C8.6 Limit Theorems and Large Deviations in Probability

BY ZHONGMIN QIAN

January 15, 2021

Contents

2021 HT. Office hours and video recording of my lectures are posted at appropriate schedule tailed with the problem sheets.

1 Several facts from analysis

Random models arising from science and engineering are time series, stochastic processes in discrete or continuous time, or more general random fields. Distributions of these models are probability measures on path spaces and in general on function spaces which are infinite dimensional spaces. Often one starts with simple discrete models with scaling parameters and takes their limits in the hope of obtaining more practical models. This is the reason why we need to study limit theories of probability measures on function spaces. In this course we develop important tools dealing with distributions on path spaces.

To develop probability limit theories covered in this course, we need several theorems from analysis which may be not covered in prerequisite options. In this first lecture, we recall several notions and theorems. The proofs of the results we review are not included, but references are given, so if you are interested in the details of a particular topic, you may study further with the indicated literature.

1.1 Topology on metric spaces

In Paper A2 Metric Spaces and Complex Analysis, we have studied the concepts of compactness and connectedness for metric spaces. Let us add three more notions about metric spaces which are useful in our discussions below.

Let (E,ρ) be a metric space. Then $B_x(r)$ denotes the open ball centered at x with radius r, that is, $\{z\in E:\rho(z,x)<r\}$. The metric ρ induces a topology on E, i.e. a collection of open subsets of E, namely a subset $U\subset E$ is open if for every $x\in E$, there is a $r>0$ such that $B_x(r)\subset U$. This topology of E is determined uniquely by the metric ρ and must be Hausdorff. The metric ρ is however not uniquely determined by its generated topology. In fact, as we have seen in Paper A2, ρ, $\rho\wedge 1$, $\frac{\rho}{1+\rho}$ and etc. determine the same topology, and therefore determine the same Borel σ-algebra on E, denoted by $\mathscr{B}(E)$. Hence, if it is necessary, we will use *bounded metrics* only.

A metric space (E,ρ) is *complete*, by definition, if every ρ-Cauchy sequence has a unique limit in E.

We say (E,ρ) is *separable* if it possesses a countable dense subset, i.e. there is a countable subset $\mathscr{Q}\subset E$, such that for every $x\in E$ and for every $\varepsilon>0$, there is $a\in\mathscr{Q}$, $x\in B_a(\varepsilon)$. From Prelims Analysis, we know that $\mathbb{R}^d$ (equipped with the standard metric) is separable and complete.

A complete and separable metric space (E,ρ) is called a *Polish space*. More proper definition of Polish spaces should be that; if you want to make your life more complicated; a topological space E is a Polish space, if there is a metric ρ on E which generates the same topology on E and (E,ρ) is complete and separable.

For a given metric space, in particular, if the space is infinite dimensional, it is often quite challenging to validate if a metric space is compact or not. The concept of *totally bounded* metric spaces is introduced to address this problem.

We recall that a metric space is compact (definition via open covers) if and only if it is sequentially compact (definition in terms of sequences).

Definition 1.1. *Let (E,ρ) be a metric space. A subset $A\subset E$ is* totally bounded*, if for every $\varepsilon>0$, there are finitely many points $x_1,\cdots,x_n\in E$ (for some $n\in\mathbb{N}$), such that $\{B_{x_i}(\varepsilon):i=1,\cdots,n\}$ is a cover of A, i.e. $\cup_{i=1}^n B_{x_i}(\varepsilon)\supset A$.*

From definition, a totally bounded metric space must be bounded and separable.

Lemma 1.2. *A metric space (E,ρ) is totally bounded, if and only if any sequence of E contains sub-sequence which is Cauchy.*

We have the following important result about compactness and totally boundedness.

Theorem 1.3. *(*Hausdorff*) Let (E,ρ) be a metric space.*

(1) If E is compact then E is totally bounded.

(2) Suppose (E,ρ) is complete. Then $A\subset E$ is compact if and only if A is totally bounded and closed.

(3) If every totally bounded and closed subset of E is compact, then (E,ρ) must be complete.

Theorem 1.4. (Ascoli-Arzelà Theorem) *Let (T,d) be a compact metric space, and $C(T;\mathbb{R}^d)$ denote the space of all continuous functions on T taking values in $\mathbb{R}^d$, equipped with the supremum norm $||x|| = \sup_{t\in T}|x(t)|$ and its induced metric.*

(1) $C(T;\mathbb{R}^d)$ is a complete metric space.

(2) A family of continuous functions, $\mathscr{L} \subset C(T;\mathbb{R}^d)$, is relatively compact, if and only if $\mathscr{L}$ is bounded

$$\sup_{x\in\mathscr{L}}\sup_{t\in T}|x(t)| < \infty,$$

and equivalently continuous in the sense that

$$\lim_{\delta\downarrow 0}\sup_{x\in\mathscr{L};d(s,s')\leq\delta}\left|x(s)-x(s')\right| = 0\,. \tag{1.1}$$

For a proof see Yosida [26, Chapter III].

This theorem may be generalized to the case where (T,d) is not necessary compact, and we will consider an important case where $T=[0,\infty)$ with the usual distance. For this case a function $x:[0,\infty)\to\mathbb{R}^d$ is called a path in $\mathbb{R}^d$. Let $\mathbb{C}(\mathbb{R}^d)$ or $C([0,\infty);\mathbb{R}^d)$ denote the space of all continuous paths in $\mathbb{R}^d$, equipped with the uniform convergence topology, i.e. the topology induced by the metric

$$\rho(x,y) = \sum_{n=1}^{\infty}\frac{1}{2^n}\left(\sup_{0\leq t\leq n}|x(t)-y(t)|\wedge 1\right), \quad \forall x,y\in C([0,\infty);\mathbb{R}^d). \tag{1.2}$$

Then $\mathbb{C}(\mathbb{R}^d)$ with metric ρ is complete and separable. About its relatively compact subsets, we have the following version of Ascoli-Arzelà theorem.

Corollary 1.5. (Ascoli-Arzelà Theorem) *$\mathscr{L}\subset\mathbb{C}(\mathbb{R}^d)$ is relatively compact, if and only if*

$$\sup_{x\in\mathscr{L}}|x(0)| < \infty$$

and

$$\lim_{\delta\downarrow 0}\sup_{x\in\mathscr{L};s,s'\leq T,|s-s'|\leq\delta}\left|x(s)-x(s')\right| = 0$$

for every $T>0$.

Since the second condition that the functions in $\mathscr{L}$ are equivalently continuous, therefore the first condition in the previous theorem may be replaced by the condition that for every $T>0$,

$$\sup_{x\in\mathscr{L},t\leq T}|x(t)| < \infty$$

which implies that the range $\{x(t): t\in[0,T], x\in\mathscr{L}\}$ is contained in a compact subset of $\mathbb{R}^d$.

1.2 Measures, integration and Daniell integration

In measure theory (paper B8.1 or A4), we start with a measure μ on a measurable space $(\Omega, \mathscr{F})$ and establish theory of integration. The procedure is simple, we first identify a class of non-negative $\mathscr{F}$-measurable *simple* functions and define integrals for these simple functions first. For a simple function $\phi = \sum_i c_i 1_{A_i}$ where $A_i \in \mathscr{F}$ and $c_i \geq 0$, i runs through a finitely many indices, its integral

$$\mu(\phi) = \int_\Omega \phi d\mu = \sum_i c_i \mu(A_i).$$

Extend the definition to a non-negative $\mathscr{F}$-measurable function f by

$$\mu(f) = \int_\Omega f d\mu = \sup\{\mu(\phi) : \phi \text{ non-negative } \mathscr{F}\text{- simple and } \phi \leq f\}.$$

For a general $\mathscr{F}$-measurable function f, if both $\mu(f^+)$ and $\mu(f^-)$ are finite, then f is integrable and define

$$\mu(f) = \int_\Omega f d\mu = \mu(f^+) - \mu(f^-).$$

Therefore, the important task is of course to construct measures. In finite dimensional case, the Lebesgue measure, constructed in Paper A4, serves our aim. In stochastic analysis, we construct measures on path spaces by studying distributions of stochastic processes and random fields. In this course we aim to provide important tools to tackle this construction problem.

Daniell integration is another way for constructing measures, which is based on the following observation.

If $(\Omega, \mathscr{F}, \mu)$ is a measure space, then $L^1(\Omega, \mathscr{F}, \mu)$, the vector space of all integrable function, is in fact a vector lattice in the sense that if $f, g \in L^1$, then $f \wedge g$ and $f \vee g$ belong to $L^1(\Omega, \mathscr{F}, \mu)$ too. The integration operation $f \to \mu(f)$ is *linear functional* on the vector space $L^1(\Omega, \mathscr{F}, \mu)$, and *positive* in the sense that $\mu(f) \geq 0$ if f is integrable and $f \geq 0$.

The following theorem establishes the equivalence between measures and Daniell integration.

Theorem 1.6. *(*Daniell-Stone's Theorem*) Let Ω be a space.*

a) Let $\mathscr{H}$ be a real vector lattice of some real valued functions on Ω, i.e. $\mathscr{H}$ is a real vector space and $\mathscr{H}$ is closed under the operations $\wedge$ and $\vee$.

b) Let $I : \mathscr{H} \to \mathbb{R}$ be a linear functional. Suppose that I is a Daniell integration *in the sense that:*
(b-i) $I(f) \geq 0$ for every $f \in \mathscr{H}$ and $f \geq 0$, and
(b-ii) I satisfies the following continuity condition (MCT): if f_n, $f \in \mathscr{H}$ and $f_n \uparrow f$, then $I(f_n) \uparrow I(f)$.

c) Let $\mathscr{F} = \sigma\{f : f \in \mathscr{H}\}$ the smallest σ-algebra such that every $f \in \mathscr{H}$ is measurable.

Then there is a measure μ on $(\Omega, \mathscr{F})$ such that $\mathscr{H} \subset L^1(\Omega, \mathscr{F}, \mu)$ and $\mu(f) = I(f)$ for every $f \in \mathscr{H}$.

Remark. The proof may be found in Dudley [13]. The continuity condition in 2) is also equivalent to that: if $f_n \in \mathscr{H}$ and $f_n \downarrow 0$, then $I(f_n) \downarrow 0$.

Remark. If there is an increasing $f_n \in \mathscr{H}$, $f_n \geq 0$ and $f_n \uparrow 1$, then the measure μ is σ-finite, and μ is unique. In particular, if $1 \in \mathscr{H}$, then μ is a finite measure, and is unique.

As an application of Daneill and Stone's theorem we may establish the Riesz representation theorem for positive functionals on a compact metric space.

Let us begin with the following

Definition 1.7. *Let E be a metric space. A function $f : E \to (-\infty,\infty]$ is called* lower semi-continuous *if $\{f > a\}$ is open for every $a \in \mathbb{R}$. If $-f$ is lower-semi-continuous, then we say f is upper semi-continuous.*

Clearly, f is lower semi-continuous if and only if $\{f \leq a\}$ is closed for every $a \in \mathbb{R}$.

Lemma 1.8. *(*Dini Lemma*) If E is a compact metric space, and $f_n \to [0,\infty)$ are* upper semi-continuous*, where $n = 1,2,\ldots$. Suppose $f_n \downarrow 0$, that is, $f_{n+1} \leq f_n$ for every n and $f_n \to 0$ as $n \to \infty$, then $f_n \to 0$ uniformly on E as $n \to \infty$.*

Proof. The Dini Lemma was proved for the case where $E = [a,b]$ in Prelims Analysis II. For any given $\varepsilon > 0$, $U_n = \{f_n < \varepsilon\}$ is open for every n. $\{U_n : n \geq 1\}$ is a open cover of E. Using compactness of E to determine N, such that $f_n(x) < \varepsilon$ for all $n \geq N$ and $x \in E$. □

Theorem 1.9. *(*F. Resiz representation theorem*) Let E be a compact metric space, and $C(E;\mathbb{R})$ denote the vector space of all continuous and real-valued functions on E. Suppose $I : C(E;\mathbb{R}) \to \mathbb{R}$ is a positive linear functional. Then I must be a Daniell integration with $\mathscr{H} = C(E;\mathbb{R})$, and therefore there is a unique finite measure μ on $(E,\mathscr{B}(E))$ such that $I(f) = \mu(f)$ for every $f \in C(E;\mathbb{R})$.*

Proof. We only need to check the continuity condition (b-ii) in the Daniell-Stone Theorem. Let $f_n \in C(E;\mathbb{R})$, $f_n \geq 0$ and $f_n \downarrow 0$. By Dini lemma, $f_n \downarrow 0$ uniformly on E. Therefore for every $\varepsilon > 0$ there is N such that $0 \leq f_n(x) < \varepsilon$ for every $x \in E$ and $n \geq N$. That is $\varepsilon 1 - f_n \geq 0$ for all $n \geq N$. Since I is positive and linear, so that

$$I(\varepsilon 1 - f_n) = \varepsilon I(1) - I(f_n) \geq 0$$

and so that

$$0 \leq I(f_n) \leq \varepsilon I(1)$$

for all $n \geq N$, which implies that $I(f_n) \to 0$. Therefore I is a Daniell integration. The other conclusions now follow from Daniell-Stone's theorem and the fact that the σ-algebra generated by all continuous functions are exactly the Borel σ-algebra on E. □

Remark. F. Resiz representation theorem for locally compact space. Suppose E is a locally compact metric space (or a locally compact Hausdorff space), then $C(E;\mathbb{R})$ is in general too big for a Daniell integration. For example, for the Lebesgue measure, not every continuous function is integrable, therefore we have to work with a smaller function space. A good and interesting one is the function space $\mathscr{H} = C_c(E;\mathbb{R})$ of all continuous functions on E with compact supports. Suppose I is a positive linear functional on $\mathscr{H}$, then one can show that I must be a Daniell integration. Hence, there is a unique σ-finite measure μ on $(E,\mathscr{B}(E))$ such that $I(f) = \mu(f)$ for every $f \in C_c(E;\mathbb{R})$. Moreover μ is *regular* in the sense that

$$\mu(U) = \sup\{\mu(K) : K \subseteq U \text{ and } K \text{ is compact}\}$$

for every open subset U, and

$$\mu(A) = \inf\{\mu(U) : U \supseteq A \text{ and } U \text{ is open}\}$$

for every $A \in \mathscr{B}(E)$. The reader should be able to find the details in standard textbook such as Dudley [13].

2 Weak convergences of probability measures

In this section we study the weak convergence of probability measures on metric spaces. Let (E,ρ) be a metric space (or more general a E is a metrizable topological space), whose Borel σ-algebra is denoted by $\mathscr{B}(E)$.

Let $C_b(E)$ (resp. $U_\rho(E)$) denote the totality of all *bounded* (resp. *bounded* and *uniformly*) continuous functions on E. Then $U_\rho(E) \subset C_b(E)$.

Definition 2.1. *A sequence (P_n) of probability measures on $(E,\mathscr{B}(E))$ converges to a probability measure P on $(E,\mathscr{B}(E))$ weakly, denoted by $P_n \to P$ weakly, if*

$$\lim_{n\to\infty} \int_E f(x)P_n(dx) = \int_E f(x)P(dx)$$

for every $f \in C_b(E)$.

The definition may be generalized to the "continuous" case. For example, suppose P_ε is a probability measure for every $\varepsilon \in (0,\delta)$, where $\delta > 0$ is some constant. Then $P_\varepsilon \to P$ weakly as $\varepsilon \downarrow 0$, if

$$\lim_{\varepsilon\downarrow 0} \int_E f(x)P_\varepsilon(dx) = \int_E f(x)P(dx)$$

for every $f \in C_b(E)$.

Example. Let P_ε be the Gaussian measure in $\mathbb{R}^d$ with mean value a and variance $\varepsilon > 0$, that is,

$$P_\varepsilon \sim (2\pi\varepsilon)^{-d/2} \exp\left(-\frac{|x-a|^2}{2\varepsilon}\right) dx.$$

Then $P_\varepsilon \to \delta_a(dx)$ weakly as $\varepsilon \downarrow 0$.

Theorem 2.2. *Let P_n and P be probability measures on $(E,\mathscr{B}(E))$, where (E,ρ) is a metric space. Then the followings are equivalent:*

(1) $P_n \to P$ weakly;

(2) $\int_E f dP_n \to \int_E f dP$ as $n \to \infty$ for every $f \in U_\rho(E)$;

(3) $\limsup_{n\to\infty} P_n(F) \le P(F)$ for every closed subset F;

(4) $\liminf_{n\to\infty} P_n(G) \ge P(G)$ for every open subset G;

(5) $\lim_{n\to\infty} P_n(A) = P(A)$ for every $A \in \mathscr{B}(E)$ such that $P(\partial A) = 0$.

Here ∂A is the boundary of A which is defined to be $\bar{A} \setminus A^o$, where A^o is the largest open subset lying inside A, that is,

$$A^o = \cup \{U : U \text{ is open and } U \subset A\},$$

and $\bar{A}$ is the closure of A, the least closed subset containing A.

The interesting point of this theorem is that the class of functions $U_\rho(E)$ may depends on the metric ρ, but the weak convergence by definition depends only on the topology which specifies the class $C_b(E)$ of bounded continuous functions on E. The advantage of item (2) lies in the fact ρ can be any metric on E which defines the same topology and thus does not alter the space $C_b(E)$.

You may recognize that (5) is the general form of convergence of distribution functions we have discussed in Part A Probability.

We develop some general tools about weak convergence of probability measures on metric spaces. A standard reference about weak convergence is the classic: *Convergence of Probability Measures* (Second Edition) by P. Billingsley, which can be your reference in your future research, but we will not follow his approach closely. The first task is to define a metric on the space of all probability measures on $(E, \mathscr{B}(E))$ whose induced topology defines the weak convergence of probability measures.

To this end we need a so-called Tychonoff's embedding theorem which says that a separable metric space E can be embedded into a compact metric space (E', ρ'). More precisely

Theorem 2.3. *(*Tychonoff's embedding theorem*) Let (E, ρ) be a* separable *metric space. Then there is a metric space (E', ρ') such that*

(1) (E', ρ') is a compact metric space,

(2) $E \subseteq E'$, and

(3) ρ and ρ' restricted on E are topologically equivalent, so if necessary, we will replace the metric ρ on E by ρ' instead.

Let $\bar{E}$ be the closure of E in (E', ρ'). Then $(\bar{E}, \rho')$ is the compact and separable *metric space.*

The proof of this theorem does not belong to this course, and is not examinable for this paper.

According to the theorem that continuous functions on a compact metric space is bounded and uniformly, we may conclude that $C(\bar{E}) = C_b(\bar{E}) = U_{\rho'}(\bar{E})$, which can be identified with uniformly continuous functions on (E, ρ). That is we have the following

Lemma 2.4. *Let (E, ρ) be a separable metric space, embedded in a compact metric space (E', ρ') as in Theorem 2.3. Without losing generality we may assume that $\rho = \rho'$ on E (otherwise we use the metric ρ' on E instead). Then every $f \in C(\bar{E})$ restricted on E is uniformly continuous w.r.t. ρ'. Conversely, every $f \in U_{\rho'}(E)$ can be extended uniquely to be a continuous function on $\bar{E}$. Therefore we will identify the space $U_{\rho'}(E)$ with $C(\bar{E})$, in particular any bounded and uniformly continuous function on E is automatically extended to be a continuous function on $\bar{E}$.*

Corollary 2.5. *Let (E, ρ) be a separable metric space. With the same notations in the previous lemma. $(\bar{E}, \rho')$ is a compact separable metric space, so that $C(\bar{E})$ equipped with the supremum norm is a separable Banach space. Therefore $U_{\rho'}(E)$ is separable.*

Example. Let $E = (a, b)$ be a open interval, where $a < b$ are two real numbers. Then $\bar{E} = [a, b]$. In Prelims Analysis II we have shown that a continuous function on (a, b) has limits at a and b (so it can be extended to be continuous function on $[a, b]$) if and only if it is uniformly continuous on (a, b).

The conclusion remains true if $a = -\infty$ or $b = \infty$, as long as you properly define the topology of the space $[-\infty, \infty]$.

Theorem 2.6. (Prohorov's metric) *Let (E, ρ) be a* separable *metric space, and $\mathscr{M}_1(E)$ denote the space of all probability measures on $(E, \mathscr{B}(E))$. By Corollary 2.5, $U_{\rho'}(E)$ is separable, so there is a countable dense subset $\{f_1, f_2, \cdots\}$ in $U_{\rho'}(E)$. Define*

$$d(P, Q) = \sum_{n=1}^{\infty} \frac{1}{2^n} (1 \wedge |P(f_n) - Q(f_n)|).$$

Then d is a metric on $\mathscr{M}_1(E)$ and $(\mathscr{M}_1(E), d)$ is separable. Let $P_n, P \in \mathscr{M}_1(E)$. Then $P_n \to P$ weakly, if and only if $d(P_n, P) \to 0$.

Proof. Proof follows from Theorem 2.2 and Corollary 2.5 immediately. □

Remark. If (E, ρ) is a *Polish space*, i.e. it is a separable and complete metric space, then $(\mathscr{M}_1(E), d)$ where d is the Prohorov's metric as constructed above is also a Polish space.

We need a characterization of compact subsets of $\mathscr{M}_1(E)$. To this end we introduce the concept of *tightness*.

Definition 2.7. *Let E be a metric space. A subset $\mathscr{L} \subset \mathscr{M}_1(E)$ is* tight*, if for every $\varepsilon > 0$, there is a compact subset $K \subset E$ such that*

$$P(K) \geq 1 - \varepsilon \quad \textit{for every } P \in \mathscr{L}. \tag{2.1}$$

We first show the following fact.

Lemma 2.8. *Let (E, ρ) be a Polish space. Then any finite subset $\mathscr{L}$ of $\mathscr{M}_1(E)$ is tight.*

Proof. By definition, we only need to show the lemma for a singleton. Let $P \in \mathscr{M}_1(E)$ and $\varepsilon > 0$. Since E is separable, for every $\delta > 0$, E can be covered by countable many balls with radius δ. Therefore, for every n, there is a sequence of *closed* balls $B_i^{(n)}$ of radius $\frac{1}{2^n}$ (where $i = 1, 2, \ldots$) such that $\cup_i B_i^{(n)} = E$ for each n. By construction

$$\lim_{k \to \infty} P\left(\cup_i^k B_i^{(n)}\right) = P(\cup_i B_i^{(n)}) = P(E) = 1.$$

Hence for each n, there is k_n such that

$$P\left(\cup_i^{k_n} B_i^{(n)}\right) > 1 - \frac{\varepsilon}{2^n}.$$

Let $K = \cap_{n=1}^{\infty} \cup_i^{k_n} B_i^{(n)}$. K is totally bounded by definition and is also closed. Since E is complete, therefore K is compact. Since

$$P(K^c) \leq \sum_{n=1}^{\infty} P\left(\left(\cup_i^{k_n} B_i^{(n)}\right)^c\right) < \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} = \varepsilon$$

and therefore $P(K) > 1 - \varepsilon$. □

We now prove the main result of this section.

Theorem 2.9. *(*Prohorov's theorem*) Let (E,ρ) be a separable metric space, and $\mathscr{L} \subset \mathscr{M}_1(E)$.*
(1) If $\mathscr{L}$ is tight, then $\mathscr{L}$ is relatively compact in $\mathscr{M}_1(E)$ equipped with Prohorov's metric.
(2) The converse of (1) is also true if E is a Polish space.

Proof. [*The proof is not examinable, too technical.*] By the Tychonoff's embedding, there is a compact metric space (E',ρ') such that $E \subset E'$, and ρ and ρ' are equivalent topologically on E. Let $\bar{E}$ be the closure of E in (E',ρ'), so that $(\bar{E},\rho')$ is compact and separable.

First prove (1). Equipped with the Prohorov's metric, $\mathscr{M}_1(E)$ is a separable metric space, hence we only need to show that $\mathscr{L}$ is sequentially relatively compact. That is, if (P_n) is any sequence in $\mathscr{L}$, we need to show one can extract a weakly convergent sub-sequence.

Since $\mathscr{B}(E) = E \cap \mathscr{B}(\bar{E})$, we can extend P_n to be a probability measure on $(\bar{E},\mathscr{B}(\bar{E}))$ by setting $Q_n(A) = P_n(A \cap E)$ for every $A \in \mathscr{B}(\bar{E})$. Since E is separable, so is $\bar{E}$, and therefore $C(\bar{E})$ is a separable too. Hence, we can choose a dense sequence $\varphi_n \in C(\bar{E})$, where $n = 1,2,\ldots$, and we can choose $\varphi_1 = 1$. For every $k = 1,2,\ldots$, $(Q_n(\varphi_k))_{n\geq 1}$ is a bounded sequence, so that it contains a convergent sub-sequence. By using Cantor's diagonal technique, we can choose a sub-sequence n_l such that $Q_{n_l}(\varphi_k) \to f_k$ as $l \to \infty$ for every k. Since for each n_l, $\phi \to Q_{n_l}(\phi)$ is linear, so that for any $\phi = \sum_{j=1}^{m} c_j\varphi_j \in \text{span}\{\varphi_k : k \geq 1\}$ then $Q_{n_l}(\phi) \to \sum_{j=1}^{m} c_j f_j = f(\phi)$. Then $|f(\phi)| \leq \|\phi\|$ as for each n_l, $|Q_{n_l}(\phi)| \leq \|\phi\|$. Clearly f is linear on the linear space $\text{span}\{\varphi_k : k \geq 1\}$. Since the closure of $\text{span}\{\varphi_k : k \geq 1\}$ is exactly $C(\bar{E})$, so that f can be uniquely extended to be a linear functional on $C(\bar{E})$, which is positive, $|f(\phi)| \leq ||\phi||$, and $f(\varphi_1) = 1$. By F. Resiz representation theorem, there is a unique probability measure Q on $(\bar{E},\mathscr{B}(\bar{E}))$, such that $Q(\phi) = f(\phi)$ for every $\phi \in C(\bar{E})$. By our construction $Q_n \to Q$ weakly as probability measures on $\bar{E}$.

Since $\mathscr{L}$ is tight, for each $m = 1,2,\ldots$ there is a compact subset K_m of E, which is also a compact subset of $\bar{E}$, such that

$$P_{n_l}(K_m) > 1 - \frac{1}{2^m} \qquad \text{for all } l = 1,2,\ldots.$$

By definition, $Q_{n_l}(K_m) = P_{n_l}(K_m)$, so it follows that (Theorem 2.2, (3))

$$Q(K_m) \geq \limsup_{l\to\infty} Q_{n_l}(K_m) = \limsup_{l\to\infty} P_{n_l}(K_m) \geq 1 - \frac{1}{2^m}$$

for every m. Therefore

$$Q(\cup_{m=1}^{\infty} K_m) = 1.$$

Define P on $(E,\mathscr{B}(E))$ to be the restriction of Q on $\mathscr{B}(E)$. Since $\cup_{m=1}^{\infty} K_m \subset E$, so $P \in \mathscr{M}_1(E)$. Let us show that $P_{n_l} \to P$ weakly.

Let F be a closed subset of E. We want to show that $P(F) \geq \limsup_{l\to\infty} P_{n_l}(F)$, which implies that $P_{n_l} \to P$ weakly. By definition, there is a closed A of $\bar{E}$ such tat $F = A \cap E$. Hence

$$\begin{aligned} P(F) &= P(A \cap E) = P(A \cap (\cup_{m=1}^{\infty} K_m) \cap E) \\ &= P(A \cap (\cup_{m=1}^{\infty} K_m)) = Q(A \cap (\cup_{m=1}^{\infty} K_m)) \\ &= Q(A) \geq \limsup_{l\to\infty} Q_{n_l}(A) = \limsup_{l\to\infty} P_{n_l}(A \cap E) \\ &= \limsup_{l\to\infty} P_{n_l}(F), \end{aligned}$$

so according to Theorem 2.2, (3), again, we can conclude that $P_n \to P$ weakly. Hence $\mathscr{L}$ is relatively compact.

Next we prove (2). Thus assume that E is a Polish space and $\mathscr{L} \subset \mathscr{M}_1(E)$ is relatively compact. For any $\delta > 0$, since E is separable, there is a sequence open balls $B_i(\delta)$ which forms a cover of E. Let $G_n = \cup_{i \le n} B_i(\delta)$. We claim that for every $\varepsilon > 0$ there is n, such that

$$\inf_{P \in \mathscr{L}} P(G_n) \ge 1 - \varepsilon.$$

Suppose it were not true, for every n, there is $P_n \in \mathscr{L}$, such that $P_n(G_n) < 1 - \varepsilon$. Since $\mathscr{L}$ is relatively compact, so there is a sub-sequence n_k such that $P_{n_k} \to P$ weakly, by Theorem 2.2, (2), since each G_n is open and $G_n \uparrow E$,

$$P(G_n) \le \liminf_{k \to \infty} P_{n_k}(G_n) \le \liminf_{k \to \infty} P_{n_k}(G_{n_k}) \le 1 - \varepsilon$$

and therefore

$$1 = P(E) = \lim_{n \to \infty} P(G_n) \le 1 - \varepsilon$$

which is impossible.

Apply the argument for $\delta = \frac{1}{2^k}$ and with ε replaced by $\varepsilon / 2^k$. Then for every $\varepsilon > 0$, and for every $k = 1, 2, \ldots$, there is an n_k, and n_k open balls $B_i^{(k)}(2^{-k})$ with radius 2^{-k} (where $i = 1, 2, \ldots, n_k$) such that

$$\inf_{P \in \mathscr{L}} P\left(G^{(k)}\right) \ge 1 - \frac{\varepsilon}{2^k}$$

where

$$G^{(k)} = \bigcup_{i=1}^{n_k} B_i^{(k)}(2^{-k}).$$

Let $G = \cap_{k=1} G^{(k)}$. Then G is totally bounded. Since E is complete, the closure of G, denoted by K, is compact in E. By definition

$$P(K) \ge P(G) \ge 1 - \varepsilon$$

for every $P \in \mathscr{L}$. This proves that $\mathscr{L}$ is tight. □

3 Weak convergence of random variables

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, and (E, ρ) be a metric space. Recall that a random variable X with values in E is a measurable mapping Ω to E, that is, for every $B \in \mathscr{B}(E)$, $X^{-1}(B) \in \mathscr{F}$. The law or distribution of X is defined to be the probability measure P_X on $(E, \mathscr{B}(E))$ defined by

$$P_X(B) = \mathbb{P}[X \in B] \quad \text{for } B \in \mathscr{B}(E).$$

We are therefore able to give the following definition of convergence for random variables *in law (or called convergence in distribution)*. Namely, we say a sequence of random variables X_1, $X_2, \ldots$ (they may be defined on different probability spaces !) taking values in a metric space E converges weakly to a random variable X, if the corresponding sequences of distributions $P_{X_n} \to P_X$ weakly. More precisely if X_n is a random variable on some probability space $(\Omega_n, \mathscr{F}_n, \mathbb{P}_n)$ with values in E, then $X_n \to X$ in distribution if for every bounded continuous function φ on E we have

$$\int_{\Omega_n} \varphi(X_n) d\mathbb{P}_n \to \int_{\Omega} \varphi(X) d\mathbb{P}$$

as $n \to \infty$.

Lemma 3.1. *Let X_n, X ($n = 1,2,\ldots$) be random variables on the same probability space $(\Omega,\mathscr{F},\mathbb{P})$ taking values in a metric space (E,ρ). If $X_n \to X$ in probability, i.e. for every $\delta > 0$,*

$$\mathbb{P}[\rho(X_n,X) > \delta] \to 0 \text{ as } n \to \infty,$$

then for every $\varphi \in U_\rho(E)$

$$\mathbb{E}[|\varphi(X_n) - \varphi(X)|] \to 0.$$

In particular $X_n \to X$ in distribution.

The proof of this lemma is left as an exercise.

Theorem 3.2. *(*Skorohod (1956) and Dudley (1968)*) Let (E,ρ) be a separable metric space, and P_n, P (where $n = 1,2,\ldots$) be probability measures on $(E,\mathscr{B}(E))$, such that $P_n \to P$ weakly. Then there is a probability space $(\Omega,\mathscr{F},\mathbb{P})$ and random variables $X_n, X : \Omega \to E$ such that $P_{X_n} = P_n$ for $n = 1,2,\ldots$, $P_X = P$, such that $X_n \to X$ almost surely.*

The proof of this important theorem may be found in Ikeda and Watanabe [16], page 9, Theorem 2.7. The proof is not examinable in C8.6.

4 Laws on spaces of continuous paths

The main reference for this part is Stroock-Varadhan [25, Section 1.3]. For simplicity, we use $\mathbb{C}(\mathbb{R}^d)$ to denote the path space $C([0,\infty);\mathbb{R}^d)$. Let us equip $\mathbb{C}(\mathbb{R}^d)$ with the following metric ρ defined by

$$\rho(x,y) = \sum_{n=1}^{\infty} \frac{1}{2^n} \left(1 \wedge \sup_{t\in[0,n]} |x(t) - y(t)| \right) \tag{4.1}$$

for any $x,y \in \mathbb{C}(\mathbb{R}^d)$. The convergence with respect to the metric ρ is the uniform convergence over any compact subset of $[0,\infty)$.

Lemma 4.1. *$\mathbb{C}(\mathbb{R}^d)$ equipped with the metric ρ given in (4.1) is a Polish space.*

The proof is left as an exercise.

The coordinate process on the path space $\mathbb{C}(\mathbb{R}^d)$ is denoted by X_t (where $t \geq 0$) which is defined by

$$X_t(w) = w(t) \quad \text{for every } w \in \mathbb{C}(\mathbb{R}^d),$$

for every $t \geq 0$. The natural filtration of σ-algebras generated by the coordinate process $\{X_t : t \geq 0\}$, is given as, by definition, $\mathscr{F}_t^0 = \sigma\{X_s : s \leq t\}$, and

$$\mathscr{F}^0 = \mathscr{F}_\infty^0 = \sigma\{X_t : t \geq 0\}.$$

Lemma 4.2. *The Borel σ-algebra on $(\mathbb{C}(\mathbb{R}^d),\rho)$ coincides with $\mathscr{F}^0 = \sigma\{X_t : t \geq 0\}$.*

The proof is left as an exercise. The following is the main result of this section.

Theorem 4.3. *Let $\mathscr{L}$ be a family of probability measures on $(\mathbb{C}(\mathbb{R}^d), \mathscr{B}(\mathbb{C}(\mathbb{R}^d)))$. Then $\mathscr{L}$ is relatively compact if and only if*

$$\lim_{L\uparrow\infty} \inf_{P\in\mathscr{L}} P[w : |w(0)| \leq L] = 1 \tag{4.2}$$

and for any $\lambda > 0$ and $T > 0$,

$$\lim_{\delta\downarrow 0} \inf_{P\in\mathscr{L}} P\left[w : \sup_{s,t\in[0,T];|t-s|<\delta} |w(t) - w(s)| \leq \lambda\right] = 1. \tag{4.3}$$

Proof. $(\mathbb{C}(\mathbb{R}^d), \rho)$ is a Polish space, so that by Prohorov's theorem, $\mathscr{L}$ is relatively compact if and only if it is tight.

Proof of Necessity. Suppose $\mathscr{L}$ is tight, then for every $\varepsilon > 0$ there is a compact subset $K_\varepsilon \subset \mathbb{C}(\mathbb{R}^d)$, such that

$$\inf_{P\in\mathscr{L}} P(K_\varepsilon) \geq 1 - \varepsilon.$$

According to the Ascoli-Arzela's theorem, a closed subset $K_\varepsilon \subset \mathbb{C}(\mathbb{R}^d)$ is compact if and only if it is bounded and equivalently continuous on any closed interval $[0, T]$, which are equivalent to the following two conditions:

$$\sup_{w\in K_\varepsilon} \sup_{t\in[0,T]} |w(t)| < \infty \tag{4.4}$$

and

$$\lim_{\delta\downarrow 0} \sup_{w\in K_\varepsilon} \sup_{s,t\in[0,T];|t-s|<\delta} |w(t) - w(s)| = 0. \tag{4.5}$$

Due to the equivalent continuity (4.5), the first condition (4.4) may be replaced by the following weaker one:

$$\sup_{w\in K_\varepsilon} |w(0)| < \infty. \tag{4.6}$$

Let $\alpha = \sup_{w\in K_\varepsilon} |w(0)|$. Then for any $L > \alpha$ and $P \in \mathscr{L}$, we have

$$P(K_\varepsilon) = P[w \in K_\varepsilon : |w(0)| < \infty] = P[w \in K_\varepsilon : |w(0)| \leq \alpha] \leq P[w : |w(0)| \leq L]$$

and therefore

$$\inf_{P\in\mathscr{L}} P[w : |w(0)| \leq L] \geq \inf_{P\in\mathscr{L}} P(K_\varepsilon) \geq 1 - \varepsilon$$

for every $\varepsilon > 0$, so (4.2) holds.

Similarly, by (4.5), for every $\lambda > 0$, there is $\delta_0 > 0$ such that

$$\sup_{w\in K_\varepsilon} \sup_{s,t\in[0,T];|t-s|<\delta_0} |w(t) - w(s)| \leq \lambda$$

so that for any $0 < \delta < \delta_0$ we have

$$\begin{aligned} P(K_\varepsilon) &\leq P\left[w \in K_\varepsilon : \sup_{s,t\in[0,T];|t-s|<\delta_0} |w(t) - w(s)| \leq \lambda\right] \\ &\leq P\left[w : \sup_{s,t\in[0,T];|t-s|<\delta} |w(t) - w(s)| < \lambda\right] \end{aligned}$$

which yields that

$$\inf_{P\in\mathscr{L}} P\left[w: \sup_{s,t\in[0,T];|t-s|<\delta}|w(t)-w(s)|\leq\lambda\right]\geq 1-\varepsilon$$

for any $\varepsilon>0$, thus (4.3) must be true.

Proof of Sufficiency. Suppose (4.2, 4.3) hold. For every $\varepsilon>0$, by (4.2), there is $L_\varepsilon>0$ such that

$$\inf_{P\in\mathscr{L}} P[w:|w(0)|\leq L_\varepsilon]>1-\frac{\varepsilon}{2},$$

and, by definition of (4.3), for every $n=1,2,\ldots$, there is $\delta_n(\varepsilon)>0$ such that

$$\inf_{P\in\mathscr{L}} P\left[w: \sup_{s,t\in[0,n];|t-s|<\delta_n(\varepsilon)}|w(t)-w(s)|\leq\frac{1}{n}\right]>1-\frac{\varepsilon}{2^{n+1}}.$$

Let

$$K_\varepsilon=\bigcap_{n=1}^{\infty}\left\{w: \sup_{s,t\in[0,n];|t-s|<\delta_n(\varepsilon)}|w(t)-w(s)|\leq\frac{1}{n}\right\}\cap\{w:|w(0)|\leq L_\varepsilon\}.$$

Then

$$P\left(\mathbb{C}(\mathbb{R}^d)\setminus K_\varepsilon\right)\leq\sum_{n=1}^{\infty}\frac{\varepsilon}{2^{n+1}}+\frac{\varepsilon}{2}=\varepsilon$$

for any $P\in\mathscr{L}$, which implies that

$$P(K_\varepsilon)\geq 1-\varepsilon \quad \text{for every } P\in\mathscr{L}.$$

According to the Ascoli-Arzela theorem, K_ε is compact in $(\mathbb{C}(\mathbb{R}^d),\rho)$. Hence $\mathscr{L}$ is tight, and therefore is relatively compact. The proof is complete. □

Let us draw some consequences which give very useful forms in applications.

Theorem 4.4. *Let (P_n) be a sequence of probability measures on $(\mathbb{C}(\mathbb{R}^d),\mathscr{B}(\mathbb{C}(\mathbb{R}^d)))$. Then (P_n) is relatively compact if and only if*

$$\lim_{L\uparrow\infty}\inf_n P_n[w:|w(0)|\leq L]=1 \tag{4.7}$$

and for any $\lambda>0$ and $T\in\mathbb{R}_+$,

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty} P_n\left[w: \sup_{s,t\in[0,T];|t-s|<\delta}|w(t)-w(s)|\geq\lambda\right]=0. \tag{4.8}$$

The following form is the most useful form to prove convergence of continuous stochastic processes.

Theorem 4.5. *Let $\{X_t^{(n)}:t\geq 0\}$ be a sequence of stochastic processes valued in $\mathbb{R}^d$ on a probability space $(\Omega,\mathscr{F},\mathbb{P})$ with* continuous sample paths. *Suppose*

$$\lim_{L\uparrow\infty}\inf_n\mathbb{P}\left[|X_0^{(n)}|\leq L\right]=1 \tag{4.9}$$

and for any $\lambda > 0$ and $T \in \mathbb{R}_+$,

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty}\mathbb{P}\left[\sup_{s,t\in[0,T];|t-s|<\delta}|X_t^{(n)}-X_s^{(n)}|\geq\lambda\right]=0. \tag{4.10}$$

Then there is a sub-sequence n_k and a continuous stochastic process X valued in $\mathbb{R}^d$, such that $X_\cdot^{(n_k)} \to X_\cdot$ in law.

As an application, we can prove the following

Theorem 4.6. *(*Kolmogorov's criterion*) Let $(X^{(n)})$ be a sequence of $\mathbb{R}^d$-valued continuous stochastic processes on a probability space $(\Omega,\mathscr{F},\mathbb{P})$ such that*

$$\lim_{L\uparrow\infty}\inf_{n}\mathbb{P}\left[|X_0^{(n)}|\leq L\right]=1 \tag{4.11}$$

and there are $\alpha > 0$ and $\beta > 0$ such that for every $T > 0$

$$\mathbb{E}\left[|X_t^{(n)}-X_s^{(n)}|^{\alpha}\right]\leq C|t-s|^{1+\beta} \tag{4.12}$$

for all $0 \leq s,t \leq T$, where $C_T > 0$ depends only on T but independent of n. Then the collection of the laws of $(X^{(n)})$ is relatively compact.

5 Compactness for laws of continuous semi-martingales

In general it is highly non trivial to estimate the probability appearing in (4.10). Here we propose a way to handle it. We will make further comments after we have considered the Skorohord topology in next section.

Let us consider the distributions on the path space $C([0,\infty);\mathbb{R}^d)$. For a path $x:[0,\infty)\to\mathbb{R}^d$, given increment size $\varepsilon > 0$ and duration $T > 0$, we may define its *entropy numbers* $\mathscr{N}(x,\varepsilon,T)$ and a *minimal gap* $\delta(x,\varepsilon,T)$. Let $T_0 = 0$,and T_{j+1} be defined inductively by $T_{j+1} = \infty$ if $T_j = \infty$, and

$$T_{j+1}=\inf\left\{t\geq T_j:|x(t)-x(T_j)|\geq\frac{\varepsilon}{4}\right\}\quad\text{ if } T_j<\infty$$

for $j = 0,1,2,\ldots$, where $\inf\emptyset = \infty$. By definition, if x is a continuous path, then

$$T_{j+1}-T_j>0\quad\text{ if } T_j<\infty,$$

and

$$|x(T_{j+1})-x(T_j)|=\frac{\varepsilon}{4}\quad\text{ if } T_{j+1}<\infty.$$

The entropy number

$$\mathscr{N}(x,\varepsilon,T)=\inf\left\{j:T_{j+1}>T\right\}$$

and the minimal gap

$$\delta(x,\varepsilon,T)=\inf\left\{T_j-T_{j-1}:1\leq j\leq\mathscr{N}(x,\varepsilon,T)\right\}.$$

Then (Exercise)

$$\sup_{s,t\leq T;|t-s|<\delta(x,\varepsilon,T)}|x(t)-x(s)|\leq\varepsilon. \tag{5.1}$$

The following lemma follows from the definition.

Lemma 5.1. *Suppose $(X_t)_{t\geq 0}$ is a stochastic process with continuous sample paths on a probability space $(\Omega,\mathscr{F},\mathbb{P})$, and we define T_j as above with $x(t)$ replaced by X_t.*

1) If $(X_t)_{t\geq 0}$ is adapted to a filtration $(\mathscr{F}_t)_{t\geq 0}$, then all T_j are $(\mathscr{F}_t)_{t\geq 0}$-stopping times.

2) Let $T>0$, $\delta>0$ and $\varepsilon>0$. Then

$$\mathbb{P}\left[\sup_{s,t\leq T;|t-s|<\delta}|X_t-X_s|>\varepsilon\right]\leq\mathbb{P}\left[\delta(X,\varepsilon,T)<\delta\right]. \tag{5.2}$$

The following lemma, adopted from the discussion in Stroock-Varadhan [25, Section 1.4], providing the necessary estimates for controlling (5.2).

Lemma 5.2. *Suppose $X=M+A$ is a d-dimensional continuous semi-martingale on a filtered probability space $(\Omega,\mathscr{F},\mathscr{F}_t,\mathbb{P})$, so that M is a continuous local martingale and A is adapted with finite variations, valued in $\mathbb{R}^d$. Assume that*

$$|dA_t|\ll Cdt \quad \text{and } |d\langle M,M\rangle|\ll Cdt$$

for some constants. Let $\varepsilon>0$ and $T>0$. Define stopping times T_j as above. Then

1) There is a constant C_1 depending only $\varepsilon>0$ and C, we have

$$\mathbb{E}\left[1_{\left\{T_j-T_{j-1}\leq\delta\right\}}|\mathscr{F}_{T_{j-1}}\right]\leq(C_1+C_2)\delta$$

on $T_{j-1}<\infty$, for every $\delta>0$.

2) There is a constant $\alpha\in(0,1)$ depending only on ε and C such that

$$\mathbb{P}\left[\mathscr{N}(x,\varepsilon,T)\geq k\right]\leq e^T\alpha^k$$

for all $k=1,2,\ldots$.

Proof. We assume that all X, M and A have zero initial value. Apply Itô's formula we obtain

$$\begin{aligned}f(X_{(T_j-T_{j-1})\wedge\delta+T_{j-1}}-X_{T_{j-1}})=&\int_{T_{j-1}}^{(T_j-T_{j-1})\wedge\delta+T_{j-1}}\nabla f(X_s-X_{T_{j-1}})dM_s\\&+\int_{T_{j-1}}^{(T_j-T_{j-1})\wedge\delta+T_{j-1}}\nabla f(X_s-X_{T_{j-1}})dA_s\\&+\frac{1}{2}\int_{T_{j-1}}^{(T_j-T_{j-1})\wedge\delta+T_{j-1}}\frac{\partial^2 f}{\partial x^i\partial x^j}(X_s-X_{T_{j-1}})d\left\langle M^i,M^j\right\rangle_s\end{aligned}$$

on $T_{j-1}<\infty$, where

$$(T_j-T_{j-1})\wedge\delta+T_{j-1}=T_j\wedge(\delta+T_{j-1}).$$

By taking conditional expectation on $\mathscr{F}_{T_{j-1}}$, we obtain that

$$\begin{aligned}\mathbb{E}\left[f(X_{T_j\wedge(T_{j-1}+\delta)}-X_{T_{j-1}})1_{\left[T_{j-1}<\infty\right]}|\mathscr{F}_{T_{j-1}}\right]=&\,\mathbb{E}\left[\int_{T_{j-1}}^{T_j\wedge(T_{j-1}+\delta)}\nabla f(X_s-X_{T_{j-1}})dA_s|\mathscr{F}_{T_{j-1}}\right]\\&+\frac{1}{2}\mathbb{E}\left[\int_{T_{j-1}}^{T_j\wedge(T_{j-1}+\delta)}\frac{\partial^2 f}{\partial x^i\partial x^j}(X_s-X_{T_{j-1}})d\left\langle M^i,M^j\right\rangle_s|\mathscr{F}_{T_{j-1}}\right].\end{aligned}$$

Let us choose a bump non-negative function $f \geq 0$ such that $f(|x|) = 1$ when $|x| = \frac{\varepsilon}{4}$, f is smooth with a compact support. f depends only on $\varepsilon > 0$.

$$\begin{aligned}
\mathbb{E}\left[1_{[T_j - T_{j-1} \leq \delta, T_{j-1} < \infty]} | \mathscr{F}_{T_{j-1}}\right] &\leq \mathbb{E}\left[f(X_{T_j \wedge (T_{j-1}+\delta)} - X_{T_{j-1}}) 1_{[T_{j-1} < \infty]} | \mathscr{F}_{T_{j-1}}\right] \\
&= \mathbb{E}\left[\int_{T_{j-1}}^{T_j \wedge (T_{j-1}+\delta)} \nabla f(X_s - X_{T_{j-1}}) dA_s | \mathscr{F}_{T_{j-1}}\right] \\
&+ \frac{1}{2}\mathbb{E}\left[\int_{T_{j-1}}^{T_j \wedge (T_{j-1}+\delta)} \frac{\partial^2 f}{\partial x^i \partial x^j}(X_s - X_{T_{j-1}}) d\left\langle M^i, M^j \right\rangle_s | \mathscr{F}_{T_{j-1}}\right]
\end{aligned}$$

By assumptions, we have

$$\mathbb{E}\left[\int_{T_{j-1}}^{T_j \wedge (T_{j-1}+\delta)} \nabla f(X_s - X_{T_{j-1}}) dA_s | \mathscr{F}_{T_{j-1}}\right] \leq C_1 \delta$$

and

$$\frac{1}{2}\mathbb{E}\left[\int_{T_{j-1}}^{T_j \wedge (T_{j-1}+\delta)} \frac{\partial^2 f}{\partial x^i \partial x^j}(X_s - X_{T_{j-1}}) d\left\langle M^i, M^j \right\rangle_s | \mathscr{F}_{T_{j-1}}\right] \leq C_2 \delta$$

for any $\delta > 0$, where C_1 and C_2 are two constants depending only on f. Then

$$\mathbb{E}\left[1_{[T_j - T_{j-1} \leq \delta, T_{j-1} < \infty]} | \mathscr{F}_{T_{j-1}}\right] \leq (C_1 + C_2)\delta$$

for every $\delta > 0$, on $T_{j-1} < \infty$.

Now we can work out the second estimate. Firstly

$$\begin{aligned}
\mathbb{E}\left[e^{-(T_{j+1} - T_j)} | \mathscr{F}_{T_j}\right] &= \mathbb{E}\left[e^{-(T_{j+1} - T_j)} 1_{[T_{j+1} - T_j < \beta]} | \mathscr{F}_{T_j}\right] \\
&+ \mathbb{E}\left[e^{-(T_{j+1} - T_j)} 1_{[T_{j+1} - T_j \geq \beta]} | \mathscr{F}_{T_j}\right] \\
&\leq \mathbb{P}\left[T_{j+1} \leq T_j + \beta | \mathscr{F}_{T_j}\right] + e^{-\beta}\mathbb{E}\left[1_{[T_{j+1} - T_j \geq \beta]} | \mathscr{F}_{T_j}\right] \\
| \mathscr{F}_{T_{j-1}} &= e^{-\beta} + (1 - e^{-\beta})\mathbb{P}\left[T_{j+1} \leq T_j + \beta | \mathscr{F}_{T_j}\right] \\
&\leq e^{-\beta} + (1 - e^{-\beta})(C_1 + C_2)\beta
\end{aligned}$$

Choose $\beta > 0$ small enough so that

$$e^{-\beta} + (1 - e^{-\beta})(C_1 + C_2)\beta = \alpha < 1.$$

Then

$$\mathbb{E}\left[e^{-(T_{j+1} - T_j)} | \mathscr{F}_{T_j}\right] \leq \alpha < 1$$

where α depends on $\varepsilon > 0$ and $C_1 + C_2$ only. It follows that

$$\begin{aligned}
\mathbb{E}\left[e^{-T_{j+1}} | \mathscr{F}_{T_j}\right] &= e^{-T_j}\mathbb{E}\left[e^{-(T_{j+1} - T_j)} | \mathscr{F}_{T_j}\right] \\
&\leq \alpha e^{-T_j}
\end{aligned}$$

so that

$$\mathbb{E}\left[e^{-T_{j+1}}|\mathscr{F}_{T_j}\right] \leq \alpha^j$$

for all j. Now

$$\begin{aligned}\mathbb{P}[\mathscr{N}(x,\varepsilon,T) \geq k] = \mathbb{P}[T_k \leq T] &\leq \mathbb{E}\left[e^{-T_k+T}\right] \\ &\leq e^T \alpha^{k-1}\end{aligned}$$

for all k. □

Theorem 5.3. *Let* $\left\{X_t^{(n)} : t \geq 0\right\}$ *be a sequence of continuous semi-martingales valued in* $\mathbb{R}^d$ *on a probability space* $(\Omega, \mathscr{F}, \mathbb{P})$. *Let* $X_t^{(n)} = X_0^{(n)} + M_t^{(n)} + A_t^{(n)}$ *be the semi-martingale decomposition. Suppose*

$$\lim_{L\uparrow\infty} \inf_n \mathbb{P}\left[|X_0^{(n)}| \leq L\right] = 1 \tag{5.3}$$

and there is a constant $C \geq 0$ *independent of n such that*

$$|dA_t^{(n)}| \ll Cdt \quad \text{ and } |d\left\langle M^{(n)}, M^{(n)}\right\rangle| \ll Cdt.$$

Then the totality of the laws of the sequence $\left\{X^{(n)} : n = 1, 2, \cdots\right\}$ *of semi-martingales is tight.*

Proof. According to the previous lemma we have for every k

$$\begin{aligned}\mathbb{P}\left[\delta(X^{(n)}, \varepsilon, T) < \delta\right] &\leq \mathbb{P}\left[\inf_{1\leq j\leq k}\left(T_j - T_{j-1}\right) \leq \delta\right] + \mathbb{P}[\mathscr{N}(x,\varepsilon,T) \geq k] \\ &\leq \sum_{j=1}^k \mathbb{P}\left[T_j - T_{j-1} \leq \delta\right] + \mathbb{P}[\mathscr{N}(x,\varepsilon,T) \geq k] \\ &\leq k(C_1 + C_2)\delta + e^T \alpha^{k-1}.\end{aligned}$$

for any $k \geq 1$. Taking sup then letting $\delta \downarrow 0$ we obtain that

$$\limsup_{\delta\downarrow 0} \sup_n \mathbb{P}\left[\delta(X^{(n)}, \varepsilon, T) < \delta\right] \leq e^T \alpha^{k-1}$$

for any k. sending $k \to \infty$ we deduce that

$$\limsup_{\delta\downarrow 0} \sup_n \mathbb{P}\left[\delta(X^{(n)}, \varepsilon, T) < \delta\right] = 0$$

which completes the proof. □

6 Wasserstein's metric and weak convergence

According to Prohorov's theorem, weak convergence of probability measures on a separable metric space (E, ρ) is a metric topology, but the Prohorov metric is less explicit and is difficult to use for analysis on the space of probability measures. The Wasserstein distance we are going to introduce is, more or less, capable of dealing with weak convergence, and much easy for analysis. This is the main reason why the Wasserstein distance plays an important rôle in the current research on probability metric spaces. We follow the exposition in Villani [30, Section 6].

Let (E,ρ) be a Polish space with its Borel σ-algebra. If $P,Q \in \mathscr{M}_1(E)$ are two probability measures on $(E,\mathscr{B}(E))$, then $\Pi(P,Q)$ denotes the collection of all probability measures Θ on the product space $E\times E$ such that the marginal distributions are P and Q.

Lemma 6.1. *(*Wasserstein's distance*) Let $P,Q \in \mathscr{M}_1(E)$ and $p \geq 1$. Define*

$$W_p(P,Q) = \left[\inf_{\Theta\in\Pi(P,Q)} \int_{E\times E} \rho(x,y)^p \Theta(dx,dy)\right]^{1/p}$$

which is non-negative (maybe even infinity). Then W_p is a metric on the space $\mathscr{P}_p(E)$ of all probability measures $P \in \mathscr{M}_1(E)$ such that

$$\int_{E\times E} \rho(x,y)^p P(dx) < \infty$$

for some $y \in E$ (and therefore for all $y \in E$).

The proof is left as an exercise. W_1 is called the Kantorovich-Rubinstein distance, which has a duality representation

$$W_1(P,Q) = \sup_{f\in \mathrm{Lip}(E), ||f||_L \leq 1} \left\{ \int_E f dP - \int_E f dQ \right\} \tag{6.1}$$

where $\mathrm{Lip}(E)$ denotes the collection of all Lipschitz functions, and $||f||_L$ denotes the Lipschitz norm of a function $f : E \to \mathbb{R}$, that is,

$$\|f\|_L = \inf\{M : |f(x)-f(y)| \leq M\rho(x,y) \text{ for any } x,y\in E\}.$$

Lemma 6.2. *Let (E,ρ) be a Polish space and $p \geq 1$. Suppose $P_n, P \in \mathscr{P}_p(E)$ and $P_n \to P$ weakly. Then the following four statements are equivalent:*

1) For some $y \in E$ (and therefore for all $y \in E$)

$$\int_E \rho(x,y)^p P_n(dx) \to \int_E \rho(x,y)^p P(dx).$$

2) For some $y \in E$ (and therefore for all $y \in E$)

$$\limsup_{n\to\infty} \int_E \rho(x,y)^p P_n(dx) \leq \int_E \rho(x,y)^p P(dx).$$

3) For some $y \in E$ (and therefore for all $y \in E$)

$$\lim_{R\uparrow\infty} \limsup_{n\to\infty} \int_{\{\rho(x,y)>R\}} \rho(x,y)^p P_n(dx) = 0.$$

4) For every $f \in C(E)$ with growth at most by $d(y,\cdot)^p$ (for some and therefore for all $y \in E$), i.e there is a constant C such that

$$|f(x)| \leq C\left[1+d(y,x)^p\right]$$

for all $x \in E$, then

$$\int_E f(x)P_n(dx) \to \int_E f(x)P(dx) \quad \text{as } n\to\infty.$$

If 4) is satisfied, then we say $P_n \to P$ weakly in $\mathscr{P}_p(E)$ as $n \to \infty$.

The main interest of the Wasserstein distance lies in the following result, which is much clean than any other tightness criterion we have done so far.

Theorem 6.3. *Let (E,ρ) be a Polish space and $p \geq 1$. Suppose $P_n \in \mathscr{P}_p(E)$ (where $n = 1,2,\ldots$) is a Cauchy sequence with respect to the Wasserstein distance W_p. Then $\{P_n : n \geq 1\}$ is tight.*

Proof. The proof is similar to the proof of Lemma 2.8. Without losing a generality we may assume that $p = 1$ as $W_1 \leq W_p$. Since (P_n) is W_1-Cauchy, so that by the Triangle inequality, we deduce that

$$\int_E \rho(a,x)P_n(dx) \leq W_1(\delta_a, P_n) \leq W_1(\delta_a, P_1) + W_1(P_1, P_n) \tag{6.2}$$

which is bounded uniformly in n, though the bound may depend on $a \in E$ which is a fixed point.

Let $\varepsilon > 0$. For every $l = 1,2,\ldots$, there is an n_l such that

$$W_1(P_i, P_j) < \frac{1}{2^{2l}}\varepsilon^2 \quad \text{for any } i,j \geq n_l. \tag{6.3}$$

Since $\{P_1, \cdots, P_{n_l}\}$ is tight (Lemma 2.8), so there is a compact subset $K_l \subset E$ such that

$$P_j(K_l) > 1 - \frac{1}{2^l}\varepsilon \quad \text{for } j = 1,\ldots,n_l. \tag{6.4}$$

For each $l = 1,2,\ldots$, K_l is compact so that it is totally bounded (Theorem 1.3), thus there are finite points $a_1^{(l)},\ldots,a_{m_l}^{(l)} \in E$, the totality of open balls centered at $x_j^{(l)}$ with radius $\varepsilon/2^l$ cover K_l, that is

$$G^{(l)} \equiv \bigcup_{j=1}^{m_l} B_{a_j^{(l)}}(2^{-l}\varepsilon) \supseteq K_l. \tag{6.5}$$

Let $U^{(l)} = \left\{ z \in E : \rho(z, G^{(l)}) < 2^{-l}\varepsilon \right\}$. Then $F^{(l)} \equiv \bigcup_{j=1}^{m_l} \overline{B_{a_j^{(l)}}(2^{-l+1}\varepsilon)} \supseteq U^{(l)}$ for each l, and by definition

$$P_n(U^{(l)}) \geq P_n(G^{(l)}) \geq P_n(K_l) > 1 - \frac{\varepsilon}{2^l}$$

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

Let

$$\phi_l(x) = 0 \vee \left(1 - \frac{\rho(x, G^{(l)})}{2^{-l}\varepsilon}\right)$$

which is Lipschitz continuous on E with Lipschitz constant smaller or equal to $2^l\varepsilon^{-1}$, and

$$1_{G^{(l)}} \leq \phi_l \leq 1_{U^{(l)}}$$

so that for $n > n_l$

$$\begin{aligned} P_n(U^{(l)}) &\geq \int_E \phi_l dP_n = \int_E \phi_l dP_{n_l} + \int_E \phi_l dP_n - \int_E \phi_l dP_{n_l} \\ &\geq \int_E \phi_l dP_{n_l} - 2^l\varepsilon^{-1} W_1(P_n, P_{n_l}) \\ &\geq P_{n_l}(G^{(l)}) - 2^l\varepsilon^{-1}\frac{\varepsilon^2}{2^{2l}} \\ &\geq 1 - 2\frac{\varepsilon}{2^l} \end{aligned}$$

where the second inequality follows from (6.1). Therefore

$$P_n(F^{(l)}) \geq P_n(U^{(l)}) \geq 1 - \frac{\varepsilon}{2^{l-1}}$$

for every $l = 1, 2, \ldots$ and for all $n = 1, 2, \ldots$. Let

$$K = \bigcap_{l=1}^{\infty} F^{(l)}.$$

Then K is closed and totally bounded, and therefore K is compact, and

$$P_n(E \setminus K) \leq \sum_{l=1}^{\infty} P_n(E \setminus F^{(l)}) \leq \sum_{l=1}^{\infty} \frac{\varepsilon}{2^{l-1}} = 2\varepsilon$$

for every $n = 1, 2, \cdots$. Therefore by definition (P_n) is tight. □

The following theorem demonstrates why Wasserstein distances are important for dealing with weak convergence.

Theorem 6.4. *Let (E, ρ) be a Polish space and $p \geq 1$. Suppose $P_n, P \in \mathscr{P}_p(E)$ and $P_n \to P$ weakly in $\mathscr{P}_p(E)$, that is, for every $f \in C(E)$ with growth at most by $d(y, \cdot)^p$ (for some and therefore for all $y \in E$), i.e. there is a constant C such that*

$$|f(x)| \leq C[1 + d(y, x)^p]$$

for all $x \in E$, then

$$\lim_{n \to \infty} \int_E f(x) P_n(dx) = \int_E f(x) P(dx),$$

if and only if

$$W_p(P_n, P) \to 0.$$

7 The Skorohod topology

Not all interesting stochastic processes have continuous sample paths. For example, the sample paths of Lèvy processes are only right continuous having left limits. In this part we study the weak convergence on such path spaces.

Let (S, d) be a metric space. We assume that d is bounded, otherwise using $d \wedge 1$ or $\frac{d}{1+d}$ instead. A function w defined on $[0, \infty)$ with values in S is called a path in S. Several quantities associated with a path can be introduced, which may measure some aspects of the regularity of a path in a metric space.

If $w : [0, \infty) \to S$ is a path in S, the *oscillation* of the path w over a subset $A \subset [0, \infty)$ may be defined by

$$\omega(w, A) = \sup_{s,t \in A} d(w(t), w(s)) \tag{7.1}$$

where supremum assumes ∞ if $\rho(w(t), w(s))$ are unbounded over $s, t \in A$. Given $T > 0$ and $\delta > 0$, the *modules of continuity* of w is given by

$$\omega(w, \delta, T) = \sup_{s,t \in [0,T]; |s-t| < \delta} d(w(t), w(s)). \tag{7.2}$$

By definition, w is continuous (so uniformly continuous) on $[0, T]$ if and only if $\omega(w, \delta, T) \to 0$ as $\delta \downarrow 0$. By using modules of continuity we may restate the Ascoli-Arzelà theorem as the following.

Theorem 7.1. *(*Ascoli-Arzel*à) A subset $K \subset \mathbb{C}(\mathbb{R}^d)$ is relatively compact if and only if*

$$\sup_{w\in K}\sup_{s\in[0,T]}|w(s)|<\infty$$

and

$$\lim_{\delta\downarrow 0}\sup_{w\in K}\omega(w,\delta,T)=0$$

for every $T>0$.

A path $w:[0,\infty)\to S$ is called a c*à*dl*à*g (this is the French abbreviation of "*continue à droite limites à gauche*") path in a metric space S, if w is right continuous on $[0,\infty)$ and has left limits at every point $t\in(0,\infty)$, that is, $\lim_{s\downarrow t}w(s)=w(t)$ for every $t\geq 0$ and $w(t-)=\lim_{s\uparrow t}w(s)$ exists for every $t>0$. $\mathbb{D}(S)$ (notation used in Jacod and Shiryaev: *Limit Theorems for Stochastic Processes*) or $D_S[0,\infty)$ (notation used in Ethier and Kurtz: *Markov Processes - Characterization and Convergence*) denotes the space of all c*à*dl*à*g paths in S.

An obvious topology on $\mathbb{D}(S)$, like on the space $\mathbb{C}(S)$, is the uniform convergence over any compact subset of $[0,\infty)$, which is the metric topology induced by the metric for example

$$\rho(w^{(1)},w^{(2)})=\sum_{n=1}^{\infty}\frac{1}{2^n}\left(1\wedge\sup_{t\in[0,n]}d(w^{(1)}(t),w^{(2)}(t))\right)$$

for $w^{(i)}$ are two paths in S. While, even if (S,d) is a Polish space, $\mathbb{D}(S)$ with the above metric ρ may be not separable. In order to study the weak convergence of laws on $\mathbb{D}(S)$ via the Prohorov theorem, we wish to introduce a metric on $\mathbb{D}(S)$ such that $\mathbb{D}(S)$ is a Polish space, and the Borel σ-algebra $\mathscr{B}(\mathbb{D}(S))$ coincides with the natural σ-algebra generated by the canonical coordinate process on $\mathbb{D}(S)$. The idea is to introduce a variation of modules of continuity, which is compatible with the c*à*dl*à*g path space $D_S[0,\infty)$. This modules of continuity will be denoted by $\omega_D(w,\delta,T)$, which should be similar to $\omega(w,\delta,T)$ (where $\delta>0$ and $T>0$), but characterizes the c*à*dl*à*g paths rather than continuous one. The simple way is to write down its definition:

$$\omega_D(w,\delta,T)=\inf_{\{0=t_0<t_1<\cdots<t_k=T\}\in\mathscr{D}_\delta[0,T]}\max_{1\leq j\leq k}\sup_{s,t\in[t_{j-1},t_j)}d(w(t),w(s)). \tag{7.3}$$

where $\mathscr{D}_\delta[0,T]$ is the collection of all finite partitions of $[0,T]$

$$0=t_0<t_1<\cdots<t_k=T.$$

such that $|t_j-t_{j-1}|>\delta$ for $j=1,\cdots,k-1$. Note that the length of the last interval $[t_{k-1},T]$ is not controlled.

Lemma 7.2. *Let $\delta>0$, $T>0$ and $w:[0,T]\to S$. Then*

$$\omega_D(w,\delta,T)\leq\omega(w,2\delta,T). \tag{7.4}$$

Proof. This is not surprising though. Consider a finite partition $0=t_0<\cdots<t_k=T$ such that $|t_j-t_{j-1}|>\delta$ for $j=1,\cdots,k-1$. If $|t_i-t_{i-1}|>2\delta$ for some $i=1,\cdots,k-1$, we can make the partition finer to be $0 = t'_0 < \cdots < t'_{k'} = \zeta$ so that $|t'_i - t'_{i-1}| \leq 2\delta$ for all $i = 1, \cdots, k'-1$ but still retain the condition that $|t'_j - t'_{j-1}| > \delta$ for $j = 1, \cdots, k'-1$. Hence

$$\begin{aligned} \omega_D(w, \delta, T) &\leq \max_{1\leq j\leq k} \sup_{s,t\in[t_j,t_{j-1})} d(w(t), w(s)) \\ &\leq \max_{1\leq j\leq k'} \sup_{s,t\in[t'_j,t'_{j-1})} d(w(t), w(s)) \\ &\leq \omega(w, 2\delta, T). \end{aligned}$$

□

The following theorem provides a characterization of paths in S which are right continuous with left-hand limits.

Lemma 7.3. *Let (S,d) be a metric space, and $w : [0,\infty) \to S$ be a path in S.*
1) If $w \in \mathbb{D}(S)$, then $\lim_{\delta\downarrow 0} \omega_D(w, \delta, T) = 0$ for any $T > 0$.
2) If in addition (S,d) is complete and $\lim_{\delta\downarrow 0} \omega_D(w, \delta, T) = 0$ for any $T > 0$, then $w \in \mathbb{D}(S)$.
3) If $w \in \mathbb{D}(S)$, then $\delta \to \omega_D(w, \delta, T)$ is right continuous at every $\delta > 0$ for every $T > 0$.

Proof. 1). Suppose w is a c*à*dl*à*g path, and $T > 0$. For $\varepsilon > 0$, let τ be the supermum of $t \in [0,T]$ such that there is a finite partition $0 = t_0 < \cdots < t_k = t$, such that

$$\max_{1\leq j\leq k} \sup_{s_1,s_2\in[t_{j-1},t_j)} d(w(s_1), w(s_2)) < \varepsilon.$$

Then $\tau > 0$ as w is right continuous at 0. Since $w(\tau-)$ exists, so that there is $\tau' \in [0,\tau)$ such that

$$\sup_{s_1,s_2\in[\tau',\tau)} d(w(s_1), w(s_2)) < \varepsilon.$$

By definition, as $\tau' < \tau$, there is a finite partition $0 = t'_0 < \cdots < t'_{k'} = \tau'$ such that

$$\max_{1\leq j\leq k'} \sup_{s_1,s_2\in[t'_{j-1},t'_j)} d(w(s_1), w(s_2)) < \varepsilon.$$

Together

$$0 = t'_0 < \cdots < t'_{k'} = \tau' < \tau$$

a partition of $[0,\tau]$ denoted by $0 = t_0 < \cdots < t_k = \tau$ and

$$\max_{1\leq j\leq k} \sup_{s_1,s_2\in[t_{j-1},t_j)} d(w(s_1), w(s_2)) < \varepsilon.$$

If $\tau < T$, then w is right continuous at τ, there is $\tau'' \in [\tau, T]$ such that

$$\sup_{s_1,s_2\in[\tau,\tau'')} d(w(s_1), w(s_2)) < \varepsilon,$$

which allows to produce a partition of $[0,\tau'']$:

$$0 = t_0 < \cdots < t_k(=\tau) < t_{k+1} = \tau''$$

such that

$$\max_{1\leq j\leq k+1}\sup_{s_1,s_2\in[t_{j-1},t_j)} d(w(s_1),w(s_2))<\varepsilon$$

which is a contradiction to the definition of τ. Hence $\tau=T$ which proves the necessity of 1).

Proof of 2). (S,d) is complete. Let $T>0$ and $\lim_{\delta\downarrow 0}\omega_D(w,\delta,T)=0$. Then, for every $\varepsilon>0$ there is a $\delta_0>0$ such that for any $0<\delta<\delta_0$, there is a finite partition of $[0,T]$: $0=t_0<\cdots<t_k=T$ such that $|t_j-t_{j-1}|>\delta$ for $j=1,\cdots,k-1$, and

$$\sup_{s_1,s_2\in[t_{j-1},t_j)} d(w(s_1),w(s_2))<\varepsilon$$

for all $j=1,\cdots,k$, which implies that w is right continuous on $[0,T)$.

Suppose by contradiction that there is $r\in(0,T]$, such that $w(r-)$ did not exist. Then, there is $\varepsilon_0>0$ and there exist $t_n,s_n\uparrow r$ such that $d(w(t_n),w(s_n))\geq\varepsilon_0$, this contradicts to the previous inequality for $\varepsilon=\varepsilon_0$.

The proof of 3) is left as an exercise. □

Example. Step functions are important examples of c*à*dl*à*g paths. For example any sequence in S (such as random sequences in discrete-time) can be considered as continuous paths if S is also a vector space, but in general a sequence can be considered as a c*à*dl*à*g path. A step function w with finitely many jumps, is a path $w:[0,\infty)\to S$ such that there is a partition $0=t_0<\ldots<t_k$ (for some k) such that $w(t)=w(t_{i-1})$ for $t\in[t_{i-1},t_i)$ for $i=1,\ldots,k$ and $w(t)=w(t_k)$ for all $t\geq t_k$. A step function may have infinitely many jumps, but the followings are typical ones. There is a partition $0=t_0<t_1<\ldots<t_j<\ldots$, where $t_j\to\infty$, and $w(t)=w(t_{j-1})$ for $t\in[t_{j-1},t_j)$ for $j=1,2,\ldots$.

Example. Let $w\in\mathbb{D}(S)$. By Lemma 7.3, for every $n=1,2,\ldots$, there is a $\delta_n>0$ such that $\omega_D(w,\delta_n,n)<\frac{1}{n}$. Hence there is a partition

$$0=t_0^{(n)}<t_1^{(n)}<\cdots<t_{k(n)}^{(n)}=n$$

such that $t_j^{(n)}-t_{j-1}^{(n)}>\delta_n$ for $j\leq k(n)-1$ and

$$\sup_{j\leq k(n)}\sup_{s,t\in[t_{j-1}^{(n)},t_j^{(n)})} d(w(t),w(s))<\frac{1}{n}.$$

Define for every $n=1,2,\ldots$ a step function $w^{(n)}$ by setting $w^{(n)}(t)=w(t_{j-1}^{(n)})$ for $t\in[t_{j-1}^{(n)},t_j^{(n)})$ and $w^{(n)}(t)=w(n)$ for $t\geq n$. Then $w^{(n)}$ is a step function with finitely many jumps, $w^{(n)}\in\mathbb{D}(S)$ and

$$\sup_{t\leq n-1} d(w(t),w^{(n)}(t))<\frac{1}{n}$$

for every n, that is $w^{(n)}\to w$ uniformly on $[0,T]$ for every $T>0$. Therefore any c*à*dl*à*g path in (S,d) is a uniform limits of step c*à*dl*à*g paths in (S,d) on any $[0,T]$.

Next we introduce the Skorohod topology on $\mathbb{D}(S)$, where (S,d) is a metric space. We may assume that the metric d is bounded by one, otherwise we may replace its (topologically) equivalent metric

$d\wedge 1$ or $\frac{d}{1+d}$ instead. Still there are slightly different ways to introduce Skorohod's topology on $\mathbb{D}(S)$, in this course we mainly follow the approach in Ethier and Kurtz [14, Chapter 3]. The easiest manner to define the Skorohod topology is to construct a metric giving rise to the required topology. Let us define such a metric on $\mathbb{D}(S)$, called the Skorohod metric.

Let Λ_0 denote the collection of *strictly increasing* function λ from $[0,\infty)$ one-to-one and onto $[0,\infty)$. Such λ must be continuous, $\lambda(0)=0$ and $\lim_{t\uparrow\infty}\lambda(t)=\infty$. The inverse $\lambda^{-1}\in\Lambda_0$ too. If $\lambda\in\Lambda_0$ then

$$\gamma(\lambda)\equiv\sup_{t>s\geq 0}\left|\log\frac{\lambda(t)-\lambda(s)}{t-s}\right|$$

which measures how close between the reparameterization $t\to\lambda(t)$ and t. If $\lambda\in\Lambda_0$ and $\gamma(\lambda)<\infty$, then we say $\lambda\in\Lambda$. Clearly $\gamma(\lambda)=\gamma(\lambda^{-1})$.

Exercise 7.4. *1) If $\lambda_i\in\Lambda$ $(i=1,2)$, then $\lambda_1\circ\lambda_2\in\Lambda$ and*

$$\gamma(\lambda_1\circ\lambda_2)\leq\gamma(\lambda_1)+\gamma(\lambda_2). \tag{7.5}$$

2) If $\lambda\in\Lambda$ then

$$\sup_{t\leq T}|\lambda(t)-t|\leq T\left(e^{\gamma(\lambda)}-1\right),\quad\forall T\geq 0. \tag{7.6}$$

3) If $\lambda_i\in\Lambda$ $(i=1,2)$, then

$$\begin{aligned}\sup_{t\leq T}\lambda_1\left(|\lambda_2(t)-t|\right) &\leq \sup_{s\geq 0}\frac{\lambda_1(s)}{s}\sup_{t\leq T}|\lambda_2(t)-t| \\ &\leq e^{\gamma(\lambda_1)}\sup_{t\leq T}|\lambda_2(t)-t| \qquad \forall T\geq 0.\end{aligned} \tag{7.7}$$

To simplify our notations, let us introduce the following convention. If β is a real or complex valued function on $[0,\infty)$, and $T>0$, then

$$\|\beta\|_T=\sup_{t\leq T}|\beta(t)|\quad\text{ and }\ \|\beta\|=\|\beta\|_\infty$$

the supremum norm of β, which may be infinity.

If $x,y:[0,\infty)\to S$ are two paths, then

$$\|d(x,y)\|_T=\sup_{t\leq T}d(x(t),y(t))$$

and

$$\|d(x,y)\|=\sup_{t\geq 0}d(x(t),y(t))=\|d(x,y)\|_\infty.$$

For $s\geq 0$ we use x^s to denote the path stopped at s, that is, $x^s(t)=x(t\wedge s)$. If $\lambda\in\Lambda_0$, then $x\circ\lambda$ is the reparameterisation of x, thus, $x\circ\lambda(t)=x(\lambda(t))$. Note the difference that $x^s\circ\lambda(t)=x(s\wedge\lambda(t))$ and $(x\circ\lambda)^s(t)=x(\lambda(t\wedge s))$ for $t\geq 0$. The subtle difference is important if the x has jumps.

Lemma 7.5. *If $x\in\mathbb{D}(S)$, $\lambda\in\Lambda$ and $s\geq 0$, then x^s and $x\circ\lambda$ also belong to $\mathbb{D}(S)$.*

Definition 7.6. *Let (S,d) be a metric space with a bounded metric d. If $x,y \in \mathbb{D}(S)$, then define*

$$\rho(x,y) = \inf_{\lambda\in\Lambda}\left[\gamma(\lambda) \vee \int_0^\infty \|d(x^s, y^s\circ\lambda)\| e^{-s} ds\right]. \tag{7.8}$$

In other words

$$\rho(x,y) = \inf_{\lambda\in\Lambda}\left[\sup_{t>s\geq 0}\left|\log\frac{\lambda(t)-\lambda(s)}{t-s}\right| \vee \int_0^\infty e^{-s}\sup_{t\geq 0} d(x(t\wedge s), y(\lambda(t)\wedge s))ds\right]. \tag{7.9}$$

Lemma 7.7. *Let (S,d) be a metric space with bonded metric d. Then ρ defined by (7.8) is a metric on $\mathbb{D}(S)$, called the Skorohod metric.*

See [14, pages 117,118] for the proof. The induced topology on $\mathbb{D}(S)$ by the metric ρ is called the *Skorohod topology*.

Remark. We can avoid the L^1-norm in the definition (7.8) if (S,d) is a normed vector space. For each $n = 1,2,\ldots$ we define a Lipschitz continuous cut-off function $k_n(t) = 1$ for $t \leq n$, $k_n(t) = 1-(t-n)$ for $t \in (n,n+1]$ and $k_n(t) = 0$ for $t > n+1$. If $x \in \mathbb{D}(\mathbb{R}^d)$, $k_n x(t) = k_n(t)x(t)$ defines a new path which coincides with $x(t)$ for $t \in [0,n]$ and vanishes on $[n+1,\infty)$. A (topologically) equivalent Skorohod metric is defined by

$$\tilde{\rho}(x,y) = \inf_{\lambda\in\Lambda}\left\{\gamma(\lambda) + \sum_{n=1}^\infty \frac{1}{2^n}(\|(k_n x)\circ\lambda - k_n y\| \wedge 1)\right\} \tag{7.10}$$

which defines the Skorohod topology on $\mathbb{D}(\mathbb{R}^d)$.

Let us discuss the convergence of sequences in $\mathbb{D}(S)$ with respect to the Skorohod topology through several lemmas, their proofs are omitted, one may refer to [14, Chapter III].

Lemma 7.8. *Let (S,d) be a metric space with a bounded metric d, and ρ the Skorohod metric on $\mathbb{D}(S)$. Suppose $x^{(n)}, x \in \mathbb{D}(S)$ (where $n = 1,2,\ldots$). Then $\rho(x^{(n)},x) \to 0$, if and only if there is a sequence $\lambda^{(n)} \in \Lambda$ such that*

$$\gamma\left(\lambda^{(n)}\right) \to 0 \tag{7.11}$$

and

$$\sup_{t\in[0,T]} d\left(x^{(n)}(t), x(\lambda^{(n)}(t))\right) \to 0 \tag{7.12}$$

as $n \to \infty$ for every $T > 0$.

Lemma 7.9. *Let (S,d) be a metric space with a bounded metric d, and ρ the Skorohod metric on $\mathbb{D}(S)$. Let $x^{(n)}$, $x \in \mathbb{D}(S)$, where $n = 1,2,\ldots$. Then $\rho(x^{(n)},x) \to 0$ as $n \to \infty$, if and only if for every $T > 0$, there is a sequence $\lambda^{(n,T)} \in \Lambda_0$ such that*

$$\sup_{t\in[0,T]}\left|\lambda^{(n,T)}(t) - t\right| \to 0$$

as $n \to \infty$, and

$$\sup_{t\in[0,T]} d\left(x^{(n)}(t), x(\lambda^{(n,T)}(t))\right) \to 0, \qquad \text{as } n \to \infty.$$

Lemma 7.10. *Let (S,d) be a metric space with a bounded metric d, and ρ the Skorohod metric on $\mathbb{D}(S)$. Let $x^{(n)}, x \in \mathbb{D}(S)$ (for $n \geq 1$). Then $\rho(x^{(n)}, x) \to 0$ as $n \to \infty$, if and only if there is a sequence $\lambda^{(n)} \in \Lambda$, such that $\gamma\left(\lambda^{(n)}\right) \to 0$, and*

$$\sup_{t \geq 0} d(x^{(n)}(t \wedge s), x(\lambda^n(t) \wedge s)) \to 0$$

as $n \to \infty$ for $s > 0$ at which x is continuous. In particular, if x is continuous at $s > 0$, then $x^{(n)}(s) \to x(s)$ and also $x^{(n)}(-s) \to x(s)$ as $n \to \infty$.

The following theorem allows fully implement Prohorov's theorems for weak convergences of distributions on $\mathbb{D}(S)$ endowed with the Skorohod topology, and its induced Borel σ-algebra $\mathscr{B}(\mathbb{D}(S))$.

Theorem 7.11. *Let S be a metric metric with a bounded metric d, and ρ be the Skorohod metric on $\mathbb{D}(S)$.*

1) If (S,d) is separable, then $(\mathbb{D}(S), \rho)$ is separable too.

2) If (S,d) is a Polish space, then $(\mathbb{D}(S), \rho)$ is separable and complete too.

We will not give the proof of the previous theorem in the lectures, see [14, Chapter III] for details.

Let $(X_t)_{t \geq 0}$ be the coordinate process on $\mathbb{D}(S)$ where for every $t \geq 0$, $X_t(w) = w(t)$ for every $w \in \mathbb{D}(S)$.

Theorem 7.12. *Let (S,d) be a Polish metric space. Then the Borel σ-algebra generated by the Skorohod topology, $\mathscr{B}(\mathbb{D}(S)) = \sigma\{X_t : t \geq 0\}$.*

Next we need to identify the relatively compact subsets of $\mathbb{D}(S)$ equipped with the Skorohod topology. To this end we introduce the following notations. If $w \in \mathbb{D}(S)$ is a step function, then there is a partition $0 = t_0 < t_1 < \cdots < t_j < \cdots$, such that

$$w(t) = w(t_{j-1}) \quad \text{if } t \in [t_{j-1}, t_j)$$

for $j = 1, 2, \cdots,$ where $t_0 = 0$ and if $t_{j-1} < \infty$ then

$$\begin{aligned} t_j &= \inf\{t > t_{j-1} : w(t) \neq w(t-)\} \\ &= \inf\{t > t_{j-1} : w(t) \neq w(t_{j-1})\}. \end{aligned}$$

Therefore $t_j = \infty$ for some j or $t_j \uparrow \infty$. To indicate the dependence on the step function $w \in \mathbb{D}(S)$, we may write t_j as $t_j(w)$.

The following give some examples of compact subsets of $\mathbb{D}(S)$.

Lemma 7.13. *Let (S,d) be a Polish space. Let $\delta > 0$ and $\Gamma \subset S$ be a compact subset. $\mathscr{A}(\Gamma, \delta)$ denotes the collection of all step functions $w \in \mathbb{D}(S)$ such that*

1) $w(t) \in \Gamma$ for all t (that is, $w \in \mathbb{D}(\Gamma)$ where (Γ, d) is a compact metric space), and

2) $t_j(w) - t_{j-1}(w) \geq \delta$ for j such that $t_{j-1}(w) < \infty$.

Then $\mathscr{A}(\Gamma, \delta)$ is compact in $\mathbb{D}(S)$ under the Skorohod topology.

Proof. Let $\left\{w^{(n)}\right\}$ be a sequence in $\mathscr{A}(\Gamma, \delta)$. If necessary, by applying Cantor's digonalization to $t_j(w^{(n)})$, we may assume that, for any $j = 0, 1, \cdots$, either 1) $t_j(w^{(n)}) = \infty$ for all n, or 2) $t_j(w^{(n)}) < \infty$, $t_j(w^{(n)}) \to t_j$ and $w^{(n)}_{t_j(w^{(n)})} \to w_{t_j}$ for some elements $w_{t_j} \in \Gamma$. Define $w \in \mathbb{D}(\Gamma)$ by $w_t = w_{t_j}$ for $t \in [t_j, t_{j+1})$ if $t_j < \infty$. Then, since $t_j(w^{(n)}) - t_{j-1}(w^{(n)}) \geq \delta$, we can verify that $w^{(n)} \to w$ in the Skorohod metric D. Clearly, $t_j - t_{j-1} \geq \delta$, so that $t_j(w) - t_{j-1}(w) \geq \delta$, and therefore $w \in \mathscr{A}(\Gamma, \delta)$. □

The following theorem is the Ascoli-Arzel*à* theorem for c*à*dl*à*gcadlag paths.

Theorem 7.14. *Let (S,d) be a Polish metric space. A subset $K \subset \mathbb{D}(S)$ is relatively compact with respect to the Skorohod topology, if and only if the following two conditions are satisfied:*

1) There is a dense set $\mathscr{Q} \subset [0,\infty)$ such that $\{w_t : w \in K\}$ is relatively compact in S for any $t \in \mathscr{Q}$.

2) For every $T > 0$

$$\lim_{\delta \downarrow 0} \sup_{w \in K} \omega_D(w, \delta, T) = 0.$$

For a proof of this theorem read [14, Chapter III].

By Prohorov's criterion for relative compactness of probability measures on Polish spaces, we then have the following tightness criterion for distributions on $\mathbb{D}(\mathbb{R}^d)$.

Theorem 7.15. *Let $\mathbb{D}(\mathbb{R}^d)$ be equipped with the Skorohod metric. Let $\mathscr{L}$ be a family of probability measures on $(\mathbb{D}(\mathbb{R}^d), \mathscr{B}(\mathbb{D}(\mathbb{R}^d)))$. Then $\mathscr{L}$ is relatively compact with respect to weak convergence topology (i.e. equivalently with respect to the Prohorov's metric topology over the space of probability measures on $\mathbb{D}(\mathbb{R}^d)$), if and only if the following two conditions are satisfied:*

$$\lim_{L \to \infty} \sup_{P \in \mathscr{L}} P\left[w : \sup_{t \in [0,T]} |w(t)| \geq L \right] = 0$$

and

$$\lim_{\delta \downarrow 0} \sup_{P \in \mathscr{L}} P[w : \omega_D(w, \delta, T) \geq \varepsilon] = 0$$

for every $T > 0$ and $\varepsilon > 0$.

Theorem 7.16. *Let $X^{(n)}$ (where $n = 1, 2, \ldots$) be a sequence of stochastic processes with values in $\mathbb{R}^d$ on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, whose sample paths are right continuous on $[0,\infty)$ with left limits on $(0,\infty)$, so that their distributions are probability measures on $(\mathbb{D}(\mathbb{R}^d), \mathscr{B}(\mathbb{D}(\mathbb{R}^d)))$. Then $(X^{(n)})$ is tight if and only if*

$$\lim_{L \to \infty} \sup_{n} \mathbb{P}\left[\sup_{t \in [0,T]} |X_t^{(n)}| \geq L \right] = 0$$

and

$$\lim_{\delta \downarrow 0} \sup_{n} \mathbb{P}\left[\omega_D(X^{(n)}, \delta, T) \geq \varepsilon \right] = 0$$

for every $T > 0$ and $\varepsilon > 0$.

The modified modules of continuity, $\omega_D(x, \delta, T)$ is not easy to estimate, so any simpler replacements will have value. Here we introduce another version of modules of continuity. If $x : [0,\infty) \to \mathbb{R}^d$, then

$$\omega_d(x, \delta, T) = \sup_{0 \leq s < t < r \leq T; r - s < \delta} \left[|x(t) - x(s)| \wedge |x(r) - x(t)| \right].$$

Theorem 7.17. *Let $\mathbb{D}(\mathbb{R}^d)$ be equipped with the Skorohod metric. Then $K \subset \mathbb{D}(\mathbb{R}^d)$ is relatively compact if and only if*

$$\sup_{w \in K} \sup_{t \leq T} |w(t)| < \infty,$$

$$\lim_{\delta \downarrow 0} \sup_{x \in K} \omega_d(w, \delta, T) = 0$$

for every $T > 0$, and

$$\lim_{\delta \downarrow 0} \sup_{w \in K} \sup_{s,t \in [0,\delta)} |w(t) - w(s)| = 0.$$

8 Cramér's theorem

The main reference is Stroock [24], and we will follow the exposition there closely.

Let (ξ_n) be an i.i.d. sequence on $(\Omega, \mathscr{F}, P)$ with a common distribution μ on R^1. For a Borel function g on R^1 such that $g(\xi_n)$ is integrable,

$$\mu(g) = \int_\Omega g(\xi_n) dP.$$

Assume that ξ_i are integrable, and set

$$a = E\xi_n = \int_{R^1} x\mu(dx).$$

Let μ_n denote the distribution of $\frac{1}{n}\sum_{i=1}^n \xi_i$. Then, according to the law of large numbers

$$\frac{1}{n}\sum_{i=1}^n \xi_i \to a \quad \text{almost surely}$$

so that $\mu_n \to \delta_a$ weakly. The probabilities of the deviations of μ_n away from δ_a is the context of Cramér's theorem.

Theorem 8.1. *(H. Cramér) Suppose for every $\lambda \in R^1$, $\int_{R^1} e^{\lambda z}\mu(dz) < \infty$, then $\{\mu_n : n \geq 1\}$ satisfies the large deviation principle with a good rate function*

$$I(x) = \sup\left\{\lambda x - \log \int_{R^1} e^{\lambda z}\mu(dz)\right\} .$$

That is, for every closed set F in R^1

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(F) \leq -\inf_F I \tag{8.1}$$

and for every open set G in R^1

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(G) \geq -\inf_G I . \tag{8.2}$$

We divide the proof of this theorem into several steps.

Lemma 8.2. *If μ is a probability measure on R^d such that*

$$M(\xi) = \int_{R^d} e^{\langle \xi, x\rangle}\mu(dx) < \infty$$

for every $\xi \in R^d$. Then $M : R^d \to (0,\infty)$ is smooth, log-convex (i.e. $\xi \to \log M(\xi)$ is convex) function. Define

$$I(x) = \sup_{\xi\in R^d} \{\langle \xi, x\rangle - \log M(\xi)\} \ ; \quad x \in R^d$$

(the Legendre transform of $\log M$*). Then I is a convex* good rate function*, in the sense that $\{x : I(x) \leq c\}$ is compact for every $c \in \mathbb{R}$. I attains its minimum zero at $a = \int_{R^d} x\mu(dx)$. In particular, if $d = 1$, then $I \uparrow$ on (a,∞) and $I \downarrow$ on $(-\infty, a)$.*

By definition, M takes values in $(0,\infty)$. A simple application of Lebesgue's dominated convergence, we may differentiate M under the integral sign

$$D^\alpha M(\xi) = \int_{R^d} \xi^\alpha e^{\langle \xi, x\rangle} \mu(dx)$$

which implies the smoothness of M. Log-convexity of M is a simple consequence of the Hölder inequality:

$$\begin{aligned}
& \int_{R^d} e^{\langle \lambda\xi_1 + (1-\lambda)\xi_2, x\rangle} \mu(dx) \\
\leq \ & \left(\int_{R^d} e^{\langle \xi_1, x\rangle}\mu(dx)\right)^\lambda \left(\int_{R^d} e^{\langle \xi_2, x\rangle}\mu(dx)\right)^{1-\lambda}
\end{aligned}$$

so that

$$\log M(\lambda\xi_1 + (1-\lambda)\xi_2) \leq \lambda \log M(\xi_1) + (1-\lambda)\log M(\xi_2) \ .$$

I is the Legendre transform of the continuous convex function $\log M$, so that I is lower semi-continuous, non-negative and convex. Finally we need to show that for every $c > 0$

$$K_c = \left\{ x \in R^d : I(x) \leq c \right\}$$

is compact. Firstly since I is lower semi-continuous so K_c is closed , thus we only need to show that K_c is bounded. Let $e_i = (0, \cdots, 1, \cdots 0)$ (1 appears in the i-th coordinate). If $x \in K_c$ then for every i

$$\pm x_i - \log M(\pm e_i) \leq c$$

which implies that every x_i is bounded, hence K_c is bounded.

By Jensen's inequality

$$\begin{aligned}
\log M(\xi) &= \log \int_{R^d} e^{\langle \xi, x\rangle} \mu(dx) \\
&\geq \int_{R^d} \langle \xi, x\rangle \mu(dx)
\end{aligned}$$

so that if a is the mean of $\mu(dx)$: $a_i = \int_{R^d} x_i \mu(dx)$, then for every ξ

$$\log M(\xi) \geq \xi_i \int_{R^d} x_i \mu(dx) = \langle a, \xi\rangle.$$

Hence

$$\langle a, \xi\rangle - \log M(\xi) \leq 0 \quad \text{for all } \xi$$

and therefore we must have $I(a) = 0$.

Now we consider the one-dimensional case, so that I is a convex, good rate function on R^1. By Jensen's inequality

$$\log \int_{R^1} e^{\lambda z} \mu(dz) \geq \lambda \int_{R^1} z\mu(dz) = a\lambda$$

so that

$$x\lambda - \log \int_{R^1} e^{\lambda z}\mu(dz) \leq (x-a)\lambda \ . \tag{8.3}$$

Since $I(a)=0$, and I is convex, I increases on $[a,\infty)$ and decreases on $(-\infty,a]$, therefore

$$\inf_{(x,y]} I = I(y) \qquad \text{if } x<y\le a$$

and

$$\inf_{[x,y)} I = I(x) \qquad \text{if } a\le x<y\,.$$

Moreover, from (8.3), if $\lambda\le 0$ and $x\ge a$ then

$$\lambda x - \log\int_{R^1} e^{\lambda z}\mu(dz)\le 0$$

so that

$$I(x)=\sup_{\lambda\ge 0}\left\{\lambda x-\log\int_{R^1} e^{\lambda z}\mu(dz)\right\} \quad \text{for} \quad x\ge a. \tag{8.4}$$

Similarly

$$I(x)=\sup_{\lambda\le 0}\left\{\lambda x-\log\int_{R^1} e^{\lambda z}\mu(dz)\right\} \quad \text{for} \quad x\le a\,. \tag{8.5}$$

Lemma 8.3. *We have*

$$\mu\left([x,\infty)\right)\le \exp\left(-I(x)\right)=\exp\left(-\inf_{[x,\infty)} I\right) \quad \textit{for } x\ge a$$

and

$$\mu\left((-\infty,x]\right)\le \exp\left(-I(x)\right)\le \exp\left(-\inf_{(-\infty,x]} I\right) \quad \textit{for } x\le a\,.$$

Proof. If $\lambda\ge 0$ and $x\ge a$

$$\begin{aligned}
\mu\left([x,\infty)\right) &= \int_{z\ge x}\mu(dz)\\
&\le \int_{z\ge x}\frac{e^{\lambda z}}{e^{\lambda x}}\mu(dz)\\
&\le \int_{R^1}\frac{e^{\lambda z}}{e^{\lambda x}}\mu(dz)\\
&= \exp\left\{-\left(\lambda x-\log\int_{R^1}e^{\lambda z}\mu(dz)\right)\right\}
\end{aligned}$$

so that

$$\begin{aligned}
\mu\left([x,\infty)\right) &= \exp\left\{-\sup_{\lambda\ge 0}\left(\lambda x-\log\int_{R^1}e^{\lambda z}\mu(dz)\right)\right\}\\
&= \exp\{-I(x)\}\,.
\end{aligned}$$

Similarly we may prove the case that $x\le a$. □

Proof of upper bound (8.1). If $F=\emptyset$ or $a\in F$ then $\inf I=0$ so that $\inf_F I=0$ the bound is trivial in this case. Therefore we assume that $a\notin F$. If $F\subset[a,\infty)$, and if $y=\inf\{z:z\in F\}$ then $F\subset[y,\infty)$ and

$$\inf_F I=I(y)=\sup_{\lambda\geq 0}\left\{\lambda y-\log\int_{R^1}e^{\lambda z}\mu(dz)\right\}. \tag{8.6}$$

However for every $\lambda>0$

$$\begin{aligned}
\mu_n(F) &\leq \mu_n([y,\infty))=P\left\{\frac{1}{n}\sum_{i=1}^n\xi_i\geq y\right\}\\
&\leq \int_{\left\{\frac{1}{n}\sum_{i=1}^n\xi_i\geq y\right\}}\frac{e^{\frac{1}{n}\lambda\sum_{i=1}^n\xi_i}}{e^{\lambda y}}dP\\
&\leq \int_\Omega\frac{e^{\frac{1}{n}\sum_{i=1}^n\lambda\xi_i}}{e^{\lambda y}}dP=\int_\Omega\frac{\prod_{i=1}^n e^{\frac{\lambda}{n}\xi_i}}{e^{\lambda y}}dP\\
&= e^{-\lambda y}\prod_{i=1}^n\int_\Omega e^{\frac{\lambda}{n}\xi_i}dP \qquad (\xi_i\text{ are i.i.d.})\\
&= e^{-\lambda y}\left(\int_{R^1}e^{\frac{\lambda}{n}z}\mu(dz)\right)^n
\end{aligned}$$

and therefore

$$\frac{1}{n}\log\mu_n(F)\leq-\left\{\frac{\lambda}{n}y-\int_{R^1}e^{\frac{\lambda}{n}z}\mu(dz)\right\}$$

for every $\lambda\geq 0$. It thus follows that

$$\begin{aligned}
\frac{1}{n}\log\mu_n(F) &\leq -\sup_{\lambda\geq 0}\left\{\lambda y-\int_{R^1}e^{\lambda z}\mu(dz)\right\}\\
&= -I(y)=-\inf_F I=-I(\min(F)).
\end{aligned}$$

We thus have proven the upper bound for the case that $F\subset[a,\infty)$.

Similarly

$$\frac{1}{n}\log\mu_n(F)\leq-\inf_F I=-I(\max(F)) \quad\text{if }\ F\subset(-\infty,a]\,.$$

Finally for an arbitrary closed set F in R^1, let $F_1=F\cap(-\infty,a]$ and $F_2=F\cap[a,\infty)$. Then

$$\frac{1}{n}\log\mu_n(F)\leq\frac{1}{n}\log(\mu_n(F_1)+\mu_n(F_2))$$

so that

$$\begin{aligned}
\limsup_{n\to\infty}\frac{1}{n}\log\mu_n(F) &\leq \max\left\{\limsup_{n\to\infty}\frac{1}{n}\log\mu_n(F_1),\ \limsup_{n\to\infty}\frac{1}{n}\log\mu_n(F_2)\right\}\\
&\leq \max\{-I(\max F_1);\ -I(\min F_2)\}\\
&= -\min\{I(\max F_1);\ I(\min F_2)\}\\
&\leq -\inf_F I
\end{aligned}$$

which is the upper bound for large deviations.

Proof of lower bound (8.2). We are going to show that for every $x \in G$

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(G) \geq -I(x) \; .$$

Obviously we only need to prove the previous inequality for those x such that $I(x) < \infty$. Need to handle two cases. Firstly let us consider the case that the supremum (which equals $I(x)$) of

$$\lambda y - \int_{R^1} e^{\lambda z} \mu(dz)$$

is not achievable. Then $x \neq a$ (as $I(a) = 0$ which is achieved when $\lambda = 0$). Without loss of generality, let us assume that $x > a$. Then we may choose a sequence of $\lambda_n > 0$ such that $\lambda_n \to +\infty$ and

$$\lambda_n x - \int_{R^1} e^{\lambda_n z} \mu(dz) \to I(x) \quad \text{as } n \to \infty \; .$$

While by Lebesgue's dominated convergence theorem

$$\lim_{n\to\infty} \int_{(-\infty,x)} e^{\lambda_n (z-x)} \mu(dz) = 0$$

and therefore

$$\begin{aligned} \lim_{n\to\infty} \int_{[x,\infty)} e^{\lambda_n (z-x)} \mu(dz) &= \lim_{n\to\infty} \int_{R^1} e^{\lambda_n (z-x)} \mu(dz) \\ &= \lim_{n\to\infty} e^{-\left\{\lambda_n x - \log \int_{R^1} \exp(\lambda_n z) \mu(dz)\right\}} \\ &= \exp\left(-I(x)\right) < +\infty. \end{aligned} \tag{8.7}$$

On the other hand, for any $\delta > 0$ we have

$$\int_{[x+\delta,\infty)} e^{\lambda_n (z-x)} \mu(dz) \geq e^{\delta \lambda_n} \mu([x+\delta,\infty))$$

so that

$$\begin{aligned} \mu([x+\delta,\infty)) &\leq e^{-\delta\lambda_n} \int_{[x+\delta,+\infty)} e^{\lambda_n (z-x)} \mu(dz) \\ &\leq e^{-\delta\lambda_n} \int_{R^1} e^{\lambda_n (z-x)} \mu(dz) \\ &\leq e^{-\delta\lambda_n} e^{-\left\{\lambda_n x - \log \int_{R^1} e^{\lambda_n z} \mu(dz)\right\}} \; . \end{aligned}$$

Letting $n \to \infty$ we conclude that

$$\begin{aligned} \mu([x+\delta,\infty)) &\leq e^{-\lim_{n\to\infty}\left\{\lambda_n x - \log \int_{R^1} e^{\lambda_n z} \mu(dz)\right\}} \lim_{n\to\infty} e^{-\delta\lambda_n} \\ &= 0 \qquad \forall \delta > 0 \; . \end{aligned}$$

Therefore $\mu((x,\infty)) = 0$, hence by (8.7)

$$\lim_{n\to\infty} \int_{[x,\infty)} e^{\lambda_n (z-x)} \mu(dz) = \mu(\{x\}) = \exp\left(-I(x)\right)$$

which follows thus that

$$\begin{aligned} \mu_n(G) &\geq \mu_n(\{x\}) \\ &\geq P\{\xi_i = x : \ i = 1, \cdots, n\} \\ &= \mu(\{x\})^n . \end{aligned}$$

Therefore

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(G) \geq \mu(\{x\}) = e^{-I(x)} .$$

Similarly one may handle the case that $x < a$.

Next we consider the case that $x \in G$ and there is λ such that

$$\begin{aligned} I(x) &= \lambda x - \log \int_{R^1} e^{\lambda z} \mu(dz) \\ &= \sup_{\eta} \left\{ \eta x - \log \int_{R^1} e^{\eta z} \mu(dz) \right\} \end{aligned}$$

where $(x-a)\lambda \geq 0$. Since λ is a critical point to

$$\eta \to \eta x - \log \int_{R^1} e^{\eta z} \mu(dz)$$

so that its derivative at λ vanishes, which yields that

$$x = \frac{\int_{R^1} z e^{\lambda z} \mu(dz)}{\int_{R^1} e^{\lambda z} \mu(dz)} . \tag{8.8}$$

Without losing generality, assume that $x \geq a$ so that $\lambda \geq 0$. Choose $\delta > 0$ such that $(x-\delta, x+\delta) \subset G$. Then

$$\begin{aligned} \mu_n(G) &= \mu_n((x-\delta, x+\delta)) \\ &= P\left\{ \left| \frac{1}{n} \sum_{i=1}^n \xi_i - x \right| < \delta \right\} \\ &\geq E\left\{ \frac{e^{\lambda \sum_{i=1}^n \xi_i}}{e^{n\lambda(x+\delta)}}; \left| \frac{1}{n} \sum_{i=1}^n \xi_i - x \right| < \delta \right\} \\ &= e^{-n\lambda(x+\delta)} E\left\{ e^{\lambda \sum_{i=1}^n \xi_i}; \left| \frac{1}{n} \sum_{i=1}^n \xi_i - x \right| < \delta \right\} \\ &= e^{-n\lambda(x+\delta)} \int_{R^n} e^{\lambda \sum_{i=1}^n z_i} 1_{\left\{ \left| \frac{1}{n} \sum_{i=1}^n z_i - x \right| < \delta \right\}} \mu(dz_1) \cdots \mu(dz_n) . \end{aligned}$$

Define a new probability measure v on R^1 by

$$\frac{d\mu}{d\nu} = \frac{e^{\lambda z}}{\int_{R^1} e^{\lambda y} \mu(dy)} \mu(dz) .$$

Then

$$
\begin{aligned}
\mu_n(G) &\geq e^{-n\lambda(x+\delta)} \left(\int_{R^1} e^{\lambda y} \mu(dy) \right)^n \int_{R^n} 1_{\left\{ \left| \frac{1}{n}\sum_{i=1}^n z_i - x \right| < \delta \right\}} \nu(dz_1) \cdots \nu(dz_n) \\
&= e^{-n\lambda(x+\delta)} \left(\int_{R^1} e^{\lambda y} \mu(dy) \right)^n P\left\{ \left| \frac{1}{n}\sum_{i=1}^n Y_i - x \right| < \delta \right\}
\end{aligned}
$$

where Y_i are i.i.d with mean (see equation (8.8))

$$
\begin{aligned}
EY_i &= \int_{R^1} z_i \nu(dz_i) = \int_{R^1} \frac{z_i e^{\lambda z_i}}{\int_{R^1} e^{\lambda y} \mu(dy)} \mu(dz_i) \\
&= \frac{1}{\int_{R^1} e^{\lambda y} \mu(dy)} \int_{R^1} z_i e^{\lambda z_i} \mu(dz_i) \\
&= x
\end{aligned}
$$

and thus by the law of large numbers

$$
P\left\{ \left| \frac{1}{n}\sum_{i=1}^n Y_i - x \right| < \delta \right\} \to 1 \text{ as } n \to \infty \, .
$$

Hence

$$
\begin{aligned}
\frac{1}{n} \log \mu_n(G) &\geq -\lambda(x+\delta) + \log \int_{R^1} e^{\lambda y} \mu(dy) \\
&\quad + \frac{1}{n} \log P\left\{ \left| \frac{1}{n}\sum_{i=1}^n Y_i - x \right| < \delta \right\} \\
&\to -\lambda(x+\delta) + \log \int_{R^1} e^{\lambda y} \mu(dy) \quad \text{as } n \to \infty \, .
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(G) &\geq -\left(\lambda x - \log \int_{R^1} e^{\lambda y} \mu(dy) \right) - \delta\lambda \\
&= -I(x) - \delta\lambda \qquad \forall \delta > 0.
\end{aligned}
$$

By letting $\delta \downarrow 0$ we obtain

$$
\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(G) \geq -I(x) \quad \text{for every } x \in G.
$$

Thus we have completed the proof of Cramér's theorem.

9 Large deviation principles, and functional integration

The speed of convergence of a sequence of distributions to their limiting distribution may be described by rate functions, which are non-negative, lower semi-continuous functions on the state space E.

Let E be a Polish space, and $\mathscr{B}(E)$ denote the Borel σ-algebra. Recall that a function $I : E \to R^1 \cup \{\infty\}$ is lower semi-continuous if for every $c \in R^1 \cup \{\infty\}$ the level set

$$I_c = I^{-1}((-\infty, c]) = \{s \in E : I(s) \leq c\}$$

is closed. In other words, I is lower semi-continuous if

$$I(s_0) \leq \liminf_{s \to s_0} I(s) .$$

A lower semi-continuous function $I : E \to [0, \infty]$ is called a *rate function*. We will see that however a necessary condition for a rate function to be the speed describing large deviations of distributions only if can I achieve its minimum zero.

A family $\{P_\varepsilon : \varepsilon > 0\}$ of probability measures on a Polish space (E, ρ) is said to satisfy the *large deviation principle* with rate function I if

$$\limsup_{\varepsilon \downarrow 0} \varepsilon \log P_\varepsilon (F) \leq -\inf_{s \in F} I(s) \tag{9.1}$$

for every closed sets F in E (called upper bound of large deviations) and

$$\liminf_{\varepsilon \downarrow 0} \varepsilon \log P_\varepsilon (G) \geq -\inf_{s \in G} I(s) \tag{9.2}$$

for every open subset G in E (that is called the lower bound of large deviations).

In this case we also say that the rate function I governs the large deviations of $\{P_\varepsilon : \varepsilon > 0\}$.

Here the following conventions are applied. Firstly, if A is empty, then $-\inf_A I = -\infty$, and $\log 0 = -\infty$.

Obviously (9.1) and (9.2) together is equivalent to

$$-\inf_{x \in B^o} I(x) \leq \liminf_{\varepsilon \downarrow 0} \varepsilon \log P_\varepsilon (B) \leq \limsup_{\varepsilon \downarrow 0} \varepsilon \log P_\varepsilon (B) \leq -\inf_{x \in \overline{B}} I(x) \tag{9.3}$$

for every Borel subset B in E. Therefore, if B is a Borel set in E such that $\inf_{\overline{B}} I = \inf_{B^o} I$ $(= \inf_B I)$ then

$$\lim_{\varepsilon \downarrow 0} \varepsilon \log P_\varepsilon (B) = -\inf_B I$$

and therefore

$$P_\varepsilon (B) \approx \exp \left\{ -\frac{\inf_B I}{\varepsilon} \right\} \qquad \text{as } \varepsilon \downarrow 0.$$

That is, unless $\inf_B I = 0$, $P_\varepsilon (B)$ tends to zero exponentially as $\varepsilon \downarrow 0$.

We will see that the lower bound of large deviations (9.2) is more or less a local property, but the upper bound (9.1) reflects in many cases the global distribution of the family $\{P_\varepsilon : \varepsilon > 0\}$, and thus is more important, and more difficult to prove. Therefore we may also introduce the so-called *weak* large deviation principle: $\{P_\varepsilon : \varepsilon > 0\}$ satisfies the weak large deviation principle with rate function I if

$$\limsup_{\varepsilon \downarrow 0} \varepsilon \log P_\varepsilon (K) \leq -\inf_{x \in K} I(x) \tag{9.4}$$

for every *compact* subset $K \subset E$ (instead of any closed set) together with the lower bound of large deviation (9.2).

Lemma 9.1. *For a given family $\{P_\varepsilon : \varepsilon > 0\}$ of probabilities on E there is at most one rate function governing the large deviations of $\{P_\varepsilon : \varepsilon > 0\}$.*

Proof. Exercise. □

We have seen that a necessary condition that a rate function I governs large deviations of a family of distributions is that the rate function I must achieve its minimum zero. Good rate functions are introduced to address this issue and other analytic difficulties.

Definition 9.2. (Good rate functions) *Let E be a Polish space. A function $I : E \to [0,\infty]$ is a* good rate function *if*

1) $\inf I < \infty$,

2) I is lower semi-continuous,

3) For any $c \geq 0$, $\{s : I(s) \leq c\}$ is compact.

Of course, condition 3 implies condition 2, and is a very strong restriction on rate functions. It follows thus that I must achieve its minimum, and indeed we can say a little bit more.

Proposition 9.3. *Let $I : E \to [0,\infty]$ be a good rate function and let $\Phi : E \to [-\infty,\infty)$ be upper semi-continuous (i.e. $-\Phi : E \to R^1 \cup \{\infty\}$ is lower semi-continuous). Then for any closed set S on which Φ is bounded above, there is a $s_0 \in S$ such that*

$$\Phi(s_0) - I(s_0) = \sup_S (\Phi - I) \ . \tag{9.5}$$

Proof. Since $\sup_S \Phi < \infty$, so $\sup_S (\Phi - I) > -\infty$, and therefore $\sup_S (\Phi - I)$ is finite. By definition, there is a sequence $\{s_n\}$ in S such that

$$\sup_S (\Phi - I) - \frac{1}{n} \leq \Phi(s_n) - I(s_n) \leq \sup_S (\Phi - I) \ .$$

In particular

$$\begin{aligned} I(s_n) &\leq \Phi(s_n) + \frac{1}{n} - \sup_S (\Phi - I) \\ &\leq \sup_S \Phi + 1 - \sup_S (\Phi - I) \end{aligned}$$

and therefore

$$\{s_n\} \subset \left\{ x : I(x) \leq \sup_S \Phi + 1 - \sup_S (\Phi - I) \right\}$$

which is compact. Hence there is a subsequence of $\{s_n\}$ (and for simplicity denoted again by $\{s_n\}$), such that $s_n \to s_0$. Since S is closed, so $s_0 \in S$. Since $\Phi - I$ is upper semi-continuous,

$$\sup_S (\Phi - I) \geq \Phi(s_0) - I(s_0) \geq \limsup_{n\to\infty} (\Phi(s_n) - I(s_n)) = \sup_S (\Phi - I)$$

so s_0 achieves the supremum of $\Phi - I$ on S. □

Proposition 9.4. *Let $I : E \to [0,\infty]$ be a good rate function on a Polish space (E,ρ), and let S be a closed, non-empty subset of E. Define for $\delta > 0$*

$$\begin{aligned} S^\delta &= \{s \in E : \rho(s,S) < \delta\} \\ &= \{s \in E : \rho(s,s') < \delta \text{ for some } s' \in S\}\,. \end{aligned}$$

Then

$$\inf_{s\in S^\delta} I(s) \uparrow \inf_{s\in S} I(s) \quad \text{as } \delta \downarrow 0\,.$$

Proof. We argue by contradiction. Suppose $\inf_{s\in S^\delta} I(s) \uparrow l$ as $\delta \downarrow 0$ for some l, but $l < \inf_{s\in S} I(s)$. Then for every n, we may choose $s_n \in S^{1/n}$ such that

$$\begin{aligned} I(s_n) &\le \inf_{s\in S^{1/n}} I(s) + \frac{1}{n} \\ &\le l + \frac{1}{n}\,. \end{aligned}$$

In particular $(s_n) \subset \{I \le l+1\}$. But I is a good rate, so $\{I < l+1\}$ is compact, hence there is a convergent sub-sequence $\{s_{n'}\}$ such that $s_{n'} \to s_0$. S is closed, so $s_0 \in S$, and therefore $I(s_0) \ge \inf_{s\in S} I(s)$. While by the lower semi-continuity

$$\begin{aligned} I(s_0) &\le \liminf_{n\to\infty} I(s_{n'}) \\ &\le l \\ &< \inf_{s\in S} I(s) \end{aligned}$$

which is a contradiction. □

A very useful tool is Varadhan's contraction principle.

Theorem 9.5. (S. R. S. Varadhan) *Let I be a good rate function governing the large deviations of $\{P_\varepsilon : \varepsilon > 0\}$ on E, and $f : E \to E'$ a continuous map (E, E' are Polish spaces). Define*

$$I'(s') = \inf\{I(s) : s \in E \text{ such that } f(s) = s'\}\,.$$

Then I' is a good function on E' that governs the large deviations of $\{P_\varepsilon \circ f^{-1} : \varepsilon > 0\}$.

Exercise in Problem Sheet 3.

Indeed all conclusions follow directly from definitions. We will give a proof of a generalization of the contraction principle.

Definition 9.6. *Let $\{P_\varepsilon : \varepsilon > 0\}$ be a family of probability measures on a Polish space (E,ρ). Then $\{P_\varepsilon : \varepsilon > 0\}$ is* exponentially tight *if for every $c > 0$ there is a compact set K_c in E such that*

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log P_\varepsilon(E \setminus K_c) \le -c. \tag{9.6}$$

Begin with an elementary fact.

Lemma 9.7. *If x_ε and y_ε are positive (for small $\varepsilon > 0$), then*

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log(x_\varepsilon + y_\varepsilon) \le \max\left\{\limsup_{\varepsilon\downarrow 0} \varepsilon \log x_\varepsilon, \limsup_{\varepsilon\downarrow 0} \varepsilon \log y_\varepsilon\right\} \tag{9.7}$$

and

$$\liminf_{\varepsilon\downarrow 0} \log(x_\varepsilon + y_\varepsilon) \ge \min\left\{\liminf_{\varepsilon\downarrow 0} \varepsilon \log x_\varepsilon,\ \liminf_{\varepsilon\downarrow 0} \varepsilon \log\ y_\varepsilon\right\}. \tag{9.8}$$

Proof. We have

$$\begin{aligned} \log(x_\varepsilon + y_\varepsilon) &\le \log\{2\max(x_\varepsilon,\ y_\varepsilon)\} \\ &= \log 2 + \log\{\max(x_\varepsilon,\ y_\varepsilon)\} \\ &= \log 2 + \max\{\log x_\varepsilon,\ \log\ y_\varepsilon\} \end{aligned}$$

so that

$$\begin{aligned} \limsup_{\varepsilon\downarrow 0} \varepsilon \log(x_\varepsilon + y_\varepsilon) &\le \limsup_{\varepsilon\downarrow 0} \varepsilon \log 2 + \limsup_{\varepsilon\downarrow 0} \varepsilon \max\{\log x_\varepsilon,\ \log\ y_\varepsilon\} \\ &= \limsup_{\varepsilon\downarrow 0} \max\{\varepsilon \log x_\varepsilon,\ \varepsilon \log\ y_\varepsilon\} \\ &\le \max\left\{\limsup_{\varepsilon\downarrow 0} \varepsilon \log x_\varepsilon,\ \limsup_{\varepsilon\downarrow 0} \varepsilon \log\ y_\varepsilon\right\}. \end{aligned}$$

Similarly

$$\begin{aligned} \log(x_\varepsilon + y_\varepsilon) &\ge \log\{\min(x_\varepsilon,\ y_\varepsilon)\} \\ &= \min\{\log x_\varepsilon,\ \log\ y_\varepsilon\} \end{aligned}$$

so that

$$\begin{aligned} \liminf_{\varepsilon\downarrow 0} \log(x_\varepsilon + y_\varepsilon) &\ge \min\left\{\liminf_{\varepsilon\downarrow 0} \varepsilon \log x_\varepsilon,\ \liminf_{\varepsilon\downarrow 0} \varepsilon \log\ y_\varepsilon\right\} \\ &\ge \min\left\{\liminf_{\varepsilon\downarrow 0} \varepsilon \log x_\varepsilon,\ \liminf_{\varepsilon\downarrow 0} \varepsilon \log\ y_\varepsilon\right\}. \end{aligned}$$

□

The following result is similar to the tightness criterion for weak convergence.

Theorem 9.8. *If $\{P_\varepsilon : \varepsilon > 0\}$ is exponentially tight and satisfies the weak large deviation principle with rate function I, then*

1) I is a good rate function.

2) I governs the large deviations of $\{P_\varepsilon : \varepsilon > 0\}$.

Proof. For every $c > 0$, we choose a compact set K_c such that (9.6) holds. Since K_c is compact, by the weak large deviation principle

$$-c \ge \liminf_{\varepsilon\downarrow 0} \varepsilon \log P_\varepsilon(E \setminus K_c) \ge -\inf_{E\setminus K_c} I$$

which yields that

$$\inf_{E\setminus K_c} I \geq c$$

so that

$$I_c = \{s : I(s) \leq c\} \subset K_{c+1}.$$

Since I is lower semi-continuous, so that I_c is a closed subset of a compact subset K_{c+1}. Thus I_c is compact.

Next we prove that the upper large deviation bound holds for any closed set S in E. In fact for any $c > 0$

$$P_\varepsilon(S) \leq P_\varepsilon(K_c \cap S) + P_\varepsilon(E \setminus K_c)$$

so that

$$\begin{aligned}
\limsup_{\varepsilon\downarrow 0} \varepsilon \log P_\varepsilon(S) &\leq \limsup_{\varepsilon\downarrow 0} \varepsilon \log \left[P_\varepsilon(K_c \cap S) + P_\varepsilon(E \setminus K_c)\right] \\
&\leq -\left[\left(\inf_{K_c\cap S} I\right) \wedge c\right] \\
&\leq -\left[\left(\inf_S I\right) \wedge c\right]
\end{aligned}$$

for every $c > 0$. Letting $c \to \infty$ we obtain

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log P_\varepsilon(S) \leq -\inf_S I$$

which is the upper bound for large deviations. □

Theorem 9.9. *(*S. R. S. Varadhan*) Let I be a rate function governing the weak large deviations of $\{P_\varepsilon : \varepsilon > 0\}$, and let $\Phi : E \to (-\infty, \infty]$ be lower semi-continuous. Then*

$$\liminf_{\varepsilon\downarrow 0} \varepsilon \log \int_E e^{\frac{\Phi}{\varepsilon}} dP_\varepsilon \geq \sup_{\Phi\wedge I<\infty} (\Phi - I). \tag{9.9}$$

Proof. As we have indicated, lower bounds for large deviations reflect only local property of the family of distributions P_ε. This point is demonstrated clearly from the proof below. For any $s \in E$ such that $\Phi(s) - I(s) < \infty$ and $\delta > 0$ we have

$$\begin{aligned}
\liminf_{\varepsilon\downarrow 0} \varepsilon \log \int_E e^{\frac{\Phi}{\varepsilon}} dP_\varepsilon &\geq \liminf_{\varepsilon\downarrow 0} \varepsilon \log \int_{B(s,\delta)} e^{\frac{\Phi}{\varepsilon}} dP_\varepsilon \\
&\geq \liminf_{\varepsilon\downarrow 0} \varepsilon \log \left\{ e^{\frac{\inf_{B(s,\delta)} \Phi}{\varepsilon}} P_\varepsilon(B(s,\delta)) \right\} \\
&= \inf_{B(s,\delta)} \Phi + \liminf_{\varepsilon\downarrow 0} \varepsilon \log P_\varepsilon(B(s,\delta)) \\
&\geq \inf_{B(s,\delta)} \Phi - \inf_{B(s,\delta)} I \\
&\geq \inf_{B(s,\delta)} \Phi - I(s)
\end{aligned}$$

Letting $\delta \downarrow 0$, since Φ is lower semi-continuous, we obtain

$$\begin{aligned}
\liminf_{\varepsilon\downarrow 0} \varepsilon \log \int_E e^{\frac{\Phi}{\varepsilon}} dP_\varepsilon &\geq \lim_{\delta\downarrow 0} \inf_{B(s,\delta)} \Phi - I(s) \\
&\geq \Phi(s) - I(s)
\end{aligned}$$

for any $s \in E$, it follows thus that

$$\liminf_{\varepsilon \downarrow 0} \varepsilon \log \int_E e^{\frac{\Phi}{\varepsilon}} dP_\varepsilon \geq \sup_{\Phi \wedge \phi < \infty} (\Phi - I).$$

□

Theorem 9.10. (S. R. S. Varadhan) *Let I be a good rate function that governs the large deviations of $\{P_\varepsilon : \varepsilon > 0\}$, and $\Phi : E \to [-\infty, \infty)$ upper semi-continuous which satisfies that*

$$\lim_{c \to \infty} \limsup_{\varepsilon \downarrow 0} \varepsilon \log \int_{\{\Phi \geq c\}} e^{\frac{\Phi}{\varepsilon}} dP_\varepsilon = -\infty. \tag{9.10}$$

Then

$$\limsup_{\varepsilon \downarrow 0} \varepsilon \log \int_E e^{\frac{\Phi}{\varepsilon}} dP_\varepsilon \leq \sup_E (\Phi - I). \tag{9.11}$$

Proof. Let us first consider the case that Φ is bounded above by $M > 0$. Given $c > 0$, the level set $K_c = \{I \leq c\}$ is compact in E. Since Φ is upper semi-continuous, for any $\delta > 0$, we can choose finite many s_i in K_c and positive numbers r_i (where $1 \leq i \leq n$) so that K_c is covered by balls $B_i \equiv B(s_i, r_i)$, and

$$\sup_{\bar{B}_i} \Phi \leq \Phi(s_i) + \delta \ ; \quad \inf_{\bar{B}_i} I \geq I(s_i) - \delta \quad \text{for all } i \, .$$

Let $G = \cup_{i=1}^n B_i$. Then

$$\begin{aligned}
\int_E e^{\frac{\Phi}{\varepsilon}} dP_\varepsilon &= \int_{E \setminus G} e^{\frac{\Phi}{\varepsilon}} dP_\varepsilon + \int_{\cup_{i=1}^n B_i} e^{\frac{\Phi}{\varepsilon}} dP_\varepsilon \\
&\leq e^{\frac{M}{\varepsilon}} P_\varepsilon(E \setminus G) + \sum_{i=1}^n \int_{B_i} e^{\frac{\Phi}{\varepsilon}} dP_\varepsilon \\
&\leq e^{\frac{M}{\varepsilon} + \log P_\varepsilon(E \setminus G)} + \sum_{i=1}^n e^{\frac{\Phi(s_i)+\delta}{\varepsilon} + \log P_\varepsilon(B_i)} \\
&\leq e^{\frac{1}{\varepsilon}(M + \varepsilon \log P_\varepsilon(E \setminus G))} \\
&\quad + \sum_{i=1}^n e^{\frac{1}{\varepsilon}((\Phi(s_i)+\delta) + \varepsilon \log P_\varepsilon(\bar{B}_i))}
\end{aligned}$$

and therefore

$$\begin{aligned}
&\limsup_{\varepsilon \downarrow 0} \varepsilon \log \int_E e^{\frac{\Phi}{\varepsilon}} dP_\varepsilon \\
\leq\ & \max \left\{ M + \limsup_{\varepsilon \downarrow 0} \varepsilon \log P_\varepsilon (E \setminus G), (\Phi(s_i) + \delta) + \limsup_{\varepsilon \downarrow 0} \varepsilon \log P_\varepsilon (\bar{B}_i) \right\} \\
\leq\ & \max \left\{ M - \inf_{E \setminus G} I, (\Phi(s_i) + \delta) - \inf_{\bar{B}_i} I \right\} \\
\leq\ & \max \{ M - c, (\Phi(s_i) + \delta) - I(s_i) + \delta \} \\
\leq\ & \max \left\{ \sup_E (\Phi - I), M - c \right\} + 2\delta \, .
\end{aligned}$$

Letting $\delta \downarrow 0$ and letting $c \to \infty$ we obtain (9.11).

For general Φ

$$\begin{aligned} \int_E e^{\frac{\Phi}{\varepsilon}} dP_\varepsilon &= \int_{\Phi<c} e^{\frac{\Phi}{\varepsilon}} dP_\varepsilon + \int_{\Phi\geq c} e^{\frac{\Phi}{\varepsilon}} dP_\varepsilon \\ &\leq \int_E e^{\frac{\Phi\wedge c}{\varepsilon}} dP_\varepsilon + \int_{\Phi\geq c} e^{\frac{\Phi}{\varepsilon}} dP_\varepsilon \end{aligned}$$

so that

$$\begin{aligned} & \limsup_{\varepsilon\downarrow 0} \varepsilon \log \int_E e^{\frac{\Phi}{\varepsilon}} dP_\varepsilon \\ \leq \;& \max\left\{ \sup_E (\Phi\wedge c - I), \limsup_{\varepsilon\downarrow 0} \varepsilon \log \int_{\Phi\geq c} e^{\frac{\Phi}{\varepsilon}} dP_\varepsilon \right\} \\ \leq \;& \max\left\{ \sup_E (\Phi - I), \limsup_{\varepsilon\downarrow 0} \varepsilon \log \int_{\Phi\geq c} e^{\frac{\Phi}{\varepsilon}} dP_\varepsilon \right\} \end{aligned}$$

and finally by letting $c \to \infty$, since

$$\lim_{c\to\infty} \limsup_{\varepsilon\downarrow 0} \varepsilon \log \int_{\Phi\geq c} e^{\frac{\Phi}{\varepsilon}} dP_\varepsilon = -\infty$$

we obtain

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log \int_E e^{\frac{\Phi}{\varepsilon}} dP_\varepsilon \leq \sup_E (\Phi - I).$$

□

Theorem 9.11. (S. R. S. Varadhan) *Let I be a good rate function that governs the large deviations of $\{P_\varepsilon : \varepsilon > 0\}$, and $\Phi : E \to R^1$ a continuous function which satisfies*

$$\lim_{c\to\infty} \limsup_{\varepsilon\downarrow 0} \varepsilon \log \int_{\{\Phi\geq c\}} e^{\frac{\Phi}{\varepsilon}} dP_\varepsilon = -\infty. \tag{9.12}$$

Then

$$\lim_{\varepsilon\downarrow 0} \varepsilon \log \int_E e^{\frac{\Phi}{\varepsilon}} dP_\varepsilon = \sup_E (\Phi - I) . \tag{9.13}$$

It follows directly from the previous two theorems.

Lemma 9.12. *If a continuous function $\Phi : E \to R^1$ satisfies*

$$\sup_{0<\varepsilon\leq 1} \left(\int_E e^{\alpha\frac{\Phi}{\varepsilon}} dP_\varepsilon \right)^{\varepsilon} < \infty$$

for some $\alpha > 1$, then (9.12) holds, that is

$$\lim_{c\to\infty} \limsup_{\varepsilon\downarrow 0} \varepsilon \log \int_{\{\Phi\geq c\}} e^{\frac{\Phi}{\varepsilon}} dP_\varepsilon = -\infty.$$

Proof. Indeed by Hölder's inequality

$$\begin{aligned}
\int_{\{\Phi\geq c\}} e^{\frac{\Phi}{\varepsilon}} dP_\varepsilon &\leq \left(\int_E e^{\alpha\frac{\Phi}{\varepsilon}} dP_\varepsilon\right)^{1/\alpha} \left(\int_{\{\Phi\geq c\}} dP_\varepsilon\right)^{1-1/\alpha} \\
&\leq \left(\int_E e^{\alpha\frac{\Phi}{\varepsilon}} dP_\varepsilon\right)^{1/\alpha} \left(e^{-\alpha\frac{c}{\varepsilon}} \int_{\{\Phi\geq c\}} e^{\alpha\frac{\Phi}{\varepsilon}} dP_\varepsilon\right)^{1-1/\alpha} \\
&\leq e^{(1-\alpha)\frac{c}{\varepsilon}} \int_E e^{\alpha\frac{\Phi}{\varepsilon}} dP_\varepsilon
\end{aligned}$$

and therefore

$$\begin{aligned}
&\varepsilon \log \int_{\{\Phi\geq c\}} e^{\frac{\Phi}{\varepsilon}} dP_\varepsilon \\
\leq\ & (1-\alpha)c + \log\left(\int_E e^{\alpha\frac{\Phi}{\varepsilon}} dP_\varepsilon\right)^{\varepsilon}.
\end{aligned}$$

The conclusion thus follows immediately. □

10 Contraction principles, continuity theorem

In practice, we more often deal with random variables directly, so we may introduce the following

Definition 10.1. *Let $\{X^\varepsilon : \varepsilon \in (0,1)\}$ be a family of random variables in a Polish space E on a probability space $(\Omega, \mathscr{F}, P)$. We say $\{X^\varepsilon : \varepsilon \in (0,1)\}$ satisfies LDP with a rate function I, if*

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log P[X^\varepsilon \in F] \leq -\inf_{s\in F} I(s) \tag{10.1}$$

for every closed subset $F \subset E$ in E, and

$$\liminf_{\varepsilon\downarrow 0} \varepsilon \log P[X^\varepsilon \in G] \geq -\inf_{s\in G} I(s) \tag{10.2}$$

for every open subset $G \subset E$.

In this section, we study the following question. If $\{X_n^\varepsilon : \varepsilon \in (0,1)\}$ $(n = 1,2,\cdots)$ are families of random variables valued in E on $(\Omega, \mathscr{F}, P)$, suppose X_n^ε converges to X^ε as $n \to \infty$ (for each ε), and suppose for each n, $\{X_n^\varepsilon : \varepsilon \in (0,1)\}$ obeys LDP with rate function I_n, under what conditions, the limiting distributions of $\{X^\varepsilon : \varepsilon \in (0,1)\}$ also satisfies LDP, and if so, how to calculate its rate function?

First of all we have the following contraction principle.

Theorem 10.2. *Let E, E' be two Polish spaces, and $f : E \to E'$ be continuous. Suppose $\{Z^\varepsilon : \varepsilon \in (0,1)\}$ satisfies a LDP with a good rate function I, then $X^\varepsilon = f(Z^\varepsilon)$ satisfies LDP with rate function*

$$I'(s') = \inf\left\{I(s) : s \in E \text{ such that } f(s) = s'\right\}.$$

A generalization of this theorem is needed in order to deal with Wiener functionals which are only measurable functions of Brownian motion sample paths. To this end, we introduce the following

Definition 10.3. *Let* $\{X_n^\varepsilon : \varepsilon \in (0,1)\}$ *and* $\{X^\varepsilon : \varepsilon \in (0,1)\}$ *be families of random variables in a Polish space* E *on* $(\Omega, \mathscr{F}, P)$*. Then we say* X_n^ε *converges to* X^ε *as* $n \to \infty$ *exponentially, if for every* $\delta > 0$

$$\lim_{n\to\infty} \limsup_{\varepsilon \downarrow 0} \varepsilon \log P\left[\rho(X_n^\varepsilon, X^\varepsilon) > \delta\right] = -\infty \tag{10.3}$$

where ρ *is the distance function on* E*.*

In Lemma 10.4 and Lemma 10.5 below, $\{X_n^\varepsilon : \varepsilon \in (0,1)\}$ and $\{X^\varepsilon : \varepsilon \in (0,1)\}$ are families of random variables in a Polish space (E, ρ) on a complete probability space $(\Omega, \mathscr{F}, P)$, X_n^ε converges to X^ε as $n \to \infty$ exponentially and for each n, $\{X_n^\varepsilon : \varepsilon > 0\}$ satisfies LDP with a good rate function I_n.

Lemma 10.4. *Let* $G \subset E$ *be open. Then*

$$\liminf_{\varepsilon\downarrow 0} \varepsilon \log P\left[X^\varepsilon \in G\right] \geq -\limsup_{n\to\infty} \inf_{B_s(\delta)} I_n \tag{10.4}$$

for every $s \in E$ *and* $\delta > 0$ *such that* $B_s(2\delta) \subset G$*.*

Proof. By the triangle inequality

$$\{X_n^\varepsilon \in B_s(\delta)\} \subset \{X^\varepsilon \in B_s(2\delta)\} \cup \{\rho(X_n^\varepsilon, X^\varepsilon) > \delta\}$$

and therefore

$$\begin{aligned} P\left[X_n^\varepsilon \in B_s(\delta)\right] &\leq P\left[X^\varepsilon \in B_s(2\delta)\right] + P\left[\rho(X_n^\varepsilon, X^\varepsilon) > \delta\right] \\ &\leq P\left[X^\varepsilon \in G\right] + P\left[\rho(X_n^\varepsilon, X^\varepsilon) > \delta\right] \end{aligned}$$

whenever $B_s(2\delta) \subset G$. Therefore

$$\varepsilon \log\left(P\left[X^\varepsilon \in G\right] + P\left[\rho(X_n^\varepsilon, X^\varepsilon) > \delta\right]\right) \geq \varepsilon \log P\left[X_n^\varepsilon \in B_s(\delta)\right].$$

By Lemma 9.7

$$\begin{aligned} &\liminf_{\varepsilon\downarrow 0} \varepsilon \log \left\{P\left[X^\varepsilon \in G\right] + P\left[\rho(X_n^\varepsilon, X^\varepsilon) > \delta\right]\right\} \\ \leq\ &\max\left\{\liminf_{\varepsilon\downarrow 0} \varepsilon \log P\left[X^\varepsilon \in G\right], \liminf_{\varepsilon\downarrow 0} \varepsilon \log P\left[\rho(X_n^\varepsilon, X^\varepsilon) > \delta\right]\right\}, \end{aligned}$$

and by assumption that $X_n^\varepsilon \to X^\varepsilon$ exponentially as $n \to \infty$, for every $c > 0$, there is an number N_1 such that

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log P\left[\rho(X_n^\varepsilon, X^\varepsilon) > \delta\right] \leq -c \quad \forall n \geq N_1.$$

Hence for $n > N_1$ we have

$$\begin{aligned} \liminf_{\varepsilon\downarrow 0} \varepsilon \log P\left[X_n^\varepsilon \in B_s(\delta)\right] &\leq \liminf_{\varepsilon\downarrow 0} \varepsilon \log\left(P\left[X^\varepsilon \in G\right] + P\left[\rho(X_n^\varepsilon, X^\varepsilon) > \delta\right]\right) \\ &\leq \liminf_{\varepsilon\downarrow 0} \varepsilon \log P\left(X^\varepsilon \in G\right) \vee (-c), \end{aligned}$$

together with the LDP lower bound for X_n^ε, we deduce that

$$\liminf_{\varepsilon\downarrow 0} \varepsilon \log P\left[X^\varepsilon \in G\right] \vee (-c) \geq -\inf_{B_s(\delta)} I_n \tag{10.5}$$

for all $n > N_1$. Hence

$$\liminf_{\varepsilon\downarrow 0} \varepsilon \log P[X^\varepsilon \in G] \vee (-c) \geq -\limsup_{n\to\infty} \inf_{B_s(\delta)} I_n \tag{10.6}$$

for every $c > 0$. If $\liminf_{\varepsilon\downarrow 0} \varepsilon \log P[X^\varepsilon \in G] = -\infty$ then

$$-c \geq -\limsup_{n\to\infty} \inf_{B_s(\delta)} I_n$$

for every $c > 0$, so that $\limsup_{n\to\infty} \inf_{B_s(\delta)} I_n(s) = \infty$. Otherwise, letting $c \to \infty$ in (10.6) one obtains that

$$\liminf_{\varepsilon\downarrow 0} \varepsilon \log P[X^\varepsilon \in G] \geq -\limsup_{n\to\infty} \inf_{B_s(\delta)} I_n,$$

which completes the proof. □

Lemma 10.5. *Let $S \subset E$ be a closed subset. Then*

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log P[X^\varepsilon \in S] \leq -\lim_{\delta\downarrow 0} \liminf_{n\to\infty} \inf_{s\in \bar{S^\delta}} I_n(s) \tag{10.7}$$

where $S^\delta = \{s \in E : \rho(s,S) < \delta\}$ for every $\delta > 0$.

Proof. By the triangle inequality we have

$$\{X^\varepsilon \in S\} \subset \left\{X_n^\varepsilon \in \bar{S^\delta}\right\} \cup \{\rho(X_n^\varepsilon, X^\varepsilon) > \delta\}$$

so that

$$P[X^\varepsilon \in S] \leq P\left[X_n^\varepsilon \in \bar{S^\delta}\right] + P[\rho(X_n^\varepsilon, X^\varepsilon) > \delta]$$

and therefore

$$\varepsilon \log P[X^\varepsilon \in S] \leq \varepsilon \log \left\{P\left[X_n^\varepsilon \in \bar{S^\delta}\right] + P[\rho(X_n^\varepsilon, X^\varepsilon) > \delta]\right\}.$$

It follows that

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log P[X^\varepsilon \in S] \leq \limsup_{\varepsilon\downarrow 0} \varepsilon \log P\left[X_n^\varepsilon \in \bar{S^\delta}\right] \vee \limsup_{\varepsilon\downarrow 0} \varepsilon \log P[\rho(X_n^\varepsilon, X^\varepsilon) > \delta]$$

for every n and $\delta > 0$. Since $X_n^\varepsilon \to X^\varepsilon$ exponentially as $n \to \infty$, so by letting $n \to \infty$ in the previous inequality we obtain

$$\begin{aligned}
\limsup_{\varepsilon\downarrow 0} \varepsilon \log P[X^\varepsilon \in S] &\leq \limsup_{n\to\infty} \limsup_{\varepsilon\downarrow 0} \varepsilon \log P\left[X_n^\varepsilon \in \bar{S^\delta}\right] \\
&\leq \limsup_{n\to\infty} \left(-\inf_{\bar{S^\delta}} I_n\right) \\
&= -\liminf_{n\to\infty} \inf_{\bar{S^\delta}} I_n
\end{aligned}$$

for every $\delta > 0$. Finally by sending $\delta \downarrow 0$ we therefore have

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log P[X^\varepsilon \in S] \leq -\lim_{\delta\downarrow 0} \liminf_{n\to\infty} \inf_{\bar{S^\delta}} I_n.$$

□

As a consequence, we have the following trivial but useful corollary.

Corollary 10.6. *Let $\{X^\varepsilon : \varepsilon \in (0,1)\}$ and $\{Y^\varepsilon : \varepsilon \in (0,1)\}$ be two families of random variables taking values in E, and Y^ε satisfy LDP with a good rate function I. Suppose X^ε and Y^ε are exponentially close in the sense that for every $\delta > 0$*

$$\limsup_{\varepsilon \downarrow 0} \varepsilon \log P[\rho(X^\varepsilon, Y^\varepsilon) > \delta] = -\infty$$

then X^ε also satisfies LDP with rate function I.

Proof. If G is open, then for every $s \in E$ and $\delta > 0$ such that $B_s(2\delta) \subset G$ by Lemma \ref{le-c-1}

$$\liminf_{\varepsilon \downarrow 0} \varepsilon \log P[X^\varepsilon \in G] \geq - \inf_{B_s(\delta)} I \tag{10.8}$$

which yields the lower bound of LDP

$$\liminf_{\varepsilon \downarrow 0} \varepsilon \log P[X^\varepsilon \in G] \geq - \inf_G I.$$

According to Lemma 10.5

$$\limsup_{\varepsilon \downarrow 0} \varepsilon \log P[X^\varepsilon \in S] \leq - \lim_{\delta \downarrow 0} \inf_{s \in \overline{S^\delta}} I(s) \tag{10.9}$$

for every closed set S. Since I is a good rate function, so that

$$\lim_{\delta \downarrow 0} \inf_{s \in \overline{S^\delta}} I(s) = \inf_S I$$

we thus also have LDP upper bound for X^ε. □

The proof of the following theorem may be also found in standard books on large deviations such as [24].

Theorem 10.7. *Let E, E' be two Polish spaces, $f_n : E \to E'$ be a sequence of continuous mappings, and I be a good rate function. Let*

$$I'_n(s') = \inf \left\{ I(s) : s \in E \text{ and } f_n(s) = s' \right\}$$

where $n = 1, 2, \cdots$. Suppose $f_n \to f$ uniformly on $I_c = \{x : I(x) \leq c\}$ for every $c \geq 0$ as $n \to \infty$, so f is well defined on $\{s \in E : I(s) < \infty\}$. Then

$$I'(s') = \inf \left\{ I(s) : s \in E \text{ such that } I(s) < \infty \text{ and } f(s) = s' \right\}$$

is a good rate function.

Suppose $\{X_n^\varepsilon : \varepsilon \in (0,1)\}$ is a family of random variables in E' satisfying LDP with rate function I'_n, and $\{X_n^\varepsilon : \varepsilon \in (0,1)\}$ converges to $\{X^\varepsilon : \varepsilon \in (0,1)\}$ exponentially

$$\lim_{n \to \infty} \limsup_{\varepsilon \downarrow 0} \varepsilon \log P\left[\rho'(X_n^\varepsilon, X^\varepsilon) > \delta\right] = -\infty, \tag{10.10}$$

where ρ' is the distance function on E', then $\{X^\varepsilon : \varepsilon \in (0,1)\}$ satisfies LDP with rate function I'.

As a special and useful example, we have the following corollary, which is called the contraction principle in LDP literature.

Theorem 10.8. *Let E, E' be two Polish spaces, and $f_n : E \to E'$ be continuous mappings. Suppose random variables $\{Z^{\varepsilon} : \varepsilon \in (0,1)\}$ in E satisfies LDP with a good rate function I. Suppose*

1) $f_n(Z^{\varepsilon}) = X_n^{\varepsilon}$ converges to X^{ε} exponentially as $n \to \infty$,

2) f_n converges to f uniformly on $I_c = \{x : I(x) \leq c\}$ for every $c \geq 0$, so f is well defined on $\{s \in E : I(s) < \infty\}$.

Then X^{ε} satisfies LDP with rate function

$$I'(s') = \inf\left\{I(s) : s \in E \text{ such that } I(s) < \infty \text{ and } f(s) = s'\right\}.$$

Proof. of Theorem (10.7). First of all we show that I' is a good rate function on E'. Let $c \geq 0$ be a constant and consider $I'_c = \{s' : I'(s') \leq c\}$. Suppose $\{s'_n\}$ is a sequence in I'_c, we want to show that it has a convergent sub-sequence with a limit in I'_c. Since I is a good rate function and f is a uniform limit of continuous mapping on $\{s \in E : I(s) \leq C\}$ for every $C \geq 0$, so for every n we may choose $s_n \in E$ such that $f(s_n) = s'_n$ and $I(s_n) \leq c + \frac{1}{n}$. In particular $\{s_n\} \subset I_{c+1}$. Since I_{c+1} is compact, so without loss of generality we may assume that $s_n \to s$ in E. Since I is lower-semi continuous

$$I(s) \leq \liminf_{n \to \infty} I(s_n) \leq c.$$

In particular $s_n \to s$ in I_{c+1}, and f is continuous on I_{c+1}, so that $f(s_n) = s'_n \to f(s) \equiv s'$. By definition of I', we also have $I'(s') \leq c$, so that $s' \in I'_c$, which proves that I'_c is compact. By definition I' is a good rate function on E'.

Next we are going to prove LDP bounds. First we prove the lower bound. We need to show that

$$\liminf_{\varepsilon \downarrow 0} \varepsilon \log P[X^{\varepsilon} \in G] \geq -\inf_{s' \in G} I'(s') \tag{10.11}$$

for every open set $G \subset E'$. According to Lemma 10.4,

$$\liminf_{\varepsilon \downarrow 0} \varepsilon \log P[X^{\varepsilon} \in G] \geq -\limsup_{n \to \infty} \inf_{B_{s'_0}(\delta)} I'_n$$

for any $\delta > 0$ such that $B_{s'_0}(\delta) \subset G$, so we only need to prove that

$$\limsup_{n \to \infty} \inf_{B_{s'_0}(\delta)} I'_n \leq \inf_{s' \in G} I'(s').$$

If $\inf_{s' \in G} I'(s') = \infty$ then there is nothing to prove, so we assume that $l = \inf_G I' < \infty$. For every $\theta > 0$, there is an $s'_0 \in G$ such that

$$l \leq I'(s'_0) \leq l + \theta.$$

Choose $\delta > 0$ smaller than θ, such that $B_{s'_0}(2\delta) \subset G$. By definition, there is an $s_0 \in E$ such that $I(s_0) < \infty$, $f(s_0) = s'_0$ and $I(s_0) = I'(s'_0)$. Hence $l \leq I(s_0) \leq l + \theta$. Since f_n converges to f uniformly on $I_{l+\theta}$, there is an N_1 such that

$$\rho'(f_n(s), f(s)) < \theta \qquad \forall n \geq N_1, s \in I_{l+\theta}.$$

In particular

$$\rho'(f_n(s_0), s'_0) < \theta \qquad \forall n \geq N_1.$$

It follows, by the definition of I'_n, that

$$\inf_{B_{s'_0}(\delta)} I'_n \leq \inf_{B_{s'_0}(\theta)} I'_n \leq I'(s'_0) = I(s_0) \leq l + \theta \quad \forall n \geq N_1$$

so that

$$\limsup_{n\to\infty} \inf_{B_{s'_0}(\delta)} I'_n \leq l + \theta.$$

Since X^ε_n converges to X^ε exponentially, by Lemma 10.4, we have

$$\liminf_{\varepsilon\downarrow 0} \varepsilon \log P[X^\varepsilon \in G] \geq -\limsup_{n\to\infty} \inf_{B_{s'_0}(\delta)} I'_n \geq -l - \theta$$

for every $\theta > 0$. Letting $\theta \downarrow 0$ to obtain the lower bound (10.11).

Next we prove the upper bound. Let $S \subset E'$ be a closed subset, and $S^\delta = \{s' \in E' : \rho'(s', S) < \delta\}$ where $\delta > 0$. Then, by Lemma 10.5,

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log P[X^\varepsilon \in S] \leq -\lim_{\delta\downarrow 0} \liminf_{n\to\infty} \inf_{\overline{S^\delta}} I'_n. \tag{10.12}$$

Let

$$l = \lim_{\delta\downarrow 0} \liminf_{n\to\infty} \inf_{\overline{S^\delta}} I'_n.$$

If $l = \infty$ then there is nothing to prove. Therefore we assume that $l < \infty$. Then, for each n there is an $s'_n \in E'$, such that $\rho'(s'_n, S) \leq \frac{1}{n}$, and

$$I'_n(s'_n) \leq l + \frac{1}{n}.$$

For each n we may choose an $s_n \in E$ such that $f_n(s_n) = s'_n$ and $I(s_n) \leq l + \frac{1}{n}$. Since I is a good function and s_n belongs to the compact set I_{l+1}, thus, without loss of generality, we may assume that $s_n \to s$. Since f_n converges uniformly on I_{l+1}, so that $f_n(s_n) \to s'$ and $f(s) = s'$. Since S is closed, $s' \in S$ and $I(s) \leq \lim_{n\to\infty} I(s_n) \leq l$, so that $I'(s') \leq l$. Hence

$$\inf_{s'\in S} I'(s') \leq I'(s') \leq l$$

and therefore

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log P[X^\varepsilon \in S] \leq -l \leq -\inf_{s'\in S} I'(s')$$

which completes the proof. □

11 Schilder's theorem

In this section we prove Schilder's large deviation principle for Brownian motion in uniform topology. There are many different proofs, we however present a proof by using an approximation procedure which can be used to prove refined versions, and to prove other large deviation problems.

Let us follow the notations used in literature on analysis of Brownian motion, such as stochastic differential equations, Malliavin calculus and ec. In literature of these theories, the space $\mathbb{C}(R^d)$ (or equivalently $C([0,\infty), R^d)$ is denoted by W^d. Since we will deal with the law of the standard Brownian motion with time duration $[0,1]$, which is supported on the continuous paths w with initial zero and time duration $[0,1]$, therefore, for simplicity, we consider the path space $\mathbb{W}^d = C_0([0,1];R^d)$ with the uniform norm $||\cdot||_{\mathbb{W}^d}$, $\mathbb{H} = H_0^1([0,1];R^d)$ and μ be the Wiener measure, i.e. the law of Brownian motion $(B_t)_{t\in[0,1]}$ with $B_0 = 0$.

$\mathbb{H}$ is the space of all R^d-valued functions h on $[0,1]$ whose generalized derivative $\dot{h} = \frac{d}{dt}h$ is square integrable, and $h(0) = 0$. The Hilbert norm of such h is given by

$$||h||_{H^1} = \left(\int_0^1 \left|\dot{h}(t)\right|^2 dt\right)^{1/2}.$$

Since

$$|h(t) - h(s)| = \left|\int_s^t \dot{h}(r)dr\right| \leq \sqrt{t-s}\,||h||_{H^1}$$

for all $0 \leq s \leq t \leq 1$, so we have a canonical continuous embedding $\mathbb{H} \hookrightarrow \mathbb{W}^d$. Therefore we will consider $\mathbb{H}$ as a subset of $\mathbb{W}^d$. That is, every element h of $\mathbb{H}$ is represented by its continuous version, still denoted by h, in its equivalence class. Note that $\mathbb{W}^d$ equipped with its uniform norm $||\cdot||_{\mathbb{W}^d}$ is a Banach space, and its dual space of continuous linear functionals on $\mathbb{W}^d$, denoted by $\mathbb{W}^{d\star}$, can be identified with a subspace of the dual space of the Hilbert space $\mathbb{H}^\star = \mathbb{H}$. Therefore we have the following canonical continuous embedding

$$\mathbb{W}^{d\star} \hookrightarrow \mathbb{H} \hookrightarrow \mathbb{W}^d.$$

The law of Brownian motion μ is then the unique probability measure on $(\mathbb{W}^d, \mathscr{B}(\mathbb{W}^d))$ such that

$$\int_{\mathbb{W}^d} e^{\sqrt{-1}l(x)}\mu(dx) = e^{-\frac{1}{2}||l||^2_{H^1}} \quad \forall l \in \mathbb{W}^{d\star} \hookrightarrow \mathbb{H}. \tag{11.1}$$

The triple $(\mathbb{W}^d, \mathbb{H}, \mu)$ is called the classical Wiener space. We won't use anything about Gaussian measures or Wiener spaces, except the quadratic function $h \to \frac{1}{2}||h||^2_{H^1}$, the co-variance function of the Wiener measure, which is identified with the rate function for Schilder's large deviation principle.

For $\varepsilon > 0$, define $\Gamma(\varepsilon) : \mathbb{W}^d \to \mathbb{W}^d$ the scaling which sends a path x to $\sqrt{\varepsilon}x$. Then, clearly $\Gamma(\varepsilon)$ is a measurable mapping from $\mathbb{W}^d$ one-to-one and onto $\mathbb{W}^d$, the pull back of the Wiener measure μ under $\Gamma(\varepsilon)$ is denoted by μ^ε, that is

$$\mu^\varepsilon(f) = \int_{\mathbb{W}^d} f(\sqrt{\varepsilon}x)\mu(dx)$$

so that μ^ε is again a Gaussian measure on $(\mathbb{W}^d, \mathscr{B}(\mathbb{W}^d))$ such that

$$\int_{\mathbb{W}^d} e^{\sqrt{-1}l(x)}\mu^\varepsilon(dx) = e^{-\frac{\varepsilon}{2}||l||^2_{H^1}} \quad \forall l \in \mathbb{H}. \tag{11.2}$$

Define $I : \mathbb{W}^d \to [0,\infty]$ by

$$I(x) = \frac{1}{2}||x||^2_{H^1} \quad \text{if } x \in \mathbb{H} \tag{11.3}$$

and $I(x) = \infty$ otherwise.

Theorem 11.1. *(*Schilder*)* $\{\mu^\varepsilon : \varepsilon > 0\}$ *satisfies the large deviation principle with rate function* I *defined by (11.3). That is*

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log \mu^\varepsilon(F) \leq -\inf_{x\in F} I(x) \tag{11.4}$$

for every closed subset $F \subset \mathbb{W}^d$ and

$$\liminf_{\varepsilon \downarrow 0} \varepsilon \log \mu^{\varepsilon}(G) \geq -\inf_{x \in G} I(x) \tag{11.5}$$

for every open subset of $G \subset \mathbb{W}^d$.

The remaining text of the notes is devoted to the proof of Schilder's large deviation principle.

For each n, let $t_n^k = \frac{k}{2^n}$ be the dyadic points in $[0,1]$, where k runs from 0 to 2^n. Let $E_n \equiv R^{2^n \times d} \equiv \prod_{k=1}^{2^n} R^d$, endowed with the Banach norm $||a||_{E_n} = \max_k |a_k|$. The natural embedding

$$g_n : E_n \to \mathbb{W}^d$$

takes an element $a = (a_1, \cdots, a_{2^n}) \in E_n$ to a continuous path $w = g_n(a) \in \mathbb{W}^d$ given by

$$w_t = a_{k-1} + 2^n(t - t_n^{k-1})(a_k - a_{k-1}), \text{ if } t \in [t_n^{k-1}, t_n^k]$$

where $a_0 = 0$.

On the other hand, there is a natural projection $\pi_n : \mathbb{W}^d \to E_n$ by sending a path $w \in \mathbb{W}^d$ to $\pi_n(w) = (a_k)$ with $a_k = w_{t_n^k}$ $(k = 1, \cdots, 2^n)$.

Both mappings g_n and π_n are continuous, and $\pi_n \circ g_n$ is the identity mapping on E_n, and $g_n \circ \pi_n$ is the identity mapping on $g_n^{-1}(E_n)$. Moreover

$$||g_n(a)||_{\mathbb{W}^d} = ||a||_{E_n} \qquad \forall a \in E_n. \tag{11.6}$$

Finally, for each n, we construct a continuous mapping $f_n : \mathbb{W}^d \to \mathbb{W}^d$ by $f_n = g_n \circ \pi_n$. That is, f_n takes each path w to its polygonal approximation $w^{(n)} = f_n(w)$ determined by

$$w_t^{(n)} = w_{t_n^{k-1}} + 2^n(t - t_n^{k-1})(w_{t_n^k} - w_{t_n^{k-1}}). \tag{11.7}$$

Applying mapping f_n to the coordinate process $\{x_t : t \in [0,1]\}$, then $x^{(n)} = f_n(x)$ where

$$x_t^{(n)} = x_{t_n^{k-1}} + 2^n(t - t_n^{k-1})(x_{t_n^k} - x_{t_n^{k-1}}), \text{ if } t \in [t_n^{k-1}, t_n^k].$$

Let μ_n^{ε} denote the distribution of $\{\sqrt{\varepsilon} x_t^{(n)} : t \in [0,1]\}$ which are probability measures on $(\mathbb{W}^d, \mathscr{B}(\mathbb{W}^d))$.

Note that for fixed n, $\{\sqrt{2^n}\left(x_{t_n^k} - x_{t_n^{k-1}}\right) : k = 1, \cdots, 2^n\}$, under the Wiener measure μ, is a family of independent random variables with the same distribution $N(0,1)$. We thus can derive a large deviation principle for the distribution family $\{\mu_n^{\varepsilon} : \varepsilon > 0\}$. More precisely, let $\{\xi_k : k = 1, \cdots, 2^n\}$ be a family of i.i.d with $N(0,1)$, and define b_k by

$$\sqrt{2^n}\,(b_k - b_{k-1}) = \xi_k, \quad k = 1, \cdots, 2^n$$

with $b_0 = 0$. That is

$$b_k = \frac{1}{\sqrt{2^n}} \sum_{l=1}^{k} \xi_l \, .$$

The distribution of $b = (b_1, \cdots, b_{2^n})$ is denoted by ν_n which is a Gaussian measure on the Euclidean space E_n. Let ν_n^{ε} be the law of the scaled Gaussian variable $\sqrt{\varepsilon}(b_1, \cdots, b_{2^n})$. The following is a direct corollary of the Cramér theorem or from a simple computation.

Lemma 11.2. *1) For each $\varepsilon > 0$ and natural number n, $g_n\left(\sqrt{\varepsilon}b\right)$ has the distribution μ_n^ε.*
2) $\{\nu_n^\varepsilon : \varepsilon > 0\}$ satisfies the large deviation principle on E_n with a good rate function

$$I_n(a) = \frac{1}{2}\frac{1}{2^n}\sum_{k=1}^{2^n}(a_k - a_{k-1})^2 \tag{11.8}$$

for any $a = (a_1, \cdots, a_{2^n}) \in E_n$ with convention that $a_0 = 0$.

Therefore, by the contraction principle, we have

Lemma 11.3. *For each n, $\{\mu_n^\varepsilon : \varepsilon > 0\}$ satisfies the large deviation principle on $\mathbb{W}^d$ with a rate function*

$$I_n'(x) = \inf\{I_n(a) : a \in E_n \text{ s. t. } g_n(a) = x\}. \tag{11.9}$$

Note that for any $x \in g_n^{-1}(E_n)$ there is exactly one solution $a \in E_n$ to the equation $g_n(a) = x$, therefore

$$I_n'(x) = \frac{1}{2}\sum_{k=1}^{2^n} 2^{-n}\left|x_{t_n^k} - x_{t_n^{k-1}}\right|^2 \tag{11.10}$$

for $x \in g_n^{-1}(E_n)$, and $I_n'(x) = \infty$ on $\mathbb{W}^d \setminus g_n^{-1}(E_n)$.

Let us note that

$$||x||_{H^1}^2 = \sum_{k=1}^{2^n} 2^{-n}\left|x_{t_n^k} - x_{t_n^{k-1}}\right|^2, \quad \forall x \in g_n^{-1}(E_n), \tag{11.11}$$

so that

$$I_n'(x) = \begin{cases} \frac{1}{2}||x||_{H^1}^2 & \forall x \in g_n^{-1}(E_n) \\ \infty & \text{otherwise.} \end{cases} \tag{11.12}$$

Since $x \in g_n^{-1}(E_n)$ if and only if there is exactly one solution $a \in E_n$ to the equation $g_n(a) = x$, we thus can rewrite

$$I_n'(x) = \inf\{I(h) : h \in \mathbb{W}^d \text{ s. t. } f_n(h) = x\} \tag{11.13}$$

where

$$I(x) = \begin{cases} \frac{1}{2}||x||_{H^1}^2 & \text{if } x \in \mathbb{H}, \\ \infty & \text{if } x \notin \mathbb{H}. \end{cases} \tag{11.14}$$

Therefore we may rewrite the previous lemma in terms of mappings f_n as the following.

Lemma 11.4. *For each n, $\{\mu_n^\varepsilon : \varepsilon > 0\}$ satisfies the large deviation principle on $\mathbb{W}^d$ with a rate function I_n' defined by (11.13).*

In order to apply the generalized contraction principle, we need to study the rate function given by (11.14).

Lemma 11.5. *The function defined by (11.14) is a good rate function on $\mathbb{W}^d$.*

Proof. Let $I_c = \{w : I(w) \le c\}$. Then for each $c \ge 0$, $I_c \subset \mathbb{H}$, and for any $w \in I_c$

$$\begin{aligned} |w_t - w_s| &= \left|\int_s^t \dot{w}_r dr\right| \\ &\le \sqrt{2c}\sqrt{t-s} \end{aligned}$$

so that I_c is bounded in $\mathbb{W}^d$, and is equi-continuous, therefore, according to Ascoli-Arzelà's theorem, I_c is pre-compact. We need to show I_c is closed in $\mathbb{W}^d$. To this end we show that for any $w \in \mathbb{W}^d$

$$I(w) = \sup_n I_n(w) \tag{11.15}$$

where

$$\begin{aligned} I_n(w) &= \frac{1}{2}||f_n(w)||^2_{H^1} \\ &= \frac{1}{2}\sum_{k=1}^{2^n} 2^{-n}\left|w_{t_n^k} - x_{t_n^{k-1}}\right|^2, \end{aligned}$$

since each I_n is continuous on E, I is lower-semi continuous, so that I_c is closed in E. Thus we only need to prove (11.15).

Let $w \in E$ and $w^{(n)} = f_n(w)$. Suppose $\sup_n I_n(w) < \infty$, then

$$\sup_n \int_0^1 \left|\dot{w}_t^{(n)}\right|^2 dt < \infty,$$

and therefore, for any $u \in C_0^\infty([0,1])$

$$\begin{aligned} \left|\int_0^1 \dot{u}_t w_t dt\right| &= \lim_{n\to\infty}\left|\int_0^1 \dot{u}_t w_t^{(n)} dt\right| \\ &= \lim_{n\to\infty}\left|\int_0^1 u_t \dot{w}_t^{(n)} dt\right| \\ &\leq \sqrt{\sup_n \int_0^1 \left|\dot{w}_t^{(n)}\right|^2 dt}\sqrt{\int_0^1 |u_t|^2 dt} \end{aligned}$$

so that $\dot{w} \in L^2([0,1])$, that is, $w \in H^1([0,T])$, and $I(w) < \infty$. Finally we prove (11.15) for $w \in H^1([0,T])$. By the fundamental theorem in calculus

$$\begin{aligned} \left|w_{t_n^k} - w_{t_n^{k-1}}\right|^2 &= \left|\int_{t_n^{k-1}}^{t_n^k} \dot{w}_t dt\right|^2 \\ &\leq 2^{-n}\int_{t_n^{k-1}}^{t_n^k} |\dot{x}_t|^2 dt \end{aligned}$$

it follows thus that

$$I_n(w) \leq \frac{1}{2}\int_0^1 |\dot{w}_t|^2 dt = I(w)\,.$$

On the other hand, by triangle inequality

$$\begin{aligned} \left|I_n(w)^{1/2} - I_n(\tilde{w})^{1/2}\right| &\leq I_n(w-\tilde{w})^{1/2} \\ &\leq \sqrt{I(w-\tilde{w})} \end{aligned}$$

so that we only need to show (11.15) for $w \in C^\infty([0,1])$. In this case we have

$$\begin{aligned} I(w) &\geq \sup_n I_n(w) \\ &\geq \frac{1}{2} \lim_{n\to\infty} \left\{ \sum_{l=1}^{2^{-n}} 2^{-n} \left| w_{t_n^k} - w_{t_n^{k-1}} \right|^2 \right\} \\ &= \lim_{n\to\infty} \int_0^1 |\dot{w}_t^{(n)}|^2 dt \\ &= I(w) . \end{aligned}$$

We have thus completed the proof. □

Lemma 11.6. *For $n \in \mathbb{N}$ and $\delta > 0$*

$$\mu_\varepsilon \left[||x - f_n(x)||_{\mathbb{W}^d} \geq \delta \right] \leq 2^n \exp\left\{ -\frac{2^n \delta^2}{8\varepsilon} \right\}. \tag{11.16}$$

In particular

$$\lim_{n\to\infty} \limsup_{\varepsilon \downarrow 0} \varepsilon \log \mu_\varepsilon \left[||x - f_n(x)||_{\mathbb{W}^d} \geq \delta \right] = -\infty. \tag{11.17}$$

Proof. Let $B_n(\delta) = \{||x - f_n(x)||_E \geq \delta\}$. If $t_n^k \leq t \leq t_n^{k+1}$, then

$$\begin{aligned} |x_t - f_n(x)_t| &= \left| x_t - x_{t_n^k} + 2^n \left(t - t_n^k \right) \left(x_{t_n^{k+1}} - x_{t_n^k} \right) \right| \\ &\leq 2 \sup_{s \leq \frac{1}{2^n}} \left| x_{t_n^k + s} - x_{t_n^k} \right| \end{aligned}$$

so that

$$\begin{aligned} \mu_\varepsilon(B_n(\delta)) &= \mu_\varepsilon \left\{ \sup_k \sup_{t_n^k \leq t \leq t_n^{k+1}} |x_t - f_n(x)_t| \geq \delta \right\} \\ &\leq \mu_\varepsilon \left\{ \sup_k \sup_{t \leq \frac{1}{2^n}} \left| x_{t_n^k + t} - x_{t_n^k} \right| \geq \frac{\delta}{2} \right\} \\ &= \mu \left\{ \sup_k \sup_{t \leq \frac{1}{2^n}} \left| x_{t_n^k + t} - x_{t_n^k} \right| \geq \frac{\delta}{2\sqrt{\varepsilon}} \right\} \\ &\leq \sum_{k=0}^{2^n - 1} \mu \left\{ \sup_{s \leq \frac{1}{2^n}} \left| x_{t_n^k + s} - x_{t_n^k} \right| \geq \frac{\delta}{2\sqrt{\varepsilon}} \right\} \\ &= 2^n \mu \left\{ \sup_{s \leq \frac{1}{2^n}} |x_s| \geq \frac{\delta}{2\sqrt{\varepsilon}} \right\} . \end{aligned}$$

Now (11.16) follows from the Gaussian tail estimate. □

Lemma 11.7. *f_n converges uniformly to the identity mapping on any level set $I_c = \{w : w \in \mathbb{H}$ s.t. $I(w) \leq c\}$.*

Proof. For any $w \in I_c$ and n we have

$$\begin{aligned}
||f_n(w) - w|| &= \max_{1\le k\le 2^n} \sup_{t\in[t_n^{k-1}, t_n^k]} |w_t - f_n(w)_t| \\
&= \max_{1\le k\le 2^n} \sup_{t\in[t_n^{k-1}, t_n^k]} |w_t - f_n(w)_t| \\
&\le 2 \max_{1\le k\le 2^n} \sup_{s\le \frac{1}{2^n}} \left| w_{t_n^{k-1}+s} - w_{t_n^{k-1}} \right| \\
&= 2 \max_{1\le k\le 2^n} \sup_{s\le \frac{1}{2^n}} \left| \int_{t_n^{k-1}}^{t_n^{k-1}+s} \dot{w}_s ds \right| \\
&\le 2\sqrt{\frac{1}{2^n}} \max_{1\le k\le 2^n} \sup_{s\le \frac{1}{2^n}} \sqrt{\int_{t_n^{k-1}}^{t_n^{k-1}+s} |\dot{w}_s|^2 ds} \\
&\le 2\sqrt{2c}\sqrt{\frac{1}{2^n}}
\end{aligned}$$

so that $f_n(w)$ goes to w uniformly on I_c. □

Now Schilder's LDP follows from the generalized contraction principle, Theorem 10.7.

References

[1] R. Azencott, Grandes déviations et applications, *Ecoles s'Été de Probabilités de Saint-Flour* VIII-1978. Lecture Notes in Math. 774, Springer, pp. 1-176 (1980). English translation is available in "Large Deviations at Saint-Flour", R. Azencott, M. I. Freidlin and S. R. S. Varadhan, Springer (2013).

[2] H. Cramér, Sur un nouveau théorème-limite de la théorie des probabilités, Actualités Scientifiques et Industrielles 736 (1938), 5-23, Colloque consacré à la théorie des probabilités, Vol. 3, Hermann, Paris.

[3] M. D. Donsker and S. R. S. Varadhan, Asymptotic evaluation of certain Wiener integrals for large time, Functional Integration and its Applications, Proceedings of the International Conference held at the Cumberland Lodge, Windsor Great Park, London, in April 1974. A. M. Arthurs, Ed., Clarendon, Oxford, 1975.

[4] M. D. Donsker and S. R. S. Varadhan, Asymptotic evaluation of certain Markov process expectations for large time, I. Comm. Pure Appl. Math. Vol. XXVIII, pp. 1-47 (1975).

[5] M. D. Donsker and S. R. S. Varadhan, Asymptotic evaluation of certain Markov process expectations for large time, II. Comm. Pure Appl. Math. Vol. XXVIII, pp. 279–301 (1975).

[6] M. D. Donsker and S. R. S. Varadhan, The asymptotics of the Wiener sausage, Comm. Pure Appl. Math. Vol. XXVIII, pp. 525-565 (1975).

[7] M. D. Donsker and S. R. S. Varadhan, The asymptotics of the Wiener sausage, Comm. Pure Appl. Math. Vol. XXIX, pp. 389-461 (1976).

[8] M. D. Donsker and S. R. S. Varadhan, On laws of iterated logarithms for local times. Comm. Pure Appl. Math. Vol. XXX, pp. 707-753 (1977).

[9] M. D. Donsker and S. R. S. Varadhan, Asymptotic evaluation of certain Markov process expectations for large time, IV. Comm. Pure Appl. Math. Vol. XXXVI, pp. 183-212 (1983).

[10] M. D. Donsker and S. R. S. Varadhan, Asymptotics for the Polaron. Comm. Pure Appl. Math. Vol. XXXVI, pp. 505-528 (1983).

[11] M. D. Donsker and S. R. S. Varadhan, Large deviations for stationary Gaussian processes. Comm. Math. Physics 97, pp. 187-210 (1985).

[12] Dudley, R. M., "Distances of probability measures and random variables." *Ann. Math. Statist.* 39, 1563-1572 (1968).

[13] Dudley, R.M., *Real Analysis and Probability*. Cambridge University Press (2002).

[14] Ethier S. N. and T. G. Kurtz, Markov Processes – Characterization and Convergence. John Wiley & Sons (1986).

[15] R. S. Ellis. Entropy, Large Deviations and Statistical Mechanics. Springer-Verlag, New York (1985).

[16] N. Ikeda and S. Watanabe, Stochastic Differential Equations and Diffusion Processes, North-Holland Publishing Company, Amsterdam, Oxford, New York (1981).

[17] Itô,K. and H. P. McKean, Jr., Diffusion Processes and their Sample Paths, Springer (1965, 1974).

[18] Kac, M., On some connections between probability theory and differential and integral equations, Proceedings of the Second Berkeley Symposium, 1950, pp. 189-215.

[19] Kurtz, T. G., "Random Time Changes and Convergence in Distribution Under the Meyer-Zheng Conditions," *The Annals of Probability*, Vol. 19, No. 3 (Jul., 1991), pp. 1010-1034.

[20] Kurtz, T. G. and P. Protter, "Weak Limit Theorems for Stochastic Integrals and Stochastic Differential Equations," *The Annals of Probability*, Vol. 19, No. 3 (Jul., 1991), pp. 1035-1070.

[21] T. Lyons and Z. Qian, System Control and Rough Paths, Oxford Science Publications, Clarendon Press, Oxford (2002).

[22] Meyer, P. A. and W. A. Zheng, "Tightness criteria for laws of semimartingales," Annales de l'I. H. P., section B, tome 20 (4) (1984), p. 353-372.

[23] Skorokhod, A. V., "Limit Theorems for Stochastic Processes", *Teor. Veroyatnost. i Primenen.,* 1:3 (1956), 289–319; *Theory Probab. Appl.*, 1:3 (1956), 261–290.

[24] Stroock, D. W., *An Introduction to the Theory of Large Deviations*, Springer-Verlag, New York (1984).

[25] Stroock, D. W and S. R. S. Varadhan, *Multidimensional Diffusion Processes*, Springer (1979, 1997).

[26] K. Yosida: Functional Analysis. 6th Edition. Springer-Verlag.

[27] S. R. S. Varadhan, Asymptotic probabilities and differential equations. Comm. Pure Appl. Math., Vol. XIX, No. 3, 1966, pp. 261-286.

[28] S. R. S. Varadhan, Large deviations and applications. Ecoles s'Été de Probabilités de Saint-Flour -. Lecture Notes in Math. , Springer, pp. 1-49 (1988).

[29] S. R. S. Varadhan, Large Deviations and Applications. SIAM, Philadelphia (1984).

[30] Villani, Cédric, *Optimal Transport – Old and New*. Springer (2009).