Compatibility Problems of Probability Measures for Stochastic Processes

by

Jie Shen

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 2018

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Shui Feng Professor, Department of Mathematics and Statistics, McMaster University
Supervisor(s):	Yi Shen Assistant Professor, Department of Statistics and Actuarial Science, University of Waterloo Ruodu Wang Associate Professor, Department of Statistics and Actuarial Science, University of Waterloo
Internal Member:	Marius Hofert Assistant Professor, Department of Statistics and Actuarial Science, University of Waterloo Alexander Schied Professor, Department of Statistics and Actuarial Science, University of Waterloo
Internal-External Member:	Alexandru Nica Professor, Pure Mathematics Department, University of Waterloo

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

A version of Chapter 2 of this thesis has been prepared as a research paper submitted and now under review for publication. A version of Chapter 3 of this thesis is going to appear in *Stochastic Processes and their Applications*. A version of Chapter 4 is in preparation. All papers are co-authored with my supervisors, where my contributions include deriving the theorems, and writing the initial draft.

Abstract

In this thesis, we address three topics in the area of compatibility for probability measures. By “compatibility”, we mean the problems concerning the existence of random variables/stochastic processes which generate certain given probability distributions in some predetermined way.

First, we study a compatibility problem for distributions on the real line and probability measures on a measurable space. For a given set of probability measures and a corresponding set of probability distributions, we propose sufficient and necessary conditions for the existence of a random variable, such that under each measure, the distribution of this random variable coincides with the corresponding distribution on the real line. Various applications in optimization and risk management are discussed.

Secondly, we investigate a compatibility problem involving periodic stationary processes. We consider a family of random locations, called intrinsic location functionals, of periodic stationary processes. We show that the set of all possible distributions of intrinsic location functionals for periodic stationary processes is the convex hull generated by a specific group of distributions. Two special subclasses of these random locations, invariant intrinsic location functionals and first-time intrinsic location functionals, are studied in more detail.

Along this direction, we proceed to propose a unified framework for random locations exhibiting some probabilistic symmetries. A theorem of Noether's type is proved, which gives rise to a conservation law describing the change of the density function of a random location as the interval of interest changes. We also discuss the boundary and near boundary behavior of the distribution of the random locations.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisors, Dr. Yi Shen and Dr. Ruodu Wang. They always support and encourage me to overcome any obstacles I have been facing through my PhD studies. Their profound knowledge, spirit of adventure and continuous support turn me to be a better person. Without their patience, guidance, and insightful advice, I would not have completed the thesis.

I would also like to thank the rest of my thesis examining committee: Dr. Shui Feng, Dr. Marius Hofert, Dr. Alexandru Nica, and Dr. Alexander Schied for taking their patience and time to read my thesis, and for all the insightful suggestions and comments they give.

Many thanks go to my dear friends, who have supported me all the way. I am grateful for their friendship and the wonderful times we spent together.

Last but certainly not least, I would like to thank my parents for their unconditional love. My special thanks go to my boyfriend, who has always been encouraging me to follow my dreams.

To my family

Table of Contents

Chapter 1

Introduction

This thesis addresses three topics in the area of compatibility for probability measures. By "compatibility", we mean the problems concerning the existence of random variables/stochastic processes which generate certain given probability distributions in some predetermined way. Different types of compatibility problems are frequently discussed and explored in existing literature. For example, necessary and sufficient conditions for the existence of a joint density with given families as its associated conditional densities were discussed in Arnold and Press (1989). Given two marginals on a Polish space, Strassen (1965) proposed a sufficient and necessary condition for a probability measure on a product space to exist. An extended result was shown in Gutmann et al. (1991) by considering the underlying probability space as Hausdorff spaces. In Joe (1997), the author explored the existence of a trivariate distribution given bivariate margins and the existence of higher-dimensional multivariate distribution under some marginal constraints.

Chapter 2 is based on the paper Shen et al. (2017), which is now under review for publication. It is dedicated to a compatibility problem for change of measures. Many change of measures problems have been investigated in previous literature, both in theory and application. For stochastic processes, Girsanov theorem describes how the distribution of a stochastic process changes under a given change of measure. It is widely used in the study of diffusions and stochastic differential equations (Revuz and Yor, 2013) and deriving the distributions of asset prices under different probability measures in financial market

(Bingham and Kiesel, 2013). The basic idea of importance sampling is also based on the probability change, which allows researchers to generate sampling points from another distribution to study the properties of a given distribution, when sampling from the given one is difficult. Importance sampling techniques are frequently used in Monte Carlo study of sequential tests (Siegmund, 1976), portfolio credit risk (Glasserman and Li, 2005), *etc.*

In this chapter, for a given set of probability measures on a probability space and a corresponding set of probability distributions on the real line, we develop sufficient and necessary conditions for the existence of a random variable, such that under each measure given on the probability space, the distribution of this random variable coincides with the corresponding distribution on the real line. More precisely, let $(F_1, \dots, F_n)$ be a set of distributions on the real line and $(Q_1, \dots, Q_n)$ be a set of probability measures on a probability space. We say $(F_1, \dots, F_n)$ and $(Q_1, \dots, Q_n)$ are compatible if there exists a random variable such that the distribution of this random variable under Q_i is F_i, and almost compatible if for any $\varepsilon > 0$, there exists a random variable such that the Kullback-Leibler divergence between F_i and $F_{i,\varepsilon}$ is smaller than ε, where $F_{i,\varepsilon}$ is the distribution of this random variable under Q_i. Assuming that all the Q_i's are atomless, we show the following equivalent condition holds: $(F_1, \dots, F_n)$ and $(Q_1, \dots, Q_n)$ are almost compatible if and only if there exist reference measures F and Q, such that

$$\left(\frac{\mathrm{d}F_1}{\mathrm{d}F}, \dots, \frac{\mathrm{d}F_n}{\mathrm{d}F}\right)\bigg|_F \prec_{\mathrm{cx}} \left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \dots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q}\right)\bigg|_Q, \tag{1.1}$$

where $\prec_{\mathrm{cx}}$ means $\leqslant$ in the convex order. If almost compatibility is replaced by compatibility, an extra condition on the existence of a continuous random variable "independent from the others" is proposed and shown to be enough to guarantee the compatibility.

Chapter 3 is based on the paper Shen et al. (2018), which is to appear in *Stochastic Processes and their Applications*. In this chapter we study a compatibility problem involving periodic stationary processes. Given a random location in a certain family called the intrinsic location functionals, a large class of random locations including and extending far beyond locations of the path supremum/infimum, hitting times, *etc*, and a probability distribution defined on an interval, we show how to decide whether there exists a periodic stationary process with certain period, such that the given distribution is the distribution of the random location for that process over the interval. The definition and properties of intrinsic location functionals for stationary processes were introduced and discussed in Samorodnitsky and Shen (2013a). In addition, Samorodnitsky and Shen (2013a) characterized stationarity by the existence of a density satisfying certain conditions for appropriate intrinsic location functionals. In Shen (2016), a characterization of stationary increments for stochastic processes was established using doubly intrinsic location functionals, a subclass of intrinsic location functional which is invariant under vertical shift of the path. In Shen (2018), given a probability distribution, the author considered the necessary condition for the existence of a self-similar process with stationary increments, such that the distribution of the location of the path supremum agrees with the given one.

In this chapter, we first investigate the properties of different types of intrinsic locations for periodic stationary processes. More precisely, denote by L an intrinsic location functional, $\mathbf{X}$ a periodic ergodic process with period 1. Let $F_{L,T}^{\mathbf{X}}$ be the distribution of L for the process $\mathbf{X}$ on the interval $[0, T]$ for $T \leqslant 1$. We show that the density f of $F_{L,T}^{\mathbf{X}}$ exists in the interior of the interval $[0, T]$, and f takes non-negative integer values. Moreover, the total variation of f on any interval is dominated by the sum of its values on the endpoints of the interval. Besides, if f is at least 1 in some small neighbourhoods of 0 and T, then f should be at least 1 on the whole interval $[0, T]$. Conversely, we show that for any F in a certain set of distributions, there exists an intrinsic location functional and a periodic stationary process with period 1, such that F is the distribution of this intrinsic location for such process on $[0, T]$. Furthermore, two special classes of intrinsic location functionals, invariant and first-time intrinsic location functionals, are explored. It is proved that for any F in a certain set, there always exists a periodic stationary process with period 1, such that the distribution of the supremum location for such process coincides with F on $[0, T]$. For the first-time intrinsic location functional, a more specific characterization compared to the general results is discussed.

Chapter 4 is an exploration of the Noether theorem for random locations, giving a description of how the distribution of a random location over an interval can change as the interval moves. This can be regarded as the compatibility problem of the distributions of this random location with different intervals. On one hand, the famous Noether theorem in mathematical physics (Noether, 1918) shows that each differentiable symmetry of a system corresponds to a conservation law. We refer to Kosmann-Schwarzbach (2011) for a thorough review of the Noether theorems. Different generalizations can be found in Yasue (1981); Thieullen and Zambrini (1997a); Baez and Fong (2013). On the other hand, the random locations of stochastic processes exhibiting certain probabilistic symmetries have been studied in a series of works in the past years. Samorodnitsky and Shen (2013a) showed that the distribution of any intrinsic random location for a stationary process must satisfy a specific set of conditions. Similar results were later established for intrinsic location functionals of stochastic processes with stationary increments and of self-similar processes with stationary increments (Shen, 2016, 2018). In a broad sense, all these processes share the common point that they exhibit a certain kind of probabilistic symmetry, which means invariance under a family of transformations.

In this chapter, we provide a framework for the aforementioned random locations and probabilistic symmetries, in which a similar result as the Noether theorem can be established. We first generalize the notion of intrinsic location functional to "intrinsic random location" by dissociating it from the paths of stochastic processes. Then we call an intrinsic random location φ-stationary, if the distribution of this location is compatible under a given flow $\varphi = \varphi^t(x)$. We construct a point process related to the φ-stationary intrinsic random location, and define the the control measure as the expectation of the point process. After establishing a connection between the distribution of any φ-stationary intrinsic random location and the control measure of a point process related to it, we show that there exists two measures, such that the density of the φ-stationary intrinsic random location is equal to the difference of these two measures, after some scaling. More precisely, denote by $L([a, b])$ a φ-stationary intrinsic random location on the interval $[a, b]$, then

$$\dot{\varphi}^0(x_2)f(x_2) - \dot{\varphi}^0(x_1)f(x_1) = \nu_\varphi^{(a,b)}((x_1, x_2]) - \mu_\varphi^{(a,b)}((x_1, x_2]),$$

for any $x_1 \leqslant x_2$, $x_1, x_2 \in (a, b)$ and some measures $\mu_\varphi^{(a,b)}$ and $\nu_\varphi^{(a,b)}$ defined in terms of the control measure of the related point process, where $\dot{\varphi}^0(x)$ is the partial derivative of φ with respect to t at time 0. With this equation, we can further derive a conservation law associated with the flow φ. This indeed gives us a unified characterization for the probabilistic symmetries of stochastic processes via certain groups of random locations. Moreover, we develop results to get the probability that a random location over an interval falls on the boundaries of the interval, as well as criteria for the density of this random location to explode near the boundaries.

In Chapter 5, we discuss three future directions: 1. The characterization of exchangeability for stochastic processes using a certain family of random locations; 2. The random locations of Max-stable processes, especially moving maximum processes, extremal Gaussian processes and Brown-Resnick processes; 3. Large deviation principles of the maximum for stochastic processes.

Chapter 2

Distributional Compatibility for Change of Measures

2.1 Introduction

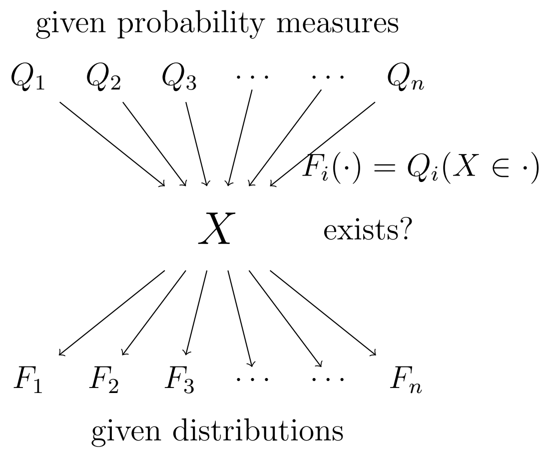
Change of probability measures is found ubiquitous in problems where multiple probability measures appear, with extensive theoretical treatment and applications in the fields of probability theory, statistics, economic decision theory, simulation, and finance.

A key feature of a change of measure is that the distribution of a random variable is transformed to another one, and this serves many theoretical as well as practical purposes, such as in the modification of a Brownian motion drift (Revuz and Yor, 2013) or in importance sampling (Siegmund, 1976; Glasserman and Li, 2005). In view of this, a question seems natural to us: *how much* would the distribution change? We formulate this question below.

(A) Given two probability measures P and Q defined on the same measurable space $(\Omega, \mathcal{A})$, suppose that a random variable $X : \Omega \to \mathbb{R}$ has a given distribution function F under P. What are the possible distributions of X under Q?

Question (A) arises naturally if one has statistical (distributional) information about a random variable X under P, but yet she is concerned about the behaviour of X under another measure Q. This includes many classic optimization problems in the literature; see Section 2.5 for more details. A general version of question (A), the vocal focus of this chapter, is the following.

(B) Given several probability measures $Q_1, \dots, Q_n$ defined on $(\Omega, \mathcal{A})$, and distribution measures $F_1, \dots, F_n$ on $\mathbb{R}$, does there exist a random variable $X : \Omega \to \mathbb{R}$ such that X has distribution F_i under Q_i for $i = 1, \dots, n$?



Question (B) is henceforth referred to as the *compatibility* problem for the n-tuples of measures $(Q_1, \dots, Q_n)$ and $(F_1, \dots, F_n)$. We give an analytical answer to question (B), and hence (A). In the main part of this chapter, we focus on n-tuples of probability measures for a positive integer n. Some results also hold for infinite (possibly uncountable) collections of probability measures; see Remark 2.2.4.

Before describing our findings, let us look at a few intuitive cases of (B). Suppose that $(Q_1, \dots, Q_n)$ and $(F_1, \dots, F_n)$ are compatible, that is, (B) has an affirmative answer. In case that $Q_1, \dots, Q_n$ are identical, it is clear that the respective distributions of a random variable under each Q_i, $i = 1, \dots, n$ are the same; thus $F_1 = \cdots = F_n$. In case that $Q_1, \dots, Q_n$ are mutually singular, the respective distributions of a random variable under Q_i, $i = 1, \dots, n$ can be arbitrary. In case that $F_1, \dots, F_n$ are mutually singular measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $Q_1, \dots, Q_n$ have to be also mutually singular. From the above observations, it then seems natural to us that whether $(Q_1, \dots, Q_n)$ and $(F_1, \dots, F_n)$ are compatible depends on the *heterogeneity* (in some sense) among $Q_1, \dots, Q_n$ compared to that of $F_1, \dots, F_n$. More precisely, $Q_1, \dots, Q_n$ need to be *more heterogeneous* than $F_1, \dots, F_n$ to allow for compatibility.

To describe the above heterogeneity mathematically, we seek help from a notion of heterogeneity order. It turns out that compatibility of $(Q_1, \dots, Q_n)$ and $(F_1, \dots, F_n)$ is closely related to multivariate convex order between the Radon-Nikodym derivatives $(\frac{\mathrm{d}F_1}{\mathrm{d}F}, \dots, \frac{\mathrm{d}F_n}{\mathrm{d}F})$ and $(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \dots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q})$, where F and Q are two "reference probability measures" on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and $(\Omega, \mathcal{A})$, respectively. In particular, we show that question (B) has an affirmative answer only if for some measures F dominating $(F_1, \dots, F_n)$ and Q dominating $(Q_1, \dots, Q_n)$,

$$\int_{\mathbb{R}} f\left(\frac{\mathrm{d}F_1}{\mathrm{d}F}, \dots, \frac{\mathrm{d}F_n}{\mathrm{d}F}\right) \mathrm{d}F \leqslant \int_{\Omega} f\left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \dots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q}\right) \mathrm{d}Q \tag{2.1}$$

for all convex functions $f : \mathbb{R}^n \to \mathbb{R}$. Furthermore, if the measurable space $(\Omega, \mathcal{A})$ is rich enough, the above necessary condition is sufficient for a positive answer to (B). We then proceed to generalize our results to random vectors and stochastic processes, and conclude the chapter with various optimization problems related to compatibility of distributions under change of measures.

Most of the results in this chapter do not rely on the structure of the measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and therefore they are valid for compatibility (the existence of a suitable mapping) of tuples of probability measures on two general measurable spaces. Some of our results turn out to be deeply related to comparison of statistical experiments (Torgersen, 1991), which shall be commented in Remark 2.3.19. For the purpose of intuitive illustration and potential probabilistic applications, we write our main results for the cases of random variables and stochastic processes.

Throughout, we work with a fixed measurable space $(\Omega, \mathcal{A})$, which allows for atomless probability measures. A probability measure Q on $(\Omega, \mathcal{A})$ is said to be atomless if for all $A \in \mathcal{A}$ with $Q(A) > 0$, there exists $B \in \mathcal{A}$, $B \subset A$ such that $0 < Q(B) < Q(A)$. Equivalently, there exists a random variable in $(\Omega, \mathcal{A})$ that is continuously distributed under Q. Let $\mathcal{F}$ be the set of probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, where $\mathcal{B}(\mathbb{R})$ stands for the Borel σ-algebra of $\mathbb{R}$, and $\mathcal{P}$ be the set of probability measures on $(\Omega, \mathcal{A})$. Let $L(\Omega, \mathcal{A})$

be the set of random variables defined on $(\Omega, \mathcal{A})$. For any measures $Q, Q_1, \ldots, Q_n$, we say that Q dominates $(Q_1, \ldots, Q_n)$, denoted by $(Q_1, \ldots, Q_n) \ll Q$, if Q dominates Q_i for each $i = 1, \ldots, n$.

Remark 2.1.1. A related question to ours is that, given a number of distributions on $\mathbb{R}$, can we construct, on some underlying space we can still choose, one random variable X and the same number of probability measures such that X has under each probability the corresponding distribution? The answer to this question is always affirmative due to the excessive flexibility in choosing the probability measures. In fact, let Ω be a topological space homomorphic to the real line, and let $\mathcal{A}$ be the Borel σ-field on Ω. For any bicontinuous mapping X from $(\Omega, \mathcal{A})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, if we define Q_i to be the image of F_i under X^{-1}, then X has distribution F_i under Q_i, $i = 1, \ldots, n$. In fact, even if $(\Omega, \mathcal{A})$ and X are both fixed, as long as the range of X covers the supports of $F_1, \ldots, F_n$, we can always choose $Q_1, \ldots, Q_n$ so that X has the corresponding distributions.

2.2 Compatibility and an equivalent condition

We first define the compatibility problem for the n-tuples of measures $(Q_1, \ldots, Q_n) \in \mathcal{P}^n$ and $(F_1, \ldots, F_n) \in \mathcal{F}^n$, the main concept of this chapter.

Definition 2.2.1. $(Q_1, \ldots, Q_n) \in \mathcal{P}^n$ and $(F_1, \ldots, F_n) \in \mathcal{F}^n$ are *compatible* if there exists a random variable X in $(\Omega, \mathcal{A})$ such that F_i is the distribution of X under Q_i for each $i = 1, \ldots, n$.

We note that F is the distribution of X under Q if and only if $F = Q \circ X^{-1}$. Below we establish our first result, which leads to an equivalent condition for compatibility of $(Q_1, \ldots, Q_n) \in \mathcal{P}^n$ and $(F_1, \ldots, F_n) \in \mathcal{F}^n$.

Theorem 2.2.2. *For $(Q_1, \ldots, Q_n) \in \mathcal{P}^n$, $(F_1, \ldots, F_n) \in \mathcal{F}^n$ and $X \in L(\Omega, \mathcal{A})$, equivalent are:*

(i) X has distribution F_i under Q_i for $i = 1, \ldots, n$.

(ii) For all $Q \in \mathcal{P}$ dominating $(Q_1, \ldots, Q_n)$, the probability measure $F = Q \circ X^{-1}$ dominates $(F_1, \ldots, F_n)$, and

$$\left(\frac{\mathrm{d}F_1}{\mathrm{d}F}, \ldots, \frac{\mathrm{d}F_n}{\mathrm{d}F}\right)(X) = \mathbb{E}^Q\left[\left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \ldots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q}\right)\Big|\sigma(X)\right]. \tag{2.2}$$

(iii) For some $Q \in \mathcal{P}$ dominating $(Q_1, \ldots, Q_n)$, the probability measure $F = Q \circ X^{-1}$ dominates $(F_1, \ldots, F_n)$, and (2.2) holds.

Proof. 1. (i)⇒(ii): By definition, X is such that $Q_i(X \in A) = F_i(A)$ for $A \in \mathcal{B}(\mathbb{R})$ and $i = 1, \ldots, n$. Let $Q \in \mathcal{P}$ such that $Q_i \ll Q$, $i = 1, \ldots, n$. Note that such Q always exists since we can take, for example, $Q = \frac{1}{n}(Q_1 + \cdots + Q_n)$. For any $A \in \mathcal{B}(\mathbb{R})$, if $F(A) = 0$, then $Q(X \in A) = 0$. Since $Q_i \ll Q$, $Q_i(X \in A) = F_i(A) = 0$, we have $F_i \ll F$ for $i = 1, \ldots, n$. We can verify that for any $A \in \mathcal{B}(\mathbb{R})$ and $i = 1, \ldots, n$,

$$\mathbb{E}^Q\left[\mathrm{I}_{\{X\in A\}}\frac{\mathrm{d}Q_i}{\mathrm{d}Q}\right] = Q_i(X \in A) = F_i(A) = \int_A \frac{\mathrm{d}F_i}{\mathrm{d}F}\mathrm{d}F = \mathbb{E}^Q\left[\mathrm{I}_{\{X\in A\}}\frac{\mathrm{d}F_i}{\mathrm{d}F}(X)\right].$$

Therefore,

$$\frac{\mathrm{d}F_i}{\mathrm{d}F}(X) = \mathbb{E}^Q\left[\frac{\mathrm{d}Q_i}{\mathrm{d}Q}\Big|\sigma(X)\right], \quad i = 1, \ldots, n.$$

2. (ii)⇒(iii): Trivial.

3. (iii)⇒(i): Suppose that (2.2) holds and F dominates $(F_1, \ldots, F_n)$. One can easily verify that, for all $A \in \mathcal{B}(\mathbb{R})$ and $i = 1, \ldots, n$,

$$\begin{aligned}\mathbb{E}^{Q_i}[\mathrm{I}_{\{X\in A\}}] = \mathbb{E}^Q\left[\mathrm{I}_{\{X\in A\}}\frac{\mathrm{d}Q_i}{\mathrm{d}Q}\right] &= \mathbb{E}^Q\left[\mathbb{E}^Q\left[\mathrm{I}_{\{X\in A\}}\frac{\mathrm{d}Q_i}{\mathrm{d}Q}\Big|X\right]\right] \\ &= \mathbb{E}^Q\left[\mathrm{I}_{\{X\in A\}}\mathbb{E}^Q\left[\frac{\mathrm{d}Q_i}{\mathrm{d}Q}\Big|X\right]\right] \\ &= \mathbb{E}^Q\left[\mathrm{I}_{\{X\in A\}}\frac{\mathrm{d}F_i}{\mathrm{d}F}(X)\right] = F_i(A).\end{aligned}$$

Therefore, X has distribution F_i under Q_i, $i = 1, \ldots, n$, thus $(Q_1, \ldots, Q_n)$ and $(F_1, \ldots, F_n)$ are compatible. □

From Theorem 2.2.2, the necessary and sufficient condition of compatibility is the existence of $X \in L(\Omega, \mathcal{A})$ satisfying (2.2) for some $Q \in \mathcal{P}$ dominating $Q_1, \dots, Q_n$. This condition is not easy to verify in general. In the next sections we explore necessary and sufficient conditions, much easier to verify, based on distributional properties of the random vectors $(\frac{\mathrm{d}F_1}{\mathrm{d}F}, \dots, \frac{\mathrm{d}F_n}{\mathrm{d}F})$ and $(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \dots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q})$, where F and Q are some measures dominating $(F_1, \dots, F_n)$ and $(Q_1, \dots, Q_n)$ respectively.

We conclude this section with the special case of $n = 2$ and $Q_1 \ll Q_2$. In this case, one can take $Q = Q_2$ in Theorem 2.2.2, and the two-dimensional equality in (2.2) reduces to a one-dimensional equality.

Corollary 2.2.3. *For $(Q_1, Q_2) \in \mathcal{P}^2$, $Q_1 \ll Q_2$ and $(F_1, F_2) \in \mathcal{F}^2$, (Q_1, Q_2) and (F_1, F_2) are compatible if and only if there exists $X \in L(\Omega, \mathcal{A})$ with distribution F_2 under Q_2, such that $F_1 \ll F_2$ and*

$$\frac{\mathrm{d}F_1}{\mathrm{d}F_2}(X) = \mathbb{E}^{Q_2}\left[\frac{\mathrm{d}Q_1}{\mathrm{d}Q_2}\middle|\sigma(X)\right].$$

Remark 2.2.4. The result in Theorem 2.2.2 can be generalized to infinite collections of probability measures. Let $\mathcal{J}$ be a (possibly uncountable) set of indices. We say that $(Q_i)_{i\in\mathcal{J}} \subset \mathcal{P}$ and $(F_i)_{i\in\mathcal{J}} \subset \mathcal{F}$ are compatible if there exists a random variable X in $(\Omega, \mathcal{A})$ such that F_i is the distribution of X under Q_i for each $i \in \mathcal{J}$. Based on a proof analogous to that of Theorem 2.2.2, we have the following result. For $(Q_i)_{i\in\mathcal{J}} \subset \mathcal{P}$ and $(F_i)_{i\in\mathcal{J}} \subset \mathcal{F}$ and $X \in L(\Omega, \mathcal{A})$, assuming that there exists a probability measure in $\mathcal{P}$ dominating $(Q_i)_{i\in\mathcal{J}}$, equivalent are:

(i) X has distribution F_i under Q_i for $i \in \mathcal{J}$.

(ii) For all $Q \in \mathcal{P}$ dominating $(Q_i)_{i\in\mathcal{J}}$, the probability measure $F = Q \circ X^{-1}$ dominates $(F_i)_{i\in\mathcal{J}}$, and

$$\frac{\mathrm{d}F_i}{\mathrm{d}F}(X) = \mathbb{E}^{Q}\left[\frac{\mathrm{d}Q_i}{\mathrm{d}Q}\middle|\sigma(X)\right] \quad \text{for all } i \in \mathcal{J}. \tag{2.3}$$

(iii) For some $Q \in \mathcal{P}$ dominating $(Q_i)_{i\in\mathcal{J}}$, the probability measure $F = Q \circ X^{-1}$ dominates $(F_i)_{i\in\mathcal{J}}$ and (2.3) holds.

2.3 Characterizing compatibility via heterogeneity order

In this section, we explore analytical conditions for compatibility of $(Q_1, \dots, Q_n)$ and $(F_1, \dots, F_n)$ based on their Radon-Nikodym derivatives with respect to some reference probability measures, which are much easier to verify than Theorem 2.2.2.

Our results in this section do not rely on the specific structure of $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, and many of them stay valid if $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is replaced by another measurable space. In particular, such results hold for compatibility defined on random vectors or stochastic processes, which will be studied in Section 2.4.

2.3.1 Preliminaries on convex order

For an arbitrary probability space $(\Gamma, \mathcal{S}, P)$, denote by $L_1^n(\Gamma, \mathcal{S}, P)$ the set of all integrable n-dimensional random vectors in $(\Gamma, \mathcal{S}, P)$. Multivariate convex order is a natural notion of heterogeneity order, as defined below.

Definition 2.3.1 (Convex order)**.** Let $(\Omega_1, \mathcal{A}_1, P_1)$ and $(\Omega_2, \mathcal{A}_2, P_2)$ be two probability spaces. For $\mathbf{X} \in L_1^n(\Omega_1, \mathcal{A}_1, P_1)$ and $\mathbf{Y} \in L_1^n(\Omega_2, \mathcal{A}_2, P_2)$, we write $\mathbf{X}|_{P_1} \prec_{\mathrm{cx}} \mathbf{Y}|_{P_2}$, if $\mathbb{E}^{P_1}[f(\mathbf{X})] \leqslant \mathbb{E}^{P_2}[f(\mathbf{Y})]$ for all convex functions $f : \mathbb{R}^n \to \mathbb{R}$.

For more on multi-dimensional convex order, we refer to Müller and Stoyan (2002, Chapter 3) and Shaked and Shanthikumar (2007, Chapter 7).

For $\mathbf{X} \in L_1^n(\Omega_1, \mathcal{A}_1, P_1)$ and $\mathbf{Y} \in L_1^n(\Omega_2, \mathcal{A}_2, P_2)$, we use $\mathbf{X}|_{P_1} \overset{\mathrm{d}}{=} \mathbf{Y}|_{P_2}$ to represent that $\mathbf{X}$ and $\mathbf{Y}$ have the same distribution under P_1 and P_2 respectively. Clearly, if $\mathbf{X}|_{P_1} \overset{\mathrm{d}}{=} \mathbf{Y}|_{P_2}$, then $\mathbf{X}|_{P_1} \prec_{\mathrm{cx}} \mathbf{Y}|_{P_2}$ and $\mathbf{Y}|_{P_2} \prec_{\mathrm{cx}} \mathbf{X}|_{P_1}$. A key feature of convex order is its connection to conditional expectations. Below in Lemma 2.3.2 we quote Theorem 7.A.1 of Shaked and Shanthikumar (2007) for this well-known result (an extension of Strassen's theorem, Strassen (1965)); one also finds a slightly simpler formulation as Theorem 3.4.2 of Müller and Stoyan (2002). See also Hirsch et al. (2011) for a construction similar to Lemma 2.3.2 for stochastic processes (termed *peacocks*).

Lemma 2.3.2. *For $\mathbf{X} \in L_1^n(\Omega_1, \mathcal{A}_1, P_1)$ and $\mathbf{Y} \in L_1^n(\Omega_2, \mathcal{A}_2, P_2)$, $\mathbf{X}|_{P_1} \prec_{\text{cx}} \mathbf{Y}|_{P_2}$ if and only if there exist a probability space $(\Omega_3, \mathcal{A}_3, P_3)$ and $\mathbf{X}', \mathbf{Y}' \in L_1^n(\Omega_3, \mathcal{A}_3, P_3)$ such that $\mathbf{X}'|_{P_3} \stackrel{\text{d}}{=} \mathbf{X}|_{P_1}$, $\mathbf{Y}'|_{P_3} \stackrel{\text{d}}{=} \mathbf{Y}|_{P_2}$, and $\mathbb{E}^{P_3}[\mathbf{Y}'|\mathbf{X}'] = \mathbf{X}'$.*

2.3.2 Heterogeneity order

As mentioned in the introduction, compatibility intuitively concerns the heterogeneity among $(Q_1, \ldots, Q_n)$ compared to $(F_1, \ldots, F_n)$. The following lemma, based on Theorem 2.2.2, yields a possible way of characterizing the comparison between the two tuples of measures. More precisely, a necessary condition for compatibility is built on a convex order relation between the random vectors $(\frac{\mathrm{d}F_1}{\mathrm{d}F}, \ldots, \frac{\mathrm{d}F_n}{\mathrm{d}F})$ and $(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \ldots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q})$ for some reference probability measures $F \in \mathcal{F}$ and $Q \in \mathcal{P}$.

Lemma 2.3.3. *If $(Q_1, \ldots, Q_n) \in \mathcal{P}^n$ and $(F_1, \ldots, F_n) \in \mathcal{F}^n$ are compatible, then for any $Q \in \mathcal{P}$ dominating $(Q_1, \ldots, Q_n)$, there exists $F \in \mathcal{F}$ dominating $(F_1, \ldots, F_n)$, such that*

$$\left.\left(\frac{\mathrm{d}F_1}{\mathrm{d}F}, \ldots, \frac{\mathrm{d}F_n}{\mathrm{d}F}\right)\right|_F \prec_{\text{cx}} \left.\left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \ldots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q}\right)\right|_Q. \tag{2.4}$$

Moreover, F in (2.4) can be taken as $Q \circ X^{-1}$, where X is a random variable with distribution F_i under Q_i, $i = 1, \ldots, n$.

Proof. This lemma is directly obtained from Theorem 2.2.2 and Lemma 2.3.2. More precisely, by Theorem 2.2.2, there exists $X \in L(\Omega, \mathcal{A})$ such that

$$\left(\frac{\mathrm{d}F_1}{\mathrm{d}F}, \ldots, \frac{\mathrm{d}F_n}{\mathrm{d}F}\right)(X) = \mathbb{E}^Q\left[\left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \ldots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q}\right) \Big| \sigma(X)\right]$$

where $F = Q \circ X^{-1}$. Therefore,

$$\left.\left(\frac{\mathrm{d}F_1}{\mathrm{d}F}, \ldots, \frac{\mathrm{d}F_n}{\mathrm{d}F}\right)\right|_F \stackrel{\text{d}}{=} \left.\mathbb{E}\left[\left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \ldots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q}\right) \Big| \sigma(X)\right]\right|_Q \prec_{\text{cx}} \left.\left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \ldots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q}\right)\right|_Q,$$

where the last inequality is by Lemma 2.3.2. $\square$

We summarize the necessary condition in Lemma 2.3.3 for compatibility by introducing the following heterogeneity order, which is shown to be a partial order in Lemma 2.3.5 below. In the following, $\mathcal{M}_1$ and $\mathcal{M}_2$ represent the sets of probability measures on two arbitrary measurable spaces, respectively.

Definition 2.3.4. We say that $(P_1, \ldots, P_n) \in \mathcal{M}_1^n$ *is dominated by* $(Q_1, \ldots, Q_n) \in \mathcal{M}_2^n$ *in heterogeneity*, denoted by $(P_1, \ldots, P_n) \prec_{\mathrm{h}} (Q_1, \ldots, Q_n)$, if

$$\left(\frac{\mathrm{d}P_1}{\mathrm{d}P}, \ldots, \frac{\mathrm{d}P_n}{\mathrm{d}P}\right)\Bigg|_P \prec_{\mathrm{cx}} \left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \ldots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q}\right)\Bigg|_Q. \tag{2.5}$$

for some $P \in \mathcal{M}_1$ dominating $(P_1, \ldots, P_n)$ and $Q \in \mathcal{M}_2$ dominating $(Q_1, \ldots, Q_n)$.

Using the language of heterogeneity order, Lemma 2.3.3 says that in order for compatibility of $(Q_1, \ldots, Q_n) \in \mathcal{P}^n$ and $(F_1, \ldots, F_n) \in \mathcal{F}^n$, a necessary condition is $(F_1, \ldots, F_n) \prec_{\mathrm{h}} (Q_1, \ldots, Q_n)$. Before discussing the sufficiency of this condition, we first establish some properties of heterogeneity order.

The following lemma implies that the choice of the reference measures P and Q in (2.5) is irrelevant; in fact, they can be conveniently chosen as the averages of the corresponding measures. It also justifies that Definition 2.3.4 defines $\prec_{\mathrm{h}}$ as a partial order.

Lemma 2.3.5. *For $(P_1, \ldots, P_n) \in \mathcal{M}_1^n$ and $(Q_1, \ldots, Q_n) \in \mathcal{M}_2^n$, let $\mathcal{M}_1^* = \{P \in \mathcal{M}_1 : (P_1, \ldots, P_n) \ll P\}$ and $\mathcal{M}_2^* = \{Q \in \mathcal{M}_2 : (Q_1, \ldots, Q_n) \ll Q\}$. The following are equivalent:*

(i) $(P_1, \ldots, P_n) \prec_{\mathrm{h}} (Q_1, \ldots, Q_n)$; that is, for some $P \in \mathcal{M}_1^*$ and $Q \in \mathcal{M}_2^*$, (2.5) holds.

(ii) For $P = \frac{1}{n}\sum_{i=1}^n P_i$ and $Q = \frac{1}{n}\sum_{i=1}^n Q_i$, (2.5) holds.

(iii) For any $Q \in \mathcal{M}_2^*$, there exists $P \in \mathcal{M}_1^*$ such that (2.5) holds.

Proof. We proceed in the order (iii)⇒(ii)⇒(i)⇒(iii).

1. (iii)⇒(ii): For $Q = \frac{1}{n}\sum_{i=1}^n Q_i$, there exists $P^* \in \mathcal{M}_1^*$ such that

$$\left(\frac{\mathrm{d}P_1}{\mathrm{d}P^*}, \ldots, \frac{\mathrm{d}P_n}{\mathrm{d}P^*}\right)\Bigg|_{P^*} \prec_{\mathrm{cx}} \left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \ldots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q}\right)\Bigg|_Q.$$

Take the convex function $f : \mathbb{R}^n \to \mathbb{R}$, $f(x_1, \ldots, x_n) = (x_1 + \cdots + x_n)^2$. It follows from the definition of convex order that

$$\mathbb{E}^{P^*}\left[\left(\frac{\mathrm{d}P_1}{\mathrm{d}P^*} + \cdots + \frac{\mathrm{d}P_n}{\mathrm{d}P^*}\right)^2\right] \leqslant \mathbb{E}^{Q}\left[\left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q} + \cdots + \frac{\mathrm{d}Q_n}{\mathrm{d}Q}\right)^2\right] = \mathbb{E}^Q[n^2] = n^2.$$

On the other hand,

$$\mathbb{E}^{P^*}\left[\frac{\mathrm{d}P_1}{\mathrm{d}P^*} + \cdots + \frac{\mathrm{d}P_n}{\mathrm{d}P^*}\right] = \mathbb{E}^{P_1}[1] + \cdots + \mathbb{E}^{P_n}[1] = n.$$

Hence, $\frac{\mathrm{d}P_1}{\mathrm{d}P^*} + \cdots + \frac{\mathrm{d}P_n}{\mathrm{d}P^*}$ has zero variance under P^*, implying that it is P^*-almost surely equal to n. In other words, $P^* = \frac{1}{n}\sum_{i=1}^n P_i$ on all sets with positive P^*-measure. Noting that P^* dominates $(P_1, \ldots, P_n)$, we have $P^* = \frac{1}{n}\sum_{i=1}^n P_i$. Therefore, (2.5) holds for $P = \frac{1}{n}\sum_{i=1}^n P_i$ and $Q = \frac{1}{n}\sum_{i=1}^n Q_i$.

2. (ii)$\Rightarrow$(i): trivial.

3. (i)$\Rightarrow$(iii): Assume (2.5) holds for some $Q \in \mathcal{M}_2^*$ and $P \in \mathcal{M}_1^*$. Let $\mathbf{Y} = (\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \ldots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q})$, $\mathbf{Z} = (\frac{\mathrm{d}P_1}{\mathrm{d}P}, \ldots, \frac{\mathrm{d}P_n}{\mathrm{d}P})$. Let Q' be another probability measure in $\mathcal{M}_2^*$. First, note that without loss of generality, we can assume that Q' is dominated by Q. Indeed, any general Q' can be decomposed as $Q' = cQ'_a + (1-c)Q'_s$, where $c \in [0,1]$, Q'_a and Q'_s are probability measures being absolutely continuous and singular with respect to Q, respectively. This implies that the distribution of $(\frac{\mathrm{d}Q_1}{\mathrm{d}Q'}, \ldots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q'})$ is a mixture of the distribution of $c^{-1}(\frac{\mathrm{d}Q_1}{\mathrm{d}Q'_a}, \ldots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q'_a})$ (with probability c) and $(0, \ldots, 0)$ (with probability $1-c$). It is easy to check that such a distribution has a larger convex order than $(\frac{\mathrm{d}Q_1}{\mathrm{d}Q'_a}, \ldots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q'_a})$. Thus, if we show (2.5) for Q'_a, the result also holds for Q'. In the sequel we assume Q' is dominated by Q, hence the random variable $X = \frac{\mathrm{d}Q'}{\mathrm{d}Q}$ is well-defined. Let a set $A = \{\mathbf{Y} \neq 0\}$. Note that since Q' dominates $(Q_1, \ldots, Q_n)$, $X > 0$ Q-almost surely on A. $(\frac{\mathrm{d}Q_1}{\mathrm{d}Q'}, \ldots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q'})$ can be then taken as $X^{-1}\mathbf{Y}$, where we define $X^{-1}\mathbf{Y} = 0$ when both X and $\mathbf{Y}$ are 0.

   By Lemma 2.3.2, there exists a probability space $(\Omega', \mathcal{A}', \eta)$ and random vectors $\mathbf{Y}', \mathbf{Z}'$, such that $\mathbf{Y}'|_\eta \stackrel{\mathrm{d}}{=} \mathbf{Y}|_Q$, $\mathbf{Z}'|_\eta \stackrel{\mathrm{d}}{=} \mathbf{Z}|_P$, and $\mathbb{E}^\eta[\mathbf{Y}'|\mathbf{Z}'] = \mathbf{Z}'$. Furthermore, we can obviously choose $(\Omega', \mathcal{A}', \eta)$ to contain a random variable X' such that $(X', \mathbf{Y}')|_\eta \stackrel{\mathrm{d}}{=}$

$(X,\mathbf{Y})|_Q$. On $(\Omega',\mathcal{A}')$, define a new probability measure η' by $\frac{\mathrm{d}\eta'}{\mathrm{d}\eta} = X'$, then $(X',\mathbf{Y}')|_{\eta'} \stackrel{\mathrm{d}}{=} (X,\mathbf{Y})|_{Q'}$. For any bounded measurable function f,

$$\mathbb{E}^{\eta}[f(\mathbf{Z}')\mathbf{Z}'] = \mathbb{E}^{\eta}[f(\mathbf{Z}')\mathbf{Y}'] = \mathbb{E}^{\eta}\left[f(\mathbf{Z}')\left(\frac{\mathbf{Y}'}{X'}\right)X'\right] = \mathbb{E}^{\eta'}\left[f(\mathbf{Z}')\left(\frac{\mathbf{Y}'}{X'}\right)\right],$$

where, again, $X'=0$ implies $\mathbf{Y}'=0$, and in this case $\frac{\mathbf{Y}'}{X'}$ is set to be 0. Hence

$$\begin{aligned}\mathbb{E}^{\eta}[f(\mathbf{Z}')\mathbf{Z}'] &= \mathbb{E}^{\eta'}\left[f(\mathbf{Z}')\mathbb{E}^{\eta'}\left[\frac{\mathbf{Y}'}{X'}\middle|\mathbf{Z}'\right]\right]\\ &= \mathbb{E}^{\eta}\left[f(\mathbf{Z}')\mathbb{E}^{\eta'}\left[\frac{\mathbf{Y}'}{X'}\middle|\mathbf{Z}'\right]X'\right]\\ &= \mathbb{E}^{\eta}\left[f(\mathbf{Z}')\mathbb{E}^{\eta'}\left[\frac{\mathbf{Y}'}{X'}\middle|\mathbf{Z}'\right]\mathbb{E}^{\eta}[X'|\mathbf{Z}']\right].\end{aligned}$$

Therefore we must have

$$\mathbb{E}^{\eta'}\left[\frac{\mathbf{Y}'}{X'}\middle|\mathbf{Z}'\right] = \frac{\mathbf{Z}'}{\mathbb{E}^{\eta}[X'|\mathbf{Z}']}$$

η-almost surely. Define measure P' by $\frac{\mathrm{d}P'}{\mathrm{d}P}(z) = \mathbb{E}^{\eta}[X'|\mathbf{Z}'=\mathbf{Z}(z)] =: V(z)$. Note that since

$$\int\frac{\mathrm{d}P'}{\mathrm{d}P}(z)\mathrm{d}P(z) = \int\mathbb{E}^{\eta}[X'|\mathbf{Z}'=\mathbf{Z}(z)]\mathrm{d}P(z) = \mathbb{E}^{\eta}\left[\mathbb{E}^{\eta}[X'|\mathbf{Z}']\right] = \mathbb{E}^{\eta}[X'] = \mathbb{E}^{Q}[X] = 1,$$

P' is a probability measure. Then we have $(\frac{\mathrm{d}P_1}{\mathrm{d}P'},\ldots,\frac{\mathrm{d}P_n}{\mathrm{d}P'}) = \frac{\mathbf{Z}}{V}$. Define probability measure η'' by $\frac{\mathrm{d}\eta''}{\mathrm{d}\eta} = \mathbb{E}^{\eta}[X'|\mathbf{Z}']$. Since the relation between $\mathbf{Z}'$, η and η'' is in parallel with that between $\mathbf{Z}$, P and P', we have

$$\left.\frac{\mathbf{Z}}{V}\right|_{P'} \stackrel{\mathrm{d}}{=} \left.\frac{\mathbf{Z}'}{\mathbb{E}^{\eta}[X'|\mathbf{Z}']}\right|_{\eta''}.$$

However, for any test function g,

$$\mathbb{E}^{\eta''}[g(\mathbf{Z}')] = \int g(\mathbf{Z}')\frac{\mathrm{d}\eta''}{\mathrm{d}\eta}\mathrm{d}\eta = \int g(\mathbf{Z}')\mathbb{E}^{\eta}[X'|\mathbf{Z}']\mathrm{d}\eta = \mathbb{E}^{\eta}[g(\mathbf{Z}')X'] = \mathbb{E}^{\eta'}(g(\mathbf{Z}')),$$

hence $\mathbf{Z}'|_{\eta'} \stackrel{\mathrm{d}}{=} \mathbf{Z}'|_{\eta''}$. Thus, $\frac{\mathbf{Z}'}{\mathbb{E}^{\eta}[X'|\mathbf{Z}']}$, as a function of $\mathbf{Z}'$, also has the same distribution under η' and η''. Consequently, we have

$$\left.\left(\frac{\mathrm{d}P_1}{\mathrm{d}P'},\ldots,\frac{\mathrm{d}P_n}{\mathrm{d}P'}\right)\right|_{P'} = \left.\frac{\mathbf{Z}}{V}\right|_{P'} \stackrel{\mathrm{d}}{=} \left.\frac{\mathbf{Z}'}{\mathbb{E}^{\eta}[X'|\mathbf{Z}']}\right|_{\eta'}.$$

Also, recalling that $(X', \mathbf{Y}')|_{\eta'} \stackrel{\mathrm{d}}{=} (X, \mathbf{Y})|_{Q'}$,

$$\left.\left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q'}, \ldots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q'}\right)\right|_{Q'} = \left.\frac{\mathbf{Y}}{X}\right|_{Q'} \stackrel{\mathrm{d}}{=} \left.\frac{\mathbf{Y}'}{X'}\right|_{\eta'}.$$

The proof is finished by noting that

$$\begin{aligned}
&\mathbb{E}^{\eta'}\left[\left.\frac{\mathbf{Y}'}{X'}\right| \frac{\mathbf{Z}'}{\mathbb{E}^{\eta}[X'|\mathbf{Z}']}\right] \\
&= \mathbb{E}^{\eta'}\left[\left.\mathbb{E}^{\eta'}\left[\left.\frac{\mathbf{Y}'}{X'}\right|\mathbf{Z}'\right]\right| \frac{\mathbf{Z}'}{\mathbb{E}^{\eta}[X'|\mathbf{Z}']}\right] \\
&= \frac{\mathbf{Z}'}{\mathbb{E}^{\eta}[X'|\mathbf{Z}']},
\end{aligned}$$

and applying Lemma 2.3.2. □

Some simple and intuitive properties of heterogeneity order are summarized in the following proposition. These properties justify the term "heterogeneity" in the order $\prec_{\mathrm{h}}$.

Proposition 2.3.6. *For $(P_1, \ldots, P_n) \in \mathcal{M}_1^n$ and $(Q_1, \ldots, Q_n) \in \mathcal{M}_2^n$, the following holds.*

(i) If $P_1, \ldots, P_n$ are identical, then $(P_1, \ldots, P_n) \prec_{\mathrm{h}} (Q_1, \ldots, Q_n)$.

(ii) If $Q_1, \ldots, Q_n$ are identical, and $(P_1, \ldots, P_n) \prec_{\mathrm{h}} (Q_1, \ldots, Q_n)$, then $P_1, \ldots, P_n$ are also identical.

(iii) If $Q_1, \ldots, Q_n$ are equivalent, and $(P_1, \ldots, P_n) \prec_{\mathrm{h}} (Q_1, \ldots, Q_n)$, then $P_1, \ldots, P_n$ are also equivalent.

(iv) If $Q_1, \ldots, Q_n$ are mutually singular, then $(P_1, \ldots, P_n) \prec_{\mathrm{h}} (Q_1, \ldots, Q_n)$.

(v) If $P_1, \ldots, P_n$ are mutually singular, and $(P_1, \ldots, P_n) \prec_{\mathrm{h}} (Q_1, \ldots, Q_n)$, then $Q_1, \ldots, Q_n$ are also mutually singular.

Proof. (i) It is straightforward to verify that

$$\left.\left(\frac{\mathrm{d}P_1}{\mathrm{d}P_1}, \ldots, \frac{\mathrm{d}P_n}{\mathrm{d}P_1}\right)\right|_{P_1} \stackrel{\mathrm{d}}{=} (1, \ldots, 1)|_{P_1} \prec_{\mathrm{cx}} \left.\left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \ldots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q}\right)\right|_{Q} \tag{2.6}$$

for any $Q \in \mathcal{M}_2$ dominating $(Q_1, \ldots, Q_n)$. Therefore, $(P_1, \ldots, P_n) \prec_{\mathrm{h}} (Q_1, \ldots, Q_n)$.

(ii) By $(P_1,\dots,P_n) \prec_{\mathrm{h}} (Q_1,\dots,Q_n)$ and Lemma 2.3.5, we have

$$\left.\left(\frac{\mathrm{d}P_1}{\mathrm{d}P},\dots,\frac{\mathrm{d}P_n}{\mathrm{d}P}\right)\right|_P \prec_{\mathrm{cx}} (1,\dots,1)|_{Q_1} \tag{2.7}$$

holds for some $P \in \mathcal{M}_1$ dominating $(P_1,\dots,P_n)$. By Lemma 2.3.2, (2.7) further implies $\mathrm{d}P_1/\mathrm{d}P = 1$ P-almost surely; thus $P_1,\dots,P_n$ are identical.

(iii) Let $P = \frac{1}{n}\sum_{i=1}^n P_i$ and $Q = \frac{1}{n}\sum_{i=1}^n Q_i$. $(P_1,\dots,P_n) \prec_{\mathrm{h}} (Q_1,\dots,Q_n)$ implies that, for each $i = 1,\dots,n$,

$$\left.\frac{\mathrm{d}P_i}{\mathrm{d}P}\right|_P \prec_{\mathrm{cx}} \left.\frac{\mathrm{d}Q_i}{\mathrm{d}Q}\right|_Q.$$

Note that $Q(\mathrm{d}Q_i/\mathrm{d}Q = 0) = 0$ as $Q_1,\dots,Q_n$ are equivalent. By Lemma 2.3.2, we know $P(\mathrm{d}P_i/\mathrm{d}P = 0) = 0$, which implies $P \ll P_i$. Thus, $P_1,\dots,P_n$ are equivalent.

(iv) As $Q_1,\dots,Q_n$ are mutually singular, there exists a partition $\{\Omega_1,\dots,\Omega_n\} \subset \mathcal{A}$ of Ω such that $Q_i(\Omega_i) = 1$, $i = 1,\dots,n$. Let $P = \frac{1}{n}\sum_{i=1}^n P_i$ and $Q = \frac{1}{n}\sum_{i=1}^n Q_i$. Note that

$$\left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q},\dots,\frac{\mathrm{d}Q_n}{\mathrm{d}Q}\right) = n \times (\mathrm{I}_{\Omega_1},\dots,\mathrm{I}_{\Omega_n})$$

takes values in the vertices of the simplex $S = \{(s_1,\dots,s_n) \in \mathbb{R}_+^n : \sum_{i=1}^n s_i = n\}$, and $(\frac{\mathrm{d}P_1}{\mathrm{d}F},\dots,\frac{\mathrm{d}P_n}{\mathrm{d}P})$ takes values in S. Furthermore, $\mathbb{E}^P[(\frac{\mathrm{d}P_1}{\mathrm{d}P},\dots,\frac{\mathrm{d}P_n}{\mathrm{d}P})] = (1,\dots,1) = \mathbb{E}^Q[(\frac{\mathrm{d}Q_1}{\mathrm{d}Q},\dots,\frac{\mathrm{d}Q_n}{\mathrm{d}Q})]$. By the Choquet-Meyer Theorem (Choquet and Meyer (1963); see Section 10 of Phelps (2001)), stating that among random vectors distributed in a simplex, the maximal elements with respect to convex order are supported over the vertices of the simplex, we have

$$\left.\left(\frac{\mathrm{d}P_1}{\mathrm{d}P},\dots,\frac{\mathrm{d}P_n}{\mathrm{d}P}\right)\right|_P \prec_{\mathrm{cx}} \left.\left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q},\dots,\frac{\mathrm{d}Q_n}{\mathrm{d}Q}\right)\right|_Q.$$

(v) Using the notation in (iv), $(\frac{\mathrm{d}P_1}{\mathrm{d}P},\dots,\frac{\mathrm{d}P_n}{\mathrm{d}P})$ takes values in the vertices of the simplex S, and $(\frac{\mathrm{d}Q_1}{\mathrm{d}Q},\dots,\frac{\mathrm{d}Q_n}{\mathrm{d}Q})$ takes values in S. Therefore, by the Choquet-Meyer Theorem again, in order for $(P_1,\dots,P_n) \prec_{\mathrm{h}} (Q_1,\dots,Q_n)$ to hold, $(\frac{\mathrm{d}Q_1}{\mathrm{d}Q},\dots,\frac{\mathrm{d}Q_n}{\mathrm{d}Q})$ has to be distributed over the vertices of the simplex S, and therefore, $Q_1,\dots,Q_n$ are mutually singular. $\square$

There exists a concept called "majorization" in statistical decision theory, which is closely related to the heterogeneity order given in Definition 2.3.4. We will discuss this link in detail in Remark 2.3.19 after we present Theorem 2.3.15, which also finds an alternative version in the context of comparison of experiments.

2.3.3 Almost compatibility

In Section 2.3.2, we see that a necessary condition for compatibility of $(Q_1, \dots, Q_n) \in \mathcal{P}^n$ and $(F_1, \dots, F_n) \in \mathcal{F}^n$ is $(F_1, \dots, F_n) \prec_{\mathrm{h}} (Q_1, \dots, Q_n)$. A natural question is whether (and with what additional assumptions) the above condition is sufficient for compatibility of $(Q_1, \dots, Q_n)$ and $(F_1, \dots, F_n)$. This boils down (via Theorem 2.2.2) to the question of, given

$$\left(\frac{\mathrm{d}F_1}{\mathrm{d}F}, \dots, \frac{\mathrm{d}F_n}{\mathrm{d}F}\right)\bigg|_F \prec_{\mathrm{cx}} \left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \dots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q}\right)\bigg|_Q,$$

where $F = \frac{1}{n}\sum_{i=1}^n F_i$ and $Q = \frac{1}{n}\sum_{i=1}^n Q_i$, constructing a random variable X with distribution F under Q such that

$$\left(\frac{\mathrm{d}F_1}{\mathrm{d}F}, \dots, \frac{\mathrm{d}F_n}{\mathrm{d}F}\right)(X) = \mathbb{E}^Q\left[\left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \dots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q}\right)\bigg|\sigma(X)\right]. \tag{2.8}$$

Such problem is similar to Lemma 2.3.2, and more generally, the martingale construction in Strassen (1965) or Hirsch et al. (2011), albeit we need to construct X in the pre-specified space $(\Omega, \mathcal{A}, Q)$. Therefore, the existence of X satisfying (2.8) naturally depends on the probability space $(\Omega, \mathcal{A}, Q)$. As a simple example, if F is a continuous distribution and one of $Q_1, \dots, Q_n$ is not atomless, then there does not exist a random variable X with distribution F under each of $Q_1, \dots, Q_n$, although $(F, \dots, F) \prec_{\mathrm{h}} (Q_1, \dots, Q_n)$ by Proposition 2.3.6 (i).

It seems then natural to assume that each of $Q_1, \dots, Q_n$ is atomless. Below we give a counter example showing that this condition is still insufficient.

Example 2.3.7. Let $\Omega = [0,1]$, $\mathcal{A} = \mathcal{B}([0,1])$, $Q_2 = \lambda$ be the Lebesgue measure, $\frac{\mathrm{d}Q_1}{\mathrm{d}Q_2}(t) = 2t$, $t \in [0,1]$, $F_2 = \lambda$ on $[0,1]$ and $\frac{\mathrm{d}F_1}{\mathrm{d}F_2}(x) = |4x-2|$, $x \in [0,1]$. For this setting we have $(F_1, F_2) \prec_{\mathrm{h}} (Q_1, Q_2)$ but (F_1, F_2) and (Q_1, Q_2) are not compatible. The details of these statements are given in Section 2.6.1.

Example 2.3.7 suggests that the heterogeneity order condition $(F_1, \dots, F_n) \prec_{\mathrm{h}} (Q_1, \dots, Q_n)$ is not sufficient for compatibility of $(Q_1, \dots, Q_n)$ and $(F_1, \dots, F_n)$. Nevertheless, in this section we show that, assuming $Q_1, \dots, Q_n$ are atomless, $(F_1, \dots, F_n) \prec_{\mathrm{h}} (Q_1, \dots, Q_n)$ is sufficient for *almost compatibility*, a weaker notion than compatibility, which we introduce below. Denote by $D_{\mathrm{KL}}(\cdot||\cdot)$ the Kullback-Leibler divergence between probability measures.

Definition 2.3.8. $(Q_1, \dots, Q_n) \in \mathcal{P}^n$ and $(F_1, \dots, F_n) \in \mathcal{F}^n$ are *almost compatible*, if for any $\varepsilon > 0$, there exists a random variable X_ε in $(\Omega, \mathcal{A})$ such that for each $i = 1, \dots, n$, the distribution of X_ε under Q_i, denoted by $F_{i,\varepsilon}$, is absolutely continuous with respect to F_i, and satisfies $D_{\mathrm{KL}}(F_{i,\varepsilon}||F_i) < \varepsilon$.

The following theorem characterizes almost compatibility via heterogeneity order in Definition 2.3.4, assuming each of $Q_1, \dots, Q_n$ is atomless.

Theorem 2.3.9. *Suppose that* $(Q_1, \dots, Q_n) \in \mathcal{P}^n$, $(F_1, \dots, F_n) \in \mathcal{F}^n$ *and each of* $Q_1, \dots, Q_n$ *is atomless.* $(Q_1, \dots, Q_n)$ *and* $(F_1, \dots, F_n)$ *are almost compatible if and only if* $(F_1, \dots, F_n) \prec_{\mathrm{h}} (Q_1, \dots, Q_n)$.

The proof of Theorem 2.3.9 is a bit lengthy, and is postponed to Section 2.6.2 of the chapter.

Remark 2.3.10. The Kullback-Leibler divergence in Definition 2.3.8 is not the only possible choice to provide an equivalent condition in Theorem 2.3.9. Indeed, the condition for necessity can be weakened to the convergence in probability of $\mathrm{d}F_{i,\varepsilon}/\mathrm{d}F_i$ to 1 as $\varepsilon \to 0$, by using Fatou's lemma and the fact that a sequence converging in probability has a subsequence converging almost surely; the proof for sufficiency implies results as strong as the uniform convergence of $\mathrm{d}F_{i,\varepsilon}/\mathrm{d}F_i$ to 1. Consequently, the Kullback-Leibler divergence used in the definition of the almost compatibility can be replaced by a series of other conditions, including:

(i) $\mathrm{d}F_{i,\varepsilon}/\mathrm{d}F_i \xrightarrow{\mathrm{p}} 1$;

(ii) $\mathrm{d}F_{i,\varepsilon}/\mathrm{d}F_i \xrightarrow{\mathrm{a.s.}} 1$;

(iii) $F_{i,\varepsilon}$ converges to F_i in total variation, and $F_{i,\varepsilon} \ll F_i$;

(iv) The Rényi divergence of order ∞ between $F_{i,\varepsilon}$ and F_i converges to 0 as $\varepsilon \to 0$,

among others, without altering the result of Theorem 2.3.9.

Almost compatibility has a practical implication for optimization problems. Suppose that $Q_1, \ldots, Q_n$ are atomless. For optimization problems of the form

$$\sup\{\phi(P \circ Y^{-1}) : Y \in L(\Omega, \mathcal{A}) \text{ has distribution } F_i \text{ under } Q_i,\ i = 1, \ldots, n\}, \tag{2.9}$$

where $\phi : \mathcal{F} \to [-\infty, \infty]$ is a functional, it suffices to consider

$$\sup\left\{\phi(F) : F \in \mathcal{F},\ (F_1, \ldots, F_n, F) \prec_{\mathrm{h}} (Q_1, \ldots, Q_n, P)\right\},$$

as long as ϕ is continuous with respect to any of the convergence types listed in Remark 2.3.10. In Section 2.5 we present detailed discussions and examples of optimization problems of the type (2.9).

2.3.4 Equivalence of heterogeneous order and compatibility

In view of the discussions in Section 2.3.3, $(F_1, \ldots, F_n) \prec_{\mathrm{h}} (Q_1, \ldots, Q_n)$ is not sufficient for compatibility of $(Q_1, \ldots, Q_n)$ and $(F_1, \ldots, F_n)$, but sufficient for almost compatibility if each of $Q_1, \ldots, Q_n$ is atomless. In this section, we seek for a slightly stronger condition on the n-tuple $(Q_1, \ldots, Q_n)$, under which compatibility and almost compatibility coincide.

Definition 2.3.11. $(Q_1, \ldots, Q_n) \in \mathcal{P}^n$ is *conditionally atomless* if there exist $Q \in \mathcal{P}$ dominating $(Q_1, \ldots, Q_n)$ and $X \in L(\Omega, \mathcal{A})$ such that under Q, X is continuously distributed and independent of $(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \ldots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q})$.

Clearly, if $(Q_1, \ldots, Q_n)$ is conditionally atomless, then each of $Q_1, \ldots, Q_n$ is atomless, since a continuous random variable under Q is also continuous under each $Q_1, \ldots, Q_n$.

Remark 2.3.12. If $Q_1, \ldots, Q_n$ are mutually singular and each of them is atomless, then $(Q_1, \ldots, Q_n)$ is conditionally atomless. This can be seen directly by constructing a uniform random variable U_i on $[0, 1]$ under Q_i for $i = 1, \ldots, n$, and writing $Q = \frac{1}{n}\sum_{i=1}^n Q_i$. As $Q_1, \ldots, Q_n$ are mutually singular, there exists a partition $\{\Omega_1, \ldots, \Omega_n\} \subset \mathcal{A}$ of Ω such that $Q_i(\Omega_i) = 1$, $i = 1, \ldots, n$. Then the random variable $U = \sum_{i=1}^n U_i \mathrm{I}_{\Omega_i}$ is uniformly distributed and independent of $(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \ldots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q})$ under Q.

Before approaching the main results of this section, we recall some basic facts about conditional distributions. For random vectors $\mathbf{T}$ and $\mathbf{S}$ defined on a probability space $(\Omega, \mathcal{A}, P)$ and taking values in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively, the conditional distribution of $\mathbf{T}$ given $\mathbf{S}$ (under P), denoted by $\mathbf{T}|\mathbf{S}$, is a mapping from $\mathcal{B}(\mathbb{R}^m) \times \Omega$ to $\mathbb{R}$, such that for each $\omega \in \Omega$, $\mathbf{T}|\mathbf{S}(\cdot, \omega)$ is a probability measure on $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$, and for each $A \in \mathcal{B}(\mathbb{R}^m)$, $\mathbf{T}|\mathbf{S}(A, \cdot) = P(\mathbf{T} \in A|\sigma(\mathbf{S}))$ P-almost surely. We write $\mathbf{T}|\mathbf{S}(\omega)$ for the probability measure $\mathbf{T}|\mathbf{S}(\cdot, \omega)$, and $\mathbf{T}|\mathbf{S}(\omega)_P$ when it is necessary to specify the probability measure P. Moreover, there exists a version of $\mathbf{T}|\mathbf{S}$ for which the conditional distribution only depends on the value of $\mathbf{S}$, i.e., $\mathbf{T}|\mathbf{S}(\omega_1) = \mathbf{T}|\mathbf{S}(\omega_2)$ whenever $\mathbf{S}(\omega_1) = \mathbf{S}(\omega_2)$. We will always use this version. For an event $E \in \mathcal{A}$, the conditional probability of E given $\mathbf{S} = s$, denoted by $P(E|\mathbf{S} = s)$, should be understood as $P[E|\sigma(\mathbf{S})](\omega)$ for ω satisfying $\mathbf{S}(\omega) = s$.

With the help of conditional distributions, we first note that the independence in Definition 2.3.11 is not essential and can be replaced by continuity of the conditional distribution. Moreover, similarly to the heterogeneity order, the reference probability measure Q can always be taken as $Q = \frac{1}{n}\sum_{i=1}^n Q_i$.

Proposition 2.3.13. *For $(Q_1, \ldots, Q_n) \in \mathcal{P}^n$, the following are equivalent:*

(i) *$(Q_1, \ldots, Q_n)$ is conditionally atomless.*

(ii) *For $Q = \frac{1}{n}\sum_{i=1}^n Q_i$, there exists a continuous random variable in $(\Omega, \mathcal{A})$ independent of $(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \ldots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q})$ under Q.*

(iii) *There exists $X \in L(\Omega, \mathcal{A})$ such that for some $Q \in \mathcal{P}$ dominating $(Q_1, \ldots, Q_n)$ (equivalently, for $Q = \frac{1}{n}\sum_{i=1}^n Q_i$), a version of the conditional distribution $X|\mathbf{Y}$ is everywhere continuous under Q where $\mathbf{Y} = (\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \ldots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q})$.*

Proof. Note that (iii) has two versions: one states the existence of Q and the other specifies Q. It is trivial to see that (ii) implies (i) and both versions of (iii). It remains to show (iii)$\Rightarrow$(i)$\Rightarrow$(ii).

We first show (i)$\Rightarrow$(ii). Assume $(Q_1, \ldots, Q_n)$ is conditionally atomless, and thus there exist $Q' \in \mathcal{P}$ and a random variable X, such that X and $\mathbf{Y} := \left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q'}, \ldots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q'}\right)$ are independent under Q'. For $i = 1, \dots, n$, any $A \in \mathcal{B}(\mathbb{R})$ and $B \in \mathcal{B}(\mathbb{R}^n)$,

$$\begin{aligned} Q_i(X \in A, \mathbf{Y} \in B) &= \mathbb{E}^{Q'}\left[\frac{\mathrm{d}Q_i}{\mathrm{d}Q'}\mathrm{I}_{\{X\in A\}}\mathrm{I}_{\{\mathbf{Y}\in B\}}\right] \\ &= \mathbb{E}^{Q'}[\mathrm{I}_{\{X\in A\}}]\mathbb{E}^{Q'}\left[\frac{\mathrm{d}Q_i}{\mathrm{d}Q'}\mathrm{I}_{\{\mathbf{Y}\in B\}}\right] = Q'(X \in A)Q_i(\mathbf{Y} \in B). \end{aligned}$$

The independence between X and $\mathbf{Y}$ also implies that

$$Q_i(X \in A) = \mathbb{E}^{Q'}\left[\frac{\mathrm{d}Q_i}{\mathrm{d}Q'}\mathrm{I}_{\{X\in A\}}\right] = \mathbb{E}^{Q'}\left[\frac{\mathrm{d}Q_i}{\mathrm{d}Q'}\right]\mathbb{E}^{Q'}[\mathrm{I}_{\{X\in A\}}] = Q'(X \in A).$$

Thus, X has the same distribution under Q_i, $i = 1, \dots, n$. Let $Q = \frac{1}{n}\sum_{i=1}^n Q_i$, and note that X also has the same distribution under Q. Moreover,

$$Q_i(X \in A, \mathbf{Y} \in B) = Q'(X \in A)Q_i(\mathbf{Y} \in B) = Q_i(X \in A)Q_i(\mathbf{Y} \in B),$$

which means that X and $\mathbf{Y}$ are independent under Q_i for $i = 1, \dots, n$. For any $A \in \mathcal{B}(\mathbb{R})$ and $B \in \mathcal{B}(\mathbb{R}^n)$,

$$\begin{aligned} Q(X \in A, \mathbf{Y} \in B) &= \frac{1}{n}\sum_{i=1}^n Q_i(X \in A, \mathbf{Y} \in B) \\ &= \frac{1}{n}\sum_{i=1}^n Q_i(X \in A)Q_i(\mathbf{Y} \in B) \\ &= Q(X \in A)\frac{1}{n}\sum_{i=1}^n Q_i(\mathbf{Y} \in B) = Q(X \in A)Q(\mathbf{Y} \in B), \end{aligned}$$

and hence X and $\mathbf{Y}$ are independent under Q. As a result, X is also independent of

$$\frac{\mathbf{Y}}{\|\mathbf{Y}\|_1} = \frac{1}{n}\left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \dots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q}\right)$$

under Q, where $||\cdot||_1$ is the Manhattan norm on $\mathbb{R}^n$. Therefore, we conclude that X and $\left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \dots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q}\right)$ are independent under Q.

Next we prove (iii)⇒(i). Assume there exists a random variable X in $L(\Omega, \mathcal{A})$ such that for some $Q \in \mathcal{P}$ dominating $Q_1, \dots, Q_n$, the conditional distribution $X|\mathbf{Y}$ is everywhere continuous under Q for $\mathbf{Y} = (\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \dots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q})$. For any $\omega \in \Omega$, let F_ω be the distribution function of $X|\mathbf{Y}(\omega)$, and define $X' : \Omega \to \mathbb{R}$ by $X'(\omega) = F_\omega(X(\omega))$. It is fundamental, though a bit lengthy, to check that X' is a random variable; moreover, $X'|\mathbf{Y}$ almost surely follows a uniform distribution on $[0, 1]$. As a result, X' is a continuous random variable independent of $(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \dots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q})$ under Q. Consequently, both versions of (iii) imply (i). $\square$

Remark 2.3.14. As a byproduct of the above proof, we note that if a random variable X is independent of $(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \dots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q})$ under a probability measure Q, then X is also independent of $(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \dots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q})$ under each of $Q_1, \dots, Q_n$. Moreover, X has the same distribution under $Q_1, \dots, Q_n$ and Q.

Now we turn back to our main target, compatibility of $(F_1, \dots, F_n)$ and $(Q_1, \dots, Q_n)$. As discussed in Section 2.3.3, to show compatibility one needs to construct a random variable X in $(\Omega, \mathcal{A})$ such that

$$\left(\frac{\mathrm{d}F_1}{\mathrm{d}F}, \dots, \frac{\mathrm{d}F_n}{\mathrm{d}F}\right)(X) = \mathbb{E}^Q\left[\left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \dots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q}\right)\Big|\sigma(X)\right].$$

It turns out that the assumption that $(Q_1, \dots, Q_n)$ is conditionally atomless allows for such a construction.

Theorem 2.3.15. *Suppose that $(Q_1, \dots, Q_n) \in \mathcal{P}^n$ is conditionally atomless and $(F_1, \dots, F_n) \in \mathcal{F}^n$. $(Q_1, \dots, Q_n)$ and $(F_1, \dots, F_n)$ are compatible if and only if $(F_1, \dots, F_n) \prec_{\mathrm{h}} (Q_1, \dots, Q_n)$.*

The key step to prove Theorem 2.3.15 is the following lemma, which might be of independent interest.

Lemma 2.3.16. *Let $\mathbf{X} = (X_1, \dots, X_m)$ and $\mathbf{Y} = (Y_1, \dots, Y_n)$ be random vectors defined on probability spaces $(\Omega_1, \mathcal{A}_1, P_1)$ and $(\Omega_2, \mathcal{A}_2, P_2)$, respectively, and f be a measurable function from $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ to $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. If $f(\mathbf{X})|_{P_1} \prec_{\mathrm{cx}} \mathbf{Y}|_{P_2}$, and there exists a continuous random variable U defined on $(\Omega_2, \mathcal{A}_2, P_2)$ independent of $\mathbf{Y}$, then there exists a random vector $\mathbf{W} = (W_1, \dots, W_m)$ defined on $(\Omega_2, \mathcal{A}_2, P_2)$, such that $\mathbf{W}|_{P_2} \overset{\mathrm{d}}{=} \mathbf{X}|_{P_1}$, and*

$$f(\mathbf{W}) = \mathbb{E}^{P_2}[\mathbf{Y}|\mathbf{W}].$$

Proof. Since $f(\mathbf{X})|_{P_1} \prec_{\mathrm{cx}} \mathbf{Y}|_{P_2}$, by Lemma 2.3.2, there exists a probability space $(\Omega', \mathcal{A}', P')$ and random vectors $\mathbf{Z}$, $\mathbf{Y}'$ defined on it and taking values in $\mathbb{R}^n$, such that $\mathbf{Z}|_{P'} \overset{\mathrm{d}}{=} f(\mathbf{X})|_{P_1}$, $\mathbf{Y}'|_{P'} \overset{\mathrm{d}}{=} \mathbf{Y}|_{P_2}$, and $\mathbf{Z} = \mathbb{E}^{P'}[\mathbf{Y}'|\mathbf{Z}]$.

Construct random vectors $\mathbf{X}'' = (X_1'', \ldots, X_m'')$ and $\mathbf{Y}'' = (Y_1'', \ldots, Y_n'')$ on a (possibly different) probability space $(\Omega'', \mathcal{A}'', P'')$, such that $\mathbf{X}''|_{P''} \overset{\mathrm{d}}{=} \mathbf{X}|_{P_1}$ and the conditional distributions satisfy $\mathbf{Y}''|\mathbf{X}''(\omega'')_{P''} = \mathbf{Y}'|\mathbf{Z}(\omega')_{P'}$ for all ω', ω'' satisfying $\mathbf{Z}(\omega') = f(\mathbf{X}''(\omega''))$. It is easy to see that $\mathbf{Y}''|_{P''} \overset{\mathrm{d}}{=} \mathbf{Y}|_{P_2}$, and

$$\mathbb{E}^{P''}[\mathbf{Y}''|\mathbf{X}''](\omega'') = \mathbb{E}^{P'}[\mathbf{Y}'|\mathbf{Z}](\omega') = \mathbf{Z}(\omega') = f(\mathbf{X}''(\omega'')), \quad \text{for } P''\text{-a.s. } \omega'' \in \Omega''.$$

What is left is therefore to construct $\mathbf{W}$ on $(\Omega_2, \mathcal{A}_2, P_2)$ such that $(\mathbf{W}, \mathbf{Y})|_{P_2} \overset{\mathrm{d}}{=} (\mathbf{X}'', \mathbf{Y}'')|_{P''}$. The idea is similar to the proof of Theorem 2.3.9. More precisely, for $l = 0, 1, \ldots$ and $h = (h_1, \ldots, h_m) \in \mathbb{Z}^m$, consider the distribution of $\mathbf{Y}''$ restricted on the event $\{X_i'' \in [h_i 2^{-l}, (h_i + 1)2^{-l}),\ 1 \leqslant i \leqslant m\}$. It has a density function, denoted by $\psi_{l,h}(\mathbf{y})$, $\mathbf{y} \in \mathbb{R}^n$, with respect to the unconditional distribution of $\mathbf{Y}''$. Without loss of generality, assume U follows a uniform distribution on $[0, 1]$. Then for each $\mathbf{y}$ and $l = 0, 1, \ldots,$ we divide $[0, 1]$ into disjoint intervals $\{I_{l,h}(\mathbf{y})\}_{h \in \mathbb{Z}^m}$, such that $|I_{l,h}(\mathbf{y})| = \psi_{l,h}(\mathbf{y})$. Moreover, we can make $\{I_{l',h}(\mathbf{y})\}_{h \in \mathbb{Z}^m}$ a refinement of $\{I_{l,h}(\mathbf{y})\}_{h \in \mathbb{Z}^m}$ for any $l' > l$. Then define random vector $\mathbf{W}_l = (W_{l,1}, \ldots, W_{l,m})$ by

$$W_{l,i} = h_i 2^{-l} \text{ for } U \in I_{l,h}(\mathbf{Y}), \quad i = 1, \ldots, m.$$

Let $\mathbf{W} = \lim_{l \to \infty} \mathbf{W}_l$. The point-wise limit exists due to the completeness of $\mathbb{R}^m$.

For any given $\mathbf{y}$, any $l = 0, 1, \ldots$ and $h \in \mathbb{Z}^m$,

$$\begin{aligned}
&P_2(\mathbf{W}_i \in [h_i 2^{-l}, (h_i + 1)2^{-l}), 1 \leqslant i \leqslant m | \mathbf{Y} = \mathbf{y}) \\
&= P_2(\mathbf{W}_{l,i} = h_i 2^{-l}, 1 \leqslant i \leqslant m | \mathbf{Y} = \mathbf{y}) \\
&= \psi_{l,h}(\mathbf{y}) = P''(\mathbf{X}_i'' \in [h_i 2^{-l}, (h_i + 1)2^{-l}), 1 \leqslant i \leqslant m | \mathbf{Y}'' = \mathbf{y}),
\end{aligned}$$

Since $\{[h_i 2^{-l}, (h_i + 1)2^{-l})\}_{h \in \mathbb{Z}^m, l = 0, 1, \ldots}$ forms a basis for $\mathcal{B}(\mathbb{R}^m)$, $\mathbf{W}|\mathbf{Y}(\omega)$ under P_2 equals $\mathbf{X}''|\mathbf{Y}''(\omega'')$ for $\omega \in \Omega$ and $\omega'' \in \Omega''$ satisfying $\mathbf{Y}(\omega) = \mathbf{Y}''(\omega'')$. Moreover, recall that $\mathbf{Y}|_{P_2} \overset{\mathrm{d}}{=} \mathbf{Y}''|_{P''}$. As a result, we conclude that $(\mathbf{W}, \mathbf{Y})|_{P_2} \overset{\mathrm{d}}{=} (\mathbf{X}'', \mathbf{Y}'')_{P''}$. $\square$

Proof of Theorem 2.3.15. Necessity is guaranteed by Lemma 2.3.3. We only show sufficiency. Suppose that $(F_1, \ldots, F_n) \prec_{\mathrm{h}} (Q_1, \ldots, Q_n)$. We shall show that $(Q_1, \ldots, Q_n)$ and $(F_1, \ldots, F_n)$ are compatible. By Lemma 2.3.5,

$$\left(\frac{\mathrm{d}F_1}{\mathrm{d}F}, \ldots, \frac{\mathrm{d}F_n}{\mathrm{d}F}\right)\bigg|_F \prec_{\mathrm{cx}} \left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \ldots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q}\right)\bigg|_Q$$

for $F = \frac{1}{n}\sum_{i=1}^n F_i$ and $Q = \frac{1}{n}\sum_{i=1}^n Q_i$. Since $(Q_1, \ldots, Q_n)$ is conditionally atomless, $Q_1, \ldots, Q_n$ are all atomless, so is Q. Hence there exists a random variable X' defined on $(\Omega, \mathcal{A})$, such that $F = Q \circ X'^{-1}$. As a result,

$$\left(\frac{\mathrm{d}F_1}{\mathrm{d}F}, \ldots, \frac{\mathrm{d}F_n}{\mathrm{d}F}\right)(X')\bigg|_Q \overset{\mathrm{d}}{=} \left(\frac{\mathrm{d}F_1}{\mathrm{d}F}, \ldots, \frac{\mathrm{d}F_n}{\mathrm{d}F}\right)\bigg|_F \prec_{\mathrm{cx}} \left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \ldots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q}\right)\bigg|_Q.$$

Applying Lemma 2.3.16 with $f(x) = \left(\frac{\mathrm{d}F_1}{\mathrm{d}F}, \ldots, \frac{\mathrm{d}F_n}{\mathrm{d}F}\right)(x)$, there exists a random variable X defined on $(\Omega, \mathcal{A})$, such that

$$\left(\frac{\mathrm{d}F_1}{\mathrm{d}F}, \ldots, \frac{\mathrm{d}F_n}{\mathrm{d}F}\right)(X) = \mathbb{E}^Q\left[\left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \ldots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q}\right)\bigg|\,\sigma(X)\right],$$

which, by Theorem 2.2.2, implies compatibility. ☐

Remark 2.3.17. As shown in Theorem 2.3.15, compatibility is closely related to heterogeneity order $\prec_{\mathrm{h}}$, and hence it defines a partial order. The direction of the order comes from the fact that a measurable mapping needs not to be a bijection. As multiple points are mapped to a same image, the "heterogeneity" between measures decreases. However, if we require the mapping to be a bijection, then compatibility becomes an equivalence relation. Indeed, in this case Theorem 2.3.15 would be applicable to both directions, which means that (2.5) holds for both directions, with $P = \frac{1}{n}\sum_{i=1}^n P_i$ and $Q = \frac{1}{n}\sum_{i=1}^n Q_i$. As a result, we must have

$$\left(\frac{\mathrm{d}P_1}{\mathrm{d}P}, \ldots, \frac{\mathrm{d}P_n}{\mathrm{d}P}\right)\bigg|_P \overset{\mathrm{d}}{=} \left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \ldots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q}\right)\bigg|_Q.$$

Moreover, the proof of Theorem 2.3.15 actually shows that, assuming both tuples of measures are conditionally atomless, the above condition is not only necessary but also sufficient to guarantee the existence of a bijection linking $(P_1, \ldots, P_n)$ to $(Q_1, \ldots, Q_n)$.

In the following corollary of our main result for $n = 2$, the heterogeneity order condition becomes one-dimensional, and is easy to check. Chapter 3 of Shaked and Shanthikumar (2007) contains several classic methods to check $X|_P \prec_{\mathrm{cx}} Y|_Q$ for arbitrary random variables X and Y and probability measures P and Q. A convenient equivalent condition of $X|_P \prec_{\mathrm{cx}} Y|_Q$ is that $\mathbb{E}^P[X] = \mathbb{E}^Q[Y]$ and $\int_y^\infty P(X > x)\mathrm{d}x \leqslant \int_y^\infty Q(Y > x)\mathrm{d}x$ for all $y \in \mathbb{R}$ (e.g. Theorem 3.A.1 of Shaked and Shanthikumar (2007)).

Corollary 2.3.18. *Suppose that $(Q_1, Q_2) \in \mathcal{P}^2$, $Q_1 \ll Q_2$, and $(F_1, F_2) \in \mathcal{F}^2$. If (Q_1, Q_2) and (F_1, F_2) are compatible, then $F_1 \ll F_2$ and*

$$\left.\frac{\mathrm{d}F_1}{\mathrm{d}F_2}\right|_{F_2} \prec_{\mathrm{cx}} \left.\frac{\mathrm{d}Q_1}{\mathrm{d}Q_2}\right|_{Q_2}. \tag{2.10}$$

Conversely, if $F_1 \ll F_2$, (2.10) holds, and in addition, (Q_1, Q_2) is conditionally atomless, then (Q_1, Q_2) and (F_1, F_2) are compatible.

Proof. Necessity follows from Corollary 2.2.3. Sufficiency follows from the simple observation that, by taking $F = F_2$, (2.10) implies $(F_1, F_2) \prec_{\mathrm{h}} (Q_1, Q_2)$. □

Below we discuss a few special cases of compatible $(Q_1, \dots, Q_n) \in \mathcal{P}^n$ and $(F_1, \dots, F_n) \in \mathcal{F}^n$ based on the heterogeneity order condition, in particular in the context of Proposition 2.3.6 and Theorem 2.3.15. We shall see how our main results are consistent with natural intuitions.

1. Assume that $Q_1, \dots, Q_n$ are identical. The natural intuition is that the respective distributions $F_1, \dots, F_n$ of a random variable under $Q_1, \dots, Q_n$ have to be identical. By Lemma 2.3.3, compatibility implies $(F_1, \dots, F_n) \prec_{\mathrm{h}} (Q_1, \dots, Q_n)$. By Proposition 2.3.6 (ii), $F_1, \dots, F_n$ are identical.

2. Assume that $Q_1, \dots, Q_n$ are mutually singular, and each of them is atomless. The natural intuition here is that the respective distributions $F_1, \dots, F_n$ of any random variable under $Q_1, \dots, Q_n$ are arbitrary. Proposition 2.3.6 (iv) suggests that $(F_1, \dots, F_n) \prec_{\mathrm{h}} (Q_1, \dots, Q_n)$ holds for any $(F_1, \dots, F_n) \in \mathcal{F}^n$. Moreover, $(Q_1, \dots, Q_n)$ is conditionally atomless, as seen in Remark 2.3.12. Therefore, by Theorem 2.3.15,

a mutually singular tuple of atomless probability measures on $(\Omega, \mathcal{A})$ is compatible with an arbitrary tuple of distributions on $\mathbb{R}$.

3. Assume that $F_1, \ldots, F_n$ are mutually singular. The natural intuition here is that the probability measures $Q_1, \ldots, Q_n$ have to be also mutually singular to allow for compatibility. Similarly to the previous case, this is justified by Theorem 2.3.15 and Proposition 2.3.6 (v).

4. Assume that $F_1, \ldots, F_n$ are identical, and $(Q_1, \ldots, Q_n)$ is conditionally atomless. Proposition 2.3.6 (i) gives $(F_1, \ldots, F_n) \prec_{\mathrm{h}} (Q_1, \ldots, Q_n)$. It follows from Theorem 2.3.15 that $(Q_1, \ldots, Q_n)$ and $(F_1, \ldots, F_n)$ are compatible. We conclude that, as long as $(Q_1, \ldots, Q_n)$ is conditionally atomless, for any distribution $F \in \mathcal{F}$, there exists a random variable X which has distribution F under each of Q_i, $i = 1, \ldots, n$. Indeed, as $(Q_1, \ldots, Q_n)$ is conditionally atomless, there exists Q dominating $(Q_1, \ldots, Q_n)$ and an F-distributed random variable X under Q independent of $(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \ldots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q})$. Remark 2.3.14 then implies that X also has distribution F under each $Q_1, \ldots, Q_n$.

5. Assume that $Q_1, \ldots, Q_n$ are equivalent. Intuitively, the respective distributions $F_1, \ldots, F_n$ of any random variable under $Q_1, \ldots, Q_n$ have to be equivalent. This fact is implied by Proposition 2.3.6 (iii).

We conclude this section by discussing the relation of Theorem 2.3.15 with some results in statistical decision theory.

Remark 2.3.19. As pointed out previously, Theorem 2.3.15 and the heterogeneity order in Definition 2.3.4 finds an important relation to comparison of statistical experiments, an area of study originated by Blackwell (Blackwell, 1951, 1953); the reader is referred to Le Cam (1996) and Torgersen (1991) for summaries. Very briefly, the question in the latter literature is to compare two experiments in terms of the information they can provide. Mathematically this translates into defining a partial order among two sets of measures of the same cardinality. Such an order is called majorization, and one way to define it is through (2.1). It is then shown that the majorization between two sets of probability measures is equivalent to the existence of a (Markov) transition kernel which turns each measure in one set into a measure in the other set. This is mathematically closely related to our definition of compatibility. As such, Theorem 2.3.15 finds a slightly different version in the context of comparison of statistical experiments. Nevertheless, the existence of a transition kernel is weaker than the existence of a point-to-point mapping; consequently, the conditionally atomless assumption does not appear in the statistical decision theory literature.

2.4 Distributional compatibility for stochastic processes

2.4.1 General results

In this section we extend our results to stochastic processes with sample paths which are continuous from right with left limits (càdlàg). For a (finite or infinite) closed interval $I \subset \mathbb{R}$, let $D(I)$ be the Skorokhod space on I, *i.e.*, the space of all càdlàg functions defined on I. Let $\mathcal{D}_I$ be the Borel σ-field of the Skorokhod topology. Denote by $\mathcal{G}_I$ the set of probability measures on $(D(I), \mathcal{D}_I)$. Our first step is to generalize the definition of compatibility to this setting, which follows in a natural way.

Definition 2.4.1. For a closed interval $I \subset \mathbb{R}$, $(Q_1, \dots, Q_n) \in \mathcal{P}^n$ and $(G_1, \dots, G_n) \in \mathcal{G}_I^n$ are *compatible* if there exists a càdlàg stochastic process $X = \{X(t)\}_{t \in I}$ defined on $(\Omega, \mathcal{A})$ such that for each $i = 1, \dots, n$, the distribution of X under Q_i is G_i.

The following result is a parallel result to Theorem 2.2.2, which shares the same proof.

Proposition 2.4.2. *Let $I \subset \mathbb{R}$ be a closed interval, $(Q_1, \dots, Q_n) \in \mathcal{P}^n$ and $(G_1, \dots, G_n) \in \mathcal{G}_I^n$. A stochastic process X has distribution G_i under Q_i for $i = 1, \dots, n$ if and only if for all $Q \in \mathcal{P}$ dominating $(Q_1, \dots, Q_n)$, $G = Q \circ X^{-1}$ dominates $(G_1, \dots, G_n)$, and*

$$\left(\frac{\mathrm{d}G_1}{\mathrm{d}G}, \dots, \frac{\mathrm{d}G_n}{\mathrm{d}G}\right)(X) = \mathbb{E}^Q\left[\left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \dots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q}\right)\Big|\sigma(X)\right].$$

In the proof of Lemma 2.3.3, no structure of the real line has been used. As a result, Lemma 2.3.3 can be directly generalized to the case of stochastic processes, with $(G_1, \dots, G_n) \in \mathcal{G}_I^n$ replacing $(F_1, \dots, F_n) \in \mathcal{F}^n$. For the other side we have, parallel to Theorem 2.3.15:

Theorem 2.4.3. *Suppose that* $(Q_1, \dots, Q_n) \in \mathcal{P}^n$ *is conditionally atomless,* $I \subset \mathbb{R}$ *is a closed interval, and* $(G_1, \dots, G_n) \in \mathcal{G}_I^n$*.* $(Q_1, \dots, Q_n)$ *and* $(G_1, \dots, G_n)$ *are compatible if and only if* $(G_1, \dots, G_n) \prec_{\mathrm{h}} (Q_1, \dots, Q_n)$*.*

Proof. The proof is similar to that of Theorem 2.3.15. The only difference is that $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is replaced by $(D(I), \mathcal{D}_I)$. A careful check of the proof of Theorem 2.3.15 shows, however, that it only relies on the completely metrizable structure of $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ to guarantee the existence and uniqueness of the limit of the constructed sequence of random variables. Since $(D(I), \mathcal{D}_I)$ is also completely metrizable, the proof naturally extends to the case of stochastic processes. More precisely, label the rational numbers in I as $\mathbb{Q} \cap I = \{t_1, t_2, \dots\}$. Then we replace the refining partition of the real line $\{[h2^{-l}, (h+1)2^{-l}),\ h \in \mathbb{Z}\}_{l=0,1,\dots}$ with the refining partition of $D(I)$: $\{X(t_i) \in [h_{l,i}2^{-l+i}, (h_{l,i}+1)2^{-l+i}),\ i = 1, \dots, l,\ h_{l,i} \in \mathbb{Z}\}_{l=1,2,\dots}$. The rest follows in the same way as in the proof of Theorem 2.3.15. $\square$

2.4.2 Relation to the Girsanov Theorem

In this section we investigate how much the drift of a Brownian motion may vary under a change of measure as in the classic Girsanov Theorem. We keep in mind that, the distribution of a Brownian motion (with respect to its natural filtration) with a deterministic drift process only depends on this drift. On the other hand, Brownian motions with stochastic drift processes are not identified by the distribution of the drift processes. Due to this reason, we consider only Brownian motions with deterministic drift processes here.

Throughout this section, let $P \in \mathcal{P}$ and $B = \{B_t\}_{t \in [0,T]}$ be a P-standard Brownian motion. Furthermore, for a $[0,T]$-square integrable deterministic process $\theta = \{\theta_t\}_{t \in [0,T]}$, define

$$\frac{\mathrm{d}Q_\theta}{\mathrm{d}P} = e^{\int_0^T \theta_t \mathrm{d}B_t - \frac{1}{2}\int_0^T \theta_t^2 \mathrm{d}t},$$

and let G_θ be the distribution measure of a Brownian motion with drift process θ. The Girsanov Theorem says that B is a Brownian motion with drift process θ and volatility 1 under Q_θ (certainly, this statement is also true for adapted drift processes). Thus, (P, Q_θ) and (G_0, G_θ) are compatible. It is clear that distribution measures of Brownian motions with different non-random volatility terms are mutually singular, and hence they are not compatible with (P, Q_θ). A next question is whether there exists a P-standard Brownian motion which has a deterministic drift process $\mu = \{\mu_t\}_{t\in[0,T]}$ under Q_θ. We are interested in the values of μ such that (G_0, G_μ) and (P, Q_θ) above are compatible. Here we do not assume that (P, Q_θ) is conditionally atomless, which means that there might not be any other random source other than B.

Theorem 2.4.4. *Suppose that the deterministic processes $\theta = \{\theta_t\}_{t\in[0,T]}$ and $\mu = \{\mu_t\}_{t\in[0,T]}$ are $[0,T]$-square integrable, and $\mu_t \neq 0$ almost everywhere on $[0,T]$. (P, Q_θ) and (G_0, G_μ) are compatible if and only if*

$$\int_0^T \mu_t^2 \mathrm{d}t \leqslant \int_0^T \theta_t^2 \mathrm{d}t.$$

Proof. (i) Necessity. By the Girsanov Theorem, we know that (G_0, G_μ) and (P, Q_μ) are compatible. Using Proposition 2.4.2 for $n = 2$, we have

$$\frac{\mathrm{d}G_\mu}{\mathrm{d}G_0}(B) = \mathbb{E}\left[\frac{\mathrm{d}Q_\mu}{\mathrm{d}P}\middle|\sigma(B)\right] = e^{\int_0^T \mu_t \mathrm{d}B_t - \frac{1}{2}\int_0^T \mu_t^2 \mathrm{d}t}.$$

Suppose that (P, Q_θ) and (G_0, G_μ) are compatible. Note that

$$e^{\int_0^T \mu_t \mathrm{d}B_t - \frac{1}{2}\int_0^T \mu_t^2 \mathrm{d}t}\Big|_P \overset{\mathrm{d}}{=} \frac{\mathrm{d}G_\mu}{\mathrm{d}G_0}(B)\Big|_P \overset{\mathrm{d}}{=} \frac{\mathrm{d}G_\mu}{\mathrm{d}G_0}\Big|_{G_0}.$$

By Theorem 2.4.3, we have

$$e^{\int_0^T \mu_t \mathrm{d}B_t - \frac{1}{2}\int_0^T \mu_t^2 \mathrm{d}t}\Big|_P \prec_{\mathrm{cx}} \frac{\mathrm{d}Q_\theta}{\mathrm{d}P}\Big|_P \overset{\mathrm{d}}{=} e^{\int_0^T \theta_t \mathrm{d}B_t - \frac{1}{2}\int_0^T \theta_t^2 \mathrm{d}t}\Big|_P.$$

Applying the convex function $x \mapsto x^2$, we have

$$e^{\int_0^T \mu_t^2 \mathrm{d}t} = \mathbb{E}[(e^{\int_0^T \mu_t \mathrm{d}B_t - \frac{1}{2}\int_0^T \mu_t^2 \mathrm{d}t})^2] \leqslant \mathbb{E}[(e^{\int_0^T \theta_t \mathrm{d}B_t - \frac{1}{2}\int_0^T \theta_t^2 \mathrm{d}t})^2] = e^{\int_0^T \theta_t^2 \mathrm{d}t}$$

and hence $\int_0^T \mu_t^2 \mathrm{d}t \leqslant \int_0^T \theta_t^2 \mathrm{d}t$.

(ii) Sufficiency. Suppose $\int_0^T \mu_t^2 \mathrm{d}t \leqslant \int_0^T \theta_t^2 \mathrm{d}t$. Define a deterministic process $\alpha = \{\alpha_t\}_{t\in[0,T]}$ by

$$\alpha_t = \inf\left\{r \geqslant 0 : \int_0^r \theta_s^2 \mathrm{d}s = \int_0^t \mu_s^2 \mathrm{d}s\right\}.$$

It is easy to see that α_t is strictly increasing in t, $\alpha_T \leqslant T$, and furthermore,

$$\theta_{\alpha_t}^2 \mathrm{d}\alpha_t = \mu_t^2 \mathrm{d}t. \tag{2.11}$$

Let a stochastic process $\hat{B} = \{\hat{B}_t\}_{t\in[0,T]}$ be given by $\mathrm{d}\hat{B}_t = \mathrm{d}B_t - \theta_t \mathrm{d}t$. By the Girsanov Theorem, $\hat{B}$ is a Q_θ-standard Brownian motion. Define

$$W_t = \int_0^t \beta_{\alpha_s} \mathrm{d}B_{\alpha_s}, \quad t \in [0,T],$$

where $\beta = \{\beta_s\}_{s\in[0,\alpha_T]}$ is given by $\beta_{\alpha_t} = \frac{\theta_{\alpha_t}}{\mu_t}$, $t \in [0,T]$. Clearly, $W = \{W_t\}_{t\in[0,T]}$ is a Gaussian process, $\mathbb{E}^P[W_t] = 0$, and

$$\mathbb{E}^P[W_t W_s] = \mathbb{E}^P[W_s^2] = \int_0^s \frac{\theta_{\alpha_u}^2}{\mu_u^2} \mathrm{d}\alpha_u = s, \quad 0 \leqslant s < t \leqslant T.$$

Therefore, W is a P-standard Brownian motion. Furthermore, for $t \in [0,T]$,

$$\begin{aligned} W_t = \int_0^t \beta_{\alpha_s} \mathrm{d}B_{\alpha_s} &= \int_0^t \beta_{\alpha_s} (\mathrm{d}\hat{B}_{\alpha_s} + \theta_{\alpha_s} \mathrm{d}\alpha_s) \\ &= \int_0^t \beta_{\alpha_s} \mathrm{d}\hat{B}_{\alpha_s} + \int_0^t \beta_{\alpha_s} \theta_{\alpha_s} \mathrm{d}\alpha_s \\ &= \int_0^t \beta_{\alpha_s} \mathrm{d}\hat{B}_{\alpha_s} + \int_0^t \mu_s \mathrm{d}s, \end{aligned}$$

where the last equality is due to (2.11). As $\int_0^t \beta_{\alpha_s} \mathrm{d}\hat{B}_{\alpha_s}$ defines a Q_θ-standard Brownian motion, we conclude that W has distribution G_μ under Q_θ, and hence (P, Q_θ) and (G_0, G_μ) are compatible. □

We list Theorem 2.4.4 for the case of a constant drift term below, and look more closely at the construction of the desired stochastic process.

Corollary 2.4.5. *Let $\theta_t = a$ and $\mu_t = b$, $t \in [0,T]$, where a and b are two constants, and $b \neq 0$. (P, Q_θ) and (G_0, G_μ) are compatible if and only if $b^2 \leqslant a^2$.*

If $b^2 \leqslant a^2$, the process which has distribution G_0 under P and distribution G_μ under Q_θ can be written in a simple explicit form. Let

$$W_t = \frac{a}{b} B_{(\frac{b}{a})^2 t}, \quad t \in [0,T].$$

It is clear that $W = \{W_t\}_{t\in[0,T]}$ is a P-Brownian motion. Furthermore,

$$W_t = \frac{a}{b} B_{(\frac{b}{a})^2 t} = \frac{a}{b}\left(\hat{B}_{(\frac{b}{a})^2 t} + a\frac{b^2}{a^2}t\right) = \frac{a}{b}\hat{B}_{(\frac{b}{a})^2 t} + bt, \quad t \in [0,T].$$

In this example, it is clear that $0 < b^2 \leqslant a^2$ is essential; otherwise W will not be well-defined.

2.5 Related optimization problems

2.5.1 General problems

In this section, we discuss some optimization problems related to compatibility of distributions and probability measures. For given $P, Q_1, \ldots, Q_n \in \mathcal{P}$, $F_1, \ldots, F_n \in \mathcal{F}$ and an objective $\phi : \mathcal{F} \to [-\infty, \infty]$, we focus on optimization problems of the form

$$\max\{\phi(P \circ Y^{-1}) : Y \in L(\Omega, \mathcal{A}) \text{ has distribution } F_i \text{ under } Q_i,\ i = 1, \ldots, n\}. \tag{2.12}$$

Here we assume $(Q_1, \ldots, Q_n)$ and $(F_1, \ldots, F_n)$ are compatible so that the above problem is properly posed, and for the sake of illustration, we assume the maximum is attained; otherwise it should be a supremum.

To simplify notation, define the set

$$L_{F_1,\ldots,F_n}(Q_1, \ldots, Q_n) = \{Y \in L(\Omega, \mathcal{A}) : Y \text{ has distribution } F_i \text{ under } Q_i,\ i = 1, \ldots, n\}.$$

Note that $L_{F_1,\ldots,F_n}(Q_1, \ldots, Q_n)$ is non-empty if and only if $(Q_1, \ldots, Q_n)$ and $(F_1, \ldots, F_n)$ are compatible. Then, (2.12) reads as

$$\max\{\phi(P \circ Y^{-1}) : Y \in L_{F_1,\ldots,F_n}(Q_1, \ldots, Q_n)\}.$$

The optimization (2.12) includes many well-known problems; see Examples 2.5.4-2.5.14 below.

As mentioned in Section 2.3.3, our main results imply that the optimization (2.12) admits an alternative form

$$\max\left\{\phi(F) : F \in \mathcal{F},\ (F_1, \ldots, F_n, F) \prec_{\mathrm{h}} (Q_1, \ldots, Q_n, P)\right\}, \tag{2.13}$$

under some continuity assumption of ϕ.

The optimization problem in (2.12) is highly challenging even for $n = 1$, and few analytical solutions are available. We first focus on the case $n = 1$. In this case, (2.12) reads as

$$\max\{\phi(P \circ Y^{-1}) : Y \in L_G(Q)\}, \tag{2.14}$$

where $P, Q \in \mathcal{P}$ and $G \in \mathcal{F}$ are given. The optimization (2.14) includes a large class of practical problems involving various types of uncertainty. For instance, P may represent the real-world probability measure, and Q represents the pricing measure in a financial market; distribution of an asset price Y under Q may be inferred from traded option prices on Y (e.g. Jarrow and Rudd (1982), Buchen and Kelly (1996)) but its distribution under P is unclear.

Assuming that (P, Q) is conditionally atomless and $P \ll Q$, our results imply the equivalent formulation of (2.14)

$$\max\left\{\phi(F) : F \in \mathcal{F},\ \frac{\mathrm{d}F}{\mathrm{d}G}\Big|_G \prec_{\mathrm{cx}} \frac{\mathrm{d}P}{\mathrm{d}Q}\Big|_Q\right\}. \tag{2.15}$$

In the next subsections, we discuss some results related to (2.12)-(2.15).

2.5.2 The set of compatible distributions and f-divergences

A straightforward consequence of our main results is that we arrive at inequalities for f-divergences, relating to some special cases of (2.14). For two probability measures P_1, P_2 on an arbitrary probability space, the *f-divergence* d_f is defined as

$$d_f(P_1, P_2) = \int f\left(\frac{\mathrm{d}P_1}{\mathrm{d}P_2}\right) \mathrm{d}P_2,$$

where f is a convex function $f : \mathbb{R}_+ \to \mathbb{R}$. The Kullback-Leibler divergence ($f(x) = x\log(x)$), the total variation distance ($f(x) = (x-1)^+$), and the Hellinger distance ($f(x) = (\sqrt{x}-1)^2$) are special cases of f-divergences. Noting that $(F, G) \prec_{\mathrm{h}} (P, Q)$ can be rewritten as an order between $d_f(F, G)$ and $d_f(P, Q)$, we have the following corollary, which is a direct consequence of Corollary 2.3.18.

Corollary 2.5.1. *Suppose that $(P,Q) \in \mathcal{P}^2$ is conditionally atomless, $P \ll Q$, and $(F,G) \in \mathcal{F}^2$. (P,Q) and (F,G) are compatible if and only if $F \ll G$ and $d_f(F,G) \leqslant d_f(P,Q)$ for all f-divergences d_f.*

In the problem (2.14) for given $P,Q \in \mathcal{P}$ and $G \in \mathcal{F}$, Corollary 2.5.1 becomes useful, as it gives conditions, by choosing suitable f-divergences, on what F may be, such that (P,Q) and (F,G) are compatible. An immediate consequence is that the set of all such F is a convex set.

Corollary 2.5.2. *Suppose that $(P,Q) \in \mathcal{P}^2$ is conditionally atomless, $P \ll Q$, and $G \in \mathcal{F}$. The set $\{F \in \mathcal{F} : (P,Q)$ and (F,G) are compatible$\}$ is convex.*

Proof. Denote by $\mathcal{F}_G = \{F \in \mathcal{F} : (P,Q)$ and (F,G) are compatible$\}$. For $F_1, F_2 \in \mathcal{F}_G$, any convex function f, and $\lambda \in [0,1]$, by Corollary 2.5.1 we have

$$\begin{aligned} d_f(\lambda F_1 + (1-\lambda)F_2, G) &= \int_{\mathbb{R}} f\left(\lambda \frac{\mathrm{d}F_1}{\mathrm{d}G} + (1-\lambda)\frac{\mathrm{d}F_2}{\mathrm{d}G}\right) \mathrm{d}G \\ &\leqslant \lambda d_f(F_1, G) + (1-\lambda) d_f(F_2, G) \\ &\leqslant \lambda d_f(P,Q) + (1-\lambda) d_f(P,Q) = d_f(P,Q), \end{aligned}$$

which, by Corollary 2.5.1 again, implies the compatibility of $(\lambda F_1 + (1-\lambda)F_2, G)$ and (P,Q), and hence $\lambda F_1 + (1-\lambda)F_2 \in \mathcal{F}_G$. $\square$

Corollary 2.5.2 will be useful in some optimization problems; see Example 2.5.13 below.

Remark 2.5.3. We make two observations regarding Corollary 2.5.2.

1. The conditionally atomless assumption is essential for Corollary 2.5.2. Note that, in a discrete probability space, a mixture of distributions may no longer be a possible distribution of a random variable in that probability space; a similar phenomenon appears if (P,Q) is not conditionally atomless.

2. Using Theorem 2.3.15, it can be checked that the statement in Corollary 2.5.2 holds in the multi-dimensional case, that is, assuming $(P, Q_1, \dots, Q_n) \in \mathcal{P}^{n+1}$ is conditionally atomless, the set $\{F \in \mathcal{F} : (P, Q_1, \dots, Q_n)$ and $(F, G_1, \dots, G_n)$ are compatible$\}$ is convex for each $(G_1, \dots, G_n) \in \mathcal{F}^n$.

In all examples of Section 2.5, for simplicity, we shall assume that $P, Q \in \mathcal{P}$ are equivalent and (P, Q) is conditionally atomless. In case (P, Q) is not conditionally atomless, the maximum should be replaced by a supremum in several places. The following example is a simple application of Corollary 2.5.1.

Example 2.5.4 (f-divergence)**.** For given $G \in \mathcal{F}$, let

$$\overline{d}_f(G) = \max_{Y \in L_G(Q)} d_f(P \circ Y^{-1}, G) \quad \text{and} \quad \underline{d}_f(G) = \min_{Y \in L_G(Q)} d_f(P \circ Y^{-1}, G).$$

Then we have

$$\overline{d}_f(G) = d_f(P, Q) \quad \text{and} \quad \underline{d}_f(G) = 0.$$

To show the first statement, take $X \in L_G(Q)$ such that $\frac{\mathrm{d}P}{\mathrm{d}Q} \in \sigma(X)$. Such X always exists if Q is atomless (e.g. Lemma A.32 of Föllmer and Schied (2016)). Let F be the distribution of X under P. By Theorem 2.2.2, we have

$$\frac{\mathrm{d}F}{\mathrm{d}G}(X) = \mathbb{E}^Q \left[\frac{\mathrm{d}P}{\mathrm{d}Q} \Big| \sigma(X) \right] = \frac{\mathrm{d}P}{\mathrm{d}Q},$$

which implies

$$\int_{\mathbb{R}} f\left(\frac{\mathrm{d}F}{\mathrm{d}G}\right) \mathrm{d}G = \int_{\Omega} f\left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right) \mathrm{d}Q.$$

Therefore, $d_f(F, G) = d_f(P, Q)$, which implies $\overline{d}_f(G) \geqslant d_f(P, Q)$. The reverse inequality $\overline{d}_f(G) \leqslant d_f(P, Q)$ is immediate from Corollary 2.5.1. The second statement is straightforward by noting that (P, Q) and (G, G) are compatible.

2.5.3 Optimization of monotone objectives

In the following, we consider the case that ϕ in (2.14) is a monotone functional with respect to the univariate stochastic order on a given probability space $(\Omega, \mathcal{A}, P)$. This setting includes many classic problems.

Definition 2.5.5 (Univariate stochastic order)**.** For $X, Y \in L(\Omega, \mathcal{A})$ and $P \in \mathcal{P}$, we write $X|_P \prec_{\mathrm{st}} Y|_P$, if $P(X > x) \leqslant P(Y > x)$ for all $x \in \mathbb{R}$. We shall also write $F \prec_{\mathrm{st}} G$ for $F, G \in \mathcal{F}$, if $F((x, \infty)) \leqslant G((x, \infty))$ for all $x \in \mathbb{R}$.

If ϕ is monotone with respect to $\prec_{\mathrm{st}}$, then to solve (2.14), it suffices to find the maximum and the minimum elements in $L_G(Q)$ with respect to stochastic order under P. In what follows, we identify a measure $F \in \mathcal{F}$ with its distribution function, and write its generalized inverse

$$F^{-1}(t) = \inf\{x \in \mathbb{R} : F(x) \geqslant t\},\ t \in (0,1].$$

Recall that, for an atomless probability measure Q and any random variable $X \in L_F(Q)$, there exists a uniform random variable on $[0,1]$, denoted by $U(X;Q)$, such that $F^{-1}(U(X;Q)) = X$ Q-almost surely.

Proposition 2.5.6. *Suppose that $P, Q \in \mathcal{P}$ are equivalent and atomless, and $G \in \mathcal{F}$. Denote by $U = U(\frac{\mathrm{d}P}{\mathrm{d}Q}; Q)$ and let $X^* = G^{-1}(U)$ and $X_* = G^{-1}(1-U)$. Then $X_*, X^* \in L_G(Q)$, and $X_*|_P \prec_{\mathrm{st}} Y|_P \prec_{\mathrm{st}} X^*|_P$ for all $Y \in L_G(Q)$.*

Proof. For $Y \in L_G(Q)$, denote by F the distribution of Y under P. Moreover, given any $a \in \mathbb{R}$, define $Z = \mathrm{I}_{\{Y>a\}}$, and let G', F' be the distributions of Z under Q and P, respectively. Then

$$\frac{\mathrm{d}F'}{\mathrm{d}G'} = \begin{cases} \frac{F(Y>a)}{G(Y>a)} & \text{if } Y > a; \\ \frac{F(Y\leqslant a)}{G(Y\leqslant a)} & \text{if } Y \leqslant a, \end{cases}$$

with probabilities $G(Y > a)$ and $G(Y \leqslant a)$ under G', respectively.

Applying Corollary 2.3.18 to (Q,P), (G',F') and convex function $f(x) = (x-b)^+$, $b \in \mathbb{R}$, we have

$$\begin{aligned} P\left(\frac{\mathrm{d}P}{\mathrm{d}Q} > b\right) - bQ\left(\frac{\mathrm{d}P}{\mathrm{d}Q} > b\right) &\geqslant G(Y>a)\left(\frac{F(Y>a)}{G(Y>a)} - b\right)^+ + G(Y\leqslant a)\left(\frac{F(Y\leqslant a)}{G(Y\leqslant a)} - b\right)^+ \\ &\geqslant F(Y>a) - bG(Y>a). \end{aligned}$$

Therefore each b gives an upper bound $P\left(\frac{\mathrm{d}P}{\mathrm{d}Q} > b\right) - bQ\left(\frac{\mathrm{d}P}{\mathrm{d}Q} > b\right) + bG(Y>a)$ for $F(Y > a)$. On the other hand, for b such that

$$Q\left(\frac{\mathrm{d}P}{\mathrm{d}Q} > b\right) \leqslant G(Y>a) \leqslant Q\left(\frac{\mathrm{d}P}{\mathrm{d}Q} \geqslant b\right),$$

it is straightforward and intuitive to see that $X^* = G^{-1}(U)$ achieves this bound, given the fact that X^* and $\frac{\mathrm{d}P}{\mathrm{d}Q}$ are comonotonic. Since this is true for all $a \in \mathbb{R}$, we conclude that $Y|_P \prec_{\mathrm{st}} X^*|_P$ for all $Y \in L_G(Q)$. The other half of the proposition can be proved symmetrically. $\square$

Remark 2.5.7. Proposition 2.5.6 can alternatively be obtained using a classic method of Fréchet-Hoeffding bounds, and it is known in the literature in a different form (see Example 2.5.8 below).

Proposition 2.5.6 yields solutions to optimization problems in (2.14), where $\phi : \mathcal{F} \to [-\infty, \infty]$ is an increasing or decreasing functional with respect to $\prec_{\mathrm{st}}$. A few classic examples are presented below. In all the following examples, $U = U(\frac{\mathrm{d}P}{\mathrm{d}Q}; Q)$, and the distribution function of $\frac{\mathrm{d}P}{\mathrm{d}Q}$ under Q is denoted by $H_{P,Q}$. Recall that we assume (P, Q) is conditionally atomless, and P, Q are equivalent.

Example 2.5.8 (Fréchet-Hoeffding)**.** The value of

$$M(G) = \max\{\mathbb{E}^P[Y] : Y \in L_G(Q)\}$$

can be obtained, via Proposition 2.5.6, as

$$M(G) = \mathbb{E}^P[X^*] = \mathbb{E}^Q\left[\frac{\mathrm{d}P}{\mathrm{d}Q} G^{-1}(U)\right] = \int_0^1 H_{P,Q}^{-1}(u) G^{-1}(u) \mathrm{d}u.$$

The value of $M(G)$ is known as the classic Fréchet-Hoeffding bound; see Remark 3.25 of Rüschendorf (2013).

Example 2.5.9 (Neyman-Pearson)**.** The value of

$$k(q) = \max\{P(A) : A \in \mathcal{A},\ Q(A) = q\}$$

can be obtained via Proposition 2.5.6. By letting $X = \mathrm{I}_A$ and G be its distribution under Q, we have

$$k(q) = \mathbb{E}^P[X^*] = \mathbb{E}^Q\left[\frac{\mathrm{d}P}{\mathrm{d}Q} G^{-1}(U)\right] = \int_{1-q}^1 H_{P,Q}^{-1}(u) \mathrm{d}u.$$

The optimal X^* has the form $X^* = \mathrm{I}_{\{U \geqslant 1-q\}}$. This result is known as the Neyman-Pearson lemma (Neyman and Pearson, 1933) in statistical hypothesis testing. Alternatively, it is known as the classic knapsack problem in a continuous setting.

Example 2.5.10 (Robust utility)**.** The value of

$$\underline{R}_u(G) = \min\{\mathbb{E}^P[u(Y)] : Y \in L_G(Q)\}$$

where $u : \mathbb{R} \to \mathbb{R}$ is an increasing utility function, can be obtained, via Proposition 2.5.6, as

$$\underline{R}_u(G) = \mathbb{E}^P[u(X_*)] = \mathbb{E}^Q\left[\frac{\mathrm{d}P}{\mathrm{d}Q}u(G^{-1}(1-U))\right] = \int_0^1 H^{-1}_{P,Q}(t)u(G^{-1}(1-t))\mathrm{d}t.$$

The value $\underline{R}_u(G)$ represents the worst-case expected utility of the random outcome Y under P if one knows the distribution of Y is G under another measure Q. The functional $\underline{R}_u$ itself is a rank-dependent utility functional in decision theory (Quiggin, 2012).

2.5.4 Optimization of non-monotone objectives

If ϕ in (2.14) is not monotone, then Proposition 2.5.6 cannot be applied directly. In such cases, we need to investigate the problem in more details, utilizing Theorem 2.3.15. In the following, $v : \mathbb{R} \to \mathbb{R}$ is a measurable function, and $\phi : \mathcal{F} \to [-\infty, \infty]$ as in (2.14), and we do not assume monotonicity of v or ϕ. For $G \in \mathcal{F}$, denote by G_v the distribution of $v(Y)$ where Y has distribution G.

Proposition 2.5.11. *Suppose that $(P, Q) \in \mathcal{P}^2$ is conditionally atomless, $P \ll Q$ and $G \in \mathcal{F}$. Then we have*

$$\sup\{\phi(P \circ (v(Y))^{-1}) : Y \in L_G(Q)\} = \sup\{\phi(P \circ Z^{-1}) : Z \in L_{G_v}(Q)\}. \tag{2.16}$$

Moreover, both supremums in (2.16) are attained simultaneously or none of them is attained.

Proof. Let $\overline{T}(G) = \sup\{\phi(P \circ (v(Y))^{-1}) : Y \in L_G(Q)\}$. First, it trivially holds that

$$\overline{T}(G) \leqslant \sup\{\phi(P \circ Z^{-1}) : Z \in L_{G_v}(Q)\}.$$

For the reverse inequality, we give an explicit construction based on the conditional atomless assumption. For any $Z \in L_{G_v}(Q)$, let F_Z be the distribution of Z under P. By Remark

2.3.14, since (P, Q) is conditionally atomless, there exist random variables U_1 and U_2 such that, under both P and Q, U_1 and U_2 are $[0,1]$-uniform, and U_1, U_2 and $\frac{\mathrm{d}P}{\mathrm{d}Q}$ are independent. Consider the measurable space $(\Omega, \mathcal{A}_0)$ where $A_0 = \sigma(\frac{\mathrm{d}P}{\mathrm{d}Q}, U_1)$, and the restricted probability measures $P' = P|_{\mathcal{A}_0}$, $Q' = Q|_{\mathcal{A}_0}$. It is clear that

$$\left(\frac{\mathrm{d}P'}{\mathrm{d}Q'}, U_1\right)\bigg|_{Q'} \stackrel{\mathrm{d}}{=} \left(\frac{\mathrm{d}P}{\mathrm{d}Q}, U_1\right)\bigg|_{Q}. \tag{2.17}$$

Note that (P, Q) and (F_Z, G_v) are compatible, as directly justified by the existence of Z. By Theorem 2.3.15, we have $(F_Z, G_v) \prec_{\mathrm{h}} (P, Q)$. As a result, $(F_Z, G_v) \prec_{\mathrm{h}} (P', Q')$ via (2.17). Moreover, the existence of U_1 and (2.17) assure that (P', Q') is conditionally atomless. Applying Theorem 2.3.15 to the measurable space $(\Omega, \mathcal{A}_0)$, we conclude that there exists an $\mathcal{A}_0$-measurable random variable Z', such that Z' has distribution F_Z under P' and G_v under Q'. As a result, Z' and Z have the same distribution under both P and Q. Furthermore, since Z' is determined by $\frac{\mathrm{d}P}{\mathrm{d}Q}$ and U_1, it is independent of U_2. Now define a random variable $Y' \in L(\Omega, \mathcal{A})$ by $Y'(\omega) = F_\omega^{-1}(U_2(\omega))$, where F_ω is the conditional distribution function of $Y \in L_G(Q)$ given $v(Y) = Z'(\omega)$. Then $Y' \in L_G(Q)$ and $v(Y') = Z' \stackrel{\mathrm{d}}{=} Z$ under both P and Q. As a result, $\overline{T}(G) \geqslant \phi(P \circ Z^{-1})$, and hence the equality holds. The above construction also justifies the statement on the attainability of supremums in (2.16). □

Proposition 2.5.11 allows us to freely transform random variables even if the transform is not one-to-one or monotone. The case of ϕ being the expectation is illustrated below.

Example 2.5.12 (Expectation of a transform)**.** The problem is to find the value of

$$\overline{R}_v(G) = \max\{\mathbb{E}^P[v(Y)] : Y \in L_G(Q)\}.$$

This example is similar to Example 2.5.10, but v is not necessarily monotone, and hence we need to utilize both Propositions 2.5.6 and 2.5.11. Using Propositions 2.5.6 and 2.5.11, we have

$$\overline{R}_v(G) = \max\{\mathbb{E}^P[Z] : Z \in L_{G_v}(Q)\} = \mathbb{E}^Q\left[\frac{\mathrm{d}P}{\mathrm{d}Q} G_v^{-1}(U)\right] = \int_0^1 H_{P,Q}^{-1}(t) G_v^{-1}(t)\mathrm{d}t.$$

One of the most common non-monotone functional ϕ on $\mathcal{F}$ is the variance. We discuss this problem below.

Example 2.5.13 (Robust variance)**.** Assume the distribution G has a finite second moment. The problem is to find the values of

$$\overline{V}(G) = \max\{\mathrm{Var}^P(Y) : Y \in L_G(Q)\} \quad \text{and} \quad \underline{V}(G) = \min\{\mathrm{Var}^P(Y) : Y \in L_G(Q)\},$$

where $\mathrm{Var}^P(Y) = \mathbb{E}^P[(Y - \mathbb{E}^P[Y])^2]$ is the variance of Y under P. For this problem, neither Proposition 2.5.6 nor Proposition 2.5.11 can be directly applied. Nevertheless, using a standard minimax argument, one can show

$$\overline{V}(G) = \max_{Y \in L_G(Q)} \min_{x \in \mathbb{R}} \{\mathbb{E}^P[(Y-x)^2]\} = \min_{x \in \mathbb{R}} \max_{Y \in L_G(Q)} \{\mathbb{E}^P[(Y-x)^2]\} = \min_{x \in \mathbb{R}} \overline{R}_x(G), \quad (2.18)$$

where

$$\overline{R}_x(G) = \max\{\mathbb{E}^P[(Y-x)^2] : Y \in L_G(Q)\},$$

which can be calculated from Proposition 2.5.11 in the same way as in Example 2.5.12. The exchangeability of minimax in (2.18) is justified by Sion's minimax theorem through the facts that the objective $\mathbb{E}^P[(Y-x)^2]$ is convex in x, linear in the distribution of Y under P, and the set of distributions of random variables in $L_G(Q)$ is convex, thanks to Corollary 2.5.2.

On the other hand,

$$\underline{V}(G) = \min_{Y \in L_G(Q)} \min_{x \in \mathbb{R}} \{\mathbb{E}^P[(Y-x)^2]\} = \min_{x \in \mathbb{R}} \min_{Y \in L_G(Q)} \{\mathbb{E}^P[(Y-x)^2]\} = \min_{x \in \mathbb{R}} \underline{R}_x(G)$$

where

$$\underline{R}_x(G) = \min\{\mathbb{E}^P[(Y-x)^2] : Y \in L_G(Q)\},$$

which can also be calculated in the same way as in Example 2.5.12.

In the simple example below, we present the maximum and minimum values of $\mathbb{E}^P[Y]$, $\mathbb{E}^P[Y^2]$ and $\mathrm{Var}^P(Y)$ for a $\mathrm{N}(0,1)$ distributed random variable Y under Q.

Example 2.5.14 (Normal distribution)**.** Let $G = \mathrm{N}(0,1)$, B_1 be $\mathrm{N}(0,1)$ distributed under P, and

$$\frac{\mathrm{d}P}{\mathrm{d}Q} = \exp\left(\frac{1}{2} - B_1\right), \quad \text{or equivalently,} \quad \frac{\mathrm{d}Q}{\mathrm{d}P} = \exp\left(B_1 - \frac{1}{2}\right).$$

This is a special case of the Girsanov change of measure in Section 2.4.2 by choosing $\theta = T = 1$. Assume that (P, Q) is conditionally atomless. Using Examples 2.5.12 and 2.5.13, we obtain

$$\max_{Y \in L_G(Q)} \mathbb{E}^P[Y] = 1,$$

which is attained by $Y^* = 1 - B_1$, and

$$\max_{Y \in L_G(Q)} \mathbb{E}^P[Y^2] = \max_{Y \in L_G(Q)} \mathrm{Var}^P(Y) = \int_0^1 \exp\left(q(u) - \frac{1}{2}\right)\left(q\left(\frac{1+u}{2}\right)\right)^2 \mathrm{d}u \approx 2.795,$$

where q is the quantile function of a standard normal distribution. On the other hand, we have

$$\min_{Y \in L_G(Q)} \mathbb{E}^P[Y] = -1,$$

which is attained by $Y^* = B_1 - 1$, and

$$\min_{Y \in L_G(Q)} \mathbb{E}^P[Y^2] = \min_{Y \in L_G(Q)} \mathrm{Var}^P(Y) = \int_0^1 \exp\left(q(u) - \frac{1}{2}\right)\left(q\left(\frac{2-u}{2}\right)\right)^2 \mathrm{d}u \approx 0.2579.$$

The details of the above calculation is given in Section 2.6.3.

From the above numbers, we note that the maximums of $\mathbb{E}^P[Y]$ and $\mathrm{Var}^P(Y)$ cannot be attained with the same random variable, whereas $\mathbb{E}^P[Y^2]$ and $\mathrm{Var}^P(Y)$ are both attained by the same random variable with mean zero. A similar observation is made for the minimums.

2.5.5 The case of mutual singularity for $n \geqslant 2$

For the case $n \geqslant 2$, the optimization problems in (2.12) are often highly difficult to solve, even if ϕ is assumed to be monotone with respect to $\prec_{\mathrm{st}}$. Results are available for the case that $Q_1, \ldots, Q_n$ are mutually singular, as presented below.

Proposition 2.5.15. *Suppose that* $(Q_1, \ldots, Q_n) \in \mathcal{P}^n$ $(F_1, \ldots, F_n) \in \mathcal{F}^n$ *are compatible,* $Q_1, \ldots, Q_n$ *are atomless and mutually singular with disjoint supports* $\Omega_1, \ldots, \Omega_n \in \mathcal{A}$, *respectively, and* $P \ll \frac{1}{n}\sum_{i=1}^n Q_i$. *For* $i = 1, \ldots, n$, *let* P_i *be given by* $P_i(A) = P(A \cap \Omega_i)$ *for* $A \in \mathcal{A}$, *and* $U_i = U(\frac{\mathrm{d}P_i}{\mathrm{d}Q_i}; Q_i)$. *Let* $X^* = \sum_{i=1}^n F_i^{-1}(U_i)\mathrm{I}_{\Omega_i}$ *and* $X_* = \sum_{i=1}^n F_i^{-1}(1 - U_i)\mathrm{I}_{\Omega_i}$. *Then* $X_*, X^* \in L_{F_1,\ldots,F_n}(Q_1, \ldots, Q_n)$, *and* $X_*|_P \prec_{\mathrm{st}} Y|_P \prec_{\mathrm{st}} X^*|_P$ *for all* $Y \in L_{F_1,\ldots,F_n}(Q_1, \ldots, Q_n)$.

Proof. Note that for $i = 1, \ldots, n$ and $B \in \mathcal{B}(\mathbb{R})$, $Q_i(X^* \in B) = Q_i(F_i^{-1}(U_i) \in B) = F_i(B)$, and therefore $X^* \in L_{F_1,\ldots,F_n}(Q_1, \ldots, Q_n)$. Moreover, Proposition 2.5.6 implies $Y|_{P_i} \prec_{\mathrm{st}} X^*|_{P_i}$ for each $i = 1, \ldots, n$, yielding $Y|_P \prec_{\mathrm{st}} X^*|_P$. The statements of X_* are analogous to those of X^*. □

We conclude this chapter by remarking that a similar result to Proposition 2.5.15 where $Q_1, \ldots, Q_n$ are not mutually singular seems extremely difficult to establish based on existing techniques.

2.6 Technical Details

2.6.1 Details in Example 2.3.7

Note that $\frac{\mathrm{d}Q_1}{\mathrm{d}Q_2}$ is uniform on $[0, 2]$ under $Q_2 = \lambda$, and $\frac{\mathrm{d}F_1}{\mathrm{d}F_2}$ is also uniform on $[0, 2]$ under $F_2 = \lambda$. Thus,

$$\left(\frac{\mathrm{d}F_1}{\mathrm{d}\lambda}, \frac{\mathrm{d}F_2}{\mathrm{d}\lambda}\right)\bigg|_\lambda \stackrel{\mathrm{d}}{=} \left(\frac{\mathrm{d}Q_1}{\mathrm{d}\lambda}, \frac{\mathrm{d}Q_2}{\mathrm{d}\lambda}\right)\bigg|_\lambda.$$

Therefore, $(F_1, F_2) \prec_{\mathrm{h}} (Q_1, Q_2)$.

Next, we will see that (Q_1, Q_2) and (F_1, F_2) are not compatible. Suppose for the purpose of contradiction that (Q_1, Q_2) and (F_1, F_2) are compatible. By Theorem 2.2.2, there exists a random variable X in $(\Omega, \mathcal{A})$ with a uniform distribution on $[0, 1]$ under $Q_2 = \lambda$ such that

$$\frac{\mathrm{d}F_1}{\mathrm{d}\lambda}(X) = \mathbb{E}^\lambda\left[\frac{\mathrm{d}Q_1}{\mathrm{d}\lambda}\bigg|\sigma(X)\right].$$

In addition,

$$\left.\frac{\mathrm{d}F_1}{\mathrm{d}\lambda}(X)\right|_\lambda \overset{\mathrm{d}}{=} \left.\frac{\mathrm{d}Q_1}{\mathrm{d}\lambda}\right|_\lambda,$$

and therefore,

$$\frac{\mathrm{d}F_1}{\mathrm{d}\lambda}(X) = \frac{\mathrm{d}Q_1}{\mathrm{d}\lambda}, \quad \lambda\text{-almost surely.}$$

From the definition of F_1 and Q_1, we have, for λ-almost surely $t \in [0,1]$, $|4X(t) - 2| = 2t$. It follows that $X(t) = (t+1)/2$ or $X(t) = (1-t)/2$ for all $t \in [0,1]$. Write

$$A = \left\{t \in [0,1] : X(t) = \frac{t+1}{2}\right\},\ B = \left\{t \in [0,1] : X(t) = \frac{1-t}{2}\right\} \text{ and } C = \left\{\frac{1-t}{2} : t \in A\right\}.$$

As X is $\mathcal{B}([0,1])$-measurable and has distribution F_2 under λ, we have $A, B \in \mathcal{B}([0,1])$ and $\lambda(A) = \lambda(B) = 1/2$. Note that $\lambda(C) = 1/4$; however $\lambda(C \cap X(A \cup B)) = 0$, contradicting the fact that X has a uniform distribution on $[0,1]$ under λ.

2.6.2 Proof of Theorem 2.3.9

Proof of Theorem 2.3.9. Necessity. Assume that $(Q_1, \dots, Q_n)$ and $(F_1, \dots, F_n)$ are almost compatible. This means that for any $\varepsilon > 0$, there exists $(F_{1,\varepsilon}, \dots, F_{n,\varepsilon})$ such that $D_{\mathrm{KL}}(F_{i,\varepsilon}||F_i) < \varepsilon$ for $i = 1, \dots, n$, and $(Q_1, \dots, Q_n)$ is compatible with $(F_{1,\varepsilon}, \dots, F_{n,\varepsilon})$. Define probability measures

$$F_\varepsilon = \frac{1}{n}(F_{1,\varepsilon} + \cdots + F_{n,\varepsilon}),$$

$$F = \frac{1}{n}(F_1 + \cdots + F_n)$$

and

$$Q = \frac{1}{n}(Q_1 + \cdots + Q_n).$$

Note that the distribution of X_ε under Q is F_ε, where X_ε is the random variable defining the compatibility between $(Q_1, \dots, Q_n)$ and $(F_{1,\varepsilon}, \dots, F_{n,\varepsilon})$. Moreover, $F_{i,\varepsilon} \ll F_\varepsilon$, $Q_i \ll Q$, and $\mathrm{d}F_{i,\varepsilon}/\mathrm{d}F_\varepsilon \leqslant n$, $\mathrm{d}Q_i/\mathrm{d}Q \leqslant n$ for $i = 1, \dots, n$. For $\varepsilon > 0$, by Lemma 2.3.3,

$$\left.\left(\frac{\mathrm{d}F_{1,\varepsilon}}{\mathrm{d}F_\varepsilon}, \dots, \frac{\mathrm{d}F_{n,\varepsilon}}{\mathrm{d}F_\varepsilon}\right)\right|_{F_\varepsilon} \prec_{\mathrm{cx}} \left.\left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q}, \dots, \frac{\mathrm{d}Q_n}{\mathrm{d}Q}\right)\right|_Q.$$

As a result, for any convex function $f : \mathbb{R}^n \to \mathbb{R}$,

$$\mathbb{E}^{F_\varepsilon}\left[f\left(\frac{\mathrm{d}F_{1,\varepsilon}}{\mathrm{d}F_\varepsilon},\dots,\frac{\mathrm{d}F_{n,\varepsilon}}{\mathrm{d}F_\varepsilon}\right)\right] \leqslant \mathbb{E}^{Q}\left[f\left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q},\dots,\frac{\mathrm{d}Q_n}{\mathrm{d}Q}\right)\right].$$

For $i = 1,\dots,n$,

$$\frac{\mathrm{d}F_{i,\varepsilon}}{\mathrm{d}F_\varepsilon} = \frac{\mathrm{d}F_i}{\mathrm{d}F}\frac{\mathrm{d}F_{i,\varepsilon}/\mathrm{d}F_i}{\mathrm{d}F_\varepsilon/\mathrm{d}F}. \tag{2.19}$$

Since $D_{\mathrm{KL}}(F_{i,\varepsilon}||F_i)$ converges to 0, by Pinsker's inequality, $F_{i,\varepsilon}$ converges to F_i in total variation, which is equivalent to $\mathrm{d}F_{i,\varepsilon}/\mathrm{d}F_i$ converging in $L^1|_{F_i}$ to 1. Hence for any sequence $\varepsilon_m \downarrow 0$, there exists a subsequence, which we still denote as $\varepsilon_m \downarrow 0$ by a slight abuse of notation, such that $\mathrm{d}F_{i,\varepsilon_m}/\mathrm{d}F_i$ converge to 1 F_i-almost surely. It is easy to check that we have $\mathrm{d}F_{\varepsilon_m}/\mathrm{d}F$ converge to 1 as well. (2.19) then implies that

$$\frac{\mathrm{d}F_{i,\varepsilon_m}}{\mathrm{d}F_{\varepsilon_m}} \to \frac{\mathrm{d}F_i}{\mathrm{d}F} \quad F_i\text{-almost surely.} \tag{2.20}$$

On any set $B \in \mathcal{B}(\mathbb{R})$ such that $F_i(B) = 0$ but $F(B) > 0$, suppose $\mathrm{d}F_{i,\varepsilon}/\mathrm{d}F_\varepsilon$ does not converge to $\mathrm{d}F_i/\mathrm{d}F = 0$ in probability under $F|_B$, the measure F restricted on B. Then there exists a positive number $\delta > 0$, and a subsequence of ε_m (again denoted as ε_m), such that $P^{F|_B}(\mathrm{d}F_{i,\varepsilon_m}/\mathrm{d}F_{\varepsilon_m} > \delta) \geqslant c$ for some constant $c > 0$. Since F_{ε_m} converges to F in total variation, for m large enough, $P^{F_{\varepsilon_m}|_B}(\mathrm{d}F_{i,\varepsilon_m}/\mathrm{d}F_{\varepsilon_m} > \delta) \geqslant c/2$. Hence $F_{i,\varepsilon_m}(B) \geqslant \delta P^{F_{\varepsilon_m}|_B}(\mathrm{d}F_{i,\varepsilon_m}/\mathrm{d}F_{\varepsilon_m} > \delta) \geqslant \frac{c\delta}{2}$, which contradicts the fact that F_{i,ε_m} converges to F_i in total variation. We conclude that $\mathrm{d}F_{i,\varepsilon}/\mathrm{d}F_\varepsilon$ converge to $\mathrm{d}F_i/\mathrm{d}F = 0$ in probability under F on set $\{\mathrm{d}F_i/\mathrm{d}F = 0\}$. Combining this result with (2.20) and taking a further subsequence allows us to replace the F_i-almost sure convergence in (2.20) by F-almost sure convergence.

For any convex function $f : \mathbb{R}^n \to \mathbb{R}$,

$$\mathbb{E}^{F_{\varepsilon_m}}\left[f\left(\frac{\mathrm{d}F_{1,\varepsilon_m}}{\mathrm{d}F_{\varepsilon_m}},\dots,\frac{\mathrm{d}F_{n,\varepsilon_m}}{\mathrm{d}F_{\varepsilon_m}}\right)\right] = \int f\left(\frac{\mathrm{d}F_{1,\varepsilon_m}}{\mathrm{d}F_{\varepsilon_m}},\dots,\frac{\mathrm{d}F_{n,\varepsilon_m}}{\mathrm{d}F_{\varepsilon_m}}\right)\mathrm{d}F_{\varepsilon_m}.$$

Since $\frac{\mathrm{d}F_{i,\varepsilon_m}}{\mathrm{d}F_{\varepsilon_m}} \in [0,n]$, and f is convex hence continuous, $\left|f\left(\frac{\mathrm{d}F_{1,\varepsilon_m}}{\mathrm{d}F_{\varepsilon_m}},\dots,\frac{\mathrm{d}F_{n,\varepsilon_m}}{\mathrm{d}F_{\varepsilon_m}}\right)\right|$ is bounded. Let b be an upper bound of it. Because F_{ε_m} converges in total variation to F, we have

$$\left|\int f\left(\frac{\mathrm{d}F_{1,\varepsilon_m}}{\mathrm{d}F_{\varepsilon_m}},\dots,\frac{\mathrm{d}F_{n,\varepsilon_m}}{\mathrm{d}F_{\varepsilon_m}}\right)\mathrm{d}F_{\varepsilon_m} - \int f\left(\frac{\mathrm{d}F_{1,\varepsilon_m}}{\mathrm{d}F_{\varepsilon_m}},\dots,\frac{\mathrm{d}F_{n,\varepsilon_m}}{\mathrm{d}F_{\varepsilon_m}}\right)\mathrm{d}F\right| \leqslant 2b\delta(F_{\varepsilon_m},F) \to 0 \tag{2.21}$$

uniformly, where $\delta(\cdot,\cdot)$ is the total variation distance. Moreover, by dominated convergence, we have

$$\int f\left(\frac{\mathrm{d}F_{1,\varepsilon_m}}{\mathrm{d}F_{\varepsilon_m}},\dots,\frac{\mathrm{d}F_{n,\varepsilon_m}}{\mathrm{d}F_{\varepsilon_m}}\right)\mathrm{d}F \to \int f\left(\frac{\mathrm{d}F_1}{\mathrm{d}F},\dots,\frac{\mathrm{d}F_n}{\mathrm{d}F}\right)\mathrm{d}F. \tag{2.22}$$

(2.21) and (2.22) together show that

$$\mathbb{E}^F\left[f\left(\frac{\mathrm{d}F_1}{\mathrm{d}F},\dots,\frac{\mathrm{d}F_n}{\mathrm{d}F}\right)\right] = \lim_{m\to\infty}\mathbb{E}^{F_{\varepsilon_m}}\left[f\left(\frac{\mathrm{d}F_{1,\varepsilon_m}}{\mathrm{d}F_{\varepsilon_m}},\dots,\frac{\mathrm{d}F_{n,\varepsilon_m}}{\mathrm{d}F_{\varepsilon_m}}\right)\right] \leqslant \mathbb{E}^Q\left[f\left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q},\dots,\frac{\mathrm{d}Q_n}{\mathrm{d}Q}\right)\right].$$

Sufficiency. Assume that $(F_1,\dots,F_n)\prec_{\mathrm{h}}(Q_1,\dots,Q_n)$. By Lemma 2.3.5, this means that

$$\left(\frac{\mathrm{d}F_1}{\mathrm{d}F},\dots,\frac{\mathrm{d}F_n}{\mathrm{d}F}\right)\bigg|_F \prec_{\mathrm{cx}} \left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q},\dots,\frac{\mathrm{d}Q_n}{\mathrm{d}Q}\right)\bigg|_Q$$

holds for $F=\frac{1}{n}\sum_{i=1}^n F_i$ and $Q=\frac{1}{n}\sum_{i=1}^n Q_i$.

By Lemma 2.3.2, there exists a probability space $(\Omega',\mathcal{A}',Q')$ and random vectors $\mathbf{Y}'=(Y_1',\dots,Y_n'),\mathbf{Z}'=(Z_1',\dots,Z_n')$ defined on $(\Omega',\mathcal{A}',Q')$, such that

$$(Y_1',\dots,Y_n')\overset{\mathrm{d}}{=}\left(\frac{\mathrm{d}Q_1}{\mathrm{d}Q},\dots,\frac{\mathrm{d}Q_n}{\mathrm{d}Q}\right)=:\mathbf{Y}=(Y_1,\dots,Y_n),$$

$$(Z_1',\dots,Z_n')\overset{\mathrm{d}}{=}\left(\frac{\mathrm{d}F_1}{\mathrm{d}F},\dots,\frac{\mathrm{d}F_n}{\mathrm{d}F}\right)=:\mathbf{Z}=(Z_1,\dots,Z_n),$$

and

$$\mathbb{E}^{Q'}[Y_i'|Z_i']=Z_i',\quad i=1,\dots,n.$$

Given $m=0,1,\dots$, define random vector $\mathbf{Y}_m=(Y_{m,1},\dots,Y_{m,n})$ by

$$Y_{m,i}=\begin{cases}0 & \text{if } Y_i=0\\ \exp(2^{-m}\lfloor 2^m\log(Y_i)\rfloor) & \text{otherwise}\end{cases}$$

for $i=1,\dots,n$. Similarly we define $\mathbf{Y}_m'$, $\mathbf{Z}_m$ and $\mathbf{Z}_m'$ for $\mathbf{Y}'$, $\mathbf{Z}$ and $\mathbf{Z}'$, respectively. Note that

$$\mathbb{E}^{Q'}\left[Y_{m,i}'|Z_{m,i}'\right]\in\left[\exp(-2^{-m})\mathbb{E}^{Q'}[Y_i'|Z_{m,i}'],\mathbb{E}^{Q'}[Y_i'|Z_{m,i}']\right]\subseteq\left[\exp(-2^{-m})Z_{m,i}',\exp(2^{-m})Z_{m,i}'\right]$$

for $i=1,\dots,n$.

Each of $Q_1, \dots, Q_n$ is atomless, and so is Q. As a result, we can divide Ω into disjoint sets $A^m_{k,j}$, $k = (k_1, \dots, k_n) \in (\mathbb{Z} \cup \{-\infty\})^n$, $j = (j_1, \dots, j_n) \in (\mathbb{Z} \cup \{-\infty\})^n$, such that $Y_{m,i}(\omega) = \exp(k_i 2^{-m})$ for $\omega \in A^m_{k,j}$ and $i = 1, \dots, n$, $Q(A^m_{k,j}) = Q'(Y'_{m,i} = \exp(k_i 2^{-m}), Z'_{m,i} = \exp(j_i 2^{-m}), i = 1, \dots, n)$. Here we follow the tradition that $\exp(-\infty) = 0$ for ease of notation. Define random vector $\mathbf{Z}''_m$ on $(\Omega, \mathcal{A}, Q)$ by $Z''_{m,i}(\omega) = \exp(j_i 2^{-m})$ for $\omega \in A^m_{k,j}$, then $(\mathbf{Y}_m, \mathbf{Z}''_m)|_Q \overset{\mathrm{d}}{=} (\mathbf{Y}'_m, \mathbf{Z}'_m)|_{Q'}$.

Let I_d be the identity random variable on $(R, \mathcal{B}(\mathbb{R}))$. For $l = 0, 1, \dots$ and $h \in \mathbb{Z}$, denote by $\varphi^m_{l,h}(z)$ the conditional probability under F that $I_d \in [h2^{-l}, (h+1)2^{-l})$ given $Z_m = z$:

$$\varphi^m_{l,h}(z) = F(I_d \in [h2^{-l}, (h+1)2^{-l})|Z_m = z).$$

Then for any $l = 0, 1, \dots,$ $A^m_{k,j}$ can be further divided into disjoint subsets $A^m_{k,j,l,h}$, such that $Q(A^m_{k,j,l,h}) = Q(A^m_{k,j})\varphi^m_{l,h}(\exp(j2^{-m}))$. Moreover, the partitions can be made such that $\{A^m_{k,j,l',h}\}_{h \in \mathbb{Z}}$ is a refinement of $\{A^m_{k,j,l,h}\}_{h \in \mathbb{Z}}$ for any $l' > l$ and any given m, k, j. Define $X_{m,l}(\omega) = h2^{-l}$ for $\omega \in A^m_{k,j,l,h}$, and $X_m = \lim_{l \to \infty} X_{m,l}$. The limit exists since it is easy to check that $X_{m,l}$ is increasing with respect to l. Note that $X_{m,l}$ is conditionally independent of $\mathbf{Y}_m$ given $\mathbf{Z}''_m$, hence X_m is also conditionally independent of $\mathbf{Y}_m$ given $\mathbf{Z}''_m$.

By construction, for any $A \in \mathbb{R}^n$, $l = 0, 1, \dots,$ and $h \in \mathbb{Z}$,

$$\begin{aligned}
& Q(\mathbf{Z}''_m \in A, X_{m,l'} \in [h2^{-l}, (h+1)2^{-l})) \\
&= Q(\mathbf{Z}''_m \in A, X_{m,l} = h2^{-l}) \\
&= \sum_{\substack{k \\ j:\exp(j2^{-m}) \in A}} Q(A^m_{k,j,l,h}) \\
&= \sum_{\substack{k \\ j:\exp(j2^{-m}) \in A}} Q(A^m_{k,j})\varphi^m_{l,h}(\exp(j2^{-m})) \\
&= \sum_{j:\exp(j2^{-m}) \in A} Q(\mathbf{Z}''_m = \exp(j2^{-m}))\varphi^m_{l,h}(\exp(j2^{-m})) \\
&= \sum_{j:\exp(j2^{-m}) \in A} F(\mathbf{Z}_m = \exp(j2^{-m}))F([h2^{-l}, (h+1)2^{-l})|\mathbf{Z}_m = \exp(j2^{-m})) \\
&= F(\mathbf{Z}_m^{-1}(A) \cap [h2^{-l}, (h+1)2^{-l}))
\end{aligned} \tag{2.23}$$

for all $l' \geqslant l$. Thus, $\mathbf{Z}_m$, restricted on interval $[h2^{-l},(h+1)2^{-l})$, has the same distribution as $\mathbf{Z}''_m$, restricted on set $X^{-1}_{m,l'}([h2^{-l},(h+1)2^{-l}))$. Note that $X^{-1}_{m,l'}([h2^{-l},(h+1)2^{-l}))$ is the same set for any $l' \geqslant l$, hence $\mathbf{Z}_m$ restricted on interval $[h2^{-l},(h+1)2^{-l})$ also has the same distribution as $\mathbf{Z}''_m$ restricted on $X^{-1}_m([h2^{-l},(h+1)2^{-l}))$ for all $m = 0,1,\dots$. Since $\{[h2^{-l},(h+1)2^{-l})\}_{h\in\mathbb{Z},l=0,1,\dots}$ forms a basis for $\mathcal{B}(\mathbb{R})$, $\mathbf{Z}_m$ restricted on any Borel set B has the same distribution as $\mathbf{Z}''_m$ restricted on $X^{-1}_m(B)$. Therefore we conclude that $\mathbf{Z}''_m = \mathbf{Z}_m \circ X_m$ Q-almost surely. Moreover, by taking $A = \mathbb{R}^n$ in (2.23), it follows that $Q(X_{m,l'} \in [h2^{-l},(h+1)2^{-l})) = F([h2^{-l},(h+1)2^{-l}))$ for all $l' \geqslant l$. A similar reasoning as above then shows that $F = Q \circ X^{-1}_m$.

For any $A \in \mathcal{B}$ and any $i = 1,\dots,n$,

$$Q_i(X_m \in A) = \int_{X_m^{-1}(A)} Y_i \mathrm{d}Q. \tag{2.24}$$

It is easy to see that

$$\int_{X_m^{-1}(A)} Y_{m,i}\mathrm{d}Q \leqslant \int_{X_m^{-1}(A)} Y_i\mathrm{d}Q \leqslant \exp(2^{-m}) \int_{X_m^{-1}(A)} Y_{m,i}\mathrm{d}Q. \tag{2.25}$$

Moreover,

$$\begin{aligned}
&\int_{X_m^{-1}(A)} Y_{m,i}\mathrm{d}Q \\
&= \sum_{k,j} e^{k_i 2^{-m}} Q\left(\mathbf{Y}_m = e^{k2^{-m}}, \mathbf{Z}''_m = e^{j2^{-m}}\right) Q\left(X_m \in A \middle| \mathbf{Y}_m = e^{k2^{-m}}, \mathbf{Z}''_m = e^{j2^{-m}}\right) \\
&= \sum_j Q\left(X_m \in A \middle| \mathbf{Z}''_m = e^{j2^{-m}}\right) \sum_k e^{k_i 2^{-m}} Q\left(\mathbf{Y}_m = e^{k2^{-m}}, \mathbf{Z}''_m = e^{j2^{-m}}\right) \\
&= \sum_j Q\left(X_m \in A \middle| \mathbf{Z}''_m = e^{j2^{-m}}\right) Q\left(\mathbf{Z}''_m = e^{j2^{-m}}\right) \mathbb{E}^Q[Y_{m,i}|\mathbf{Z}''_m = e^{j2^{-m}}] \\
&= \sum_j Q\left(X_m \in A \middle| \mathbf{Z}''_m = e^{j2^{-m}}\right) Q\left(\mathbf{Z}''_m = e^{j2^{-m}}\right) \mathbb{E}^{Q'}[Y'_{m,i}|\mathbf{Z}'_m = e^{j2^{-m}}] \\
&\geqslant \sum_j Q\left(X_m \in A \middle| \mathbf{Z}''_m = e^{j2^{-m}}\right) Q\left(\mathbf{Z}''_m = e^{j2^{-m}}\right) \exp\left(j_i 2^{-m} - 2^{-m}\right) \\
&= \sum_j F\left(A \middle| \mathbf{Z}_m = e^{j2^{-m}}\right) F\left(\mathbf{Z}_m = e^{j2^{-m}}\right) \exp\left(j_i 2^{-m} - 2^{-m}\right)
\end{aligned}$$

$$\begin{aligned}
&\geqslant \exp\left(-2^{-m}\right) \sum_j \exp\left(j_i 2^{-m}\right) F\left(A \cap \{\mathbf{Z}_m = e^{j2^{-m}}\}\right) \\
&= \exp\left(-2^{-m}\right) \int_A Z_{m,i} \mathrm{d}F \\
&\geqslant \exp\left(-2^{-m+1}\right) \int_A Z_i \mathrm{d}F \\
&= \exp\left(-2^{-m+1}\right) F_i(A),
\end{aligned}$$

where the second equality holds since X_m is independent of $\mathbf{Y}_m$ given $\mathbf{Z}''_m$, and the fifth equality holds because $Q \circ X_m^{-1} = F$ and $\mathbf{Z}_m \circ X_m = \mathbf{Z}''_m$. Symmetrically,

$$\int_{X_m^{-1}(A)} Y_{m,i} \mathrm{d}Q \leqslant \exp(2^{-m}) F_i(A). \tag{2.26}$$

Combining (2.24)-(2.26), we have

$$Q_i(X_m \in A) \in [\exp(-2^{-m+1}) F_i(A), \exp(2^{-m+1}) F_i(A)].$$

Since this holds for any $A \in \mathcal{B}(\mathbb{B})$, we conclude that $Q_i \circ X_m^{-1}$ is absolutely continuous with respect to F_i, and $\mathrm{d}Q_i \circ X_m^{-1}/\mathrm{d}F_i \in [\exp(-2^{-m+1}), \exp(2^{-m+1})]$. It is easy to see that $D_{\mathrm{KL}}(Q_i \circ X_m^{-1} || F_i)$ converges to 0 as $m \to \infty$. □

2.6.3 Details in Example 2.5.14

We only present details for the maximum values, as the case for the minimum values is analogous.

Note that by the Girsanov Theorem, $X = B_1 - 1$ is $\mathrm{N}(0,1)$ distributed under Q. Denote by $H_{P,Q}$ the distribution of $\frac{\mathrm{d}P}{\mathrm{d}Q}$ under Q, which is clearly a log-normal distribution with parameter $(-1/2, 1)$. Using Example 2.5.8, we can calculate the maximum value of $\mathbb{E}^P[Y]$, as

$$\begin{aligned}
\max_{Y \in L_G(Q)} \mathbb{E}^P[Y] = \int_0^1 H_{P,Q}^{-1}(u) q(u) \mathrm{d}u &= \int_0^1 \exp\left(q(u) - \frac{1}{2}\right) q(u) \mathrm{d}u \\
&= \int_{-\infty}^{\infty} \exp\left(s - \frac{1}{2}\right) s \mathrm{d}q^{-1}(s) \\
&= \int_{-\infty}^{\infty} \frac{s}{\sqrt{2\pi}} \exp\left(-\frac{(s-1)^2}{2}\right) \mathrm{d}s = 1.
\end{aligned}$$

Noting that $1 - B_1$ is $\mathrm{N}(0,1)$ distributed under Q, the above maximum value is attained by $Y^* = 1 - B_1$.

We proceed to calculate $\max_{Y \in L_G(Q)} \mathbb{E}^P[Y^2]$. Denote by G_0 the distribution of X^2 under Q, which is a chi-square distribution with 1 degree of freedom. Using Example 2.5.12, we have

$$\max_{Y \in L_G(Q)} \mathbb{E}^P[Y^2] = \int_0^1 H_{P,Q}^{-1}(u) G_0^{-1}(u) \mathrm{d}u = \int_0^1 \exp\left(q(u) - \frac{1}{2}\right)\left(q\left(\frac{1+u}{2}\right)\right)^2 \mathrm{d}u.$$

The numerical value of the above integral is 2.795.

Finally, we investigate $\max_{Y \in L_G(Q)} \mathrm{Var}^P(Y)$. Using Example 2.5.13, we have

$$\max_{Y \in L_G(Q)} \mathrm{Var}^P(Y) = \min_{x \in \mathbb{R}} \overline{R}_x(G),$$

where for $x \in \mathbb{R}$,

$$\overline{R}_x(G) = \max_{Y \in L_G(Q)} \mathbb{E}^P[(Y - x)^2].$$

Using Example 2.5.12 again, we have, for $x \in \mathbb{R}$,

$$\overline{R}_x(G) = \int_0^1 H_{P,Q}^{-1}(u) G_x^{-1}(u) \mathrm{d}u,$$

where G_x is the distribution of $(X - x)^2$ under Q. Clearly,

$$G_x(t) = \int_{-\sqrt{t}+x}^{\sqrt{t}+x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{s^2}{2}\right) \mathrm{d}s, \quad t \geqslant 0,$$

and hence G_x is symmetric in x. Moreover, it is easy to see that, for $x > 0$ and $t \geqslant 0$, $G_x(t)$ is decreasing in x. As a consequence, $G_0^{-1}(s) \leqslant G_x^{-1}(s)$ for $s \in [0,1]$ and $x \in \mathbb{R}$. Therefore, the minimum value of $\overline{R}_x(G)$ is attained by $x = 0$, namely

$$\max_{Y \in L_G(Q)} \mathrm{Var}^P(Y) = \overline{R}_0(G) = \max_{Y \in L_G(Q)} \mathbb{E}^P[Y^2].$$

Chapter 3

Random Locations of Periodic Stationary Processes

3.1 Introduction

Random locations of stationary processes have been studied for a long time, and various results exist for special random locations and processes. For example, the results regarding the hitting time for Ornstein-Uhlenbeck processes date back to Breiman's paper in 1967 (Breiman, 1967), with recent developments made by Leblanc et al. (Leblanc et al., 2000) and Alili et al. (Alili et al., 2005). Early discussions about the location of path supremum over an interval can be found in the work of Leadbetter et al (Leadbetter et al., 1983). The book by Lindgren Lindgren (2012) provides an excellent summary of general results in stationary processes.

Recently, properties of possible distributions of the location of the path supremum have been obtained, and the sufficiency of the properties was proven (Samorodnitsky and Shen, 2012, 2013b). In Samorodnitsky and Shen (2013a), Samorodnitsky and Shen proceeded to introduce a general type of random locations called intrinsic location functionals, including but also extending far beyond the random locations mentioned above. In Shen (2016), equivalent representations of intrinsic location functionals were established using partially ordered random sets and piecewise linear functions.

In this chapter, we study intrinsic location functionals of periodic stationary processes, and characterize all the possible distributions of these random locations. The periodic setting leads to new properties along with challenges, which are the focus of this chapter. The periodicity also adds a discrete flavor to the problem, which, surprisingly, suggests a link with other well-studied properties such as joint mixability (Wang and Wang, 2016).

The motivation of this work is twofold. From the general theoretical perspective, since the study of continuous-time stationary processes requires a differentiable manifold structure to apply analysis techniques as well as a group structure to define stationarity, the most general and natural framework under which the random locations of stationary processes can be considered is an Abelian Lie group. It is well known that any connected Abelian Lie group can be represented as the product of real lines and one-dimensional torus, *i.e.*, circles. In other words, the real line $\mathbb{R}$ and one-dimension circle S_1 are building blocks for connected Abelian Lie groups. Therefore, in order to understand the properties of random locations of stationary processes in the general setting, it is crucial to study their behaviors on $\mathbb{R}$ and S_1 first. While the case for $\mathbb{R}$ was done in Samorodnitsky and Shen (2013b), this chapter deals with the circular case, which is equivalent to imposing a periodic condition on the stationary processes over the real line.

A more specific motivation comes from a problem in the extension of the so-called "relatively stationary process". A relatively stationary process is, briefly speaking, a stochastic process only defined on a compact interval, the finite dimensional distribution of which is invariant under translation, as long as all the time indices in the distribution remain inside the interval. Parthasarathy and Varadhan (Parthasarathy and Varadhan, 1964) showed that a relatively stationary process can always be extended to a stationary process over the whole real line. A question to ask as the next step is when such an extension can be periodic. Equivalently, if the relatively stationary process is defined on an arc of a circle instead of the compact interval on the real line, can it always be extended to a stationary process over the circle? This chapter will provide an answer to this question.

The rest of the chapter is organized as follows. In Section 3.2, we introduce some notation and assumptions for intrinsic location functionals and stationary and ergodic processes. In Section 3.3, we show some general results on intrinsic location functionals of periodic stationary processes. Sufficient and necessary conditions are established to characterize the distributions of these random locations. The following two sections are devoted to two special types of intrinsic location functionals. In Section 3.4, the class of *invariant intrinsic location functionals* is studied. The density of any invariant intrinsic location functional has a uniform lower bound, and such a distribution can always be constructed via the location of the path supremum over the interval. In Section 3.5, we show that the density of a *first-time intrinsic location functional* is non-increasing, and establish a link between the structure of the set of first-time intrinsic locations' distributions and the joint mixability of some distributions.

3.2 Notation and preliminaries

Throughout the chapter, $\mathbf{X} = \{X(t),\ t \in \mathbb{R}\}$ will denote a periodic stationary process. Without loss of generality, assume $\mathbf{X}$ has period 1. Moreover, for simplicity, we assume the sample function $X(t)$ is continuous unless specified otherwise. Indeed, all the arguments in the following parts also work for $\mathbf{X}$ with càdlàg sample paths.

As mentioned in the Introduction, an equivalent description of a periodic stationary stochastic process is a stationary process on a circle. That is, consider $\{X(t),\ t \in \mathbb{R}\}$ as a process defined on S_1, where S_1 is a circle with perimeter 1.

Let H be a set of functions on $\mathbb{R}$ with period 1, and assume it is invariant under shifts. The latter means that for all $g \in H$ and $c \in \mathbb{R}$, the function $\theta_c g(x) := g(x + c)$, $x \in \mathbb{R}$ belongs to H. We equip H with its cylindrical σ-field. Let $\mathcal{I}$ be the set of all compact, non-degenerate intervals in $\mathbb{R}$: $\mathcal{I} = \{[a, b] : a < b,\ [a, b] \subset \mathbb{R}\}$. We first define intrinsic location functional, the primary object of this chapter.

Definition 3.2.1. (Samorodnitsky and Shen, 2013a) 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 mapping $L(\cdot, I) : H \to \mathbb{R} \cup \{\infty\}$ is measurable.

2. For every $g \in H$ and $I \in \mathcal{I}$, $L(g, I) \in I \cup \{\infty\}$.

3. (Shift compatibility) For every $g \in H$, $I \in \mathcal{I}$ and $c \in \mathbb{R}$,

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

where $I - c$ is the interval I shifted by $-c$, and by convention, $\infty + c = \infty$.

4. (Stability under restrictions) For every $g \in H$ and $I_1, I_2 \in \mathcal{I}$, $I_2 \subseteq I_1$, if $L(g, I_1) \in I_2$, then $L(g, I_2) = L(g, I_1)$.

5. (Consistency of existence) For every $g \in H$ and $I_1, I_2 \in \mathcal{I}$, $I_2 \subseteq I_1$, if $L(g, I_2) \neq \infty$, then $L(g, I_1) \neq \infty$.

All the conditions in Definition 3.2.1 being natural and general, the family of intrinsic location functionals is a very large family of random locations, including and extending far beyond the location of the path supremum/infimum, the first/last hitting times, the location of the first/largest jump, *etc.*

Remark 3.2.2. ∞ is added to the range of the intrinsic location functionals to deal with the issue that some intrinsic location functionals may not be well defined for certain paths in some intervals. The σ-field on $\mathbb{R} \cup \{\infty\}$ is then given by treating $\{\infty\}$ as a separate point and taking the σ-field generated by the Borel sets in $\mathbb{R}$ and $\{\infty\}$.

It turns out that with the presence of a period, the relation between stationary processes and ergodic processes plays a crucial role in analyzing the distributions of the random locations. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Recall that a measurable function f is called T-*invariant* for a measurable mapping $T : \Omega \to \Omega$, if

$$f(T\omega) = f(\omega) \quad P\text{-almost surely.}$$

For a stationary process $\mathbf{X} = \{X(t),\ t \in \mathbb{R}\}$, let $\tilde{\Omega}$ be its canonical space equipped with the cylindrical σ-field $\tilde{\mathcal{F}}$, and θ_t be the shift operator as defined earlier. That is,

$$\theta_t \tilde{\omega}(s) = \tilde{w}(s + t),\ \text{for } \tilde{\omega} \in \tilde{\Omega}.$$

Denote by $P_{\mathbf{X}}(\cdot) = P(\mathbf{X} \in \cdot)$ the distribution of $\mathbf{X}$ on $(\tilde{\Omega}, \tilde{\mathcal{F}})$. A stationary process $\{X(t),\ t \in \mathbb{R}\}$ is called ergodic, if each measurable function f defined on $(\tilde{\Omega}, \tilde{\mathcal{F}})$ which is θ_t-invariant for every t is constant $P_{\mathbf{X}}$-almost surely.

It is known that the set of the laws of all stationary processes is a convex set and the extreme points of this set are the laws of the ergodic processes. Thus, we have the ergodic decomposition for stationary processes:

Theorem 3.2.3. *(Theorem A.1.1, Kifer (Kifer, 1988)) Let $\mathcal{M}$ be the space of all stationary probability measures, and $\mathcal{M}_e$ the subset of $\mathcal{M}$ consisting of all ergodic probability measures. Equip $\mathcal{M}$ and $\mathcal{M}_e$ with the natural σ-field: $\sigma(\mu \to \mu(A) : A \in \mathcal{F})$. For any stationary probability measure $\mu_{\mathbf{X}} \in \mathcal{M}$, there exists a probability measure λ on $\mathcal{M}_e$ such that*

$$\mu_{\mathbf{X}} = \int_{\rho \in \mathcal{M}_e} \rho \mathrm{d}\lambda.$$

The following proposition shows that for periodic stationary processes, ergodicity simply means that all the paths are the same up to translation. This simple fact will be used later in showing the main results of this chapter.

We say a probability space $(\Omega, \mathcal{F}, P)$ can be extended to a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$, if there exists a measurable mapping π from $(\tilde{\Omega}, \tilde{\mathcal{F}})$ to $(\Omega, \mathcal{F})$ satisfying $P = \tilde{P} \circ \pi^{-1}$. In this case, the process $\tilde{\mathbf{X}}$ defined on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ by $\tilde{\mathbf{X}}(\tilde{\omega}) = \mathbf{X}(\pi(\tilde{\omega}))$ will be identified with the original process $\mathbf{X}$.

Proposition 3.2.4. *For any continuous periodic ergodic process $\mathbf{X}$ with period 1, there exists a deterministic function g with period 1, such that $X(t) = g(t + \tilde{U})$ for $t \in \mathbb{R}$ almost surely on an extended probability space, in which $\tilde{U}$ follows a uniform distribution on $[0, 1]$.*

Proof. Let $C_1(\mathbb{R})$ be the space of continuous functions with period 1. For $h \geqslant 0$, define set $B_h := \{g \in C_1(\mathbb{R}) : \sup_{t \in \mathbb{R}} |g(t)| \leqslant h\}$. Note that B_h is in the invariant σ-algebra, and hence by ergodicity, $P(\mathbf{X} \in B_h)$ is either 0 or 1 for any h. Consequently, there exists h_0 (depending on $\mathbf{X}$) such that $P(\mathbf{X} \in B_{h_0}) = 1$.

Similarly, for function $\delta : [0, \infty) \to [0, \infty)$, define set

$$C_\delta := \{g \in C_1(\mathbb{R}) : |g(x) - g(y)| < \varepsilon \text{ for any } \varepsilon > 0 \text{ and all } |x - y| < \delta(\varepsilon)\},$$

then C_δ is in the invariant σ-algebra, $P(\mathbf{X} \in C_\delta) \in \{0,1\}$, and there exists function δ_0 such that $P(\mathbf{X} \in C_{\delta_0}) = 1$.

Furthermore, for any n, $\mathbf{t} = (t_1, ..., t_n)$ and $\mathbf{A} = (A_1, ..., A_n)$, where $t_1 < t_2 < \cdots < t_n$ and $A_1, ..., A_n$ are non-degenerate closed intervals, define sets

$$H_{\mathbf{t},\mathbf{A}} := \{g \in C_1(\mathbb{R}) : g(t_1) \in A_1, \ldots, g(t_n) \in A_n\}$$

and

$$H^0_{\mathbf{t},\mathbf{A}} := \{g \in C(\mathbb{R}) : \text{ there exists a constant } c,\ \theta_c g \in H_{\mathbf{t},\mathbf{A}}\}.$$

Again, $H^0_{\mathbf{t},\mathbf{A}}$ is in the invariant σ-algebra, and hence by ergodicity $P(\mathbf{X} \in H^0_{\mathbf{t},\mathbf{A}})$ is either 0 or 1 for any n, $t_1, ..., t_n$ and $A_1, ..., A_n$.

For $m = 0, 1, ...$, let $n_m = 2^m$ and $t_i^m = (i-1)2^{-m}$ for $i = 1, ..., n_m$. Then there exists $A_1^m, ..., A_{n_m}^m$ of the form $A_i^m = [k_i 2^{-m}, (k_i+1)2^{-m}]$, $k_i \in \mathbb{Z}, i = 1, ..., n_m$, such that $P(\mathbf{X} \in H^0_{\mathbf{t}^m,\mathbf{A}^m}) = 1$, where $\mathbf{t}^m = (t_1^m, ..., t_{n_m}^m)$, $\mathbf{A}^m = (A_1^m, ..., A_{n_m}^m)$. Moreover, we can choose the sets such that $\{H^0_{\mathbf{t}^m,\mathbf{A}^m}\}_{m=0,1,...}$ form a decreasing sequence, *i.e.*, $H^0_{\mathbf{t}^{m_1},\mathbf{A}^{m_1}} \supseteq H^0_{\mathbf{t}^{m_2},\mathbf{A}^{m_2}}$ if $m_1 \leqslant m_2$.

Consider the sequence of sets $\{H^0_{\mathbf{t}^m,\mathbf{A}^m} \cap B_{h_0} \cap C_{\delta_0}\}_{m=0,1,...}$. Each set in this sequence is closed and consists of functions which are uniformly bounded and equicontinuous. By Arzelà-Ascoli Theorem and the fact that we are looking at functions with period 1, which can be 1-1 mapped to $\{g \in C([0,1]) : g(0) = g(1)\} \subset C([0,1])$, the sets in this sequence are compact. As a result, the intersection of all the sets is non-empty. Moreover, there exists a single deterministic function with period 1, denoted by g, such that for any f in the intersection, $f(t) = g(t+c)$ for some $c \in \mathbb{R}$. Indeed, assume this is not the case, i.e., there exists f_1, f_2 both in $H^0_{\mathbf{t}^m,\mathbf{A}^m} \cap B_{h_0} \cap C_{\delta_0}$ for all $m = 0, 1, ...$, yet $f_1 \neq \theta_c f_2$ for any c, then fundamental analysis shows that

$$\inf_{c\in\mathbb{R}} \sup_{i\in\mathbb{Z}} |f_1(i2^{-m}) - \theta_c f_2(i2^{-m})| \geqslant \frac{1}{2} \inf_{c\in\mathbb{R}} \sup_{t\in\mathbb{R}} |f_1(t) - \theta_c f_2(t)| > 0$$

for m large enough, hence f_1 and f_2 will eventually be separated by some $H^0_{\mathbf{t}^m,\mathbf{A}^m}$. Thus, we conclude that $X(t) = g(t+V)$ almost surely for some random variable V.

The last step is to show that there exists an extended probability space and a uniform $[0,1]$ random variable $\tilde{U}$ defined on that space, such that $X(t) = g(t+\tilde{U})$ almost surely.

First, suppose there exists a uniform $[0,1]$ random variable U in some probability space, then $\{X(t), t \in \mathbb{R}\} \stackrel{d}{=} \{g(t+U), t \in \mathbb{R}\}$. Indeed, since the equality is in the distributional sense, we can assume that U is independent of everything else by considering, for example, the product space of the original probability space and $[0,1]$ equipped with the Borel σ-field and the Lebesgue measure. Then by stationarity and ergodicity, we have

$$\begin{aligned}\{X(t), t \in \mathbb{R}\} &\stackrel{d}{=} \{X(t+U), t \in \mathbb{R}\}\\ &= \{g(t+V+U), t \in \mathbb{R}\}\\ &\stackrel{d}{=} \{g(t+U), t \in \mathbb{R}\}.\end{aligned}$$

Moreover, the mapping $h : [0,1] \to C([0,1])$ given by $h(x) = \{g(t+x), t \in [0,1]\}$ is continuous, hence measurable. (Note that the Borel σ-field and the cylindrical σ-field coincide on $C([0,1])$.) As a result, there exists an extended probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ with a uniform [0,1] random variable $\tilde{U}$ defined on that, such that $\{X(t), t \in \mathbb{R}\} = h(\tilde{U}) = \{g(t+\tilde{U}), t \in \mathbb{R}\}$ almost surely on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$.

$\square$

3.3 Distributions of intrinsic location functionals

In this section, we characterize (properties of) intrinsic location functionals of periodic stationary processes. For a compact interval $[a,b]$, denote the value of an intrinsic location functional L for the process $\mathbf{X}$ on that interval by $L(\mathbf{X}, [a,b])$. Since $\mathbf{X}$ is stationary and L is shift compatible, the distribution of $L - a$ depends solely on the length of the interval. Thus, we can focus on the intervals starting from 0, in which case $L(\mathbf{X}, [0,b])$ is abbreviated as $L(\mathbf{X}, b)$. Furthermore, with the 1-periodicity of $\mathbf{X}$, it turns out that the only interesting cases are those with $b \leqslant 1$. In the following we assume $b \leqslant 1$ throughout. The case where $b > 1$ will be briefly discussed in Remark 3.3.4, after the introduction of a representation result for intrinsic location functional.

Denote by $F^{\mathbf{X}}_{L,[a,b]}$ the law of $L(\mathbf{X}, [a,b])$. It is a probability measure supported on $[a,b] \cup \{\infty\}$.

It was shown in Samorodnitsky and Shen (2013a) that the distribution of an intrinsic location functional for any stationary process over the real line, not necessarily periodic, possesses a specific group of properties. Adding periodicity obviously will not change these results. Here we present a simplified version of the original theorem for succinctness.

Proposition 3.3.1. *Let L be an intrinsic location functional and $\{X(t),\ t \in \mathbb{R}\}$ a stationary process. The restriction of the law $F_{L,T}^{\mathbf{X}}$ to the interior $(0, T)$ of the interval is absolutely continuous. Moreover, there exists a càdlàg version of the density function, denoted by $f_{L,T}^{\mathbf{X}}$, which satisfies the following conditions:*
(a) The limits

$$f_{L,T}^{\mathbf{X}}(0+) = \lim_{t \downarrow 0} f_{L,T}^{\mathbf{X}}(t) \text{ and } f_{L,T}^{\mathbf{X}}(T-) = \lim_{t \uparrow T} f_{L,T}^{\mathbf{X}}(t) \tag{3.1}$$

exist.
(b)

$$\mathrm{TV}_{(t_1,t_2)}(f_{L,T}^{\mathbf{X}}) \leqslant f_{L,T}^{\mathbf{X}}(t_1) + f_{L,T}^{\mathbf{X}}(t_2)$$

for all $0 < t_1 < t_2 < T$, where

$$\mathrm{TV}_{(t_1,t_2)}(f_{L,T}^{\mathbf{X}}) = \sup \sum_{i=1}^{n-1} \left| f_{L,T}^{\mathbf{X}}(s_{i+1}) - f_{L,T}^{\mathbf{X}}(s_i) \right|$$

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

Note that we have $\int_0^T f_{L,T}^{\mathbf{X}}(s)\mathrm{d}s < 1$ if there exists a point mass at ∞ or at the boundaries 0 and T.

We call the condition (b) in Proposition 3.3.1 "Condition (TV)", or the "variation constraint", because it puts a constraint on the total variation of the density function. It is not difficult to show that Condition (TV) is equivalent to the following Condition (TV'):

There exists a sequence $\{t_n\}$, $t_n \downarrow 0$, such that

$$\mathrm{TV}_{(t_n,T-t_n)}(f) \leqslant f(t_n) + f(T - t_n),\ n \in \mathbb{N}.$$

The above general result about the distribution of the intrinsic location functionals for stationary processes over the real line is still valid for periodic stationary processes, and serves as a basis for further exploration. It is, however, not the focus of this chapter. For the rest of the chapter we will concentrate on the new properties introduced by the periodicity assumption, which do not hold in the general case.

For any intrinsic location functional L and $T \leqslant 1$, let $I_{L,T}$ be the set of probability distributions $F^{\mathbf{X}}_{L,T}$ for periodic stationary processes $\mathbf{X}$ with period 1 on $[0, T]$. Our goal is to understand the structure of the set $I_{L,T}$, and the conditions that the distributions in $I_{L,T}$ need to satisfy. To this end, note that since ergodic processes are extreme points of the set of stationary processes, the extreme points of the set $I_{L,T}$ can only be the distributions of L for periodic ergodic processes with period 1. The next proposition gives a list of properties for these distributions.

Proposition 3.3.2. *Let L be an intrinsic location functional, $\mathbf{X}$ be a periodic ergodic process with period 1, and $T \leqslant 1$. Then $F^{\mathbf{X}}_{L,T}$ and its càdlàg density function on $(0, T)$, denoted by f, satisfy:*

1. *f takes values in non-negative integers;*
2. *f satisfies the condition (TV);*
3. *If $F^{\mathbf{X}}_{L,T}[0, T] > 0$, and there does not exist $t \in (0, T)$ such that $F^{\mathbf{X}}_{L,T}[0, t] = 1$ or $F^{\mathbf{X}}_{L,T}[t, T] = 1$, then $f(t) \geqslant 1$ for all $t \in (0, T)$. If furthermore, $F^{\mathbf{X}}_{L,T}(\{\infty\}) > 0$, then $f - 1$ also satisfies the condition (TV).*

Note that the condition in the first part of property 3 can be translated into requiring either a positive but smaller than 1 mass at ∞, or a positive point mass or a positive limit of the density function at each of the two boundaries 0 and T.

The proof of Proposition 3.3.2 relies on the following representation result given in Shen (2016).

Proposition 3.3.3. *A mapping $L(g, I) : H \times \mathcal{I} \to \mathbb{R} \cup \{\infty\}$ is an intrinsic location functional if and only if*

1. *$L(\cdot, I)$ is measurable for $I \in \mathcal{I}$;*

2. *There exists a subset of $\mathbb{R}$ determined by g, denoted as $S(g)$, and a partial order $\preceq$ on it, satisfying:*

   (1) For any $c \in \mathbb{R}$, $S(g) = S(\theta_c g) + c$;

   (2) For any $c \in \mathbb{R}$ and $t_1, t_2 \in S(g)$, $t_1 \preceq t_2$ implies $t_1 - c \preceq t_2 - c$ in $S(\theta_c g)$,

such that for any $I \in \mathcal{I}$, either $S(g) \cap I = \emptyset$, in which case $L(g, I) = \infty$, or $L(g, I)$ is the unique maximal element in $S(g) \cap I$ according to $\preceq$.

Such a pair $(S, \preceq)$ in the above proposition is called a *partially ordered random set representation of L*. Intuitively, this representation result shows that a random location is an intrinsic location functional if and only if it always takes the location of the maximal element in a random set of points, according to some partial order. Both the random set and the order are determined by the path and are shift-invariant.

Remark 3.3.4. By Proposition 3.3.3, for a function g with period 1, $t \in S(g)$ implies $t + c \in S(\theta_{-c} g) = S(g)$ for any $c \in \mathbb{Z}$. Moreover, if $t + 1 \preceq t$, then $t + c_2 \preceq t + c_1$ for all $c_1, c_2 \in \mathbb{Z}, c_2 > c_1$. As a result, for an interval $[a, b]$ with length greater than 1, only the points in the leftmost cycle $[a, a+1)$ can have the maximal order. Thus, the location of the intrinsic location functional on $[a, b]$ will be the same as on $[a, a+1]$. Symmetrically, if $t \preceq t + 1$, then the location of the intrinsic location functional on $[a, b]$ will be the same as on $[b-1, b]$. Hence we only need to consider the intervals with length no larger than 1.

Proof of Proposition 3.3.2. Property 2 directly comes from Proposition 3.3.1. We only need to check properties 1 and 3.

Property 1. Since $\mathbf{X}$ is a periodic ergodic process with period 1, by Proposition 3.2.4, there exists a periodic deterministic function g with period 1 such that $X(t) = g(t + U)$ for $t \in \mathbb{R}$, where U follows a uniform distribution on $[0, 1]$. In other words, all the sample paths of $\mathbf{X}$ are the same up to translation. Let $(S, \preceq)$ be a partially ordered random set representation of L. For any $s \in S(g)$, define

$$a_s := \sup\{\Delta s \in \mathbb{R} : r \preceq s \text{ for all } r \in (s - \Delta s, s) \cap S(g)\},$$
$$b_s := \sup\{\Delta s \in \mathbb{R} : r \preceq s \text{ for all } r \in (s, s + \Delta s) \cap S(g)\},$$

and define $\sup \emptyset = \infty$ by convention. By a slight abuse of notation, we also use a_s and b_s to denote the same quantity for $s \in S(\mathbf{X})$. Intuitively, a_s and b_s are the largest distance by which we can go to the left and right of the point s without passing a point with higher order than s according to $\preceq$, respectively. Thus, for $0 < t < t + \Delta t < T$, we have

$$\begin{aligned} & P\left(\text{there exists}\, s \in [t, t+\Delta t] \cap S(\mathbf{X}) : a_s > t + \Delta t,\ b_s > T - t\right) \\ \leqslant{} & P\left(t \leqslant L(\mathbf{X}, (0,T)) \leqslant t + \Delta t\right) \\ \leqslant{} & P\left(\text{there exists}\, s \in [t, t+\Delta t] \cap S(\mathbf{X}) : a_s \geqslant t,\ b_s \geqslant T - t - \Delta t\right). \end{aligned} \tag{3.2}$$

Seeing that $X(t) = g(t + U)$, $S(\mathbf{X}) = S(g) - U$. By change of variable $s \to s - U$,

$$\begin{aligned} & P\left(\text{there exists}\, s \in [t, t+\Delta t] \cap S(\mathbf{X}) : a_s > t + \Delta t,\ b_s > T - t\right) \\ ={} & P\left(\text{there exists}\, s \in S(g) : a_s > t + \Delta t,\ b_s > T - t,\ s - U \in [t, t+\Delta t]\right). \end{aligned}$$

Note the values of a_s and b_s remain unchanged, since they are defined with respect to $\mathbf{X}$ on the left hand side, and with respect to g on the right hand side.

Since $S(g)$ has period 1, $s \in S(g)$ if and only if $s - \lfloor s \rfloor \in S(g) \cap [0, 1)$. Moreover, since $s - U$ and $s - \lfloor s \rfloor - U - \lfloor s - \lfloor s \rfloor - U \rfloor$ share the same fractional part and are both in $[0, 1)$, $s - U = s - \lfloor s \rfloor - U - \lfloor s - \lfloor s \rfloor - U \rfloor$. Thus, by another change of variable $s - \lfloor s \rfloor \to s$, we have

$$\begin{aligned} & P\left(\text{there exists}\, s \in S(g) : a_s > t + \Delta t,\ b_s > T - t,\ s - U \in [t, t+\Delta t]\right) \\ ={} & P\big(\text{there exists}\, s \in S(g) \cap [0, 1) \\ & \quad \text{such that}\, a_s > t + \Delta t,\ b_s > T - t,\ \text{and}\ s - U - \lfloor s - U \rfloor \in [t, t+\Delta t]\big). \end{aligned}$$

Therefore, for Δt small enough,

$$\begin{aligned} & P\left(\text{there exists}\, s \in [t, t+\Delta t] \cap S(\mathbf{X}) : a_s > t + \Delta t,\ b_s > T - t\right) \\ ={} & \left|\{s \in S(g) \cap [0, 1) : a_s > t + \Delta t,\ b_s > T - t\}\right| \cdot \Delta t, \end{aligned}$$

where $|A|$ denotes the cardinal of set A. Thus, we have

$$\begin{aligned} f(t) &= \lim_{\Delta t \to 0} \frac{P\left(t \leqslant L(\mathbf{X}, (0,T)) \leqslant t + \Delta t\right)}{\Delta t} \\ &\geqslant \left|\{s \in S(g) \cap [0, 1) : a_s > t,\ b_s > T - t\}\right|. \end{aligned} \tag{3.3}$$

Symmetrically,

$$
\begin{aligned}
f(t) &= \lim_{\Delta t \to 0} \frac{P\left(t \leqslant L(\mathbf{X},(0,T)) \leqslant t+\Delta t\right)}{\Delta t} \\
&\leqslant |\{s \in S(g) \cap [0,1) : a_s \geqslant t,\ b_s \geqslant T-t\}|. \qquad (3.4)
\end{aligned}
$$

Moreover, it is easy to see that the set $\Sigma := \{s \in S(g) \cap [0,1) : a_s > 0 \text{ and } b_s > 0\}$ is at most countable, then $\{t : a_s = t \text{ or } b_s = T-t \text{ for some } s \in \Sigma\}$ is also at most countable. Hence the density can be taken as the càdlàg modification of $|\{s \in S(g) \cap [0,1) : a_s \geqslant t,\ b_s \geqslant T-t\}|$, which only takes values in non-negative integers.

Property 3. Assume $F^{\mathbf{X}}_{L,T}[0,T] > 0$ and there does not exist $t \in (0,T)$, such that $F^{\mathbf{X}}_{L,T}[0,t] = 1$ or $F^{\mathbf{X}}_{L,T}[t,T] = 1$. There are two possible cases depending on whether $F^{\mathbf{X}}_{L,T}$ has a point mass at ∞.

First suppose $F^{\mathbf{X}}_{L,T}(\{\infty\}) \in (0,1)$. Then by the partially ordered random set representation, there exists an interval $[s_\infty, t_\infty]$ (depending on g) satisfying $t_\infty - s_\infty \geqslant T$, such that $S(g) \cap [s_\infty, t_\infty] = \emptyset$. Since g has period 1, $S(g) \cap [s_\infty + 1, t_\infty + 1] = \emptyset$ as well. Let $\tau = L(g, [t_\infty, s_\infty + 1])$. Since L is not identically ∞, such a finite τ must exist. Moreover note that there is no point of $S(g)$ in $[s_\infty, t_\infty]$ and $[s_\infty + 1, t_\infty + 1]$, hence τ is actually the maximal element in $S(g)$ according to $\preceq$ on the interval $[s_\infty, t_\infty + 1]$. Thus, $a_\tau > \tau - s_\infty = \tau - t_\infty + t_\infty - s_\infty \geqslant T$, and symmetrically $b_\tau \geqslant T$. Consequently, $\tau - \lfloor \tau \rfloor$ is in the set $\{s \in S(g) \cap [0,1) : a_s \geqslant t,\ b_s \geqslant T-t\}$ for all $t \in (0,T)$. Since the density function $f(t)$ can be taken as the càdlàg modification of $|\{s \in S(g) \cap [0,1) : a_s \geqslant t,\ b_s \geqslant T-t\}|$, $f(t) \geqslant 1$ for all $t \in (0,T)$.

For the second possibility, suppose now there is either a positive mass or a positive limit of the density function on each of the two boundaries 0 and T. Suppose for the purpose of contradiction that there exists a non-degenerate interval $[u, T-v]$ such that $f(t) = 0$ for all $t \in [u, T-v]$. For $t \in S(g)$, we distinguish four different types: $A := \{t \in S(g) : a_t \leqslant u,\ b_t > T-u-\varepsilon\}$, $B := \{t \in S(g) : a_t > T-v-\varepsilon,\ b_t \leqslant v\}$, $C := \{t \in S(g) : a_t > u,\ b_t > v,\ a_t + b_t > T\}$ and $D := \{t \in S(g) : a_t > u,\ b_t > v,\ a_t + b_t = T\}$, where $0 < \varepsilon < \frac{T-u-v}{2}$. Sets A, B, C and D are disjoint, and for any $t \in S(g)$ such that $t = L(g, I)$ for some interval I with length T, $t \in A \cup B \cup C \cup D$. By the assumption about f, it is easy to see that $A \neq \emptyset$, $B \neq \emptyset$ and $C = \emptyset$.

We claim that for any $x \in A$ and $y \in B$, if $x > y$, then $x - y > T$. Suppose it is not true. For interval $I = [t, t+T]$, where t satisfies $0 \leqslant y - t < T - v - \varepsilon$ and $0 \leqslant t + T - x < T - u - \varepsilon$, let z be the maximal element in $S(g) \cap I$ according to $\preceq$. Note that the choice of t guarantees that $x, y \in I$, hence $S(g) \cap I \neq \emptyset$, z always exists. Moreover, $x \preceq z$ and $y \preceq z$. Because $y \in B$, y is larger in $\preceq$ than any point to its left within a distance smaller than $T - v - \varepsilon$, which contains $[t, y]$. Thus, z cannot be in this part of the interval I. Similarly, z cannot be in $[x, t+T]$, hence $z \in [y, x]$. For such z,

$$a_z \geqslant a_y > T - v - \varepsilon > u, \quad b_z \geqslant b_x > T - u - \varepsilon > v,$$

and $a_z + b_z > T - v - \varepsilon + T - u - \varepsilon > T$, which means $z \in C$. However, $C = \emptyset$ by assumption. Therefore, for any $x \in A$, $y \in B$ and $x > y$, we have $x - y > T$.

On the other hand, we show in the following paragraphs that for any point $y \in B$, there exists another point $y' \in B$, such that $\frac{u}{2} < y' - y \leqslant T$. To this end, consider a number of intervals $[y - \varepsilon_i, y - \varepsilon_i + T]$ given any arbitrary point $y \in B$ and $\varepsilon_i = \frac{1}{2i} u$ for $i = 1, 2, \ldots$. Denote l_i as the maximal element in $[y - \varepsilon_i, y - \varepsilon_i + T] \cap S(g)$ according to $\preceq$. Notice that since $y \in S(g)$, l_i always exists. Seeing that $a_y > T - v - \varepsilon > u$, l_i must be in $[y, y+T]$. Since $l_i - y \leqslant T$, l_i must be in the set $B \cup D$.

Next, we show that there exists i such that $l_i \in B$. Suppose $l_i \in D$ for all i. If there exist $l_i = l_j \in D$ for some $i < j$, then l_i is the maximal element in both $[y - \varepsilon_i, y - \varepsilon_i + T] \cap S(g)$ and $[y - \varepsilon_j, y - \varepsilon_j + T] \cap S(g)$. As a result, we have $a_{l_i} \geqslant l_i - y + \varepsilon_i$, and $b_{l_i} \geqslant y - \varepsilon_j + T - l_i$. However, this leads to

$$a_{l_i} + b_{l_i} \geqslant T + \varepsilon_i - \varepsilon_j > T,$$

hence l_i cannot be in D. Thus, for any $i \neq j$, $l_i \neq l_j$. By the fact that $a_{l_i} > u$ and $b_{l_i} > v$, there are at most $\frac{T}{\min\{u,v\}}$ points in the set $D \cap [y, y+T]$, which contradicts the assumption that $l_i \in D \cap [y, y+T]$ for all $i = 1, 2, \ldots$. As a result, there always exists at least one point $l_i \in B$.

Furthermore, for such l_i, if $l_i - y \leqslant \frac{u}{2}$, then

$$b_{l_i} \geqslant T - \frac{u}{2} - \varepsilon_i \geqslant T - u > v,$$

which contradicts the fact that $l_i \in B$. Therefore for any $y \in B$, there always exists a point $y' = l_i \in B$, such that

$$\frac{u}{2} < y' - y \leqslant T.$$

As a result, for any periodic function g with period 1, there exists $y_1 \in B$ and then a sequence of points $\{y_i,\ i = 2, \ldots, k\}$ in B such that for $i = 1, \ldots, k-1$,

$$\frac{u}{2} < y_{i+1} - y_i \leqslant T,$$

and k is chosen such that

$$y_{k-1} < 1 + y_1 \leqslant y_k.$$

However, since g is a periodic function with period 1 and $A \neq \emptyset$, this means that there must exist some points $x \in A$ and $y \in B$ such that $x - y \leqslant T$, which contradicts the result we derived before. Therefore, we conclude that there does not exist a non-degenerate interval $[u, T - v]$ such that $f(t) = 0$ for all $t \in [u, T - v]$, if the condition in the first part of property 3 holds.

Finally we turn to the second part in property 3. Assume $F^{\mathbf{X}}_{L,T}(\{\infty\}) > 0$, then we show that $f - 1$ will satisfy the condition (TV). Recall that a positive probability at ∞ for $F^{\mathbf{X}}_{L,T}$ implies the existence of a maximal interval $[s_\infty, t_\infty]$ depending on g satisfying $t_\infty - s_\infty \geqslant T$ and $S(g) \cap [s_\infty, t_\infty] = \emptyset$. Indeed, the inequality $t_\infty - s_\infty \geqslant T$ can be strengthened to $t_\infty - s_\infty > T$, since otherwise its contribution to the point mass at ∞ will be 0, even though it allows one particular value of U such that $g(t + U) \cap [0, T] = \emptyset$. Consider an interval $[u, v] \subset (0, T)$, such that f is flat on $[u, v]$. Since f takes integer values and satisfies the variation constraint, such an interval always exists. Define

$$S'(g) = S(g) \cup \{s_\infty + v - \varepsilon + C : C \in \mathbb{Z}\} \cup \bigcup_{C \in \mathbb{Z}} (s_\infty + T + \varepsilon + C, t_\infty + C)$$

for ε small enough, and extend the order $\preceq$ to $S'(g)$ (still denoted by $\preceq$) by setting $s_\infty + v - \varepsilon + C \preceq t_1 \preceq t_2 \preceq t$ for any $C \in \mathbb{Z}$, $t_1, t_2 \in (s_\infty + T + \varepsilon + C, t_\infty + C)$, $t_1 < t_2$, and any $t \in S(g)$. Intuitively, the extended order assigns the minimal order to $s_\infty + v - \varepsilon$, then an increasing order to the points in $(s_\infty + T + \varepsilon, t_\infty)$, while keeping the order for the added points always inferior to the original points in $S(g)$, and is finally completed by a periodic extension to $\mathbb{R}$. Let L' be an intrinsic location functional having $(S'(g), \preceq)$ as its partially ordered random set representation, and denote by f' the density of $F^{\mathbf{X}}_{L',T}$. It is easy to see that $f' = f + \mathbb{I}_{(v-2\varepsilon, v-\varepsilon]}$. Hence for ε small enough and $t_n \downarrow 0$ with t_1 being small enough, $\mathrm{TV}_{(t_n, T-t_n)}(f') = \mathrm{TV}_{(t_n, T-t_n)}(f) + 2$ for any n. Since f' satisfies the condition (TV), we must have $\mathrm{TV}_{(t_n, T-t_n)}(f) + 2 \leqslant f(t_n) + f(T - t_n)$. Thus $\mathrm{TV}_{(t_n, T-t_n)}(f - 1) \leqslant (f(t_n) - 1) + (f(T - t_n) - 1)$, which is the variation constraint for $f - 1$. $\square$

With the properties of the distributions of L for periodic ergodic processes with period 1 at hand, we proceed to study the structure of $I_{L,T}$, the set of all distributions of L for periodic stationary processes. Denote by E_T the collection of probability distributions on $[0, T] \cup \{\infty\}$ satisfying the three properties listed in Proposition 3.3.2, and let $\mathcal{P}_T$ be the collection of all probability distributions on $[0, T] \cup \{\infty\}$ which are absolutely continuous on $(0, T)$. For the rest of the chapter, denote by $C(A)$ the convex hull generated by a set $A \subseteq \mathcal{P}_T$ under the weak topology.

Theorem 3.3.5. *$I_{L,T}$ is a convex subset of $\mathcal{P}_T$. Moreover, $I_{L,T} \subseteq C(E_T)$.*

Proof. The convexity of $I_{L,T}$ is obvious. If $F_1, F_2 \in I_{L,T}$, then there exist stationary processes with period 1, denoted by $\mathbf{X}_1, \mathbf{X}_2$, such that $F_1 = F^{\mathbf{X}_1}_{L,T}$ and $F_2 = F^{\mathbf{X}_2}_{L,T}$. For any $a \in [0, 1]$, $aF_1 + (1 - a)F_2 = F^{\mathbf{X}}_{L,T}$, where the process $\mathbf{X}$ is a mixture of $\mathbf{X}_1$ and $\mathbf{X}_2$, with weights a and $1 - a$, respectively.

Next we show $I_{L,T} \subseteq C(E_T)$. By ergodic decomposition, any $F \in I_{L,T}$ can be written as $F = \int_{G \in E_T} G \mathrm{d}\lambda$, where λ is a probability measure on E_T. The integration holds in the sense of mixture of probability measures, i.e.,

$$\int_{x \in [0,T] \cup \{\infty\}} h(x) \mathrm{d}F(x) = \int_{G \in E_T} \int_{x \in [0,T] \cup \{\infty\}} h(x) \mathrm{d}G(x) \mathrm{d}\lambda$$

for all bounded and continuous function h defined on $[0, T] \cup \{\infty\}$. Since the set of probability measures on $[0, T] \cup \{\infty\}$ equipped with the weak topology is separable, we conclude that $F \in C(E_T)$. $\square$

The converse of Theorem 3.3.5, that for an arbitrarily given intrinsic location functional L and any distribution $F \in C(E_T)$ there exists a periodic stationary process $\mathbf{X}$ such that $F = F^{\mathbf{X}}_{L,T}$, is not true in general. For example, it can be easily checked that $L(g, I = [a, b]) := a$ is an intrinsic location functional. Yet the only possible distribution for L on $[0, T]$ is a Dirac measure on the boundary 0. However, the next result shows that the converse does hold if we do not focus on any particular L, but collect the possible distributions for all the intrinsic locations functionals. In other words, any member in $C(E_T)$ can be the distribution of some intrinsic location functional on $[0, T]$ and some periodic stationary process with period 1. More formally, define $I_T = \bigcup_L I_{L,T}$ to be the set of all possible distributions of intrinsic location functionals on $[0, T]$, then $I_T = C(E_T)$. Here and throughout the chapter, when we discuss the existence of a stochastic process without specifying the underlying probability space, the existence should be understood as that of the process together with the existence of a probability space on which the process is defined.

Theorem 3.3.6. *For any $F \in C(E_T)$, there exist an intrinsic location functional and a periodic stationary process with period* 1*, such that F is the distribution of this intrinsic location for such process on* $[0, T]$*.*

The proof of Theorem 3.3.6 consists of three parts. The main steps of the proof are presented in Part I below. Parts II and III are put in Sections 3.4 and 3.5, respectively, due to the explicit construction required for specific types of intrinsic location functionals.

Proof of Theorem 3.3.6, Part I. We define an intrinsic location functional $L = L(g, I)$ as

$$L(g, I) = \begin{cases} L_1(g, I) & \text{if } g(t) \geqslant 0 \text{ for all } t \in \mathbb{R}, \\ L_2(g, I) & \text{if there exists } t \in \mathbb{R} \text{ such that } g(t) = -1, \\ L_3(g, I) & \text{otherwise,} \end{cases}$$

where

$$L_1(g, I) = \inf \left\{ t \in I : g(t) = \sup_{s \in I} g(s), g(t) \geqslant \frac{1}{2} \right\},$$
$$L_2(g, I) = \inf\{t \in I : g(t) = -1\},$$

and

$$L_3(g,I) = \sup\{t \in I : g(t) = -2\}.$$

Intuitively, L_1 is based on the location of the path supremum, but truncated at level $\frac{1}{2}$. L_2 and L_3 are first and last hitting times, respectively.

We first show that such L is an intrinsic location functional, by using the partially ordered random set representation of intrinsic location functionals. It is not difficult to verify that L_1, L_2 and L_3 are all intrinsic location functionals, and hence they all have their own partially ordered random set representations, denoted as $(S_1(g), \preceq_1)$, $(S_2(g), \preceq_2)$ and $(S_3(g), \preceq_3)$. For positive sample paths, L has $(S_1, \preceq_1)$ as its partially ordered random representation; otherwise for sample paths reaching level -1, L has $(S_2, \preceq_2)$; otherwise, L has $(S_3, \preceq_3)$. Combining the three cases gives a complete partially ordered random set representation for L. Thus, L is an intrinsic location functional.

Next, we need to show that for any $F \in E_T$, there exists a periodic ergodic process with period 1 such that F is the distribution of L over $[0,T]$ for such process. For any $F \in E_T$, let f be its density function on $(0,T)$. We discuss two possible scenarios depending on whether $f(t) \geqslant 1$ for all t or not.

1. If $f(t) \geqslant 1$ for all $t \in (0,T)$, we are going to show that there exists a periodic ergodic process with period 1 and positive sample paths, such that F is the distribution of L_1 on $[0,T]$ for that process. Since L_1 is a modified version of the location of the path supremum, this part of the proof is postponed and will be resumed right after the proof of Theorem 3.4.7, in which we focus on the distribution of the location of the path supremum.

2. Otherwise, $f(t) = 0$ for some t. Recall from the definition of E_T that if $f(0+) \geqslant 1$ and $f(T-) \geqslant 1$, then $f(t) \geqslant 1$ for all $t \in (0,T)$. Hence in this case we must have $f(0+) = 0$ or $f(T-) = 0$. Assume $f(T-) = 0$ for example. Take $u := \inf\{t \in (0,T) : f(t) = 0\}$ and a sequence $\{t_n \in (u,T)\}_{n\in\mathbb{N}}$ such that $t_n \uparrow T$ as $n \to \infty$ and $f(t_n) = 0$ for all n. The variation constraint applied to the intervals $(0,u)$ and (u,t_n) implies that f is non-increasing in $(0,u)$ and that $f(t) = 0$ for $f \in [u,T)$, respectively. Symmetric results hold for the case where $f(0+) = 0$. To summarize, if f is the density function for a distribution in E_T and $f(t) = 0$ for some t, we have

(1) f takes values in non-negative integers;

(2) Either there exists $u \in (0, T)$ such that f is a non-increasing function in the interval $(0, u)$ and $f(t) = 0$ for $t \in [u, T)$, or there exists $v \in (0, T)$ such that f is a non-decreasing function in the interval $[v, T)$ and $f(t) = 0$ for $t \in (0, v)$.

By symmetry, we only prove the case where f is non-increasing in the interval $(0, u)$ and $f(t) = 0$ for $t \in [u, T)$. Since the intrinsic location functional that we are going to use in this case, L_2, is a first hitting time, this part of the proof is postponed and will be resumed right after the proof of Proposition 3.5.4, which deals with this type of intrinsic location functionals. □

Remark 3.3.7. The proof of Theorem 3.3.6 actually implies a stronger result: all the distributions in $C(E_T)$ can be generated by a single intrinsic location functional, which is the location L defined in the proof of the theorem.

Remark 3.3.8. Among the three conditions defining the set E_T, the condition (TV) is stable under convex combination, while the other two, integer values and a lower bound at level 1 under some conditions, are not. Therefore when passing from ergodic processes to stationary processes, these two conditions will not persist. However, this does not mean that they will simply disappear. They still affect the structure of the set of all possible distributions $I_T = C(E_T)$, but in a complicated way. While an explicit, analytical description of I_T is not known, we point out in the following example that I_T is indeed a proper subset of the set of all distributions solely satisfying condition (TV).

Denote by A_T the class of probability distributions on $[0, T] \cup \{\infty\}$ with densities satisfying the variation constraint (TV). Let $T = 1$ and consider a probability distribution F with density function

$$f(t) = \begin{cases} \frac{4}{3}, & t \in (0, \frac{3}{4}), \\ 0, & t \in [\frac{3}{4}, 1). \end{cases}$$

From the construction of f, it is easy to check that $F \in A_T$. Suppose F is also in the set I_T, then it can be written as an integral of the elements in the set E_T with respect to a probability measure on E_T, as discussed in the proof of Theorem 3.3.5. Since $f(t) = 0$ for all $t \in [\frac{3}{4}, 1)$, the variation constraint implies that any candidate density g to construct f must be non-increasing on the interval $(0, \frac{3}{4})$ and $g(t) = 0$ for all $t \in [\frac{3}{4}, 1)$. Moreover, g takes integer values, so there exists g such that $g(t) = 2$ for $t \in (0, \frac{3}{4})$. However, the integral of g is

$$\int_0^T g(t)\mathrm{d}t = \frac{3}{2} > 1,$$

which means that there does not exist a distribution in E_T such that g is its density function. Therefore, $F \notin C(E_T)$, hence I_T is a proper subset of A_T.

3.4 Invariant intrinsic location functionals

In this section, we consider a special type of intrinsic location functionals, referred to as the *invariant intrinsic location functionals*.

Definition 3.4.1. An intrinsic location functional L is called *invariant*, if it satisfies

1. $L(g, I) \neq \infty$ for any compact interval I and $g \in H$.
2. $L(g, [0, 1]) = L(g, [a, a+1]) \mod 1$, for any $a \in \mathbb{R}$ and $g \in H$.

Remark 3.4.2. Invariance is a natural requirement for an intrinsic location functional on S_1. The projection of an interval with length of 1 in S_1 forms a loop, with the starting and ending points being mapped to the same point. The above definition then requires that the location over the whole circle is always well-defined, and does not depend on the location of the starting/ending point.

Example 3.4.3. It is easy to see that the location of the path supremum

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

is an invariant intrinsic location functional, provided that the path supremum is uniquely achieved.

Besides the location of the path supremum, other invariant intrinsic location functionals include the location of the point with the largest/smallest slope (if the sample paths are in C^1), the location of the point with the largest/smallest curvature (if the sample paths are in C^2), *etc*, provided the uniqueness of these locations. The related criteria for uniqueness often go back to checking the uniqueness of the path supremum/infimum in one period. Indeed, if the a periodic stationary process has sample paths in C^1 (resp. C^2), then its first (resp. second) derivative is again a periodic stationary process. For a Gaussian process $\mathbf{X}$, its derivative $\mathbf{X}'$ is still Gaussian, and Kim and Pollard (Kim and Pollard, 1990) showed that the supremum is almost surely achieved at a unique point if $\mathrm{Var}(X'(s), X'(t)) \neq 0$ for $s \neq t$. In our periodic case, this means that the process has no period smaller than 1. Another condition was developed by Pimentel (Pimentel, 2014) for general processes with continuous sample paths.

For an invariant intrinsic location functional, we have the following lower bound for its density function.

Proposition 3.4.4. *For $T \in (0, 1]$, any invariant intrinsic location functional L and any periodic stationary process $\mathbf{X}$ with period 1, the density $f^{\mathbf{X}}_{L,T}$ of L on $(0, T)$ satisfies*

$$f^{\mathbf{X}}_{L,T}(t) \geqslant 1 \quad \textit{for all } t \in (0, T). \tag{3.5}$$

Proof. Let $0 < a < b < 1$. Since $\mathbf{X}$ is stationary, we have

$$P(L(\mathbf{X}, [0, 1]) \in (0, b - a)) = P(L(\mathbf{X}, [a, a + 1]) \in (a, b)). \tag{3.6}$$

By the assumption of invariant intrinsic location functionals, for any $a \in \mathbb{R}$,

$$L(\mathbf{X}, [0, 1]) = L(\mathbf{X}, [a, a + 1]) \mod 1.$$

Then

$$\begin{aligned} P(L(\mathbf{X}, [0, 1]) \in (0, b - a)) &= P(L(\mathbf{X}, [a, a + 1]) \in (a, b)) \\ &= P(L(\mathbf{X}, [0, 1]) \in (a, b)). \end{aligned}$$

It means that $L(\mathbf{X}, [0,1])$ follows a uniform distribution on the interval $[0,1]$. Thus, for any $t \in (0,1)$,

$$f^{\mathbf{X}}_{L,[0,1]}(t) = 1.$$

For any Borel set $B \in \mathcal{B}([0,T])$, $T \leqslant 1$, by condition 4 (stability under restrictions) in Definition 3.2.1,

$$F^{\mathbf{X}}_{L,[0,T]}(B) \geqslant F^{\mathbf{X}}_{L,[0,1]}(B).$$

Therefore, for any $0 < t < T$,

$$f^{\mathbf{X}}_{L,T}(t) \geqslant f^{\mathbf{X}}_{L,1}(t) = 1.$$

$\square$

For a given invariant intrinsic location functional L and $T \leqslant 1$, let $I^1_{L,T}$ be the collection of probability distributions of L on $[0,T]$ for periodic stationary processes with period 1. Let E^1_T be the collection of probability distributions with no point mass at ∞, and (càdlàg) densities f on $(0,T)$ satisfying:

1. f takes values in positive integers for all $t \in (0,T)$;
2. f satisfies the condition (TV).

Then we have the following result regarding the structure of the set $I^1_{L,T}$, parallel to the result for general intrinsic location functionals, Theorem 3.3.5.

Corollary 3.4.5. *$I^1_{L,T}$ is a convex subset of $\mathcal{P}_T$. Moreover, $I^1_{L,T} \subseteq C(E^1_T)$.*

Proof. By Proposition 3.4.4, the density f for any periodic ergodic process $\mathbf{X}$ with period 1 satisfies $f(t) \geqslant 1$ for all $t \in (0,T)$. The rest of the proof follows in the same way as that of Theorem 3.3.5. $\square$

Before proceeding to the next result, Theorem 3.4.7, which gives the other direction of the relation between $C(E_T^1)$ and the set of all possible distributions, we note that the definition of the location of the path supremum can be extended to the processes with càdlàg sample paths. This extension will be helpful in the proof of Theorem 3.4.7.

Remark 3.4.6. For any periodic stationary process $\mathbf{X}$ with period 1 and càdlàg sample paths, let $X'(t) = \limsup_{s\to t} X(s)$, $t \in \mathbb{R}$. Then $\mathbf{X'} = \{X'(t),\ t \in \mathbb{R}\}$ has upper semi-continuous sample paths and its supremum over the interval can be attained. As a result, for any $\mathbf{X}$ with càdlàg sample paths, the location of the path supremum for $\mathbf{X}$ can be defined as

$$\tau_{\mathbf{X},T} := \inf \left\{ t \in [0,T] : X'(t) = \sup_{s\in[0,T]} X'(s) \right\}.$$

Denote by $\mathcal{L}_I$ the set of invariant intrinsic location functionals. Let $I_T^1 = \bigcup_{L\in\mathcal{L}_I} I_{L,T}^1$ be the collection of all the possible distributions for invariant intrinsic location functionals and periodic stationary processes with period 1 on $[0,T]$. The next result, in combination with Corollary 3.4.5, shows that $I_T^1 = C(E_T^1)$.

Theorem 3.4.7. *For any $F \in C(E_T^1)$, there exists an invariant intrinsic location functional and a periodic stationary process with period 1, such that F is the distribution of this invariant intrinsic location functional for such process.*

Proof. It suffices to show that for any distribution $F \in E_T^1$, there exists a periodic ergodic process $\mathbf{Y}$ with period 1 such that F is the distribution of the unique location of the path supremum for $\mathbf{Y}$ on $[0,T]$. By Proposition 3.3.2, the density function of F, denoted by f, takes non-negative integer values and satisfies the condition (TV). As a result, f must be a piecewise constant function and has a unique decomposition

$$f(t) = \sum_{i=1}^{m} \mathbb{I}_{(u_i,v_i]}(t), \tag{3.7}$$

where m can be infinity and the intervals are maximal, in the sense that for any $i,j = 1,\dots,m$, $(u_i,v_i]$ and $(u_j,v_j]$ have only three possible relations:

$$(u_i,v_i] \subset (u_j,v_j], \quad \text{or} \quad (u_j,v_j] \subset (u_i,v_i], \quad \text{or} \quad [u_i,v_i] \cap [u_j,v_j] = \emptyset.$$

According to whether $u_i = 0$ or $v_i = T$, we call the intervals of the form $(0, T]$, $(0, v_i]$, $(u_i, T]$ and $(u_i, v_i]$ the base, left, right and central block(s), respectively. Observe that properties 1 and 2 in the definition of E_T^1 are equivalent to requiring that there is at least one base block, and the number of the central blocks does not exceed the number of the base blocks.

We construct the stationary process in spirit of Proposition 3.2.4. That is, first construct a periodic deterministic function g, and then uniformly shift its starting point to get $Y(t) = g(t + U)$, where U is a uniform random variable on $[0, 1]$. Let m_1 be the number of the base blocks in the collection. We group the entire collection of blocks into m_1 components by assigning to each base block at most one central block, and assigning the left and the right blocks in an arbitrary way. Assume $a = F(0) > 0$ and $b = 1 - F(T) > 0$. Let

$$d_1 = \frac{1}{m_1} a \text{ and } d_2 = \frac{1}{m_1} b.$$

For $j = 1, \ldots, m_1$, let

$$L_j = d_1 + \text{the total length of the blocks in the } j\text{th component} + d_2,$$

then $\sum_{i=1}^{m_1} L_i = 1$. Set $g(0) = 2$ and $g(L_1) = 2$. Using the blocks of the first component, we will define the function g on the interval $(0, L_1]$. If the first component has l left blocks, r right blocks and a central block, where l and r can potentially be infinity, we denote them by $(0, v_j]$, $j = 1, \ldots, l$, $(u_k, T]$, $k = 1, \ldots, r$ and $(u, v]$ respectively. The case where a central block does not exist corresponds to letting $u = v$. Set

$$g\left(\sum_{i=1}^{j-1} v_i + \sum_{i=1}^{j} \frac{1}{2^{i+1}} d_1\right) = g\left(\sum_{i=1}^{j} v_i + \sum_{i=1}^{j} \frac{1}{2^{i+1}} d_1\right) = 1 + 2^{-j}, \ j = 1, \ldots, l, \tag{3.8}$$

$$g\left(d_1 + \sum_{i=1}^{l} v_i\right) = g\left(d_1 + \sum_{i=1}^{l} v_i + v\right) = g\left(d_1 + \sum_{i=1}^{l} v_i + v + T - u\right) = \frac{1}{2},$$

and

$$\begin{aligned} g\left(L_1 - \sum_{i=1}^{j} \frac{1}{2^{i+1}} d_2 - \sum_{i=1}^{j-1} (T - u_i)\right) &= g\left(L_1 - \sum_{i=1}^{j} \frac{1}{2^{i+1}} d_2 - \sum_{i=1}^{j} (T - u_i)\right) \\ &= 1 + 2^{-j}, \ j = 1, \ldots, r. \end{aligned} \tag{3.9}$$

Next, if the values of g at two adjacent points constructed above, $t_1 < t_2$, are equal, we join them by a V-shaped curve satisfying some Lipschitz condition. We complete the function g by filling in the other gaps with straight lines between adjacent points (with different values). With the similar construction, we can also define g on the interval $[L_i, L_{i+1}]$, for $i = 1, \dots, m_1 - 1$. Then g is well defined on the interval $[0, 1]$ and we extend g as a periodic function with period 1. If a or b is equal to 0, we take (the càdlàg version of) the limit of the corresponding construction with $a \downarrow 0$ or $b \downarrow 0$. We have a periodic ergodic process Y as $Y(t) = g(t + U)$ for $t \in \mathbb{R}$, where U is uniformly distributed on $[0, 1]$. It is straightforward, though lengthy, by tracking the value of $L(g(t + U), [0, T])$ as a function of U, to see that the distribution of the location of the path supremum for $\mathbf{Y}$ is F. The proof is finally complete with an application of ergodic decomposition. $\square$

Remark 3.4.8. Since the only random location used in the proof of Theorem 3.4.7 is the location of the path supremum, we actually showed that the set of all possible distributions for invariant intrinsic location functionals is contained in the set of possible distributions solely for the location of path supremum. In this sense, the location of path supremum is a representative of the invariant intrinsic location functionals. This fact is related to the partially ordered random set representation of the intrinsic location functionals.

Remark 3.4.9. In the part of introduction we mentioned the question as whether every relatively stationary process defined on an interval $[0, T]$ can always be extended to a periodic stationary process with a given period $T' > T$. Proposition 3.4.4, together with Theorem 3.4.7, gives a negative answer to this question. To see this, let $T' = 1$, and consider the location of the path supremum denoted as τ. Let $T'' > 1$. As a result of Theorem 3.4.7, a simple scaling shows that for a probability distribution F on $[0, T]$ with its density function f on $(0, T)$, as long as f only takes values in positive multiples of $\frac{1}{T''}$ and satisfies the variation constraint (TV), there exists a periodic ergodic process $\mathbf{X}$ with period T'', such that F is the distribution of τ over the interval $[0, T]$ for $\mathbf{X}$. In particular, the value of $f(t)$ can be as small as $\frac{1}{T''}$ for some $t \in (0, T)$. Consider $\mathbf{X}|_{[0,T]}$, the restriction of $\mathbf{X}$ on $[0, T]$. It is a relatively stationary process. Suppose it can be extended to a periodic stationary process with period 1, denoted by $\mathbf{Y}$. Then by Proposition 3.4.4, the density of τ on $(0, T)$ for Y is bounded from below by 1. Since $\mathbf{Y}$ agrees with $\mathbf{X}|_{[0,T]}$ on $[0, T]$, the lower bound 1 is also valid for $\mathbf{X}|_{[0,T]}$, hence $\mathbf{X}$ as well. This contradicts the fact that $f(t)$ can take value $\frac{1}{T''}$. We therefore conclude that the relatively stationary process $\mathbf{X}|_{[0,T]}$ does not have a stationary extension with period 1.

We now turn back to the second part of the proof of Theorem 3.3.6 which we promised in the previous section.

Proof of Theorem 3.3.6, Part II. Recall that an intrinsic location functional L_1 is defined as follows:

$$L_1(g, I) = \inf \left\{ t \in I : g(t) = \sup_{s \in I} g(s),\ g(t) \geqslant \frac{1}{2} \right\},$$

and our goal in this part is to show that for any probability distribution $F \in E_T$ such that $f(t) \geqslant 1$ for all $t \in (0, T)$, there exists a periodic ergodic process with period 1 and non-negative sample paths, such that F is the distribution of L_1 on $[0, T]$ for that process.

Comparing the conditions for the distribution F and those for the distributions that we constructed in Theorem 3.4.7, the only difference is that F allows a possible point mass at ∞ while the distributions in Theorem 3.4.7 do not, because the location of the path supremum will always exist for processes with upper semi-continuous paths. This is the reason for which a modification is necessary. The way to construct the process changes accordingly, but not much. More precisely, let F be our target distribution, with possible point masses a and b at the two boundaries 0 and T, respectively. Additionally, it has a possible point mass c at ∞. Since the case where $c = 0$ has been covered in the proof of Theorem 3.4.7, here we focus on $c > 0$. Note that since $f - 1$ also satisfies the variation constraint in this case, there exists at least one component which does not have a central block. Set this component as the first component. The construction of the process $X(t) = g(t + U)$, hence the function g, goes exactly in the same way as in the proof of Theorem 3.4.7, except for that now for this first component, instead of building the central block by setting

$$g\left(d_1 + \sum_{i=1}^{l} v_i\right) = g\left(d_1 + \sum_{i=1}^{l} v_i + v\right) = g\left(d_1 + \sum_{i=1}^{l} v_i + v + T - u\right) = \frac{1}{2},$$

we set

$$g\left(d_1 + \sum_{i=1}^{l} v_i\right) = g\left(d_1 + \sum_{i=1}^{l} v_i + T + c\right) = \frac{1}{2},$$

and join them using a V-shaped curve as in the other cases. The construction of the rest of this component is shifted correspondingly. It is not difficult to verify that this part will contribute the desired mass at ∞. $\square$

The variation constraint (TV) implies an upper bound for the density for intrinsic location functionals and stationary processes:

$$f^{\mathbf{X}}_{L,T}(t) \leqslant \max\left(\frac{1}{t}, \frac{1}{T-t}\right), \quad 0 < t < T. \tag{3.10}$$

Moreover, such an upper bound was proved to be optimal (Samorodnitsky and Shen, 2013b). With periodicity and the invariance property, we can now improve the above bound, and show that the improved upper bound is also optimal.

Proposition 3.4.10. *Let L be an invariant intrinsic location functional, $\mathbf{X}$ be a periodic stationary process with period 1, and $T \in (0, 1]$. Then the density $f^{\mathbf{X}}_{L,T}$ satisfies*

$$f^{\mathbf{X}}_{L,T}(t) \leqslant \max\left(\lfloor\frac{1-T}{t}\rfloor, \lfloor\frac{1-T}{T-t}\rfloor\right) + 2. \tag{3.11}$$

Moreover, for any $t \in (0, \frac{T}{2})$ such that $\frac{1-T}{t}$ is not an integer and $t \in [\frac{T}{2}, T)$ such that $\frac{1-T}{T-t}$ is not an integer, there exists an invariant intrinsic location functional L and a periodic stationary process $\mathbf{X}$ with period 1, such that the equality in (3.11) is achieved at t.

Proof. Let $g^{\mathbf{X}}_{L,T}(t) = f^{\mathbf{X}}_{L,T}(t) - 1$, then for every $0 < t_1 < t_2 < T$, the variation constraint will be

$$\mathrm{TV}_{(t_1,t_2)}(g^{\mathbf{X}}_{L,T}) = \mathrm{TV}_{(t_1,t_2)}(f^{\mathbf{X}}_{L,T}) \leqslant f^{\mathbf{X}}_{L,T}(t_1) + f^{\mathbf{X}}_{L,T}(t_2) = g^{\mathbf{X}}_{L,T}(t_1) + g^{\mathbf{X}}_{L,T}(t_2) + 2.$$

Denote $a = \inf_{0<s\leqslant t} g^{\mathbf{X}}_{L,T}(s)$, $b = \inf_{t\leqslant s<T} g^{\mathbf{X}}_{L,T}(s)$. For any given $\varepsilon > 0$, there exists $u \in (0, t]$ such that

$$g^{\mathbf{X}}_{L,T}(u) \leqslant a + \varepsilon,$$

and there exists $v \in [t, T)$ such that

$$g^{\mathbf{X}}_{L,T}(v) \leqslant b + \varepsilon.$$

Note that

$$at + b(T-t) \leqslant \int_0^T g^{\mathbf{X}}_{L,T}(s)\mathrm{d}s = \int_0^T (f^{\mathbf{X}}_{L,T}(s) - 1)\mathrm{d}s \leqslant 1 - T. \tag{3.12}$$

Now applying the variation constraint to the interval $[u, v]$, we have

$$\begin{aligned} a + b + 2\varepsilon &\geqslant g^{\mathbf{X}}_{L,T}(u) + g^{\mathbf{X}}_{L,T}(v) \\ &\geqslant |g^{\mathbf{X}}_{L,T}(t) - g^{\mathbf{X}}_{L,T}(u)| + |g^{\mathbf{X}}_{L,T}(v) - g^{\mathbf{X}}_{L,T}(t)| - 2 \\ &\geqslant (g^{\mathbf{X}}_{L,T}(t) - a - \varepsilon)_+ + (g^{\mathbf{X}}_{L,T}(t) - b - \varepsilon)_+ - 2. \end{aligned}$$

By the definition of a and b, $a \leqslant g^{\mathbf{X}}_{L,T}(t)$ and $b \leqslant g^{\mathbf{X}}_{L,T}(t)$. Letting $\varepsilon \to 0$, we have

$$g^{\mathbf{X}}_{L,T}(t) \leqslant a + b + 1. \tag{3.13}$$

Combining (3.12) and (3.13) leads to

$$g^{\mathbf{X}}_{L,T}(t) \leqslant \max\left(\frac{1-T}{t}, \frac{1-T}{T-t}\right) + 1.$$

Then for every $0 < t < T$, an upper bound of $f^{\mathbf{X}}_{L,T}(t)$ is

$$f^{\mathbf{X}}_{L,T}(t) \leqslant \max\left(\frac{1-T}{t}, \frac{1-T}{T-t}\right) + 2.$$

By Proposition 3.3.2, $f^{\mathbf{Y}}_{L,T}$ takes integer values for any periodic ergodic process $\mathbf{Y}$ with period 1. Through ergodic decomposition, we further have the upper bound:

$$f^{\mathbf{X}}_{L,T}(t) \leqslant \max\left(\lfloor\frac{1-T}{t}\rfloor, \lfloor\frac{1-T}{T-t}\rfloor\right) + 2.$$

It remains to prove that such upper bound can be approached. For any $t \in (0, \frac{T}{2})$ such that $\frac{1-T}{t}$ is not an integer, define f by

$$f(s) = \begin{cases} 1 + \lfloor\frac{1-T}{t}\rfloor, & s \in (0, t), \\ 2 + \lfloor\frac{1-T}{t}\rfloor, & s \in [t, t+\varepsilon), \\ 1, & s \in [t+\varepsilon, T), \end{cases}$$

where ε is small enough so that $\int_0^T f(s)\mathrm{d}s \leqslant 1$. As f takes integer values and satisfies the condition (TV), by Theorem 3.4.7, there exists an invariant intrinsic location functional L and a periodic ergodic stationary process with period 1 such that f is the density of L for such process. By similar construction, we can also find an invariant intrinsic location functional L and a periodic ergodic process with period 1 such that the density of L for such process approaches $\lfloor \frac{1-T}{T-t} \rfloor + 2$ at point t for $t \in [\frac{T}{2}, T)$ satisfying $\frac{1-T}{T-t}$ is not an integer. $\square$

We end this section by comparing the upper bound (3.11) with the result (3.10) for general stationary processes. For $t \leqslant \frac{T}{2}$, the following inequality holds between these two bounds:

$$\max\left\{\lfloor \frac{1-T}{t} \rfloor, \lfloor \frac{1-T}{T-t} \rfloor\right\} + 2 \leqslant \frac{1-T}{t} + 2 \leqslant \frac{1}{t} = \max\left\{\frac{1}{t}, \frac{1}{T-t}\right\}.$$

For $t \geqslant \frac{T}{2}$,

$$\max\left\{\lfloor \frac{1-T}{t} \rfloor, \lfloor \frac{1-T}{T-t} \rfloor\right\} + 2 \leqslant \frac{1-T}{T-t} + 2 \leqslant \frac{1}{T-t} = \max\left\{\frac{1}{t}, \frac{1}{T-t}\right\}.$$

Therefore, the upper bound in (3.11) is always sharper than that in (3.10). The improvement is most significant when T is close to 1 and t is close to 0 or T.

3.5 First-time intrinsic location functionals

In this section, we introduce another type of intrinsic location functionals called the *first-time intrinsic location functionals* via the partially ordered random set representation.

Definition 3.5.1. An intrinsic location functional L is called a *first-time intrinsic location functional*, if it has a partially ordered random set representation $(S(\mathbf{X}), \preceq)$ such that for any $t_1, t_2 \in S(\mathbf{X})$, $t_1 \leqslant t_2$ implies $t_2 \preceq t_1$.

It is easy to see that the notion of the first-time intrinsic location functionals is a generalization of the first hitting times. As its name suggests, it contains all the intrinsic location functionals which can be defined as "the first time" that some condition is met.

Proposition 3.5.2. *Let $\mathbf{X}$ be a periodic stationary process with period 1, and L be a first-time intrinsic location functional. Fix $T \in (0,1]$. Then the density of L on $(0,T)$ for $\mathbf{X}$ is non-increasing.*

Proof. By ergodic decomposition, it suffices to prove the result for periodic ergodic process $\mathbf{X}$ with period 1 having the representation $X(t) = g(t+U)$, where U is a uniform random variable on $[0,1]$. Let $(S, \preceq)$ be a partially ordered random set representation for L. By a similar argument as the discussion below (3.4), we have for $t \in (0,T)$,

$$f(t) = \left|\{s \in S(g) \cap (0,1] : a_s \geqslant t,\ b_s \geqslant T-t\}\right|,$$

where $a_s = \sup\{\Delta s \in \mathbb{R} : r \preceq s \text{ for all } r \in (s-\Delta s, s) \cap S(g)\}$, $b_s = \sup\{\Delta s \in \mathbb{R} : r \preceq s \text{ for all } r \in (s, s+\Delta s) \cap S(g)\}$. By the definition of first-time intrinsic location functionals and that of b_s, we have

$$b_s = \infty, \quad \text{for any } s \in S(g).$$

Thus for $t_1 \leqslant t_2$,

$$f(t_2) = \left|\{s \in S(g) \cap (0,1] : a_s \geqslant t_2\}\right| \text{ and } f(t_1) = \left|\{s \in S(g) \cap (0,1] : a_s \geqslant t_1\}\right|.$$

If there exists $s \in S(g) \cap (0,1]$ such that $a_s \geqslant t_2$, then $a_s \geqslant t_2 \geqslant t_1$, which means that $f(t_1) \geqslant f(t_2)$. As a result, f is non-increasing on the interval $(0,T)$. ☐

For any first-time intrinsic location functional L and $T \leqslant 1$, let $I_{L,T}^M$ be the collection of the probability distributions of L on $[0,T]$ for all periodic stationary processes with period 1. Denote by E_T^M the subset of E_T consisting of the distributions with non-increasing density functions on $(0,T)$ and no point mass at T. Then we have the following result of the structure of $I_{L,T}^M$, parallel to Section 4.

Proposition 3.5.3. *$I_{L,T}^M$ is a convex subset of $\mathcal{P}_T$ and $I_{L,T}^M \subseteq C(E_T^M)$.*

The proof of Proposition 3.5.3 follows in a similar way to that of Theorem 3.3.5 and is omitted.

As in the previous cases, the other direction also holds.

Proposition 3.5.4. *For any $F \in C(E_T^M)$, there exists a first-time intrinsic location functional and a periodic stationary process with period* 1*, such that F is the distribution of this first-time intrinsic location functional for such process.*

Proof. We can actually use a single first-time intrinsic location functional for the proof. For example, let $L(g, I) = L_2(g, I) = \inf\{t \in I : g(t) = -1\}$ as defined in the proof of Theorem 3.3.6. By ergodic decomposition, it suffices to show the result for distributions in E_T^M. Let F be a probability distribution in E_T^M. Equivalently, F is a probability distribution supported on $[0, T] \cup \{\infty\}$, with a possible point mass a at 0, a possible point mass at ∞, and a non-increasing density function f which takes non-negative integer values. Our goal is to show that there exists a periodic ergodic process with period 1 such that the distribution of the first time reaching level -1 between 0 and T for such process is F. For ease of exposition, assume the point masses at 0 and at ∞ are both positive. The degenerate cases can be handled in a similar way. Since f is non-increasing on $(0, T)$ with non-negative integer values, it can be written as

$$f(t) = \sum_{i=0}^{\infty} \mathbb{I}_{(0,u_i)}(t),$$

where $u_i \geqslant u_{i+1}$. Define $s_i = \sum_{k=1}^{i} u_k$, $i = 1, 2, ...$ and $s_0 = 0$. Let

$$g(s_i) = -1, \text{ for } i = 0, 1, \ldots$$

In addition to $s_0, s_1, \ldots$, we set $g(t) = -1$ for $t \in [s_\infty, s_\infty + a]$ and $g(1) = -1$. Note that since $\int_0^1 f(t)\mathrm{d}t \leqslant 1$, $0 \leqslant s_\infty \leqslant s_\infty + a \leqslant 1$. Next we join the consecutive points $(s_i, -1)$ and $(s_{i+1}, -1)$, $i = 0, 1, \ldots$ using V-shaped curves satisfying some Lipschitz condition with, for example, Lipschitz constant 1. Similarly, use a V-shaped curve to join $(s_\infty + a, -1)$ and $(1, -1)$. Therefore, we can construct a periodic deterministic function g with period 1, and the required periodic ergodic process can be written as $X(t) = g(t + U)$ for $t \in \mathbb{R}$, where U follows a uniform distribution on $[0, 1]$. It is then routine to check that the distribution of L is exactly F by expressing the value of L as a function of U. $\square$

We have now all the pieces to complete the proof of Theorem 3.3.6.

Proof of Theorem 3.3.6, Part III. Let $F \in E_T$, and f be its density function on $(0, T)$. Recall that our goal in this part is to show that if f is non-increasing with $\sup\{t : f(t) > 0\} < T$, then for the intrinsic location functional $L_2(g, I) = \inf\{t \in I : g(t) = -1\}$, there exists a periodic ergodic process $\mathbf{X}$, such that F is the distribution of L_2 on $[0, T]$ for $\mathbf{X}$. Note that since $f(t)$ takes value 0 as t approaches T, by the definition of E_T, F do not have a point mass at T. As a result, $F \in E_T^M$. Thus, by the proof of Proposition 3.5.4, F is the distribution of L_2 for some periodic ergodic process with period 1. $\square$

Denote by $\mathcal{L}_M$ the set of first-time intrinsic location functionals. Let $I_T^M = \bigcup_{L \in \mathcal{L}_M} I_{L,T}^M$ be the collection of all the possible distributions for first-time intrinsic location functionals and periodic stationary processes with period 1 on $[0, T]$. Denote by A_T^M the class of probability distribution on $(0, T)$ with the properties that the corresponding density is càdlàg and non-increasing. We would like to give a verification whether a function in A_T^M is also in I_T^M. The recently developed concept of joint mixability (Wang et al., 2013) is helpful.

In the following part, for any set A of distributions, we write $f \in_d A$, if there exists $F \in A$ such that f is the corresponding density part of F.

In the definition below, we slightly generalize the concept of joint mixability to the case of possibly countably many distributions. In the following N is either a positive integer or it is infinity. If $N = \infty$, we interpret any tuple $(x_1, \ldots, x_N)$ as $(x_i,\ i = 1, 2, \ldots)$. Joint mixability and intrinsic location functionals are connected in Proposition 3.5.6 below.

Definition 3.5.5. (Wang et al., 2013) Suppose $N \in \mathbb{N} \cup \{\infty\}$. A random vector $(X_1, \ldots, X_N)$ is said to be a joint mix if $P(\sum_{i=1}^N X_i = C) = 1$ for some $C \in \mathbb{R}$. An N-tuple of distributions $(F_1, \ldots, F_N)$ is said to be jointly mixable if there exists a joint mix $\mathbf{X} = (X_1, \ldots, X_N)$ such that $X_i \sim F_i$, $i = 1, \ldots, N$.

Proposition 3.5.6. *For any $f \in_d A_T^M$, let $N = \lceil f(0+) \rceil$, and define the distribution functions*

$$F_i : \mathbb{R} \to [0, 1], \quad x \mapsto \min\{(i - f(x)\mathbb{I}_{\{x<T\}})_+, 1\}\mathbb{I}_{\{x \geqslant 0\}}, \quad i = 1, \ldots, N. \tag{3.14}$$

Then $f \in_d I_T^M$ if there exists a random vector $\mathbf{X} = (X_1, \dots, X_N)$ such that $X_i \sim F_i$, $i = 1, \dots, N$ and $P(\sum_{i=1}^N X_i \leqslant 1) = 1$. In particular, $f \in_d I_T^M$ if $(F_1, \dots, F_N)$ is jointly mixable.

Proof. Suppose that there exists a random vector $\mathbf{X} = (X_1, \dots, X_N)$ such that $X_i \sim F_i$, $i = 1, \dots, N$ and $P(\sum_{i=1}^N X_i \leqslant 1) = 1$. For $\mathbf{x} = (x_1, \dots, x_N)$ satisfying $\sum_{i=1}^N x_i \leqslant 1$, define

$$f_{\mathbf{x}} : [0, T] \to \mathbb{R}_+, \quad y \mapsto \sum_{i=1}^N \mathbb{I}_{\{y \leqslant x_i\}}.$$

Obviously $f_{\mathbf{x}}$ is a non-increasing function and we can check

$$\int_0^T f_{\mathbf{x}}(y)\mathrm{d}y = \sum_{i=1}^N \int_0^T \mathbb{I}_{\{y \leqslant x_i\}}\mathrm{d}y = \sum_{i=1}^N x_i \leqslant 1.$$

Thus, $f_{\mathbf{x}}$ is a non-increasing function on $[0, T]$ taking values in $\mathbb{N}_0$, $\int_0^T f_{\mathbf{x}}(y)\mathrm{d}y \leqslant 1$, and hence $f_{\mathbf{x}} \in_d E_T^M$. Moreover, for $y \in [0, T]$,

$$\begin{aligned}\mathbb{E}[f_{\mathbf{X}}(y)] &= \mathbb{E}\left[\sum_{i=1}^N \mathbb{I}_{\{y \leqslant X_i\}}\right] \\ &= \lfloor f(y) \rfloor + \mathbb{E}\left[\mathbb{I}_{\{y \leqslant X_{\lfloor f(y) \rfloor}\}}\right] = \lfloor f(y) \rfloor + (f(y) - \lfloor f(y) \rfloor) = f(y).\end{aligned}$$

Therefore, we conclude that $f \in_d I_T^M$ since it is a convex combination of $f_{\mathbf{x}}$.

Now suppose that $(F_1, \dots, F_N)$ is jointly mixable. Then there exists a joint mix $\mathbf{X} = (X_1, \dots, X_N)$ such that $X_i \sim F_i$, $i = 1, \dots, N$ and $P(\sum_{i=1}^N X_i = C) = 1$ for some $C \in \mathbb{R}$. It suffices to verify that $C \leqslant 1$, which follows from

$$\begin{aligned}C = \sum_{i=1}^N \mathbb{E}[X_i] &= \sum_{i=1}^N \int_0^T (1 - F_i(x))\mathrm{d}x \\ &= \sum_{i=1}^N \int_0^T \min\{(f(x) - i + 1)_+, 1\}\mathrm{d}x = \int_0^T f(x)\mathrm{d}x \leqslant 1. \qquad (3.15)\end{aligned}$$

This completes the proof. $\square$

Remark 3.5.7. In this section, N might be infinity. It can be easily checked that in the case of $N = \infty$, the limit $\sum_{i=1}^{N} X_i$ in the above proof is well-defined since $\sum_{i=1}^{N} \mathbb{E}[X_i] \leqslant 1$ and $X_i \geqslant 0$, $i = 1, \ldots, N$.

Corollary 3.5.8. *For a given density function $f \in_d A_T^M$, if there exists a step function $g \in_d E_T^M$ such that*

$$g(t) \geqslant f(t), \text{ for all } t \in (0, T),$$

then $f \in_d I_T^M$.

Proof. For any $f \in_d A_T^M$, take N and F_i, $i = 1, \ldots, N$ as defined in Proposition 3.5.6. Let $\mathbf{X} = (X_1, \ldots, X_N)$ be a random vector such that $X_i \sim F_i$, $i = 1, \ldots, N$. Then we have

$$\sum_{i=1}^{N} X_i \leqslant \sum_{i=1}^{N} f^{-1}(i-1) \leqslant \int_0^T g(t)\mathrm{d}t \leqslant 1$$

hold almost surely. Thus, $f \in_d I_T^M$ by Proposition 3.5.6. □

Corollary 3.5.9. *Suppose that $f \in_d A_T^M$ is convex on $[0, T]$ and*

$$\sum_{i=0}^{N} f^{-1}(i) \leqslant 1 + f^{-1}(1). \tag{3.16}$$

Then $f \in_d I_T^M$.

Proof. Let $N = \lceil f(0+) \rceil$ and F_i, $i = 1, \ldots, N$ be as in (3.14). Denote by μ_i the mean of F_i for $i = 1, \ldots, N$. Apparently F_i has a non-increasing density supported in $[f^{-1}(i), f^{-1}(i-1)]$ for each $i = 1, \ldots, N$. By the convexity of f, we have

$$\sum_{i=1}^{N} f^{-1}(i) + \max\{f^{-1}(i-1) - f^{-1}(i) : i = 1, \ldots, N\} = \sum_{i=0}^{N} f^{-1}(i) - f^{-1}(1) \leqslant 1.$$

Since each F_i has non-increasing densities, conditions in Corollary 4.7 of Jakobsons et al. (2016) are satisfied, giving that there exists $\mathbf{X} = (X_1, \ldots, X_N)$ such that $X_i \sim F_i$, $i = 1, \ldots, N$ and

$$\text{ess-sup}\left(\sum_{i=1}^{N} X_i\right) = \max\left\{\sum_{i=1}^{N} f^{-1}(i) + \max_{i=1,\ldots,N}\{f^{-1}(i-1) - f^{-1}(i)\}, \sum_{i=1}^{N} \mu_i\right\} \leqslant 1.$$

The corollary follows from Proposition 3.5.6. □

Remark 3.5.10. Formally, Corollary 4.7 of Jakobsons et al. (2016) only gives, for any $\varepsilon > 0$ and $N \in \mathbb{N}$, the existence of $\mathbf{X} = (X_1, \dots, X_N)$ such that

$$\text{ess-sup}\left(\sum_{i=1}^{N} X_i\right) < \max\left\{\sum_{i=1}^{N} f^{-1}(i) + \max_{i=1,\dots,N}\{f^{-1}(i-1) - f^{-1}(i)\}, \sum_{i=1}^{N} \mu_i\right\} + \varepsilon.$$

A standard compactness argument would justify the case $\varepsilon = 0$ and $N = \infty$. Corollary 4.7 of Jakobsons et al. (2016) requires the joint mixability of non-increasing densities; see Theorem 3.2 of Wang and Wang (2016). For $f \in_d A_T^M$, there is generally no constraints (except for location constraints) on the distributions $F_1, \dots, F_N$. It is a difficult task to analytically verify whether a given tuple of distributions is jointly mixable. For some other known necessary and sufficient conditions for joint mixability, see Wang and Wang (2016).

Corollary 3.5.11. *Suppose that* $f \in_d A_T^M$ *is linear on its essential support* $[0, b]$ *and* $f(b) = 0$. *Then* $f \in_d I_T^M$.

Proof. Obviously the slope of the linear function f on its support is not zero.

1. $\int_0^T f(x)\mathrm{d}x = 1$. In this case, f is convex on $[0, T]$. We only need to verify (3.16) in Corollary 3.5.9. Since $T < 1$ and since f integrates to 1, we have $N \geqslant 3$. Note that, from integration by parts and change of variables, $\int_0^N f^{-1}(t)\mathrm{d}t = \int_0^T f(x)\mathrm{d}x = 1$. It follows from the linearity of f that

$$\begin{aligned}\sum_{i=0}^{N} f^{-1}(i) - f^{-1}(1) &= \sum_{i=3}^{N} f^{-1}(i) + f^{-1}(0) + f^{-1}(2) \\ &= \sum_{i=3}^{N} f^{-1}(i) + \int_0^2 f^{-1}(t)\mathrm{d}t \\ &\leqslant \int_2^N f^{-1}(t)\mathrm{d}t + \int_0^2 f^{-1}(t)\mathrm{d}t = 1.\end{aligned}$$

   The desired result follows from Corollary 3.5.9.

2. $\int_0^T f(x)\mathrm{d}x < 1$. This case can be obtained from a mixture of (a) and $g \in_d E_T^M$ where $g : [0, T] \to \{0\}$. □

When $\int_0^T f(x)\mathrm{d}x < 1$, we obtain a sufficient condition for $f \in_d A_M^T$ to be $f \in_d I_T^M$ using Proposition 3.5.6 together with a result in Embrechts et al. (2015).

Corollary 3.5.12. *For any $f \in_d A_T^M$, let $N = \lceil f(0+) \rceil$. Then $f \in_d I_T^M$ if*

$$\max_{i=1,\dots,N} \{f^{-1}(i-1) - f^{-1}(i)\} \leqslant 1 - \int_0^T f(x)\mathrm{d}x.$$

Proof. Let F_i, $i = 1, \dots, N$ be as in (3.14). Apparently F_i is supported in $[f^{-1}(i), f^{-1}(i-1)]$ for each $i = 1, \dots, N$. Denote $L = \max\{f^{-1}(i-1) - f^{-1}(i) : i = 1, \dots, N\}$. From Corollary A.3 of Embrechts et al. (2015), there exists a random vector $\mathbf{X} = (X_1, \dots, X_N)$ such that $X_i \sim F_i$, $i = 1, \dots, N$ and

$$P\left(\left|\sum_{i=1}^N X_i - \sum_{i=1}^N \mathbb{E}[X_i]\right| \leqslant L\right) = 1.$$

From (3.15), we have $\sum_{i=1}^N \mathbb{E}[X_i] = \int_0^T f(x)\mathrm{d}x$ and therefore,

$$P\left(\sum_{i=1}^N X_i \leqslant 1\right) \geqslant P\left(\sum_{i=1}^N X_i \leqslant L + \int_0^T f(x)\mathrm{d}x\right) = 1.$$

□

Chapter 4

Noether Theorem for Random Locations

4.1 Introduction

The famous Noether theorem in mathematical physics Noether (1918) shows that each differentiable symmetry of a system corresponds to a corresponding conservation law. The most important and immediate examples include translation in space and the conservation of momentum, translation in time and the conservation of energy, rotation in space and the conservation of angular momentum, *etc.* A thorough review of the Noether theorem can be found in the book by Kosmann-Schwarzbach (Kosmann-Schwarzbach, 2011).

Since the last two decades of the twentieth century, various works have been carried out to extend the Noether theorem to stochastic settings. Just to name a few, Yasue (Yasue, 1981) proposed a theory for stochastic calculus of variations, and got a corresponding generalization of the Noether theorem. Misawa (Misawa, 1994) considered the conservative quantities and symmetry for stochastic dynamical systems described by certain type of stochastic differential equations. Thieullen and Zambrini proved a version of the Noether theorem, in which they associated a function giving a martingale to each family of transformations exhibiting certain symmetry (Thieullen and Zambrini, 1997b). They also extended the Noether theorem to diffusion processes in $\mathbb{R}^3$ whose diffusion matrix is proportional to identity (Thieullen and Zambrini, 1997a). Entering the new century, van Casteren (Van Casteren, 2003) obtained a version of the stochastic Noether theorem using the ideas and backgrounds from stochastic control. More recently, Baez and Fong (Baez and Fong, 2013) considered Markov processes and found an analogy of the classical Noether theorem in this setting. Along this direction, Gough, Ratiu and Smolyanov (Gough et al., 2015) gave a Noether theorem for dissipative quantum dynamical semi-groups. Another scenario where an external random force exists was studied by Luzcano and de Oca (Lezcano and de Oca, 2018).

The random locations of stochastic processes exhibiting certain probabilistic symmetries have been studied in a series of works in the past years. In Samorodnitsky and Shen (2013a), Samorodnitsky and Shen introduced a large family of random locations called "intrinsic location functionals", which include the location of the path supremum, the first/last hitting time to a fixed level, *etc.* It was shown that the distribution of any random location in this family for a stationary process must satisfy a specific set of conditions. Similar results were later established between a subclass of intrinsic location functionals and stochastic processes with stationary increments (Shen, 2016). In Shen (2018), the stochastic processes combining both a scaling symmetry and a stationarity of the increments were studied, and it is shown that stronger conditions hold for the distribution of its path supremum over an interval.

As the research of random locations progressed, it became clearer and clearer that there is a general correspondence between probabilistic symmetries and classes of random locations, such that the distributions of the random locations behave in a very specific way under the corresponding symmetry. Indeed, it is not difficult to see that the setting for the random locations of stochastic processes having probabilistic symmetries is similar to the settings in which the Noether theorems hold, in that they are both systems with infinitesimally generated symmetries. This observation leads to the question as whether a result of the Noether type exists for the random locations. There is, however, a critical difference: in the case of random locations, the symmetries are only in the distributional sense. While the overall distribution of the processes, hence also the distributions of the random locations, remain invariant after the corresponding transformations, the values of the locations do evolve after the transformations in each realization. As a result, the mathematical tools used to derive the Noether theorems for deterministic systems can not be applied to get similar results here. It turns out that the methods developed in the literature previously mentioned are not helpful as well.

The goal of this chapter is, therefore, to provide a framework which contains the aforementioned random locations and probabilistic symmetries as special cases, and in which a Noether theorem can be established. To this end, we generalize the notion of random location by dissociating it from the paths of stochastic processes. More precisely, the random locations are no longer functionals of the paths as in Samorodnitsky and Shen (2013a); Shen (2016, 2018), but special elements in a point process which may or may not be related to a stochastic process in continuous time. Another point process is then constructed, and we show that the distribution of the random locations can be expressed in terms of the control measure of the latter point processes. Finally, a conservation law appears using a function derived from the control measure.

The rest of this chapter is organized as follows. In section 4.2 we introduce the basic settings and definitions, with examples making connections to the existing literature. In section 4.3 we state and prove the main results, including the Noether theorem as a conservation law when the interval of interest moves along a flow, and its consequences, such as a constraint on the total variation of the density function of the random locations. Section 4.4 discusses the boundary and near-boundary behavior of the random locations.

4.2 Basic settings

Here and throughout the chapter, let $\mathcal{I}$ be the collection of all the non-degenerate compact intervals on $\mathbb{R}$. Let $\bar{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$, and equip it with the σ-field $\bar{\mathcal{B}} = \sigma(\mathcal{B}(\mathbb{R}), \{\infty\})$. That is, we treat ∞ as a separate point and take the Borel σ-field of the extended topology.

Definition 4.2.1. A stochastic process $\{L(I)\}_{I\in\mathcal{I}}$ indexed by compact intervals and taking values in $\bar{R}$ is called an intrinsic random location, if it satisfies the following conditions:

1. For every $I \in \mathcal{I}$, $L(I) \in I \cup \{\infty\}$.

2. (Stability under restriction) For every $I_1, I_2 \in \mathcal{I}$, $I_2 \subseteq I_1$, if $L(I_1) \in I_2$, then $L(I_1) = L(I_2)$.

3. (Consistency of existence) For every $I_1, I_2 \in \mathcal{I}$, $I_2 \subseteq I_1$, if $L(I_2) \neq \infty$, then $L(I_1) \neq \infty$.

Intuitively, the value ∞ is typically used to deal with the case where a random location is not well-defined on a given interval for certain realization. For example, if the random location is defined as the first hitting time of a continuous-time stochastic process to certain level, then it is possible that the process does not hit the level in the given interval. In this case we will assign ∞ as the value of the random location.

Let $\varphi = \{\varphi^t\}_{t\in\mathbb{R}}$ be a flow on $\mathbb{R}$. That is, $\{\varphi^t\}_{t\in\mathbb{R}}$ is a family of real-valued functions defined on $\mathbb{R}$, satisfying $\varphi^0 = Id$ and $\varphi^s \circ \varphi^t = \varphi^{s+t}$ for $s, t \in \mathbb{R}$. We further assume that

$$\varphi(x,t) = \varphi^t(x) \in C^{1,1}(\mathbb{R} \times \mathbb{R}); \tag{4.1}$$

$$\text{the fixed points } \Phi_0 := \{x : \varphi^t(x) \equiv x\} \text{ are isolated.} \tag{4.2}$$

In many cases, it will be convenient to consider the extended real line $\mathbb{R} \cup \{-\infty, \infty\}$ and the set of extended fixed points $\bar{\Phi}_0 = \Phi_0 \cup \{-\infty, \infty\}$. Two points $\alpha, \beta, \alpha < \beta$ are called consecutive in $\bar{\Phi}_0$, if $\alpha, \beta \in \bar{\Phi}_0$, and $(\alpha, \beta) \cap \bar{\Phi}_0 = \phi$. Note that since there is no fixed point between α and β, and φ is continuous, $\varphi^t(x)$ must be monotone in t for any fixed $x \in (\alpha, \beta)$ and increasing in x for any fixed $t \in \mathbb{R}$. In particular, for every fixed $x \in (\alpha, \beta)$, $\varphi^{\cdot}(x)$ is a bijection from $\mathbb{R}$ to (α, β).

An intrinsic random location is called φ-stationary, if its distribution is compatible with the flow φ, more precisely, if $\varphi^t(L([a,b])) \stackrel{d}{=} L([\varphi^t(a), \varphi^t(b)])$ for every $t \in \mathbb{R}$ and $a, b \in \mathbb{R}, a < b$. It is called stationary if the flow is the translation $\varphi^t(x) = x + t$.

Remark 4.2.2. Due to the continuity of φ, a φ-stationary intrinsic random location, restricted to the open interval between two consecutive extended fixed points of φ, can be easily transformed into a stationary intrinsic random location using a transformation. More precisely, let L be a φ-stationary intrinsic random location and α, β be two consecutive points in $\bar{\Phi}_0$. Fix any $x_0 \in (\alpha, \beta)$. Then $\varphi^t(x_0)$ is a continuous monotone function in t

with $\lim_{t\to-\infty}\varphi^t(x_0)=\alpha$ and $\lim_{t\to\infty}\varphi^t(x_0)=\beta$, or symmetrically, $\lim_{t\to-\infty}\varphi^t(x_0)=\beta$ and $\lim_{t\to\infty}\varphi^t(x_0)=\alpha$. As a result, we can define a transform $\tau:(\alpha,\beta)\to\mathbb{R}$ by

$$\varphi^{\tau(x)}(x_0)=x.$$

That is, $\tau(x)$ is the time it takes to go from x_0 to x following the flow φ, or from x to x_0 if its value is negative. Note that we have identity between τ and φ:

$$\tau(x)=\tau((\varphi^t)^{-1}(x))+t \tag{4.3}$$

for $x\in(\alpha,\beta)$ and $t\in\mathbb{R}$.

Since τ is a bijection, its inverse τ^{-1} is well-defined. Define L' by

$$L'(I)=\tau(L(\tau^{-1}(I))),\quad I\in\mathcal{I},$$

then it is elementary to check that such defined L' is a stationary intrinsic random location. Consequently, all the results regarding a φ-stationary intrinsic random location can be transformed into corresponding results regarding stationary intrinsic random locations, and we only need to prove the latter ones.

As explained in Introduction, the definition of intrinsic random location is motivated by the random locations of stochastic processes studied in previous literature (Samorodnitsky and Shen, 2013a; Shen, 2016); Shen (2018). Therefore, it is not surprising that one important way to obtain φ-stationary intrinsic random locations is through the stochastic processes exhibiting some probabilistic symmetry under φ, and to define the random location as a functional which is determined by the path of the process and compatible with φ. For example, let the flow be the translation $\varphi^t(x)=x+t$. Correspondingly, we have the (strictly) stationary processes as the family of processes whose distributions are invariant under φ. In this case, let H be a space of functions closed under translation, equipped with the cylindrical σ-field, and consider a mapping $L_H:\mathcal{I}\times H\to\bar{\mathbb{R}}$ satisfying

1. $L_H(I,\cdot):H\to\bar{\mathbb{R}}$ is measurable;
2. $L_H(I,f)\in I\cup\{\infty\}$ for every $f\in H$;

3. For every $I_1, I_2 \in \mathcal{I}$, $I_2 \subseteq I_1$ and every $f \in H$, if $L_H(I_1, f) \in I_2$, then $L_H(I_2, f) = L_H(I_1, f)$;

4. For every $I_1, I_2 \in \mathcal{I}$, $I_2 \subseteq I_1$ and every $f \in H$, if $L_H(I_2, f) \neq \infty$, then $L_H(I_1, f) \neq \infty$;

5. $L_H(I, f) = L(I - t, f \circ \varphi^t) + t$ for any $f \in H$, where $I - t := \{x \in \mathbb{R} : x + t \in I\}$.

Conditions 2, 3 and 4 correspond to the three conditions in the definition for intrinsic random locations, while condition 5 requires the random location to be compatible with translation. Then it is easy to check that the random location L defined by

$$L(I)(\omega) = L_H(I, X(\cdot, \omega))$$

is a stationary intrinsic random location if $\mathbf{X} = \{X(t, \omega)\}_{t\in\mathbb{R}}$ is a stationary process with sample paths in H. Such a mapping like L_H was introduced in Samorodnitsky and Shen (2013a), where its relation to stationarity has also been studied in detail.

Other probabilistic symmetries of stochastic processes which can be used to define intrinsic random locations stationary with respect to certain flow include self-similarity, isometry (in higher dimension), stationarity of the increments, *etc.* They have been discussed respectively in the sequence of papers (Shen, 2016, 2013, 2018). Two cases are special and worth some more attention.

First, even for the same φ, there can be various ways to construct φ-stationary intrinsic random locations from stochastic processes. For instance, still consider the translation. If instead of the distribution of the process, we only require the distribution of the increments of the process to be translation invariant, then the resulting family of processes is the family of processes with stationary increments, which is strictly larger compared to the family of stationary processes. As a price for the relaxation of the condition on the side of processes, a stronger assumption needs to be imposed to the mapping L_H. More precisely, L_H now needs to be invariant under vertical shift of the path: $L_H(I, f) = L_H(I, f+c)$ for any $f \in H$ and $c \in \mathbb{R}$. It has been shown in Shen (2016) that similar results as in Samorodnitsky and Shen (2013a) hold between such random locations and stochastic processes with stationary increments.

Second, different symmetries can be combined together. For instance, self-similarity by itself does not give any new result in nature, due to the Lamperti transformation (see, for example, Embrechts and Maejima (2002)). However, as shown in Shen (2018), when it is combined with the stationarity of the increments, stronger distributional properties can be derived for the random locations which are compatible with both scaling and translation.

It should be pointed out that although many φ-intrinsic random locations are defined using certain continuous-time stochastic processes, such processes are not an indispensable part of the construction. It is in this sense that the current framework is a generalization of those used in previous works, where the definition of the random location does require a continuous-time process.

Example 4.2.3. Let $\{(X_i, Y_i)\}_{i\in\mathbb{Z}}$ be a point process in $\mathbb{R}^2$, where $X_{i+1} - X_i$ are independent and identically distributed positive random variables, and $\{Y_i\}_{i\in\mathbb{Z}}$ is a stationary sequence. Then one can define random locations such as

$$L_1(I) = \sup\{X_i : X_i \in I\}$$

and

$$L_2(I) = \inf\{X_i : X_i \in I, Y_i = \sup_{j:X_j\in I} Y_j\},$$

where the tradition $\inf(\phi) = \sup(\phi) = \infty$ is used. Intuitively, among all the points with the first coordinate in I, L_1 takes the largest first coordinate, while L_2 takes the first coordinate of the point with the largest second coordinate. The infimum in the definition of L_2 is to deal with the case where the supremum is achieved in multiple points. If in addition, we have $P(Y_i = Y_j) = 0$ for all i, j, then the infimum can be removed. It is easy to check that both L_1 and L_2 are stationary intrinsic random locations.

The point process in example 4.2.3 can be regarded as a one-dimensional point process $\{X_i\}_{i\in\mathbb{Z}}$ in which each point X_i also gets a label Y_i in a stationary way. The following example is more "higher dimensional" and geometrical in nature.

Example 4.2.4. Consider a stationary random tessellation of $\mathbb{R}^2$ such as the Gilbert tessellation. Let I' be a fixed compact interval. For any compact intervals I and I', among all the pieces of the tessellation for which the geometric center is located in $I \times I'$, take the one with the largest area. Then the first or the second coordinate of its geometric center is a stationary intrinsic random location indexed by I or I', respectively, if we again follow the tradition to assign value ∞ when no piece has its center in $I \times I'$.

4.3 Main results

We start this section by introducing some preparatory results.

The stability under restriction property in Definition 4.2.1 implies the following trivial comparison lemma.

Lemma 4.3.1. *Let L be an intrinsic random location. Then for any $I_1, I_2 \in \mathcal{I}$, $I_2 \subseteq I_1$ and any $I \subseteq I_2$, $P(L(I_1) \in I) \leqslant P(L(I_2) \in I)$.*

Proof. By stability under restriction, $L(I_1) \in I \subseteq I_2$ implies $L(I_2) = L(I_1) \in I$, hence the result. □

The distribution of a stationary intrinsic random location $L = L(I)$ is absolutely continuous in the interior of the interval I. Indeed, the next proposition does not only show the absolute continuity, but also provides an upper bound for the density. It was first proved in Samorodnitsky and Shen (2013a) for the stationary processes and random locations which are compatible with translation. Here we include a short proof for a modified version for the sake of completeness.

Proposition 4.3.2. *Let L be a stationary intrinsic random location. For any $a < b$ and $0 < \varepsilon < \min\{x-a, b-x\}$,*

$$P(L([a,b]) \in (x, x+\varepsilon]) \leqslant 2\varepsilon \max\{\frac{1}{x-a}, \frac{1}{b-x}\}. \tag{4.4}$$

Proof. Suppose that, to the contrary, (4.4) fails for some a, b, x and ε. That is,

$$P(L([a,b]) \in (x, x+\varepsilon]) > 2\varepsilon \max\{\frac{1}{x-a}, \frac{1}{b-x}\}.$$

Without loss of generality, assume $x - a \leqslant b - x$. Then

$$\begin{aligned} P\left(L([a,x]) \in (x - y_i, x + \varepsilon - y_i]\right) &= P\left(L([a + y_i, x + y_i]) \in (x, x + \varepsilon]\right) \\ &\geqslant P(L([a,b]) \in (x, x + \varepsilon]) \\ &> 2\varepsilon \max\{\frac{1}{x-a}, \frac{1}{b-x}\} \\ &= \frac{2\varepsilon}{x-a} \end{aligned}$$

for $y_i = i\varepsilon$, $i = 1, \ldots, \lfloor \frac{x-a}{\varepsilon} \rfloor$. Since $\frac{x-a}{\varepsilon} \geqslant 1$, $\lfloor \frac{x-a}{\varepsilon} \rfloor \geqslant \frac{x-a}{2\varepsilon}$. Hence we have

$$\begin{aligned} 1 &\geqslant \sum_{i=1}^{\lfloor \frac{x-a}{\varepsilon} \rfloor} P\left(L([a,x]) \in (x - y_i, x + \varepsilon - y_i]\right) \\ &> \lfloor \frac{x-a}{\varepsilon} \rfloor \frac{2\varepsilon}{x-a} \geqslant 1. \end{aligned}$$

Contradiction. A similar contradiction can be derived for the case where $x - a > b - x$. Hence (4.4) is proved. $\square$

As a consequence of Proposition 4.3.2, we also have the following continuity result.

Lemma 4.3.3. *Let L be a stationary intrinsic random location, then for any $u, v \in \mathbb{R}, u < v$, $P(L([a,b]) \in [u,v])$ is continuous in a and b for $a < u$ and $b > v$.*

Proof. By symmetry it suffices to prove $P(L([a,b]) \in [u,v])$ is continuous in a for $a < u$. For $\varepsilon \in \left(0, \frac{u-a}{2}\right)$, we have

$$\begin{aligned} 0 \leqslant& P(L([a+\varepsilon,b]) \in [u,v]) - P(L([a-\varepsilon,b]) \in [u,v]) \\ =& (P(L([a-\varepsilon,b]) \in [a-\varepsilon,u)) - P(L([a+\varepsilon,b]) \in [a+\varepsilon,u))) \\ &- (P(L([a+\varepsilon,b]) \in (v,b]) - P(L([a-\varepsilon,b]) \in (v,b])) \\ &- (P(L([a+\varepsilon,b]) = \infty) - P(L([a-\varepsilon,b]) = \infty)) \\ \leqslant& P(L([a-\varepsilon,b]) \in [a-\varepsilon,u)) - P(L([a+\varepsilon,b]) \in [a+\varepsilon,u)), \end{aligned}$$

where the inequality comes from Lemma 4.3.1 and the consistence of existence property in Definition 4.2.1. Also, by stationarity and Lemma 4.3.1,

$$\begin{aligned}P(L([a+\varepsilon,b])\in[a+\varepsilon,u)) &= P(L([a-\varepsilon,b-2\varepsilon])\in[a-\varepsilon,u-2\varepsilon))\\ &\geqslant P(L([a-\varepsilon,b])\in[a-\varepsilon,u-2\varepsilon)).\end{aligned}$$

Hence

$$\begin{aligned}&P(L([a+\varepsilon,b])\in[u,v])-P(L([a-\varepsilon,b])\in[u,v])\\ \leqslant& P(L([a-\varepsilon,b])\in[a-\varepsilon,u))-P(L([a-\varepsilon,b])\in[a-\varepsilon,u-2\varepsilon))\\ =&P(L([a-\varepsilon,b])\in[u-2\varepsilon,u)).\end{aligned} \tag{4.5}$$

By Proposition 4.3.2, $P(L([a-\varepsilon,b])\in[u-2\varepsilon,u))\leqslant P(L([a,b])\in[u-2\varepsilon,u))\to 0$ as $\varepsilon\to 0$. Thus we conclude that $P(L([a,b])\in[u,v])$ is continuous in a for $a<u$. □

In order to introduce a point process which will play an essential role in deriving the main results, we first show that each intrinsic random location gives a partial order among the potential values of the random location. Similar idea originated in Shen (2016). The proof is however different due to the difference in settings. More precisely, let L be an intrinsic random location. Define the random set $S:=\{x\in\mathbb{R}: x=L(I) \text{ for some } I\in\mathcal{I}\}$. Define a binary relation "$\preceq$" on S:

$$x\preceq y \quad \text{if there exists } I\in\mathcal{I}, \text{ such that } x,y\in I, L(I)=y.$$

Intuitively, $x\preceq y$ if both points are in an interval, and the location falls on y, not on x.

Lemma 4.3.4. *$\preceq$ is a partial order.*

Proof. It is easy to see that $\preceq$ is reflexive. It is antisymmetric since for any I containing x and y and satisfying $L(I)=x$ or $L(I)=y$, $L(I)=L([x\wedge y,x\vee y])$ by the stability under restriction property in Definition 4.2.1. As a result, $x\preceq y$ if and only if $L([x\wedge y,x\vee y])=y$. Finally, if $x\preceq y$ and $y\preceq z$, then by Definition 4.2.1,

$$L([x\wedge y,x\vee y]\cup[y\wedge z,y\vee z])\in\{L([x\wedge y,x\vee y]),L([y\wedge z,y\vee z])\}=\{y,z\}\subset[y\wedge z,y\vee z]$$

Again by the stability under restriction property, we must have $L([x\wedge y,x\vee y]\cup[y\wedge z,y\vee z])=L([y\wedge z,y\vee z])=z$, hence $x\preceq z$. □

For each $x \in S$, define $l_x := \sup\{y \in S : y < x, x \preceq y\}$ and $r_x := \inf\{y \in S : y > x, x \preceq y\}$. Intuitively, l_x and r_x are the farthest locations to the left and to the right of the point x such that no point in S between this location and x has a higher order than x according to $\preceq$. It is easy to see that if in addition, there exists $[a, b] \in \mathcal{I}$ such that $x = L([a, b])$ and $x \in (a, b)$, then $l_x \leqslant a < x$ and $r_x \geqslant b > x$. Thus, for every such x, the point in $\mathbb{R}^3$ defined by $\varepsilon_x := (l_x, x, r_x)$ falls in the area $E := \{(z_1, z_2, z_3) : z_1 < z_2 < z_3\}$. Let $\mathcal{E}$ be the collection of such points, then the (random) counting measure determined by $\mathcal{E}$, denoted by $\xi := \sum_{\varepsilon_x \in \mathcal{E}} \delta_{\varepsilon_x}$, forms a point process in E. Since $l_x < a, r_x > b$ and $x \in (a, b)$ implies $L([a, b]) = x$, $\mathcal{E}$ has at most one point in $(-\infty, a) \times (a, b) \times (b, \infty)$ for any $a, b \in \mathbb{R}, a < b$, hence the point process ξ is σ-finite. Denote by η its control measure, *i.e.*, $\eta(A) = \mathbb{E}(\xi(A))$ for any $A \in \mathcal{B}(E)$, where $\mathcal{B}(E)$ is the Borel σ-field on E.

Theorem 4.3.5. *Let L be a stationary intrinsic random location, and η be the control measure of the point process ξ just defined for L. Then for any $a < u < v < b$,*

$$P(L([a, b]) \in [u, v]) = \eta((-\infty, a) \times (u, v) \times (b, \infty)) = \eta((-\infty, a] \times [u, v] \times [b, \infty)). \quad (4.6)$$

Remark 4.3.6. Theorem 4.3.5 serves for three purposes. First, it builds a connection between the distribution of a stationary intrinsic random location and the control measure of the point process related to it. Second, it also shows that the planes in E with one of the three coordinates fixed are always null sets under η. As a result, one does not need to pay special attention to the openness/closedness of the boundaries of the intervals for the coordinates. Finally, since L is stationary, *i.e.*, $P(L([a, b]) \in [u, v]) = P(L([a + c, b + c]) \in [u+c, v+c])$ for all $a \leqslant u < v \leqslant b$ and $c \in \mathbb{R}$, and the sets of the form $(-\infty, a] \times [u, v] \times [b, \infty)$ generate $\mathcal{B}(E)$, the Borel σ-field on E, the measure η is also invariant under translation along the direction $(1, 1, 1)$. We formulate this result as the following corollary, the proof of which is obvious and omitted.

Corollary 4.3.7. *Let $A \in \mathcal{B}(E)$. Then $\eta(A) = \eta(A + c)$ for any $c \in \mathbb{R}$, where $A + c = \{(z_1, z_2, z_3) : (z_1 - c, z_2 - c, z_3 - c) \in A\}$.*

Proof of Theorem 4.3.5. If $x = L([a, b]) \in [u, v]$, then $x \in S$, $l_x \leqslant a$, and $r_x \geqslant b$. Note that it is possible that $l_x = a$ (resp. $r_x = b$), since a (resp. b) can be the limit of an increasing (resp. decreasing) sequence of points in S with higher orders than x according to $\preceq$, while the endpoint itself is not in S or does not have a higher order than x. Meanwhile, if there exists $x \in [u, v] \cap S$ such that $l_x < a$ and $r_x > b$, then we must have $x = L([a, b])$. Therefore,

$$\begin{aligned} &P(L([a-\varepsilon, b+\varepsilon]) \in [u, v]) \leqslant \eta((-\infty, a) \times [u, v] \times (b, \infty)) \leqslant P(L([a, b]) \in [u, v]) \\ \leqslant &\eta((-\infty, a] \times [u, v] \times [b, \infty)) \leqslant P(L([a+\varepsilon, b-\varepsilon]) \in [u, v]). \end{aligned}$$

The control measure η appears in the above expression because there can be at most one point in $\mathcal{E}$ in the area $(-\infty, a] \times [u, v] \times [b, \infty)$. In this case the expectation coincides with the corresponding probability.

By Lemma 4.3.3,

$$\lim_{\varepsilon \downarrow 0} P(L([a-\varepsilon, b+\varepsilon]) \in [u, v]) = \lim_{\varepsilon \downarrow 0} P(L([a+\varepsilon, b-\varepsilon]) \in [u, v]),$$

hence we must have

$$\eta((-\infty, a) \times [u, v] \times (b, \infty)) = P(L([a, b]) \in [u, v]) = \eta((-\infty, a] \times [x, y] \times [b, \infty)). \quad (4.7)$$

Finally, by Proposition 4.3.2, $L([a, b])$ is continuously distributed on (a, b), hence $P(L([a, b]) \in [u, v])$ is continuous in u and v, so is $\eta((-\infty, a) \times [u, v] \times (b, \infty))$. Therefore, $\eta((-\infty, a) \times [u, v] \times (b, \infty)) = \eta((-\infty, a) \times (u, v) \times (b, \infty))$. $\square$

Our last preparation before proceeding to prove the main result of this chapter is the following proposition.

For a stationary intrinsic random location L, $a < u < v < b$ and any $\varepsilon > 0$, define

$$M_{\varepsilon,[u,v]} = P(L([a, b]) \in [a, a+\varepsilon), L([a+\varepsilon, b+\varepsilon]) \in [u, v])$$

and

$$N_{\varepsilon,[u,v]} = P(L([a+\varepsilon, b+\varepsilon]) \in (b, b+\varepsilon], L([a, b]) \in [u, v]).$$

Further define $\mu_{\varepsilon,[u,v]}$ to be the conditional distribution of $L([a+\varepsilon, b+\varepsilon])$ given $L([a, b]) \in [a, a+\varepsilon)$ and $L([a+\varepsilon, b+\varepsilon]) \in [u, v]$, and $\nu_{\varepsilon,[u,v]}$ to be the conditional distribution of $L([a, b])$ given $L([a+\varepsilon, b+\varepsilon]) \in (b, b+\varepsilon]$ and $L([a, b]) \in [u, v]$, if $M_{\varepsilon,[u,v]}$ and $N_{\varepsilon,[u,v]}$ are strictly positive. If $M_{\varepsilon,[u,v]} = 0$ or $N_{\varepsilon,[u,v]} = 0$, define the corresponding $\mu_{\varepsilon,[u,v]}$ or $\nu_{\varepsilon,[u,v]}$ to be the null measure.

Let $\mu^{(a,b)}$ and $\nu^{(a,b)}$ be measures on (a,b) (equipped with the Borel σ-field) given by

$$\mu^{(a,b)}([w,y)) = \eta\left((z_1,z_2,z_3) : z_1 \in [a,a+1), z_2 \in [z_1+w-a, z_1+y-a), z_3 \in (z_1+b-a,\infty)\right) \tag{4.8}$$

and

$$\nu^{(a,b)}([w,y)) = \eta\left((z_1,z_2,z_3) : z_1 \in (-\infty, z_3+a-b), z_2 \in [z_3+w-b, z_3+y-b), z_3 \in (b,b+1]\right) \tag{4.9}$$

for all $w,y \in (a,b)$, $w<y$, where η is the control measure of the point process ξ corresponding to L as defined previously. Denote by $\mu^{(a,b)}|_{[u,v]}$ and $\nu^{(a,b)}|_{[u,v]}$ the restriction of the measures $\mu^{(a,b)}$ and $\nu^{(a,b)}$ on $[u,v]$, respectively.

Proposition 4.3.8. *Let L be a stationary intrinsic random location. For $a<u<v<b$, let $M_{\varepsilon,[u,v]}$, $N_{\varepsilon,[u,v]}$, $\mu_{\varepsilon,[u,v]}$ and $\nu_{\varepsilon,[u,v]}$ be defined as above. Then $\frac{1}{\varepsilon}M_{\varepsilon,[u,v]}\mu_{\varepsilon,[u,v]}$ and $\frac{1}{\varepsilon}N_{\varepsilon,[u,v]}\nu_{\varepsilon,[u,v]}$ converge vaguely as $\varepsilon \to 0$ to $\mu^{(a,b)}|_{[u,v]}$ and $\nu^{(a,b)}|_{[u,v]}$, respectively.*

Proof. By symmetry it suffices to prove the convergence for $\frac{1}{\varepsilon}M_{\varepsilon,[u,v]}\mu_{\varepsilon,[u,v]}$ as $\varepsilon \to 0$. For any $\varepsilon \in (0,b-a)$, define measure λ_ε on $[a+\varepsilon,b)$ by

$$\lambda_\varepsilon(A) = P(L([a,b]) \in [a,a+\varepsilon), L([a+\varepsilon,b+\varepsilon]) \in A), \quad A \in \mathcal{B}([a+\varepsilon,b)),$$

then it is easy to see that for any $\varepsilon < u-a$ and $A' \in \mathcal{B}([u,v])$,

$$M_{\varepsilon,[u,v]}\mu_{\varepsilon,[u,v]}(A') = P(L([a,b]) \in [a,a+\varepsilon), L([a+\varepsilon,b+\varepsilon]) \in A') = \lambda_\varepsilon(A').$$

Hence it suffices to prove that $\frac{1}{\varepsilon}\lambda_\varepsilon([w,y))$ converges to $\mu^{(a,b)}([w,y))$ for any $w,y \in (a,b), w < y$.

Note that $L([a,b]) \in [a,a+\varepsilon)$ and $L([a+\varepsilon,b+\varepsilon]) \in [w,y)$ implies that there exists a point $x \in [w,y) \cap S$, such that $l_x \in [a,a+\varepsilon]$ and $r_x \in [b+\varepsilon,\infty)$. Meanwhile, the existence of a $x \in [w,y)\cap S$ satisfying $l_x \in (a,a+\varepsilon)$ and $r_x \in (b+\varepsilon,\infty)$ would guarantee that $L([a,b]) \in [a,a+\varepsilon)$ and $L([a+\varepsilon,b+\varepsilon]) \in [w,y)$. Therefore, we have

$$\eta((a,a+\varepsilon)\times[w,y)\times(b+\varepsilon,\infty)) \leqslant \lambda_\varepsilon([w,y)) \leqslant \eta([a,a+\varepsilon]\times[w,y)\times[b+\varepsilon,\infty)).$$

By Theorem 4.3.5, the boundaries of the intervals are negligible under η. Hence

$$\lambda_\varepsilon([w,y)) = \eta([a,a+\varepsilon)\times[w,y)\times(b+\varepsilon,\infty)).$$

For $\varepsilon = \frac{1}{n}$, $n\in\mathbb{N}$, by Corollary 4.3.7, we have

$$\begin{aligned}
&\frac{1}{\varepsilon}\eta([a,a+\varepsilon)\times[w,y)\times(b+\varepsilon,\infty))\\
=&n\eta\left(\left[a,a+\frac{1}{n}\right)\times[w,y)\times\left(b+\frac{1}{n},\infty\right)\right)\\
=&\sum_{i=0}^{n-1}\eta\left(\left[a+\frac{i}{n},a+\frac{i+1}{n}\right)\times\left[w+\frac{i}{n},y+\frac{i}{n}\right)\times\left(b+\frac{i+1}{n},\infty\right)\right).
\end{aligned}$$

Note that the set

$$\bigcup_{i=0}^{n-1}\left(\left[a+\frac{i}{n},a+\frac{i+1}{n}\right)\times\left[w+\frac{i}{n},y+\frac{i}{n}\right)\times\left(b+\frac{i+1}{n},\infty\right)\right)$$

contains

$$\left\{(z_1,z_2,z_3): z_1\in[a,a+1), z_2\in\left[z_1+w-a, z_1+y-a-\varepsilon\right), z_3\in\left(z_1+b-a+\varepsilon,\infty\right)\right\},$$

and is contained in

$$\left\{(z_1,z_2,z_3): z_1\in[a,a+1), z_2\in\left[z_1+w-a-\varepsilon, z_1+y-a\right), z_3\in\left(z_1+b-a,\infty\right)\right\}.$$

Moreover, these bounds naturally extend to the case where ε is any positive rational number. Indeed, let $\varepsilon = \frac{m}{n}$, $m,n\in\mathbb{N}$. Then a similar reasoning as above leads to

$$\begin{aligned}
&\eta\left(\left\{(z_1,z_2,z_3): z_1\in[a,a+m), z_2\in\left[z_1+w-a, z_1+y-a-\varepsilon\right), z_3\in\left(z_1+b-a+\varepsilon,\infty\right)\right\}\right)\\
\leqslant&\frac{m}{\varepsilon}\eta([a,a+\varepsilon)\times[w,y)\times(b+\varepsilon,\infty))\\
\leqslant&\eta\left(\left\{(z_1,z_2,z_3): z_1\in[a,a+m), z_2\in\left[z_1+w-a-\varepsilon, z_1+y-a\right), z_3\in\left(z_1+b-a,\infty\right)\right\}\right).
\end{aligned}$$

Then by Corollary 4.3.7,

$$\begin{aligned}
&\eta\left(\left\{(z_1,z_2,z_3): z_1\in[a,a+1), z_2\in\left[z_1+w-a, z_1+y-a-\varepsilon\right), z_3\in\left(z_1+b-a+\varepsilon,\infty\right)\right\}\right)\\
\leqslant&\frac{1}{\varepsilon}\eta([a,a+\varepsilon)\times[w,y)\times(b+\varepsilon,\infty))\\
\leqslant&\eta\left(\left\{(z_1,z_2,z_3): z_1\in[a,a+1), z_2\in\left[z_1+w-a-\varepsilon, z_1+y-a\right), z_3\in\left(z_1+b-a,\infty\right)\right\}\right)
\end{aligned}$$

for any positive rational $\varepsilon > 0$. Since $\frac{1}{\varepsilon}\eta([a, a+\varepsilon) \times [w, y) \times (b+\varepsilon, \infty))$ is continuous in ε, by the continuity of measure, we have

$$\frac{1}{\varepsilon}\eta([a, a+\varepsilon) \times [w, y) \times (b+\varepsilon, \infty))$$
$$\to \eta\left((z_1, z_2, z_3) : z_1 \in [a, a+1), z_2 \in [z_1 + w - a, z_1 + y - a), z_3 \in (z_1 + b - a, \infty)\right)$$

as $\varepsilon \to 0$. This is exactly $\mu^{(a,b)}|_{[u,v]}([w, y))$ defined in (4.8). The convergence to $\nu^{(a,b)}|_{[u,v]}$ can be shown symmetrically. $\square$

We now prove the main result of this chapter. Denote by $\mathring{I}$ the interior of the compact interval I, and let $\dot{\varphi}^t(x) = \frac{\partial \varphi(x,t)}{\partial t}|_{x,t}$. In addition, for any flow φ on $\mathbb{R}$ satisfying Assumptions (4.1) and (4.2) and a given interval $[a, b]$ between two consecutive extended fixed points of φ, we define measures $\mu_\varphi^{(a,b)}$ and $\nu_\varphi^{(a,b)}$ as the pull-backs of $\mu^{(a,b)}$ and $\nu^{(a,b)}$ under the bijection τ. More precisely, assuming that τ is increasing, then define measure $\mu_\varphi^{(a,b)}$ on (a, b) by

$$\mu_\varphi^{(a,b)}([w, y)) := \eta\left((z_1, z_2, z_3) : \tau(z_1) \in [\tau(a), \tau(a+1)),\right.$$
$$\left.\tau(z_2) \in [\tau(z_1) + \tau(w) - \tau(a), \tau(z_1) + \tau(y) - \tau(a)), \tau(z_3) \in (\tau(z_1) + \tau(b) - \tau(a), \infty)\right)$$

for all $w, y \in (a, b), w < y$. $\nu_\varphi^{(a,b)}$ is defined similarly. The case where τ is decreasing is symmetric.

Theorem 4.3.9. *Let φ be a flow on $\mathbb{R}$ satisfying Assumptions (4.1) and (4.2), and L be a φ-stationary intrinsic random location. Let α, β be two consecutive points in $\bar{\Phi}_0$. Then for any $I = [a, b] \subset (\alpha, \beta)$, the distribution of $L(I)$ is absolutely continuous in $\mathring{I}$, and it has a càdlàg density function, denoted by f. Moreover, f satisfies*

$$\dot{\varphi}^0(x_2)f(x_2) - \dot{\varphi}^0(x_1)f(x_1) = \nu_\varphi^{(a,b)}((x_1, x_2]) - \mu_\varphi^{(a,b)}((x_1, x_2])$$

for any $x_1 \leqslant x_2$, $x_1, x_2 \in \mathring{I}$.

Proof. By Remark (4.2.2), it suffices to prove the result for $\varphi^t(x) = x + t$, where $\dot{\varphi}^0(x)$ becomes the constant 1, and $\mu_\varphi^{(a,b)}$ and $\nu_\varphi^{(a,b)}$ are simply $\mu^{(a,b)}$ and $\nu^{(a,b)}$ defined before Proposition 4.3.8.

Let $C_C^\infty((u,v))$ be the set of smooth functions from $\bar{\mathbb{R}}$ to $\mathbb{R}$ with support in (u,v), and g be any function in $C_C^\infty((u,v))$. By stationarity, for any $\varepsilon > 0$, we have

$$\mathbb{E}[g(L([a+\varepsilon, b+\varepsilon]))] = \mathbb{E}[g(L([a,b]) + \varepsilon)],$$

hence

$$\mathbb{E}[g(L([a+\varepsilon, b+\varepsilon]))] - \mathbb{E}[g(L([a,b]))] = \mathbb{E}[g(L([a,b]) + \varepsilon)] - \mathbb{E}[g(L([a,b]))]. \tag{4.10}$$

Denote by F the distribution of $L([a,b])$, then the right hand side of (4.10) can be rewritten as

$$\int_a^b (g(s+\varepsilon) - g(s))\mathrm{d}F(s).$$

Since g is smooth and compactly supported, g' is bounded, hence g is uniformly Lipschitz. As a result, Dominated Convergence Theorem applies and we have

$$\begin{aligned}
&\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left(\mathbb{E}[g(L([a,b]) + \varepsilon)] - \mathbb{E}[g(L([a,b]))]\right) \\
=&\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_a^b (g(s+\varepsilon) - g(s))\mathrm{d}F(s) \\
=&\int_a^b g'(s)\mathrm{d}F(s) = \int_u^v g'(s)\mathrm{d}F(s).
\end{aligned} \tag{4.11}$$

For the left hand side of (4.10), we have

$$\begin{aligned}
&\mathbb{E}[g(L([a+\varepsilon, b+\varepsilon]))] - \mathbb{E}[g(L([a,b]))] \\
=&\mathbb{E}[g(L([a+\varepsilon, b+\varepsilon])); L([a,b]) \in [a, a+\varepsilon)] - \mathbb{E}[g(L([a,b])); L([a+\varepsilon, b+\varepsilon]) \in (b, b+\varepsilon]] \\
&+ \mathbb{E}[g(L([a+\varepsilon, b+\varepsilon])); L([a,b]) \in [a+\varepsilon, b]] - \mathbb{E}[g(L([a,b])); L([a+\varepsilon, b+\varepsilon]) \in [a+\varepsilon, b]],
\end{aligned}$$

where the notation $\mathbb{E}[X; A]$ stands for the expectation of X restricted on A, *i.e.*, $\mathbb{E}[X;A] = \mathbb{E}[X\mathbf{1}_A]$. Since g is supported on $[u,v] \subset (a,b)$, for $\varepsilon < u - a$,

$$\begin{aligned}
&\mathbb{E}[g(L([a+\varepsilon, b+\varepsilon])); L([a,b]) \in [a+\varepsilon, b]] \\
=&\mathbb{E}[g(L([a+\varepsilon, b+\varepsilon])); L([a,b]) \in [a+\varepsilon, b], L([a+\varepsilon, b+\varepsilon]) \in [a+\varepsilon, b]] \\
=&\mathbb{E}[g(L([a,b])); L([a,b]) \in [a+\varepsilon, b], L([a+\varepsilon, b+\varepsilon]) \in [a+\varepsilon, b]] \\
=&\mathbb{E}[g(L([a,b])); L([a+\varepsilon, b+\varepsilon]) \in [a+\varepsilon, b]],
\end{aligned}$$

where the equality in the middle comes from the stability under restriction property of L. Therefore, we have

$$\begin{aligned}
&\mathbb{E}[g(L([a+\varepsilon,b+\varepsilon]))]-\mathbb{E}[g(L([a,b]))]\\
=&\mathbb{E}[g(L([a+\varepsilon,b+\varepsilon]));L([a,b])\in[a,a+\varepsilon)]-\mathbb{E}[g(L([a,b]));L([a+\varepsilon,b+\varepsilon])\in(b,b+\varepsilon]]\\
=&\mathbb{E}[g(L([a+\varepsilon,b+\varepsilon]));L([a,b])\in[a,a+\varepsilon),L([a+\varepsilon,b+\varepsilon])\in[u,v]]\\
&-\mathbb{E}[g(L([a,b]));L([a+\varepsilon,b+\varepsilon])\in(b,b+\varepsilon],L([a,b])\in[u,v]] \qquad (4.12)\\
=&\int_u^v g(s)M_{\varepsilon,[u,v]}\mathrm{d}\mu_{\varepsilon,[u,v]}(s)-\int_u^v g(s)N_{\varepsilon,[u,v]}\mathrm{d}\nu_{\varepsilon,[u,v]}(s). \qquad (4.13)
\end{aligned}$$

Combining (4.13) with Proposition 4.3.8, we have

$$\begin{aligned}
&\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(\mathbb{E}[g(L([a+\varepsilon,b+\varepsilon]))]-\mathbb{E}[g(L([a,b]))]\right)\\
=&\int_u^v g(s)\mathrm{d}(\mu^{(a,b)}-\nu^{(a,b)})(s),
\end{aligned}$$

hence by (4.10) and (4.11),

$$\int_u^v g'(s)\mathrm{d}F(s)=\int_u^v g(s)\mathrm{d}(\mu^{(a,b)}-\nu^{(a,b)})(s)$$

for all $g\in C_C^\infty((u,v))$. This means, the signed measure on (u,v) given by $\mathrm{d}(\nu^{(a,b)}-\mu^{(a,b)})(s)$ is a derivative of the measure given by $dF(s)$ in the sense of generalized function. (Generalized functions are alternatively called distributions. In this chapter we would use the term "generalized functions" to avoid confusion with the probability distributions of the random locations. Readers are referred to Barros-Neto (1973) for an overview of the generalized functions.) Consequently, we have

$$F((u,x])=\int_u^x \nu^{(a,b)}((u,s])-\mu^{(a,b)}((u,s])+c\ \mathrm{d}s$$

for all $x\in(u,v)$ and some constant c. As a result, F is differentiable on (u,v); its derivative, denoted as f, satisfies

$$f(x)=\nu^{(a,b)}((u,x])-\mu^{(a,b)}((u,x])+c,$$

for almost all x in (u, v). It is easy to see that f is càdlàg on (u, v). Taking $u \downarrow a$ and $v \uparrow b$ shows that F is absolutely continuous on (a, b), and $f(x) = \nu^{(a,b)}((x_0, x]) - \mu^{(a,b)}((x_0, x]) + c$, $x \in (a, b)$ is a càdlàg version of the density of F on (a, b). Here x_0 is an arbitrary fixed point in (a, b), and $\nu^{(a,b)}((x_0, x])$ (resp. $\mu^{(a,b)}((x_0, x])$) is understood as $-\nu^{(a,b)}((x, x_0])$ (resp. $-\mu^{(a,b)}((x, x_0])$) when $x < x_0$. Moreover, taking $x = x_0$ leads to $c = f(x_0)$. Therefore, we have

$$f(x) = f(x_0) + \nu^{(a,b)}((x_0, x]) - \mu^{(a,b)}((x_0, x]), \quad x \in (a, b),$$

or alternatively,

$$f(x_2) - f(x_1) = \nu^{(a,b)}((x_1, x_2]) - \mu^{(a,b)}((x_1, x_2]), \quad x_1, x_2 \in (a, b), x_1 \leqslant x_2.$$

We complete the proof by applying the change of variable given in Remark (4.2.2) for general flow φ satisfying Assumptions (4.1) and (4.2). □

A simple rewrite of the result in Theorem 4.3.9 gives rise to a conservation law when the interval of interest moves according to the flow φ, which indicates clearly that what we obtained is, by its nature, a Noether theorem. More precisely, consider a given interval $[a_0, b_0]$ between two consecutive extended fixed points, α and β, of φ. Let L be a φ-stationary intrinsic random location. For any $x \in (\alpha, \beta)$ and $t \in \mathbb{R}$ such that $x \in (\varphi^t(a_0), \varphi^t(b_0))$, denote by $f_t(x)$ the density of $L([\varphi^t(a_0), \varphi^t(b_0)])$ at point x. Moreover, define the single-variable function $K(y) = \nu_\varphi^{(a_0,b_0)}((x_0, y]) - \mu_\varphi^{(a_0,b_0)}((x_0, y])$ for $y \in (a_0, b_0)$, where $x_0 \in (a_0, b_0)$, and $\nu_\varphi^{(a_0,b_0)}((x_0, y])$ (resp. $\nu_\varphi^{(a_0,b_0)}((x_0, y])$) is understood as $-\nu_\varphi^{(a_0,b_0)}((y, x_0])$ (resp. $-\nu_\varphi^{(a_0,b_0)}((y, x_0])$) for $y < x_0$. Then we have

Corollary 4.3.10.

$$\dot{\varphi}^0(x) f_t(x) - K((\varphi^t)^{-1}(x))$$

is a constant for t satisfying $x \in (\varphi^t(a_0), \varphi^t(b_0))$.

Proof. Since L is φ-stationary, by the change of variable formula and (4.3),

$$\dot{\varphi}^0(x) f_t(x) = f_t'(\tau(x)) = f_0'(\tau((\varphi^t)^{-1}(x))) = \dot{\varphi}^0((\varphi^t)^{-1}(x)) f_0((\varphi^t)^{-1}(x)),$$

where f_t' is the density function of the stationary intrinsic random location L' defined by

$$L'(I) = \tau(L(\tau^{-1}(I)))$$

on interval $I = [\tau(a_0) + t, \tau(b_0) + t]$.

By Theorem 4.3.9, we have

$$\begin{aligned}
&\dot{\varphi}^0((\varphi^t)^{-1}(x))f_0((\varphi^t)^{-1}(x)) \\
=&\dot{\varphi}^0(x_0)f_0(x_0) + \nu_\varphi^{(a_0,b_0)}((x_0, (\varphi^t)^{-1}(x)]) - \mu_\varphi^{(a_0,b_0)}((x_0, (\varphi^t)^{-1}(x)]) \\
=&\dot{\varphi}^0(x_0)f_0(x_0) + K((\varphi^t)^{-1}(x)),
\end{aligned}$$

hence

$$\dot{\varphi}^0(x)f_t(x) - K((\varphi^t)^{-1}(x)) = \dot{\varphi}^0(x_0)f_0(x_0),$$

which is a constant for t satisfying $x \in (\varphi^t(a_0), \varphi^t(b_0))$. □

Also as a consequence of Theorem 4.3.9, we have the following result, which shows that the total variation of $\dot{\varphi}^0(x)f(x)$ is bounded by its values and limits. Special cases for stationary processes, processes with stationary increments and self-similar processes with stationary increments have been studied in Shen (2013, 2016); Shen (2018).

Denote by $\mathrm{TV}^+_{(u,v)}(f), \mathrm{TV}^-_{(u,v)}(f)$ and $\mathrm{TV}_{(u,v)}(f)$ the positive variation, negative variation and total variation of the function f on the interval (u, v), respectively. That is,

$$\mathrm{TV}^+_{(u,v)}(f) := \sup_{u<x_1<\cdots<x_n<v} \sum_{i=1}^{n-1}(f(x_{i+1}) - f(x_i))^+,$$

$$\mathrm{TV}^-_{(u,v)}(f) := \sup_{u<x_1<\cdots<x_n<v} \sum_{i=1}^{n-1}(f(x_{i+1}) - f(x_i))^-,$$

and

$$\mathrm{TV}_{(u,v)}(f) := \sup_{u<x_1<\cdots<x_n<v} \sum_{i=1}^{n-1}|f(x_{i+1}) - f(x_i)|,$$

where the suprema are taken over all the partitions of (u, v). Define $f(x-) = \lim_{y\uparrow x} f(y)$ to be the left limit of a càdlàg function.

Corollary 4.3.11. *Let φ be a flow on $\mathbb{R}$ satisfying Assumptions (4.1) and (4.2), and L be a φ-stationary intrinsic random location. Let α, β be two consecutive points in $\bar{\Phi}_0$. Then for any $I = [a, b] \in \mathcal{I}$ such that $I \subset (\alpha, \beta)$ and $u, v \in (a, b)$, $u < v$, the càdlàg density function f of $L(I)$ on (a, b) satisfies*

$$\mathrm{TV}^{+}_{(u,v)}(\dot{\varphi}^0(\cdot) f(\cdot)) \leqslant \dot{\varphi}^0(v) \min\{f(v), f(v-)\}, \tag{4.14}$$

$$\mathrm{TV}^{-}_{(u,v)}(\dot{\varphi}^0(\cdot) f(\cdot)) \leqslant \dot{\varphi}^0(u) \min\{f(u), f(u-)\}, \tag{4.15}$$

and

$$\mathrm{TV}_{(u,v)}(\dot{\varphi}^0(\cdot) f(\cdot)) \leqslant \dot{\varphi}^0(u) \min\{f(u), f(u-)\} + \dot{\varphi}^0(v) \min\{f(v), f(v-)\}. \tag{4.16}$$

Remark 4.3.12. One of the main results in Samorodnitsky and Shen (2013a) and Shen (2016) was the so-called "total variation constraints", which states that the density f of the distribution of a random location compatible with translation, for stationary or stationary increment processes, satisfies

$$\mathrm{TV}^{+}_{(u,v)}(f) \leqslant \min\{f(v), f(v-)\},$$

$$\mathrm{TV}^{-}_{(u,v)}(f) \leqslant \min\{f(u), f(u-)\},$$

and

$$\mathrm{TV}_{(u,v)}(f) \leqslant \min\{f(u), f(u-)\} + \min\{f(v), f(v-)\}.$$

Now it becomes clear that they are special cases of Corollary 4.3.11 where $\varphi^t(x) = x + t$, hence consequences of the Noether theorem for random locations.

The proof of Corollary 4.3.11 mainly relies on the following proposition, which gives upper bounds for the mass that $\mu^{(a,b)}$ and $\nu^{(a,b)}$ can put on an interval. For simplicity, the proposition is presented using stationary intrinsic random locations. It is straightforward to extend all the definitions and results to general φ-stationary intrinsic random locations if needed.

Proposition 4.3.13. *Let L be a stationary intrinsic random location. Under the same setting as before, $\mu^{(a,b)}([u, v]) \leqslant f(u-)$, $\nu^{(a,b)}([u, v]) \leqslant f(v)$.*

Proof. Take $v' \in (v, b)$, then

$$\mu^{(a,b)}([u,v]) \leqslant \mu^{(a,b)}([u,v')) = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} M_{\varepsilon,[u,v']}\mu_{\varepsilon,[u,v']}([u,v')) \leqslant \limsup_{\varepsilon\to 0} \frac{1}{\varepsilon} M_{\varepsilon,[u,v']},$$

since $\mu_{\varepsilon,[u,v']}$ is a probability measure.

On the other hand, by definition,

$$\begin{aligned} M_{\varepsilon,[u,v']} &= P(L([a,b]) \in [a, a+\varepsilon), L[a+\varepsilon, b+\varepsilon] \in [u, v')) \\ &\leqslant P(L([a,b]) \in [a, a+\varepsilon), L[a+\varepsilon, b] \in [u, v')) \\ &= P(L([a+\varepsilon, b]) \in [u, v')) - P(L([a,b]) \in [u, v')). \end{aligned}$$

Moreover, recall that by (4.5) and Lemma 4.3.1, we have, for ε small enough,

$$P(L([a+\varepsilon, b]) \in [u, v')) - P(L([a,b]) \in [u, v')) \leqslant P(L([a,b]) \in [u-\varepsilon, u)).$$

Hence

$$\limsup_{\varepsilon\to 0} \frac{1}{\varepsilon} M_{\varepsilon,[u,v']} \leqslant \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} P(L([a,b]) \in [u-\varepsilon, u)) = f(u-).$$

The bound for $\nu^{(a,b)}([u,v])$ can be derived symmetrically. □

Proof of Corollary 4.3.11. For simplicity we only prove the result for $\varphi^t(x) = x + t$. The general case then follows by the change of variable discussed in Remark 4.2.2.

In this case, by Theorem 4.3.9, we have

$$f(x_2) - f(x_1) = \nu^{(a,b)}((x_1, x_2]) - \mu^{(a,b)}((x_1, x_2])$$

for any $x_1, x_2 \in [u, v], x_1 < x_2$.

Hence

$$(f(x_2) - f(x_1))^+ \leqslant \nu((x_1, x_2]).$$

Therefore, for any partition $u < x_1 < \cdots < x_n < v$ of (u, v),

$$\sum_{i=1}^{n-1} (f(x_{i+1}) - f(x_i))^+ \leqslant \nu((u, v]) \leqslant f(v)$$

by Proposition 4.3.13. Taking supremum over all partitions of (u, v) on the left hand side leads to

$$\mathrm{TV}^{+}_{(u,v)}(f) \leqslant f(v).$$

Moreover, since f is càdlàg, we also have

$$\mathrm{TV}^{+}_{(u,v)}(f) = \lim_{y \uparrow v} \mathrm{TV}^{+}_{(u,y)}(f) \leqslant \lim_{y \uparrow v} f(y) = f(v-),$$

hence

$$\mathrm{TV}^{+}_{(u,v)}(f) \leqslant \min\{f(v-), f(v)\}.$$

The result for $\mathrm{TV}^{-}_{(u,v)}(f)$ can be proved symmetrically. Finally, adding the two inequalities (4.14) and (4.15) gives (4.16). □

4.4 Boundary and near-boundary behavior

In Section 4.3, we mainly focus on the behavior of the distribution of a φ-stationary intrinsic random location L in the interior of the interval of interest $I = [a, b]$. We have seem that a càdlàg density, denoted by f, exists on (a, b). Indeed, (4.4) gives an upper bound for $f(x), x \in (a, b)$. Such a bound, however, diverges as the x approaches a or b. Moreover, there may also be point masses on the two boundaries of the interval, which were not studied in Section 4.3. Now we provide these missing pieces by discussing the boundary and near-boundary behavior of L.

For simplicity, in this section we always assume that L is a stationary intrinsic random location. The results can be easily generalized to the case where L is φ-stationary.

Recall that $S = \{x \in \mathbb{R} : x = L(I) \text{ for some } I \in \mathcal{I}\}$, $l_x = \sup\{y \in S : y < x, x \preceq y\}$ and $r_x = \inf\{y \in S : y > x, x \preceq y\}$, where "$\preceq$" is the partial order determined by L. For any $T > 0$, define $S_{l,T} := \{x \in S : l_x = x, r_x \geqslant x + T\}$ and $S_{r,T} := \{x \in S : r_x = x, l_x \leqslant x - T\}$. Denote by $Leb(\cdot)$ the Lebesgue measure on $\mathbb{R}$. Then we have

Proposition 4.4.1. *For $I = [a, b]$,*

$$P(L(I) = a) = P(a \in S_{l,b-a}) = \mathbb{E}(Leb(S_{l,b-a} \cap [0, 1))), \tag{4.17}$$

$$P(L(I) = b) = P(b \in S_{r,b-a}) = \mathbb{E}(Leb(S_{r,b-a} \cap [0, 1))). \tag{4.18}$$

Proof. By symmetry, it suffices to prove 4.17. Note that for $x \in S$, $l_x = x, r_x > x + b - a$ implies that $L([x, x + b - a]) = x$, which in turn implies that $l_x \leqslant x, r_x \geqslant x + b - a$. Hence we have

$$P(a \in S, l_a = a, r_a > b) \leqslant P(L([a, b]) = a) \leqslant P(a \in S, l_a \leqslant a, r_a \geqslant b).$$

However,

$$P(a \in S, l_a < a, r_a \geqslant b) = \eta((-\infty, a) \times \{a\} \times [b, \infty)) = 0,$$

since the plane with the second coordinate fixed is a $\eta-$null set, according to Theorem 4.3.5. Therefore for $\varepsilon > 0$,

$$\begin{aligned} &P(L([a, b + \varepsilon]) = a) \leqslant P(a \in S, l_a = a, r_a > b) \\ &\qquad \leqslant P(L([a, b]) = a) \leqslant P(a \in S, l_a = a, r_a \geqslant b) \leqslant P(L([a, b - \varepsilon]) = a). \end{aligned} \tag{4.19}$$

Next, by a similar reasoning as in the proof of Lemma 4.3.3, $P(L([a, b]) = a)$ is continuous in b for $b > a$. Indeed, for $b' > b$,

$$\begin{aligned} &P(L([a, b]) = a) - P(L([a, b']) = a) \\ =&P(L([a, b']) \in (b, b')) - [P(L([a, b] \in (a, b))) - P(L([a, b']) \in (a, b))] \\ &- [P(L([a, b]) = b) - P(L([a, b']) = b')] - [P(L([a, b]) = \infty) - P(L([a, b']) = \infty)] \\ \leqslant&P(L([a, b']) \in (b, b')) \leqslant P(L([a + b' - b, b']) \in (b, b')) = P(L([a, b]) \in (2b - b', b)) \to 0 \end{aligned}$$

as $b' \downarrow b$, where the inequalities follow from Lemma 4.3.1, and the convergence is due to the existence of a density of $L([a, b])$ on (a, b) given by Theorem 4.3.9.

Thus, we have $P(L(I) = a) = P(a \in S_{l,b-a})$ by taking $\varepsilon \to 0$ in (4.19) and applying the continuity result proved above. The second equality in (4.17) then follows naturally by the fact that $P(x \in S_{l,b-a})$ is a constant in x, due to the equality $P(L(I) = a) = P(a \in S_{l,b-a})$ and the fact that L is a stationary intrinsic random location. $\square$

We now turn to the near-boundary behavior of the distribution of $L(I)$, $I = [a, b]$. More precisely, we would like to know when the density $f(x)$ will explode as x approaches the boundaries of the interval I. Clearly, by Theorem 4.3.9, $\lim_{x\downarrow a} f(x) = \infty$ if and only if $\mu^{(a,b)}((a, x_0]) = \infty$ for some (equivalently, any) $x_0 \in (a, b)$. By (4.8), this means

$$\eta(\{(z_1, z_2, z_3) : z_1 \in [a, a+1), z_2 \in (z_1, z_1 + x_0 - a], z_3 \in (z_1 + b - a, \infty)\}) = \infty. \quad (4.20)$$

Similarly, $\lim_{x\uparrow b} f(x) = \infty$ if any only if

$$\eta(\{(z_1, z_2, z_3) : z_1 \in (-\infty, z_3 + a - b), z_2 \in (z_3 + x_0 - b, z_3), z_3 \in (b, b+1]\}) = \infty. \quad (4.21)$$

Define set

$$S_1 := \{x \in [0, 1) : l_x < x, r_x > x, r_x - l_x > b - a\},$$

then (4.20) or (4.21) would require $\mathbb{E}(|S_1|) = \infty$, where $|S_1|$ is the cardinal number of S_1, with the convention that $|\cdot| = \infty$ for any infinite set. Indeed, assume (4.20) holds for example. Then by Corollary 4.3.7 and taking $x_0 = \min\{a + \frac{1}{2}, b\}$, (4.20) holds if and only if

$$2\eta(\{(z_1, z_2, z_3) : z_1 \in [a, a + 1/2)\,, z_2 \in (z_1, z_1 + \min\{1/2, b - a\}]\,, z_3 \in (z_1+b-a, \infty)\}) = \infty. \quad (4.22)$$

Since the set $\{(z_1, z_2, z_3) : z_1 \in [a, a + 1/2)\,, z_2 \in (z_1, z_1 + \min\{1/2, b - a\}]\,, z_3 \in (z_1 + b - a, \infty)\}$ is a subset of $\{(z_1, z_2, z_3) : z_1 < z_2, z_2 \in [a, a + 1), z_3 > z_2, z_3 - z_1 > b - a\}$, (4.22) implies that the latter set must also have measure ∞ under η. Then

$$\eta(\{(z_1, z_2, z_3) : z_1 < z_2, z_2 \in [0, 1), z_3 > z_2, z_3 - z_1 > b - a\}) = \mathbb{E}(|S_1|) = \infty$$

by Corollary 4.3.7.

Although not necessary, one direct and simple way leading to $\mathbb{E}(|S_1|) = \infty$ is, of course, to have S_1 to be infinite with positive probability. The next proposition gives a necessary and sufficient condition for S_1 to be infinite.

Proposition 4.4.2. *The set S_1 has infinite number of elements if and only if at least one of the following four scenarios is true:*

(1) There exists an increasing sequence $\{x_n\}_{n=1,2,\ldots}$ in $S \cap [0, 1)$, such that for each n, $x_{n+1} \preceq x_n$, $l_{x_n} < x_n$, and $r_{x_n} \geqslant x_n + b - a$;

(2) *There exists an decreasing sequence* $\{x_n\}_{n=1,2,...}$ *in* $S \cap [0,1)$*, such that for each* n*,* $x_n \preceq x_{n+1}$*,* $l_{x_n} < x_n$*, and* $r_{x_n} \geqslant x_n + b - a$*;*

(3) *There exists an decreasing sequence* $\{x_n\}_{n=1,2,...}$ *in* $S \cap [0,1)$*, such that for each* n*,* $x_{n+1} \preceq x_n$*,* $r_{x_n} > x_n$*, and* $l_{x_n} \leqslant x_n - b + a$*;*

(4) *There exists an increasing sequence* $\{x_n\}_{n=1,2,...}$ *in* $S \cap [0,1)$*, such that for each* n*,* $x_n \preceq x_{n+1}$*,* $r_{x_n} > x_n$*, and* $l_{x_n} \leqslant x_n - b + a$*.*

Proof. The "if" part is trivial. For the "only if" part, assume $|S_1| = \infty$. Then there exists a monotone sequence of points in S_1. Without loss of generality, assume the sequence is increasing, and denote it by $\{x_n\}_{n=1,2,...}$, with $\lim_{n\to\infty} x_n = x_\infty$, which is not necessarily in S_1. Moreover, x_1 can be chosen so that $x_\infty - x_1 < b - a$.

Next, the sequence can be taken such that for any $n = 1, 2, ...$, either $x_n \preceq x_{n+1}$ or $x_n \succeq x_{n+1}$, which is not trivial since "$\preceq$" is only a partial order. To see this, consider the set of indices $J = \{j : x_j \npreceq x_{j+1}, x_j \nsucceq x_{j+1}\}$. For any $n \in J$, let $y_n = L([x_n, x_{n+1}])$, then $y_n \in (x_n, x_{n+1})$. As such, we have $r_{x_n} \leqslant y_n$. By the definition of S_1, this implies that $l_{y_n} \leqslant l_{x_n} < r_{x_n} - (b-a) \leqslant y_n - (b-a)$. Symmetrically, $r_{y_n} > y_n + (b-a)$. This means, for any $n_1, n_2 \in J$, $|y_{n_1} - y_{n_2}| \geqslant b - a$, which guarantees that J is a finite set. Taking the subsequence of $\{x_n\}$ starting from $n_0 = \max\{j : j \in J\} + 1$ gives a new sequence for which either $x_n \preceq x_{n+1}$ or $x_n \succeq x_{n+1}$.

For such a sequence, it is clear that for any $n \geqslant 2$, $x_n \preceq x_{n-1}$ and $x_n \preceq x_{n+1}$ can not hold at the same time, since otherwise $r_{x_n} - l_{x_n} \leqslant x_{n+1} - x_{n-1} < b - a$, implying that x_n can not be in S_1. Thus, either $\{x_n\}_{n=1,2,...}$ is monotone according to $\preceq$, or there exists n_0, such that $x_1 \preceq x_2 \preceq \cdots \preceq x_{n_0}$ and $x_{n_0} \succeq x_{n_0+1} \succeq \ldots$. As a result, there always exists a subsequence of $\{x_n\}_{n=1,2,...}$, still denoted as $\{x_n\}_{n=1,2,...}$ by a slight abuse of notation, which is monotone according to $\preceq$. Next we discuss the two possible cases.

Case 1: $x_{n+1} \preceq x_n$ for any n. In this case note that $l_{x_n} \in [x_{n-1}, x_n)$, hence $\lim_{n\to\infty} l_{x_n} = x_\infty$. Moreover, since x_n is decreasing in n according to $\preceq$ and $r_{x_n} > l_{x_n} + b - a \geqslant x_{n-1} + b - a > x_\infty$ for any $n \geqslant 2$, r_{x_n} is non-increasing in n for $n \geqslant 2$. Therefore,

$$r_{x_n} \geqslant \lim_{n\to\infty} r_{x_n} \geqslant \lim_{n\to\infty} l_{x_n} + b - a = x_\infty + b - a > x_n + b - a$$

for $n = 2, \ldots$. Thus, scenario (1) in the proposition holds for $\{x_n\}_{n=2,3,\ldots}$.

Case 2: $x_n \preceq x_{n+1}$ for any n. Then $r_{x_1} < x_\infty$, hence $l_{x_1} < x_\infty - b + a$. By a similar reasoning as in case 1, l_{x_n} is non-increasing in n, so $l_{x_n} \leqslant l_{x_1}$. Recall that x_n is increasing and $\lim_{n\to\infty} x_n = x_\infty$, therefore, there exists n_0, such that $x_\infty - x_n < x_\infty - b + a - l_{x_1}$ for any $n > n_0$, which implies $l_{x_n} \leqslant l_{x_1} < x_n - b + a$ for $n \geqslant n_0$. Taking the subsequence of $\{x_n\}_{n=1,2,\ldots}$ starting from x_{n_0} leads to scenario (4).

Scenarios (2) and (3) can be derived symmetrically by assuming that the sequence $\{x_n\}_{n=1,2,\ldots}$ is decreasing. $\square$

With Proposition 4.4.2 proved, it is then obvious that scenarios (1) and (2) corresponds to the explosion of the density f near the boundary a, while scenarios (3) and (4) corresponds to the explosion of f near the boundary b.

Corollary 4.4.3. *Under the same setting as in Proposition 4.4.2, (1) or (2) implies that* $\lim_{x\downarrow a} f(x) = \infty$*, (3) or (4) implies that* $\lim_{x\uparrow b} f(x) = \infty$*.*

Proof. We prove that scenario (1) implies $\lim_{x\downarrow a} f(x) = \infty$. The other cases are similar.

In scenario (1), for any $n \geqslant 2$, $0 \leqslant x_{n-1} \leqslant l_{x_n} < x_n < 1$, and $r_{x_n} > x_n + b - a > l_{x_n} + b - a$. Moreover, $x_n - l_{x_n} \leqslant x_n - x_{n-1} \to 0$ as $n \to \infty$. Hence scenario (1) happens with positive probability implies that

$$\eta(\{(z_1, z_2, z_3) : z_1 \in [z_2 - \Delta, z_2), z_2 \in [0, 1), z_3 \in (z_1 + b - a, \infty)\}) = \infty$$

for any $\Delta > 0$. In particular,

$$\eta(\{(z_1, z_2, z_3) : z_1 \in [z_2 - x_0 + a, z_2), z_2 \in [0, 1), z_3 \in (z_1 + b - a, \infty)\}) = \infty.$$

Note that

$$\begin{aligned}
&\{(z_1, z_2, z_3) : z_1 \in [z_2 - x_0 + a, z_2), z_2 \in [0, 1), z_3 \in (z_1 + b - a, \infty)\} \\
\subset &\{(z_1, z_2, z_3) : z_1 \in [-x_0 + a, 1), z_2 \in (z_1, z_1 + x_0 - a], z_3 \in (z_1 + b - a, \infty)\}.
\end{aligned}$$

Thus (4.20) holds:

$$\begin{aligned}&\eta(\{(z_1,z_2,z_3):z_1\in[a,a+1),z_2\in(z_1,z_1+x_0-a],z_3\in(z_1+b-a,\infty)\})\\=&\frac{1}{1+x_0-a}\eta(\{(z_1,z_2,z_3):z_1\in[-x_0+a,1),z_2\in(z_1,z_1+x_0-a],z_3\in(z_1+b-a,\infty)\})\\=&\infty,\end{aligned}$$

where the first equality follows from Corollary 4.3.7. □

As an application of Proposition 4.3.2 and Corollary 4.4.3, consider the location of the path supremum of a stochastic process $\mathbf{X}=\{X(t)\}_{t\in\mathbb{R}}$ with continuous sample paths, formally defined as

$$\tau_{\mathbf{X},I}:=\inf\{t\in I:X(t)=\sup_{s\in I}X(s)\}.$$

The infimum is used to choose the leftmost point among all the points where $\sup_{s\in I}X(s)$ is achieved, in the case where there are more than one such point. If we further assume that

Assumption U. For any $I\in\mathcal{I}$,

$$P(\text{there exist } t_1,t_2\in I,t_1\neq t_2,\text{ such that } X(t_1)=X(t_2)=\sup_{s\in I}X(s))=0,$$

$i.e.$, the location of the path supremum is almost surely unique, then the infimum in the definition of $\tau_{\mathbf{X},I}$ can be removed.

Most of the commonly used processes do satisfy Assumption U. It is proved in Kim and Pollard (1990) that for a Gaussian process $\mathbf{X}$, Assumption U holds if and only if $Var(X(t)-X(s))\neq 0$ for any $s\neq t$. A necessary and sufficient condition for more general processes with continuous sample paths can be found in Pimentel (2014).

Note that in the case of the location of the path supremum, the random set S, as defined before Lemma 4.3.4, takes the form

$$S=\{t:\text{ there exists }\Delta>0,\text{ such that }X(t)=\sup_{s\in[t-\Delta,t]}X(s)\text{ or }X(t)=\sup_{s\in[t,t+\Delta]}X(s)\},$$

and the partial order $\preceq$ is the natural order of the value of the process $X(t)$.

Corollary 4.4.4. *Let $\mathbf{X} = \{X(t)\}_{t\in\mathbb{R}}$ be a stochastic process with continuous sample paths and stationary increments. Assume $\mathbf{X}$ satisfies Assumption U. If the local maxima of $\mathbf{X}$ is dense in $[a,b]$ with positive probability, then the density of $\tau_{\mathbf{X},I}$, denoted by f, satisfies $\lim_{t\downarrow a} f(t) = \infty$ or $\lim_{t\uparrow b} f(t) = \infty$.*

Proof. By the stationarity of the increments, it suffices to prove the results for the case where $a = 0$. Denote by D the event that the local maxima of $\mathbf{X}$ is dense. Let $\tau' = \tau_{\mathbf{X},[0,4b]}$, then $r_{\tau'} - l_{\tau'} \geqslant 4b$. Therefore, $P(D, r_{\tau'} - \tau' \geqslant 2b) > 0$ or $P(D, \tau' - l_{\tau'} \geqslant 2b) > 0$. Without loss of generality, assume that $P(D, r_{\tau'} - \tau' \geqslant 2b) > 0$. As a result,

$$P(D, \text{there exists } t \in S \cap [0,4b], r_t - t \geqslant 2b) > 0,$$

hence also

$$P(D, \text{there exists } t \in S \cap [0,b), r_t - t \geqslant 2b) > 0$$

by the stationarity of the increments.

Let $t_\infty = \tau_{\mathbf{X},[0,b]}$. From now on we focus on the event

$$\{D,\ \text{there exists } t \in S \cap [0,b), r_t - t \geqslant 2b\}.$$

In this case, $t_\infty = \tau_{\mathbf{X},[0,2b]} < b$, and $r_{t_\infty} \geqslant 2b$. By Assumption U, there exists $\varepsilon \in (0, b - t_\infty)$, such that $\inf_{s\in[t_\infty, t_\infty+\varepsilon]} X(s) > \sup_{s\in[b,2b]} X(s)$. For $n = 1, 2, ...$, let $t_n = \tau_{\mathbf{X},[t_\infty + \frac{1}{n+1}\varepsilon, 2b]}$. Then t_n is a non-increasing sequence satisfying $\lim_{n\to\infty} t_n = t_\infty$, and $X(t_n) \leqslant X(t_{n+1})$ for all n. Moreover, since

$$\sup_{s\in[t_\infty+\frac{1}{n+1}\varepsilon, b]} X(s) \geqslant \sup_{s\in[t_\infty+\frac{1}{n+1}\varepsilon, t_\infty+\varepsilon]} X(s) > \sup_{s\in[b,2b]} X(s),$$

$t_n \in [t_\infty + \frac{1}{n+1}\varepsilon, b]$, and $r_{t_n} \geqslant 2b \geqslant t_n + b$. By removing all equal terms in $\{t_n\}_{n=1,2,...}$ and all the terms in $\{t_n\}_{n=1,2,...}$ at which the values of $\mathbf{X}$ are equal, we get a decreasing sequence $\{t_n\}_{n=1,2,...}$, satisfying $\lim_{n\to\infty} t_n = t_\infty$ and $X(t_n) < X(t_{n+1})$, hence $t_n \preceq t_{n+1}$, for all n. Since the local maxima are dense and the sample paths are continuous, such a sequence can be approached by a sequence of local maxima $\{t'_n\}_{n=1,2,...}$, while all the properties derived above still hold. In addition, as all the points in the new sequence are local maxima, we have $l_{t'_n} < t'_n$, $n = 1, 2, ...$ By the stationarity of the increments, this is scenario (2) in Proposition 4.4.2. Symmetrically, if $P(\tau' - l_{\tau'} \geqslant 2b) > 0$, then scenario (4) in Proposition 4.4.2 happens with positive probability. $\square$

The following result is a direct application of Corollary 4.4.4 and Proposition 4.3.2. The processes for which the result applies include Brownian motion, Ornstein-Uhlenbeck processes, or more generally, any process $\{X(t)\}_{t\geqslant 0}$ satisfying Assumption U and of the form

$$X(t) = \int_0^t Y(s)dB_s,$$

where $\{Y(t)\}_{t\geqslant 0}$ is a predictable stationary process which is independent of the standard Brownian motion $\{B_t\}_{t\geqslant 0}$, and for which the above stochastic integral is well-defined.

Corollary 4.4.5. *Let* $\mathbf{X} = \{X(t)\}_{t\geqslant 0}$ *be a continuous semimartingale with stationary increments, satisfying Assumption U. Assume that the local martingale part of* $\mathbf{X}$ *almost surely does not have any flat part. For any* $I = [a, b] \in \mathcal{I}$*, let* $\tau_{\mathbf{X},I}$ *be defined as previously, and* f *be its density on* (a, b)*. Then* $P(\tau_{\mathbf{X},I} = a) = P(\tau_{\mathbf{X},I} = b) = 0$*, and* $\lim_{t\downarrow a} f(t) + \lim_{t\uparrow b} f(t) = \infty$*.*

Proof. Since the $\mathbf{X}$ is a semimartingale and has a local martingale part which is nowhere flat, it is of unbounded variation over any interval, hence the local maxima and the local minima of $\mathbf{X}$ are almost surely dense in any interval. Thus, Corollary 4.4.4 applies. Moreover, also because of the unbounded variation and the continuity of the path, with probability 1, a is an accumulation point, both from the left and from the right, of the level set $\{t \in \mathbb{R} : X(t) = X(a)\}$. As a result, for any $\varepsilon > 0$, there exists $t \in (a, a+\varepsilon]$, such that $X(t) \geqslant X(a)$. If the equality holds for all such $t \in (a, b]$, then Assumption U is violated. Hence almost surely there exists $t \in (a, b]$ such that $X(t) > X(a)$. Thus, $P(\tau_{\mathbf{X},I} = a) = 0$. The case for the right boundary b is symmetric. $\square$

Chapter 5

Future Works

In Chapter 2, we proposed a sufficient and necessary condition for the compatibility of probability measures on a probability space with their corresponding distributions on the real line. In Chapter 3, we studied the existence of a periodic stationary process such that the distribution of an intrinsic location functional for this process coincides with given distribution. In Chapter 4, we gave a unified framework for random locations with probabilistic symmetries. In this chapter, we are mainly focused on three future directions: random locations for exchangeable processes, random locations, especially the location of the path supremum, of max-stable processes, and large deviation of the maximum for stochastic processes.

5.1 Random locations for exchangeable processes

In this part we will study another probabilistic symmetry: exchangeability. For $0 \leqslant a \leqslant b$, let transposition $T_{a,b}(t)$ be

$$T_{a,b}(t) = \begin{cases} t+b-a, & t \leqslant a, \\ t-a, & t \in (a,b], \\ t, & t > b. \end{cases}$$

An $\mathbb{R}$-valued process X on $\mathbb{R}^+$ with $X_0 = 0$ is said to be *exchangeable* if $X \circ T_{a,b}^{-1} \overset{d}{=} X$ for all $a, b \in \mathbb{R}^+$. Here the notation

$$(X \circ f^{-1})_t = \int \mathbb{I}_{\{s \in \mathbb{R}, f(s) \leqslant t\}} \mathrm{d}X_s, \text{ for } t \in \mathbb{R}.$$

I plan to work on the properties of a subclass of intrinsic location functionals for exchangeable processes. Conversely, I will proceed to characterizing the exchangeability for stochastic processes using this family of random locations. More precisely, for any intrinsic location functional, a partially ordered random set representation was proposed in Shen (2016). Consider the intrinsic location functionals whose corresponding partial order at each point solely depends on a small neighbourhood of this point. It should not be difficult to show that the distribution of such intrinsic location for any exchangeable process on a compact interval converges to the uniform distribution on this interval, as the length of the neighbourhood goes to 0.

For the other direction, I expect to propose sufficient conditions for the exchangeability of stochastic processes using the distribution of a certain set of random locations. For simplicity, I will start with point processes and further extend the results to general continuous processes.

5.2 Max-stable processes

Max-stable processes arise as the limit of maxima of independent and identically distributed processes, under appropriate normalization. The theoretical properties of max-stable processes have been extensively studied, including the finite dimension distribution, connection with Poisson point processes, the spectral representations and association with α-stable processes (De Haan, 1984; Stoev and Taqqu, 2005), while little attention has been paid to the random locations of these processes. Thus, it will be interesting to see how the random locations are distributed for the max-stable processes, with special attention to the random locations naturally related to the maximum, such as the location of the path supremum over an interval. With the spectral representation, I can start with the location of the path supremum of each component over an interval, and hopefully I can discover interesting results for some special max-stable processes, such as moving maximum processes, extremal Gaussian processes and Brown-Resnick processes. Another interesting aspect of max-stable processes is stationarity. For the stationarity of a max-stable process, Engelke and Kabluchko (2016) explores a special form of $\eta(t) = \max_{i\in\mathbb{N}} U_i e^{<X_i,t>-\kappa(t)}$, where $\{U_i, i \in \mathbb{N}\}$ are Poisson processes, $\{X_i, i \in \mathbb{N}\}$, are independent copies of a random vector X, and $\kappa : \mathbb{R}^d \to \mathbb{R}$ is a function. I am interested in the stationarity of more general max-stable processes, and will try to come up with a set of simple sufficient and necessary conditions.

5.3 Large deviation of the maximum for stochastic processes

Large deviation theory has been developed in various fields, ranging from queuing theory to statistics and from interacting particle systems to superexponential estimates, see Varadhan (2008) for a survey of large deviation. In risk models, for example, large deviation principles are widely used to estimate the asymptotic behavior and the exponential rate of the total claims probability (Klüppelberg and Mikosch, 1997; Tang et al., 2001). I am interested in the application of large deviation principles to the maximum of the stochastic processes, which is naturally related to the problems about the random locations discussed in previous chapters, including the location of the path supremum, the first hitting time to high levels, etc. For example, for a stochastic process, one can consider how the distributions of first hitting times to a level over an interval evolve when both the length of the interval and the level grow to infinity in an appropriate way. I expect that large deviation principles to be a powerful tool to investigate the asymptotic behaviors of the set of first hitting times. This can be regarded as a compatibility problem of the distributions of this random location with extreme values and larger and larger intervals.

Index

References

Alili, L., Patie, P., and Pedersen, J. L. (2005). Representations of the first hitting time density of an Ornstein-Uhlenbeck process. *Stochastic Models*, 21(4):967–980.

Arnold, B. C. and Press, S. J. (1989). Compatible conditional distributions. *Journal of the American Statistical Association*, 84(405):152–156.

Baez, J. C. and Fong, B. (2013). A Noether theorem for Markov processes. *Journal of Mathematical Physics*, 54(1):013301.

Barros-Neto, J. (1973). *An introduction to the theory of distributions*. M. Dekker.

Bingham, N. H. and Kiesel, R. (2013). *Risk-neutral valuation: Pricing and hedging of financial derivatives*. Springer Science & Business Media.

Blackwell, D. (1951). Comparison of experiments. Technical report, Howard University Washington United States.

Blackwell, D. (1953). Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, 24(2):265–272.

Breiman, L. (1967). First exit times from a square root boundary. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 2, pages 9–16. University of California Press.

Buchen, P. W. and Kelly, M. (1996). The maximum entropy distribution of an asset inferred from option prices. *Journal of Financial and Quantitative Analysis*, 31(1):143–159.

Choquet, G. and Meyer, P.-A. (1963). Existence et unicité des représentations intégrales dans les convexes compacts quelconques. *Ann. Inst. Fourier (Grenoble)*, 13:139–154.

De Haan, L. (1984). A spectral representation for max-stable processes. *The Annals of Probability*, 12(4):1194–1204.

Embrechts, P. and Maejima, M. (2002). *Selfsimilar processes*. Princeton University Press.

Embrechts, P., Wang, B., and Wang, R. (2015). Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance and Stochastics*, 19(4):763–790.

Engelke, S. and Kabluchko, Z. (2016). A characterization of the normal distribution using stationary max-stable processes. *Extremes*, 19(1):1–6.

Föllmer, H. and Schied, A. (2016). *Stochastic Finance. An Introduction in Discrete Time*. Walter de Gruyter, Berlin, Fourth Edition.

Glasserman, P. and Li, J. (2005). Importance sampling for portfolio credit risk. *Management science*, 51(11):1643–1656.

Gough, J. E., Ratiu, T. S., and Smolyanov, O. G. (2015). Noether's theorem for dissipative quantum dynamical semi-groups. *Journal of Mathematical Physics*, 56(2):022108.

Gutmann, S., Kemperman, J., Reeds, J., and Shepp, L. (1991). Existence of probability measures with given marginals. *The Annals of Probability*, 19(4):1781–1797.

Hirsch, F., Profeta, C., Roynette, B., and Yor, M. (2011). *Peacocks and associated martingales, with explicit constructions*. Springer Science & Business Media.

Jakobsons, E., Han, X., and Wang, R. (2016). General convex order on risk aggregation. *Scandinavian Actuarial Journal*, 2016(8):713–740.

Jarrow, R. and Rudd, A. (1982). Approximate option valuation for arbitrary stochastic processes. *Journal of Financial Economics*, 10(3):347–369.

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. CRC Press.

Kifer, Y. (1988). *Ergodic theory of random transformations.* Birkhauser Boston, Inc., Dunod.

Kim, J. and Pollard, D. (1990). Cube root asymptotics. *The Annals of Statistics*, 18(1):191–219.

Klüppelberg, C. and Mikosch, T. (1997). Large deviations of heavy-tailed random sums with applications in insurance and finance. *Journal of Applied Probability*, 34(2):293–308.

Koralov, L. and Sinai, Y. G. (2007). *Theory of probability and random processes.* Springer Science & Business Media.

Kosmann-Schwarzbach, Y. (2011). The Noether Theorems. In *The Noether Theorems*, chapter 2, pages 55–64. Springer.

Le Cam, L. (1996). Comparison of experiments: A short review. *Lecture Notes-Monograph Series*, 30:127–138.

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and related properties of random sequences and processes.* Springer Science & Business Media.

Leblanc, B., Renault, O., and Scaillet, O. (2000). A correction note on the first passage time of an ornstein-uhlenbeck process to a boundary. *Finance and Stochastics*, 4(1):109–111.

Lezcano, A. G. and de Oca, A. C. M. (2018). A stochastic version of the Noether Theorem. *Foundations of Physics*, 48:726–746.

Lindgren, G. (2012). *Stationary stochastic processes: theory and applications.* CRC Press.

Misawa, T. (1994). Conserved quantities and symmetry for stochastic dynamical systems. *Physics Letters A*, 195(3-4):185–189.

Müller, A. and Stoyan, D. (2002). *Comparison methods for statistical models and risks*, volume 389. Wiley New York.

Neyman, J. and Pearson, E. (1933). On the problem of the most efficient tests of statistical inference. *Biometrika A*, 20:175–240.

Noether, E. (1918). Invariant variation problems. *Gott. Nachr.*, 1918:235–257. [Transp. Theory Statist. Phys.1,186(1971)].

Parthasarathy, K. R. and Varadhan, S. S. (1964). Extension of stationary stochastic processes. *Theory of Probability & Its Applications*, 9(1):65–71.

Phelps, R. R. (2001). *Lectures on Choquet's theorem. Second ed.* Springer-Verlag Berlin Heidelberg.

Pimentel, L. P. (2014). On the location of the maximum of a continuous stochastic process. *Journal of Applied Probability*, 51(1):152–161.

Puccetti, G. and Wang, R. (2015). Extremal dependence concepts. *Statistical Science*, 30(4):485–517.

Quiggin, J. (2012). *Generalized expected utility theory: The rank-dependent model*. Springer Science & Business Media.

Revuz, D. and Yor, M. (2013). *Continuous martingales and Brownian motion*. Springer Science & Business Media.

Rüschendorf, L. (2013). Mathematical risk analysis. *Springer Ser. Oper. Res. Financ. Eng. Springer, Heidelberg*.

Samorodnitsky, G. and Shen, Y. (2012). Distribution of the supremum location of stationary processes. *Electronical Journal of Probability*, 17(42):1–17.

Samorodnitsky, G. and Shen, Y. (2013a). Intrinsic location functionals of stationary processes. *Stochastic Processes and their Applications*, 123(11):4040–4064.

Samorodnitsky, G. and Shen, Y. (2013b). Is the location of the supremum of a stationary process nearly uniformly distributed? *The Annals of Probability*, 41(5):3494–3517.

Shaked, M. and Shanthikumar, G. (2007). *Stochastic orders*. Springer Science & Business Media.

Shen, J., Shen, Y., Wang, B., and Wang, R. (2017). Distributional compatibility for change of measures. *arXiv preprint arXiv:1706.01168*.

Shen, J., Shen, Y., and Wang, R. (2018). Random locations of periodic stationary processes. *Stochastic Processes and their Applications*. (to appear).

Shen, Y. (2013). *Stationarity and random locations*. PhD thesis, Cornell University.

Shen, Y. (2016). Random locations, ordered random sets and stationarity. *Stochastic Processes and their Applications*, 126(3):906–929.

Shen, Y. (2018). Location of the path supremum for self-similar processes with stationary increments. *Annales de l'Institut Henri Poincaré (B)*. (to appear).

Siegmund, D. (1976). Importance sampling in the monte carlo study of sequential tests. *The Annals of Statistics*, 4(4):673–684.

Stoev, S. A. and Taqqu, M. S. (2005). Extremal stochastic integrals: a parallel between max-stable processes and α-stable processes. *Extremes*, 8(4):237–266.

Strassen, V. (1965). The existence of probability measures with given marginals. *The Annals of Mathematical Statistics*, 36(2):423–439.

Tang, Q., Su, C., Jiang, T., and Zhang, J. (2001). Large deviations for heavy-tailed random sums in compound renewal model. *Statistics & Probability Letters*, 52(1):91–100.

Thieullen, M. and Zambrini, J.-C. (1997a). Probability and quantum symmetries. i. the theorem of noether in schrödinger's euclidean quantum mechanics. *Ann. Inst. H. Poincaré Phys. Théor*, 67(3):297–338.

Thieullen, M. and Zambrini, J.-C. (1997b). Symmetries in the stochastic calculus of variations. *Probability Theory and Related Fields*, 107(3):401–427.

Torgersen, E. (1991). *Comparison of statistical experiments*, volume 36. Cambridge University Press.

Van Casteren, J. A. (2003). The hamilton-jacobi-bellman equation and the stochastic Noether theorem. In *Evolution Equations: Applications to Physics, Industry, Life Sciences and Economics*, pages 375–401. Springer.

Varadhan, S. (2008). Large deviations. *The Annals of Probability*, 36(2):397–419.

Wang, B. and Wang, R. (2016). Joint mixability. *Mathematics of Operations Research*, 41(3):808–826.

Wang, R., Peng, L., and Yang, J. (2013). Bounds for the sum of dependent risks and worst value-at-risk with monotone marginal densities. *Finance and Stochastics*, 17(2):395–417.

Yasue, K. (1981). Stochastic calculus of variations. *Journal of Functional Analysis*, 41(3):327–340.