r\,.$$

Note that the last point obtained in this step is

$$x\left((l+r+1)d + \sum_{i=1}^{l} v_i + v + T - u + \sum_{i=1}^{r}(T - u_j)\right) = 2 - 2^{-(r-1)}\,.$$

Step 4 We add just one more point at distance d from the last point of the previous step by setting

$$x\left((l+r+2)d+\sum_{i=1}^{l} v_i+v+T-u+\sum_{i=1}^{r}(T-u_j)\right)=2\,.$$

Note that this point coincides with L_1 as defined in (4.7).

If the component has no left blocks, then Step 1 above is skipped, and Step 2 becomes the initial step with

$$x(d)=x(d+v)=x(d+v+T-u)=\frac{1}{2}\,.$$

If the component has no right blocks, then Step 3 above is skipped, and at Step 4 we add the distance d to the final point of Step 2, that is we set

$$x\left((l+2)d+\sum_{i=1}^{l} v_i+v+T-u\right)=2\,.$$

If the component has no central block, then Step 2 is skipped, but we do add the distance T to the last point of Step 1. That is, the first point obtained at Step 3 becomes

$$x\left((l+1)d+\sum_{i=1}^{l} v_i+T\right)=1\,,$$

if there are any left blocks, with the obvious change if $l=0$. Finally, if there is neither central block, nor any right blocks, then both Step 2 and Step 3 are skipped, and Step 4 just adds $d+T$ to the last point of Step 1, i.e. it becomes

$$x\left((l+1)d+\sum_{i=1}^{l} v_i+T\right)=2\,,$$

once again with the obvious change if $l=0$. It is easy to check that in any case Step 4 sets $x(L_1)=2$, with L_1 as defined in (4.7). In particular, $L_1>T$.

Finally, we specify the piecewise linear rule by which we complete the construction of the function x on the interval $[0, L_1]$. The function has been defined on a finite set of points and we proceed from left to right, starting with $x(0) = 2$, to fill the gap between one point in the finite set and the adjacent point from the right, until we reach $x(L_1) = 2$. By the construction, there are pairs of adjacent points in which the values of x coincide, and pairs of adjacent points in which the values of x are different. In most cases only adjacent points at the distance d have equal values of x, but if, e.g. a central block is missing, then at a pair of adjacent points at a distance T, or $d + T$, the values of x coincide as well.

In any case, if the values of x at two adjacent points are different, we define the values of x between these two points by linear interpolation. If the values of x at two adjacent points, say, a and b with $a < b$, are equal to, say, y we define the function x between these two points by

$$x(t) = \max\bigl(y - (t - a)/d,\, y - (b - t)/d\bigr)$$

provided the value at the midpoint, $y - (b - a)/2d \geq -1$. If this lower bound fails, we define the values of x between the points $a + dy$ and $b - dy$ by

$$x(t) = \max\bigl(-\tau(t - (a + dy)),\, -\tau((b - dy) - t)\bigr)\,,$$

for an arbitrary $\tau > 0$ such that both $\tau \leq 1/d$ and the value at the midpoint, $-\tau\bigl((b - a)/2 - dy\bigr) \geq -1$. The reason for this slightly cumbersome definition is the need to ensure that x is nowhere constant, while keeping the lower bound of x and its Lipschitz constant under control. We note, at this point, that, since in all cases $b - a \leq T + d$, we can choose, for a fixed T, the value of τ so that $\tau \geq \tau_d > 0$, where the constant τ_d stays bounded away from zero for d in a compact interval.

Now that we have defined a periodic function $(x(t),\ t \in \mathbb{R})$ with period HT, we define a stationary process $\mathbf{X}$ by $X(t) = x(t-U)$, $t \in \mathbb{R}$, where U is uniformly distributed between 0 and HT. The process is, clearly, sample continuous and satisfies Assumption L. We observe, further, that, if the supremum in the interval $[0, T]$ is achieved in the interior of the interval, then it is achieved at a local maximum of the function x. If the value at the local maximum is equal to 2, then it is due to an endpoint of a component, and, since the contribution of any component has length exceeding T, this supremum is unique. If the value at the local maximum is smaller than 2, then that local maximum is separated from the nearest local maximum with the same value of x by at least the distance induced by Step 2, which is T. Consequently, in this case the supremum over $[0, T]$ is unique as well. Similarly, if the supremum is achieved at one of the endpoints of the interval, it has to be unique as well, on a set of probability 1. Therefore, the process $\mathbf{X}$ satisfies Assumption U$_T$.

Example We interrupt the exposition for a moment to demonstrate a simple special case of the construction of the process $\mathbf{X}$ to help the reader to visualize the procedure. Consider a candidate density function

$$f(t) = \begin{cases} \frac{2}{HT} & \text{if } t \in (0, v_1) \cup [u, v), \\ \frac{1}{HT} & \text{if } t \in [v_1, u) \cup [v, T) \end{cases}$$

for $0 < v_1 < u < v < T$, with $H > 1 + \frac{v_1 + v - u}{T}$. This corresponds to a proper collection of three blocks: a base block $(0, T)$, a central block (u, v), and a left block $(0, v_1)$. Hence the total number of blocks $m = 3$, and $d = \frac{1}{3}(HT - T - v_1 - (v - u)) > 0$. Since there is only one base block, we use one component, of the length $L_1 = HT$.

The construction of the deterministic function $x(t)$ on $[0, HT]$ is as follows.

The starting point is $x(0) = 2$. Then step 1, dealing with the left block $(0, v_1)$, assigns value $2 - 2^{1-1} = 1$ to points d and $d + v_1$. Step 2 continues to set

$$x(2d + v_1) = x(2d + v_1 + v) = x(2d + v_1 + v + T - u) = \frac{1}{2}.$$

Step 3 is skipped since there is no right block. Finally, the end point of this component, $x(3d+v_1+v+T-u) = 2$ is added in step 4. Since $3d+v_1+v+T-u = HT$, this is the end of the cycle.

To demonstrate the the piecewise linear interpolation rule between these special points, we choose specific values $T = 6, H = 2, v_1 = 1, u = 3$ and $v = 5$. This implies $d = 1$. Firstly, between the pairs of points with the t coordinates 0 and $d = 1$, $d + v_1 = 2$ and $2d + v_1 = 3$, $2d + v_1 + v + T - u = 11$ and $HT = 12$ we use linear interpolation. Consider the segment between the points $d = 1$ and $d + v_1 = 2$, at which x has the common value $y = 1$. The general rule checks the value of the interpolation at the midpoint of the segment, which is $1 - \frac{v_1}{2d} = 1/2$. It is greater than -1, so no modification is necessary. Same procedure applies to the segments between the points $2d + v_1 = 3$ and $2d + v_1 + v = 8$, and the points $2d + v_1 + v = 8$ and $2d + v_1 + v + T - u = 11$. Only on the interval $(3, 8)$ the interpolation procedure has to be modified. We set $\tau = 1/2$ (so that the value of the lowest point is exactly -1) and obtain

$$x(t) = \begin{cases} 3.5 - t & \text{if } 3 \leq t \leq 3.5 \\ (3.5 - t)/2 & \text{if } 3.5 \leq t \leq 5.5 \\ (7.5 - t)/2 & \text{if } 5.5 \leq t \leq 7.5 \\ 7.5 - t & \text{if } 7.5 \leq t \leq 8 \end{cases}.$$

The figure below shows the density f and the function $x(t)$ within one cycle.

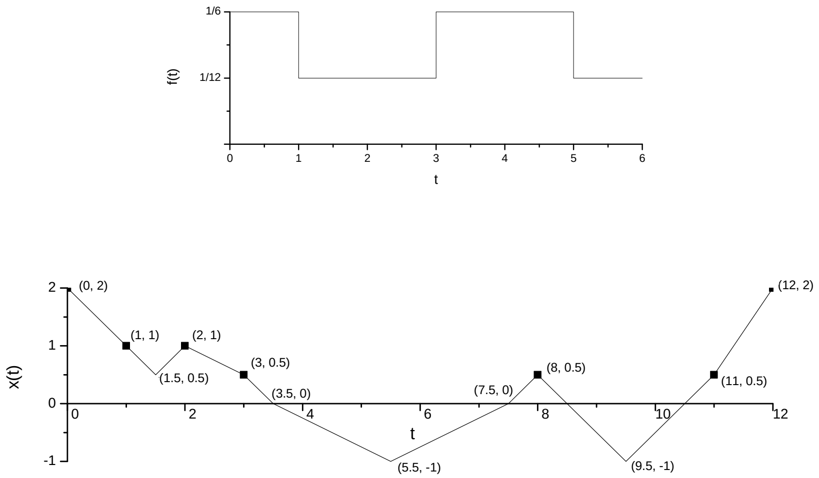


Fig 1. The functions f and x in the special case.

Finally, we extend the function $x(t)$ periodically with period HT to the whole real line. The process $\mathbf{X}$ is then defined by $\mathbf{X}(t) = x(t-U)$, where U is uniformly distributed between 0 and HT.

We now return to the general case considered in the theorem. We first show that for the process $\mathbf{X}$ constructed above, the density $f_{\mathbf{X},T}$ coincides with the function f given in (4.5), with which the construction was performed. According to the above analysis, we need to account for the contribution of each local maximum of the function x over its period to the density $f_{\mathbf{X},T}$. The local maxima may appear in Step 1 of the construction, and then they are due to left blocks. They may apear in Step 3 of the construction, and then they are due to right blocks. They may appear Step 2 of the construction, and then they are due to central blocks. Finally, the points where x has value 2 are always local maxima. We will see that they are due to base blocks. We start with the latter local maxima. Clearly, each such local maximum is, by periodicity, equal to one of the B values, $\sum_{j=1}^{i} L_j - HT$, $i = 1, \ldots, B$. The ith of these points becomes the global maximum of $\mathbf{X}$ over $[0, T]$ if and only if

$$U \in \left(HT - \sum_{j=1}^{i} L_j,\ (H+1)T - \sum_{j=1}^{i} L_j \right) ,$$

and the global maximum is then located at the point $\sum_{j=1}^{i} L_j - HT + U$. Therefore, the contribution of each such local maximum to the density is $1/HT$ at each $0 < t < T$, and overall the points where x has value 2 contribute to $f_{\mathbf{X},T}$

$$f_{\text{base}}(t) = \frac{B}{HT},\ 0 < t < T\,. \tag{4.8}$$

Next, we consider the contribution to $f_{\mathbf{X},T}$ of the local maxima due to left blocks. For simplicity of notation we consider only the left blocks in the first component. Then the local maximum due to the jth left block is at the point $jd + \sum_{i=1}^{j} v_i$. As before, we need to check over what interval of the values of U this local maximum becomes the global maximum of $\mathbf{X}$ over $[0, T]$. The relevant values of U must be such that the time interval $\left(jd + \sum_{i=1}^{j-1} v_i,\ jd + \sum_{i=1}^{j} v_i\right)$ is shifted to cover the origin, and this corresponds to an interval of length v_j of the values of U. The shifted local maximum itself will then be located within the interval $(0, v_j)$, which contributes $1/HT$ at each $0 < t < v_j$. Overall, the local maxima due to left blocks contribute to $f_{\mathbf{X},T}$

$$f_{\text{left}}(t) = \frac{1}{HT} \sum_{\text{left blocks}} \mathbf{1}_{(0,v_i)}(t),\ 0 < t < T\,. \tag{4.9}$$

Similarly, the local maxima due to right blocks contribute to $f_{\mathbf{X},T}$

$$f_{\text{right}}(t) = \frac{1}{HT} \sum_{\text{right blocks}} \mathbf{1}_{(u_i,T)}(t),\ 0 < t < T\,. \tag{4.10}$$

Finally, we consider the central blocks. If the first component has a central block, then the local maximum due to the central block is at the point $(l+1)d + \sum_{i=1}^{l} v_i + v$. Any value of U that makes this local maximum the global maximum over $[0,T]$ must be such that the time interval $\big((l+1)d + \sum_{i=1}^l v_i,\ (l+1)d + \sum_{i=1}^l v_i + v\big)$ is shifted to cover the origin. Furthermore, that value of U must also be such that the time interval $\big((l+1)d + \sum_{i=1}^l v_i + v,\ (l+1)d + \sum_{i=1}^l v_i + v + T - u\big)$ is shifted to cover the right endpoint T. If we think of shifting the origin instead of shifting x, the origin will have to be located in the interval $\big((l+1)d+\sum_{i=1}^l v_i,\ (l+1)d+\sum_{i=1}^l v_i+v-u\big)$. This corresponds to a set of values of U of measure $v-u$, and the shifted local maximum will then be located within the interval (u,v), which contributes $1/HT$ at each $u<t<v$ to the density. Overall, the local maxima due to central blocks contribute to $f_{\mathbf{X},T}$

$$f_{\text{central}}(t) = \frac{1}{HT} \sum_{\text{central blocks}} \mathbf{1}_{(u,v)}(t),\ 0<t<T\,. \tag{4.11}$$

Since

$$f_{\mathbf{X},T}(t) = f_{\text{base}}(t) + f_{\text{left}}(t) + f_{\text{right}}(t) + f_{\text{central}}(t),\ 0<t<T\,,$$

we conclude by (4.8) - (4.11) that $f_{\mathbf{X},T}$ indeed coincides with the function f given in (4.5). Therefore, we have proved the converse part of the theorem in the case when the candidate density f is of the form (4.5).

We now prove the converse part of the theorem for a general f with properties (a)-(c) in the statement of the theorem. Recall that we need only to treat the second possibility in property (c). In order to construct a stationary process $\mathbf{X}$ for which $f_{\mathbf{X},T} = f$, we will approximate the candidate density f by functions of the form (4.5). Since we will need to deal with convergence of a sequence of continuous stationary processes we have just constructed in the case when the candidate density is of the form (4.5), we record, at this point, several properties of the stationary periodic process $X(t) = x(t-U)$, $t \in \mathbb{R}$ constructed above.

Property 1 *The process* $\mathbf{X}$ *is uniformly bounded:* $-1 \le X(t) \le 2$ *for all* $t \in \mathbb{R}$.

Property 2 *The process* $\mathbf{X}$ *is Lipschitz continuous, and its Lipschitz constant does not exceed* $3/2d$.

Property 3 *The process* $\mathbf{X}$ *is differentiable except at countably many points, at which* $\mathbf{X}$ *has left and right derivatives. On the set* $D_0 = \{t : X(t) > 0\}$ *the derivatives satisfy*

$$|X'(t)| \geq \frac{1}{2^N d}$$

(where the bound applies to both left and right derivatives if t *is not a differentiability point). Here* N *is the bigger of the largest number of left blocks any component has, and the largest number of the right blocks any component has. Similarly, on the set* $D_1 = \{t : X(t) \leq 0\}$ *the derivatives satisfy*

$$|X'(t)| \geq \tau_d\,,$$

where $\tau_d > 0$ *stays bounded away from zero for* d *in a compact interval.*

Property 4 *The distance between any two local maxima of* $\mathbf{X}$ *cannot be smaller than* d*. At its local maxima,* $\mathbf{X}$ *takes values in a finite set of at most* $N + 3$ *elements. Moreover, the absolute difference in the values of the process* $\mathbf{X}$ *in two local maxima in the interval* $(0, T)$ *is at least* 2^{-N}*, where* N *is as above.*

All these properties follow from the corresponding properties of the function x by considering the possible configuration of the blocks in a component.

We will now construct a sequence of approximations to a candidate density f as above. Let $n = 1, 2, \ldots$. It follows from the general properties of càdlàg functions (see e.g. [5]) that there is a finite partition $0 = t_0 < t_1 < \ldots < t_k = T$ of the interval $[0, T]$ such that

$$|f(s) - f(t)| \leq \frac{1}{nT} \text{ for all } t_i \leq s, t < t_{i+1},\, i = 0, \ldots, k-1. \tag{4.12}$$

We define a piecewise constant function $\tilde{f}_n$ on $(0,T)$ by setting, for each $i = 1,\dots,k$, the value of $\tilde{f}_n$ for $t_{i-1} \le t < t_i$ to be

$$\tilde{f}_n(t) = \frac{1}{knT} \max\left\{ j = 0,1,\dots : \, f(s) \ge \frac{j}{knT} \text{ for all } t_{i-1} \le s < t_i \right\} .$$

By definition and (4.12) we see that

$$f(t) - \frac{2}{nT} \le \tilde{f}_n(t) \le f(t), \; 0 < t < T \, . \tag{4.13}$$

Next, we notice that for every $i = 1,\dots,k-1$ there are points $s_i \in (t_{i-1}, t_i)$ and $s_{i+1} \in (t_i, t_{i+1})$ such that

$$\bigl| f(s_i) - f(s_{i+1}) \bigr| \ge \bigl| \tilde{f}_n(t_i-) - \tilde{f}_n(t_i) \bigr| - \frac{1}{knT} \, .$$

Therefore,

$$TV_{(0,T)}(\tilde{f}_n) \le TV_{(0,T)}(f) + \frac{1}{nT} \, . \tag{4.14}$$

We now define

$$f_n(t) = \tilde{f}_n(t) + \frac{1}{nT} \;\; 0 < t < T \, .$$

Clearly, the function f_n is càdlàg, has bounded variation on $(0,T)$ and is bounded away from zero. By (4.14), f_n also satisfies (4.3) since f does. Finally, since $\int_0^T f_{\mathbf{X},T}(t)\, dt \; < \; 1$, we see by (4.13) that, for all n large enough, $\int_0^T f_{\mathbf{X},T}(t)\, dt \; < \; 1$ as well. Therefore, for such n the function f_n has properties (a)-(c) in the statement of the theorem, and in the sequel we will only consider n large as above. We finally notice that f_n takes finitely many different values, all of which are in the set $\{j/knT, \; j = 1,2,\dots\}$. Therefore, f_n can be written in the form (4.5), with $H = kn$. Indeed, the blocks can be built by combining into a block all neighboring intervals where the value of f_n is the smallest, subtracting $1/knT$ from the value of f_n in the constructed block and iterating the procedure.

We have already proved that for any function of the type (4.5) there is a stationary process required in the statement of the theorem. Recall that a construction of this stationary process depends on assignment of blocks in a proper collection to components, and we would like to make sure that no component has "too many" left or right blocks. To achieve this, we need to distribute the left and right blocks as evenly as possible between the components. Two observations are useful here. First of all, it follows from the definition of f_n and (4.5) that

$$\frac{1}{k_n nT}(L_n + B_n) = f_n(0+) \leq f(0+) + \frac{1}{k_n nT} \leq f(0+) + 1$$

for n large enough (we are writing k_n instead of k to emphasize the dependence of k on n), where L_n and B_n are the numbers of the the left and base blocks in the nth collection. On the other hand, similar considerations tell us that

$$\frac{1}{k_n nT} B_n = \inf_{0<t<T} f_n(t) \geq \inf_{0<t<T} f(t) - \frac{2}{nT} \geq \frac{1}{2} \inf_{0<t<T} f(t)\,,$$

once again for n large enough, where we have used property (b) of f. Therefore, for such n,

$$\frac{L_n}{B_n} \leq 2\frac{f(0+) + 1}{\inf_{0<t<T} f(t)}\,, \tag{4.15}$$

and the right hand side is a finite quantity depending on f, but not on n. Performing a similar analysis for the right blocks, and recalling that we are distributing the left and right blocks as evenly as possible between the components, we see that there is a number $\Delta_f \in (0, \infty)$ such that for all n large enough, no component in the nth collection has more than Δ_f left blocks or Δ_f right blocks.

We will also need bounds on the important parameter $d = d_n$ appearing in the construction of a stationary process corresponding to functions of the type (4.5); these bounds do not depend on a particular way we assigns blocks to different components. Recall that

$$d_n = \frac{k_n nT}{m_n}\left(1 - \int_0^T f_n(t)\,dt\right)\,, \tag{4.16}$$

where $m_n = B_n + L_n + R_n + C_n$ (in the obvious notation) is the total number of blocks in the nth collection. Since

$$\frac{1}{k_n nT}\big(B_n + \max(L_n, R_n, C_n)\big) = \sup_{0<t<T} f_n(t), \;\; \frac{1}{k_n nT} B_n = \inf_{0<t<T} f_n(t)\,,$$

we see that

$$\inf_{0<t<T} f_n(t) \leq \frac{1}{k_n nT} m_n \leq 3 \sup_{0<t<T} f_n(t)\,. \tag{4.17}$$

We also know by the uniform convergence that $\int_0^T f_n \to \int_0^T f$. Therefore, by (4.16) and (4.17) we obtain that, for all n large enough,

$$\frac{1 - \int_0^T f(t)\,dt}{4 \sup_{0<t<T} f(t)} \leq d_n \leq \frac{2}{\inf_{0<t<T} f(t)}\,. \tag{4.18}$$

An immediate conclusion is the following fact. By construction, the distribution of $X_n(0)$ is absolutely continuous; let g_n denote the right continuous version of its density. Since $\mathbf{X}_n$ is obtained by uniform shifting of a piecewise linear periodic function with period H_nT, the value of the density $g_n(v)$ at each point v times the length of the period does not exceed the total number of the linear pieces in a period divided by the smallest absolute slope of any linear piece. The former does not exceed $2m_n$, and by **Property 3** and the above, the latter cannot be smaller than

$$\min\left(\frac{1}{2^{\Delta_f} d_n},\, \tau_{d_n}\right)\,.$$

Since, by (4.18), d_n is uniformly bounded from above, we conclude, for some finite positive constant $c = c(f)$, $g_n(v) \leq c(f) m_n / H_n$. Further, by the definition of d_n,

$$m_n d_n = H_n T P\big(\tau_{\mathbf{X}_n,T} \in \{0, T\}\big) \leq H_n T\,.$$

Once again, since by (4.18), d_n is uniformly bounded from below, we conclude that

$$g_n(v) \text{ is uniformly bounded in } v \text{ and } n. \tag{4.19}$$

Let $\mathbf{X}_n$ be the stationary process corresponding to f_n constructed above. We view $\mathbf{X}_n$ as a random element of the space $C(\mathbb{R})$ of continuous functions on $\mathbb{R}$ which we endow with the metric

$$\rho(\mathbf{x}, \mathbf{y}) = \sum_{m=1}^{\infty} 2^{-m} \bigl(\sup_{|t| \leq m} |x(t) - y(t)| \bigr) .$$

Let μ_n be the law of $\mathbf{X}_n$ on $C(\mathbb{R})$, $n = 1, 2, \ldots$ (but large enough, as needed). By **Property 1** and **Property 2** of the processes $\mathbf{X}_n$ and the lower bound in (4.18), these processes are uniformly bounded and equicontinuous. Therefore, by Theorem 7.3 in [5], for every fixed $m = 1, 2, \ldots$ the restrictions of the measures μ_n to the interval $[-m, m]$ form a tight family of probability measures. Let $n_{1j} \to \infty$ be a sequence positive integers such that the restrictions of $\mu_{n_{1j}}$ to $[-1, 1]$ converge weakly to a probability measure ν_1 on $C([-1, 1])$. Inductively define for $m = 2, 3, \ldots$ $n_{mj} \to \infty$ to be a subsequence of the sequence $n_{m-1,j} \to \infty$ such that the restrictions of $\mu_{n_{mj}}$ to $[-m, m]$ converge weakly to a probability measure ν_m on $C([-m, m])$. Then the "diagonal" sequence of measures $\bigl(\mu_{n_{jj}}, \, j = 1, 2, \ldots\bigr)$ is such that the restrictions of these measures to each interval $[-m, m]$ converge weakly to ν_m on $C([-m, m])$. By the Kolmogorov existence theorem, there is a (cylindrical) probability measure ν on functions on $\mathbb{R}$ whose restrictions to each interval $[-m, m]$ coincide with ν_m (considered now as a cylindrical measure). Since each probability measure ν_m is supported by $C([-m, m])$, the measure ν itself is supported by functions in $C(\mathbb{R})$. By construction, the measure ν is shift invariant. If $\mathbf{X}$ is the canonical stochastic process defined on $\bigl(C(\mathbb{R}), \nu\bigr)$, then $\mathbf{X}$ is a sample continuous stationary process. In the remainder of the proof we will show that $\mathbf{X}$ satisfies Assumption L and Assumption U$_T$, and that $f_{\mathbf{X},T} = f$.

We start with proving that Assumption L holds for $\mathbf{X}$. It is, clearly, enough to prove that, on a set of probability 1,

$$\text{any two local maxima of } \mathbf{X} \text{ are at least } \theta := \frac{1 - \int_0^T f(t)\,dt}{5\sup_{0<t<T} f(t)} \text{ apart.} \tag{4.20}$$

Suppose that (4.20) fails. Then there is m sucht that, on an event of positive probability, two local maxima of $\mathbf{X}$ closer than θ exist in the time interval $[-m, m]$. Recall that a subsequence of the sequence of the (laws of) $\mathbf{X}_n$ converges weakly in the uniform topology on $C([-m, m])$ to the (law of) $\mathbf{X}$. For notational simplicity we will identify that subsequence with the entire sequence $(\mathbf{X}_n)$. By the Skorohod representation theorem (Theorem 6.7 in [5]), we may define the processes $(\mathbf{X}_n)$ on some probability space so that $\mathbf{X}_n \to \mathbf{X}$ a.s. in $C([-m, m])$. Fix ω for which this convergence holds, and for which $\mathbf{X}$ has two local maxima closer than θ exist in the time interval $[-m, m]$. It straightforward to check that the uniform convergence and **Property 3** above imply that for all n large enough, the processes $\mathbf{X}_n$ will have two local maxima closer than $5\theta/4$. This is, of course, impossible, due to **Property 4** and (4.18). The resulting contradiction proves that $\mathbf{X}$ satisfies Assumption L.

Next, we prove that Assumption U$_T$ holds for $\mathbf{X}$. Since the process $\mathbf{X}$ satisfies Assumption L, by Lemma 3.2.2 in Chapter 3, it has finitely many local maxima in the interval $(0, T)$ (in fact, by (4.20), it cannot have more than $\lceil T/\theta \rceil$ local maxima). Clearly, the values of $\mathbf{X}$ at the largest local maximum and the second largest local maximum (if any) are well defined random variables. We denote by (M_1, M_2) the largest and the second largest among $X(0)$, $X(T)$ and the values of $\mathbf{X}$ at the largest local maximum and the second largest local maximum (if any). The fact that Assumption U$_T$ holds for $\mathbf{X}$ will follow once we prove that

$$P\big(M_1 = M_2\big) = 0\,. \tag{4.21}$$

We proceed similarly to the argument in the proof of Assumption L. We may assume that $\mathbf{X}_n \to \mathbf{X}$ a.s. in $C[0,T]$. Fix ω for which this convergence holds. The uniform convergence and **Property 3** of the processes $(\mathbf{X}_n)$, together with the uniform upper bound on d_n in (4.18), show that, for every local maximum t_ω of $\mathbf{X}$ in the interval $(0,T)$ and any $\delta > 0$, there is $n(\omega,\delta)$ such that for all $n > n(\omega,\delta)$, the process $\mathbf{X}_n$ has a local maximum in the interval $(t_\omega - \delta, t_\omega + \delta)$. This immediately implies that

$$M_1 - M_2 \geq \limsup_{n\to\infty} \bigl(M_1^{(n)} - M_2^{(n)}\bigr)$$

a.s., where the random vector $(M_1^{(n)}, M_2^{(n)})$ is defined for the process $\mathbf{X}_n$ in the same way as the random vector (M_1, M_2) is defined for the process $\mathbf{X}$, $n = 1, 2, \ldots$. In particular, for any $\varepsilon > 0$,

$$P\bigl(M_1 - M_2 < \varepsilon\bigr) \leq \limsup_{n\to\infty} P\bigl(M_1^{(n)} - M_2^{(n)} < \varepsilon\bigr)\,. \tag{4.22}$$

As a first step, notice that, by **Property 4** of the processes $(\mathbf{X}_n)$, for any $\varepsilon < \Delta_f$,

$$P\Bigl(M_1^{(n)} - M_2^{(n)} < \varepsilon, \tag{4.23}$$

$$\text{both } M_1^{(n)} \text{ and } M_2^{(n)} \text{ achieved at local maxima}\Bigr) = 0$$

for each n. Next, since by **Property 4**, at its local maxima the process $\mathbf{X}_n$ can take at most $\Delta_f + 3$ possible values, we conclude by (4.19) that for all $\varepsilon > 0$,

$$P\Bigl(M_1^{(n)} - M_2^{(n)} < \varepsilon, \text{ one of } M_1^{(n)} \tag{4.24}$$

$$\text{and } M_2^{(n)} \text{ is achieved at a local maximum, and one at an endpoint}\Bigr) \leq c_f\varepsilon\,,$$

for some $c_f \in (0,\infty)$. Finally, we consider the case when both $M_1^{(n)}$ and $M_2^{(n)}$ are achieved at the endpoints of the interval $[0,T]$. In that case, it is impossible that $\mathbf{X}_n$ has a local maximum in $(0,T)$, since that would force time 0 to belong to one of the decreasing linear pieces of the process due to left blocks, and time T to belong one of the increasing linear pieces of the process due to right blocks. By construction, the distance between any two points belonging to such intervals is larger than T. That forces $X_n(t), 0 \le t \le T$ to consist of at most two linear pieces. By **Property 3** of the process $\mathbf{X}_n$, in order to achieve $|X_n(0) - X_n(T)| \le \varepsilon$, each block of the proper collection generating $\mathbf{X}_n$ contributes at most an interval of length $\varepsilon / \min(1/(2^\Delta d_n), \tau_{d_n})$ to the set of possible shifts U. Recall that there are m_n blocks in the collection. By the uniform bounds (4.18) we conclude that for all $\varepsilon > 0$,

$$P\Big(M_1^{(n)} - M_2^{(n)} < \varepsilon, \tag{4.25}$$

$$M_1^{(n)} \text{ and } M_2^{(n)} \text{ achieved at the endpoints}\Big)$$

$$\le \varepsilon \frac{m_n}{H_n T} \frac{1}{\min(1/(2^\Delta d_n), \tau_{d_n})} \le \varepsilon \frac{1}{d_n \min(1/(2^\Delta d_n), \tau_{d_n})} \le \tilde{c}_f \varepsilon \, ,$$

for some $\tilde{c}_f \in (0, \infty)$.

Combining (4.22), (4.23), (4.24) and (4.25) we see that for all $\varepsilon > 0$ small enough,

$$P\big(M_1 - M_2 < \varepsilon\big) \le (c_f + \tilde{c}_f)\varepsilon \, .$$

Letting $\varepsilon \downarrow 0$ we obtain (4.21), so that the process $\mathbf{X}$ satisfies Assumption U_T.

It is now a simple manner to finish the proof of the theorem. Assume, once again, that $\mathbf{X}_n \to \mathbf{X}$ a.s. in $C[0, T]$. Fix ω for which this convergence holds, and both $\mathbf{X}$ and each $\mathbf{X}_n$ have a unique supremum in the interval $[0, T]$. It follows from the uniform convergence that $\tau_{\mathbf{X}_n, T} \to \tau_{\mathbf{X}, T}$ as $n \to \infty$. Therefore, we also have that $\tau_{\mathbf{X}_n, T} \Rightarrow \tau_{\mathbf{X}, T}$ (weakly). However, by construction, $f_n(t) \to f(t)$ for every $0 < t < T$. This implies that f is the density of $\tau_{\mathbf{X}, T}$, and the proof of the theorem is complete. □

4.4 Long intervals

In spite of the broad range of possibilities for the distribution of the supremum location shown in the previous section, it turns out that, when the length of an interval becomes large, and the process satisfies a certain strong mixing assumption, uniformity of the distribution of the supremum location becomes visible at certain scales. We make this statement precise in this section.

In this section we allow a stationary process $\mathbf{X}$ to have upper semi-continuous, not necessarily continuous, sample paths. Moreover, we will not generally impose either Assumption U$_T$, or Assumption L. Without Assumption U$_T$, the supremum may not be reached at a unique point, so we will work with the leftmost supremum location defined in Section 4.2.

Recall that a stationary stochastic process $\mathbf{X} = (X(t),\, t \in \mathbb{R})$ is called strongly mixing (or α-mixing) if

$$\sup\Big\{ \big|P(A\cap B) - P(A)P(B)\big| :\, A \in \sigma\big(X(s),\, s\leq 0\big),\, B \in \sigma\big(X(s),\, s \geq t\big)\Big\}$$

$$\to 0 \text{ as } t \to \infty;$$

see e.g. [17], p. 195. Sufficient conditions on the spectral density of a stationary Gaussian process that guarantee strong mixing were established in [14].

Let $\mathbf{X}$ be an upper semi-continuous stationary process. We introduce a "tail version" of the strong mixing assumption, defined as follows.

Assumption TailSM: there is a function $\varphi :\, (0,\infty) \to \mathbb{R}$ such that

$$\lim_{t\to\infty} P\big(\sup_{0\leq s\leq t} X(s) \geq \varphi(t)\big) = 1$$

and

$$\sup\Big\{\big|P(A\cap B)-P(A)P(B)\big|:\,A\in\sigma\big(X(s)\mathbf{1}(X(s)\geq\varphi(t)),\\ s\leq 0\big),\,B\in\sigma\big(X(s)\mathbf{1}(X(s)\geq\varphi(t)),\,s\geq t\big)\Big\}\to 0 \text{ as } t\to\infty.$$

It is clear that if a process is strongly mixing, then it also satisfies Assumption TailSM. The point of the latter assumption is that we are only interested in mixing properties of the part of the process "responsible" for its large values. For example, the process

$$X(t)=\begin{cases} Y(t) & \text{if } Y(t)>1 \\ Z(t) & \text{if } Y(t)\leq 1\end{cases}\,,\ t\in\mathbb{R}\,,$$

where $\mathbf{Y}$ is a strongly mixing process such that $P(Y(0)>1)>0$, and $\mathbf{Z}$ an arbitrary stationary process such that $P(Z(0)<1)=1$, does not have to be strongly mixing, but it clearly satisfies Assumption TailSM with $\varphi\equiv 1$.

We will impose one more assumption on the stationary processes we consider in this section. It deals with the size of the largest atom the distribution of the supremum of the process may have.

Assumption A:

$$\lim_{T\to\infty}\sup_{x\in\mathbb{R}}P\big(\sup_{t\in[0,T]}X(t)=x\big)=0\,.$$

In Theorem 4.4.1 below Assumption A could be replaced by requiring Assumption U_T for all T large enough. We have chosen Assumption A instead since for many important stationary stochastic processes the supremum distribution is known to be atomless anyway; see e.g. [22] for continuous Gaussian processes and [6] for certain stable processes. The following sufficient condition for Assumption A is also elementary: suppose that the process $\mathbf{X}$ is ergodic. If for some $a \in \mathbb{R}$, $P\big(\sup_{t\in[0,1]} X(t) = x\big) = 0$ for all $x > a$ and $P(X(0) > a) > 0$, then Assumption A is satisfied.

Theorem 4.4.1. *Let $\mathbf{X} = (X(t),\, t \in \mathbb{R})$ be a stationary sample upper semi-continous process, satisfying Assumption TailSM and Assumption A. The density $f_{\mathbf{X},T}$ of the supremum location satisfies*

$$\lim_{T\to\infty} \sup_{\varepsilon \le t \le 1-\varepsilon} \Big| T f_{\mathbf{X},T}(tT) - 1 \Big| = 0 \tag{4.26}$$

for every $0 < \varepsilon < 1/2$. In particular, the law of $\tau_{\mathbf{X},T}/T$ converges weakly to the uniform distribution on $(0,1)$.

Proof. It is obvious that (4.26) implies weak convergence of the law of $\tau_{\mathbf{X},T}/T$ to the uniform distribution. We will, however, prove the weak convergence first, and then use it to derive (4.26).

We start with a useful claim that, while having nothing to do with any mixing by itself, will be useful for us in a subsequent application of Assumption TailSM. Let $T_n,\ d_n \uparrow \infty$, $d_n/T_n \to 0$ as $n \to \infty$. We claim that for any $\delta \in (0,1)$,

$$P\Big(\delta T_n - d_n \le \tau_{\mathbf{X},T_n} \le \delta T_n + d_n\Big) = 0\,. \tag{4.27}$$

To see this, simply note that by (4.1), the probability in (4.27) is bounded from above by

$$2d_n \sup_{\delta T_n - d_n \le t \le \delta T_n + d_n} f_{\mathbf{X},T_n}(t) \le 2d_n \max\left(\frac{1}{\delta T_n - d_n},\, \frac{1}{(1-\delta)T_n - d_n}\right) \to 0$$

as $n \to \infty$.

The weak convergence stated in the theorem will follow once we prove that for any rational number $r \in (0,1)$, we have $P\big(\tau_{\mathbf{X},T} \le rT\big) \to r$ as $T \to \infty$. Let

$r = m/k$, $m, k \in \mathbb{N}$, $m < k$ be such a rational number. Consider T large enough so that $T > k^2$, and partition the interval $[0, T]$ into subintervals

$$C_i = \left[(T + \sqrt{T})\frac{i}{k},\, (T + \sqrt{T})\frac{i+1}{k} - \sqrt{T} \right],\ i = 0, 1, \ldots, k-1\,,$$

$$D_i = \left[(T + \sqrt{T})\frac{i}{k} - \sqrt{T},\, (T + \sqrt{T})\frac{i}{k} \right],\ i = 1, \ldots, k-1\,,$$

and observe that by (4.27),

$$P\Big(\tau_{\mathbf{X},T} \in \bigcup_{i=1}^{k-1} D_i \Big) \to 0 \text{ as } T \to \infty.$$

Therefore,

$$P\big(\tau_{\mathbf{X},T} \le rT\big) = P\Big(\max_{0 \le i \le m-1} M_{i,T} \ge \max_{m \le i \le k-1} M_{i,T} \Big) + o(1) \tag{4.28}$$

as $T \to \infty$, where $M_{i,T} = \sup_{t \in C_i} X(t)$, $i = 0, 1, \ldots .k-1$.

Let φ be the function given in Assumption TailSM. Then

$$\begin{aligned} & P\Big(\max_{0 \le i \le m-1} M_{i,T} \ge \max_{m \le i \le k-1} M_{i,T} \Big) \\ & = P\Big(\max_{0 \le i \le m-1} V_{i,T} \ge \max_{m \le i \le k-1} V_{i,T} \Big) + o(1)\,, \end{aligned} \tag{4.29}$$

where $V_{i,T} = \sup_{t \in C_i} X(t) \mathbf{1}\big(X(t) > \varphi(\sqrt{T})\big)$, $i = 0, 1, \ldots .k-1$.

Denote by G_T the distribution function of each one of the random variables $V_{i,T}$, and let $W_{i,T} = G_T(V_{i,T})$, $i = 0, 1, \ldots, k-1$. It is clear that

$$\begin{aligned} & P\Big(\max_{0 \le i \le m-1} V_{i,T} \ge \max_{m \le i \le k-1} V_{i,T} \Big) \\ & = P\Big(\max_{0 \le i \le m-1} W_{i,T} \ge \max_{m \le i \le k-1} W_{i,T} \Big)\,. \end{aligned} \tag{4.30}$$

Notice, further, that by Assumption TailSM, for every $0 < w_i < 1, i = 0, 1, \ldots, k-1$,

$$\lim_{T \to \infty} \Big| P\Big(W_{i,T} \le w_i,\ i = 0, 1, \ldots, k-1 \Big) - \prod_{i=0}^{k-1} P\Big(W_{i,T} \le w_i \Big) \Big| = 0\,. \tag{4.31}$$

Let

$$D(T) = \sup_{x\in\mathbb{R}} P\big(\sup_{t\in C_0} X(t) = x\big) + P\big(\sup_{t\in C_0} X(t) \le \varphi(\sqrt{T})\big)\,.$$

By Assumption A, $D(T) \to 0$ as $T \to \infty$. Since for every $0 < w < 1$,

$$w - D(T) \le P\Big(W_{0,T} \le w\Big) \le w\,,$$

we conclude by (4.31) that the law of the random vector $\big(W_{0,T}, \ldots, W_{k-1,T}\big)$ converges weakly, as $T \to \infty$, to the law of a random vector $(U_0, \ldots, U_{k-1})$ with independent standard uniform components. Since this limiting law does not charge the boundary of the set $\{(w_0, w_1, \ldots, w_{k-1}) : \max_{0\le i\le m-1} w_i \le \max_{m\le i\le k-1} w_i\}$, we conclude by (4.28), (4.29) and (4.30) that

$$P\big(\tau_{\mathbf{X},T} \le rT\big) \to P\big(\max_{0\le i\le m-1} U_i \ge \max_{m\le i\le k-1} U_i\big) = m/k = r\,,$$

and so we have established the weak convergence claim of the theorem.

We now prove the uniform convergence of the densities in (4.26). Suppose that the latter fails for some $0 < \varepsilon < 1/2$. There are two possibilities. Suppose first that there is $\theta > 0$, a sequence $T_n \to \infty$ and a sequence $t_n \in [\varepsilon, 1-\varepsilon]$ such that for every n, $T_n f_{\mathbf{X},T_n}(t_nT_n) \ge 1+\theta$. By compactness we may assume that $t_n \to t_* \in [\varepsilon, 1-\varepsilon]$ as $n \to \infty$. By Lemma 4.2.1 and the regularity properties of the density, for every n and every $0 < \tau, \delta < 1$ such that

$$\big(1 - (1-\tau)/t_n\big)_+ < \delta < \min\big(\tau/t_n, 1\big) \tag{4.32}$$

we have

$$T_n f_{\mathbf{X},(1-\tau)T_n}(t_n(1-\delta)T_n) \ge T_n f_{\mathbf{X},T_n}(t_nT_n) \ge 1+\theta\,.$$

Since $t_n \to t_*$, there is a choice of $0 < \tau < 1$ such that

$$1+\theta > \frac{1}{1-\tau} \tag{4.33}$$

and, moreover, the range in (4.32) is nonempty for all n large enough. Furthermore, we can find $0 < a < b < 1$ such that

$$\big(1-(1-\tau)/t_n\big)_+ < a < b < \min\big(\tau/t_n, 1\big)$$

for all n large enough. Therefore, for such n

$$(1+\theta)(b-a) \le \int_a^b T_n f_{\mathbf{X},(1-\tau)T_n}(t_n(1-\delta)T_n)\, d\delta$$

$$= \frac{1}{t_n} P\Big(\tau_{\mathbf{X},(1-\tau)T_n} \in \big((1-b)t_nT_n,\ (1-a)t_nT_n\big)\Big) \to \frac{1}{1-\tau}(b-a)$$

as $n \to \infty$ by the already established weak convergence. This contradicts the choice (4.33) of τ.

The second way (4.26) can fail is that there is $0 < \theta < 1$, a sequence $T_n \to \infty$ and a sequence $t_n \in [\varepsilon, 1-\varepsilon]$ such that for every n, $T_n f_{\mathbf{X},T_n}(t_nT_n) \le 1-\theta$. We can show that this option is impossible as well by appealing, once again, to Lemma 3.3.1 and using an argument nearly identical to the one described above. Therefore, (4.26) holds, and the proof of the theorem is complete. $\square$

The following corollary is an immediate conclusion of Theorem 4.4.1. It shows the uniformity of the limiting conditional distribution of the location of the supremum given that it belongs to a suitable subinterval of $[0, T]$.

Corollary 4.4.2. *Let* $\mathbf{X} = (X(t),\, t \in \mathbb{R})$ *be a stationary sample upper semi-continous process, satisfying Assumption TailSM and Assumption A. Let* $0 < a_T \le a'_T < b'_T \le b_T < T$ *be such that*

$$\liminf_{T\to\infty} \frac{a_T}{T} > 0,\ \ \limsup_{T\to\infty} \frac{b_T}{T} < 1,\ \ \lim_{T\to\infty} \frac{b'_T - a'_T}{b_T - a_T} = \theta.$$

Then

$$\lim_{T\to\infty} P\Big(\tau_{\mathbf{X},T} \in \big(a'_T, b'_T\big)\Big|\tau_{\mathbf{X},T} \in \big(a_T, b_T\big)\Big) = \theta\,.$$

CHAPTER 5

INTRINSIC LOCATION FUNCTIONALS OF STATIONARY PROCESSES

5.1 Introduction

We consider a large family of measurable functionals of the sample paths of a stochastic process restricted to a compact interval in the real line. The functionals are “intrinsically” connected to the sample path in the sense that they shift together with the path; this is why we call them intrinsic location functionals. They include various first/last hitting times, first/last locations of the largest value/largest jump of the process, and many others. These functionals are often highly discontinuous functions of the sample path, and for a specific process their distribution is either very difficult to derive, or else rests on a very specific property of the process, such as a Markov property.

In this chapter we study the distribution of such functionals from a different point of view. Instead of looking at a specific stochastic process, we study the general question of how the stationarity of a stochastic process affects the distribution of an intrinsic location functional. Specifically, we show that the laws of any such functionals are absolutely continuous when restricted to the interior of the interval, and their densities have a version that satisfies very specific total variation constraints. For one very specific functional, the leftmost location of the supremum over an interval, such total variation constraints were established in Chapter 3, but in this chapter we show that this behaviour is universal, in the sense that the constraints are shown to hold for a large variety of functionals. This universality turns out to be a characterization of stationarity. That is, given a fixed stochastic process, if for a rich enough subfamily of intrinsic location functionals, the distribution of the functional has a density within each interval that satisfies the total variation constraints, then the process has to be stationary.

We study the structure of the family of the probability distributions characterized by the total variation constraints. We determine its extreme points and show how this can be used so solve certain extremal problems.

The rest of this chapter is organized as follows. In Section 5.2 we define the intrinsic location functionals and consider a number of examples. A description of the very specific features of the laws of intrinsic location functionals of stationary processes is stated and proved in Section 5.3, where we also include a discussion showing that all the defining properties of intrinsic location functionals are necessary for the conclusions of the theorem to hold. In Section 5.4 we discuss the structure of the set of all possible distributions of intrinsic location functionals and use it to solve certain extremal problems related to these functionals. In Section 5.5 we establish that the total variation constraints characterize stationarity of the process. The results of this section are refined and generalized in Section 5.6.

5.2 Intrinsic Location Functionals

Let H be a set of functions on $\mathbb{R}$, invariant under shifts. That is, for any $f \in H$ and $c \in \mathbb{R}$ the function $\theta_c f$ defined by $\theta_c f(x) = f(x+c)$, $x \in \mathbb{R}$ belongs to H. We equip H with its cylindrical σ-field. Let $\mathcal{I}$ be the set of all compact, non-degenerate intervals in $\mathbb{R}$: $\mathcal{I} = \{[a,b] : a < b, [a,b] \subset \mathbb{R}\}$.

Definition 1. *A mapping $L : H \times \mathcal{I} \to \mathbb{R} \cup \{\infty\}$ is called an intrinsic location* *functional, if it satisfies the following conditions.*

1. *For every $I \in \mathcal{I}$ the map $L(\cdot, I) : H \to \mathbb{R} \cup \{\infty\}$ is measurable.*
2. *For every $f \in H$ and $I \in \mathcal{I}$, $L(f, I) \in I \cup \{\infty\}$.*
3. *(Shift compatibility) For every $f \in H$, $I \in \mathcal{I}$ and $c \in \mathbb{R}$,*

   $$L(f, I) = L(\theta_c f, I - c) + c,$$

   where $I - c$ is the interval I shifted by $-c$, and $\infty + c = \infty$.
4. *(Stability under restrictions) For every $f \in H$ and $I_1, I_2 \in \mathcal{I}$, $I_2 \subseteq I_1$,*

   $$\text{if } L(f, I_1) \in I_2, \text{ then } L(f, I_2) = L(f, I_1).$$

5. *(Consistency of existence) For every $f \in H$ and $I_1, I_2 \in \mathcal{I}$, $I_2 \subseteq I_1$,*

   $$\text{if } L(f, I_2) \neq \infty, \text{ then } L(f, I_1) \neq \infty.$$

We associate the possibility of an infinite value of L with "non-existence": a certain condition is never satisfied over the interval I if $L(f, I) = \infty$. Otherwise, $L(f, I) \in I$. The shift compatibility requirement is the reason for the adjective "intrinsic". The stability under restrictions property asserts the global nature of L over the interval I. Finally, the consistency of existence property says that, if a certain condition is satisfied somewhere over a small interval, it is definitely satisfied somewhere over a larger interval as well.

Example 5.2.1. Let H be the space of the upper semi-continuous functions. Then the leftmost location of the supremum over the interval, defined as

$$\tau_{f,[a,b]} := \inf\{s \in [a, b] : f(s) = \sup_{t \in [a,b]} f(t)\}$$

is an intrinsic location functional. As we have already seen, this functional was considered in detail in Chapters 3 and 4. It is an example of intrinsic location functional that does not take an infinite value. Of course, a similarly defined rightmost location of the supremum over the interval is an intrinsic location functional as well.

Example 5.2.2. Let H be the space of continuous functions $\mathcal{C}(\mathbb{R})$. Then the first hitting time of certain level l, defined as

$$T^l_{f,[a,b]} := \inf\{s \in [a,b] : f(s) = l\}$$

is an intrinsic location functional. Replacing in this definition infimum by supremum leads to the last exit time of the level l, which is also an intrinsic location functional. In both cases an infinite value is a possibility.

It is easy to think of many other examples of intrinsic location functionals. A few further examples are the leftmost/rightmost point with the largest/smallest slope for C^1 functions, or the leftmost/rightmost location of the largest jump/the jump whose size is the closest to a given number for càdlàg functions. On the other hand, certain natural functionals fail to be intrinsic location functionals, as the following examples show.

Example 5.2.3. Let $H = \mathcal{C}(\mathbb{R})$. The first hitting time of a level l after a given time point t:

$$T^l_{t,f,[a,b]} := \inf\{s \in [a,b], s \geq t : f(s) = l\}$$

is not an intrinsic location functional, since it involves a fixed point t and, therefore, is not shift compatible.

Example 5.2.4. Let H be the set of all continuous functions on $\mathbb{R}$ with separated local maxima. That is, for every $f \in H$ and compact interval $[a,b]$ there is $\delta > 0$ so that $|t_1 - t_2| \geq \delta$ for any two different local maxima t_1, t_2 of f in $[a,b]$.

Given a function $f \in H$ and an interval $[a, b]$, denote by $A = \{t_1, t_2, ...\}$ the set of local maxima of f on $[a, b]$. Then the leftmost largest local maximum

$$M^1_{f,[a,b]} := \inf\{s \in A : f(s) = \sup_{t \in A} f(t)\}$$

is an intrinsic location functional; it is just the leftmost location of supremum over the interval of Example 5.2.1. However, the location of the leftmost second largest local maximum

$$M^2_{f,[a,b]} := \inf\{s \in A \setminus \{M^1_{f,[a,b]}\} : f(s) = \sup_{t \in A \setminus \{M^1_{f,[a,b]}\}} f(t)\}$$

is not an intrinsic location functional, even though it is shift compatible. On a smaller interval, the second largest local maximum of the larger interval may become the largest local maximum. Therefore this functional is not stable under restrictions.

Example 5.2.5. Let $H = \mathcal{C}(\mathbb{R})$. Then the first hitting time of certain level l within a fixed distance d to the right endpoint of the interval, defined as

$$T^{l,d}_{f,[a,b]} := \inf\{s \in [a, b], s \geq b - d : f(s) = l\}$$

is not an intrinsic functional. Although it is both shift compatible and stable under restrictions, it does not possess consistency of existence: such a hitting time may exist on a smaller interval, but disappear on a larger interval since the original location is now too far from the right endpoint of the interval.

In the remainder of this chapter $\mathbf{X} = (X(t), t \in \mathbb{R})$ is a stationary process defined on some probability space $(\Omega, \mathcal{F}, P)$, and having sample paths in H. For a compact interval $[a, b]$, we will denote the value of an intrinsic location functional L evaluated on the process $\mathbf{X}$ on that interval by $L(\mathbf{X}, [a, b])$. Note that our assumptions imply that $L(\mathbf{X}, [a, b])$ is a well defined $[a, b] \cup \{\infty\}$-valued random variable. Stationarity of the process and shift compatibility of L, clearly, imply that the distribution of L on an interval, relatively to its left endpoint, depends only on the length of the interval. Thus we will often study intervals of the type $[0, b]$, in which case, we will use the corresponding single variable notation $L(\mathbf{X}, b)$.

We denote by $F_{\mathbf{X},[a,b]}$ the law of $L(\mathbf{X}, [a, b])$; it is a probability measure supported on the set $[a, b] \cup \{\infty\}$. Again, if the interval is of the type $[0, b]$, the corresponding notation is $F_{\mathbf{X},b}$. We preserve the same notation for the cumulative distribution function, i.e. we will write $F_{\mathbf{X},[a,b]}(t)$ for the value $F_{\mathbf{X},[a,b]}$ assigns to the interval $[a, t]$, $a \leq t \leq b$, with the corresponding single variable notation if $a = 0$.

5.3 Properties of the distributions of intrinsic location functionals of stationary processes

The main result of this section is an extension of most parts of Theorem 3.3.1 from the special case of the leftmost location of the supremum to the general intrinsic location functionals defined in the previous section.

Theorem 5.3.1. *Let L be an intrinsic location functional and $\mathbf{X} = (X(t), t \in \mathbb{R})$ a stationary process. Then the restriction of the law $F_{\mathbf{X},T}$ to the interior $(0, T)$ of the interval is absolutely continuous. The density, denoted by $f_{\mathbf{X},T}$, can be taken to be equal to the right derivative of the cdf $F_{\mathbf{X},T}$, which exists at every point in the interval $(0, T)$. In this case the density is right continuous, has left limits, and has the following properties.*

(a) The limits

$$f_{\mathbf{X},T}(0+) = \lim_{t\to 0} f_{\mathbf{X},T}(t) \text{ and } f_{\mathbf{X},T}(T-) = \lim_{t\to T} f_{\mathbf{X},T}(t)$$

exist.

(b) The density has a universal upper bound given by

$$f_{\mathbf{X},T}(t) \le \max\left(\frac{1}{t}, \frac{1}{T-t}\right),\ 0 < t < T\,. \tag{5.1}$$

(c) The density has a bounded variation away from the endpoints of the interval. Furthermore, for every $0 < t_1 < t_2 < T$*,*

$$TV_{(t_1,t_2)}(f_{\mathbf{X},T}) \le \min\bigl(f_{\mathbf{X},T}(t_1), f_{\mathbf{X},T}(t_1-)\bigr) + \min\bigl(f_{\mathbf{X},T}(t_2),\ f_{\mathbf{X},T}(t_2-)\bigr)\,, \tag{5.2}$$

where

$$TV_{(t_1,t_2)}(f_{\mathbf{X},T}) = \sup\sum_{i=1}^{n-1}\bigl|f_{\mathbf{X},T}(s_{i+1}) - f_{\mathbf{X},T}(s_i)\bigr|$$

is the total variation of $f_{\mathbf{X},T}$ *on the interval* (t_1, t_2)*, and the supremum is taken over all choices of* $t_1 < s_1 < \ldots < s_n < t_2$*.*

(d) The density has a bounded positive variation at the left endpoint and a bounded negative variation at the right endpoint. Furthermore, for every $0 < \varepsilon < T$*,*

$$TV^{+}_{(0,\varepsilon)}(f_{\mathbf{X},T}) \le \min\bigl(f_{\mathbf{X},T}(\varepsilon), f_{\mathbf{X},T}(\varepsilon-)\bigr) \tag{5.3}$$

and

$$TV^{-}_{(T-\varepsilon,T)}(f_{\mathbf{X},T}) \le \min\bigl(f_{\mathbf{X},T}(T-\varepsilon), f_{\mathbf{X},T}(T-\varepsilon-)\bigr)\,, \tag{5.4}$$

where for any interval $0 \le a < b \le T$*,*

$$TV^{\pm}_{(a,b)}(f_{\mathbf{X},T}) = \sup\sum_{i=1}^{n-1}\bigl(f_{\mathbf{X},T}(s_{i+1}) - f_{\mathbf{X},T}(s_i)\bigr)_{\pm}$$

is the positive (negative) variation of $f_{\mathbf{X},T}$ *on the interval* (a, b)*, and the supremum is taken over all choices of* $a < s_1 < \ldots < s_n < b$*.*

(e) The limit $f_{\mathbf{X},T}(0+) < \infty$ *if and only if* $TV_{(0,\varepsilon)}(f_{\mathbf{X},T}) < \infty$ *for some (equivalently, any)* $0 < \varepsilon < T$*, in which case*

$$TV_{(0,\varepsilon)}(f_{\mathbf{X},T}) \le f_{\mathbf{X},T}(0+) + \min\bigl(f_{\mathbf{X},T}(\varepsilon), f_{\mathbf{X},T}(\varepsilon-)\bigr) . \tag{5.5}$$

Similarly, $f_{\mathbf{X},T}(T-) < \infty$ *if and only if* $TV_{(T-\varepsilon,T)}(f_{\mathbf{X},T}) < \infty$ *for some (equivalently, any)* $0 < \varepsilon < T$*, in which case*

$$TV_{(T-\varepsilon,T)}(f_{\mathbf{X},T}) \le \min\bigl(f_{\mathbf{X},T}(T-\varepsilon), f_{\mathbf{X},T}(T-\varepsilon-)\bigr) + f_{\mathbf{X},T}(T-) . \tag{5.6}$$

The proof of Theorem 5.3.1 is parallel to the proof of Theorem 3.3.1; we provide an outline here. In particular, we have to verify that the possibility of an infinite value (impossible in the earlier work) is consistent with the argument.

Proof. We start with a lemma that is a counterpart of Lemma 3.2.1.

Lemma 5.3.1. (i) *For any* $\Delta \in \mathbb{R}$,

$$F_{\mathbf{X},[\Delta,T+\Delta]}(\cdot) = F_{\mathbf{X},T}(\cdot - \Delta) .$$

(ii) *For any intervals* $[c,d] \subseteq [a,b]$,

$$F_{\mathbf{X},[a,b]}(B) \le F_{\mathbf{X},[c,d]}(B) \text{ for any Borel set } B \subset [c,d].$$

(iii) *For any intervals* $[c,d] \subseteq [a,b]$,

$$F_{\mathbf{X},[a,b]}(\{\infty\}) \le F_{\mathbf{X},[c,d]}(\{\infty\}).$$

Clearly, the three statements of Lemma 5.3.1 are directly implied by, respectively, shift compatibility, stability under restrictions and consistency of existence properties of intrinsic location functionals.

Choose $0 < \delta < T/2$. Using shift compatibility and stability under restrictions together with the stationarity of the process, the argument in Chapter 3 shows that for every $\delta \le t \le T-\delta$, for every $\rho > 0$ and every $0 < \varepsilon < \delta\rho/(1+\rho)$

$$P\bigl(t < L(\mathbf{X},T) \le t+\varepsilon\bigr) \le \varepsilon(1+\rho)\max\left(\frac{1}{t},\frac{1}{T-t}\right)\,; \tag{5.7}$$

a possibility of an infinite value does not play a role in this argument. Obviously, (5.7) implies absolute continuity of $F_{\mathbf{X},T}$ on the interval $(\delta, T-\delta)$ and, since $\delta > 0$ can be taken to be arbitrarily small, also on $(0,T)$. The version of the density given by

$$f_{\mathbf{X},T}(t) = \limsup_{\varepsilon\downarrow 0}\frac{1}{\varepsilon}P\bigl(t < L(\mathbf{X},T) \le t+\varepsilon\bigr),\ 0 < t < T\,,$$

automatically satisfies the bound (5.1).

The second important ingredient in the proof of the theorem is the following lemma, which is analogous to Lemma 3.3.1. Here the infinite value does play a role, so the consistency of existence property of intrinsic functionals has to be used.

Lemma 5.3.2. *Let $0 \le \Delta < T$. Then for every $0 \le \delta \le \Delta$, $f_{\mathbf{X},T-\Delta}(t) \ge f_{\mathbf{X},T}(t+\delta)$ almost everywhere in $(0, T-\Delta)$. Furthermore, for every such δ and every $\varepsilon_1, \varepsilon_2 \ge 0$, such that $\varepsilon_1 + \varepsilon_2 < T - \Delta$,*

$$\int_{\varepsilon_1}^{T-\Delta-\varepsilon_2}\bigl(f_{\mathbf{X},T-\Delta}(t) - f_{\mathbf{X},T}(t+\delta)\bigr)\,dt \tag{5.8}$$

$$\le \int_{\varepsilon_1}^{\varepsilon_1+\delta} f_{\mathbf{X},T}(t)\,dt + \int_{T-\Delta-\varepsilon_2+\delta}^{T-\varepsilon_2} f_{\mathbf{X},T}(t)\,dt\,.$$

Proof. The statement $f_{\mathbf{X},T-\Delta}(t) \ge f_{\mathbf{X},T}(t+\delta)$ almost everywhere in $(0, T-\Delta)$ follows from Lemma 5.3.1 as in Chapter 3. For (5.8), we have

$$\int_{\varepsilon_1}^{T-\Delta-\varepsilon_2}\bigl(f_{\mathbf{X},T-\Delta}(t) - f_{\mathbf{X},T}(t+\delta)\bigr)\,dt$$

$$= P\big(L(\mathbf{X}, T-\Delta) \in (\varepsilon_1, T-\Delta-\varepsilon_2)\big) - P\big(L(\mathbf{X}, T) \in (\varepsilon_1+\delta, T-\Delta-\varepsilon_2+\delta)\big)$$

$$= P\big(L(\mathbf{X}, T) \notin (\varepsilon_1+\delta, T-\Delta-\varepsilon_2+\delta)\big) - P\big(L(\mathbf{X}, T-\Delta) \notin (\varepsilon_1, T-\Delta-\varepsilon_2)\big)$$

$$= P\big(L(\mathbf{X}, T) \in [0, \varepsilon_1+\delta)\big) + P\big(L(\mathbf{X}, T) \in (T-\Delta-\varepsilon_2+\delta, T]\big)$$

$$+P\big(L(\mathbf{X}, T) = \infty\big) - P\big(L(\mathbf{X}, T-\Delta) \in [0, \varepsilon_1)\big)$$

$$-P\big(L(\mathbf{X}, T-\Delta) \in (T-\Delta-\varepsilon_2, T-\Delta]\big) - P\big(L(\mathbf{X}, T-\Delta) = \infty\big)$$

$$= P\big(L(\mathbf{X}, T) \in (\varepsilon_1, \varepsilon_1+\delta)\big) + \Big(P\big(L(\mathbf{X}, T) \in [0, \varepsilon_1)\big) - P\big(L(\mathbf{X}, T-\Delta) \in [0, \varepsilon_1)\big)\Big)$$

$$+P\big(L(\mathbf{X}, T) \in (T-\Delta-\varepsilon_2+\delta, T-\varepsilon_2)\big) + \Big(P\big(L(\mathbf{X}, T) = \infty\big) - P\big(L(\mathbf{X}, T-\Delta) = \infty\big)\Big)$$

$$+\Big(P\big(L(\mathbf{X}, T) \in (T-\varepsilon_2, T]\big) - P\big(L(\mathbf{X}, [\Delta, T]) \in (T-\varepsilon_2, T]\big)\Big)$$

$$\leq P\big(L(\mathbf{X}, T) \in (\varepsilon_1, \varepsilon_1+\delta)\big) + P\big(L(\mathbf{X}, T) \in (T-\Delta-\varepsilon_2+\delta, T-\varepsilon_2)\big)$$

$$= \int_{\varepsilon_1}^{\varepsilon_1+\delta} f_{\mathbf{X},T}(t)\, dt + \int_{T-\Delta-\varepsilon_2+\delta}^{T-\varepsilon_2} f_{\mathbf{X},T}(t)\, dt\,,$$

since by Lemma 5.3.1, all the differences of probabilities above are non-positive.

□

Lemmas 5.3.1 and 5.3.2 are the only tools needed to complete the proof of Theorem 5.3.1 as in Chapter 3. □

Absence of even one of the three defining properties of an intrinsic location functional will, generally, void the conclusions of Theorem 5.3.1. To demonstrate that, we will use examples 5.2.3, 5.2.4 and 5.2.5 above. In all cases we will use a very simple periodic stationary process $X_{\text{per}}(t) = \sin(t+U)$, $t \in \mathbb{R}$, where U is uniformly distributed between 0 and 2π. We will also use a simple device to show a failure of the conclusions of Theorem 5.3.1: suppose that for some $0 < a < b < T$ we have $P\big(L(\mathbf{X}, T) \in [a, b]\big) = 1$. Then a density with the prescribed total variation properties cannot exist. Indeed, take $0 < t_1 < a$,

$b < t_2 < T$. Then the right hand side of (5.2) vanishes. On the other hand, the largest value of the density over the interval $[a, b]$ cannot be smaller than $1/(b-a)$, so the left hand side of (5.2) cannot be smaller than $2/(b-a)$.

Example 5.3.2. The first hitting time after a given time defined in Example 5.2.3: $T^l_{t,f,[a,b]} := \inf\{s \in [a,b], s \geq t : f(s) = l\}$ satisfies stability under restrictions and consistency of existence, but not shift compatibility. Take $l = 0$, $t > 0$ and $T > t+\pi$. Then for the periodic process $\mathbf{X}_{\text{per}}$ above, $P(T^l_{t,\mathbf{X}_{\text{per}},[0,T]} \in [t, t+\pi]) = 1$, and the conclusions of Theorem 5.3.1 cannot hold.

Example 5.3.3. The leftmost second largest local maximum functional $M^2_{f,[a,b]}$ of Example 5.2.4 satisfies shift compatibility and consistency of existence, but not stability under restrictions. Let $T > 2\pi$. For the periodic process $\mathbf{X}_{\text{per}}$ above, $P\big(M^2_{\mathbf{X}_{\text{per}},[0,T]} \in [\pi, 2\pi]\big) = 1$, so the conclusions of Theorem 5.3.1 cannot hold.

Example 5.3.4. The first hitting time of a level l within a fixed distance d to the right endpoint of the interval, $T^{l,d}_{f,[a,b]}$ of Example 5.2.5, satisfies shift compatibility and stability under restrictions, but not consistency of existence. Let $l = 0$ and $T > d > \pi$. Then for the periodic process $\mathbf{X}_{\text{per}}$ above, $P(T^{l,d}_{\mathbf{X}_{\text{per}},[0,T]} \in [T-d, T-d+\pi]) = 1$. Once again, the conclusions of Theorem 5.3.1 cannot hold.

We end this section by showing the following result, which introduces ergodicity into the scenario and explains how it will affect the infimum of the density function.

Theorem 5.3.5. *Let* $\mathbf{X} = \{X(t)\}_{t\in\mathbb{R}}$ *be an ergodic stationary process with path space* H. *Let* L *be an intrinsic location functional defined on* $H \times \mathcal{I}$. *Denote by* $f(t)$ *the density function of* $L(\mathbf{X}, [0,T])$ *for some positive real number* T. *If* $f(0+) > 0$ *and* $f(T-) > 0$, *then* $\inf_{t\in(0,T)} f(t) > 0$.

Proof. Suppose the theorem is not true. Then there exist an ergodic stationary process $\mathbf{X}$, an intrinsic location functional L, a positive number T_0 and the density function f of $L(\mathbf{X}, [0, T_0])$, such that $f(0+) > 0$, $f(T_0-) > 0$, and $\inf_{t\in(0,T_0)} f(t) = 0$. Since f is càdlàg, there exists $s \in (0, T_0)$, such that $\min\{f(s), f(s-)\} = 0$. The limits $f(0+) > 0$ and $f(T_0-) > 0$ guarantees that $P(L(\mathbf{X}, [0, T_0]) \in (0, s)) > 0$ and $P(L(\mathbf{X}, [0, T_0]) \in (s, T_0)) > 0$. Consider interval $[0, T]$ with $T > T_0$, and denote by f_T the corresponding density function. By Lemma 5.3.1, $f_T(t) = 0$ for $t \in (s, s + T - T_0)$, $f_T(t) \le f(t)$ for $t \in (0, s)$, and $f_T(t) \le f(t - T + T_0)$ for $t \in (s + T - T_0, T)$. Moreover, the point masses on the boundaries satisfy

$$P(L(\mathbf{X}, [0, T_0]) = 0) \ge P(L(\mathbf{X}, [0, T]) = 0),$$

$$P(L(\mathbf{X}, [0, T_0]) = T_0) \ge P(L(\mathbf{X}, [0, T]) = T).$$

On the other hand, Lemma 5.3.1 also implies that

$$P(L(\mathbf{X}, [0, T]) \in (0, s) \cup (s + T - T_0, T)) + P(L(\mathbf{X}, [0, T] = 0 \text{ or } T)$$

$$= P(L(\mathbf{X}, [0, T]) \ne \infty) \ge P(L(\mathbf{X}, [0, T_0]) \ne \infty)$$

$$= P(L(\mathbf{X}, [0, T_0]) \in (0, s) \cap (s, T_0)) + P(L(\mathbf{X}, [0, T_0]) = 0 \text{ or } T_0).$$

Combining these results leads to $f(t) = f_T(t), \forall t \in (0, s)$ and $f(t) = f_T(t + T - T_0), \forall t \in (s, T_0)$. We also have the integrated version $P(L(\mathbf{X}, [0, T_0]) \in (0, s)) = P(L(\mathbf{X}, [0, T]) \in (0, s))$, and $P(L(\mathbf{X}, [0, T_0]) \in (s, T_0)) = P(L(\mathbf{X}, [0, T]) \in [s + T - T_0, T])$. By (cite lemma) the last two equalities, satisfied for arbitrary $T > T_0$, imply that with probability 1, either $L(\mathbf{X}, [0, T]) \in (0, s), \forall T > T_0$, or $L(\mathbf{X}, [0, T]) \in (s + T - T_0, T), \forall T > T_0$. Denote these two events by A and B. Notice that both of them are of strictly positive probabilities.

Now consider a functional $g : H \to \mathbb{R}$, defined by $g(X) = 1_{\{L(X,[0,T_0]) \in (0,s)\}}$. Since L is measurable, g is clearly also measurable. Then

$$E(g(\mathbf{X})) = P(L(\mathbf{X}, [0, T_0]) \in (0, s)) = P(A).$$

According to Birkhoff's ergodic theorem, we should have with probability 1

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T g \circ \theta_u(\mathbf{X}) du = E(g(\mathbf{X})) = P(A) > 0,$$

where θ_u is the shift operator: $\theta_u X(t) = X(t + u)$. However, on event B, since $L(\mathbf{X}, [0, T]) \in (s + T - T_0, T), \forall T > T_0$, the stability under restriction property of intrinsic location functional requires $L(\mathbf{X}, [T - T_0, T]) \in (s + T - T_0, T), \forall T > T_0$, thus $L(\mathbf{X}, [T - T_0, T]) \notin (T - T_0, T - T_0 + s), \forall T > T_0$. As a result,

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T g \circ \theta_u(\mathbf{X}) = 0$$

on B. Contradiction. Therefore the assumption at the beginning of the proof is false and the theorem is proved.

□

5.4 Structure of the set of all possible distributions

Theorem 5.3.1 of the previous section shows that the distribution of $L(\mathbf{X}, T)$ for any intrinsic location functional L, any stationary process $\mathbf{X}$ and any positive real number T is of a very special type. In this section we study the fine structure of this class of laws.

We denote by A_T the class of probability measures F on $[0, T] \cup \{\infty\}$ with the following properties.

1. The restriction of F to the interior $(0,T)$ of the interval is absolutely continuous.
2. A version of the density is given by the right derivative of the cdf $F\big([0,t]\big)$, $0<t<T$, which exists at every point in the interval $(0,T)$.
3. This density f is right continuous, has left limits, and satisfies the total variation constraints (5.2), (5.3), (5.4), (5.5) and (5.6).

It is elementary to check that the total variation constraints (5.2) imply the upper bound (5.1) on the densities of all laws in A_T.

We endow the set $[0,T]\cup\{\infty\}$ with the topology obtained by treating the infinite point as an isolated point of the set. Let $\mathcal{P}_T$ be the collection of all probability measures on $[0,T]\cup\{\infty\}$.

Theorem 5.4.1. *The set A_T is a weakly closed convex subset of $\mathcal{P}_T$. Moreover, for any $0<\varepsilon<T/2$, the restrictions of the laws in A_T to the interval $(\varepsilon, T-\varepsilon)$ form a compact in total variation family of finite measures.*

Proof. The convexity of A_T is obvious. Fix $0<\varepsilon<T/2$, and let f be the version of the density of an arbitrary member of the class A_T in the interior of the interval $[0,T]$ described in the definition of that class. For $x>0$ small enough, we have

$$\int_\varepsilon^{T-\varepsilon}\big|f(x+y)-f(y)\big|\,dy=\sum_{j=1}^{\lfloor (T-2\varepsilon)/x\rfloor}\int_{\varepsilon+(j-1)x}^{\varepsilon+jx}\big|f(x+y)-f(y)\big|\,dy$$

$$+\int_{\varepsilon+\lfloor (T-2\varepsilon)/x\rfloor x}^{T-\varepsilon}\big|f(x+y)-f(y)\big|\,dy$$

$$\le\int_0^x\sum_{j=1}^{\lfloor (T-2\varepsilon)/x\rfloor}\big|f(\varepsilon+jx+y)-f(\varepsilon+(j-1)x+y)\big|\,dy+\max\left(\frac{1}{\varepsilon},\frac{1}{T-\varepsilon}\right)x$$

$$\leq TV_{(\varepsilon,T-\varepsilon)}(f)\,x + \max\left(\frac{1}{\varepsilon}, \frac{1}{T-\varepsilon}\right)x \leq 3\max\left(\frac{1}{\varepsilon}, \frac{1}{T-\varepsilon}\right)x$$

by (5.2) and (5.1). Since the final upper bound converges to 0 as $x \to 0$ uniformly over the entire class A_T, we conclude by Theorem 20, p. 298 in [11] that the family of the densities of the laws in A_T is relatively compact in $L_1(\varepsilon, T-\varepsilon)$, for each $0 < \varepsilon < T/2$.

Next, let F_n, $n = 1, 2, \dots$ be a sequence of probability measures in A_T such that $F_n \Rightarrow F$ for some $F \in \mathcal{P}_T$. For $n \geq 1$ we denote by f_n the version of the density of F_n in the interior of the interval $[0, T]$ described in the definition of the class A_T. Let $0 < t < T$. For $0 < \varepsilon < \min(t, T-t)$ we have

$$F\big((t-\varepsilon, t+\varepsilon)\big) \leq \liminf_{n\to\infty} F_n\big((t-\varepsilon, t+\varepsilon)\big) \leq \int_{t-\varepsilon}^{t+\varepsilon} \max\left(\frac{1}{s}, \frac{1}{T-s}\right) ds\,.$$

This implies that F is absolutely continuous in the interior of the interval $[0, T]$ with a density f satisfying

$$f(t) \leq \max\left(\frac{1}{t}, \frac{1}{T-t}\right),\ 0 < t < T\,.$$

Since for every $0 < \varepsilon < T/2$ the sequence (f_n) is relatively compact in $L_1(\varepsilon, T-\varepsilon)$, we conclude that

$$f_n \to f \ \text{ in } L_1(\varepsilon, T-\varepsilon). \tag{5.9}$$

Fix once again $0 < \varepsilon < T/2$, and notice that, according to (5.9), there is a subsequence (f_{n_k}) with $n_k \to \infty$ such that

$$f_{n_k} \to f \ \text{ a.e. in } (\varepsilon, T-\varepsilon). \tag{5.10}$$

In the computations in the sequel we will identify, for typographical convenience, the subsequence (f_{n_k}) with the entire sequence (f_n). Let A_* be the set of $\varepsilon < t < T-\varepsilon$ of full measure for which the convergence in (5.10) takes place.

The next step is to show that for every $\varepsilon < t < T-\varepsilon$,

$$\lim_{s\downarrow t,\, s\in A_*} f(s) \text{ exists, and } \lim_{s\uparrow t,\, s\in A_*} f(s) \text{ exists.} \tag{5.11}$$

We will prove the first statement in (5.11); the second one is analogous. Suppose that, to the contrary, for some $\varepsilon < t < T-\varepsilon$ the limit from the right does not exist. Then there are sequences in A_*, $s_m \downarrow t$ and $v_m \downarrow t$, such that

$$b := \lim_{m\to\infty} f(s_m) > a := \lim_{m\to\infty} f(v_m)\,.$$

We may, of course, assume that $s_1 > v_1 > s_2 > v_2 > \ldots > t$. Let $\tau = b - a > 0$, and take M so large that

$$f(s_m) > b - \tau/6,\ f(v_m) < a + \tau/6 \text{ for all } m > M. \tag{5.12}$$

Choose K so large that

$$(2K-1)\tau > 6\max\left(\frac{1}{\varepsilon}, \frac{1}{T-\varepsilon}\right)\,,$$

and choose n so large that

$$\bigl|f_n(v_m) - f(v_m)\bigr| \le \tau/6,\ \bigl|f_n(s_m) - f(s_m)\bigr| \le \tau/6 \tag{5.13}$$

for each $m = M+1, \ldots, M+K$; this is possible to achieve since each s_m and each v_m is in the set A_*. It follows from (5.12) and (5.13) that

$$f_n(s_m) > b - \tau/3,\ f_n(v_m) < a + \tau/3 \text{ for each } m = M+1, \ldots, M+K,$$

so that

$$\sum_{m=M+1}^{m+K} \bigl|f_n(s_m) - f_n(v_m)\bigr| + \sum_{m=M+1}^{m+K-1} \bigl|f_n(v_m) - f_n(s_{m+1})\bigr| > (2K-1)\tau/3\,.$$

By the choice of K, however, this contradicts the total variation constraint (5.2) since, by (5.1),

$$\max\Bigl(f_n(s_{M+1}),\ f_n(v_{M+K})\Bigr) \le \max\left(\frac{1}{\varepsilon}, \frac{1}{T-\varepsilon}\right)\,.$$

Therefore, (5.11) holds.

Next, we show that the set

$$B_* = \Big\{ t \in A_* : \ f(t) \neq \lim_{s \downarrow t,\, s \in A_*} f(s) \Big\}$$

is, at most, countable, which will follow once we check that for any $\theta > 0$ the set

$$B_*(\theta) = \Big\{ t \in A_* : \ \Big| f(t) - \lim_{s \downarrow t,\, s \in A_*} f(s) \Big| > \theta \Big\}$$

is finite. Specifically, we will show that the cardinality of $B_*(\theta)$ does not exceed

$$\frac{6}{\theta} \max\left(\frac{1}{\varepsilon}, \frac{1}{T-\varepsilon} \right) .$$

Indeed, suppose that, to the contrary, there are points $\varepsilon < v_1 < v_2 < \ldots < v_K < T - \varepsilon$ in $B_*(\theta)$ for some

$$K > \frac{6}{\theta} \max\left(\frac{1}{\varepsilon}, \frac{1}{T-\varepsilon} \right) .$$

For each $m = 1, \ldots, K$ choose $s_m \in A_*$, $v_m < s_m < v_{m+1}$ (with $v_{K+1} = T - \varepsilon$) such that

$$\big| f(v_m) - f(s_m) \big| > \theta \, .$$

Finally, choose n so large that

$$\big| f_n(v_m) - f(v_m) \big| \leq \theta/3, \ \big| f_n(s_m) - f(s_m) \big| \leq \theta/3, \ m = 1, \ldots, K \, .$$

Then for every $m = 1, \ldots, K$ we have

$$\big| f_n(v_m) - f_n(s_m) \big| > \theta/3 \, ,$$

so that by the choice of K,

$$\sum_{m=1}^{K} \big| f_n(s_m) - f_n(v_m) \big| > 2 \max\left(\frac{1}{\varepsilon}, \frac{1}{T-\varepsilon} \right) .$$

Once again, this is incompatible with the combination of the total variation constraint (5.2) and the upper bound (5.1). The resulting contradiction proves that the set B_* is, at most, countable.

The standard diagonal argument now allows us to get rid of $\varepsilon > 0$ in the above conclusions: there is a subsequence (f_{n_k}) with $n_k \to \infty$ such that $f_{n_k}(t) \to f(t)$ for almost every $0 < t < T$, say, for $t \in A_*$. Furthermore, for every $0 < t < T$ (5.1) holds. Finally, the set B_* (defined now for the entire interval $(0, T)$) is at most countable. We are in a position to define now

$$g(t) = \lim_{s \downarrow t,\, s \in A_*} f(s), \ 0 < t < T\,. \tag{5.14}$$

The resulting function is automatically right continuous with left limits. Moreover, g coincides with f on $A_* \setminus B_*$, i.e. g is a version of f, hence a density of the limiting law F in the interior of the interval $[0, T]$. The right continuity of g shows that the right derivative of F exists at every point in $(0, T)$ and coincides with g at that point. By construction, g satisfies the total variation constraints (5.2), (5.3), (5.4), (5.5) and (5.6). This proves that A_T is weakly closed.

Finally, let $0 < \varepsilon < T/2$, and let (F_n) be a sequence in A_T. By the weak compactness of $\mathcal{P}_T$, we can choose a subsequence (F_{n_k}) with $n_k \to \infty$ weakly converging in $\mathcal{P}_T$ to some F; since we already know that A_T is weakly closed, $F \in A_T$. Let f be some version of the density of F in $(0, T)$. We have established in the course of the proof that the densities (f_{n_k}) of the laws (F_{n_k}) form a relatively compact family in $L_1(\varepsilon, T - \varepsilon)$. Since f can be the only limit point, we conclude that $f_{n_k} \to f$ in $L_1(\varepsilon, T - \varepsilon)$. This, of course, means that the restrictions of the laws (F_{n_k}) to the interval $(\varepsilon, T - \varepsilon)$ converge in total variation to the restriction of the law F to the same interval, so the last statement of the theorem has been proved. $\square$

Note that the set of finite signed measures on $[0, T] \cup \{\infty\}$ equipped with the topology of weak convergence is a locally convex topological vector space. According to Theorem 5.4.1, the set A_T is a compact convex subset of that space. By the Krein-Milman theorem, the set A_T is equal to the closed convex hull of its extreme points; see e.g. Theorem 4, p. 440 in [11]. Our next result describes the extreme points of the set A_T.

Theorem 5.4.2. *The extreme points of the set A_T are:*

(1) the measures μ_t, $t \in (0, T)$ concentrated on $(0, T)$, absolutely continuous with respect to the Lebesgue measure on $(0, T)$, with density functions $f_{\mu_t} = \frac{1}{t}\mathbf{1}_{(0,t)}$, $0 < t < T$;

(2) the measures ν_t, $t \in (0, T)$ concentrated on $(0, T)$, absolutely continuous with respect to the Lebesgue measure on $(0, T)$, with density functions $f_{\nu_t} = \frac{1}{T-t}\mathbf{1}_{(t,T)}$, $0 < t < T$;

(3) the point masses δ_0, δ_T and δ_∞.

Proof. Since any probability measure m in A_T admits a unique decomposition of the type $m = \alpha_1\delta_0 + \alpha_2\delta_T + \alpha_3\delta_\infty + \beta m_{AC}$, where $\alpha_1, \alpha_2, \alpha_3, \beta \geq 0$, $\alpha_1 + \alpha_2 + \alpha_3 + \beta = 1$, and m_{AC} is an absolutely continuous measure on $(0, T)$, it is enough to prove that the first two cases in the theorem describe all the extreme points of A_T that are concentrated on $(0, T)$ and are absolutely continuous there.

Let f be the density of such a measure as described in the definition of the class A_T. We start by showing that f must be monotone. To this end, define functions $f_1(t) = TV^+_{(0,t]}(f)$ and $f_2(t) = TV^-_{(t,T)}$, $t \in (0, T)$. By (5.3) and (5.4) these functions are well-defined and nonnegative. Moreover, f_1 is a nondecreasing càdlàg function with $f_1(0+) = 0$, while f_2 is a nonincreasing càdlàg function with $f_2(T-) = 0$. It also follows from (5.3) and (5.4) that $f(t) \geq \max\bigl(f_1(t), f_2(t)\bigr)$ for $0 < t < T$.

Choose $0 < t_1 < T$, and note that for every $t_1 < t < T$,

$$f(t) = f(t_1) + TV^+_{(t_1,t]}(f) - TV^-_{(t_1,t]}(f)\,,$$

while

$$f_1(t) = f_1(t_1) + TV^+_{(t_1,t]}(f),\ \ f_2(t) = f_2(t_1) - TV^-_{(t_1,t]}(f)\,.$$

Therefore,

$$f(t) = f_1(t) + f_2(t) + \bigl(f(t_1) - f_1(t_1) - f_2(t_1)\bigr) := f_1(t) + f_2(t) + C(t_1)\,.$$

From here we immediately conclude that $C(t_1)$ is independent of t_1 and, hence, is equal to some constant C. Since $C \geq -f_1(t)$ for any $0 < t < T$, we can let $t \to 0$ to conclude that $C \geq 0$, so we have $f = f_1 + f_2'$, where $f_2' = f_2 + C$. If f is not monotone, then both $\int_0^T f_1(s)ds > 0$ and $\int_0^T f_2'(s)ds > 0$. Hence

$$f(t) = \int_0^T f_1(s)ds \cdot \frac{f_1(t)}{\int_0^T f_1(s)ds} + \int_0^T f_2'(s)ds \cdot \frac{f_2'(t)}{\int_0^T f_2'(s)ds},\ 0 < t < T\,,$$

a convex combination of two monotone densities, which are automatically densities of some laws in A_T. That is, the law corresponding to such f cannot be an extreme point of A_T.

Therefore, the density f must be monotone. Suppose that there are points t_1, t_2 in $(0,T)$ such that $f(t_1) = a_1$, $f(t_2) = a_2$ for some $0 < a_1 < a_2$. Let $f_1(t) = \max\bigl(f(t) - a_1,\ 0\bigr)$ and $f_2(t) = f(t) - f_1(t)$, $0 < t < T$. Since f is monotone, so are both f_1 and f_2. Once again, this allows us to represent

$$f(t) = \int_0^T f_1(s)ds \cdot \frac{f_1(t)}{\int_0^T f_1(s)ds} + \int_0^T f_2(s)ds \cdot \frac{f_2(t)}{\int_0^T f_2(s)ds},\ 0 < t < T\,,$$

showing that the law corresponding to such f cannot be an extreme point of A_T.

Therefore, the density f can take at most one non-zero value. In order to conclude that it must be of the form f_{μ_t} or f_{ν_t} described in (1) or (2) in the theorem, we only need to observe that the only remaining possibility, $f \equiv 1$, does not correspond to an extreme point of A_T since this constant density can be written in the form $f_{\mu_{T/2}}/2 + f_{\nu_{T/2}}/2$.

It remains to prove that for each $0 < t < T$, the densities f_{μ_t} and f_{ν_t} do correspond to extreme points of A_T. We will consider f_{μ_t}; the argument for f_{ν_t} is similar. Suppose that there are two different laws in A_T that are concentrated on $(0, T)$, with the corresponding densities g_1 and g_2, as described in the definition of the class A_T, such that

$$f_{\mu_t}(s) = pg_1(s) + (1-p)g_2(s), \ 0 < s < T \, , \tag{5.15}$$

for some $0 < p < 1$. There must be a point $0 < s_i < t$ such that $g_i(s_i) > 1/t$, $i = 1, 2$. Since $g_i(t) = 0$, $i = 1, 2$, the total variation requirement forces

$$g_i(0-) \geq g_i(s_i) > \frac{1}{t}, \ i = 1, 2 \, ,$$

so that

$$pg_1(0-) + (1-p)g_2(0-) > \frac{1}{t} \, .$$

This means that (5.15) is violated in a neighbourhood of the left endpoint. This contradiction completes the proof. □

Knowing the set of all extreme points of the set A_T allows us to obtain universal bounds on the expectation of functions of intrinsic location functionals.

Corollary 5.4.3. *Let g be a bounded, or nonnegative, measurable function on* $[0, T] \cup$

$\{\infty\}$. *Then for any stationary process* $\mathbf{X}$ *and intrinsic location functional* L,

$$\begin{aligned}
& \min\Big\{g(0), g(T), g(\infty), \inf_{t\in(0,T)} \tfrac{1}{t}\int_0^t g(s)ds, \inf_{t\in(0,T)} \tfrac{1}{T-t}\int_t^T g(s)ds\Big\} \\
\leq\ & \mathbb{E}\big[g\big(L(\mathbf{X},[0,T])\big)\big] \\
\leq\ & \max\Big\{g(0), g(T), g(\infty), \sup_{t\in(0,T)} \tfrac{1}{t}\int_0^t g(s)ds, \sup_{t\in(0,T)} \tfrac{1}{T-t}\int_t^T g(s)ds\Big\}.
\end{aligned}$$

The bounds obtained in Corollary 5.4.3 can sometimes be improved if one is interested only in certain subsets of all intrinsic location functionals. We describe now one such situation.

We call an intrinsic location functional $L : H \times \mathcal{I} \to \mathbb{R} \cup \{\infty\}$ *an earliest occurrence intrinsic location functional* if it has the following property: for every $a < b < c$ and $f \in H$,

$$\text{if } L(f,[a,b]) \in [a,b] \text{ then } L(f,[a,c]) = L(f,[a,b]).$$

The first hitting time $T^l_{f,[a,b]}$ of Example 5.2.2 is, clearly, an earliest occurrence intrinsic location functional.

Proposition 5.4.4. *For every* $T > 0$ *the distribution of* $L(\mathbf{X}, T)$ *for any earliest occurrence intrinsic location functional* L *and any stationary process* $\mathbf{X}$ *belongs to the set* A^{e}_T *consisting of all laws in* A_T *that do not put any mass at the right endpoint of the interval, and whose density in* $(0,T)$ *is nonincreasing. This set is weakly closed in* $\mathcal{P}_T$*, and its extreme points are the point masses* δ_0 *and* δ_∞*, as well as the measures* μ_t, $t \in (0,T]$*, concentrated on* $(0,T)$*, absolutely continuous with respect to the Lebesgue measure on* $(0,T)$*, with density functions* $f_{\mu_t} = \frac{1}{t}\mathbf{1}_{(0,t)}$, $0 < t \leq T$.

Remark 5.4.5. Note that, while some of the extreme points of A_T are no longer in A^{e}_T, the latter subset of A_T does have one extreme point that is not an extreme point of A_T, specifically the measure μ_T.

Proof of Proposition 5.4.4. Let $0 < t_1 < t_2 < T$, and take $0 < \varepsilon < t_1$. Using successively the stability under restrictions, the earliest occurrence property, and the shift compatibility, together with the stationarity of $\mathbf{X}$, we have

$$P\big(L(\mathbf{X},T) \in (t_2-\varepsilon, t_2+\varepsilon)\big) \leq P\big(L(\mathbf{X},[t_2-t_1,T]) \in (t_2-\varepsilon, t_2+\varepsilon)\big)$$

$$\leq P\big(L(\mathbf{X},[t_2-t_1,T+t_2-t_1]) \in (t_2-\varepsilon, t_2+\varepsilon)\big) = P\big(L(\mathbf{X},T) \in (t_1-\varepsilon, t_1+\varepsilon)\big)\,.$$

If $f_{\mathbf{X},T}$ is the version of the density described in the definition of the class A_T, we see that $f_{\mathbf{X},T}(t_2) \leq f_{\mathbf{X},T}(t_1)$, so the density must be nonincreasing. Similarly,

$$P\big(L(\mathbf{X},T) = T\big) \leq P\big(L(\mathbf{X},2T) = T\big) = 0$$

because laws in A_{2T} cannot have a mass in the interior of an interval. Therefore, no mass at the right endpoint of the interval is possible.

To see that A_T^{e} is weakly closed, note that by Theorem 5.4.1, any weakly convergent sequence in A_T^{e} has its limit in A_T. Since by the proof of Theorem 5.4.1 pointwise convergence of densities takes place in $(0,T)$ apart from a set of Lebesgue measure 0, and the limiting density is right continuous, the limiting density must be nonincreasing. Additionally, the density f of every law F in A_T^{e} satisfies $f(t) \leq 2/T$ for every $T/2 \leq t < T$ by monotonicity, so that

$$F\big((T-\varepsilon,T]\big)\big) \leq 2\varepsilon/T,\ 0 < \varepsilon < T/2\,,$$

and the weak limit of a sequence in A_T^{e} has the same property, possibly apart from a countable set of ε. Letting $\varepsilon \to 0$, while keeping away from the exceptional set, shows that the weak limit does not put any mass at T and, hence, is in A_T^{e}.

It remains to describe the extreme points of A_T^{e} and, as in Theorem 5.4.2, the only non-trivial case is that of the extreme points of A_T^{e} that are concentrated on

$(0, T)$ and are absolutely continuous there. The same argument as in the proof of Theorem 5.4.2 shows for any such extreme point the density can take at most one non-zero value, so it has to be of the form f_{μ_t}, $0 < t \le T$. For $t < T$ the latter laws are extreme points of A_T, hence of A_T^{e} as well. To see that the same is true for f_{μ_T} suppose, to the contrary, that there are two different laws in A_T^{e} that are concentrated on $(0, T)$, with the corresponding densities g_1 and g_2, such that

$$pg_1(s) + (1-p)g_2(s) = \frac{1}{T},\ 0 < s < T\,, \tag{5.16}$$

for some $0 < p < 1$. Once again, there are points $0 < s_i < T$ such that $g_i(s_i) > 1/T$, $i = 1, 2$, and the monotonicity of g_1 and g_2 forces

$$g_i(0-) \ge g_i(s_i) > \frac{1}{T},\ i = 1, 2\,.$$

Therefore, (5.16) is violated in a neighbourhood of the left endpoint of the interval. $\square$

Proposition 5.4.4 immediately implies the following counterpart of Corollary 5.4.3.

Corollary 5.4.6. *Let g be a bounded, or nonnegative, measurable function on $[0, T] \cup \{\infty\}$. Then for any stationary process $\mathbf{X}$ and earliest occurrence intrinsic location functional L,*

$$\begin{aligned} & \min\Big\{g(0), g(\infty), \inf_{t\in(0,T)} \tfrac{1}{t}\textstyle\int_0^t g(s)ds\Big\} \\ \le\ & \mathbb{E}\big[g\big(L(\mathbf{X}, [0, T])\big)\big] \\ \le\ & \max\Big\{g(0), g(\infty), \sup_{t\in(0,T)} \tfrac{1}{t}\textstyle\int_0^t g(s)ds\Big\}. \end{aligned}$$

Remark 5.4.7. The class A_T is the smallest class containing all possible distributions of $L(\mathbf{X}, T)$ for any intrinsic location functional L and any stationary process $\mathbf{X}$, while the class A_T^{e} is the smallest class containing all possible distributions of $L(\mathbf{X}, T)$ for any earliest occurrence intrinsic location functional L

and any stationary process $\mathbf{X}$, as easy examples show. In particular, the bounds obtained in Corollaries 5.4.3 and 5.4.6 are the tightest bounds possible.

The proposition below presents one application of the bounds given in corollaries 5.4.3 and 5.4.6.

Proposition 5.4.8. *For any stationary process* $\mathbf{X}$*, intrinsic location functional* L*,* $T > 0$ *and* $0 < c < d < T$*,*

$$P(L(\mathbf{X}, T) \in [c, d]) \leq \frac{d - c}{\min(T - c, d)}. \tag{5.17}$$

If the functional is an earliest occurrence intrinsic location functional, then

$$P(L(\mathbf{X}, T) \in [c, d]) \leq \frac{d - c}{d}. \tag{5.18}$$

Proof. One simply uses the upper bounds in corollaries 5.4.3 and 5.4.6 with the function $g = \mathbf{1}_{[c,d]}$. □

Remark 5.4.9. It is interesting that the upper bounds in the proposition are optimal even for very specific intrinsic location functionals. For example, it follows from the results in Chapter 4 that the upper bound in (5.17) is optimal for the leftmost location of the supremum $\tau_{f,[a,b]}$ of Example 5.2.1.

On the other hand, consider the first hitting time $T^{l}_{f,[a,b]}$ of Example 5.2.2. For the continuous stationary periodic process $\mathbf{X}(t) = \sin(t\pi/d + U) + l$, $t \in \mathbb{R}$, with U uniformly distributed on $[0, 2\pi]$, the first hitting time is uniformly distributed between 0 and d and, hence, achieves equality in (5.18).

For certain intrinsic location functionals L and certain stationary processes $\mathbf{X}$ the law of $L(\mathbf{X}, T)$ is symmetric around the mid-point of the interval $[0, T]$, i.e.

$$P\big(L(\mathbf{X}, T) \in B\big) = P\big(L(\mathbf{X}, T) \in T - B\big) \tag{5.19}$$

for any Borel subset B of $[0, T/2)$. This happens, for example, when the process $\mathbf{X}$ is time reversible, i.e. if $(X(-t),\ t \in \mathbb{R}) \overset{d}{=} (X(t),\ t \in \mathbb{R})$, while the functional L has a certain uniqueness property associated with it. The quintessential example of such a situation is the leftmost location of the supremum of Example 5.2.1 evaluated at a continuous stationary Gaussian process. Such a process is always time reversible and, as long as $X(t) \neq X(0)$ a.s. for $0 < t \leq T$, the supremum is achieved, with probability 1, at a unique point of the interval $[0, T]$, so that (5.19) holds; see Lemma 2.6 in [13].

We denote by A_T^{S} the set of all symmetric laws in A_T, i.e. the set of all laws in A_T satisfying (5.19). The following theorem describes the structure of this set of probability laws. Its proof is similar to that of Theorem 5.4.2 and is omitted.

Theorem 5.4.10. *The set A_T^{S} is a weakly closed convex subset of $\mathcal{P}_T$. The extreme points of this set are:*

(1) the measures ρ_t, $t \in (0, T/2)$ concentrated on $(0, T)$, absolutely continuous with respect to the Lebesgue measure on $(0, T)$, with density functions

$$f_{\rho_t}(s) = \begin{cases} \frac{1}{2t} & 0 < s < t \\ \frac{1}{2t} & T - t \leq s < T \\ 0 & \textit{otherwise} \end{cases} ;$$

(2) the measures ξ_t, $t \in (0, T/2)$ concentrated on $(0, T)$, absolutely continuous with respect to the Lebesgue measure on $(0, T)$, with density functions

$$f_{\xi_t}(s) = \begin{cases} \frac{1}{2(T-t)} & 0 < s < t \\ \frac{1}{T-t} & t \leq s < T - t \\ \frac{1}{2(T-t)} & T - t \leq s < T \end{cases} ;$$

(3) the discrete measures $(\delta_0 + \delta_T)/2$ and δ_∞.

Remark 5.4.11. Interestingly, the uniform distribution on $(0,T)$, $\rho_{T/2} = \xi_{T/2}$, is not an extreme point of A_T^{S}, because for every $0 < t < T/2$, the mixture

$$\frac{t}{T}\rho_t + \frac{T-t}{T}\xi_t$$

coincides with the uniform distribution.

The following corollary is a counterpart of Corollary 5.4.3 to the symmetric case. We restrict ourselves (without loss of generality) to symmetric functions, i.e. functions satisfying $g(T/2-t) = g(T/2+t)$, $0 \le t \le T/2$.

Corollary 5.4.12. *Let g be a bounded, or nonnegative, measurable symmetric function on $[0,T] \cup \{\infty\}$. Then for any stationary process $\mathbf{X}$ and intrinsic location functional L, satisfying the symmetry assumption* (5.19)*,*

$$\begin{aligned}
& \min\Big\{ g(0), g(\infty), \inf\nolimits_{t\in(0,T/2)} \tfrac{1}{t}\textstyle\int_0^t g(s)ds, \\
& \qquad \inf\nolimits_{t\in(0,T/2)} \Big[\tfrac{1}{T-t}\textstyle\int_0^t g(s)ds + \tfrac{2}{T-t}\textstyle\int_t^{T/2} g(s)ds \Big] \Big\} \\
\le \;\; & \mathbb{E}\big[g\big(L(\mathbf{X},[0,T])\big)\big] \\
\le \;\; & \max\Big\{ g(0), g(\infty), \sup\nolimits_{t\in(0,T/2)} \tfrac{1}{t}\textstyle\int_0^t g(s)ds, \\
& \qquad \sup\nolimits_{t\in(0,T/2)} \Big[\tfrac{1}{T-t}\textstyle\int_0^t g(s)ds + \tfrac{2}{T-t}\textstyle\int_t^{T/2} g(s)ds \Big] \Big\}.
\end{aligned}$$

Corollary 5.4.12 implies the following upper bounds on the mass the law of an intrinsic location functional assigns, in the symmetric case, to any subinterval of $(0,T)$.

Proposition 5.4.13. *For any stationary process $\mathbf{X}$ and intrinsic location functional L,* *satisfying the symmetry assumption* (5.19)*, and* $0 < c < d < T$*,*

$$P(L(\mathbf{X},T) \in [c,d]) \le \begin{cases} \frac{d-c}{2d} & c+2d \le T \\ \frac{d-c}{T-c} & c+2d > T,\, c+d \le T,\, 2d-c \le T \\ \frac{3d-c-T}{2d} & c+d \le T,\, 2d-c > T \\ \frac{T+d-3c}{2(T-c)} & c+d > T,\, d > 2c \\ \frac{d-c}{d} & c+d > T,\, d \le 2c,\, 2c+d \le 2T \\ \frac{d-c}{2(T-c)} & 2c+d > 2T \end{cases}.$$

Proof. One uses the upper bounds in Corollary 5.4.12 with the symmetrized indicator function $g = \big(\mathbf{1}_{[c,d]} + \mathbf{1}_{[T-d,T-c]}\big)/2$, and straightforward optimization over $t \in (0, T/2)$. □

Remark 5.4.14. Once again, the upper bounds we have obtained are optimal even for the leftmost location of the supremum $\tau_{f,[a,b]}$ of Example 5.2.1, when the supremum is unique, and the stationary process is reversible. This follows from the results in Chapter 4.

5.5 Characterizing stationarity

Much of the previous discussion in this chapter centered around the basic property of the intrinsic location functionals of stationary processes evaluated on some interval: the fact that their law must be absolutely continuous in the interior of the interval, and have a density satisfying the total variation constraints described in Theorem 5.3.1. The nature of this fact is itself interesting and, intuitively at least, intimately related to the stationarity of the underlying process: an intrinsic location functional "is shifted together with the process". Since the latter is stationary, one expects a shift to have only a limited effect on the law of the functional, hence its density does not change much from point to point. For certain intrinsic location functionals one can even be forgiven for believing that the density has to be constant; this is, for instance, the situation with the leftmost location of the supremum $\tau_{f,[a,b]}$ of Example 5.2.1, when the supremum is unique. We know that the density does not need to be constant, but the total variation constraints on the density may be viewed as restricting how different from a constant can the density be.

In this section we make this intuition precise. It turns out that existence of a density satisfying the total variation constraints for each appropriate intrinsic location functional (or even only those in a certain subclass of intrinsic location functionals) requires stationarity of the stochastic process. The theorem below is formulated for the processes with continuous sample paths and, correspondingly, to intrinsic location functionals on the space $H = C(\mathbb{R})$. Note, however, that the proof of the fact that (2) implies (1) in that theorem is valid for any space H for which the functional defined in (5.20) is measurable. This is the case, for instance for the space $H = D(\mathbb{R})$, the space of all càdlàg functions. In Section 5.6 we also extend the fact that (3) implies (1) to other spaces, in particular to the space $H = D(\mathbb{R})$. The following theorem is the main result of this section.

Theorem 5.5.1. *Let $\mathbf{X}$ be a stochastic process with continuous sample paths. The following statements are equivalent.*

1. *The process $\mathbf{X}$ is stationary.*
2. *For some (equivalently, any) $\Delta > 0$, any intrinsic location functional $L : C(\mathbb{R}) \times \mathcal{I} \to \mathbb{R} \cup \{\infty\}$, the law of $L(\mathbf{X}, I) - a$, $I = [a, a + \Delta] \in \mathcal{I}$, does not depend on a.*
3. *For any intrinsic location functional $L : C(\mathbb{R}) \times \mathcal{I} \to \mathbb{R} \cup \{\infty\}$, any interval $I = [a, b] \in \mathcal{I}$, the law of $L(\mathbf{X}, I)$ is absolutely continuous on (a, b) and has a*

density satisfying the total variation constraints.

Proof. The fact that (1) implies (2) is obvious, while the fact that (1) implies (3) follows from the discussion in Section 5.3.

To see that (2) implies (1), let $\Delta > 0$. Take any $n = 1, 2, ...$, time points $0 < t_1 < ... < t_n$ and closed intervals $I_1, \ldots, I_n$. Then

$$L(f, I) := \inf\{t \in I : X(t + t_i) \in I_i, \ i = 1, ..., n\}, \ I \in \mathcal{I}, \tag{5.20}$$

is, clearly, an intrinsic location functional on $C(\mathbb{R})$. Furthermore, for any real $a \in \mathbb{R}$,

$$P\big(X(a + t_i) \in I_i, \ i = 1, ..., n\big) = P\big(L(\mathbf{X}, [a, a + \Delta]) = a\big)$$
$$= P\big(L(\mathbf{X}, [a, a + \Delta]) - a = 0\big),$$

which is independent of a by the assumption. We conclude that

$$\big(X(a + t_i), \ i = 1, ..., n\big) \overset{d}{=} \big(X(t_i), \ i = 1, ..., n\big)$$

for all real a. Since this is true for all $n = 1, 2, \ldots$ and $0 < t_1 < ... < t_n$, the process $\mathbf{X}$ is stationary.

In the remainder of the proof we show that (3) implies (1). Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function. For $n = 1, 2, ...$, $h \in \mathbb{R}$, $d \geq 0$, $\mathbf{t} = (t_1, ..., t_n)$ such that $0 < t_1 < ... < t_n$ and a collection of open intervals $\mathbf{I} = \big(I_1, ..., I_n\big)$, define a set of points by

$$A^{h,d}_{\mathbf{t},\mathbf{I}}(f) = \{s \in \mathbb{R} : f(s) = h, \inf\{r > s : f(r) = h\} > s + d, \tag{5.21}$$
$$f(s + t_i) \in I_i, \ i = 1, ..., n\}.$$

We start with recording a simple fact.

Lemma 5.5.1. *Let $\mathbf{X}$ be a continuous stochastic process. If (3) is satisfied, then for any $h, \mathbf{t}, \mathbf{I}$, and $d > 0$,*

$$\mathbb{P}\big(a \in A_{\mathbf{t},\mathbf{I}}^{h,d}(\mathbf{X})\big) = 0$$

for any $a \in \mathbb{R}$.

Proof. The functional $L(f, [a, b]) = \inf\big(A_{\mathbf{t},\mathbf{I}}^{h,d}(f) \cap [a, b]\big)$ is easily seen to be an intrinsic location functional on $C(\mathbb{R})$. Since by the definition, if $a \in A_{\mathbf{t},\mathbf{I}}^{h,d}(f)$, then

$$A_{\mathbf{t},\mathbf{I}}^{h,d}(f) \cap (a, a + d] = \emptyset\,,$$

we obtain

$$\mathbb{P}\big(a \in A_{\mathbf{t},\mathbf{I}}^{h,d}(\mathbf{X})\big) \leq \mathbb{P}\big(L(\mathbf{X}, [a - d, a + d] = a\big) = 0$$

by the absolute continuity property in (3). □

The following lemma shows a feature of the random sets $A_{\mathbf{t},\mathbf{I}}^{h,d}(\mathbf{X})$ implied by the total variation property; note that *if we knew that the process $\mathbf{X}$ was stationary*, its statement would follow from the ergodic decomposition.

Lemma 5.5.2. *Let $\mathbf{X}$ be a continuous stochastic process. If (3) is satisfied, then for any $h, d, \mathbf{t}$ and $\mathbf{I}$, with probability 1, $A_{\mathbf{t},\mathbf{I}}^{h,d}(\mathbf{X})$ is either the empty set or both $\inf\big(A_{\mathbf{t},\mathbf{I}}^{h,d}(\mathbf{X})\big) = -\infty$ and $\sup\big(A_{\mathbf{t},\mathbf{I}}^{h,d}(\mathbf{X})\big) = \infty$.*

Proof of Lemma 5.5.2. It is easy to check that the collections of outcomes we are discussing are measurable. Suppose, for example, that, to the contrary, there exist $h, d, \mathbf{t}$ and $\mathbf{I}$ such that, on event of positive probability, $-\infty < \inf\big(A_{\mathbf{t},\mathbf{I}}^{h,d}(\mathbf{X})\big) < \infty$; the supremum can be dealt with similarly. The functional $L(f, [a, b]) = \inf\big(A_{\mathbf{t},\mathbf{I}}^{h,d}(f) \cap [a, b]\big)$ is, again, an intrinsic location functional on $C(\mathbb{R})$. Choose an interval $[a_1, b_1]$, such that $\mathbb{P}\big[\inf\big(A_{\mathbf{t},\mathbf{I}}^{h,d}(\mathbf{X})\big) \in [a_1, b_1]\big] =: c > 0$. For any

$a < a_1$ and $b > b_1$, $\inf\bigl(A^{h,d}_{\mathbf{t},\mathbf{I}}(\mathbf{X})\bigr) \in [a_1, b_1]$ implies $L(\mathbf{X}, [a,b]) = \inf\bigl(A^{h,d}_{\mathbf{t},\mathbf{I}}(\mathbf{X})\bigr)$, so $\mathbb{P}(L(\mathbf{X}, [a,b]) \in [a_1, b_1]) \geq c$. However, the upper bounds (3.1) on the density (following from the assumption (3)) require that probability to converge to zero as $a \to \infty$ and $b \to \infty$. The resulting contradiction proves the lemma. $\square$

For any $h \in \mathbb{R}$ the set

$$\mathcal{C}_h = \bigl\{ f \in C(\mathbb{R}) : \inf\bigl(A^{h,0}(f)\bigr) = -\infty, \ \sup\bigl(A^{h,0}(f)\bigr) = \infty \bigr\}$$

(meaning that the vector $\mathbf{t}$ is empty) is a cylindrical set. Let $h_i, i = 1, 2, \ldots$ be an enumeration of the rationals in $\mathbb{R}$, and construct inductively a subsequence $h_{i_j}, j = 1, 2, \ldots$ according to the rule $i_1 = \inf\bigl\{ i \geq 1 : \mathbb{P}\bigl(\mathbf{X} \in \mathcal{C}_{h_i}\bigr) > 0 \bigr\}$, while for $j \geq 2$,

$$i_j = \inf\Bigl\{ i > i_{j-1} : \mathbb{P}\Bigl(\mathbf{X} \in \mathcal{C}_{h_i} \setminus \bigl(\cup_{k=0}^{j-1} \mathcal{C}_{h_{i_k}}\bigr)\Bigr) > 0 \Bigr\}.$$

Let $\mathcal{C}'_1 = \mathcal{C}_{h_{i_1}}$, $\mathcal{C}'_j = \mathcal{C}_{h_{i_j}} \setminus \bigl(\cup_{k=0}^{j-1} \mathcal{C}_{h_{i_k}}\bigr)$, $j \geq 2$. If the process $\mathbf{X}$ is, with positive probability, a constant process, we also define $\mathcal{C}'_0$ to be the collection of constant functions in $C(\mathbb{R})$. Then the sets $\bigl(\mathcal{C}'_j\bigr)$ are disjoint, $\mathbb{P}\bigl(\mathbf{X} \in \mathcal{C}'_j\bigr) > 0$ for each j, while by Lemma 5.5.2, $\mathbb{P}\bigl(\mathbf{X} \notin \cup_j \mathcal{C}'_j\bigr) = 0$. Let $\mathbf{X}_j$ be a continuous stochastic process whose law is the conditional law of $\mathbf{X}$ given $\mathbf{X} \in \mathcal{C}'_j$. Note that, if each $\mathbf{X}_j$ is a stationary process, then so is $\mathbf{X}$ itself, as a mixture of stationary processes (note that each set $\mathcal{C}'_j$ is shift invariant). Since a constant process is, obviously, stationary, we only need to establish stationarity of each process $\mathbf{X}_j$, $j \geq 1$. We also claim that the statement (3) of the theorem is satisfied for each one of these processes. To see that, let L be any intrinsic location functional on $C(\mathbb{R})$. Define

$$L_j(f, [a,b]) = \begin{cases} L(f, [a,b]) & \text{if } f \in \mathcal{C}'_j \\ \infty & \text{if } f \notin \mathcal{C}'_j \end{cases}.$$

Then L_j is also an intrinsic location functional on $C(\mathbb{R})$. Further, for any $a < c <$

$d < b$ we have

$$\mathbb{P}\big(L(\mathbf{X}_j, [a,b]) \in [c,d]\big) = \frac{1}{\mathbb{P}\big(\mathbf{X} \in \mathcal{C}_j'\big)} \mathbb{P}\big(L_j(\mathbf{X}, [a,b]) \in [c,d]\big)\,.$$

Therefore, the restriction of the law of $L(\mathbf{X}_j, [a,b])$ to the interior of the interval differs only by a multiplicative constant from the restriction of the law of $L_j(\mathbf{X}, [a,b])$ to the interior of that interval. It follows that the statement (3) of the theorem is satisfied for the process $\mathbf{X}_j$.

In the remainder of the proof, therefore, we will establish stationarity of the process $\mathbf{X}_j$, $j \geq 1$. For notational convenience we will still call it $\mathbf{X}$, with the understanding that the process has its sample paths in $\mathcal{C}_h$ for some $h \in \mathbb{R}$. Fixing such h, we denote for $a \in \mathbb{R}$ and $\Delta > 0$,

$$p^{h,d}_{\mathbf{t},\mathbf{I},a,\Delta}(\mathbf{X}) = \mathbb{P}\big(A^{h,d}_{\mathbf{t},\mathbf{I}}(\mathbf{X}) \cap [a, a+\Delta] \neq \emptyset\big). \tag{5.22}$$

The proof of the theorem will be completed by the next two lemmas.

Lemma 5.5.3. *If for any $\Delta > 0$, $d \geq 2\Delta$, $\mathbf{t}$ and $\mathbf{I}$, the probability $p^{h,d}_{\mathbf{t},\mathbf{I},a,\Delta}(\mathbf{X})$ is independent of a, then the process $\mathbf{X}$ is stationary.*

Proof. We note that, by Lemma 5.5.1, the probability $p^{h,d}_{\mathbf{t},\mathbf{I},a,\Delta}(\mathbf{X})$ does not change if either or both of the endpoints of the interval $[a, a+\Delta]$ are removed.

Fix $\mathbf{t} = (t_1, ..., t_n)$ and intervals $\mathbf{I} = \big(I_1, ..., I_n\big)$ such that both

$$P\big(X(t_j) \in \partial(I_j)\big) = 0 \ \text{ and } h \notin \bar{I}_j, \ j = 1, \ldots, n\,. \tag{5.23}$$

For $m = 1, 2, ...$, let $a_j^m = t_1 - (j+1)2^{-m}$ be the left endpoint of the interval $T_j^m = [t_1 - (j+1)2^{-m}, t_1 - j2^{-m})$, $j = 0, 1, ...$, of the length $\Delta_m = 2^{-m}$. Consider the sum

$$S_m(\mathbf{t}, \mathbf{I}) := \sum_{j=0}^{\infty} p^{h,(j+2)\Delta_m}_{\mathbf{t}-a_j^m,\mathbf{I},a_j^m,\Delta_m}(\mathbf{X}),$$

with $\mathbf{t} - a_j^m = (t_1 - a_j^m, ..., t_n - a_j^m)$. Notice that the values of $d = (j+2)\Delta_m$ are chosen in such a way that this sum is the sum of probabilities of disjoint events. Moreover,

$$\begin{aligned}&\Big\{A^{h,(j+2)\Delta_m}_{\mathbf{t}-a_j^m,\mathbf{I}}(\mathbf{X}) \cap T_j^m \neq \emptyset \text{ for some } j\Big\}\\ &\subseteq \Big\{X(t_i+\delta) \in I_i, i = 1, \ldots, n, \text{ for some } \delta \in [0, \Delta_m].\Big\}.\end{aligned}$$

That is,

$$S_m(\mathbf{t}, \mathbf{I}) \leq \mathbb{P}\Big(X(t_i+\delta) \in I_i, i = 1, \ldots, n, \text{ for some } \delta \in [0, \Delta_m].\Big).$$

Taking limit on both sides gives us

$$\begin{aligned}\limsup_{m\to\infty} S_m(\mathbf{t}, \mathbf{I}) &\leq \mathbb{P}\big(X(t_i) \in \bar{I}_i,\ i = 1, 2, \ldots, n\big) \qquad (5.24)\\ &= \mathbb{P}\big(X(t_i) \in I_i,\ i = 1, 2, \ldots, n\big)\end{aligned}$$

by (5.23).

On the other hand, consider the event

$$A = \big\{X(t_i) \in I_i,\ i = 1, 2, \ldots, n\big\}.$$

On this event we can define three random variables as follows. First, let $l = l(\omega) := \sup\{s < t_1 : X(s) = h, \inf\{t > s : X(t) = h\} > s\}$ and $r = r(\omega) := \inf\{s > t_1 : X(s) = h, \inf\{t > s : X(t) = h\} > s\}$. Next, let

$$\epsilon_0 = \epsilon_0(\omega) := \inf\big\{\theta > 0 :\ X(t_i+\theta) \in \partial(I_i), \text{ some } i = 1, \ldots, n.\big\}$$

For $\omega \in A$ take any m satisfying $\Delta_m < \min\{\epsilon_0, \frac{r-t_1}{2}\}$, and let $j = j(m, \omega)$ be such that $l \in T_j^m$. It follows from (5.23) that l and r are consecutive points in the set $\{s : X(s) = h, \inf\{t > s : X(t) = h\} > s\}$. Therefore, for some $M(\omega) < \infty$, for all $m > M(\omega)$,

$$l(\omega) \in A^{h,(j(m,\omega)+2)\Delta_m}_{\mathbf{t}-a^m_{j(m,\omega)},\mathbf{I}} \cap T^m_{j(m,\omega)}.$$

We conclude that

$$A \subseteq \liminf_{m\to\infty} \bigcup_{j=0}^{\infty} \Big\{ A^{h,(j+2)\Delta_m}_{\mathbf{t}-a_j^m,\mathbf{I}}(\mathbf{X}) \cap T_j^m \neq \emptyset \Big\}$$

and, hence,

$$\begin{aligned} & \mathbb{P}\big(X(t_i) \in I_i,\ i = 1,2,\ldots,n\big) = \mathbb{P}(A) \\ \leq\ & \liminf_{m\to\infty} \mathbb{P}\left(\bigcup_{j=0}^{\infty} \Big\{ A^{h,(j+2)\Delta_m}_{\mathbf{t}-a_j^m,\mathbf{I}}(\mathbf{X}) \cap T_j^m \neq \emptyset \Big\} \right) \\ =\ & \liminf_{m\to\infty} S_m(\mathbf{t},\mathbf{I})\,. \end{aligned}$$

Together with (5.24) this proves that

$$\mathbb{P}\big(X(t_i) \in I_i,\ i = 1,2,\ldots,n\big) = \lim_{m\to\infty} S_m(\mathbf{t},\mathbf{I})\,. \tag{5.25}$$

Let now $u \in \mathbb{R}$, and impose an extra assumption on the intervals:

$$P\big(X(t_j + u) \in \partial(I_j)\big) = 0,\ j = 1,\ldots,n\,. \tag{5.26}$$

Then (5.25) implies that

$$\mathbb{P}\big(X(t_i + u) \in I_i,\ i = 1,2,\ldots,n\big) = \lim_{m\to\infty} S_m(\mathbf{t}+u,\mathbf{I})\,.$$

However, by the assumptions of the lemma,

$$S_m(\mathbf{t}+u,\mathbf{I}) = \sum_{j=0}^{\infty} p^{h,(j+2)\Delta_m}_{\mathbf{t}-a_j^m,\mathbf{I},a_j^m+u,\Delta_m}(\mathbf{X}) = S_m(\mathbf{t},\mathbf{I})\,,$$

$m = 1,2,\ldots$. Therefore

$$\mathbb{P}\big(X(t_i + u) \in I_i,\ i = 1,2,\ldots,n\big) = \mathbb{P}\big(X(t_i) \in I_i,\ i = 1,2,\ldots,n\big)$$

for all open intervals $I_1,\ldots,I_n$ satisfying (5.23) and (5.26). This implies that

$$\big(X(t_1+u),\ldots,X(t_n+u)\big) \overset{d}{=} \big(X(t_1),\ldots,X(t_n)\big)\,.$$

Since this is true for all $n \geq 1$, $t_1 < \ldots < t_n$ and $u \in \mathbb{R}$, the process $\mathbf{X}$ is stationary. □

The following lemma shows that condition (3) of the theorem implies the key assumption of Lemma 5.5.3.

Lemma 5.5.4. *Suppose that (3) of the theorem is satisfied. Then* $p^{h,d}_{\mathbf{t},\mathbf{I},a,\Delta}(\mathbf{X})$ *is independent of* a *for any* $\Delta, \mathbf{t}, \mathbf{I}$ *and* $d \geq 2\Delta$*.*

Proof. We start by showing that $p^{h,d}_{\mathbf{t},\mathbf{I},a,\Delta}(\mathbf{X})$ is a non-increasing function of a. If this is not the case, then there are $a_1 < a_2$ such that

$$p^{h,d}_{\mathbf{t},\mathbf{I},a_1,\Delta}(\mathbf{X}) < p^{h,d}_{\mathbf{t},\mathbf{I},a_2,\Delta}(\mathbf{X})\,. \tag{5.27}$$

By splitting the interval $[a_1, a_2]$ into two intervals of equal length, and repeating the procedure as many times as necessary, we can achieve the above inequality with $0 < a_2 - a_1 < \Delta$, so we simply assume that this constraint already holds. In this case, $[a_1, a_1 + \Delta] \cap [a_2, a_2 + \Delta] = [a_2, a_1 + \Delta] \neq \emptyset$, $[a_1, a_1 + \Delta] \cup [a_2, a_2 + \Delta] = [a_1, a_2 + \Delta]$, and the length of this union $a_2 + \Delta - a_1 < d$.

Recall the distance between any two points in set $A^{h,d}_{\mathbf{t},\mathbf{I}}(\mathbf{X})$ must be at least d. Therefore, any interval of the length smaller than d, contains at most one point of $A^{h,d}_{\mathbf{t},\mathbf{I}}(\mathbf{X})$. We call this "self-excluding" property. As a result of this property, inside any interval with length not exceeding d, the probability $p^{h,d}_{\mathbf{t},\mathbf{I},a,\Delta}(\mathbf{X})$ is, actually, additive. Specifically, let $I = [a, a + \Delta]$, $\Delta \leq d$, and let $I_1 = [a_1, a_1 + \Delta_1], I_2 = [a_2, a_2 + \Delta_2], \ldots$ be disjoint subintervals of I. Then

$$p^{h,d}_{\mathbf{t},\mathbf{I},a,\Delta}(\mathbf{X}) = \sum_{i=1}^{\infty} p^{h,d}_{\mathbf{t},\mathbf{I},a_i,\Delta_i}(\mathbf{X}).$$

Therefore, by (5.27),

$$\begin{aligned} 0 &< p^{h,d}_{\mathbf{t},\mathbf{I},a_2,\Delta}(\mathbf{X}) - p^{h,d}_{\mathbf{t},\mathbf{I},a_1,\Delta}(\mathbf{X}) \\ &= \left(p^{h,d}_{\mathbf{t},\mathbf{I},a_2,a_1+\Delta-a_2}(\mathbf{X}) + p^{h,d}_{\mathbf{t},\mathbf{I},a_1+\Delta,a_2-a_1}(\mathbf{X})\right) \\ &- \left(p^{h,d}_{\mathbf{t},\mathbf{I},a_1,a_2-a_1}(\mathbf{X}) + p^{h,d}_{\mathbf{t},\mathbf{I},a_2,a_1+\Delta-a_2}(\mathbf{X})\right) \\ &= p^{h,d}_{\mathbf{t},\mathbf{I},a_1+\Delta,a_2-a_1}(\mathbf{X}) - p^{h,d}_{\mathbf{t},\mathbf{I},a_1,a_2-a_1}(\mathbf{X})\,, \end{aligned}$$

so that

$$p^{h,d}_{\mathbf{t},\mathbf{I},a_1,a_2-a_1}(\mathbf{X}) < p^{h,d}_{\mathbf{t},\mathbf{I},a_1+\Delta,a_2-a_1}(\mathbf{X}) .$$

Consider again the intrinsic location functional

$$L(f,I) := \inf\{t : t \in A^{h,d}_{\mathbf{t},\mathbf{I}}(f) \cap I\} .$$

Take $I = [a_1, a_1 + D]$ for $D > d$. By the self-excluding property,

$$\mathbb{P}(L(\mathbf{X},I) \in [a, a+\delta]) = p^{h,d}_{\mathbf{t},\mathbf{I},a,\delta}(\mathbf{X})$$

for any a and δ satisfying $a \geq a_1$ and $a + \delta \leq a_1 + d$. In particular, the density of the law of $L(\mathbf{X}, I)$ in $(a, a+\delta)$, which exists by the condition (3), can be chosen independent of the length D of the interval I. Since

$$\mathbb{P}(L(\mathbf{X},I) \in [a_1,a_2]) = p^{h,d}_{\mathbf{t},\mathbf{I},a_1,a_2-a_1}(\mathbf{X})$$
$$< p^{h,d}_{\mathbf{t},\mathbf{I},a_1+\Delta,a_2-a_1}(\mathbf{X}) = \mathbb{P}(L(\mathbf{X},I) \in [a_1+\Delta, a_2+\Delta]),$$

there are $s_1 \in [a_1, a_2]$ and $s_2 \in [a_1 + \Delta, a_2 + \Delta]$, independent of D, such that $c := f_{\mathbf{X},I}(s_2) - f_{\mathbf{X},I}(s_1) > 0$. By the total mass considerations, there is $t = t_D \in (a_1 + D/2, a_1 + D)$ such that $f_{\mathbf{X},I}(t) \leq 2/D$, so that by the total variation constraint on the interval $[s_1, t]$ we have

$$\big(f_{\mathbf{X},I}(s_2) - f_{\mathbf{X},I}(t)\big) + \big(f_{\mathbf{X},I}(s_2) - f_{\mathbf{X},I}(s_1)\big) \leq TV_{[s_1,t]}(f_{\mathbf{X},I}) \leq f_{\mathbf{X},I}(s_1) + f_{\mathbf{X},I}(t) .$$

Rearranging the terms gives us

$$\frac{2}{D} \geq f_{\mathbf{X},I}(t) \geq f_{\mathbf{X},I}(s_2) - f_{\mathbf{X},I}(s_1) = c > 0 .$$

This relation, however, cannot hold for D large enough. The resulting contradiction proves that $p^{h,d}_{\mathbf{t},\mathbf{I},a,\Delta}(\mathbf{X})$ is a non-increasing function of a.

We can repeat the above argument by considering instead the intrinsic location functional,

$$L_1(f,I) := \sup\{t : t \in A^{h,d}_{\mathbf{t},\mathbf{I}}(f) \cap I\}.$$

This will show that $p^{h,d}_{\mathbf{t},\mathbf{I},a,\Delta}$ is a non-decreasing function of a. It follows that $p^{h,d}_{\mathbf{t},\mathbf{I},a,\Delta}$ must be independent of a. $\square$

The combination of the last two lemmas, obviously, completes the proof of Theorem 5.5.1. $\square$

5.6 Intrinsic locations sets

The arguments in the proof of Theorem 5.5.1 establishing the stationarity of the process $\mathbf{X}$ used only intrinsic location functionals of a special kind. In this section we concentrate on these special functionals. This will allow us both to relax the assumptions of Theorem 5.5.1 and to extend its statement to stochastic processes with sample paths in certain spaces different from the space of continuous functions.

Let $\mathcal{V}$ denote the collection of all subsets of $\mathbb{R}$. We equip $\mathcal{V}$ with the σ-field $\mathcal{F}_{\mathcal{V}}$ generated by the sets

$$\Big\{A \subseteq \mathbb{R} : \ A \cap I = \emptyset\Big\}, \ I = [a, b], \ -\infty < a < b < \infty \,.$$

Definition 2. *Let H be a set of functions on $\mathbb{R}$, invariant under shifts, equipped with its cylindrical σ-field. An **intrinsic location set** A is a measurable mapping from H to $\mathcal{V}$ that satisfies*

$$A(\theta_c f) = A(f) - c$$

for every $c \in \mathbb{R}$.

The following example shows that the set that played the crucial role in the proof of Theorem 5.5.1 is an intrinsic location set on $C(\mathbb{R})$.

Example 5.6.1. Let $H = C(\mathbb{R})$, and consider the set $A^{h,d}_{\mathbf{t},\mathbf{I}}$ defined in (5.21). We will check that it is an intrinsic location set. Clearly, only measurability needs to be checked. To this end, for $h, r \in \mathbb{R}$ define a map $\tau_{h,r} :\ C(\mathbb{R}) \to [-\infty, r]$ by

$$\tau_{h,r}(f) = \sup\bigl\{t \le r :\ f(t) = h\bigr\},\ f \in C(\mathbb{R})\,.$$

Since for each $t \in (-\infty, r]$,

$$\begin{aligned}\bigl\{f \in C(\mathbb{R}) :\ \tau_{h,r}(f) \ge t\bigr\} &= \bigl\{f \in C(\mathbb{R}) :\ f(s) = h \text{ for some } t \le s \le r\bigr\} \\ &= \bigcap_{k=1}^{\infty} \bigcup_{q \in [t,r] \cap \mathbb{Q}} \bigl\{f \in C(\mathbb{R}) :\ |f(q) - h| \le 1/k\bigr\}\,,\end{aligned}$$

where $\mathbb{Q}$ is the set of rational numbers, the map $\tau_{h,r}$ is measurable. Since for any $-\infty < a < b < \infty$

$$\begin{aligned}&\Bigl\{f \in C(\mathbb{R}) :\ A^{h,d}_{\mathbf{t},\mathbf{I}}(f) \cap [a,b] \ne \emptyset\Bigr\} \\ = &\bigcup_{\substack{r \in [a,b] \cap \mathbb{Q} \\ \text{or } r = a \text{ or } r = b}} \Bigl\{f \in C(\mathbb{R}) :\ \tau_{h,r}(f) \ge a,\ f(t) \ne h \text{ on } \bigl(\tau_{h,r}(f), \tau_{h,r}(f) + d + \varepsilon\bigr) \\ &\qquad \text{for some } \varepsilon > 0,\ f\bigl(\tau_{h,r}(f) + t_i\bigr) \in I_i,\ i = 1, \ldots, n\Bigr\}\,,\end{aligned}$$

the measurability of $A^{h,d}_{\mathbf{t},\mathbf{I}}$ will follow once we check that every set (say, B_r) in the union is measurable. To see that this last statement is true, denote for an interval $I = (c_1, c_2)$ and $\delta > 0$, $I_\delta = (c_1 + \delta, c_2 - \delta)$ if $c_1 + \delta < c_2 - \delta$, and set $I_\delta = \emptyset$ otherwise. Then

$$\begin{aligned}B_r = &\bigcup_{j=1}^{\infty} \bigcup_{k=1}^{\infty} \bigcup_{M=1}^{\infty} \bigcap_{m=M}^{\infty} \bigcup_{i=1}^{m} \\ &\Bigl\{f \in C(\mathbb{R}) :\ \tau_{h,r}(f) \in \bigl[a + (i-1)(r-a)/m, a + i(r-a)/m\bigr], \\ &f(t) \ne h \text{ on } \bigl[a + i(r-a)/m, a + i(r-a)/m + d + 1/j\bigr], \\ &f\bigl(a + i(r-a)/m + t_l\bigr) \in (I_l)_{1/k},\ l = 1, \ldots, n\Bigr\}\,,\end{aligned}$$

which makes it clear that B_r is a cylindrical set.

Therefore, $A^{h,d}_{\mathbf{t},\mathbf{I}}$ is indeed an intrinsic location set.

Given an intrinsic location set A on H, the functionals

$$L_1(f,I) := \inf\{t \in I \cap A(f)\},\ \ L_2(f,I) := \sup\{t \in I \cap A(f)\},\ f \in H\,, \qquad (5.28)$$

turn out to be intrinsic location functionals. Indeed, only their measurability is not immediately clear. However, if $I = [a,b]$, then

$$\big\{f \in H:\ L_1(f,I) = \infty\big\} = \big\{f \in H:\ A(f) \cap [a,b] = \emptyset\big\}\,,$$

while for $a < c < b$,

$$\big\{f \in H:\ L_1(f,I) \in (c,b] \cup \{\infty\}\big\} = \bigcup_{k=1}^{\infty} \big\{f \in H:\ A(f) \cap [a, c+1/k] = \emptyset\big\}\,.$$

Since these subsets of H are measurable, $L_1(\cdot, I)$ is measurable. Measurability of $L_2(\cdot, I)$ can be established in a similar way.

Notice that the functional $L_1(\cdot, I)$ is an earliest occurrence intrinsic location functional in the sense introduced in Section 5.4 . Similarly, the functional $L_2(\cdot, I)$ is *a latest occurrence intrinsic location functional*, i.e. a functional with the following property: for every $a < b < c$ and $f \in H$,

$$\text{if } L(f,[b,c]) \in [b,c] \text{ then } L(f,[a,c]) = L(f,[b,c]).$$

We already know that, if the process is stationary, then the distribution of $L(X,T)$ for any earliest occurrence intrinsic location functional L does not put any mass at the right endpoint of the interval, and its density in $(0,T)$ is nonincreasing (Proposition 5.4.4). This applies, in particular, to the functionals L_1 in (5.28). In a similar way we can show that distribution of $L(X,T)$ for any latest occurrence intrinsic location functional L does not put any mass at the left endpoint of the interval, and its density in $(0,T)$ is nondecreasing. This applies, in particular, to the functionals L_2 in (5.28). Moreover, only functionals of the type (5.28) were used in the proof of Theorem 5.5.1. We obtain, therefore, the following corollary.

Corollary 5.6.2. *Suppose that for any intrinsic location functional L on the space $C(\mathbb{R})$, of the type* (5.28)*, any interval $I = [a, b] \in \mathcal{I}$, the law of $L(\mathbf{X}, I)$ is absolutely continuous on (a, b) and has a density satisfying the total variation constraints. Then the process $\mathbf{X}$ is stationary.*

We can use the idea of the intrinsic location set to extend the results of Theorem 5.5.1 and Corollary 5.6.2 to processes with sample paths in certain spaces other than the space of continuous functions. We will call a set H of functions on $\mathbb{R}$ an *LI* set (from *locally integrable*) if it has following properties:

- H is invariant under shifts;
- H is equipped with its cylindrical σ-field $\mathcal{C}_H$;
- the map $H \times \mathbb{R} \to \mathbb{R}$ defined by $(f, t) \to f(t)$ is measurable;
- any $f \in H$ is locally integrable.

An example of an *LI* set is the space $D(\mathbb{R})$ of càdlàg functions on $\mathbb{R}$. Note that, by Fubini's theorem, for any $\delta > 0$ and an *LI* set H, the map $T_\delta : H \to C(\mathbb{R})$, defined by

$$T_\delta(f) = \int_t^{t+\delta} f(s)\, ds,\ t \in \mathbb{R} \tag{5.29}$$

is $\mathcal{C}_H/\mathcal{C}_{C(\mathbb{R})}$-measurable.

Let now $A : C(\mathbb{R}) \to \mathcal{V}$ be an intrinsic location set on $C(\mathbb{R})$. If H is an *LI* set, then for $\delta > 0$ we can define a mapping $B_\delta : H \to \mathcal{V}$ by $B_\delta = A \circ T_\delta$. By definition, B_δ is measurable. Further, for any $c \in \mathbb{R}$, $T_\delta(\theta_c f) = \theta_c(T_\delta f)$ for any $f \in H$ (using the same notation for the shift operator on different spaces). Therefore, for $f \in H$,

$$B_\delta(\theta_c f) = A(T_\delta(\theta_c f)) = A(\theta_c(T_\delta f))$$

$$= A(T_\delta f) - c = B_\delta(f) - c\,,$$

so that B_δ is an intrinsic location set on H.

Let $\mathbf{X}$ be a stochastic process with sample paths in H. For every $\delta > 0$, we view $\mathbf{Y}_\delta = T_\delta \mathbf{X}$ as a stochastic process with sample paths in $C(\mathbb{R})$. For any intrinsic location set A on $C(\mathbb{R})$ we have, in the above notation

$$\inf\{t \in I \cap A(\mathbf{Y}_\delta)\} = \inf\{t \in I \cap B_\delta(\mathbf{X})\}$$

for any interval I and, similarly, with the functionals of the type L_2 in (5.28). Therefore, if we assume that for any intrinsic location functional L on the space H, of the type (5.28), any interval $I = [a, b] \in \mathcal{I}$, the law of $L(\mathbf{X}, I)$ is absolutely continuous on (a, b) and has a density satisfying the total variation constraints, then, for every $\delta > 0$, the continuous process $\mathbf{Y}_\delta$ satisfies the assumptions of Corollary 5.6.2 and, hence, is stationary.

By Fubini's theorem we know that there is a Borel subset $\mathbb{R}_0$ of $\mathbb{R}$ of Lebesgue measure zero, such that for any $t \notin \mathbb{R}_0$,

$$nY_{1/n}(t) \to X(t) \text{ a.s..}$$

Combining this with the stationarity of $\mathbf{Y}_\delta$ for each $\delta > 0$ tells us that

$$\big(X(t_1 + h), \ldots, X(t_k + h)\big) \overset{d}{=} \big(X(t_1), \ldots, X(t_k)\big) \tag{5.30}$$

for any $k = 1, 2 \ldots$ and $t_1, \ldots, t_k$ and h such that none of the times in (5.30) is in the null set $\mathbb{R}_0$. For certain spaces H this implies stationarity of the process $\mathbf{X}$; by the right continuity this is, certainly, true for the space $D(\mathbb{R})$. Hence, we obtain the following result.

Proposition 5.6.3. *Let H be an LI set, and let $\mathbf{X}$ be a stochastic process with sample paths in H. Suppose that, for any intrinsic location functional L on the space H, of the* *type* (5.28)*, any interval $I = [a, b] \in \mathcal{I}$, the law of $L(\mathbf{X}, I)$ is absolutely continuous on (a, b) and has a density satisfying the total variation constraints. If for any process with sample paths in the set H,* (5.30) *implies stationarity, then $\mathbf{X}$ is stationary. This is the case, in particular, if $H = D(\mathbb{R})$.*

CHAPTER 6

INTRINSIC LOCATION FUNCTIONALS OF STATIONARY RANDOM FIELDS

6.1 Introduction

In Chapter 5, we have seen how the common properties of a large family of random locations in dimension one can be extracted out to form the notion intrinsic location functional. Briefly speaking, an intrinsic location functional is a mapping from $H \times \mathcal{I}$ to $\mathbb{R}$, where H is the path space and $\mathcal{I}$ is the set of all compact intervals in dimension 1. For each path f in H and each interval I in $\mathcal{I}$, an intrinsic location functional either gives a location on I, or returns value ∞ if such a location is not well-defined for that path. There are also some other mild conditions in the definition, which became the reason for the adjective "intrinsic". The family of one-dimensional intrinsic location functionals includes interesting random locations such as the location of the path supremum/infimum over an interval, the first/last hitting time of certain level, the point with the largest slope over an interval, etc. It is proved that the distribution of any intrinsic location functional is absolutely continuous inside the interval of interest, and that despite of the different definitions and origins of different location functionals in this family, the density functions of their distributions always satisfy the same total variation constraints. See Chapter 5 for details.

It then becomes natural to ask what will happen in higher dimension. Can we extend the definition of intrinsic location functional to higher dimensional case? Shall we still have the continuities and total variation constraints? If yes, what exact form will they take? There are abundant motivations for such an extension, since many concepts only make sense or become interesting in higher dimensional case. For example, the locations and values of the critical points of a random field is related to the homology of its excursion sets. For a detailed treatment, see [2]. However, in one dimension, all the information that the homology groups can provide is simply the connectedness, while in higher dimensions the homology groups provide more interesting information, such as whether the excursion set contains a loop or a hole, etc. Therefore an understanding of the distribution of intrinsic location functionals in higher dimensional case will be helpful to the study of the random homology of excursion sets. Another example could be the locations of the saddle points, which do not exist in dimension 1.

The goal of this chapter is therefore to answer the questions posed above. We will see that such an extension is feasible, and that most of the results obtained in one-dimensional case in Chapter 5 do have their higher dimensional counterparts. Moreover, in higher dimensional case, in addition to translation, another rigid body movement–rotation, becomes possible, and the corresponding probabilistic notion is isotropy. Thus it is also interesting to see what more conditions we can get with isotropy in hand.

The rest of this chapter is organized as follows: In Section 6.2 we will define the intrinsic location functionals in higher dimension, and introduce notations needed for later use. Section 6.3 contains the statement and proof of the main theorem, which guarantees both the continuity properties and the total variation constraints. An analysis on the isotopic random fields is performed in Section 6.4, which gives rise to the so-called "angular total variation constraints".

6.2 Definition and Notation

Let H be a space of functions defined on $\mathbb{R}^n$. Typically, H could be the space of all the continuous functions, the space of all smooth functions, or, more generally, the space of all upper semi-continuous functions, etc. Let $\mathcal{I}$ be a set of subsets in $\mathbb{R}^n$. In stationary case, we often take $\mathcal{I}$ to be the set of all compact, non-degenerate hypercubes in $\mathbb{R}^n$: $\mathcal{I} = \{\prod_{i=1}^n [a_i, b_i] : a_i < b_i, \forall i\}$. For any vector $c \in \mathbb{R}^n$, define θ_c to be the shift operator on H: $\forall f \in H$, $\forall x \in \mathbb{R}^n$, $\theta_c f(x) := f(x + c)$.

Definition 3. *A mapping $L : (H \times \mathcal{I}) \to \mathbb{R}^n \cup \{\infty\}$ is called an intrinsic location functional, if it satisfies the following conditions.*

1. *For every $I \in \mathcal{I}$ the map $L(\cdot, I) : H \to \mathbb{R}^n \cup \{\infty\}$ is measurable.*
2. *For every $f \in H$ and $I \in \mathcal{I}$, $L(f, I) \in I \cup \{\infty\}$.*
3. *(Shift compatibility) For every $f \in H$, $I \in \mathcal{I}$ and $c \in \mathbb{R}^n$,*

   $$L(f, I) = L(\theta_c f, I - c) + c,$$

   where $I - c$ is the hypercube I shifted by c, and $\infty + c = \infty$.
4. *(Stability under restrictions) For every $f \in H$ and $I_1, I_2 \in \mathcal{I}$, $I_2 \subseteq I_1$, if $L(f, I_1) \in I_2$, then $L(f, I_2) = L(f, I_1)$.*
5. *(Consistency of existence) For every $f \in H$ and $I_1, I_2 \in \mathcal{I}$, $I_2 \subseteq I_1$, if $L(f, I_2) \neq \infty$, then $L(f, I_1) \neq \infty$.*

This is clearly a generalization of the intrinsic location functional in one dimension, defined in Chapter 5. The motivation and intuition behind each condition are also the same as the one dimensional case: condition (1) clarifies the measurability issue; condition (2) justifies the name "location functional" by guaranteeing that it does return a location on the hypercube of interest, or infinity if such a location is not well-defined; condition (3) is the reason for the adjective "intrinsic"; condition (4) makes the location of an hypercube also the corresponding location for any smaller hypercube containing that point; finally, condition (5) requires that a location already existing for an hypercube will not disappear in any larger hypercube, although its value may change.

Example 6.2.1. Let H be the space of all the upper semi-continuous functions on $\mathbb{R}^n$. Let $\tau_{f,I} = \inf\{t \in I : f(t) = \sup_{s \in I} f(s)\}$ be the location of supremum of a function $f \in H$ on I, where the infimum in the definition is taken in lexicographic order. Then it is easy to check that $\tau_{f,I}$ satisfies all the conditions listed above. Thus the location of supremum in higher dimensional case, just as its one dimensional counterpart analyzed in Chapters 3 and 4, is an intrinsic location functional. In this case, $\tau_{f,I}$ always takes a value in I.

Example 6.2.2. Let $H = \mathcal{C}(\mathbb{R}^n)$. The hitting times in one dimension now become level sets $H_l := \{t \in I : f(t) = l\}$, and are therefore not random locations. However, we can take a certain point in H^l according to some rule. For instance, for any linear objective function $g \in \mathcal{C}(\mathbb{R}^n)$, take the point in H^l which maximizes its value. In case of a tie, use, for example, lexicographic order to break it. Then such defined point is an intrinsic location functional, and it can be regarded as an extension of hitting times in higher dimension. Since it is totally possible that the function does not hit level l on I, this is an example of intrinsic location functional which can possibly take value ∞.

Example 6.2.3. Let $H = \mathcal{C}(\mathbb{R}^n)$, the space of all continuous function on $\mathbb{R}^n$. Let $S_{f,I}$ be the set of all saddle points of function $f \in H$ in the hypercube I. Define $s_{f,I} = \inf S_{f,I}$, where the infimum is taken in lexicographic order. Then $s_{f,I}$ is an intrinsic location functional. Again, $s_{f,I}$ can take ∞ as its value.

Example 6.2.4. Let H and $S_{f,I}$ be the same as in the previous example. Now instead of ordering the saddle points according to their own coordinate in lexicographic order, we order them by their relative location to the geometric center a of the hypercube I. More precisely, for every $s \in S_{f,I}$, compute its distance to a, $||s - a||$, where $|| \cdot ||$ is the n-dimensional Euclidean norm, and take the point s which minimizes this distance. In case of a tie, we use again the lexicographic order to break it. It is not hard to see that such defined location functional is not stable under restriction, and is therefore not an intrinsic location functional. This example shows that the selection rule over a discrete point set does matter in determining whether a random location is an intrinsic location functional or not. As we have seen, the lexicographic order is in general good as a criterion for selection, while the distance to the center of the hypercube is not. The other selection rules which does not satisfy the conditions for intrinsic location functionals includes:

– Taking the point which is closest to a fixed point, for instance, the origin: under this selection rule the location functional is stable under restriction and its existence is consistent, but it is not shift invariant.

– Only taking a point if it is within certain distance to a corner/face of the hypercube: under this selection rule the location functional is shift invariant and stable under restriction, but it is not consistent in existence.

For the remainder of this chapter $\mathbf{X} = (X(t), t \in \mathbb{R}^n)$ is a stationary random field defined on some probability space $\big(\Omega, \mathcal{F}, P\big)$, and having paths in H. For a compact hypercube $I = \prod_{i=1}^n [a_i, b_i] \subset \mathbb{R}^n$, we will denote the value of a location functional L on that hypercube by $L(\mathbf{X}, I)$.

Let $I = \prod_{i=1}^{n}[a_i, b_i]$ be an element in $\mathcal{I}$. For $k = 0, 1, ..., n$, the (open) k-faces of this hypercube are all the k-dimensional hypercubes having the form

$$\{t = (t_1, ..., t_n) : t_{\sigma(1)} = a_{\sigma(1)}, ..., t_{\sigma(m)} = a_{\sigma(m)}, t_{\sigma(m+1)} = b_{\sigma(m+1)}, ...,$$

$$t_{\sigma(n-k)} = b_{\sigma(n-k)}, t_{\sigma(n-k+1)} \in (a_{\sigma(n-k+1)}, b_{\sigma(n-k+1)}), ..., t_{\sigma(n)} \in (a_{\sigma(n)}, b_{\sigma(n)})\}$$

for some $m = 0, ..., n-k$, where $\{\sigma(i)\}_{i=1,...,n}$ is a permutation of $\{1, ..., n\}$. Denote by I^k the union of all the k-faces of I. Then $I^i \cap I^j = \phi$ for any $i \neq j$, and $I = \cup_{i=0}^{n} I^i$. We will also need to pick a direction $j \in \{1, ..., n\}$. In this case the set I^k can be further divided into two parts: $I^k = I^{k,j} \cup I^{k,\hat{j}}$, where $I^{k,j}$ is the union of the k-faces in which $t_j \in (a_j, b_j)$ is free, and $I^{k,\hat{j}}$ is the union of the k-faces in which either $t_j = a_j$ or $t_j = b_j$. Also for this direction j and $s \in (a_j, b_j)$, denote by $I^k_{t_j=s}$ the intersection of the hyperplane $t_j = s$ and $I^{k,j}$, and I^k_j the projection of $I^{k,j}$ to the $n-1$ dimensional subspace $t_j = 0$.

For any $I = \prod_{i=1}^{n}[a_i, b_i]$, we denote by $F_{\mathbf{X},I}$ the law (distribution function) of $L(\mathbf{X}, I)$, then $F_{\mathbf{X},I}$ is a distribution supported on the set $I \cup \{\infty\}$.

Recall that for a function f defined on $\mathbb{R}$, the total variation of f over an interval (t_1, t_2) is

$$TV_{(t_1,t_2)}(f_{\mathbf{X},T}) = \sup \sum_{i=1}^{n-1} \left| f_{\mathbf{X},T}(s_{i+1}) - f_{\mathbf{X},T}(s_i) \right|,$$

where the supremum is taken over all choices of $t_1 < s_1 < \ldots < s_n < t_2$. Similarly, define the positive/negative variation to be

$$TV^{\pm}_{(t_1,t_2)}(f_{\mathbf{X},T}) = \sup \sum_{i=1}^{n-1} \left(f_{\mathbf{X},T}(s_{i+1}) - f_{\mathbf{X},T}(s_i) \right)_{\pm}.$$

Again, the supremum is taken over all choices of $t_1 < s_1 < \ldots < s_n < t_2$. In higher dimensional case, for $j = 1, ..., n$, $s = (s_1, ..., s_{n-1}) \in \mathbb{R}^{n-1}$, and for a function f defined on $\mathbb{R}^n$, we can define the "directional total variation" of f on direction j as

$$TV^j_{(t_1,t_2)}(f|_s) := TV_{(t_1,t_2)}(f(s_1, ..., s_{j-1}, \cdot, s_j, ..., s_{n-1})).$$

Similarly,

$$TV^{j\pm}_{(t_1,t_2)}(f|_s) := TV^{\pm}_{(t_1,t_2)}(f(s_1, ..., s_{j-1}, \cdot, s_j, ..., s_{n-1})).$$

6.3 Main results in higher dimensions

Theorem 6.3.1. *Let L be an intrinsic location functional on some path space H, $\mathbf{X} = (X(t),\, t = (t_1, ..., t_n) \in \mathbb{R}^n)$ be a stationary random field such that $\mathbb{P}(\mathbf{X} \in H) = 1$. Let $I = \prod_{i=1}^n [0, b_i]$. Then the restriction of the law $F_{\mathbf{X},I}$ to I^k is absolutely continuous with respect to the k-dimensional Lebesgue measure m^k for any $k = 1, ..., n$. Moreover, given any $j = 1, ..., n$, there exists a version of (k-dimensional) densities of $F_{\mathbf{X},I}$ restricted to I^k, denoted by $\{f^{k,j}_{\mathbf{X},I}\}_{k=1,...,n}$, which satisfies the following conditions.*

(a) For $k = 1, ..., n$, $f^{k,j}_{\mathbf{X},I}(t)$ is upper semi-continuous and almost everywhere continuous on t_j.

(b) For $k = 1, ..., n$, the limits

$$f^{k,j,0+}_{\mathbf{X},I}(s) := \lim_{t_j \downarrow 0, p^j t = s} f^{k,j}_{\mathbf{X},I}(t)$$

and

$$f^{k,j,b_j-}_{\mathbf{X},I}(s) := \lim_{t_j \uparrow b_j, p^j t = s} f^{k,j}_{\mathbf{X},I}(t)$$

exist almost everywhere on I^k_j with respect to the $k-1$ dimensional Lebesgue measure m^{k-1}, where p^j is the orthogonal projection operator to the subspace $t_j = 0$.

(c) For $k = 1, ..., n$, the density $f^{k,j}_{\mathbf{X},I}$ has a universal upper bound given by

$$f^{k,j}_{\mathbf{X},I}(t) \leq \prod_{i=1}^{k} \max\left(\frac{1}{t_{\sigma(i)}}, \frac{1}{b_{\sigma(i)} - t_{\sigma(i)}}\right),\ 0 < t < T\,, \tag{6.1}$$

where $\sigma(1), ..., \sigma(k)$ are the k free coordinates of t: $\{\sigma(i)\}_{i=1,...,k} = \{m \in \{1, ..., n\} : t_m \in (0, b_m)\}$.

(d) For every $0 < t_1 < t_2 < b_j$,

$$\sum_{k=1}^{n} \int_{I^k_j} TV^{j}_{(t_1,t_2)}(f^{k,j}_{\mathbf{X},I}|_s) m^{k-1}(ds) \leq \sum_{k=1}^{n} \int_{I^k_{t_j=t_1} \cup I^k_{t_j=t_2}} f^{k,j}_{\mathbf{X},I}(t) m^{k-1}(dt)\,. \tag{6.2}$$

(e) The densities have bounded positive variations almost everywhere at the side $t_j = 0$ and bounded negative variations almost everywhere at the side $t_j = b_j$. Furthermore, for every $0 < \varepsilon < T$,

$$\sum_{k=1}^{n} \int_{I^k_j} TV^{j+}_{(0,\varepsilon)}(f^{k,j}_{\mathbf{X},I}|_s) m^{k-1}(ds) \leq \sum_{k=1}^{n} \int_{I^k_{t_j=\varepsilon}} f^{k,j}_{\mathbf{X},I}(t) m^{k-1}(dt), \tag{6.3}$$

and

$$\sum_{k=1}^{n} \int_{I^k_j} TV^{j-}_{(b_j-\varepsilon,b_j)}(f^{k,j}_{\mathbf{X},I}|_s) m^{k-1}(ds) \leq \sum_{k=1}^{n} \int_{I^k_{t_j=b_j-\varepsilon}} f^{k,j}_{\mathbf{X},I}(t) m^{k-1}(dt). \tag{6.4}$$

(f)

$$\sum_{k=1}^{n} \int_{I^k_j} f^{k,j,0+}_{\mathbf{X},I}(s) m^{k-1}(ds) < \infty$$

if and only if

$$\sum_{k=1}^{n} \int_{I^k_j} TV^{j}_{(0,\varepsilon)}(f^{k,j}_{\mathbf{X},I}|_s) m^{k-1}(ds) < \infty$$

for some (equivalently, any) $0 < \varepsilon < b_j$, in which case

$$\begin{aligned} &\textstyle\sum_{k=1}^{n} \int_{I^k_j} TV^{j}_{(0,\varepsilon)}(f^{k,j}_{\mathbf{X},I}|_s) m^{k-1}(ds) \\ < \ &\textstyle\sum_{k=1}^{n} \left(\int_{I^k_j} f^{k,j,0+}_{\mathbf{X},I}(s) m^{k-1}(ds) + \int_{I^k_{t_j=\varepsilon}} f^{k,j}_{\mathbf{X},I}(t) m^{k-1}(dt) \right). \end{aligned} \tag{6.5}$$

Similarly,

$$\sum_{k=1}^{n} \int_{I_j^k} f_{\mathbf{X},I}^{k,j,b_j-}(s) m^{k-1}(ds) < \infty$$

if and only if

$$\sum_{k=1}^{n} \int_{I_j^k} TV^j_{(b_j-\varepsilon,b_j)}(f_{\mathbf{X},I}^{k,j}|_s) m^{k-1}(ds) < \infty$$

for some (equivalently, any) $0 < \varepsilon < b_j$*, in which case*

$$\begin{aligned} & \textstyle\sum_{k=1}^{n} \int_{I_j^k} TV^j_{(b_j-\varepsilon,b_j)}(f_{\mathbf{X},I}^{k,j}|_s) m^{k-1}(ds) \\ \leq \;& \textstyle\sum_{k=1}^{n} \left(\int_{I_j^k} f_{\mathbf{X},I}^{k,j,b_j-}(s) m^{k-1}(ds) + \int_{I^k_{t_j=b_j-\varepsilon}} f_{\mathbf{X},I}^{k,j}(t) m^{k-1}(dt) \right) . \end{aligned} \tag{6.6}$$

Remark 6.3.2. It is not hard to see that all the properties in the definition of intrinsic location functional are indispensable for the total variation constraints to hold. One can construct examples to show this using the counterexamples provided in Example 6.2.4 and simple (for instance, periodic) stationary random fields.

Before we start the proof of the theorem, let us firstly announce the following simple yet important lemma.

Lemma 6.3.1. (i) *For any* $c \in \mathbb{R}^n$*, any* $I \in \mathcal{I}$*,*

$$F_{\mathbf{X},I+c}(\cdot) = F_{\mathbf{X},I}(\cdot - c) .$$

(ii) *For any* $I_1, I_2 \in \mathcal{I}, I_2 \subseteq I_1$*,*

$$F_{\mathbf{X},I_1}(B) \leq F_{\mathbf{X},I_2}(B) \text{ for any Borel set } B \subseteq I_2 .$$

(iii) *For any* $I_1, I_2 \in \mathcal{I}, I_2 \subseteq I_1$*,*

$$F_{\mathbf{X},I_1}(\{\infty\}) \leq F_{\mathbf{X},I_2}(\{\infty\}).$$

As in the one dimensional case, the three points in this lemma are, respectively, direct results of shift invariance, stability under restriction and consistency of existence properties of intrinsic location functionals.

Now let us start the proof of the theorem.

Proof. Given any $k = 1, ..., n$, choose $\{\delta_i\}_{i=1,...,k}$ such that $0 < \delta_i < b_i/2$ for any $i = 1, ..., k$. We prove that if L is the location of the supremum, then for any $t = (t_1, ..., t_k) \in \mathbb{R}^k$ such that $\delta_i \le t_i \le b_i - \delta_i, \forall i = 1, ..., k$, for every $\rho > 0$ and every $\varepsilon = (\varepsilon_1, ..., \varepsilon_k)$ such that $0 < \varepsilon_i < \delta_i\rho/(1+\rho)$ for any $i = 1, ..., k$,

$$P\big(t_i < L(\mathbf{X}, I)_i \le t_i + \varepsilon_i, i = 1, ..., k\big) \le (1+\rho)^k \prod_{i=1}^{k} \left(\varepsilon_i \max\left(\frac{1}{t_i}, \frac{1}{b_i - t_i}\right)\right), \quad (6.7)$$

where $L(\mathbf{X}, I)_i$ is the i-th component of $L(\mathbf{X}, I) \in \mathbb{R}^n \cup \{\infty\}$, with the tradition $\infty_i = \infty, \forall i$.

By symmetry between different directions and between 0 and $b = (b_1, ...b_n)$, this inequality, once proved, will imply the absolute continuity of $F_{\mathbf{X},I}$ on I^k with respect to the k-dimensional Lebesgue measure. Moreover, it shows that the version of density given by

$$f^k_{\mathbf{X},I}(t) = \limsup_{\varepsilon \to 0, \varepsilon > 0} \prod_{i=1}^{k} \frac{1}{\varepsilon_{\sigma(i)}} P\left(t_{\sigma(i)} < L(\mathbf{X}, I) \le t_{\sigma(i)} + \varepsilon_{\sigma(i)}\right), t \in I^k,$$

where $\sigma(1), ..., \sigma(k)$ are the k free coordinates of t, satisfies the bound (6.1).

Now we look at the proof of (6.7). Suppose that the result is not true. That means there exists some $t \in \prod_{i=1}^k [\delta_i, b_i - \delta_i]$ and $\varepsilon \in \prod_{i=1}^k (0, \delta_i\rho/(1+\rho))$, for which (6.7) fails. Choose $\theta \in \mathbb{R}^k$, such that

$$\varepsilon_i < \theta_i < \frac{\rho}{1+\rho}\delta_i, \forall i = 1, ..., k$$

and $a_1 = (a_{11}, ..., a_{1k}), a_2 = (a_{21}, ..., a_{2k}) \in \mathbb{R}^k$ such that for $i = 1, ..., k$, $0 < a_{1i} < t_i < a_{2i} < b_i$ and

$$\min(t_i, b_i - t_i) - \theta_i < a_{2i} - a_{1i} < \min(t_i, b_i - t_i) - \varepsilon_i.$$

For $s \in \prod_{i=1}^k [a_{1i}, a_{2i}]$, by stationarity, we have

$$P\left(s_i < L(\mathbf{X}, I + s - t)_i \le s_i + \varepsilon_i, i = 1, ..., k\right) > (1+\rho)^k \prod_{i=1}^k \left(\varepsilon_i \max\left(\frac{1}{t_i}, \frac{1}{b_i - t_i}\right)\right), \tag{6.8}$$

where $I + s - t$ is the hypercube I shifted by $(s_1 - t_1, ..., s_k - t_k, 0, ..., 0)$.

Further, let $s_1, s_2 \in \prod_{i=1}^k [a_{1i}, a_{2i}]$ such that

$$\prod_{i=1}^k (s_{1i}, s_{1i} + \varepsilon_i] \cap \prod_{i=1}^k (s_{2i}, s_{2i} + \varepsilon_i] = \phi.$$

Then we check that

$$\{L(\mathbf{X}, I - s_j + t) \in \prod_{i=1}^k (s_{ji}, s_{ji} + \varepsilon_i], j = 1, 2\} = \phi. \tag{6.9}$$

Indeed, let Ω_{s_1,s_2} be the event in (6.9). Note that the hypercubes $\prod_{i=1}^k (s_{1i}, s_{1i} + \varepsilon_i]$ and $\prod_{i=1}^k (s_{2i}, s_{2i} + \varepsilon_i]$ are disjoint and, by the choice of the parameters a_1 and a_2, each of these two hypercubes is a subset of both $\prod_{i=1}^k [s_{1i} - t_i, s_{1i} - t_i + b_i]$ and $\prod_{i=1}^k [s_{2i} - t_i, s_{2i} - t_i + b_i]$, and therefore also their joint $I_{s_1,s_2} := \prod_{i=1}^k [\max(s_{1i}, s_{2i}) - t_i, \min(s_{1i}, s_{2i}) - t_i + b_i]$. Now consider $L(\mathbf{X}, I_{s_1,s_2})$. On the event Ω_{s_1,s_2}, by stability under restriction, we have both

$$L(\mathbf{X}, I_{s_1,s_2}) = L(\mathbf{X}, \prod_{i=1}^k [s_{1i} - t_i, s_{1i} - t_i + b_i]) \in \prod_{i=1}^k (s_{1i}, s_{1i} + \varepsilon_i]$$

and

$$L(\mathbf{X}, I_{s_1,s_2}) = L(\mathbf{X}, \prod_{i=1}^k [s_{2i} - t_i, s_{2i} - t_i + b_i]) \in \prod_{i=1}^k (s_{2i}, s_{2i} + \varepsilon_i],$$

which are obviously contradictory. Thus (6.9) is established.

We now apply (6.8) and (6.9) to the points $s_{\bar{j}} := s_{(j_1,...,j_k)} = (a_{11}+j_1\varepsilon_1, ..., a_{1k}+j_k\varepsilon_k)$, for $j_i = 0, 1, \ldots, \lceil (a_{2i}-a_{1i})/\varepsilon_i \rceil - 1$, $i = 1, ..., k$. We have

$$\begin{aligned}
1 &\geq P\left(\bigcup_{\bar{j}} \{L(\mathbf{X}, \prod_{i=1}^{k}[(s_{\bar{j}}-t)_i, (s_{\bar{j}}-t+b)_i]) \in \prod_{i=1}^{k}((s_{\bar{j}})_i, (s_{\bar{j}}+\varepsilon)_i]\}\right) \\
&= \sum_{\bar{j}} P\left(L(\mathbf{X}, \prod_{i=1}^{k}[(s_{\bar{j}}-t)_i, (s_{\bar{j}}-t+b)_i]) \in \prod_{i=1}^{k}((s_{\bar{j}})_i, (s_{\bar{j}}+\varepsilon)_i]\right) \\
&> (1+\rho)^k \prod_{i=1}^{k}\left(\frac{a_{2i}-a_{1i}}{\varepsilon_i}\varepsilon_i \max\left(\frac{1}{t_i}, \frac{1}{b_i-t_i}\right)\right) \\
&> (1+\rho)^k \prod_{i=1}^{k}\left(\Big(\min\big(t_i, b_i-t_i\big) - \theta_i\Big) \max\left(\frac{1}{t_i}, \frac{1}{b_i-t_i}\right)\right) \\
&> \prod_{i=1}^{k}\left(1 - \frac{\delta_i}{\min(t_i, b_i-t_i)}\frac{\rho}{1+\rho}\right)(1+\rho)^k \geq \left(1-\frac{\rho}{1+\rho}\right)^k (1+\rho)^k = 1
\end{aligned}$$

by the choice of θ. This contradiction proves (6.7).

Once the existence of a k-dimensional density function, denoted by $f^k_{\mathbf{X},I}$, is guaranteed on I^k, Lebesgue differentiation theorem in high dimension assures that

$$f^k_{\mathbf{X},I}(t) = \lim_{r\to 0}\frac{1}{m^k(B^k(0,r))}\int_{B^k(t,r)} f^k_{\mathbf{X},I}(s)ds \tag{6.10}$$

for almost every $t \in I^k$, where $B^k(t,r)$ is the k-dimensional open ball in I^k with center t and radius r. Define set $A^{k'}$ to be the set of all points in I^k for which (6.10) holds, then $A^{k'} \subseteq I^k$ and $m^k(I^k \backslash A^{k'}) = 0$. Define set A^k to be the set of all the continuous points of $f^k_{\mathbf{X},I}$ in $A^{k'}$: $A^k := \{t \in A^{k'} : f^k_{\mathbf{X},I}(t) = \lim_{s\to t, s\in A^{k'}} f^k_{\mathbf{X},I}(s)\}$. In the following, we prove that

Proposition 6.3.3. $m^k(I^k) = m^k(A^{k'}) = m^k(A^k)$.

Proof. For any positive integer h, define

$$A^k_h := \{t \in A^{k'} : \limsup_{s\to t, s\in A^{k'}} f^k_{\mathbf{X},I}(s) - \liminf_{s\to t, s\in A^{k'}} f^k_{\mathbf{X},I}(s) > \frac{1}{h}\}.$$

Then it is easy to see that $A^{k'} = A^k \cup \left(\cup_{h=1}^{\infty} A_h^k\right)$. Thus we need only to show $m^k(A_h^k) = 0$ for any k and h. Moreover, notice that I^k can be decomposed as the union of different k-dimensional open faces, each of which can be further represented by a pair $(\bar{\sigma}, m)$, where $m = 1, ..., n-k$, and $\bar{\sigma} = (\sigma(1), ..., \sigma(m), \sigma(m+1), ..., \sigma(n-k), \sigma(n-k+1), ..., \sigma(n))$ is a permutation of $\{1, ..., n\}$. More precisely, for δ small enough, define

$$I^k_{(\bar{\sigma},m,\delta)} := \{t \in I^k : t_{\sigma(i)} = 0 \text{ for } i \le m, t_{\sigma(j)} = b_j \text{ for } m+1 \le j \le n-k,$$

$$t_{\sigma(l)} \in (\delta, b_l - \delta) \text{ for } l \ge n-k+1\},$$

then

$$I^k = \cup_{(\bar{\sigma},m,\delta)} I^k_{(\bar{\sigma},m,\delta)}.$$

We are going to prove that $m^k(A_h^k \cap I^k_{(\bar{\sigma},m,\delta)}) = 0$ for any given $k = 1, ..., n$, $m = 1, ..., k$, $\delta > 0$ and any possible $\bar{\sigma}$.

In order to lighten the notation, in the following we take k, m, δ and $\bar{\sigma}$ as given, and drop them from the index. For $\varepsilon < \delta$, define

$$I_{+\varepsilon} = \{t \in \mathbb{R}^n : t_{\sigma(i)} \in [0, b_{\sigma(i)}], \forall i = 1, ..., n-k, t_{\sigma(j)} \in [-\varepsilon, b_{\sigma(j)}+\varepsilon], \forall j = n-k+1, ..., n\},$$

$$I_{-\varepsilon} = \{t \in \mathbb{R}^n : t_{\sigma(i)} \in [0, b_{\sigma(i)}], \forall i = 1, ..., n-k, t_{\sigma(j)} \in [\varepsilon, b_{\sigma(j)}-\varepsilon], \forall j = n-k+1, ..., n\}.$$

Notice that $I^k_{(\bar{\sigma},m,\delta)}$ is a subset of the corresponding open faces of both $I_{+\varepsilon}$ and $I_{-\varepsilon}$. That is, define

$$I_{\pm} := \{t \in I_{\pm\varepsilon} : t_{\sigma(i)} = 0 \text{ for } i \le m, t_{\sigma(j)} = b_j \text{ for } m+1 \le j \le n-k,$$

$$t_{\sigma(l)} \in (\mp\varepsilon, b_{\sigma(l)} \pm \varepsilon) \text{ for } l \ge n-k+1\},$$

then $I^k_{(\bar{\sigma},m,\delta)} \subset I_- \subset I_+$. Denote by $f_{\pm\varepsilon}$ the (k-dimensional) density functions of $L(\mathbf{X}, I_{\pm\varepsilon})$ on the faces $I_{\pm}$. We have the following lemma, which can be considered as a stronger version of Lemma 6.3.1 (ii).

Lemma 6.3.2.

$$f_{+\varepsilon}(t) \leq \liminf_{s\to t, s\in A^{k'}} f^k_{\mathbf{X},I}(s) \leq \limsup_{s\to t, s\in A^{k'}} f^k_{\mathbf{X},I}(s) \leq f_{-\varepsilon}(t)$$

for almost every $t \in I_-$.

Proof. Since the exception set of $A^{k'}$ is always a null set for any hypercube I, we only need to prove the result for all the $t \in I_-$ such that both

$$f^k_{\mathbf{X},I_{+\varepsilon}}(t) = \lim_{r\to 0} \frac{1}{m^k(B^k(0,r))} \int_{B^k(t,r)} f^k_{\mathbf{X},I_{+\varepsilon}}(s)ds \tag{6.11}$$

and

$$f^k_{\mathbf{X},I_{-\varepsilon}}(t) = \lim_{r\to 0} \frac{1}{m^k(B^k(0,r))} \int_{B^k(t,r)} f^k_{\mathbf{X},I_{-\varepsilon}}(s)ds$$

hold. Indeed, let $\liminf_{s\to t, s\in A^{k'}} f^k_{\mathbf{X},I}(s) = a$. Then for any $\eta > 0$, there exists a point $s \in I_-$ and an arbitrarily small radius r, such that $||s-t||_\infty < \varepsilon$ and $\frac{1}{m^k(B^k(0,r))}\int_{B^k(s,r)} f^k_{\mathbf{X},I}(u)du \leq a + \eta$. Then by Lemma 6.3.1 (ii),

$$\frac{1}{m^k(B^k(0,r))} \int_{B^k(t,r)} f^k_{\mathbf{X},I_{+\varepsilon}}(u)du \leq \frac{1}{m^k(B^k(0,r))} \int_{B^k(s,r)} f^k_{\mathbf{X},I}(u)du \leq a + \eta.$$

Taking $r \to 0$, (6.11) implies immediately that $f^k_{\mathbf{X},I_{+\varepsilon}}(t) \leq a + \eta$. Since η can be any positive number, we conclude that $f^k_{\mathbf{X},I_{+\varepsilon}}(t) \leq \liminf_{s\to t, s\in A^{k'}} f^k_{\mathbf{X},I}(s)$. The inequality on the right can be proved in the symmetric way. $\square$

Suppose there are $k, h, \bar{\sigma}, m$ and δ, such that $m^k(A^k_h \cap I^k_{(\bar{\sigma},m,\delta)}) > 0$. Then Lemma 6.3.2 implies that for any $\varepsilon < \delta$,

$$P(L(\mathbf{X}, I_{-\varepsilon}) \in I_-) - P(L(\mathbf{X}, I_{+\varepsilon}) \in I_-) = \int_{I_-} \left(f^k_{\mathbf{X},I_-}(t) - f^k_{\mathbf{X},I_+}(t)\right) m^k(dt)$$

$$\geq \int_{A^k_h \cap I^k_{(\bar{\sigma},m,\delta)}} \left(f^k_{\mathbf{X},I_-}(t) - f^k_{\mathbf{X},I_+}(t)\right) m^k(dt) \geq \frac{1}{h} m^k(A^k_h \cap I^k_{(\bar{\sigma},m,\delta)}) > 0,$$

thus also

$$\lim_{\varepsilon\to 0} P(L(\mathbf{X}, I_{-\varepsilon}) \in I_-) - P(L(\mathbf{X}, I_{+\varepsilon}) \in I_-) > 0. \tag{6.12}$$

This, however, contradicts with

Lemma 6.3.3.

$$\lim_{\varepsilon \to 0} P(L(\mathbf{X}, I_{-\varepsilon}) \in I_-) - P(L(\mathbf{X}, I_{+\varepsilon}) \in I_-) = 0.$$

Proof. Recall that the free coordinates in $I_\pm$ are $\sigma(n-k+1), ..., \sigma(n)$. Define

$$I_{i,\varepsilon} = \{t \in \mathbb{R}^n : t_{\sigma(i)} \in [0, b_{\sigma(i)}], \forall j = 1, ..., n-k,$$

$$t_{\sigma(j)} \in [-\varepsilon, b_{\sigma(j)} + \varepsilon], \forall j = n-k+1, ..., n-k+i,$$

$$t_{\sigma(j)} \in [\varepsilon, b_{\sigma(j)} - \varepsilon], \forall j = n-k+i+1, ..., n\},$$

then $I_{0,\varepsilon} = I_{-\varepsilon}$ and $I_{k,\varepsilon} = I_{+\varepsilon}$. By increasing i we "expand" the range of the coordinates one by one from $[\varepsilon, b_{\sigma(j)} - \varepsilon]$ to $[-\varepsilon, b_{\sigma(j)} + \varepsilon]$, and find a "path" from $I_{-\varepsilon}$ to $I_{+\varepsilon}$. Notice that

$$\begin{aligned} & P(L(\mathbf{X}, I_{-\varepsilon}) \in I_-) - P(L(\mathbf{X}, I_{+\varepsilon}) \in I_-) \\ = & \sum_{i=0}^{k-1} \left(P(L(\mathbf{X}, I_{i,\varepsilon}) \in I_-) - P(L(\mathbf{X}, I_{i+1,\varepsilon} \in I_-)) \right). \end{aligned}$$

Thus in order to prove the result of the lemma it suffices to prove it for one step in the path. Without loss of generality, let us look at the first step $I_{0,\varepsilon}$ to $I_{1,\varepsilon}$ and prove

$$\lim_{\varepsilon \to 0} \left(P(L(\mathbf{X}, I_{0,\varepsilon}) \in I_-) - P(L(\mathbf{X}, I_{1,\varepsilon}) \in I_-) \right) = 0. \tag{6.13}$$

According to the value of the coordinate $\sigma(n-k+1)$, the one being expanded in this step, we divide $I_{1,\varepsilon}$ into four parts:

$$D_1 := \{t \in I_{1,\varepsilon} : t_{\sigma(n-k+1)} \in (\varepsilon, b_{\sigma(n-k+1)} - \varepsilon)\},$$

$$D_2 := \{t \in I_{1,\varepsilon} : t_{\sigma(n-k+1)} = \varepsilon \text{ or } t_{\sigma(n-k+1)} = b_{\sigma(n-k+1)} - \varepsilon\},$$

$$D_3 := \{t \in I_{1,\varepsilon} : t_{\sigma(n-k+1)} \in (-\varepsilon, \varepsilon) \text{ or } t_{\sigma(n-k+1)} \in (b_{\sigma(n-k+1)} - \varepsilon, b_{\sigma(n-k+1)} + \varepsilon)\},$$

and

$$D_4 := \{t \in I_{1,\varepsilon} : t_{\sigma(n-k+1)} = -\varepsilon \text{ or } t_{\sigma(n-k+1)} = b_{\sigma(n-k+1)} + \varepsilon\}.$$

Thus $I_{0,\varepsilon} = D_1 \cup D_2$, $I_{1,\varepsilon} = D_1 \cup D_2 \cup D_3 \cup D_4$, and $I_- \subset D_1$. By Lemma 6.3.1 (ii),

$$\begin{aligned} & P(L(\mathbf{X}, I_{0,\varepsilon}) \in I_-) - P(L(\mathbf{X}, I_{1,\varepsilon}) \in I_-) \\ \leq\ & P(L(\mathbf{X}, I_{0,\varepsilon}) \in D_1) - P(L(\mathbf{X}, I_{1,\varepsilon}) \in D_1) \\ =\ & P(L(\mathbf{X}, I_{1,\varepsilon}) \in D_2) + P(L(\mathbf{X}, I_{1,\varepsilon}) \in D_3) + P(L(\mathbf{X}, I_{1,\varepsilon}) \in D_4) \\ +\ & P(L(\mathbf{X}, I_{1,\varepsilon}) = \infty) - P(L(\mathbf{X}, I_{0,\varepsilon}) \in D_2) - P(L(\mathbf{X}, I_{0,\varepsilon}) = \infty). \end{aligned}$$

By absolute continuity proved before, $P(L(\mathbf{X}, I_{1,\varepsilon}) \in D_2) = 0$, because D_2 is a null set in $I_{1,\varepsilon}$ and is not the boundary. Moreover, if we further divide D_2 and D_4 into two parts: $D_2 = D_{21} \cup D_{22}$ and $D_4 = D_{41} \cup D_{42}$, where

$$D_{21} := \{t \in I_{1,\varepsilon} : t_{\sigma(n-k+1)} = \varepsilon\},$$

$D_{22} = D_2 \backslash D_{21}$ and D_{41} and D_{42} are defined in the similar way, then it becomes clear that $D_{21} = D_{41} + 2\varepsilon e_{\sigma(k+1)}$ and $D_{22} = D_{42} - 2\varepsilon e_{\sigma(k+1)}$, where e_i is the unit vector on coordinate i. Again by Lemma 6.3.1 (ii), $P(L(\mathbf{X}, I_{1,\varepsilon}) \in D_{4i}) \leq P(L(\mathbf{X}, I_{0,\varepsilon}) \in D_{2i})$ for $i = 1, 2$. As a result, $P(L(\mathbf{X}, I_{1,\varepsilon}) \in D_4) \leq P(L(\mathbf{X}, I_{0,\varepsilon}) \in D_2)$. Moreover, by Lemma 6.3.1 (iii) $P(L(\mathbf{X}, I_{1,\varepsilon}) = \infty) - P(L(\mathbf{X}, I_{0,\varepsilon}) = \infty) < 0$. Hence we have

$$P(L(\mathbf{X}, I_{0,\varepsilon}) \in I_-) - P(L(\mathbf{X}, I_{1,\varepsilon}) \in I_-) \leq P(L(\mathbf{X}, I_{1,\varepsilon}) \in D_3).$$

However, $m^k(D_3 \cap I^k_{1,\varepsilon}) \to 0$ as $\varepsilon \to 0$ for any k, where $I^k_{1,\varepsilon}$ is the union of the k-dimensional open faces of $I_{1,\varepsilon}$, defined in the same way as I^k. Thus by absolute continuity, $\lim_{\varepsilon \to 0} P(L(\mathbf{X}, I_{1,\varepsilon}) \in D_3) = 0$. (6.13) is proved. $\square$

The contradiction between (6.12) and Lemma 6.3.3 shows that the assumption $m^k(A^k_h \cap I^k_{(\bar{\sigma},m,\delta)}) > 0$ will not hold for any k, h, m and $\bar{\sigma}$, thus Proposition 6.3.3 is proved.

$\square$

We are now ready to construct the density function $f^{k,j}_{\mathbf{X},I}$ for a given direction $j = 1, ..., n$. Recall that $I^{k,j}$ and $I^{k,\hat{j}}$ are respectively the union of k-dimensional faces of I for which the coordinate j is free and is not free, and I^k_j is the orthogonal projection of $I^{k,j}$ on the subspace $t_j = 0$. For any $s \in I^k_j$, let $A^{k,j} = A^k \cap I^{k,j}$ and let $A^{k,j}|_s = \{t \in A^{k,j} : p^j t = s\}$, the section of A^k on the one dimensional subspace $p^j t = s$. Let set $G^{k,j} := \{s \in I^k_j : m(A^{k,j}|_s) = b_j\}$, that is, the set of the one-dimensional subspaces along direction j, on which A^k hold almost surely. Since $m^k(I^k \backslash A^k) = 0$, we know $m^{k-1}(I^k_j \backslash G^{k,j}) = 0$. Therefore the set $E^{k,j}_0 := \{t \in I^k : p^j t \notin G^{k,j}\}$ is a null set, and we can define the density function on $E^{k,j}_0$ arbitrarily, as long as it is upper semi-continuous on t_j and satisfies (6.1). For instance, the constant 0 is a valid choice. On $E^{k,j}_1 := (E^{k,j}_0)^c \cap A^{k,j}$, define $f^{k,j}_{\mathbf{X},I}(t) = \lim_{r \to 0} \frac{1}{m^k(B^k(0,r))} \int_{B^k(t,r)} f^k_{\mathbf{X},I}(s) ds$, which exists since $A^k \subseteq A^{k'}$. Finally, on $E^{k,j}_2 := I^{k,j} \backslash (E^{k,j}_0 \cup E^{k,j}_1)$, define

$$f^{k,j}_{\mathbf{X},I}(t) = \limsup_{s \to t, s \in E^{k,j}_1, p^j(s-t)=0} f^{k,j}_{\mathbf{X},I}(s).$$

Now it is easy check that such defined $f^{k,j}_{\mathbf{X},I}$ is upper semi-continuous and almost everywhere continuous on the coordinate t_j. Clearly there is no problem for $t \in E^{k,j}_0$. For $t \in E^{k,j}_1$, by continuity on A^k, for any $\varepsilon > 0$, there exists $\delta > 0$, such that for any $s \in E^{k,j}_1$ satisfying $p^j s = p^j t$ and $s_j \in (t_j - \delta, t_j + \delta)$, $|f^{k,j}_{\mathbf{X},I}(s) - f^{k,j}_{\mathbf{X},I}(t)| \leq \varepsilon$. Then for any $s \in E^{k,j}_2$ satisfying $p^j s = p^j t$ and $s_j \in (t_j - \delta, t_j + \delta)$, since the value of the density function is defined as the limsup, it also satisfies $|f^{k,j}_{\mathbf{X},I}(s) - f^{k,j}_{\mathbf{X},I}(t)| \leq \varepsilon$. Thus the density $f^{k,j}_{\mathbf{X},I}$ is continuous for any $t \in E^{k,j}_1$ on t_j. By similar argument, we can check that $f^{k,j}_{\mathbf{X},I}$ is upper semi-continuous on set $E^{k,j}_2$ on t_j. Since $E^{k,j}_2$ is a null set, it follows that $f^{k,j}_{\mathbf{X},I}$ is everywhere upper semi-continuous and almost everywhere continuous on the coordinate t_j, which is part (a) of the theorem. Indeed, we have

$$f^{k,j}_{\mathbf{X},I}(t) = \limsup_{s\to t, p^j(s-t)=0} f^{k,j}_{\mathbf{X},I}(s)$$

for every $t \in I^k$. This regularity condition implies that for this version of density, bound (6.1) is not only satisfied almost everywhere but actually everywhere. Hence the part (c) of the theorem is also proved.

Now we address the variation of the version of density function $f^{k,j}_{\mathbf{X},I}$. Actually we only need to prove the following lemma, which can be regarded as a higher dimensional generalization of Lemma 5.3.2.

Lemma 6.3.4. *Let* $0 \le \Delta < b_j$. *Define* $I' = \{t \in I : t_j \le b_j - \Delta\}$. *Then for every* $0 \le \delta \le \Delta$, $f^{k,j}_{\mathbf{X},I'}(t) \ge f^{k,j}_{\mathbf{X},I}(t+\delta e_j)$ *almost everywhere on* $\{t \in I' : t_j \in (0, b_j - \Delta)\}$ *with respect to the* k*-dimensional Lebesgue measure. Furthermore, for every such* δ *and every* $\varepsilon_1, \varepsilon_2 \ge 0$, *such that* $\varepsilon_1 + \varepsilon_2 < b_j - \Delta$,

$$\sum_{k=1}^{n} \int_{D_1\cap I^{k,j}} \left(f^{k,j}_{\mathbf{X},I'}(t) - f^{k,j}_{\mathbf{X},I}(t+\delta e_j)\right) m^k(dt) \tag{6.14}$$

$$\le \sum_{k=1}^{n} \int_{D_2\cap I^{k,j}} f^{k,j}_{\mathbf{X},I}(t) m^k(dt) + \sum_{k=1}^{n} \int_{D_3\cap I^{k,j}} f^{k,j}_{\mathbf{X},I}(t) m^k(dt),$$

where $D_1 = \{t \in I : t_j \in (\varepsilon_1, b_j - \Delta - \varepsilon_2)\}$, $D_2 = \{t \in I : t_j \in (\varepsilon_1, \varepsilon_1 + \delta)\}$, *and* $D_3 = \{t \in I : t_j \in (b_j - \Delta - \varepsilon_2 + \delta, b_j - \varepsilon_2)\}$.

Proof. The comparison part $f^{k,j}_{\mathbf{X},I'}(t) \ge f^{k,j}_{\mathbf{X},I}(t+\delta e_j)$ follows directly from part (ii) of Lemma 6.3.1. For (6.14), notice that

$$\sum_{k=1}^{n} \int_{D_1\cap I^{k,j}} \left(f^{k,j}_{\mathbf{X},I'}(t) - f^{k,j}_{\mathbf{X},I}(t+\delta e_j)\right) m^k(dt)$$

$$= P\left(L(\mathbf{X}, I')_j \in (\varepsilon_1, b_j - \Delta - \varepsilon_2)\right) - P\left(L(\mathbf{X}, I)_j \in (\varepsilon_1 + \delta, b_j - \Delta - \varepsilon_2 + \delta)\right)$$

$$= P\left(L(\mathbf{X}, I)_j \notin (\varepsilon_1 + \delta, b_j - \Delta - \varepsilon_2 + \delta)\right) - P\left(L(\mathbf{X}, I')_j \notin (\varepsilon_1, b_j - \Delta - \varepsilon_2)\right)$$

$$
\begin{aligned}
&= P\left(L(\mathbf{X},I)_j \in [0,\varepsilon_1+\delta)\right) + P\left(L(\mathbf{X},I)_j \in (b_j-\Delta-\varepsilon_2+\delta, b_j]\right) \\
&\quad + P(L(\mathbf{X},I)=\infty) - P\left(L(\mathbf{X},I')_j \in [0,\varepsilon_1)\right) \\
&\quad - P\left(L(\mathbf{X},I')_j \in (b_j-\Delta-\varepsilon_2, b_j-\Delta]\right) - P(L(\mathbf{X},I')=\infty) \\
&= P\left(L(\mathbf{X},I)_j \in (\varepsilon_1,\varepsilon_1+\delta)\right) + P\left(L(\mathbf{X},I)_j \in (b_j-\Delta-\varepsilon_2+\delta, b_j-\varepsilon_2)\right) \\
&\quad + \left(P\left(L(\mathbf{X},I)_j \in [0,\varepsilon_1)\right) - P\left(L(\mathbf{X},I')_j \in [0,\varepsilon_1)\right)\right) \\
&\quad + \left(P\left(L(\mathbf{X},I)_j \in (b_j-\varepsilon_2,b_j]\right) - P\left(L(\mathbf{X},I'+\Delta e_j)_j \in (b_j-\varepsilon_2,b_j]\right)\right) \\
&\quad + \left(P\left(L(\mathbf{X},I)=\infty\right) - P\left(L(\mathbf{X},I')=\infty\right)\right).
\end{aligned}
$$

By Lemma 6.3.1 (ii) we know

$$\left(P\left(L(\mathbf{X},I)_j \in [0,\varepsilon_1)\right) - P\left(L(\mathbf{X},I')_j \in [0,\varepsilon_1)\right)\right) \leq 0$$

and

$$\left(P\left(L(\mathbf{X},I)_j \in (b_j-\varepsilon_2,b_j]\right) - P\left(L(\mathbf{X},I'+\Delta e_j)_j \in (b_j-\varepsilon_2,b_j]\right)\right) \leq 0,$$

while by Lemma 6.3.1 (iii),

$$\left(P\left(L(\mathbf{X},I)=\infty\right) - P\left(L(\mathbf{X},I')=\infty\right)\right) \leq 0.$$

Thus we are left with

$$
\begin{aligned}
&\sum_{k=1}^{n} \int_{D_1 \cap I^{k,j}} \left(f^{k,j}_{\mathbf{X},I'}(t) - f^{k,j}_{\mathbf{X},I}(t+\delta e_j)\right) m^k(dt) \\
&\leq P\left(L(\mathbf{X},I)_j \in (\varepsilon_1,\varepsilon_1+\delta)\right) + P\left(L(\mathbf{X},I)_j \in (b_j-\Delta-\varepsilon_2+\delta, b_j-\varepsilon_2)\right) \\
&= \sum_{k=1}^{n} \int_{D_2 \cap I^{k,j}} f^{k,j}_{\mathbf{X},I}(t) m^k(dt) + \sum_{k=1}^{n} \int_{D_3 \cap I^{k,j}} f^{k,j}_{\mathbf{X},I}(t) m^k(dt),
\end{aligned}
$$

as required. □

Once Lemma 6.3.4 is proved, by focusing on direction j, the rest of the proof of the theorem follows in almost exactly the same way as in the one dimensional case in Chapter 5 or, originally, as in Chapter 3. □

6.4 Isotropic random fields and angular total variation constraints

As we have seen, stationarity–the invariance of probability distribution under translation, is closely related to the total variation constraints of the density functions of intrinsic location functionals. There are, of course, many other symmetries besides the symmetry under translation, each one related to certain movement/operation. In particular, if we look at rotation and add the invariance under rotation to the list of assumptions, the result is then the well-known notion of isotropy. It is therefore interesting to see what will happen under isotropy for the distribution of the intrinsic location functionals. On the other hand, such an investigation only becomes possible recently, when we know enough about the higher dimensional case, since rotation only makes sense in dimension greater than or equal to 2. In this section, we will make a brief introduction to the distribution of the intrinsic location functionals of isotropic random fields.

Since isotropy takes stationarity as a premise, the distributions of the intrinsic location functionals of isotropic random fields automatically inherit all the properties that we proved for stationary random fields. The new properties come from the rotational invariance. To make life easier, in this part we will focus on the 2-dimensional case. The higher dimensional case are similar, but with much heavier notations and expressions.

Since we will deal with rotation, the most convenient region to consider is no longer an hypercube, but a cone in a ball. In particular, it becomes a sector in dimension 2. Thus $\mathcal{I}$ is no longer the set of all hypercubes, but becomes the set of all sectors. Using polar coordinate system, we can express a sector as $A = \{(r, \theta) : r \leq R, \theta \in [0, \Theta]\}$. Define the one dimensional faces $D_0 = \{(r, \theta) \in A : \theta = 0, r \in (0, R)\}$, $D_1 = \{(r, \theta) \in A : \theta = \Theta, r \in (0, R)\}$, $D_2 = \{(r, \theta) \in A : r = R, \theta \in (0, \Theta)\}$, and the two dimensional face, which is also the interior of the sector, $E = \{(r, \theta) \in A : r \in (0, R), \theta \in (0, \Theta)\}$. Define set $D = D_0 \cup D_1 \cup D_2$. Similar to the stationary case, denote by $F_{\mathbf{X},A}$ the probability distribution of $L(\mathbf{X}, A)$. For a function $f(r, \theta)$, define the "angular total variation" with radial parameter r and between two angles θ_1 and θ_2 as

$$TV^{\theta}_{(\theta_1,\theta_2)}(f|_r) := TV_{(\theta_1,\theta_2)}(f(r, \cdot)),$$

and the positive/negative "angular variation" as

$$TV^{\theta\pm}_{(\theta_1,\theta_2)}(f|_r) := TV^{\pm}_{(\theta_1,\theta_2)}(f(r, \cdot)).$$

We have the following result parallel to the stationary case:

Theorem 6.4.1. *Let L be an intrinsic location functional on some path space H, $\mathbf{X} = (X(t),\, t = (r, \theta))$ be an isotropic random field such that $\mathbb{P}(\mathbf{X} \in H) = 1$. Let $A = \{(r, \theta) : r \leq R, \theta \in [0, \Theta]\}$, D, E as defined before. Then the restriction of the law $F_{\mathbf{X},A}$ to D is absolutely continuous with respect to the 1-dimensional Lebesgue measure, and the restriction of $F_{\mathbf{X},A}$ to E is absolutely continuous with respect to the 2-dimensional Lebesgue measure. Moreover, there exists a version of densities of $F_{\mathbf{X},A}$ restricted to D and E in polar coordinate system, denoted by $f^D_{\mathbf{X},A}$ and $f^E_{\mathbf{X},A}$, which satisfy the following conditions.*

(a) $f^D_{\mathbf{X},A}$ and $f^E_{\mathbf{X},A}$ are upper semi-continuous and almost everywhere continuous on θ.

(b) $\forall r \in (0, R)$, the limits

$$f^{E,0+}_{\mathbf{X},A}(r) := \lim_{\theta \downarrow 0} f^E_{\mathbf{X},A}(r, \theta)$$

and

$$f_{\mathbf{X},A}^{E,\Theta-}(r) := \lim_{\theta\uparrow\Theta} f_{\mathbf{X},A}^{E}(r,\theta)$$

exist almost everywhere on $(0,R)$ with respect to the 1 dimensional Lebesgue measure, and the limits

$$f_{\mathbf{X},A}^{D,0+} := \lim_{\theta\downarrow 0} f_{\mathbf{X},A}^{D}(R,\theta)$$

and

$$f_{\mathbf{X},A}^{D,\Theta-} := \lim_{\theta\uparrow\Theta} f_{\mathbf{X},A}^{D}(R,\theta)$$

exist.

(c) For every $0<\theta_1<\theta_2<\Theta$,

$$\int_{(0,R)} TV_{(\theta_1,\theta_2)}^{\theta}(f_{\mathbf{X},A}^{E}|_r)dr + TV_{(\theta_1,\theta_2)}^{\theta}(f_{\mathbf{X},A}^{D}|_R) \tag{6.15}$$

$$\leq \int_{(0,R)} \left(f_{\mathbf{X},A}^{E}(r,\theta_1) + f_{\mathbf{X},A}^{E}(r,\theta_2)\right) dr + f_{\mathbf{X},A}^{D}(R,\theta_1) + f_{\mathbf{X},A}^{D}(R,\theta_2).$$

(d) The density $f_{\mathbf{X},A}^{E}$ has bounded positive angular variation almost everywhere at the side $\theta=0$ and a bounded negative angular variation almost everywhere at the side $\theta=\Theta$. The density $f_{\mathbf{X},A}^{D}$ has bounded positive angular variation at the side $\theta=0$ and a bounded negative angular variation at the side $\theta=\Theta$. Furthermore, for every $0<\varepsilon<\Theta$,

$$\int_0^R TV_{(0,\varepsilon)}^{\theta+}(f_{\mathbf{X},A}^{E}|_r)dr + TV_{(0,\varepsilon)}^{\theta+}(f_{\mathbf{X},A}^{D}|_R) \leq \int_0^R f_{\mathbf{X},A}^{E}(r,\varepsilon)dr + f_{\mathbf{X},A}^{D}(R,\varepsilon), \tag{6.16}$$

and

$$\int_0^R TV_{(\Theta-\varepsilon,\Theta)}^{\theta-}(f_{\mathbf{X},A}^{E}|_r)dr + TV_{(\Theta-\varepsilon,\Theta)}^{\theta-}(f_{\mathbf{X},A}^{D}|_R) \leq \int_0^R f_{\mathbf{X},A}^{E}(r,\Theta-\varepsilon)dr + f_{\mathbf{X},A}^{D}(R,\Theta-\varepsilon). \tag{6.17}$$

(e)

$$\int_0^R f_{\mathbf{X},A}^{E,0+}(r)dr + f_{\mathbf{X},A}^{D,0+} < \infty$$

if and only if

$$\int_0^R TV^{\theta}_{(0,\varepsilon)}(f^{E}_{\mathbf{X},A}|_r)dr + TV^{\theta}_{(0,\varepsilon)}(f^{D}_{\mathbf{X},A}|_R) < \infty$$

for some (equivalently, any) $0 < \varepsilon < \Theta$*, in which case*

$$\begin{aligned} &\int_0^R TV^{\theta}_{(0,\varepsilon)}(f^{E}_{\mathbf{X},A}|_r)dr + TV^{\theta}_{(0,\varepsilon)}(f^{D}_{\mathbf{X},A}|_R) \\ \leq &\int_0^R \left(f^{E,0+}_{\mathbf{X},A}(r) + f^{E}_{\mathbf{X},A}(r,\varepsilon)\right) dr + f^{D,0+}_{\mathbf{X},A} + f^{D}_{\mathbf{X},A}(R,\varepsilon). \end{aligned} \tag{6.18}$$

Similarly,

$$\int_0^R f^{E,\Theta-}_{\mathbf{X},A}(r)dr + f^{D,\Theta-}_{\mathbf{X},A} < \infty$$

if and only if

$$\int_0^R TV^{\theta}_{(\Theta-\varepsilon,\Theta)}(f^{E}_{\mathbf{X},A}|_r)dr + TV^{\theta}_{(\Theta-\varepsilon,\Theta)}(f^{D}_{\mathbf{X},A}|_R) < \infty$$

for some (equivalently, any) $0 < \varepsilon < T$*, in which case*

$$\begin{aligned} &\textstyle\int_0^R TV^{\theta}_{(\Theta-\varepsilon\Theta)}(f^{E}_{\mathbf{X},A}|_r)dr + TV^{\theta}_{(\Theta-\varepsilon,\Theta)}(f^{D}_{\mathbf{X},A}|_R) \\ \leq \quad &\textstyle\int_0^R f^{E,\Theta-}_{\mathbf{X},A}(r) + f^{E}_{\mathbf{X},A}(r,\Theta-\varepsilon)dr + f^{D,\Theta-}_{\mathbf{X},A} + f^{D}_{\mathbf{X},A}(R,\Theta-\varepsilon). \end{aligned} \tag{6.19}$$

Due to the limited space and the fact that the proof of Theorem 6.4.1 proceeds in a similar way as the proof of Theorem 6.3.1, we will not prove Theorem 6.4.1 in detail, but will only give a proof sketch here.

The absolute continuity results on E, D_0 and D_1 can be achieved by bounding the probabilities from above by the corresponding probabilities in smaller rectangular areas, the absolute continuity of which is assured by Theorem 6.3.1. More precisely, let set $E_{-\varepsilon} = \{t \in E : d(t, D) \geq \varepsilon\}$ be the closed set of all points in E which is at least ε away from the boundary of E. Then there exists a finite open cover of $E_{-\varepsilon}$ by open rectangles which are subsets of E. Denote one family of such rectangles by $\{I_i\}_{i=1,\ldots,N}$. Lemma 6.3.1 (ii) implies that $F_{\mathbf{X},A}$ is dominated by $F_{\mathbf{X},I_i}$ on I_i, while $F_{\mathbf{X},I_i}$ is absolutely continuous on I_i by Theorem 6.3.1. Thus $F_{\mathbf{X},A}$ is absolutely continuous on each I_i, so it is absolutely continuous on the whole set $E_{-\varepsilon}$. Taking union of $E_{-\varepsilon}$ with a sequence of $\varepsilon \to 0$ shows that $F_{\mathbf{X},A}$ is absolute continuous on E. For D_0, take $0 < r_1 < r_2 < R$. For h small enough, the rectangle I, represented in Cartesian coordinate system as $I = [r_1, r_2] \times [0, h]$, is a subset of the sector A. By Theorem 6.3.1, $F_{\mathbf{X},I}$ is absolutely continuous on the interval $\{(x, 0) : x \in (r_1, r_2)\}$ with respect to 1-dimensional Lebesgue measure. Thus $F_{\mathbf{X},A}$ must also be absolute continuous on this interval. Taking union on $r_1 \to 0$ and $r_2 \to R$ gives the (one-dimensional) absolute continuity on D_0. The absolute continuity on D_1 follows immediately by isotropy and symmetry.

The proof of the rest of Theorem 6.4.1, including the absolute continuity on D_2 and the angular total variation constraints, relies on the following transformation. Take $x = \theta$, $y = r$ and redraw $\mathbf{X}'(x, y) = \mathbf{X}(r, \theta)$ on 2-dimensional Cartesian coordinate system. Then the fan shaped areas $\{(r, \theta) : r \in (r_1, r_2), \theta \in (\theta_1, \theta_2)\}$ are transformed into rectangles. Clearly, the transformed random field $\mathbf{X}'$ is no longer stationary. However, it remains stationary on direction x, and this directional stationarity is already enough for all of our purposes. A minor issue arises when we construct the version of density function having continuity properties on direction x: since we only have one directional stationarity, it is not clear that the set A^k defined before will still be almost sure on D or E, since A^k requires continuity on all directions, the proof of which then requires the stationarity on all directions. Nevertheless, we can replace A^k by a new set $A^{k,x}$, which is defined in the similar way as A^k, but only requires continuity on direction x. Following a similar way as in the proof of Theorem 6.3.1, it is not hard to check that $A^{k,x}$ is almost everywhere. All the construction procedures then can be carried out as before, using $A^{k,x}$ instead of A^k. At last, the proof of the total variation constraints on direction x remains intact.

CHAPTER 7

RANDOM LOCATIONS OF STATIONARY INCREMENT PROCESSES

7.1 Introduction

In Chapter 5, we defined a class of locations called "intrinsic location functionals". This is a large class which includes most commonly used random locations, such as the location of the path supremum over an interval, the first/last hitting time to a level over an interval, etc. It is shown that if the process is stationary, then the concept of intrinsic location functional provides a structural and unified way to understand the distributions of the random locations of the process over any interval. More precisely, the distribution of any intrinsic location functional must be absolutely continuous in the interior of the interval, and the density function satisfies the so-called "total variation constraints". This implies that the set of all possible distributions for any given intrinsic location functional will be convex, which in turn automatically sets bounds for related expectations. A higher dimensional extension of this result, along with its counterpart for isotropic case, is stated and proved in Chapter 6. It is also discovered in Chapter 5 that the total variation constraint is not merely a property coming from stationarity; instead, the total variation constraints for all the intrinsic location functionals are actually enough to guarantee the stationarity. In this sense, the total variation constraint is indeed the stationarity itself, viewed from a different angle. Briefly speaking, there is equivalence between three things: the stationarity of the process, the invariance under translation of the distributions of intrinsic location functionals, and the total variation constraints of the intrinsic location functionals. It reveals a deep relationship between stationarity and total variation.

As a relative of stationary process, stationary increment process also exhibits certain stationarity, but only for the increments instead of the process itself. It then becomes interesting to ask what we can say about the distribution of random locations of stationary increment processes, and whether there is also some equivalence between the stationary increment property and total variation constraints of certain random locations. These are the questions that we try to answer in this chapter. Since the class of stationary increment processes is much larger than the class of stationary processes, it is not realistic to expect that the total variation constraints still hold for all intrinsic location functionals. Instead, in this chapter we propose another notion "doubly intrinsic location functionals", as a subclass of intrinsic location functionals. Roughly speaking, a doubly intrinsic location functional is an intrinsic location functional whose value does not change under vertical shift of the path. This concept is extracted out from the random locations such as the location of the path supremum/infimum over an interval, the first/last hitting time of the derivative for $\mathcal{C}^1$ pathes, among others. Interestingly, it can be proved that to stationary increment processes, doubly intrinsic location functionals play the same role as intrinsic location functionals to stationary processes. There is an equivalence between the stationarity of increments, the invariance under translation of the distributions of doubly intrinsic location functionals, and the total variation constraints of doubly intrinsic location functionals.

The rest of this chapter is organized in the following way. In Section 7.2 we introduce the notion of doubly intrinsic location functional and see examples of it. Section 7.3 is dedicated to formulating, explaining and then proving the main theorem of equivalence. Section 7.4 extends the results to higher dimensional cases.

7.2 Notation and definition

Firstly let us recall the basic notations. Let H be a space of functions from $\mathbb{R}$ to $\mathbb{R}$ satisfying certain properties, *e.g.*, continuous functions, càdlàg functions, or more generally, upper semi-continuous functions, *etc*. On H we have the (left) shift operator θ_c for any $c \in \mathbb{R}$:

$$\theta_c f(x) := f(x+c), \quad \forall x \in \mathbb{R}.$$

Let $\mathcal{I}$ be the collection of all the compact intervals on $\mathbb{R}$. For any $I \in \mathcal{I}$ and any $c \in \mathbb{R}$, denote by $I - c$ the interval shifted by c:

$$I - c := \{x - c : x \in I\}.$$

In Chapter 5 we defined the intrinsic location functionals. Here, in order to benefit from the vertical shift invariance property possessed by a subclass of intrinsic location functionals such as the location of the supremum/infimum over an interval, we add the vertical shift invariance to the definition to form the new notion of "doubly intrinsic location functional".

Definition 4. *An intrinsic location functional L is called doubly intrinsic, if for every function $f \in H$, every interval $I \in \mathcal{I}$ and every $c \in \mathbb{R}$,*

$$L(f, I) = L(f + c, I).$$

Denote by $\mathcal{D}$ the set of all doubly intrinsic location functionals defined on H.

The word "doubly" in the name refers to the fact that L is both "horizontally shift invariant", in the sense that it moves along with the function and interval horizontally, and "vertically shift invariant", in the sense that it does not move along with the function vertically.

In general, once we verify that certain location is an intrinsic location functional, it is very easy to check whether it is doubly intrinsic or not. Intuitively, an intrinsic location functional is doubly intrinsic if and only if its value only depends on the "shape" of the function and does not depend on the "height" of the function. Here are some most natural and important examples of doubly intrinsic location functionals.

Example 7.2.1. Let H be the space of all the upper (lower) semi-continuous functions. Then the location of the supremum (infimum) over an interval

$$\tau_{f,I} := \inf\{t \in I : f(t) = \sup(\inf)_{s\in I} f(s)\}$$

is a doubly intrinsic location functional. The infimum outside means that in case of a tie, we always chose the leftmost point among all the points achieving the path supremum (infimum).

Example 7.2.2. Let H be the space of all càdlàg functions. Then the starting point of the largest shortfall during a short period

$$S^d_{f,[a,b]} := \inf\{s : s \in [a,b], t \in (s, s+d], f(s) - f(t) = \sup_{s'\in[a,b], t'\in(s,s+d]} (f(s') - f(t'))\}$$

is a doubly intrinsic location functional.

Needless to say, any random location which only depends on the value of the first derivative of $\mathcal{C}^1$ functions is also doubly intrinsic. For instance, the location of the first local maxima, the first time that the derivative hits certain level, *etc*.

The class of doubly intrinsic location functionals extends, however, far beyond these "natural" examples. Actually, let H, H' be two spaces of functions, and φ be a mapping from H to H' which is interchangeable with translation:

$$\forall f \in H, \forall c \in \mathbb{R}, \varphi(\theta_c f) = \theta_c(\varphi f) \tag{7.1}$$

and consistent with vertical shift:

$$\forall f \in H, \forall c \in \mathbb{R}, \exists c' \in \mathbb{R}, \varphi(f + c) = \varphi(f) + c'. \tag{7.2}$$

If L' is a doubly intrinsic location functional in $H' \times \mathcal{I}$, then the functional L on $H \times \mathcal{I}$, defined by

$$L(f, I) := L'(\varphi f, I), \quad \forall f \in H, \forall I \in \mathcal{I},$$

is also a doubly intrinsic location functional, provided that the measurability condition is satisfied. We call it the doubly intrinsic location functional induced by φ. This procedure allows us to associate random locations which are originally only well-defined for "nice" functions to the functions which does not possess the required properties. The transforms satisfying (7.1) and (7.2) include many commonly used operations such as convolution, differentiation, moving average, moving difference, *etc*.

Example 7.2.3. Let ψ be the classical mollifier:

$$\psi(x) = \begin{cases} e^{-1/(1-|x|^2)} & \text{if } |x| < 1 \\ 0 & \text{if } |x| \geq 1 \end{cases},$$

then the operation of convolution with ψ transforms any measurable function to smooth function. That is, let f be any measurable function, then $f * \psi$ is a smooth function, where "$*$" denotes convolution. This convolution is obviously interchangeable with translation. It is easy to see that the location of the first hitting time of the derivative to level h over an interval:

$$L'(g, I) := \inf\{t \in I : g'(t) = h\}$$

(following the tradition that $\inf(\sup)\phi = \infty$) is a doubly intrinsic location functional on the space of all smooth functions. If, moreover, the space H is an LI set as defined in Chapter 5, then similar reasoning as(5.29) guarantees the measurability issue for the induced location functional

$$L(f, I) := L'(f * \psi, I).$$

Thus L is also a doubly intrinsic location functional, now defined on any LI set. The doubly intrinsic location functionals of this kind will play an important role later in the proof of the main theorem.

7.3 Doubly intrinsic location functionals and stationary increment processes

In this section we state the main theorem and prove it. To assure the measurability, in this section we assume that the path space H is an *LI* set as defined in Chapter 5. Again, this is a quite mild condition, which is satisfied by the spaces such as the space of continuous functions or the space of càdlàg functions. It is not difficult to see that this assumption will guarantee that the mapping $f \to f * \phi$ is $\mathcal{C}_H/\mathcal{C}_{C(\mathbb{R})}$ measurable.

Theorem 7.3.1. *Let* $\mathbf{X}$ *be a stochastic process having path in* H *with probability 1. Then the followings are equivalent.*

1. *The process $\mathbf{X}$ is of stationary increments.*

2. *For some (equivalently, any) $\Delta > 0$, any doubly intrinsic location functional $L : H \times \mathcal{I} \to \mathbb{R} \cup \{\infty\}$, the law of $L(\mathbf{X}, I) - a$, $I = [a, a + \Delta] \in \mathcal{I}$, does not depend on a.*

3. *For any doubly intrinsic location functional $L : H \times \mathcal{I} \to \mathbb{R} \cup \{\infty\}$, any interval $I = [a, b] \in \mathcal{I}$, The law of $L(\mathbf{X}, I)$ is absolutely continuous on (a, b) and has a density satisfying the total variation constraints* (5.2)*,* (5.3)*,* (5.4)*,* (5.5) *and* (5.6)*.*

Similar to the case of intrinsic location functionals and stationary processes, this theorem shows that there is a deep and fundamental relationship between the stationarity of increments, the shift invariance of the distributions of doubly intrinsic locations, and the total variation constraints. The most surprising part is that the total variation constraints alone are enough to imply the stationarity of increments, even there is no distributional invariance explicitly formulated at all. Intuitively, it seems to be totally possible that all the doubly intrinsic location functionals always satisfy the total variation constraints, yet their distributions change over different period. This theorem, however, tells us that this will never happen. The total variation constraints automatically lead to the distributional invariance under translation. It could be the case that for some doubly intrinsic location functional, its distribution varies over time while always keeping the total variation constraints true; but then there must be some other doubly intrinsic location functional, for which the total variation constraints are violated. As a family of random locations, the doubly intrinsic location functional is rich enough such that the total variation constraints on this family provide enough information to guarantee the stationarity of the increment of the process.

It is also interesting to make a comparison between Theorem 7.3.1 and its stationary counterpart, Theorem 5.5.1. In each of these cases, we have two spaces: the space of processes and the space of location functionals. In Theorem 5.5.1, the space of processes is the stationary processes, and the corresponding space of location functionals is the intrinsic location functionals. The two spaces are related one to each other via the total variation constrains. In this sense, the total variation constraints introduces a "duality" between the space of processes and the space of random locations. In Theorem 7.3.1, the space of processes becomes the stationary increment processes. Notice that since stationary processes are automatically of stationary increments but the converse is not true, the space of stationary increment processes is strictly larger than the space of stationary processes. Therefore we should expect a smaller space of the locations on the other side of the duality. It is indeed the case here, since doubly intrinsic location functionals is by definition a proper subset of intrinsic location functionals. In conclusion, Theorem 7.3.1 and Theorem 5.5.1 have the same nature, with different sizes of the sets on both sides of the duality.

Now let us turn to the proof of Theorem 7.3.1. The proof actually highly resembles the corresponding proofs in Chapter 5. The full proof will have four directions: $(1) \to (2)$, $(1) \to (3)$, $(2) \to (1)$ and $(3) \to (1)$. Given the fact that the proofs for some directions are very long, we will not include everything in the proof below, but will refer to the same proofs in Chapter 5 when it is possible. Many lemmas and settings, however, require changes and reverification.

First of all, notice the following lemma:

Lemma 7.3.1. *Let $\mathbf{X}$ be a stationary increment process with paths in H almost surely. Let $L \in \mathcal{D}$ and denote by $F_{\mathbf{X},I}(\cdot)$ the distribution of $L(\mathbf{X}, I)$. Then*

(i) *For any* $\Delta \in \mathbb{R}$,

$$F_{\mathbf{X},[\Delta,T+\Delta]}(\cdot) = F_{\mathbf{X},[0,T]}(\cdot - \Delta)\,.$$

(ii) *For any intervals* $[c,d] \subseteq [a,b]$,

$$F_{\mathbf{X},[a,b]}(B) \leq F_{\mathbf{X},[c,d]}(B) \textit{ for any Borel set } B \subset [c,d].$$

(iii) *For any intervals* $[c,d] \subseteq [a,b]$,

$$F_{\mathbf{X},[a,b]}(\{\infty\}) \leq F_{\mathbf{X},[c,d]}(\{\infty\}).$$

Proof. The point (ii) and (iii) are direct results of the stability under restriction and the consistency of existence in the definition of intrinsic location functionals, respectively. For (i), define a new process $\mathbf{Y}$ by $\mathbf{Y}(t) := \mathbf{X}(t) - \mathbf{X}(\Delta) + \mathbf{X}(0), t \in \mathbb{R}$, then the stationarity of the increments implies that the process $\mathbf{Y}(\cdot + \Delta)$ has the same distribution as $\mathbf{X}(\cdot)$. Thus

$$F_{\mathbf{X},[\Delta,T+\Delta]}(\cdot) = F_{\mathbf{Y},[0,T]}(\cdot).$$

Although $\mathbf{Y}(t) - \mathbf{X}(t) = \mathbf{X}(0) - \mathbf{X}(\Delta)$ is random and depends on the realization, it is a constant over time. Thus

$$L(\mathbf{X},[0,T]) = L(\mathbf{Y},[0,T]),$$

hence

$$F_{\mathbf{X},[0,T]}(\cdot) = F_{\mathbf{Y},[0,T]}(\cdot).$$

□

The rest of the proof in the direction $(1) \to (2)$ and $(1) \to (3)$ follows in the same way as in Chapter 5.

To prove that $(2) \to (1)$, consider the following location functional:

$$G_{\mathbf{t},\mathbf{I}}(\mathbf{X}, [a, a + \Delta]) := \inf\{t \in [a, a + \Delta] : t \in S(\mathbf{X}, \mathbf{t}, \mathbf{I})\},$$

where the random set of points S is defined by

$$S(\mathbf{X}, \mathbf{t}, \mathbf{I}) := \{t \in \mathbb{R} : \mathbf{X}(t + t_i) - \mathbf{X}(t) \in I_i, \forall i = 1, ..., n\},$$

n is a positive integer, $\mathbf{t} = (t_1, ..., t_n)$ such that $0 < t_1 < ... < t_n$, and $\mathbf{I} = I_1 \times ... \times I_n \in \mathcal{I}^n$. It is then easy to check that such defined $G_{\mathbf{t},\mathbf{I}}$ is a doubly intrinsic location functional for any $n = 1, 2, ...$, any $\mathbf{t}$ and $\mathbf{I}$. Moreover, $G_{\mathbf{t},\mathbf{I}}(\mathbf{X}, [a, a + \Delta]) = a$ if and only if

$$\mathbf{X}(a + t_i) - \mathbf{X}(a) \in I_i, \forall i = 1, ..., n.$$

If the distribution of $G_{\mathbf{t},\mathbf{I}}$ does not change with a, the probability that $\mathbf{X}(a + t_i) - \mathbf{X}(a) \in I_i, \forall i = 1, ..., n.$ can not change over a. Since this shift invariance holds for all $n, \mathbf{t}$ and $\mathbf{I}$, the stationarity of the increments is guaranteed.

We are now left with the proof that $(3) \to (1)$. The main object that we are going to consider are the doubly intrinsic location functionals of the type of Example 7.2.3, but slightly more complicated. More precisely, let the function ψ as defined in Example 7.2.3. Define process $\mathbf{Y} := \mathbf{X} * \psi$, then $\mathbf{Y}$ is a stationary increment process with smooth path. Consequently, $\mathbf{Z} = \mathbf{Y}'$, the derivative of $\mathbf{Y}$, is a smooth stationary process. For any $n = 1, 2, ...$, any $h > 0$, $d \geq 0$, any $\mathbf{t} = (t_0, t_1, ..., t_n)$ such that $0 < t_0 < t_1 < ... < t_n$ and any $\mathbf{I} = I_1 \times ... \times I_n \in \mathcal{I}^n$, define the random set of points

$$A^{h,d}_{\mathbf{t},\mathbf{I}}(\mathbf{X}) = \{s \in \mathbb{R} : \mathbf{Z}(s) = h, \inf\{r > s : \mathbf{Z}(r) = h\} > t + d,$$

$$\mathbf{X}(s + t_i) - \mathbf{X}(s + t_0) \in I_i, \forall i = 1, ..., n\}.$$

Notice that the LI setting guarantees the measurability. This set seems to be a little strange at the first glance, since the points are marked according to the process $\mathbf{Z}$, but then filtered using conditions on the original process $\mathbf{X}$. However, since $\mathbf{Z}$ is transformed from $\mathbf{X}$ and both the operation of convolution and differentiation are interchangeable with translation, the location

$$L(\mathbf{X}, I) := \inf\{t : t \in A^{h,d}_{\mathbf{t},\mathbf{I}}(\mathbf{X}) \cap I\}$$

is an intrinsic location functional. Moreover, since the points are marked on the derivative $\mathbf{Z}$ and then filtered using conditions only on the increments $\mathbf{X}(s+t_i) - \mathbf{X}(s+t_0)$, the location $L(\mathbf{X}, I)$ is invariant under vertical shift. Hence $L(\mathbf{X}, I)$ is a doubly intrinsic location functional. After defining

$$p^{h,d}_{\mathbf{t},\mathbf{I},a,\Delta}(\mathbf{X}) = \mathbb{P}(A^{h,d}_{\mathbf{t},\mathbf{I}}(\mathbf{X}) \cap [a, a+\Delta] \neq \phi), \tag{7.3}$$

we are totally back to the track of the proof for the stationary case (Theorem 5.5.1). Here we list the corresponding forms that the lemmas should take under the stationarity of increments.

Lemma 7.3.2. *Let* $\mathbf{X}$ *be a stochastic process. If condition (3) in Theorem 7.3.1 is satisfied, then for any* $h, d, \mathbf{t}$ *and* $\mathbf{I}$ *as defined before, with probability 1,* $A^{h,d}_{\mathbf{t},\mathbf{I}}(\mathbf{X})$ *is either the empty set or an infinite set, in which case* $\inf(A^{h,d}_{\mathbf{t},\mathbf{I}}(\mathbf{X})) = -\infty$ *and* $\sup(A^{h,d}_{\mathbf{t},\mathbf{I}}(\mathbf{X})) = \infty$.

Lemma 7.3.3. *Given* $h \in \mathbb{R}$*, if for any* $\Delta > 0$*,*$d \geq 2\Delta$*,* $\mathbf{t}$ *and* $\mathbf{I}$ *as defined above,* $p^{h,d}_{\mathbf{t},\mathbf{I},a,\Delta}(\mathbf{X})$ *is always constant on* a*, then the process* $\mathbf{X}$ *is of stationary increments.*

Lemma 7.3.4. *Assume that for any doubly intrinsic location functional* $L \in \mathcal{D}$*, any interval* $I \in \mathcal{I}$*,* $L(\mathbf{X}, I)$ *admits a density function* $f_{\mathbf{X},I}(t)$ *in* $\mathring{I}$*, which satisfies the total variation constraints on* I*. Then* $p^{h,d}_{\mathbf{t},\mathbf{I},a,\Delta}(\mathbf{X})$ *is constant on* a *for any* $\Delta > 0$*,* $d \geq 2\Delta$*,* $\mathbf{t}$ *and* $\mathbf{I}$ *as defined before.*

Lemma 7.3.2 gives us the right to decompose the path space and focus on only one given h. Lemma 7.3.4 and Lemma 7.3.3 then lead to the desired result in a straightforward way.

Just like Corollary 5.6.2 in the stationary case, there exists a stronger version of Theorem 7.3.1, which only uses a special class of doubly intrinsic location functionals to derive the stationarity of increments. Due to the high similarity between this result and Corollary 5.6.2, we will skip it here.

7.4 Higher dimensional cases

The result of Theorem 7.3.1 can be generalized to higher dimensional domain (random fields) in a way parallel to Chapter 6.

Both the notion of doubly intrinsic location functional and the definition of *LI* sets expands to higher dimension without any difficulty. All one needs to do is to replace the intervals to their natural counterparts–hypercubes, and to replace corresponding scalers to vectors. Denote by $\mathcal{I}^n$ the collection of compact hypercubes in $\mathbb{R}^n$.

The hypercube $\mathbf{I}$ can be then structurized as a stratified manifold, namely, a collection of different dimensional faces, plus a n-dimensional interior. For a fixed direction $j = 1, ..., n$ and a dimension k, denote by I_j^k the projection on the hyperplane $t_j = 0$ of the k-dimensional faces of I for which the jth coordinate t_j is free. Further let $I_{t_j=t}^k$ be the union of the $k-1$ dimensional faces of the hypercube $\mathbf{I}\cap\{\mathbf{t} : t_j = t\}$. The directional total variation of a function $f : \mathbb{R}^n \to \mathbb{R}$ on direction j, at the point $\mathrm{s} = (s_1, ..., s_{n-1})$ between two real numbers t_1 and t_2,

$t_1 < t_2$, is defined as

$$TV^j_{(t_1,t_2)}(f|_s) := TV_{(t_1,t_2)}(f(s_1, ..., s_{j-1}, \cdot, s_j, ..., s_{n-1})),$$

where $TV_{(t_1,t_2)}(f(\cdot))$ is the normal definition of one dimensional total variation of function f between t_1 and t_2.

The existence of a probability density function in one dimension now becomes the existence of a family of probability density functions with different dimensions on different dimensional faces, respectively. Then we say the (directional) total variation constraints are satisfied for $L(\mathbf{X}, \mathbf{I})$, if for any direction j, there exist versions of density functions of $L(\mathbf{X}, \mathbf{I})$ on the k dimensional faces for $k = 1, ..., n$, denoted by $f^{k,j}_{\mathbf{X},\mathbf{I}}$, such that

$$\sum_{k=1}^{n} \int_{I^k_j} TV^j_{(t_1,t_2)}(f^{k,j}_{\mathbf{X},\mathbf{I}}|_s) m^{k-1}(ds) \leq \sum_{k=1}^{n} \int_{I^k_{t_j=t_1} \cup I^k_{t_j=t_2}} f^{k,j}_{\mathbf{X},\mathbf{I}}(t) m^{k-1}(dt)\,, \tag{7.4}$$

where m^{k-1} is the $k-1$ dimensional Lebesgue measure.

Then we have the following equivalence, as the higher dimensional counterpart of Theorem 7.3.1.

Theorem 7.4.1. *Let $\mathbf{X}$ be a random field, with paths in an LI set H almost surely. Then the followings are equivalent.*

1. *The random field $\mathbf{X}$ is of stationary increments.*
2. *For some (equivalently, any) $\mathbf{\Delta} = (\Delta_1, ..., \Delta_n), \Delta_i > 0$ for all $i = 1, .., n$, any doubly intrinsic location functional $L : H \times \mathcal{I}^n \to \mathbb{R}^n \cup \{\infty\}$, the law of $L(\mathbf{X}, \mathbf{I}) - a$, $a = (a_1, ..., a_n) \in \mathbb{R}^n$, $\mathbf{I} = \prod_{i=1}^n [a_i, a_i + \Delta_i] \in \mathcal{I}^n$ does not depend on a.*
3. *For any doubly intrinsic location functional $L : H \times \mathcal{I}^n \to \mathbb{R}^n \cup \{\infty\}$, any $\mathbf{I} = I_1 \times ... \times I_n$ and any dimension $k = 1, ..., n$, the law of $L(\mathbf{X}, \mathbf{I})$ restricted*

to I^k *is absolutely continuous, and has a density satisfying the total variation constraints.*

CHAPTER 8

REPRESENTATION AND LOCALLY INTRINSIC LOCATION FUNCTIONALS

8.1 Introduction

In Chapter 5, we introduced a new notion called "intrinsic location functional", as an abstraction of the most commonly used random locations such as the location of the path supremum/infimum over an interval, the first/last hitting time, etc. It turns out that, despite the huge variety of the origins and natures of the random locations in this family, the common points that they share, now summarized in the definition of intrinsic location functional, are sufficient to guarantee many interesting and important properties of their distributions for stationary processes. The key properties are called "total variation constraints", since they put constraints on the total variation of the density functions. It was then proved that the total variation constraints of the intrinsic location functionals are not merely a group of properties of stationary processes: they are actually the stationarity itself, viewed from a different angle. On the other hand, there has not been many characterization and representation results discovered for the new object of intrinsic location functional. In this chapter, we will develop equivalent descriptions, as well as an important generalization, of intrinsic location functionals. These new results will be highly helpful for a better and more comprehensive understanding of this notion.

The rest of the chapter is organized in the following way. In Section 8.2, we introduce a generalization of intrinsic location functional called "local intrinsic location functional", which allows one to define a random location only for intervals with a single fixed length. Then we derive descriptions for it and also for intrinsic location functionals using partially ordered random point sets. The relation between local intrinsic location functional and intrinsic location functional is investigated in Section 8.3, showing that the former naturally inherits most of the properties of the latter. We provide yet another description in Section 8.4, which focuses on characterizing the value of a (local) intrinsic location functional as a function of the starting point of the interval of interest when the length of the interval is fixed.

8.2 Definition of local intrinsic location functional and representation by ordered set

The results proved in Chapter 5 showed how closely the concept of intrinsic location functional is related to stationarity. In some sense, the total variation constraints for intrinsic location functionals are just stationarity itself viewed from a different perspective. However, if one only considers the total variation constraint for intervals with a particular length, condition (4) and (5) in the definition of intrinsic location functional may appear a little bit too strong: in order to get the total variation constraint for the intervals with this length, one needs to introduce the relationships between intervals with different lengths. Therefore, it is interesting to check if we can adjust the definition of intrinsic location functional, while assuring that the total variation constraint hold for the intervals with a fixed length. It turns out that a reasonable way for this purpose is to define the following object, which we name "local intrinsic location functional".

Definition 5. *Fix $T > 0$. A mapping $L_T : H \times \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ is called a local intrinsic* *location functional with related length T, if it satisfies the following conditions.*

1. *For every $a \in \mathbb{R}$, the map $L_T(\cdot, a):\ H \to \mathbb{R} \cup \{\infty\}$ is measurable.*
2. *For every $f \in H$ and $a \in \mathbb{R}$, $L_T(f, a) \in [a, a+T] \cup \{\infty\}$.*
3. *For every $f \in H$, $a \in \mathbb{R}$ and $c \in \mathbb{R}$,*

   $$L_T(f, a) = L_T(\theta_c f, a-c) + c,$$

   where $\infty + c = \infty$.
4. *For every $f \in H$ and $a, b \in \mathbb{R}$, $L_T(f, a) \in [a, a+T] \cap [b, b+T]$ implies that either $L_T(f, b) = L_T(f, a)$, or $L_T(f, b) \in [b, b+T] \backslash [a, a+T]$.*

The first three conditions in this definition are the same as in the definition of intrinsic location functional. The condition (4) is new and replaces both condition (4) and (5) in Definition 1. Intuitively, it first requires that if the locations for two intervals with the same length both fall into the intersection of these two intervals, then they must agree. This is a counterpart of condition (4) (stability under restriction) in the definition of intrinsic location functional, but now only explicitly involving intervals with one fixed length. The second possibility in condition (4) says that if the location for the first interval is located in the second interval yet is no longer the corresponding location for the second interval, then it must be replaced by another point which is located in the second interval but outside the first interval. In particular, the corresponding location for the second interval can not take value ∞. In this sense, the second part of condition (4) actually serves as an alternative of condition (5) (consistency of existence) in the definition of intrinsic location functional.

It is not difficult to see that if we restrict the definition of an intrinsic location functional to intervals with a fixed length, then it automatically gives out a local intrinsic location functional:

Example 8.2.1. Let $L : H \times \mathcal{I} \to \mathbb{R} \cup \{\infty\}$ be an intrinsic location functional. Then it is easy to check that for any fixed length $T > 0$, L_T defined by

$$L_T(f, a) = L(f, [a, a + T]),$$

$f \in H, a \in \mathbb{R}$ is a local intrinsic location functional.

On the other hand, a natural "extension" of the definition of a local intrinsic location functional to intervals with different lengths does not necessarily give out an intrinsic location functional, as shown by the following example.

Example 8.2.2. Let $H = \mathcal{C}(\mathbb{R})$, $l > 0$, $L_T(f, a)$ be the first hitting time to a fixed level h in the interval [a,a+T], provided that its distance to the left end point of the interval is at most l. That is,

$$L_T(f, a) = \inf\{t \in [a, a + T] : f(t) = h, t \leq a + l\}.$$

Then L_T is a local intrinsic location functional. However, its "natural" extension, $L(f, [a, b]) := \inf\{t \in [a, b] : f(t) = h, t \leq a + l\}$ is not an intrinsic location functional. To see this, notice that the existence of such a location in an interval with length T does not guarantee its existence for all the larger intervals containing it, since the location may fail to remain close enough to the left end point when the interval expands.

It turns out that despite the large variety covered by the concept of local intrinsic location functional, they all correspond to the idea of taking the maximal element in a random set, ordered according to some specific rule.

Theorem 8.2.3. *Let H be defined as before. A mapping $L_T = L_T(f, a)$ from $H \times \mathbb{R}$ to $\mathbb{R} \cup \{\infty\}$ is a local intrinsic location functional with related length T, if and only if*

1. *$L_T(\cdot, a)$ is measurable for $a \in \mathbb{R}$;*
2. *For each function $f \in H$, there exists a subset of $\mathbb{R}$ denoted as $S(f)$ and a partial order $\preceq$ on it, satisfying:*
   - *(a) For any $c \in \mathbb{R}$, $S(f) = S(\theta_c f) + c$;*
   - *(b) For any $c \in \mathbb{R}$ and any $t_1, t_2 \in S(f)$, $t_1 \preceq t_2$ implies $t_1 - c \preceq t_2 - c$ in $S(\theta_c f)$,*

   such that for any $a \in \mathbb{R}$, either $S(f) \cap [a, a+T] = \phi$, in which case $L_T(f, a) = \infty$, or $L_T(f, a) \in S(f) \cap [a, a+T]$ and $s \preceq L_T(f, a)$ for all $s \in S(f) \cap [a, a+T]$.

Proof. It is easy to check that the measurability of $L_T(\cdot, a)$ for $a \in \mathbb{R}$ and the existence of such an ordered set $S(f)$ for $f \in H$ guarantee that L_T is a local intrinsic location functional. For the other direction, let L_T be a local intrinsic location functional with related length T. For each path f, define a set

$$S(f) = \{t \in \mathbb{R} : t = L_T(f, a) \text{ for some } a \in \mathbb{R}\}.$$

Thus $S(f)$ is the set of all the points which is chosen as the location for some interval with length T. From now on we fix the function f and simplify the notation $S(f)$ as S. We introduce the following partial binary relation on S. For two points $x, y \in S$, say $x \preceq_0 y$ if and only if there exists an interval $I_{x,y} = [a_{x,y}, a_{x,y} + T]$, such that $x, y \in I_{x,y}$ and $L_T(f, a_{x,y}) = y$. In another word, $x \preceq_0 y$ if and only if some interval with length T containing both of them "chooses" y rather than x to be its corresponding location. Then we complete $\preceq_0$ by taking the smallest transitive binary relation containing it, denoted as $\preceq$. We claim that such defined $\preceq$ is actually a partial order on the set S.

The reflexivity is clear: by definition, $x \preceq x, \forall x \in S$. The transitivity is also guaranteed by construction. Therefore the only thing left is to check the antisymmetry: if $x \preceq y$ and $y \preceq x$, then $x = y$. To this end, firstly notice that the construction of the binary relation $\preceq_0$ guarantees that it is always antisymmetric before being extended to $\preceq$. That is, $x \preceq_0 y$ and $y \preceq_0 x$ implies $x = y$. Now assume $x \neq y$, $x \preceq y$ and $y \preceq x$, then there is a loop: $x = t_0 \preceq_0 t_1 \preceq_0 ... \preceq_0 y = t_n \preceq_0 t_{n+1} \preceq_0 ... \preceq_0 t_{n+m-1} \preceq_0 t_{n+m} = x$ for some positive integers m, n, and points $t_0, t_1, ..., t_{n+m-1}, t_{n+m} = t_0$ satisfying $|t_{i+1} - t_i| \leq T$ for any $i = 0, ..., n + m - 1$.

To deal with this loop, notice that we have the proposition below, which states that if two points within distance less than or equal to T have a relation $\preceq$ between them, then there must be a direct relation given by $\preceq_0$. They can not be only related through a chain of "$\preceq_0$" via other points.

Lemma 8.2.1. *Let the relations $\preceq_0$ and $\preceq$ be as defined above. Then $t_1 \preceq t_2$ and $|t_2 - t_1| \leq T$ imply $t_1 \preceq_0 t_2$ or $t_2 \preceq_0 t_1$.*

Proof. Proof by contradiction. Without loss of generality, assume there are two points $t_1, t_2 \in S$, $t_1 < t_2$, $t_2 - t_1 \leq T$, there exist points $s_0, s_1, ..., s_n, s_{n+1}$ such that $s_0 = t_1 \preceq_0 s_1 \preceq_0 ... \preceq_0 s_n \preceq_0 s_{n+1} = t_2$, however, there is no direct relation given by $\preceq_0$ between t_1 and t_2. That is, every interval with length T containing the interval $[t_1, t_2]$ have neither t_1 nor t_2 as its corresponding location. Since $t_1 \in S$, there is $a \in \mathbb{R}$, such that $L_T(f, a) = t_1$. The interval $[a, a + T]$ can not include t_2, otherwise $t_2 \preceq_0 t_1$. Therefore $a + T < t_2$. Consider $L_T(f, t_2 - T)$. Because

$L_T(f,a) = t_1 \in [a, a+T] \cap [t_2 - T, t_2]$, the condition (4) in the definition of local intrinsic location functional rules out the possibility that $L_T(f, t_2 - T) = \infty$ or $L_T(f, t_2 - T) \in [a, a+T] \cap [t_2 - T, t_2]$. Thus $L_T(f, t_2 - T) \in (a+T, t_2] \subseteq (t_1, t_2]$. It can not be t_2 either since then $t_1 \preceq_0 t_2$. As a result, $L_T(f, t_2 - T) \in (t_1, t_2)$. Denote $L_T(f, t_2 - T)$ by t_3. Then $t_3 \in S$ and by definition $t_1 \preceq_0 t_3$ and $t_2 \preceq_0 t_3$.

Consider the intervals $[s_j, s_{j+1})$ for $j = 0, ..., n$ which satisfies $s_j < s_{j+1}$. Clearly, their union covers the interval $[t_1, t_2)$, therefore also the point t_3. Assume $t_3 \in [s_k, s_{k+1})$. There are two cases. Case 1: $s_{k+1} \leq t_2$. Since $s_k \preceq_0 s_{k+1}$, there is a real number a_1, such that $s_k \in [a_1, a_1 + T]$ and $L_T(f, a_1) = s_{k+1}$. Similarly, since $t_2 \preceq_0 t_3$, there is a real number a_2 such that $t_2 \in [a_2, a_2 + T]$ and $L_T(f, a_2) = t_3$. However $s_{k+1} \leq t_2$ implies that both s_{k+1} and t_3 are in the interval $[a_1, a_1 + T] \cap [a_2, a_2 + T]$, thus contradicting with the definition of local intrinsic location functional. Case 2: $s_{k+1} > t_2$. In this case, notice that $t_2 \in S$, so there exists a_3 such that $L_T(f, a_3) = t_2$. However, since $\preceq_0$ is antisymmetric, $t_2 \preceq_0 t_3$ implies that $t_3 \not\preceq_0 t_2$, so $a_3 > t_3$. Now both t_2 and s_{k+1} are in the interval $[a_1, a_1 + T] \cap [a_3, a_3 + T]$, yet L_T gives out different locations, contradiction again. To conclude, the assumption at the beginning of the proof can not hold, and the lemma is proved. □

Now we turn back to the loop and prove the following result: there exist $i_1, i_2, i_3 \in \{0, ..., n+m-1\}$, such that $t_{i_1} \preceq_0 t_{i_2} \preceq_0 t_{i_3} \preceq_0 t_{i_1}$. Consider the rightmost point in set $\{t_i\}_{i=0,\ldots,n+m-1}$, denoted as $t_j := \max_{i=0}^{n+m-1} t_i$. Notice that $t_{j-1} < t_j$, $t_{j+1} < t_j$, therefore $|t_{j+1} - t_{j-1}| < T$, and $t_{j-1} \preceq_0 t_j \preceq_0 t_{j+1}$ (define $t_{-1} = t_{n+m-1}$). By Lemma 8.2.1 there is a relation $\preceq_0$ between t_{j+1} and t_{j-1}. If $t_{j+1} \preceq_0 t_{j-1}$, we already have a loop with three terms as desired. If $t_{j-1} \preceq_0 t_{j+1}$, then consider the set $\{t_i\}_{i=0,\ldots,n+m-1, i \neq j}$. It is also a loop as the set $\{t_i\}_{i=0,\ldots,n+m-1}$

by which we started, now with one less term. An iteration of this procedure finally decreases the size of the set to 3, so we find a loop with 3 terms again.

The existence of a loop with 3 terms, however, contradicts with the definition of the relation $\preceq_0$. To see this, without loss of generality, suppose that we have $t_1 < t_2 < t_3$ satisfying $t_1 \preceq_0 t_2 \preceq_0 t_3 \preceq_0 t_1$ and $t_3 - t_1 \leq T$. This means that there exists $a, b \in \mathbb{R}$, such that $t_1, t_3 \in [a, a+T]$ and $L_T(f, a) = t_1$, $t_1, t_2 \in [b, b+T]$ and $L_T(f, b) = t_2$. However, the fact that $t_1, t_2 \in [a, a+T] \cap [b, b+T]$, yet $L_T(f, a)$ and $L_T(f, b)$ are not equal contradicts with the definition of local intrinsic location functional.

In total, we have seen that a loop of relation $\preceq_0$, therefore also $\preceq$, is not possible. Thus the antisymmetry is proved. The relation $\preceq$ is a partial order. Finally, it is clear by the construction of the partially ordered set $(S(f), \preceq)$ that either $S(f) \cap [a, a+T] = \phi$, in which case $L_T(f, a) = \infty$, or $L_T(f, a) \in S(f) \cap [a, a+T]$, in which case $s \preceq L_T(f, a)$ for all $s \in S(f) \cap [a, a+T]$. $\square$

Remark 8.2.4. The partial order in the theorem has the special property that there exists a unique maximal element over any interval with length T. In this sense it behaves like a total order. Indeed, by order extension principle, the partial order $\preceq$ can always be extended to a total order on $S(f)$, and it is clear that we can do it in a shift-invariant way, so that the resulting total order also satisfies the conditions in Theorem 8.2.3. Nonetheless, here we would like to keep $\preceq$ a partial order for generality.

A similar reasoning allows us to derive the ordered set representation for intrinsic location functionals.

Corollary 8.2.5. *Let $H, \mathcal{I}$ be defined as before. A mapping $L = L(f, I)$ from $H \times \mathcal{I}$ to*

$\mathbb{R} \cup \{\infty\}$ *is an intrinsic location functional if and only if*

1. *$L(\cdot, I)$ is measurable for $I \in \mathcal{I}$;*
2. *For each function $f \in H$, there exists a partially ordered subset of $\mathbb{R}$, denoted as $(S(f), \preceq_1)$, satisfying:*
   (a) *For any $c \in \mathbb{R}$, $S(f) = S(\theta_c f) + c$;*
   (b) *For any $c \in \mathbb{R}$ and any $t_1, t_2 \in S(f)$, $t_1 \preceq_1 t_2$ implies $t_1 - c \preceq_1 t_2 - c$ in $S(\theta_c f)$,*

   such that for any $I \in \mathcal{I}$, either $S(f) \cap I = \phi$, in which case $L(f, I) = \infty$, or $L(f, I) \in S(f) \cap I$ and $s \preceq_1 L(f, I)$ for all $s \in S(f) \cap I$.

Proof. Again, it is routine to check the "if" direction. For the other direction, define $S(f) := \{t \in \mathbb{R} : L(f, I) = t \text{ for some } I \in \mathcal{I}\}$ and the binary relation $\preceq_1$ on $S(f)$: $x \preceq_1 y$ if and only if there exists an interval $I \in \mathcal{I}$ such that $x, y \in I$ and $L(f, I) = y$. The argument goes through in the same way, and is actually simpler, since such defined $\preceq_1$ is now directly a partial order. $\square$

Example 8.2.6. Let H be the space of all upper semi-continuous functions on $\mathbb{R}$. The location of the path supremum $\tau_{f,I} := \inf\{t \in I : f(t) = \sup_{s\in I} f(s)\}$, $f \in H, I \in \mathcal{I}$ is an intrinsic location functional. It corresponds to an ordered set $(S(f), \preceq)$, where $S(f) = S^1(f) \cup S^2(f)$, $S^1(f)$ is the union of the set of local maxima of f, and $S^2(f) := \{t \in \mathbb{R} : t = \sup_{s\in[t-T,t]}(f(s)) \text{ or } t = \sup_{s\in[t,t+T]}(f(s))\}$. "$\preceq$" is firstly ordered by the value of the function at the points and in case of a tie, inversely ordered by the location (that is, the left locations receive high orders).

Example 8.2.7. Let H be the space of all continuous functions on $\mathbb{R}$. The first hitting time of a level l over an interval I, defined by $T^l_{f,I} := \inf\{t \in I : f(t) = l\}$

is also an intrinsic location functional. The ordered set $(S(f), \preceq)$ is now given by $S(f) = f^{-1}(l)$ and the inverse order on the real line.

It is clear that the partially ordered random set representation of a local intrinsic location functional or an intrinsic location functional can not be unique, since one can always add irrelevant points to $S(f)$ and assign them very low orders, so that the added points are actually never chosen as the location for any interval. However, there exists a unique minimal representation, as indicated by the proof of Theorem 8.2.3.

Corollary 8.2.8. *Let L be a local intrinsic location functional (resp. intrinsic location functional) with path space H. There exists a partially ordered set $(S(f), \preceq)$ for each function $f \in H$, satisfying the conditions in Theorem 8.2.3 (resp. Corollary 8.2.5), such that for any other partially ordered set $(S'(f), \preceq')$ also satisfying the same conditions,*

$$S(f) \subseteq S'(f)$$

and

$$s_1, s_2 \in S(f), s_1 \preceq s_2 \textit{ implies } s_1 \preceq' s_2 \textit{ in } S'(f).$$

The proof is very easy and omitted here.

8.3 Extension and restriction

The ordered set representation provides powerful tools for us to clarify the link between intrinsic location functional and local intrinsic location functional. The theorem below shows that a local intrinsic location functional is "almost" just a "local" version of an intrinsic location functional.

We call a mapping L from $H \times \mathcal{I}$ to $\mathbb{R} \cup \{\infty\}$ a "pre-intrinsic location functional", if it satisfies all the defining properties in Definition 1 except for the measurability condition (1). In another word, a pre-intrinsic location functional becomes an intrinsic location functional once it is measurable for any fixed interval $I \in \mathcal{I}$.

Theorem 8.3.1. *Let L be an intrinsic location functional. Then for any $T > 0$,*

$$L_T(f, a) := L(f, [a, a + T]) \tag{8.1}$$

is a local intrinsic location functional. Conversely, let L_T be a local intrinsic location functional. Then there exists a pre-intrinsic location functional L, such that (8.1) holds for all $f \in H$ and $a \in \mathbb{R}$.

Proof. The fact that a restricted intrinsic location functional is a local intrinsic location functional can be easily checked either by their definitions or by the ordered set representation. For the other direction, suppose we have a local intrinsic location functional L_T, with the partially ordered set $(S(f), \preceq)$ for each $f \in H$. By the order extension principle, $(S(f), \preceq)$ can always be extended, in a shift-invariant way, to a totally ordered set $(S(f), \preceq_1)$, which is, of course, a special partially ordered set. Define $L(f, I)$ for any $I \in \mathcal{I}$ by taking the maximal element in I of $S(f)$ according to $\preceq_1$: $L(f, I) \in S(f)$ and $s \preceq_1 L(f, I)$ for all $s \in S(f) \cap I$, then by Corollary 8.2.5 such defined L is a pre-intrinsic location functional. □

Notice, however, that we have not touched the measurability issue and claimed that each local intrinsic location functional necessarily has an intrinsic location functional extension. The problem of measurability is highly nontrivial and in general, the measurability of a local intrinsic location functional for intervals with a single fixed length may not be enough to guarantee the measurability of its extensions with all different interval lengths. Instead, we prove the following result, which shows that there always exists an intrinsic location functional which agrees almost surely with the given local intrinsic location functional for any stationary process in the interior of any interval with the fixed length.

Proposition 8.3.2. *Let $L_T : H \times \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ be a local intrinsic location functional with related length T. Then there exists an intrinsic location functional $L : H \times \mathcal{I} \to \mathbb{R} \cup \{\infty\}$, such that for any $a \in \mathbb{R}$ and stationary process $\mathbf{X}$ with paths in H,*

$$\mathbb{P}[L_T(\mathbf{X}, a) \neq L(\mathbf{X}, [a, a+T]),$$

$$L_T(\mathbf{X}, a) \in (a, a+T) \text{ or } L(\mathbf{X}, [a, a+T]) \in (a, a+T)] = 0.$$

Before we go to the proof of Proposition 8.3.2, let us first look at a useful lemma.

Lemma 8.3.1. *Let L_T be a local intrinsic location functional defined on $H \times \mathbb{R}$. Then*

1. *For any $f \in H$, any $a < b$ such that $L_T(f, a) \neq \infty$ and $L_T(f, b) \neq \infty$, $L_T(f, a) \leq L_T(f, b)$.*
2. *If $L_T(f, a) = L_T(f, b) = t \neq \infty$, then $L_T(f, c) = t$ for any $c \in [a, b]$.*
3. *If $a < b$, $b - a \leq T$ and $L_T(f, a) = L_T(f, b) = \infty$, then $L_T(f, c) = \infty$ for all $c \in [a, b]$.*

Proof. Suppose for some $a < b$, $L_T(f, b) < L_T(f, a) < \infty$. Then both $L_T(f, a)$ and $L_T(f, b)$ are in the interval $[a, a+T] \cap [b, b+T]$. However, by the definition of local intrinsic location functional, this implies that they must be equal. Thus the first claim of the proposition is proved. Now assume $L_T(f,a) = L_T(f,b) = t \neq \infty$. Then $t \in [a, a+T] \cap [b, b+T] = [b, a+T] \neq \phi$. For any $c \in [a,b]$, $[c, c+T] \supseteq [b, a+T]$, hence $t \in [a, a+T] \cap [c, c+T]$. By definition of local intrinsic location functional, $L_T(f,c) \neq \infty$. Then by the first claim of the proposition, $t = L_T(f,a) \leq L_T(f,c) \leq L_T(f,b) = t$. Therefore $L_T(f,c) = t$ as well. Finally, if $a < b, b-a \leq T$, then for any $c \in [a,b]$, $[c, c+T] \subset [a, a+T] \cup [b, b+T]$. If $L_T(f,a) = L_T(f,b) = \infty$, then by Theorem 8.2.3, $[a, a+T] \cap S(f) = [b, b+T] \cap S(f) = \phi$, where $S(f)$ is a set of points corresponding to L_T. As a result, $[c, c+T] \cap S(f) = \phi$, which, going back to L_T, means that $L_T(f,c) = \infty$.

□

Proof of Proposition 8.3.2. For any function $f \in H$, define the sets

$$S_1(f) := \{t \in \mathbb{R} : \exists (x,y) \subset \mathbb{R}, \text{ s.t. } L_T(f,a) = t, \forall a \in (x,y)\},$$

$$S_2(f) := \{t \in \mathbb{R} \setminus S_1(f) : L_T(f,t) = t \text{ or } L_T(f, t-T) = t\}$$

and $S'(f) = S_1(f) \cup S_2(f)$.

On $S'(f)$ assign a binary relation $\preceq_0$: $t_1 \preceq_0 t_2$ if and only if $|t_2 - t_1| < T$ and there exists a real number a satisfying $t_1, t_2 \in [a, a+T]$ such that $L_T(f,a) = t_2$. Notice that the set $S'(f)$ is a subset of the set we constructed in the proof of Theorem 8.3.1, and $\preceq_0$ is also a restriction of the corresponding binary relation that we saw before. As a result, one can again extend $\preceq_0$ to a smallest partial order, still denoted by $\preceq$.

For function $f \in H$ and a compact interval I, define $L(f,I)$ to be the first element in $S'(f)$ which is maximal in I:

$$L(f,I) = \inf\{t \in S'(f) \cap I : t' \in S'(f) \cap I \text{ and } t \preceq t' \text{ implies } t' = t\}.$$

We can denote the set on the right hand side of the definition above, namely, the set of all the maximal in I points in $S'(f)$, by $M_{f,I}$. Then $L(f, I)$ is simply $\inf(M_{f,I})$, with the tradition that $\inf(\phi) = \infty$. Indeed, this way of choosing the first maximal element is equivalent to assigning an additional order among the maximal elements according to their location, with the left receiving the higher order and the right lower. The resulting new order will then satisfy all the conditions listed in Corollary 8.2.5, which assures that such defined $L(f, I)$ is a pre-intrinsic location functional. Thus all that is left is to check the measurability.

Fix $I = [a, b]$ with $|I| = b - a > T$ and $f \in H$. The event $\{L(f, I) \leq s\}$ is $\{a \in M_{f,I}\}$ if $s = a$, $\{a \in M_{f,I}\} \cup \{M_{f,I} \cap (a, s] \neq \phi\}$ if $s \in (a, b)$, and $\{a \in M_{f,I}\} \cup \{M_{f,I} \cap (a, s) \neq \phi\} \cup \{b \in M_{f,I}\}$ if $s = b$. Therefore it suffices to verify the measurability for each of these sets.

Lemma 8.3.2. *For* $I = [a, b]$, $b - a > T$, $t \in M_{f,I} \cap (a, b)$ *if and only if for some sequences* $\{t_{1n}\}_{n=1,2,...}$ *and* $\{t_{2n}\}_{n=1,2,...}$ *such that* $t_{1n} \to t$ *and* $t_{2n} \to t$ *as* $n \to \infty$, $L_T(f, a \vee (t_{1n} - T)) = L_T(f, (b - T) \wedge t_{2n}) = t$ *holds for* $n = 1, 2, ...$

Proof. Firstly assume that $t \in M_{f,I} \cap (a, b)$. If $a \leq t - T$, then for any $s \in (t, (t + T) \wedge b)$, $[s - T, s] \subset (a, b)$, and $t \in (s - T, s)$. By the maximality of t under the partial order $\preceq$, $L_T(f, s - T) = t$. Therefore we only need to take $\{t_{1n}\}_{n=1,2,...}$ a decreasing sequence converging to t with $t_{11} < (t + T) \wedge b$ to have $L_T(f, t_{1n} - T) = t$. If $a > t - T$, then the maximality implies that $L_T(f, a) = t$. Combining these two cases, $L_T(f, a \vee (t_{1n} - T)) = t$. Symmetrically we have $L_T(f, (b - T) \wedge t_{2n}) = t$.

Now suppose $t \in S'(f) \cap (a, b)$ but $t \notin M_{f,I}$. Then there exists $s \in (t - T, t + T) \cap [a, b]$ such that $t \preceq_0 s$. Without loss of generality, assume that $s < t$. Then for any $r \in [t-T, s)$, $L_T(f,r) \neq t$, since otherwise $s \preceq_0 t$. Therefore there does not exist a sequence $\{t_{1n}\}_{n=1,2,...}$, such that $L_T(f, a \vee (t_{1n} - T)) = t$. Finally, it is clear that if $t \notin S'(f) \cap (a,b)$, the sequence $\{t_{1n}\}_{n=1,2,...}$ and $\{t_{2n}\}_{n=1,2,...}$, either. The lemma is proved. $\square$

For any x, y such that $a \leq x < y \leq b$, denote by $E_I(x,y)$ the event $L_T(f, a \vee (y-T)) = L_T(f, (b-T) \wedge x) \neq \infty$. For $r, s \in (a,b)$ and $m = 1, 2, ...$, define event $E_{I,m}(r,s) = \bigcup_{i=1}^{2^m-1} E_I(r + \frac{(i-1)(s-r)}{2^m}, r + \frac{(i+1)(s-r)}{2^m})$. Consider the set

$$\begin{aligned} E(I,r,s) &:= \textstyle\bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} E_{I,m}(r,s) \\ &= \textstyle\bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \bigcup_{i=1}^{m-1} E_I(r + \frac{(i-1)(s-r)}{2^m}, r + \frac{(i+1)(s-r)}{2^m}). \end{aligned}$$

It is clearly measurable. Suppose there is a point $t \in (r,s)$ in $M_{f,I}$. For any m large enough, let i' be an index satisfying $t \in (r + \frac{(i'-1)(s-r)}{2^m}, r + \frac{(i'+1)(s-r)}{2^m})$. Then event $E_I(r + \frac{(i'-1)(s-r)}{2^m}, r + \frac{(i'+1)(s-r)}{2^m})$ holds. Consequently $E_{I,m}(r,s)$ holds hence $E(I,r,s)$ also holds. Thus $\{M_{f,I} \cap (r,s) \neq \phi\} \subseteq E(I,r,s)$. On the other hand, suppose $E(I,r,s)$ is realized. Then for all m large enough, $E_I(r + \frac{(i-1)(s-r)}{2^m}, r + \frac{(i+1)(s-r)}{2^m})$ holds for some $i = 1, ..., 2^m - 1$. Denote by J_m the set of indices $i = 1, ..., 2^m - 1$ for which $E_I(r + \frac{(i-1)(s-r)}{2^m}, r + \frac{(i+1)(s-r)}{2^m})$ holds, and $B_m := \bigcup_{i \in J_m} [r + \frac{(i-1)(s-r)}{2^m}, r + \frac{(i+1)(s-r)}{2^m}]$. It is easy to check by definition that B_m is a decreasing sequence of closed sets, thus there exists some point $t \in [r,s]$ which is covered by infinite members in $\{B_m\}_{m=0,1,...}$, therefore also infinite number of intervals forming $B_m, m = 0, 1, ...$ Let $\{I_{m_j} = [a_{m_j}, b_{m_j}]\}_{j=1,2,...}$ be such a sequence always covering t. Notice that $a_{m_j} \to t$ and $b_{m_j} \to t$ as $j \to \infty$. Moreover, $E_I(a_{m_j}, b_{m_j})$ holds for all $j = 1, 2, ...$ by construction. Thus $t \in M_{f,I}$. Thus we have

$$\{M_{f,I} \cap (r,s) \neq \phi\} \subseteq E(I,r,s) \subseteq \{M_{f,I} \cap [r,s] \neq \phi\},$$

which implies

$$E(I,r,s) \cup \{r \in M_{f,I}\} \cup \{s \in M_{f,I}\} = \{M_{f,I} \cap [r,s] \neq \phi\}.$$

It is easy to check that $\{r \in M_{f,I}\}$ and $\{s \in M_{f,I}\}$ are measurable. $\{r \in M_{f,I}\}$, for example, can only happen if $r \in M_{f,I} \cap S_1(f)$, which is then equivalent to $\cup_{n=1}^{\infty} E_I(r - \frac{1}{n}, r + \frac{1}{n})$. As a result, $\{M_{f,I} \cap [r,s] \neq \phi\}$ is measurable for any r, s in the interior of $[a,b]$. It is then trivial to see the measurability of $\{M_{f,I} \cap (a,s] \neq \phi\}$ for $s \in (a,b)$ or $\{M_{f,I} \cap (a,s) \neq \phi\}$ for $s = b$ by taking a countable union. The case for the two endpoints a and b can be checked directly. The measurability of $a \in M_{f,I}$, for instance, is verified once we observe that $a \in M_{f,I} \cap S_1(f)$ if and only if there exists a sequence $\{s_n\}_{n=1,2...}$, such that $s_n \uparrow a$ and $L_T(f, s_n) = a$ for $n = 1, 2, ...$ $a \in M_{f,I} \cap S_2(f)$, of course, if and only if $L_T(f,a) = a$.

For the case of $I = [a,b]$ with $|I| = T$, the key is to notice that $L(f,I) = t \in (a,b)$ if and only if there exists a positive integer n such that $L(f, I_{1n}) = t$ or $L(f, I_{2n}) = t$, where $I_{1n} = [a - \frac{1}{n}, b]$, $I_{2n} = [a, b + \frac{1}{n}]$, and $L(f, I_{1n})$ and $L(f, I_{2n})$ are defined as above for $|I| > T$. Thus for any $s \in (a,b)$,

$$\{L(f,I) \in [a,s])\} = \{a \in M_{f,I}\} \cup \left(\bigcup_{n=1}^{\infty} \{L(f, I_{in}) \in (a,s], i = 1 \text{ or } 2\} \right)$$

is measurable. The cases with $s = a$ or $s = b$ are not much different from before.

Finally if $I = [a,b]$ with $|I| < T$, $L(f,I) = t \in (a,b)$ is equivalent to the existence of three points $x, y \in \mathbb{Q}$ and $z \in (\mathbb{Q} \cap [b-T, a]) \cup \{a, b-T\}$, such that $L_T(f,x) = L_T(f,y) = L_T(f,z) = t$. It is not difficult to check this equivalence. Intuitively, the existence of x and y assures that $t \in S'(f)$, while the existence of z guarantees the maximality of t in I. The countability of the rational set then leads to the measurability. We skip the details.

Combining the three cases proves the measurability of $L(\cdot, I)$ for any compact interval I, as desired. L is thus an intrinsic location functional. The last thing in the proof is therefore to show the relationship between $L_T(\mathbf{X}, a)$ and $L(\mathbf{X}, [a, a+T])$ claimed in the proposition.

Let $\mathbf{X}$ be any stationary process with paths in H. Firstly, assume $L(\mathbf{X}, [a, a + T]) = t \in (a, a + T)$ but $L_T(\mathbf{X}, a) \neq t$. Then $L_T(\mathbf{X}, a) \notin S'(\mathbf{X})$, since otherwise t and $L_T(\mathbf{X}, a)$ are both in $S'(\mathbf{X})$, $|t - L_T(\mathbf{X}, a)| < T$ and by definition of $\preceq_0$, $t \preceq_0 L_T(\mathbf{X}, a)$, contradicting the maximality of $L(\mathbf{X}, [a, a+T])$. By the same reasoning, if $L_T(\mathbf{X}, a) \in (a, a + T)$ but $L(\mathbf{X}, [a, a + T]) \neq L_T(\mathbf{X}, a)$ then $L_T(\mathbf{X}, a) \notin S'(\mathbf{X})$. Together, we have

$$\{L_T(\mathbf{X}, a) \neq L(\mathbf{X}, [a, a + T]),$$

$$L_T(\mathbf{X}, a) \in (a, a + T) \text{ or } L(\mathbf{X}, [a, a + T]) \in (a, a + T)\}$$

$$\subseteq \{L_T(\mathbf{X}, a) \notin S'(\mathbf{X})\}.$$

Notice that if $L_T(\mathbf{X}, a) = a$ or $L_T(\mathbf{X}, a) = a + T$, then $L_T(\mathbf{X}, a) \in S'(\mathbf{X})$ automatically. By the definition of $S'(X)$, $L_T(\mathbf{X}, a) \notin S'(\mathbf{X})$ if and only if $L_T(\mathbf{X}, a) \in (a, a + T)$ and $L_T(\mathbf{X}, a) \neq L_T(\mathbf{X}, b)$ for any $b \neq a$, which is equivalent to $L_T(\mathbf{X}, a) \neq L_T(\mathbf{X}, b)$ for any $b \neq a$, $b \in \mathbb{Q}$ by Lemma 8.3.1. Thus $\{L_T(\mathbf{X}, a) \neq S'(\mathbf{X})\}$ is measurable. Now we show that $\mathbb{P}(L_T(\mathbf{X}, a) \notin S'(\mathbf{X})) = 0$. Assume $\mathbb{P}(L_T(\mathbf{X}, a) \in (a, a + T) \setminus S'(\mathbf{X})) > 0$. Then there exists $\Delta > 0$, such that $\mathbb{P}(L_T(\mathbf{X}, a) \in (a + \Delta, a + T - \Delta) \setminus S'(\mathbf{X})) =: \delta > 0$. Take $\epsilon < \Delta/(\lfloor 1/\delta \rfloor)$ and compact intervals $I_i = [a + i\epsilon, a + i\epsilon + T]$ for $i = 0, 1, ..., \lfloor 1/\delta \rfloor$, where "$\lfloor \cdot \rfloor$" refers to the largest integer smaller or equal to the argument. By construction, for any $i, j = 0, 1, ..., \lfloor 1/\delta \rfloor$, $I_i \cap I_j \supset [a + i\epsilon + \Delta, a + i\epsilon + T - \Delta] \cup [a + j\epsilon + \Delta, a + j\epsilon + T - \Delta]$. This, however, implies that the events $E_i := \{L_T(\mathbf{X}, a + i\epsilon) \in (a + i\epsilon + \Delta, a + i\epsilon + T - \Delta) \setminus S'(\mathbf{X})\}$ must be disjoint for different i. Otherwise, suppose E_i and E_j holds for some $i < j$. Then since both $L_T(\mathbf{X}, a + i\epsilon)$ and $L_T(\mathbf{X}, a + j\epsilon)$ are in the intersection of I_i and I_j, they must be equal. Lemma 8.3.1 then implies that $L_T(\mathbf{X}, a') = L_T(\mathbf{X}, a + i\epsilon)$ for all $a' \in [a + i\epsilon, a + j\epsilon]$. This contradicts with E_i, which requires that $L_T(\mathbf{X}, a + i\epsilon) \notin S'(\mathbf{X})$. By stationarity,

$\mathbb{P}(E_i) = \mathbb{P}(L_T(\mathbf{X}, a) \in (a + \Delta, a + T - \Delta) \setminus S(\mathbf{X})) = \delta, i = 0, 1, ..., \lfloor 1/\delta \rfloor$. Then

$$\mathbb{P}(\bigcup_{i=0}^{\lfloor 1/\delta \rfloor} E_i) = \delta \cdot (\lfloor 1/\delta \rfloor + 1) > 1,$$

which clearly shows the existence of a contradiction. As a result, $\mathbb{P}(L_T(\mathbf{X}, a) \notin S'(\mathbf{X})) = 0$ and the proof of the proposition is complete. □

The importance of Proposition 8.3.2 resides in the fact that most of the distributional properties of intrinsic location functionals proved in [18] can now be transformed automatically to local intrinsic location functionals. In particular, local intrinsic location functionals always satisfy the total variation constraints. Thus the equivalence between the stationarity, the total variation constraints and the shift invariance of the distributions can be extended to local intrinsic location functionals.

Corollary 8.3.3. *Let $\mathbf{X}$ be a stochastic process with continuous paths. Let $\mathcal{L}_{loc,T}$ be the set of all local intrinsic location functionals in $\mathcal{C}(\mathbb{R})$ with related length T. Then the followings are equivalent:*

1. *The process $\mathbf{X}$ is stationary.*
2. *For any $T > 0$, any local intrinsic location functional $L_T \in \mathcal{L}_{loc,T}$, the distribution of $L_T(\mathbf{X}, a) - a$ does not depend on a.*
3. *For any $T > 0$, any local intrinsic location functional $L_T \in \mathcal{L}_{loc,T}$ and any $a \in \mathbb{R}$, $L_T(\mathbf{X}, a)$ admits a density function $f_{\mathbf{X},a,T}(t)$ in $(a, a+T)$, which satisfies the total variation constraint on $[a, a + T]$.*

Remark 8.3.4. A closer examination of the proof of the equivalence theorem in [18] shows that the length of the interval does not play any crucial role in the proof of the equivalence between (1) and (2). As a result, (2) in Corollary 8.3.3 is also equivalent to:

$(2')$ For a fixed $T > 0$, any local intrinsic location functional $L_T \in \mathcal{L}_{loc,T}$, the distribution of $L_T(\mathbf{X}, a) - a$ does not depend on a.

To sum up, while the equivalence between the stationarity and the total variation constraints of the intrinsic location functionals have been established in [18], we just extended this result to local intrinsic location functionals, which is more generally defined compared to intrinsic location functionals. Moreover, the local intrinsic location functionals are further identified with the shift-compatible ordered sets of points $(S(\cdot), \preceq)$ on $\mathbb{R}$ as path functionals. Such an identification provides a particularly convenient way to define local intrinsic location functionals.

We complete this section by the following corollary, which examines the relation between intrinsic location functionals and local intrinsic location functionals, from the perspective of the partially ordered sets they correspond to.

Corollary 8.3.5. *Let H, $\mathcal{I}$ be defined as before. Let $L : H \times \mathcal{I} \to \mathbb{R} \cup \{\infty\}$ be a mapping satisfying that $L(\cdot, I)$ is measurable for any $I \in \mathcal{I}$. Then L is an intrinsic location functional if and only if*

1. *$L_T : H \times \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ defined by $L_T(f, a) = L(f, [a, a+T])$ is a local intrinsic location functional for each $T > 0$.*
2. *Let $(S_T(\cdot), \preceq_T)$ be the minimal ordered random set representation for $L_T, T > 0$. Then there exists a partially ordered random set $(S(\cdot), \preceq_1)$ satisfying condition (2) in Corollary 8.2.5, such that for any $T > 0$, $f \in H$, $S_T(f) \subseteq S(f)$, and $t_1, t_2 \in S_T(f), t_1 \preceq_T t_2$ implies $t_1 \preceq_1 t_2$ in $S(f)$.*

Proof. The "if" direction is directly from Corollary 8.2.5. In the following we focus on the "only if" direction. Suppose L is an intrinsic location functional with path space H. Then L_T defined as in the corollary are automatically local intrinsic location functionals, as previously proved. Assume that there does not exist a partially ordered set $(S(f), \preceq_1)$ satisfying (2). In particular, we can take $(S(f), \preceq_1)$ to be a partially ordered set representation of L, and assume that it does not satisfy (2) as described in Corollary 8.2.5. Then there must exist $T > 0$, $f \in H$, such that either $S_T(f) \nsubseteq S(f)$, or $S_T(f) \subseteq S(f)$ but there are $t_1, t_2 \in S_T(f), t_1 \preceq_T t_2$ but $t_1 \npreceq_1 t_2$ in $S(f)$. The first possibility is easily eliminated: $t \in S_T(f)$ implies that $L_T(f, a) = L(f, [a, a+T]) = t$ for some $a \in \mathbb{R}$, thus by definition of $S(f)$, it must contain t as well. Now assume we do have $t_1, t_2 \in S_T(f), t_1 \preceq_T t_2$ but $t_1 \npreceq_1 t_2$ in $S(f)$. Without lose of generality, assume $t_1 < t_2$. By the minimality of $(S_T(f), \preceq_T)$, this means that there exist $s_i \preceq_0 s_{i+1}$ for $i = 1, ..., m-1$, where $t_1 = s_1 < s_2 < ... < s_m = t_2$ are all the points in $S_T(f) \cap [t_1, t_2]$, and $\preceq_0$ is defined as before: $x \preceq_0 y$ if $L_T(f, a) = y$ for some a such that $x, y \in [a, a+T]$. Consider $L(f, [t_1, t_2])$. If $t_2 - t_1 \leq T$, then $t_1 \preceq_0 t_2$, by definition $t_1 \preceq_1 t_2$, contradiction. If $t_2 - t_1 > T$, we know that $L(f, [t_1, t_2]) \in S_T(f) \cap [t_1, t_2]$. It can not be t_2, otherwise $t_1 \preceq_1 t_2$. Thus $L(f, [t_1, t_2]) = s_i$ for some $i = 1, ..., m-1$. However, by stability under restriction, $L(f, I) = s_i$ for any interval $I \in \mathcal{I}$ such that $|I| = T$, $s_i \in I$ and $I \subset [t_1, t_2]$. Thus $s_i \npreceq_0 s_j$ for any $j = 1, ..., m$. Contradiction. Therefore the assumption can not hold, and the proof is completed. $\square$

8.4 Path characterization

Let L_T be a local intrinsic location functional with related length T. Given any $f \in H$, define $g(x) := L_T(f,x) - x, \forall x \in \mathbb{R}$. Thus $g(x)$ is the relative location of L_T with respect to the starting point x. The following result gives out a characterization of the path of the function g. In another word, it answers the question how we can tell whether a random location is a local intrinsic location functional by looking at the change of its place relative to the interval as the interval shifts.

We call a partition satisfying certain property the "roughest", if all the other partitions satisfying the property is a refinement of the given partition.

Theorem 8.4.1. *Let L_T be a local intrinsic location functional discussed before, and g be the function defined above. Then for any $f \in H$, there exists a roughest partition of the real line by intervals (the intervals can be degenerated, and the boundaries of the intervals can be open or closed), such that for any member $I = (a,b), (a,b], [a,b)$ or $[a,b]$ of this partition, exactly one of the followings is true.*

(1) $b - a \leq T$, and $g(x) = d - x$ for some $d \in [b, a+T]$ and all $x \in I$.

(2) $g(x) = \infty$ for all $x \in I$.

Moreover, if $g(a) \neq T$ (resp. $g(b) \neq 0$), then $\lim_{x\uparrow a} g(x) = 0$ (resp. $\lim_{x\downarrow b} g(x) = T$). If I is open on a (resp. b), then $g(a) = 0$ (resp. $g(b) = T$).

On the other hand, let L_T be a mapping from $H \times \mathbb{R}$ to $\mathbb{R} \cup \{\infty\}$ such that $L_T(\cdot, a)$ is measurable for any $a \in \mathbb{R}$, and $L_T(f,a) = L_T(\theta_c f, a-c) + c$ for any $a, c \in \mathbb{R}$. If for any function $f \in H$, there always exists a partition P of the real line by intervals satisfying the properties listed above, then L_T is a local intrinsic location functional with related length T.

Roughly speaking, Theorem 8.4.1 tells us that the function g consists of linear pieces with slope -1 and intervals with value ∞. The pieces are combined together following the rule that when the interval $[x, x+T]$ shifts along the real line, a location can "disappear" in the interior of the interval only if it is replaced by another location appearing at the right endpoint $x+T$. Symmetrically, a location can only "appear" in the interior of the interval only if it is replacing another location disappearing at the left endpoint x. The actual scenario is a little bit more complicated, since both the replaced and replacing "location" can be indeed the limit of a sequence of locations, where comes the limits in the formulation of the theorem.

Proof. Let L_T be a local intrinsic location functional with related length T. By Theorem 8.2.3 and Remark 8.2.4, for each $f \in H$, there is a set $S(f) \subseteq \mathbb{R}$ and a partial order $\preceq$ on it, satisfying $S(\theta_c f) = S(f) - c$ and $t_1 \preceq t_2$ in $S(f)$ if and only if $t_1 - c \preceq t_2 - c$ in $S(\theta_c f)$ for any $c \in \mathbb{R}$, such that $L_T(f, x)$ is the unique maximal element by $\preceq$ in $S(f) \cap [x, x+T]$ for any $f \in H$ and any $x \in \mathbb{R}$, provided that it exists. For any fixed $x \in \mathbb{R}$, there are two cases. Case 1: $S(f) \cap [x, x+T] = \phi$. In this case define $a = \sup\{S(f) \cap (-\infty, x)\}$ and $b = \inf\{S(f) \cap (x+T, \infty)\} - T$. Then a, b are clearly the two boundaries of the largest interval containing x on which $L_T(f, \cdot) = \infty$. Notice that it is possible to have $a = b$, in which case the interval becomes degenerate. Case 2: $S(f) \cap [x, x+T] \neq \phi$. In this case define $a = \max\{L_T(f,x) - T, \sup\{y \in \mathbb{R} : y \in S(f), y < L_T(f,x), L_T(f,x) \preceq y\}\}$ and $b = \min\{L_T(f,x), \inf\{y \in \mathbb{R} : y \in S(f), y > L_T(f,x), L_T(f,x) \preceq y\} - T\}$. Then $L_T(f,x)$ will remain the same when and only when x moves between a and b. That is, $L_T(f,y) = L_T(f,x)$ for $y \in I$, $I = [a,b], [a,b), (a,b]$ or (a,b), whether the boundary is closed or open being determined by which one is larger/smaller in the max and min in the definition of a and b, and whether the supremum and infimum are achieved by a single point or only by a sequence of points. As a result, for any $y \in I$, $g(y) = L_T(f, y) - y = L_T(f, x) - y = d - y$ where $d := L_T(f, x) \in \cap_{y \in I}[y, y + T] \subseteq [b, a + T]$. Thus case 2 corresponds to scenario (1) and case 1 corresponds to scenario (2) in Theorem 8.4.1.

Next we check the combination rule, that is, the sentence below the two scenarios in the theorem. Firstly assume $g(a) \neq T$. Hence either $g(a) < T$ or $g(a) = \infty$. If $g(a) < T$, consider $g(x)$ for $x \in (a - T + g(a), a)$. Notice that $x + T > a + g(a) = L_T(f, a)$. However, $L_T(f, x)$ can not be equal to $L_T(f, a)$, since otherwise by Lemma 8.3.1 a will not be the left endpoint of a largest interval on which $g(\cdot)$ is linear. Hence $L_T(f, x) \in [x, x + T] \backslash [a, a + T] = [x, a)$. Since x can be arbitrarily close to a, this implies $g(x) \to 0$ as $x \uparrow a$.

The argument for the possibility $g(a) = \infty$ is similar. For any $x < a$, $L_T(f, x) \in [x, a]$ or $L_T(f, x) = \infty$. The last instance, however, is not possible when $x > a - T$, since otherwise by Lemma 8.3.1 the interval I will not be the largest interval on which g is ∞. Thus $L_T(f, x) \in [x, a]$, which then implies that $g(x) \to 0$ as $x \uparrow a$.

In the same spirit, we can show that if I is open at a, then $g(a) = 0$. Assume it is not the case. Then $g(a) = \infty$ or $0 < g(a) \leq T$. If $g(a) = \infty$ and $g(x)$ is also infinity on (a, b) or $(a, b]$, the maximality of the interval I is violated; if $g(x) = d - x$ for any $x \in I$ and some $d \in [b, a + T]$, then $L_T(f, x) = d \in [a, a + T]$, which contradicts with $L_T(f, a) = g(a) + a = \infty$ according to the definition of local intrinsic location functional. Hence we must have $0 < g(a) \leq T$. Consider a point $s \in (a, \min(a + g(a), b))$. $L_T(f, s) = d \in [b, a + T] \subseteq [s, a + T]$. However $L_T(f, a) = a + g(a) \in [s, a + T]$, thus $L_T(f, s) = L_T(f, a)$, contradicting with the openness of I on a. Therefore both of the two possibilities fail and $g(a)$ must take value 0.

Now let us turn to the other direction of the proof. The measurability and shift invariance is already given. The value range $L_T(f,a) \in [a, a+T] \cup \{\infty\}$ for any $f \in H$ and $a \in \mathbb{R}$ is easy to check. It remains condition (4) in Definition 5. Before we proceed, notice that the combination rule determines the following fact:

Lemma 8.4.1. *Let $g : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ be a function satisfying the combination rule. Then for $x, y \in \mathbb{R}, x < y$ satisfying $g(x) \neq \infty$ and $g(y) \neq \infty$, $g(x) - g(y) \leq y - x$. The equality holds if and only if x and y are in the same maximal interval in Theorem 8.4.1. Equivalently, let $L_T(t) = g(t) + t$ for $t \in \mathbb{R}$, then for $x, y \in \mathbb{R}, x < y$ satisfying $L_T(x) \neq \infty$ and $L_T(y) \neq \infty$, $L_T(x) \leq L_T(y)$. The equality holds if and only if x and y are in the same maximal interval in Theorem 8.4.1.*

The proof of this lemma is easy and omitted here.

Let $y_1 < y_2$ be two arbitrary points on real line. We can assume that $y_2 - y_1 \leq T$, since otherwise the condition $L_T(f, y_2) \in [y_1, y_1 + T]$ can never be satisfied. There are two cases. Case 1: y_1 and y_2 are in the same interval I, on which $g(x) = d - x$ or $g(x) = \infty$. Clearly, in this case $L_T(f, y_1) = L_T(f, y_2)$. Case 2: y_1 and y_2 are not in the same interval. Say, $y_2 \in I_2$ and $y_1 \notin I_2$, where $I_2 = [a_2, b_2], [a_2, b_2), (a_2, b_2]$ or (a_2, b_2) is the largest interval containing y_2 on which $g(x) = d - x$ or $g(x) = \infty$. Notice that $L_T(f, y_1) \neq L_T(f, y_2)$, since otherwise the monotonicity implies that $L_T(f, x) = L_T(f, y_2)$ for all $x \in [y_1, y_2]$, contradicting with the assumption that I is the largest interval. Our goal is therefore to prove that in this case, $L_T(f, y_2) \in [y_1, y_1 + T] \cap [y_2, y_2 + T] = [y_2, y_1 + T]$ implies $L_T(f, y_1) \in [y_1, y_2)$.

Firstly, $L_T(f, y_1)$ can not be infinity. Otherwise, let I_1 be the largest interval containing y_1 on which the location takes value ∞. By the combination rule $\lim_{y\downarrow b_1} g(y) = T$, where b_1 is the right endpoint of I_1. $b_1 \geq y_1$ so $y_2 - b_1 \leq y_2 - y_1$. Meanwhile $L_T(f, y_2) \in [y_2, y_1+T]$ implies that $g(y_2) = L_T(f, y_2)-y_2 \leq y_1+T-y_2$, thus $\lim_{y\downarrow b_1} g(y) - g(y_2) = T - g(y_2) \geq y_2 - y_1$. If equality actually holds for both this inequality and the previous one, then $y_1 = b_1$, and $\lim_{y\downarrow b_1} g(y) - g(y_2) = y_2 - y_1$, hence also $\lim_{y\downarrow y_1} g(y) - g(y_2) = y_2 - y_1$. By Lemma 8.4.1, $y_1 \geq a_2$, where a_2 is the left endpoint of the maximal interval I_2 containing y_2. Since $y_1 \notin I_2$, $y_1 = a_2$ and I_2 is open at y_1. However, by combination rule, this implies that $g(y_1) = T \neq \infty$, contradiction. Thus the two inequalities can not be equalities at the same time. As a result, $\lim_{y\downarrow b_1} g(y) - g(y_2) > y_2 - b_1$, which, however, contradicts with Lemma 8.4.1. Thus $L_T(f, y_1) \neq \infty$.

Next, notice that $\lim_{y\downarrow a_2} L_T(f, y) = L_T(f, y_2) \in [y_2, y_1 + T]$. If $g(a_2) = T$, then $g(a_2) - \lim_{g\downarrow} g(y) \geq 0 = \lim_{y\downarrow a_2} -a_2$. According to Lemma 8.4.1, this can only happen if $a_2 \in I_2$. However, $y_1 \leq a_2 \leq L_T(f, a_2) \leq y_1 + T$ and $T = g(a_2) = L_T(f, a_2) - a_2$ implies that $y_1 = a_2$. Together we have $y_1 \in I_2$, contradiction. Thus $g(a_2) \neq T$. Therefore by combination rule, $\lim_{y\uparrow a_2} g(y) = 0$. If $a_2 < y_2$, then by the monotonicity of $L_T(f, \cdot)$ given by Lemma 8.4.1, $L_T(f, y_1) \leq \lim_{y\uparrow a_2} L_T(f, y) = a_2 \in [y_1, y_2)$. Therefore we only need to consider the case where $a_2 = y_2$. Suppose that in this case $L_T(f, y_1) = a_2 = y_2$. By the monotonicity of $L_T(f, \cdot)$ and the fact that $\lim_{y\uparrow a_2} L_T(f, y) = a_2 = y_2$, b_1 must be equal to y_2, where b_1 is the right endpoint of the maximal interval I_1 containing y_1. $g(b_1) = g(y_2) \neq 0$, otherwise $L_T(f, y_1) = L_T(f, y_2)$, implying that y_1 and y_2 are in the same maximal interval. The combination rule then implies that $\lim_{y\downarrow y_2} g(y) = T$. Moreover, since I_1 is open at y_2, $g(y_2) = T$. This contradicts with the observation that $L_T(f, y_2) = g(y_2) + y_2 \in [y_2, y_1 + T]$. To conclude,

$L_T(f, y_1) < y_2$, hence $L_T(f, y_1) \in [y_1, y_2)$. The other direction of Theorem 8.4.1 is therefore proved.

$\square$

CHAPTER 9

CONTINUITY IN DISTRIBUTION AND APPROXIMATION

9.1 Introduction

One of the main difficulties of working with the locations is the fact that they are seldom continuous as functionals of the path under most commonly used norms. For example, a small perturbation of the path can totally change the location of the path supremum, especially if the original supremum location has a close competitor. Consequently, the validity of any approximation becomes unclear, since approximating the path, no matter how close, does not necessarily result in approximating the corresponding location at the same time. Even when we go to the distributional level, it is in general hard to tell whether two stationary processes with similar distributions will also have similar distributions for the random locations.

In this chapter we will try to give some answer to the question of continuity and approximation. Due to the difficulty discussed above, we will concentrate ourselves on a special case: the location of the path supremum over a compact interval or hypercube for stationary Gaussian processes or random fields satisfying some conditions. This is, of course, restrictive; however, it may give us an idea on how this type of questions can be solved in general for other random locations. The Gaussian restriction also provides a convenient representation of the processes by their covariance functions, or, equivalently, by their spectral measures.

The rest of this chapter is organized as follows. In Section 9.2 we introduce the basic setting and notation for this chapter. In Section 9.3 the first version of the continuity result is stated and proved. The assumptions are checked and simplified in Section 9.4, where another version of the continuity theorem is formulated.

9.2 Notation and Settings

Let $\{X^n(\mathbf{t})\}_{n=0,1,...}$ be a family of $d-$dimensional centered and normalized stationary Gaussian random fields, with covariance functions h^n. That is, all the marginal distributions have expectation 0 and variance 1, and $h^n(\mathbf{s}) = Cov(X^n(\mathbf{t}), X^n(\mathbf{t}+\mathbf{s})) = E(X^n(\mathbf{t})X^n(\mathbf{t}+\mathbf{s}))$ for any $\mathbf{s}$ and $\mathbf{t} \in \mathbb{R}^d$. We are interested in the location of the supremum of $X^n(\mathbf{t})$ over the hypercube $\mathbf{I} := \prod_{i=1}^d [0, T_i]$. Further, we will start with the following three assumptions:

A1) For any n, X^n has finite fourth order spectral moments. As a result, it almost surely has C^1 paths and finite number of local maxima;

A2) There does not exist a finite set $\{\mathbf{t_1}, ..., \mathbf{t_k}\}$, such that $\{X^0(\mathbf{t_1}), ..., X^0(\mathbf{t_k})\}$ are linearly dependent;

A3) Denote by $M^n \geq M_2^n$ the two largest values among all the local maxima of $X^n(\mathbf{t})$ on the hypercube $\mathbf{I}$. We assume that as $\epsilon \to 0$, $P(M^n - M_2^n \leq \epsilon)$ converge to 0 uniformly in n. As a result, for any n, X^n almost surely has a unique global maximum on the hypercube. We denote its location by τ^n. Thus τ^n is the only root for $X^n(t) = M^n$. The reason why we need the convergence to be uniform will be clear later (in the proof). Also, denote the corresponding distribution of τ^n by F^n.

9.3 A First Result on the Continuity in Distribution of the Location of the Path Supremum

Theorem 9.3.1. *Under assumptions A1)-A3), if* $h^n \to h^0$ *pointwisely, and the fourth order spectral moments of* $\{X^n\}_{n=0,1,...}$ *are bounded from above, then* $\tau^n \to \tau^0$ *in distribution.*

Proof. Here we prove the result for the case of dimension 1. We will see that the generalization to the multi-dimensional case is very natural. From now on, we consider a family of one-dimensional stationary Gaussian processes $\{X^n(t)\}_{n=0,1,..}$, restricted to the interval $[0,T]$.

Before we start, let us fix some terminology. By a "δ-mesh" we mean a finite set of points in $[0,T]$, $A = \{t_0, ..., t_K\}, 0 = t_0 < t_1 < ... < t_K = T$, such that $\sup_{1 \le i \le K}(t_i - t_{i-1}) \le \delta$. In the proof we actually need a sequence of δ_m-meshes A_m, such that $\delta_m \to 0$ as $m \to \infty$. Associated to the δ_m-mesh, for n=0,1,... define τ^{nm} to be the location of the global maximum over $\{X^n(A_m)\}$, with the corresponding distribution function F^{nm}. We say a δ_m-mesh "catches" the global maximum τ^n, if $\tau^{nm} = \lfloor \tau^n \rfloor_m$ or $\tau^{nm} = \lceil \tau^n \rceil_m$, where $\lfloor x \rfloor_m = \sup\{t \in A_m : t \le x\}$ and $\lceil x \rceil_m = \inf\{t \in A_m : t > x\}$. Otherwise we say the δ_m-mesh "misses" it.

Now we start the proof. We proceed through the three following steps:

1. Prove that

$$P(\text{the } \delta_m\text{-mesh catches the global maximum } \tau^n) = 1 - P(\text{the } \delta_m\text{-mesh miss it}) \to 1$$

uniformly in n as $m \to \infty$.

2. Prove that for any $t \in [0, T]$,

$$|F^{nm}(t) - F^n(t)| \to 0$$

uniformly in n as $m \to \infty$.

3. Prove that for any $t \in [0, T]$, any given m, we have

$$|F^{nm}(t) - F^{0m}(t)| \to 0, \text{ as } n \to \infty.$$

Step 1 is a preparatory step for step 2, while the other two steps are directly related to the proof of the final result. Actually, once we have finished these three steps, the final result follows easily: Given $t \in [0, T]$, for any $\epsilon > 0$, by step 2 we can take m large enough, such that

$$\sup_n |F^{nm}(t) - F^n(t)| < \frac{\epsilon}{3};$$

meanwhile, by step 3, there exists a $N(m)$, such that

$$|F^{nm}(t) - F^{0m}(t)| < \frac{\epsilon}{3}, \ \forall n \geq N(m).$$

Thus $\forall n \geq N(m)$, we have

$$\begin{aligned} |F^n(t) - F^0(t)| &\leq |F^n(t) - F^{nm}(t)| + |F^{nm(t)} - F^{0m}(t)| + |F^{0m}(t) - F^0(t)| \\ &\leq \tfrac{\epsilon}{3} + \tfrac{\epsilon}{3} + \tfrac{\epsilon}{3} \\ &\leq \epsilon. \end{aligned}$$

Now we go back and see the proofs of steps 1-3.

Proof of step 1: $\forall n, m, \epsilon > 0$,

$$\begin{aligned} & \{\text{the } \delta_m\text{-mesh misses } \tau^n\} \\ \subset \ & \{M^n - M_2^n \leq \epsilon\} \cup \{M^n - \max\{X^n(\lfloor \tau^n \rfloor_m), X^n(\lceil \tau^n \rceil_m)\} > \epsilon\} \\ \subset \ & \{M^n - M_2^n \leq \epsilon\} \cup \{\sup_{t \in [0,T]} |\dot{X}^n(t)| > \tfrac{\epsilon}{\delta_m}\} \end{aligned}$$

where the $\dot{X}^n(t)$ is the first order derivative in the classical sense of X^n at the point t. Since the fourth spectral moments are finite, the processes also have first order quadratic mean derivatives with probability 1, and the two type of derivatives should be equal. Thus,

$$\{\text{the } \delta_m\text{-mesh misses } \tau^n\} \subset \{M^n - M_2^n \leq \epsilon\} \cup \{\sup_{t\in[0,T]} |X^{n'}(t)| > \frac{\epsilon}{\delta_m}\},$$

where $X^{n'}(t)$ denotes the first order derivative in quadratic mean, here and later. Therefore,

$$\begin{aligned} P(\text{the } \delta_m\text{-mesh misses } \tau^n) &\leq P(M^n - M_2^n \leq \epsilon) \\ &+ P(\sup_{t\in[0,T]} |X^{n'}(t)| > \tfrac{\epsilon}{\delta_m}, \text{ the } \delta_m\text{-mesh misses } \tau^n). \end{aligned}$$

Now take $\epsilon = \sqrt{\delta_m}$. As $\delta_m \to 0$ when $m \to \infty$, $\epsilon \to 0$, and $\frac{\epsilon}{\delta_m} = \frac{1}{\sqrt{\delta_m}} \to \infty$. By assumption A3), $P(M^n - M_2^n \leq \epsilon) \to 0$ uniformly in n. Hence it suffices to focus on the second term on the right hand side. For that, notice that if the δ_m-mesh misses the global maximum τ^n, then τ^n can not be 0 or T, thus τ^n is in the interior of $[0, T]$. Then at τ^n the first order derivative $X^{n'}(\tau^n) = 0$. Because the processes have C^1 paths almost surely, $\sup_{t\in[0,T]} |X^{n'}(t)| > \frac{\epsilon}{\delta_m}$ implies that the first order derivative $X^{n'}$ need to have at least one upcrossing of the level $\frac{\epsilon}{\delta_m}$, or one downcrossing of the level $-\frac{\epsilon}{\delta_m}$. Therefore, we have

$$\begin{aligned} & P(\text{the } \delta_m\text{-mesh misses } \tau^n) - P(M^n - M_2^n \leq \epsilon) \\ \leq\ & P(\text{upcrossing of } \tfrac{\epsilon}{\delta_m} \text{ or downcrossing of } -\tfrac{\epsilon}{\delta_m} \text{ by } X^{n'} \text{ on } [0,T]) \\ =\ & 2P(N^n_{\frac{\epsilon}{\delta_m}} \geq 1) \\ \leq\ & 2E(N^n_{\frac{\epsilon}{\delta_m}}), \end{aligned}$$

where $N^n_{\frac{\epsilon}{\delta_m}}$ is the number of upcrossings of $X^{n'}(t)$ of the level $\frac{\epsilon}{\delta_m}$ on $[0, T]$. By Rice's formula,

$$E\left(N^n_{\frac{\epsilon}{\delta_m}}\right) = \frac{T}{2\pi} (\lambda_4^n)^{\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\frac{\epsilon}{\delta_m}\right)^2\right).$$

As $\{\lambda_4^n\}$ is bounded, $E\left(N^n_{\frac{\epsilon}{\delta_m}}\right) \to 0$ as $m \to \infty$ uniformly in n. Thus we have proved the step 1: $P(\text{the } \delta_m\text{-mesh misses } \tau^n) \to 0$ uniformly in n.

Now we start to prove the step 2: $\forall t \in [0, T]$, $F^{nm}(t) \to F^n(t)$ uniformly in n.

First notice that the result holds trivially for $t = T$. For any given $t \in (0, T)$, the proof proceeds in the following way:

Notice that on $\{\text{the } \delta_m\text{-mesh catches } \tau^n\}$, $\tau^{nm} \leq t$ implies that $\tau^n \leq \lceil t \rceil_m$, while $\tau^n \leq \lfloor t \rfloor_m$ in turn implies $\tau^{nm} \leq t$. For each given m, we decompose each distribution function $F^{nm}(t)$ (resp. $F^n(t)$) into two parts, the first part $F_1^{nm}(t)$ (resp. $F_1^n(t)$) is the probability that the global maximum happens before t, and the δ_m-mesh catches τ^n, and the second part $F_2^{nm}(t)$ (resp. $F_2^n(t)$) is the remaining part, that is, the probability that a global maximum happens before t, and the δ_m-mesh misses τ^n. The key idea for the proof is that F_2 is always smaller than $P(\text{the } \delta_m\text{-mesh misses } \tau^n)$ (denoted as P_m), which converge uniformly to 0 as $m \to \infty$. More precisely, we have

$$F^{nm}(t) = F_1^{nm}(t) + F_2^{nm}(t).$$

As

$$F_1^n(\lfloor t \rfloor_m) \leq F_1^{nm}(t) \leq F_1^n(\lceil t \rceil_m)$$

and

$$0 \leq F_2^{nm}(t) \leq P_m,$$

$$F_1^n(\lfloor t \rfloor_m) \leq F^{nm}(t) \leq F_1^n(\lceil t \rceil_m) + P_m.$$

On the other hand, for $F_1^n(t)$, we have

$$F_1^n(\lfloor t \rfloor_m) \leq F_1^n(t) \leq F_1^n(\lceil t \rceil_m).$$

Hence

$$|F_1^n(t) - F^{nm}(t)| \leq F_1^n(\lceil t \rceil_m) - F_1^n(\lfloor t \rfloor_m) + P_m.$$

Moreover,

$$F_1^n(\lceil t \rceil_m) \leq F^n(\lceil t \rceil_m),$$

$$F_1^n(\lfloor t \rfloor_m) = F^n(\lfloor t \rfloor_m) - F_2^n(\lfloor t \rfloor_m) \geq F^n(\lfloor t \rfloor_m) - P_m,$$

thus

$$|F_1^n(t) - F^{nm}(t)| \leq F^n(\lceil t \rceil_m) - F^n(\lfloor t \rfloor_m) + 2P_m.$$

Finally, noticing that

$$|F_1^n(t) - F^n(t)| \leq P_m,$$

we have

$$|F^n(t) - F^{nm}(t)| \leq F^n(\lceil t \rceil_m) - F^n(\lfloor t \rfloor_m) + 3P_m.$$

By step 1, P_m converges to 0 uniformly on n when $m \to \infty$. For the first term on the right hand side, remember that we have the universal bound of the density function (3.1) given in Chapter 3, thus $F^n(\lceil t \rceil_m) - F^n(\lfloor t \rfloor_m)$ converges to 0 uniformly in n as m goes to infinity. Therefore the uniform convergence is proved.

We turn to the step 3 now. Notice that $X^n(A_m)$, $n = 1, 2, ...$, as a sequence Gaussian random vectors, converge to $X^0(A_m)$ in distribution. Thus by Skorohod embedding theorem, there exists another sequence of Gaussian random variables $Y^n(A_m)$, $n = 1, 2, ...$, such that $X^n(A_m)$ and $Y^n(A_m)$ have the same distribution for any fixed n, and $Y^n(A_m)$ converge to $X^0(A_m)$ almost surely. By Assumption A2), the covariance matrix of $X^0(A_m)$ is not degenerate, thus $X^0(A_m)$ has almost surely a unique maximal coordinate. Therefore the maximal coordinate of $Y^n(A_m)$ will converge to the maximal coordinate of $X^0(A_m)$ almost surely, hence also in distribution. This means, back to X^n, that the maximal coordinate of $X^n(A_m)$ also converge to the maximal coordinate of $X^0(A_m)$ in distribution, which clearly implies that $|F^{nm}(t) - F^{0m}(t)| \to 0$ for any $t \in [0, T]$ and any given m. Thus we have finished the proof of step 3, so is the whole proof. $\square$

9.4 An alternative version of the continuity result

The condition in Assumption A3) that the convergence of $P(M^n - M_2^n \leq \epsilon)$ to 0 as $\epsilon \to 0$ be uniform on n is not directly given in terms of the standard characterizations of a stationary Gaussian process or random fields. It is desirable to replace it by conditions easier to check, such as conditions on spectral measures. Moreover, it turns out that Assumption A2) can be dropped by adding a small perturbation to the original process if A2) is violated. In this section, we will make these two points precise, and give out a more succinct version of the continuity result proved in the last section. For the rest of this section we work in the case of dimension 1, where I is the compact interval $[0, T]$.

Let us recall Assumption U_T for a stochastic process X appeared in Chapter 3:

Assumption U$_T$: $P(\Omega_T) = 0$, where

$$\Omega_T = \big\{\omega \in \Omega : \, X(t_i) = \sup_{t \in [0,T]} X(t) \text{ for at least two different } t_1, t_2 \in [0, T]\big\}.$$

Under Assumption U_T for the process X^n, $n = 0, 1, \ldots$, we have the following theorem:

Theorem 9.4.1. *Let $\{X^n(t)\}_{n=0,1,...}$ be a family of one dimensional standardized Gaussian processes. Assume for $n = 0, 1, ...$, X^n satisfies Assumption U_T, and has finite sixth spectral moment. If the spectral measures of $X^n(t), n = 1, 2, ...$ converge to the spectral measure of $X^0(t)$ weakly, then τ_n converge to τ_0 in distribution.*

Proof. First of all, notice the following facts:

1. Since X^n always has finite sixth spectral moments, they have almost surely C^2 paths. Reader can refer to Chapter 7 in [15] for details.
2. Since X^0 satisfies Assumption U_T, its second spectral moment $\lambda_2^0 > 0$, and $\lambda_4^0 - (\lambda_2^0)^2 > 0$. (Otherwise, X^0 will be a single triangular wave, for which Assumption U_T is violated.) Moreover, since the spectral measures of $\{X^n\}_{n=1,2,...}$ converge to the spectral measure of X^0, for n large enough, $\{\lambda_2^n\}$ and $\{\lambda_4^n - (\lambda_2^n)^2\}$ are bounded away from 0.
3. Since X^n has finite sixth spectral moment for $n = 0, 1, ...$ and the spectral measures of X^n converge to the spectral measure of X^0 weakly, their sixth spectral moments are bounded from above. Consequently, their fourth spectral moments are also bounded from above.

Having these facts in mind, it is not difficult to see that in order to prove Theorem 9.4.1, all we need is to prove the following two lemmas.

Lemma 9.4.1. *Under the assumption A1) and A2), if we have, moreover, that both the second spectral moments λ_2^n and the quantity $\lambda_4^n - (\lambda_2^n)^2$ are bounded from 0, and that the sixth spectral moments λ_6^n are bounded from above, then we can drop the condition that the convergence of $P(M^n - M_2^n \leq \epsilon)$ to 0 as $\epsilon \to 0$ be uniform on n in Theorem 9.3.1.*

Remark 9.4.2. The conditions on the second and fourth spectral moments in this lemma are natural and should be expected, since what they basically say is just that the processes should not be "too near" to the single sin/cos function case.

Proof. From the proof of Theorem 9.3.1, we know that we do not really need the convergence of $P(\text{the } \delta_m\text{-mesh misses } \tau^n)$ to 0 as $m \to \infty$ to be uniform on n. What we actually need is that for any $\epsilon > 0$, there exist a m and a $N(m)$, such that $P(M^0 - M_2^0 \leq \sqrt{\delta_m}) \leq \epsilon$ and $P(M^n - M_2^n \leq \sqrt{\delta_m}) \leq \epsilon$ for any $n \geq N(m)$. A closer check of the proof for step 1 in Theorem 9.3.1 shows that the changes start at the decomposition

$$\begin{aligned} & P(\text{the } \delta_m\text{-mesh misses } \tau^n) \\ \leq \ & P(M^n - M_2^n \leq \sqrt{\delta_m}, \text{ and the } \delta_m\text{-mesh misses } \tau^n) \\ & + P(\sup_{t\in[0,T]} |X^{n'}(t)| > \tfrac{1}{\sqrt{\delta_m}}, \text{ and the } \delta_m\text{-mesh misses } \tau^n). \end{aligned}$$

The uniform convergence of the second term on the right hand is still valid, thus to show the result above, we need to show for all $\epsilon > 0$, there exists m_0, such that for all $m \geq m_0$, we have

$$P(M^0 - M_2^0 \leq \sqrt{\delta_m} \text{ and the } \delta_m\text{-mesh misses } \tau^n) < \epsilon,$$

and there exists $N(m)$, such that for all $n \geq N(m)$,

$$P(M^n - M_2^n \leq \sqrt{\delta_m} \text{ and the } \delta_m\text{-mesh misses } \tau^n) < \epsilon.$$

That is,

$$\lim_{m\to\infty} P(M^0 - M_2^0 \leq \sqrt{\delta_m} \text{ and the } \delta_m\text{-mesh misses } \tau^n) = 0,$$

and

$$\lim_{m\to\infty} \lim_{n\to\infty} P(M^n - M_2^n \leq \sqrt{\delta_m} \text{ and the } \delta_m\text{-mesh misses } \tau^n) = 0.$$

Firstly, let us fix some more terminology. We call a point in a discretized process X^{nm} to be a local m-maximum, if it is larger than both of its two neighbors in the δ_m-mesh. A local maximum of X^n, whose location denoted by l, is said to be "detected" in the δ_m-mesh, if either $\lfloor l \rfloor_m$ or $\lceil l \rceil_m$ is a local m-maximum. Define the event

$$D_{nm} := \{\text{both } \tau^n \text{ and } \tau_2^n \text{ are detected by the } \delta_m\text{-mesh}\},$$

where τ_2^n is the location of the second largest local maximum (potentially including 0 and T). Denote the value of the two largest m-local maxima of X^{nm} to be M^{nm} and M_2^{nm}, respectively.

We know that when $M^n - M_2^n \leq \sqrt{\delta_m}$, if both τ^n and τ_2^n are detected, with the corresponding locations τ^{nm}, τ_2^{nm} (notice that they are not necessarily the location of the two largest local m-maxima), and if the sample path is Lipschitz with constant $\frac{1}{\sqrt{\delta_m}}$, then

$$M^n - \delta_m \cdot \frac{1}{\sqrt{\delta_m}} \leq X^n(\tau^{nm}) \leq M^{nm} \leq M^n.$$

Moreover, we are only paying attention to the set {the δ_m-mesh misses τ^n}, so if τ^n is detected, its value can not be larger than M_2^{nm}. Thus we actually have

$$M^n - \delta_m \cdot \frac{1}{\sqrt{\delta_m}} \leq X^n(\tau^{nm}) \leq M_2^{nm} \leq M^{nm} \leq M^n,$$

which implies $0 \leq M^{nm} - M_2^{nm} \leq \sqrt{\delta_m}$.

The key to the proof is therefore the decomposition

$$\begin{aligned}
& P(M^n - M_2^n \leq \sqrt{\delta_m} \text{ and the } \delta_m\text{-mesh misses } \tau^n) \\
\leq\ & P(0 \leq M^{nm} - M_2^{nm} \leq \sqrt{\delta_m}) \\
+\ & P(\tau^n \text{ or } \tau_2^n \text{ is not detected by the } \delta_m\text{-mesh}) \\
+\ & P(X^n(t) \text{ is not Lipschitz with constant } \tfrac{1}{\sqrt{\delta_m}}, \text{ and the } \delta_m\text{-mesh misses } \tau^n).
\end{aligned}$$

We start by the easiest term, the third term on the right hand side. As X^n almost surely has C^1 path, the sample function is not Lipschitz with constant $\frac{1}{\sqrt{\delta_m}}$, if and only if there exist $t \in [0, T]$, such that $|X^{n'}(t)| > \sqrt{\delta_m}$. On the other hand, if the global maximum is missed, then it can not take place at 0 or T, so it is a local maximum, having derivative 0. By continuity of the derivative we know there is at least one upcrossing of the first derivative to the level $\sqrt{\delta_m}$, or one downcrossing of the first derivative to the level $-\sqrt{\delta_m}$. As in the proof of the theorem, a simple application of Rice's formula shows that this term converges to 0 uniformly on n when $m \to \infty$, provided that the fourth spectral moments $\{\lambda_4^n\}$ are bounded from above.

Now for the second term on the right hand side,

$$\begin{aligned} & P(\tau^n \text{ or } \tau_2^n \text{ is not detected by the } \delta_m\text{-mesh}) \\ \leq \;& P(\text{there exists at least on local maximum of } X^n(t), \text{ with width less than } \delta_m), \end{aligned}$$

where the "width" of a local maximum located at l is defined as the largest length x, such that the path is increasing on $[l - x, l]$, and decreasing on $[l, l + x]$. This, in turn, implies that the derivative $X^{n'}$ has two zeros in a distance smaller than δ_m, moreover, their second derivatives (Notice that as the sixth spectral moments are finite, the second derivative almost surely exists and is continuous.) have different signs, or at least one of them is 0. Then there are two possibilities: 1. the seconde derivative $X^{n''}$ is either larger than $\frac{1}{\sqrt{\delta_m}}$ or smaller than $-\frac{1}{\sqrt{\delta_m}}$ for some points on $[0, T]$; 2. otherwise, there is a local maximum or minimum of $X^{n'}$, whose absolute value is smaller or equal to $\delta_m \cdot \frac{1}{\sqrt{\delta_m}} = \sqrt{\delta_m}$. Since the second derivatives at the two zeros of the first derivative have different signs or at least one of them is 0, by continuity there exist a point on $[0, T]$, for which the second derivative takes value 0. Combining this information, under possibility 1 we must have at least one upcrossing of the second derivative $X^{n''}$

to the level $\frac{1}{\sqrt{\delta_m}}$, or one downcrossing to the level $-\frac{1}{\sqrt{\delta_m}}$. Therefore

$$\begin{aligned}
& P(\tau^n \text{ or } \tau_2^n \text{ is not detected by the } \delta_m\text{-mesh}) \\
\leq \;& P(\text{there exists at least one local maximum of } X^n(t), \text{ with width less than } \delta_m) \\
\leq \;& P(X^{n''} \text{ upcrosses the level } \tfrac{1}{\sqrt{\delta_m}} \text{ or downcrosses the level } -\tfrac{1}{\sqrt{\delta_m}}) \\
+ \;& P(X^{n'} \text{ has a local extreme, whose absolute value is smaller or equal to } \sqrt{\delta_m}).
\end{aligned}$$

Remember that when $m \to \infty$, $\sqrt{\delta_m} \to 0$ and $\frac{1}{\sqrt{\delta_m}} \to \infty$. Thus similar to the proof of step 1 in Theorem 9.3.1, since the sixth spectral moments $\{\lambda_6^n\}$ are bounded, the first term in the last expression converge to 0 as $m \to \infty$ uniformly on n. For the second term, it suffices to apply the result (7.6.3) on page 161 in [15]:

Assume $\xi(t)$ is standardized (with mean 0 and unit variance), stationary Gaussian process, with finite second and fourth spectral moments λ_2 and λ_4. If moreover $(\xi(t), \xi'(t), \xi''(t))$ have a nonsingular distribution, then the number of global maxima of $\xi(t)$ between (0,T) for which the value of the process is greater than u, denoted as $N'_u(T)$, satisfies

$$E(N'_u(T)) = \frac{T}{2\pi}\left\{\left(\frac{\lambda_4}{\lambda_2}\right)^{\frac{1}{2}}\left(1 - \Phi\left(u\left(\frac{\lambda_4}{D}\right)^{\frac{1}{2}}\right)\right) + (2\pi\lambda_2)^{\frac{1}{2}}\phi(u)\Phi\left(\frac{u\lambda_2}{D^{\frac{1}{2}}}\right)\right\},$$

where $D = \lambda_4 - \lambda_2^2$.

Notice that $(\xi(t), \xi'(t), \xi''(t))$ are jointly normal with mean zero and the covariance matrix

$$\begin{bmatrix} \lambda_0 & 0 & -\lambda_2 \\ 0 & \lambda_2 & 0 \\ -\lambda_2 & 0 & \lambda_4 \end{bmatrix},$$

therefore they have nonsingular distribution if and only if $D > 0$, which is indicated here by the assumption that $\{\lambda_4 - \lambda_2^2\}$ is bounded from 0.

This result show that, under the assumptions of Lemma 9.4.1,

$$\begin{aligned} & P(X^{n'} \text{ has a local extreme, whose absolute value is smaller or equal to } \sqrt{\delta_m}) \\ \leq \ & 2P(X^{n'} \text{ has a local maximum, whose value is between } -\sqrt{\delta_m} \text{ and } \sqrt{\delta_m}) \\ \leq \ & 2(E(N'_{\sqrt{\delta_m}}) - E(N'_{-\sqrt{\delta_m}})), \end{aligned}$$

which converge to 0 as $m \to \infty$ uniformly on n.

Finally we turn to the first term on the right hand side. We start by looking at the case where $n = 0$.

$$\begin{aligned} & P(0 \leq M^{0m} - M_2^{0m} \leq \sqrt{\delta_m}) \\ \leq \ & P(0 \leq M^{0m} - M_2^{0m} \leq \sqrt{\delta_m}, \text{ and } X^n(t) \text{ is Lipschitz with constant } \tfrac{1}{\sqrt{\delta_m}}) \\ + \ & P(X^n(t) \text{ is not Lipschitz with constant } \tfrac{1}{\sqrt{\delta_m}}) \\ \leq \ & P(\{0 \leq M^{0m} - M_2^{0m} \leq \sqrt{\delta_m}, \\ & \quad \text{and } X^n(t) \text{ is Lipschitz with constant } \tfrac{1}{\sqrt{\delta_m}}\} \cap D_{0m}) \\ + \ & 1 - P(D_{0m}) + P(X^n(t) \text{ is not Lipschitz with constant } \tfrac{1}{\sqrt{\delta_m}}). \end{aligned}$$

As we have just seen, $1 - P(D_{0m}) \to 0$ when $m \to \infty$, so is the term

$$P(X^n(t) \text{ is not Lipschitz with constant } \frac{1}{\sqrt{\delta_m}}).$$

For

$$P(\{0 \leq M^{0m} - M_2^{0m} \leq \sqrt{\delta_m}, \text{ and } X^n(t) \text{ is Lipschitz with constant } \frac{1}{\sqrt{\delta_m}}\} \cap D_{0m}),$$

notice that since both L^0 and L_2^0 are detected, and the sample function is Lipschitz with constant $\frac{1}{\sqrt{\delta_m}}$,

$$M^{0m} \leq M^0 \leq M^{0m} + \delta_m \cdot \frac{1}{\sqrt{\delta_m}} = M^{0m} + \sqrt{\delta_m},$$

and

$$M_2^{0m} \leq M_2^0 \leq M^{0m} + \sqrt{\delta_m}.$$

As a result,

$$\begin{aligned} & P(\{0 \leq M^{0m} - M_2^{0m} \leq \sqrt{\delta_m}, \text{ and } X^n(t) \text{ is Lipschitz with constant } \tfrac{1}{\sqrt{\delta_m}}\} \cap D_{0m}) \\ \leq \; & P(M^0 - M_2^0 \leq 2\sqrt{\delta_m}), \end{aligned}$$

which converges to 0 as $m \to \infty$.

For $n = 1, 2, ...$, the idea is to compare them to the case where $n = 0$, and show that their difference can be arbitrarily small when n is large enough.

Recall that we always have

$$\begin{aligned} & P(M^n - M_2^n \leq \sqrt{\delta_m}) \\ \leq \; & P(0 \leq M^{nm} - M_2^{nm} \leq \sqrt{\delta_m}) \\ + \; & P(\tau^n \text{ or } \tau_2^n \text{ is not detected by the } \delta_m\text{-mesh}) \\ + \; & P(X^n(t) \text{ is not Lipschitz with constant } \frac{1}{\sqrt{\delta_m}}, \text{ and the } \delta_m\text{-mesh misses } \tau^n). \end{aligned} \tag{9.1}$$

As we have seen, the second and the third term on the right hand side converge as $m \to \infty$ uniformly on n, so there is no problem for them. Thus for any $\epsilon > 0$, there exists a m, such that

$$\begin{aligned} & P(\tau^n \text{ or } \tau_2^n \text{ is not detected by the } \delta_m\text{-mesh}) \\ + \; & P(X^n(t) \text{ is not Lipschitz with constant } \tfrac{1}{\sqrt{\delta_m}}, \text{ and the } \delta_m\text{-mesh misses } \tau^n). \\ \leq \; & \tfrac{\epsilon}{3} \end{aligned}$$

for any $n = 1, 2....$ For the first term, similar to the step 3 of the proof to the theorem, it can be shown that for any given m, there exists $N(m)$, such that for all $n \geq N(m)$,

$$|P(M^n - M_2^n \leq \sqrt{\delta_m}) - P(M^0 - M_2^0 \leq \sqrt{\delta_m})| \leq \frac{\epsilon}{3}.$$

We have seen that $P(M^0 - M_2^0 \leq \sqrt{\delta_m}) \to 0$ when $m \to \infty$, so there exists a m_0, such that for any $m \geq m_0$,

$$P(M^0 - M_2^0 \leq \sqrt{\delta_m}) \leq \frac{\epsilon}{3}.$$

Therefore for any $\epsilon > 0$, there is a m_0, such that for any $m \geq m_0$, there exists $N(m)$, satisfying for any $n \geq N(m)$,

$$\begin{aligned}
& P(M^n - M_2^n \leq \sqrt{\delta_m}) \\
\leq\ & P(0 \leq M^{nm} - M_2^{nm} \leq \sqrt{\delta_m}) \\
+\ & P(\tau^n \text{ or } \tau_2^n \text{ is not detected by the } \delta_m\text{-mesh}) \\
+\ & P(X^n(t) \text{ is not Lipschitz with constant } \tfrac{1}{\sqrt{\delta_m}}, \text{ and the } \delta_m\text{-mesh misses } \tau^n) \\
\leq\ & P(M^0 - M_2^0 \leq \sqrt{\delta_m}) + |P(M^n - M_2^n \leq \sqrt{\delta_m}) - P(M^0 - M_2^0 \leq \sqrt{\delta_m})| \\
+\ & P(\tau^n \text{ or } \tau_2^n \text{ is not detected by the } \delta_m\text{-mesh}) \\
+\ & P(X^n(t) \text{ is not Lipschitz with constant } \tfrac{1}{\sqrt{\delta_m}}, \text{ and the } \delta_m\text{-mesh misses } \tau^n) \\
\leq\ & \tfrac{\epsilon}{3} + \tfrac{\epsilon}{3} + \tfrac{\epsilon}{3} \\
=\ & \epsilon.
\end{aligned}$$

Since in (9.1), the two other terms all converge to 0 when $m \to \infty$ uniformly on n, adding them will not change the result above. Lemma 9.4.1 is therefore proven. □

Lemma 9.4.2. *Let $\{X_n\}_{n=0,1,\ldots}$ be a family of standardized Gaussian processes, satisfying all the conditions listed in Lemma 9.4.1 except for Assumption A2), then the spectral measures of $X^n, n = 1, 2, \ldots$ still converge to the spectral measure of X^0 in distribution.*

Proof. The idea of the proof is to add a "small" disturbance to the original processes when A2) is violated, to transform it to a case where A2) holds. First, we know that a sufficient condition for A2) to hold is that the spectral measure has a continuous part, see page 203 in [9]. Thus if A2) does not hold, then the spectral measure of $X^0(t)$ must be purely discrete. Let $Y(t)$ be another standardized stationary Gaussian process, with continuous spectral measure, finite second, fourth and sixth spectral moments, and independent of all $X^n(t)$. Denote its second, fourth and sixth spectral moments to be λ_2^Y, λ_4^Y and λ_6^Y, respectively.

Let δ be a strictly positive number. By adding $\delta Y(t)$ to all the processes $X^n(t)$, we get a new sequence of stationary Gaussian processes, denoted as $Z^n(t)$. It is easy to see that the covariance function of $Z^n(t)$ is

$$\begin{aligned} R^{Zn}(t) &= \text{Cov}(X^n(0) + \delta Y(0), X^n(t) + \delta Y(t)) \\ &= \text{Cov}(X^n(0), X^n(t)) + \delta^2 \text{Cov}(Y(0), Y(t)) \\ &= R^n(t) + \delta^2 R^Y(t), \end{aligned}$$

where $R^Y(t)$ is the covariance function of $Y(t)$. Therefore if originally the covariance functions of $X^n(t)$ converge to the covariance function of $X^0(t)$, it is also the case for $Z^n(t)$ and $Z^0(t)$. Now as $Z^n(t)$ and $Z^0(t)$ have continuous part in the spectral measures, bounded second, fourth and sixth spectral moments satisfying all the conditions needed for the theorem, the result of the theorem holds. That is, the location of the global maximum of Z^n converge in distribution to the location of the global maximum of Z^0. The next thing which needs to do, is to show that by choosing a δ "small enough", with a high probability we are just changing the location of the global maximum by a small amount.

From the proof of Lemma 9.4.1, we know that $P(M^{Z0} - M_2^{Z0} \leq \delta_1)$ and $\lim_{n\to\infty} P(M^{Zn} - M_2^{Zn} \leq \delta_1)$ can be arbitrarily small as long as δ_1 and δ are small enough, where M^{Zn} is the global maximum of $Z^n(t)$ on $[0, T]$, and M_2^{Zn} is the largest value among the other local maxima of Z^n (potentially including 0 or T). Notice that as λ_2^Y is finite, by Rice's formula, for any $\epsilon > 0$ and δ_1 given, there exists a $\delta > 0$, such that

$$P(\delta |M|^Y \geq \frac{\delta_1}{2}) \leq \frac{\epsilon}{3},$$

where $|M|^Y$ is the global maximum of the absolute value of $Y(t), t \in [0, T]$. Now in the case where $\delta |M|^Y < \frac{\delta_1}{2}$, and $M^{Z0} - M_2^{Z0} > \delta_1$, the global maximum of $X^0(t)$ can not "jump" from one peak to another after adding the component $\delta Y(t)$. Thus intuitively, the only possibility for a big change on the location of the global maximum by adding a small component, is that the sample path is very "flat" around the global maximum. More precisely,

$$
\begin{aligned}
& P(|\tau^{Z0} - \tau^0| \geq \delta_1^{\frac{1}{3}}) \\
\leq \;& P(\delta |M|^Y \geq \tfrac{\delta_1}{2}) \\
+ \;& P(M^{Z0} - M_2^{Z0} \leq \delta_1) \\
+ \;& P(M^0 - \max\{X^0(\tau^0 - \delta_1^{\frac{1}{3}}), X^0(\tau^0 + \delta_1^{\frac{1}{3}})\} \leq \delta_1, \text{ and } M^{Z0} - M_2^{Z0} > \delta_1),
\end{aligned}
$$

where $X^0(\tau^0 - \delta_1^{\frac{1}{3}})$ is defined as $X^0(0)$ if $\tau^0 - \delta_1^{\frac{1}{3}} < 0$, similar for the second term in the max and T. We have seen that the first two terms on the right hand side can be arbitrarily small by properly choosing δ_1 and δ. For the last term, denote it as P_{δ_1}. Now suppose that $\lim_{\delta_1 \to 0} P_{\delta_1} \neq 0$, then the events $\{M^0 - \max\{X^0(\tau^0 - \delta_1^{\frac{1}{3}}), X^0(\tau^0 + \delta_1^{\frac{1}{3}})\} \leq \delta_1\}$ must happen infinitely often as $\delta_1 \to 0$ with a strictly positive probability. By the existence of the second derivative of the sample function, it implies that $P(X^{0''}(\tau^0) = 0) > 0$. However, recall that $(X^0(t), X^{0'}(t), X^{0''}(t))$ forms a centered Gaussian vector with covariance matrix

$$
\begin{bmatrix}
\lambda_0^0 & 0 & -\lambda_2^0 \\
0 & \lambda_2^0 & 0 \\
-\lambda_2^0 & 0 & \lambda_4^0
\end{bmatrix},
$$

and that this matrix is supposed to be non-degenerate, the probability that their exists a local maximum, whose second derivative is 0, is 0. Thus $P(X^{0''}(\tau^0) = 0) = 0$, contradiction. As a result, the assumption that $\lim_{\delta_1 \to 0} P_{\delta_1} \neq 0$ is false, $P_{\delta_1} \to 0$ when $\delta_1 \to 0$.

Now for any given $\epsilon > 0$, it suffices to take δ_1 and δ, such that $P(M^{Z0} - M_2^{Z0} \leq \delta_1) < \frac{\epsilon}{3}$, $P(M^0 - \max\{X^0(\tau^0 - \delta_1^{\frac{1}{3}}), X^0(\tau^0 + \delta_1^{\frac{1}{3}})\} \leq \delta_1) < \frac{\epsilon}{3}$, and $P(\delta |M|^Y \geq \frac{\delta_1}{2}) <$

$\frac{\epsilon}{3}$, to have what we want:

$$P(|\tau^{Z0} - \tau^0| \geq \delta_1^{\frac{1}{3}}) < \epsilon.$$

For the case n is not 0 but large enough, following the decomposition

$$\begin{aligned} & P(|\tau^{Zn} - \tau^n| \geq \delta_1^{\frac{1}{3}}) \\ \leq \; & P(\delta |M|^Y \geq \tfrac{\delta_1}{2}) \\ + \; & P(M^{Zn} - M_2^{Zn} \leq \delta_1) \\ + \; & P(M^n - \max\{X^n(\tau^n - \delta_1^{\frac{1}{3}}), X^n(\tau^n + \delta_1^{\frac{1}{3}})\} \leq \delta_1, \text{ and } M^{Zn} - M_2^{Zn} > \delta_1), \end{aligned}$$

we can see that the first term on the right hand side does not depend on n, so it can be arbitrarily small, just as shown for the case $n = 0$. By Lemma 9.4.1 the second term is also arbitrarily small for n large enough. For the last term, for any m,

$$\begin{aligned} & P(M^n - \max\{X^n(\tau^n - \delta_1^{\frac{1}{3}}), X^n(\tau^n + \delta_1^{\frac{1}{3}})\} \leq \delta_1, \text{ and } M^{Zn} - M_2^{Zn} > \delta_1) \\ \leq \; & P(M^{nm} - I^{nm} \leq \delta_1), \end{aligned}$$

where the quantity $I^{nm} := \max\{\lceil \tau^n - \delta_1^{\frac{1}{3}} \rceil_m, \lfloor \tau^n + \delta_1^{\frac{1}{3}} \rfloor_m\}$. For any given m, we are now familiar with the fact that $P(M^{nm} - I^{nm} \leq \delta_1)$ can be arbitrarily close to $P(M^{0m} - I^{0m} \leq \delta_1)$, as long as n is large enough. However,

$$P(M^{0m} - I^{0m} \leq \delta_1) \to P(M^0 - \max\{X^0(\tau^0 - \delta_1^{\frac{1}{3}}), X^0(\tau^0 + \delta_1^{\frac{1}{3}})\} \leq \delta_1)$$

when the discretization index $m \to \infty$. Since δ_1 is chosen such that $P(M^0 - \max\{X^0(\tau^0 - \delta_1^{\frac{1}{3}}), X^0(\tau^0 + \delta_1^{\frac{1}{3}})\} \leq \delta_1) < \frac{\epsilon}{3}$, there always exists m large enough and the corresponding $N(m)$, such that for any $n \geq N(m)$,

$$P(M^n - \max\{X^n(\tau^n - \delta_1^{\frac{1}{3}}), X^n(\tau^n + \delta_1^{\frac{1}{3}})\} \leq \delta_1, \text{ and } M^{Zn} - M_2^{Zn} > \delta_1) < \frac{\epsilon}{3}.$$

Combining this with the previous case where $n = 0$, we have proved that for any $\epsilon > 0$, and any δ_1 small enough, there always exists a coefficient $\delta > 0$, such that $P(|\tau^{Z0} - \tau^0| \geq \delta_1^{\frac{1}{3}}) < \epsilon$, and $P(|\tau^{Zn} - \tau^n| \geq \delta_1^{\frac{1}{3}}) < \epsilon$ for any n large enough. This leads to the convergence of τ^n to τ^0 in distribution by an argument similar to the step 2 of the proof for Theorem 9.3.1. $\square$

$\square$

BIBLIOGRAPHY

[1] R. Adler. *An introduction to continuity, extrema and related topics for general Gaussian processes. IMS Lectures Notes, Volume 12*. Institute of Mathematical Statistics, Hayward, 1990.

[2] R. Adler and J. Taylor. *Random Fields and Geometry*. Springer, New York, 2007.

[3] J.M. Azaïs and M. Wschebor. The distribution of the maximum of a gaussian process: Rice method revisited. *In and out of equilibrium: Probability with a physical flavour*, Birkhauser, Boston: 321–348, 2002.

[4] J.M. Azaïs and M. Wschebor. *Level Sets and Extrema of Random Processes and Fields*. John Wiley & Sons, Hoboken, 2009.

[5] P. Billingsley. *Convergence of Probability Measures*. Wiley, New York, 2nd edition, 1999.

[6] T. Byczkowski and K. Samotij. Absolute continuity of stable seminorms. *The Annals of Probability*, 14:289–312, 1986.

[7] G. Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46:255–308, 2009.

[8] M.K. Chung *et al*. Topological characterization of signal in brain images using min-max diagrams. *Proceeding MICCAI '09 Proceedings of the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention: Part II*, pages 158–166, 2009.

[9] H. Cramér and M.R. Leadbetter. *Sationary and Related Stochastic Processes*. Wiley, New York, 1967.

[10] R. Curry, J. Ghrist and M. Robinson. Euler calculus and its applications to signals and sensing. *Proceedings of Symposia in Applied Mathematics*, to appear, 2012.

[11] N. Dunford and J. Schwartz. *Linear Operators, Part 1: General Theory*. Wiley, New York, 1988.

[12] O. Kallenberg. *Probabilistic Symmetries and Invariance Principles*. Springer, 2005.

[13] J. Kim and D. Pollard. Cube root asymptotics. *The Annals of Statistics*, 18:191–219, 1990.

[14] A. Kolmogorov and Y. Rozanov. On a strong mixing condition for a stationary gaussian process. *Teoria Veroyatnostei i Primeneniya*, 5:222–227, 1960.

[15] G. Leadbetter, M. Lindgren and H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes*. Springer Verlag, New York, 1983.

[16] G. Lindgren. *Sationary Stochastic Processes: Theory and Applications*. Chapman & Hall/CRC, Boca Raton, 2012.

[17] M. Rosenblatt. *Random Processes*. Oxford University Press, New York, 1962.

[18] G. Samorodnitsky and Y. Shen. Intrinsic location functionals of stationary processes. Submitted.

[19] G. Samorodnitsky and Y. Shen. Is the location of the supremum of a stationary process nearly uniformly distributed? *The Annals of Probability*, accepted.

[20] G. Samorodnitsky and Y. Shen. Distribution of the supremum location of stationary processes. *Electronic Journal of Probability*, 17:1–17, 2012.

[21] J. Taylor and R. Adler. Euler characteristics for gaussian fields on manifolds. *The Annals of Probability*, 31(2):533–563, 2003.

[22] N. Ylvisaker. The expected number of zeros of a stationary gaussian process. *Annals of Mathematical Statistics*, 36:1043–1046, 1965.

INDEX