AN ABSTRACT OF THE THESIS OF

Dae-Sik Hwang for the degree of Doctor of Philosophy in Mathematics presented on March 12, 1991

Title : Large Deviation Principles for Random Measures

Abstract approved : Signature redacted for privacy.

Robert M. Burton

Large deviation theory has experienced much development and interest in the last two decades. A large deviation principle is the exponential decay of the probability of increasingly rare events and the computation of a rate or entropy function which measures the rate of decay. Within the probability literature there has been much use made of these rates in diverse applications. These large deviation principles have been discovered for independent and identically distributed random variables, as well as random vectors and these have been extended to some cases of weak dependence.

In this thesis we prove large deviation principles for finite dimensional distributions of scaling limits of random measures. Functional approaches to large deviation theory using test functions as dual objects to random measures are also developed. These results are applied to some important classes of models, in particular Poisson point processes, Poisson center cluster processes and doubly stochastic point processes.

Large Deviation Principles
for Random Measures

by

Dae-Sik Hwang

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Completed March 12, 1991
Commencement June 1991

APPROVED :

Signature redacted for privacy.

Professor of Mathematics in charge of major

Signature redacted for privacy.

Chair of department of Mathematics

Signature redacted for privacy.

Dean of Graduate School

Date thesis is presented March 12, 1991

Typed by the researcher Dae-Sik Hwang

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, professor Robert M. Burton for his guidance, helpful discussion, encouragement and support which made this dissertation possible.

I also thank my committee members, professor Timothy A. Budd, Larry Chen, Mina E. Ossiander, Harold R. Parks and Edward C. Waymire for their support and cooperation during my graduate study.

Also, thanks are due to the Headquarters of Korea Air Force and Korea Air Force Academy which gave me an opportunity to study abroad.

Additional thanks go to my fellow graduate students, John A. Crow and Marc R. Goulet for their friendship and kindness.

With sincere appreciation, I thank my mother-in-law, brothers, sisters and brothers-in-law for their concern and unending encouragement.

I would like to express special appreciation to my lovely two daughters, Hyelyung (Haley in school) and Injung for being well in school and at home.

In particular, my wife Jiyeon has contributed much to the successful completion of my graduate study with her warm love, encouragement and endless patience since studying from Seoul National University to Oregon State University.

TABLE OF CONTENTS

LARGE DEVIATION PRINCIPLES FOR RANDOM MEASURES

1. INTRODUCTION

1.1. Introduction

Cramér presented the first large deviation theorem at a probability symposium in 1937. Since 1937, this theory has undergone an extensive development and this original work was extended in various directions. There have been many developments in the theory of large deviations over the last two decades.

Primary and most significant is the work of M.D. Donsker and S.R.S. Varadhan [10, 11] who have developed a powerful machinery in a series of papers to deal with many old and new problems in probability where precise estimates of large deviation probabilities play an important role. Gärtner (1977 [16]) and Ellis (1984 [13]) have developed useful and surprising generalizations with assumptions about the dependence of the random variables and the moment generating functions. Also large deviation principles (L.D.P.) have found many applications in statistics (Groeneboom [17]), statistical mechanics (Ellis [12], Lanford [23]) and in stochastic processes [10, 11]. Consequently, much attention has been devoted to establishing such L.D.P. for a wide variety of processes.

Up to now, the most important asymptotic theorems which we have met have been the law of large numbers, weak and strong and the central limit theorem.

Let us consider a sequence of random variables $\hat{S}_1, \hat{S}_2, \ldots, \hat{S}_n, \ldots$, which converges in probability to a real constant number m. This means that $P(|\hat{S}_n - m| > \epsilon) \to 0$ as $n \to \infty$ for every $\epsilon > 0$. These variables are said to satisfy a large deviation principle if this convergence is exponentially rapid.

Let $X_1, X_2, \ldots$ be a sequence of real-valued independent and identically distributed (i.i.d.) random variables with $E[X_1] = m$.

Let

(1.1) $\hat{S}_n = \frac{S_n}{n}$, where $S_n = X_1 + \ldots + X_n$.

The law of large numbers says that for every $\epsilon > 0$,

(1.2) $P\Big(|\hat{S}_n - m| > \epsilon\Big) \to 0$ in probability or a.s. as $n \to \infty$.

Now, since $P(|\hat{S}_n - m| > \epsilon)$ converges to zero, for each $\epsilon > 0$, we can also observe the rate of convergence to 0 of such probability in (1.2). It is often the case that it converges exponentially rapidly as $n \to \infty$. That is,

(1.3) $P(|\hat{S}_n - m| > \epsilon) \approx F(\epsilon, m, n)e^{-n \cdot I(m, \epsilon)}$,

where $I(\cdot, \cdot)$ is a positive quantity and $F(\cdot, \cdot, \cdot)$ is a slowly varying function of n.

Roughly speaking, if such type of (1.3) is obtained, we say that $\{\hat{S}_n\}$ satisfies a large deviation principle. Moreover, since the law of large numbers says convergence of some probability and a large deviation principle is related to the rate of convergence, a large deviation principle can rightly be considered as an extension or generalization of the law of large numbers.

Also if the moment generating function, $M_n(t) = E[e^{tS_n}]$, is finite for each real number t in $\mathbb{R}$, then it turns out that the probability in (1.3) converges exponentially rapidly to zero as $n \to \infty$.

In this dissertation, in Chapter II, we present some preliminaries and backgrounds of ideas which will be used throughout this thesis later.

Suppose that $X_1, X_2, \ldots$ be a sequence of random vectors taking values in $\mathbb{R}^d$, $d \in \{1, 2, \ldots\}$. Let $S_n = X_1 + \ldots + X_n$ be the partial sums. Ellis proved large deviation theorems for $\frac{S_n}{n}$ of a general class of random vectors in his paper [13].

Also we assume that X is Poisson center cluster process with centers U, Poisson process with intensity α, and member V with finite total expected mass ζ. Note that $E[X(A)] = \alpha\zeta|A|$. The ergodic theorem says that $\frac{X[0, \lambda]}{\lambda}$ converges to $\alpha\zeta$. Burton and Dehling proved the large deviation principle for $\frac{X[0, \lambda]}{\lambda}$. That is,

$$(1.4) \qquad \lim_{\lambda \to \infty} \frac{-1}{\lambda} \cdot \log P\left(\left|\frac{X[0, \lambda]}{\lambda} - \alpha\zeta\right| > x\right) \geq I(x), \text{ where } I(x) \text{ is a rate function.}$$

In Chapter III, we will extend the above property (1.4) to the large deviation principle for the finite dimensional distributions of random measures. That is, let X be random measure on $\mathbb{R}^d$ and $A_1, A_2, \ldots, A_n$ disjoint and bounded Borel subsets of $\mathbb{R}^d$. Then the random vector obtained by rescaling the random measure

$$(1.5) \qquad \frac{\left(X_\lambda(A_1), \ldots, X_\lambda(A_n)\right)}{\lambda^d} \text{ satisfies a large deviation principle,}$$

where $X_\lambda(A) = X(\lambda A)$ for a Borel subset A in $\mathbb{R}^d$.

This reduces to (1.4) if we take $n = 1$, $d = 1$ and A = [0, 1] in (1.5). Also we apply this extension to Poisson point processes, Poisson center cluster random measures and doubly stochastic processes.

In Chapter IV, we introduce a functional approach to large deviation theory for random measures. Here we use an appropriate space of test functions considered as "dual" objects to random measures.

We denote the collection of Radon measures on $\mathbb{R}$ by $M(\mathbb{R})$, which is endowed with the weak topology. Let M[0, L] be the restriction of measures on $M(\mathbb{R})$ to $[0, L] \subseteq \mathbb{R}$. Let X be a random measure in $M(\mathbb{R})$ and X^L a random measure in M[0, L] obtained by restricting X to [0, L]. Denote $\tilde{X}_\lambda = \left(X_\lambda\right)^L$. We consider the following properties.

(*Large Deviation Upper Bound*) ; For any closed subset F in M[0, L],

$$\limsup_{\lambda\to\infty} \frac{1}{\lambda}\cdot\log P\left(\frac{\tilde{X}_\lambda}{\lambda} \in F\right) \le -I_L(F).$$

(*Large Deviation Lower Bound*) ; For any open subset G in M[0, L],

$$\liminf_{\lambda\to\infty} \frac{1}{\lambda}\cdot\log P\left(\frac{\tilde{X}_\lambda}{\lambda} \in G\right) \ge -I_L(G).$$

Above, we denote $I_L(A) = \inf_{\mu\in A} I_L(\mu)$ for a Borel subset A and

$$I_L(\mu) = \sup_{f\in C[0,L]}\left\{\mu(f) - \lim_{\lambda\to\infty} \frac{1}{\lambda}\cdot\log E[\exp(\tilde{X}_\lambda(f))]\right\}.$$

We give conditions on X that guarantee the upper bound and lower bound properties. It is shown that the above classes of random measures satisfy these conditions.

1.2. Large Deviation for Bernoulli Random Variables

Let us consider the large deviation properties of independent Bernoulli random variables. Let a triple $(\Omega, \mathcal{A}, P)$ be a probability space with Bernoulli random variables X_k, which are assumed to be independent and identically distributed with

$$P(X_k = 1) = p \text{ and } P(X_k = 0) = 1 - p, \text{ where } k = 1, 2, \ldots, n.$$

It is natural to think of X_k as describing the result of an experiment at the k-th stage. Recall (1.1) for Bernoulli random variables X_n as follows.

Let

$$S_0 = 0 \quad \text{and} \quad S_n = X_1 + \ldots + X_n.$$

Then, $E[S_n] = np$ and consequently $E[\frac{S_n}{n}] = p$. That is to say, the mean value of the frequency of success coincides with the probability p of success. Here the question is how much the frequency $\frac{S_n}{n}$ of success differs from its probability p. That is to say,

$$(1.6) \qquad |\frac{S_n}{n} - p| > \epsilon \text{ for every real number } \epsilon > 0.$$

It is natural to expect that the total probability of the events for which (1.6) will also be small for sufficiently large n.

It is meaningful for us to try to estimate the probability of the event such that $\left\{|\frac{S_n}{n} - p| > \epsilon\right\}$. For this, we have already known that

$$(1.7) \qquad P\left(|\frac{S_n}{n} - p| \geq \epsilon\right) \leq \frac{\mathrm{Var}\left(\frac{S_n}{n}\right)}{\epsilon^2} \text{ by } \textit{Chebyshev's inequality}$$

$$= \frac{p(1-p)}{n\epsilon^2}$$

$$(1.7)' \qquad \leq \frac{1}{4n\epsilon^2}.$$

Therefore, we see, for sufficiently large n that there is small probability that the frequency $\frac{S_n}{n}$ of success deviates from the probability p by more than ϵ.

If S_n is the total number of heads, then $\frac{S_n}{n}$ is the relative frequency of heads in n trials coin tossing. For sufficiently large n, it is reasonable to expect that $\frac{S_n}{n}$ will be close to p. The laws of large numbers make this statement mathematically precise. According to W.L.L.N. (Theorem 1.1), the probability that $\frac{S_n}{n}$ differs from p is very small as the number of trials is very large, that is, $\frac{S_n}{n} \to p$ in probability as $n \to \infty$. This may be proved by letting $n \to \infty$ in (1.7)$'$.

S.L.L.N. (Theorem 1.2) gives a stronger assertion, that is to say $\frac{S_n}{n} \to p$ with probability one as $n \to \infty$.

The classical limit theorems of probability describe the asymptotic behavior of the partial sums S_n defined in (1.1) suitably normalized.

There are the most basic three theorems as follows :

Now, first of all, let us consider the classical limit theorems. The asymptotic theories which we have met up to now depend upon the law of large numbers of the forms (Theorem 1.1 and 1.2).

Theorem 1.1 (*Weak Law of Large Numbers*) Let $\{X_k\}$ be a sequence of independent random variables with the same distribution. If the expected value $E[X_k] = m$ is finite, then for every $\epsilon > 0$,

$$(1.8) \qquad P\left(\left| \frac{X_1 + \ldots + X_n}{n} - m \right| > \epsilon \right) \to 0 \text{ as } n \to \infty.$$

This means that the probability that the sample average $\frac{S_n}{n}$ will differ from the expected value by less than ϵ tends to one as n goes to infinity.

Theorem 1.2 (*Strong Law of Large Numbers*) Let $\{X_k\}$ be a sequence of independent and identically distributed random variables. If the expected value $E[X_k] = m$ is finite, then, with probability one,

$$(1.9) \qquad \frac{X_1 + \ldots + X_n}{n} \to m \quad \text{as } n \to \infty.$$

Strong law of large numbers means that two notions of the sample average and the expected value coincide asymptotically. That is, for sufficiently large number n, the sample average of $X_1, \ldots, X_n$ approximates arbitrarily closely the common expected value $E[X_n]$. The importance of this fact stems from the observation that one can calculate sample averages.

On the other hand, the most important and well-known theorem of the classical limit theorems of probability theory is the central limit theorem. This result states that the sum of a large number of independent and identically distributed random variables with zero mean and finite variance is approximately standard normal (or Gaussian), no matter what the common distribution of those random variables is.

Theorem 1.3 (*Central Limit Theorem*) Let $\{X_k\}$ be a sequence of independent and identically distributed random variables. If the expected value $E[X_k] = m$ and the variance $Var(X_k) = \sigma^2$ are both finite, then for every real number a,

$$(1.10) \qquad \lim_{n\to\infty} P\left(\frac{X_1 + \ldots + X_n - nm}{n} \le a\right) = \frac{1}{2}\int_{-\infty}^{a} e^{-\frac{t^2}{2}}\, dt.$$

Here, for a sequence $\{X_k\}$ of Bernoulli random variables and every $\epsilon > 0$, we have known that

$$(1.11) \qquad P\left(|\tfrac{S_n}{n} - p| > \epsilon\right) \to 0 \text{ as } n \to \infty \text{ by L.L.N..}$$

So, we want to estimate the rate of convergence to zero of such probability. This is an example of a large deviation probability.

Now let us consider a coin tossing random variable X_k defined as in the above. Without doubt, the frequency of winning will be close to p. This closeness of $\frac{S_n}{n}$ and p appears first of all through the formulae

$$(1.12) \qquad E[\tfrac{S_n}{n}] = p \quad \text{and} \quad Var\left(\tfrac{S_n}{n}\right) = \frac{p(1-p)}{n}.$$

By recalling (1.7)$'$, we obtain that

$$(1.13) \qquad P\left(|\tfrac{S_n}{n} - p| > \epsilon\right) \to 0 \quad \text{as} \quad n \to \infty.$$

This result is well-known as the weak law of large numbers. Now let us observe the rate of convergence to zero of such probability as follows :

<u>Case of $t \geq 0$</u> :

Define

$$\Phi(t) = E[e^{tX_1}].$$

So,

$$\Phi(t) = 1 - p + pe^{t} \quad \text{and} \quad E[e^{tS_n}] = [\Phi(t)]^{n}.$$

Consider the probability $P\left(\frac{S_n}{n} \geq a\right)$ for every real number a ;

$$P\left(\tfrac{S_n}{n} \geq a\right) = P\left(e^{t(\frac{S_n}{n} - a)} \geq 1\right)$$

$$\leq E[e^{t(\frac{S_n}{n} - a)}] \text{ by } \textit{Markov's inequality}$$

$$= [\Phi(\tfrac{t}{n})]^{n} e^{-at}.$$

Here, replacing t/n by t we can derive that

$$(1.14) \qquad P\left(\tfrac{S_n}{n} \geq a\right) \leq e^{-n\left(at - \operatorname{Log}\Phi(t)\right)}.$$

Since the above inequality (1.14) holds for every $t > 0$, we have that

$$(1.15) \qquad P\left(\tfrac{S_n}{n} \geq a\right) \leq e^{-n\left\{\sup_{t>0}\left(at - \operatorname{Log}\Phi(t)\right)\right\}}.$$

Letting $f(t) = at - \text{Log}\Phi(t)$, $f(t)$ is concave downward. Moreover, since $f(0) = 0$ and $f'(0) = a - p$, $f(t)$ has a strictly positive maximum $r(a)$ attained for $t > 0$ according to whether we assume $a > p$.

<u>Case of $t \leq 0$</u> :

Likewise in (1.15), we have that

(1.16) $$P\left(\frac{S_n}{n} \leq a\right) \leq e^{-n\left\{\sup\limits_{t<0}\left(at - \text{Log}\Phi(t)\right)\right\}}.$$

Also f(t) has a strictly positive maximum r(a) attained for $t < 0$ accordingly to whether we assume $a < p$.

Up to now, we have proved the following theorem for large deviations.

<u>Theorem 1.4</u> (<u>*Large Deviation Theorem*</u>) Let $\epsilon > 0$. Let $r(p,\epsilon)$ and $r(p,-\epsilon)$ be the strictly positive numbers defined above. Then

(1.17) $$P\left(\frac{S_n}{n} \geq p + \epsilon\right) \leq e^{-n\cdot r(p,\epsilon)},$$

and

(1.18) $$P\left(\frac{S_n}{n} \leq p - \epsilon\right) \leq e^{-n\cdot r(p,-\epsilon)}.$$

Moreover, letting $R(p,\epsilon) = \inf\left\{r(p,\epsilon),\ r(p,-\epsilon)\right\}$, we then have

(1.19) $$P\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) \leq 2e^{-n\cdot R(p,\epsilon)}.$$

We say that $\frac{S_n}{n}$ converges to p at an exponential rate $R(p,\epsilon)$.

2. MATHEMATICAL PRELIMINARIES

2.1. Preliminaries of Random Measure and Point Process

Let S be a locally compact Polish (i.e. there exists some separable and complete metrization ρ of S) space. Let $\mathcal{S}$ be the Borel σ-field in S and $\mathcal{S}_0$ the bounded sets in $\mathcal{S}$ (i.e. sets with compact closures). $C_C(S)$ will denote the collection of all continuous functions $f : S \to \mathbb{R}$ with compact support. A measure μ on $(S, \mathcal{S})$ is called a Radon measure if $\mu(B) < \infty$ for all bounded subsets $B \in \mathcal{S}_0$. Let $\mathfrak{M}(S)$ be the collection of all Radon measures on $(S, \mathcal{S})$ and $\mathfrak{N}(S)$ the subcollection of all measures $\mu \in \mathfrak{M}(S)$ satisfying $\mu(B) \in \mathbb{Z}^+ = \{0, 1, \ldots\}$ for $B \in \mathcal{S}$ (we write $\mathfrak{M} = \mathfrak{M}(S)$ and $\mathfrak{N} = \mathfrak{N}(S)$ for notational simplicity). Let $\mathcal{M}$ be the σ-algebra in $\mathfrak{M}$ generated by sets of the form $\{\mu \in \mathfrak{M} : \mu(B) < r\}$ for $B \in \mathcal{S}_0$ and $0 \leq r < \infty$. Likewise, $\mathfrak{N}$ is a closed subspace of $\mathfrak{M}$. So, note that $\mathcal{N} \in \mathcal{M}$ and $\mathcal{N}$ is the restriction of $\mathcal{M}$ to $\mathfrak{N}$.

Given $\mu \in \mathfrak{M}$ and $f \in C_C(S)$, we can define the integral $\mu(f)$ of f with respect to μ by

$$\mu(f) = \int_S f(s)\mu(ds) = \int f d\mu. \tag{2.1}$$

In particular, we introduce the above properties in the special case where $S = \mathbb{R}^d$. Let $\mathbb{R}^d$ be d-dimensional Euclidean space. Let $\mathfrak{B}^d$ be the Borel σ-field in $\mathbb{R}^d$ and $\mathfrak{B}_0{}^d$ the subset of $\mathfrak{B}^d$ consisting of bounded sets (i.e. sets with compact closures). A measure μ on $(\mathbb{R}^d, \mathfrak{B}^d)$ is called a Radon measure if $\mu(B) < \infty$ for all bounded subsets $B \in \mathfrak{B}_0{}^d$ (see Kallenberg [19] or Karr [21] for a detailed exposition).

From now on, we take $(\mathfrak{M}, \mathcal{M})$ on $(\mathbb{R}^d, \mathfrak{B}^d)$ instead of on $(S, \mathcal{S})$. We also take $\mathfrak{N}$ which is a closed subspace of such $\mathfrak{M}$. The formula (2.1) is rewritten as follows :

Given $\mu \in \mathfrak{M}$ and $f \in C_C(\mathbb{R}^d)$, we denote

$$\mu(f) = \int_{\mathbb{R}^d} f(x)\mu(dx) = \int f d\mu. \tag{2.2}$$

The vague topology on $\mathfrak{M}$, that is, the coarsest topology making the mapping $\mu \to \mu(f)$, $f \in \mathfrak{F}_c$, continuous, is the topology generated by the class of all finite intersections of subsets of $\mathfrak{M}$ of the form $\{\mu \in \mathfrak{M} : s \leq \int f d\mu \leq t\}$ as subbase for $s, t \in \mathbb{R}$ and every function $f \in \mathfrak{F}_c$, where

$$\mathfrak{F}_c = \Big\{f : \mathbb{R}^d \to [0, \infty] \text{ ; f is continuous function with compact support}\Big\}.$$

Likewise, the vague topology on $\mathfrak{N}$ is also defined similarly.

Let $(\Omega, \mathcal{A}, P)$ be a probability space.

Definition 2.1 A measurable mapping X from $(\Omega, \mathcal{A}, P)$ into $(\mathfrak{M}, \mathcal{M})$ is called a random measure. The induced measure $P_X = P \circ X^{-1}$ on $(\mathfrak{M}, \mathcal{M})$ is called the distribution of the random measure X.

We distinguish the set of counting measure by

$$\mathfrak{N} = \Big\{\mu \in \mathfrak{M} : \mu(A) \in \mathbb{Z}^+ \text{ for each } A \in \mathfrak{B}^d\Big\}, \text{ where } \mathbb{Z}^+ = \{0, 1, 2, \ldots\}.$$

Definition 2.2 A measurable mapping X from a probability space $(\Omega, \mathcal{A}, P)$ into $(\mathfrak{N}, \mathcal{N})$ is called a point random field (point process). The induced measure P_X is called the distribution of the point process X.

Example 2.3 The most important point process is a Poisson point process. The distribution of the Poisson point process X with intensity $\Lambda \in \mathfrak{M}$ is determined by

(1) For every subset $A \in \mathfrak{B}^d$, X(A) has a Poisson distribution with mean $\Lambda(A)$, that is,

$$P\Big\{X(A) = k\Big\} = e^{-\Lambda(A)}\frac{\big(\Lambda(A)\big)^k}{k!}, \quad k = 0, 1, \ldots .$$

(2) If $A_1, A_2, \ldots$ are mutually disjoint bounded subsets of $\mathfrak{B}^d$, then the random variables $X(A_1), X(A_2), \ldots$ are mutually independent.

That is , a Poisson point process X with intensity Λ is a point process such that

$$P\Big\{X(A_1) = k_1, \ldots , X(A_n) = k_n\Big\} = \frac{\big(\Lambda(A_1)\big)^{k_1}}{k_1!}e^{-\Lambda(A_1)} \cdots \frac{\big(\Lambda(A_n)\big)^{k_n}}{k_n!}e^{-\Lambda(A_n)}$$

for all $k_1, \ldots , k_n \in \mathbb{Z}^+$ and $A_1, \ldots , A_n$, bounded and disjoint subsets of $\mathfrak{B}^d$.

When the intensity Radon measure Λ is taken to be a multiple of Lebesgue measure, that is, $\Lambda(A) = \alpha|A|$ for some $0 < \alpha < \infty$, we obtain a stationary Poisson point process which we will consider later in Remark 3.4.

Theorem 2.4 (*Law of Rare Events*) Let X be a point process satisfying the following conditions such that

(a) (*completely random*) $X(A_1), \ldots , X(A_n)$ are independent for disjoint Borel subsets $A_1, \ldots , A_n$,

(b) (*without multiple occurrences*) $P\big(X(A) \geq 2\big) = o(|A|)$ as $|A| \to 0$,

(c) (*homogeneous*) $P\big(X(A) = 1\big) = \alpha|A| + o(|A|)$ as $|A| \to 0$,

where $0 < \alpha \leq \infty$ and $|A|$ is the Lebesque measure of A.

Then X is a stationary Poisson point process with intensity $\alpha|A|$.

2.2. Generating Functionals

We begin by considering the finite dimensional case of the random measure X introduced in Section 2.1. in order to discover what the natural extension should be. We seek then to extend the notion of multivariate probability generating functions and characteristic functions to the more general setting of point processes. For complete discussion of probability generating functionals for point processes, the reader is referred to Fisher's papers [15], Gupta and Waymire's paper [18], and Westcott's paper [32].

Let $\mathbb{R}^d$ be d-dimensional Euclidean space and $\mathfrak{B}^d$ denote the collection of Borel subsets of $\mathbb{R}^d$. Let X be a point process. Also let $A_1, \ldots, A_n$ be bounded and disjoint subsets of $\mathbb{R}^d$. Then the joint distribution of the random vector $\big(X(A_1), \ldots, X(A_n)\big)$ is uniquely determined by its multivariate probability generating function

$$(2.3)\qquad g(s_1, \ldots, s_n) = E[s_1^{X(A_1)} \cdot \ldots \cdot s_n^{X(A_n)}],\ 0 \le s_i \le 1 \text{ and } 1 \le i \le n.$$

The formula (2.3) can be expressed as

$$(2.4)\qquad g(s_1, \ldots, s_n) = E[e^{\log\left(s_1^{X(A_1)} \cdot \ldots \cdot s_n^{X(A_n)}\right)}]$$

$$(2.5)\qquad = E[e^{\sum_{i=1}^{n} \log(s_i)X(A_i)}]$$

$$= E[e^{X\left(\sum_{i=1}^{n} \log(s_i)1_{A_i}\right)}].$$

Since X can be thought of as corresponding directly to a sequence of points $w = \{x_i\}$ in $\mathbb{R}^d$, it induces a counting measure $N(A)(w)$ = the random number of

$\{i : x_i \in A_i\}$ on $\mathbb{R}^d$. Thus, given a real-valued measurable function f on $\mathbb{R}^d$, we may define for each w the integral of f with respect to the process X as follows :

$$\int_{\mathbb{R}^d} f(x)dX(x) = \sum_i f(x_i). \tag{2.6}$$

Now, define the function ξ on $\mathbb{R}^d$ by

$$\xi(x) = \begin{cases} s & \text{for } x \in A_i,\ 1 \leq i \leq n \\ 1 & \text{for } x \notin \bigcup_{i=1}^{n} A_i \end{cases}. \tag{2.7}$$

Note that the finite dimensional distributions of X can be recovered by taking the above function ξ of the form (2.7). From (2.7), we have

$$\sum_{i=1}^{n} \log(s_i)X(A_i) = \int_{\mathbb{R}^d} \log \xi(x)dX(x). \tag{2.8}$$

Substituting (2.8) into (2.5),

$$g(s_1, \ldots, s_n) = E[\exp\Big(\int_{\mathbb{R}^d} \log \xi(x)dX(x)\Big)]. \tag{2.9}$$

Consequently, we can lead up to the probability generating functional G_X somewhat naturally through the above discussion.

Definition 2.5 Let $\mathcal{V}$ be the collection of arbitrary real-valued, measurable function ξ on $\mathbb{R}^d$ satisfying

(a) $0 \leq \xi(x) \leq 1$ for all $x \in \mathbb{R}^d$.

(b) $\xi(x) = 1$ on the complement of a bounded subset of $\mathbb{R}^d$.

Then the probability generating functional G_X corresponding to a point process X is defined by

$$G_X(\xi) = E[e^{X(\log(\xi))}] = E\left\{\exp\left(\int_{\mathbb{R}^d} \log \xi(x) dX(x)\right)\right\}, \quad \xi \in \mathcal{V}. \tag{2.10}$$

Theorem 2.6 Let w be the point configuration $\{x_i\}$ corresponding to X. The probability generating functional G_X of a point process X can be expressed as

$$G_X(\xi) = E[\prod_{x_i \in w} \xi(x_i)]. \tag{2.11}$$

Proof. $G_X(\xi) = E[\exp\left(\int_{\mathbb{R}^d} \log \xi(x) dX(x)\right)]$ from the formula (2.10)

$= E[\exp\left(\sum_{i=1}^{n} \log(s_i)\right)]$ by the formula (2.6)

$= E[\prod_{x_i \in w} e^{\log \xi(x_i)}] = E[\prod_{x_i \in w} \xi(x_i)]$.

Remark 2.7 Let $X_1(\cdot), \ldots, X_n(\cdot)$ be point processes. The superposition X_S of such n point processes means the total aggregation of all their points and we denote X_S by

$$X_S = \sum_{i=1}^{n} X_i(\cdot). \tag{2.12}$$

Theorem 2.8 Let $X_1(\cdot), \ldots, X_n(\cdot)$ be n independent point processes with probability generating functional $G_{X_1}, \ldots, G_{X_n}$, respectively. Then the probability generating functional $G_{X_S}(f) = G_{X_1}(f) \cdot \ldots \cdot G_{X_n}(f)$ for every $f \in \mathcal{V}$.

Theorem 2.9 Let X, $X_1(\cdot)$, $X_2(\cdot)$, . . . be a sequence of point processes with probability generating functional G_X, G_{X_1}, G_{X_2}, . . . , respectively. Then the sequence converges to a point process X if and only if $\lim_{n\to\infty} G_{X_n}(f) = G_X(f)$ for every $f \in \mathcal{V}$.

Definition 2.10 The characteristic functional C_X of a point process X is defined by

$$C_X(\phi) = E[\exp\Big(i\int_{\mathbb{R}^d} \phi(x)dX(x)\Big)]$$

for ϕ bounded, measurable and real-valued functions which have compact support.

Definition 2.11 The Laplace functional L_X of a point process X is defined by

$$L_X(\xi) = E[e^{-X(\xi)}] \text{ for all } \xi \in \mathcal{V}.$$

Note that the characteristic functional is a special case of the probability generating functional since we get $G_X(\xi) = G_X(e^{i\phi}) = C_X(\phi)$ if we take complex-valued functions ξ such that $\xi(x) = \exp\big(i\phi(x)\big)$. Likewise, we also have the relation $G_X(\xi) = L_X\big(-\log(\xi)\big)$.

Example 2.12 Let X be a Poisson point process with intensity $\Lambda \in \mathfrak{M}$. Then

(1) The probability generating functional of X is

$$G_X(\xi) = e^{\int\big(\xi(x) - 1\big)d\Lambda(x)}$$

and

(2) The characteristic functional of X is

$$C_X(\phi) = e^{\int\big(e^{i\phi(x)} - 1\big)d\Lambda(x)}.$$

2.3. Weak Convergence

In this section we discuss the general idea of weak convergence which is a generalization of convergence in distribution and briefly summarize some standard results in the theory of weak convergence of probability measures.

Let S be a metric space. The Borel σ-field $\mathcal{S}$ of S is the smallest σ-field generated by the open subsets of S.

First, we shall study weak convergence (or convergence in distribution) on probability measures on the σ-field $\mathcal{S}$.

To begin with, we refer to the notions of convergence of a sequence of measures on a metric space as follows.

Definition 2.13 Let μ and $\{\mu_n\}$ be finite measures on the metric space S. Then

(a) $\mu_n \to \mu$ weakly if $\int f d\mu_n \to \int f d\mu$ for all bounded continuous functions f on S.

(b) $\mu_n \to \mu$ vaguely if $\int f d\mu_n \to \int f d\mu$ for all bounded continuous functions f on S which have compact support.

(c) $\mu_n \to \mu$ strongly if $||\mu_n - \mu|| \to 0$ (where $||\ ||$ denote the total variation norm).

Let $\mathcal{P}_n$ and $\mathcal{P}$ be probability measures on (S, $\mathcal{S}$).

Each $\mathcal{P}$ is uniquely determined by the set of integrals, $\left\{\int f d\mathcal{P} : f \in C(S)\right\}$ (see Billingsley [2]), where C(S) is the space of bounded continuous real-valued functions f on S.

We will always use the supremum norm $\| f \|_\infty = \sup_{x \in S} |f(x)|$ on C(S).

Definition 2.14 If $\int f d\mathcal{P}_n \to \int f d\mathcal{P}$ for every $f \in C(S)$, then we say that $\mathcal{P}_n$ converges weakly to $\mathcal{P}$ and denote $\mathcal{P}_n \Rightarrow \mathcal{P}$.
This is a specialization of 2.13 (a).

Sometimes it may be difficult to check weak convergence using the definition. So, it is often convenient to use the following theorem. The following theorem provides useful conditions which are equivalent to weak convergence.

Theorem 2.15 Let $\mathcal{P}_n$ and $\mathcal{P}$ be probability measures on (S, $\mathcal{S}$). The following five conditions are equivalent :

(a) $\mathcal{P}_n \Rightarrow \mathcal{P}$.

(b) $\lim_{n\to\infty} \int f d\mathcal{P}_n = \int f d\mathcal{P}$ for all bounded and uniformly continuous real-valued functions $f \in C(S)$.

(c) $\limsup_{n\to\infty} \mathcal{P}_n(F) \leq \mathcal{P}(F)$ for every closed subset F in S.

(d) $\liminf_{n\to\infty} \mathcal{P}_n(G) \geq \mathcal{P}(G)$ for every open subset G in S.

(e) $\lim_{n\to\infty} \mathcal{P}_n(A) = \mathcal{P}(A)$ for all $\mathcal{P}$-continuity sets A in S (where A is a $\mathcal{P}$-continuity set if $\mathcal{P}(\partial A) = 0$ and ∂A is the boundary of A).

Proof. See Billingsley [2], Theorem 2.1 for detailed proof.

Secondly, we shall study weak convergence (or convergence in distribution) of random elements in a metric space S. We follow the same notations, (S, $\mathcal{S}$) as before.

Note that weak convergence can be reparaphased in terms of the convergence of the distribution of random elements.

Let $(\Omega, \mathcal{A}, P)$ be any probability space.

<u>Definition</u> <u>2.16</u> Let X be a mapping from a probability space $(\Omega, \mathcal{A}, P)$ into a metric space S. X is called a <u>random</u> <u>element</u> if X is measurable, that is, $X^{-1}(A) \in \mathcal{A}$ for every $A \subseteq S$.

The <u>distribution</u> of a random element X is the probability measure $P \circ X^{-1}$ induced by X on $(S, \mathcal{S})$ and we denote $\mathcal{P}_X = P \circ X^{-1}$. That is, $\mathcal{P}_X(A) = P \circ X^{-1}(A) = P\{w : X(w) \in A\} = P(X \in A)$ for each $A \in \mathcal{S}$.

<u>Remark</u> <u>2.17</u> The following examples are the most common ones of a random element X.

(1) If $S = R^1$, then X is called a random variable.

(2) If $S = R^d$(in the cases of $d \geq 2$), then X is called a random vector.

(3) If $S = C[0, \infty)$, the space of real-valued continuous functions on$[0, \infty)$, then X is called a random function (process with continuous paths).

(4) If $S = \mathfrak{M}$, then X is called a random measure.

(5) If $S = \mathfrak{N}$, then X is called a point process.

(6) If $S = D[0, \infty)$, the space of real-valued right continuous functions on $[0, \infty)$ with finite left limits existing on $(0, \infty)$, then X is called a random process with jump discontinuities.

Note that we can understand weak convergence in terms of the convergence of the distributions of random elements.

Definition 2.18 A sequence $\{X_n\}$ of random elements converges in distribution to the random element X, and we write $X_n \xrightarrow{\mathfrak{D}} X$, if the distributions $\mathcal{P}_{X_n}$ of X_n converge weakly to the distribution $\mathcal{P}_X$ of X, that is, $\mathcal{P}_{X_n} \Rightarrow \mathcal{P}_X$.

Remark 2.19 For convergence in distribution, Theorem 2.15 restates the equivalence of the following five statements as follows :

(a) $X_n \xrightarrow{\mathfrak{D}} X$.

(b) $\lim_{n\to\infty} E[f(X_n)] = E[f(X)]$ for all bounded and uniformly continuous real-valued functions f.

(c) $\limsup_{n\to\infty} P(X_n \in F) \leq P(X \in F)$ for all closed subset F in S.

(d) $\liminf_{n\to\infty} P(X_n \in G) \geq P(X \in A)$ for all open subset G in S.

(e) $\lim_{n\to\infty} P(X_n \in A) = P(X \in A)$ for all X-continuity sets A in S (where a set A is an X-continutiy set if $P(X \in \partial A) = 0$ and ∂A is the boundary of A).

Recall that we denote $\mathcal{P}_X(B) = P(X \in B) = P\{w : X(w) \in B\}$ for each $B \in \mathcal{S}$.

In the cases of the random element $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}^d, \mathfrak{B}^d)$ (where $\mathfrak{B}^d$ denotes the Borel σ-field of $\mathbb{R}^d$) with $S = \mathbb{R}^d$, where $d = 1, 2, \ldots$, the convergence in distribution of random variables (in the case $d = 1$) and random vectors (in the case when $d \geq 2$) is usually considered in terms of their distribution functions.

Definition 2.20 The distribution function of a random vector X is the function $F(\cdot) : \mathbb{R}^d \to [0, 1]$ defined by $F(\underset{\sim}{x}) = P\{\underset{\sim}{y} \in \mathbb{R}^d : \underset{\sim}{y} \leq \underset{\sim}{x}\}$ for every $\underset{\sim}{x} \in \mathbb{R}^d$, where $\underset{\sim}{y} \leq \underset{\sim}{x}$ means that $y_i \leq x_i$ for $i = 1, 2, \ldots, d$.

Note that the function F defined as above satisfies the following properties:

(1) $\lim_{\underset{\sim}{x} \to \infty} F(\underset{\sim}{x}) = 1$ and $\lim_{\underset{\sim}{x} \to \infty} F(\underset{\sim}{x}) = 0$.

(2) F is a non-decreasing function, that is, if $\underset{\sim}{x} \leq \underset{\sim}{y}$, then $F(\underset{\sim}{x}) \leq F(\underset{\sim}{y})$.

(3) F is right continuous, that is, for any x and the decreasing sequence $\{\underset{\sim}{x}_n : n \geq 1\}$, which converges to $\underset{\sim}{x}$, $\lim_{n \to \infty} F(\underset{\sim}{x}_n) = F(\underset{\sim}{x})$.

In particular, F is continuous at $\underset{\sim}{x}$ if and only if $\{\underset{\sim}{y} \in \mathbb{R}^d : \underset{\sim}{y} \leq \underset{\sim}{x}\}$ is an X-continuity set.

Definition 2.21 Let X be a random vector on $\mathbb{R}^d$. Then we define the characteristic function $\Phi_X(\underset{\sim}{t})$ by $\Phi_X(\underset{\sim}{t}) = E[\exp(i<\underset{\sim}{t} \cdot \underset{\sim}{x}>]$ for every $\underset{\sim}{t} \in \mathbb{R}^d$, where $<\underset{\sim}{t} \cdot \underset{\sim}{x}> = \sum_{i=1}^{d} t_i x_i$ with $\underset{\sim}{t} = (t_1, \ldots, t_d)$ and $\underset{\sim}{x} = (x_1, \ldots, x_d)$.

The characteristic function $\Phi_X(\underset{\sim}{t})$ of X is often used to understand some information about the distribution of a random vector X.

Remark 2.22 For distribution functions, Theorem 2.15 can be restated as the equivalent statements as follows :

(a) $X_n \xrightarrow{\mathfrak{D}} X$.

(b) $\lim_{n \to \infty} E[f(X_n)] = E[f(X)]$ for all bounded, uniformly continuous real-valued functions f.

(c) $\limsup_{n \to \infty} P(X_n \in F) \leq P(X \in F)$ for all closed subsets F in $\mathbb{R}^d$.

(d) $\liminf_{n \to \infty} P(X_n \in G) \leq P(X \in G)$ for all open subsets G in $\mathbb{R}^d$.

(e) $\lim_{n \to \infty} F_n(\underset{\sim}{x}) = F(\underset{\sim}{x})$ for all continuity points $\underset{\sim}{x}$ of $F(\underset{\sim}{x})$ for $\underset{\sim}{x} \in \mathbb{R}^d$.

(f) If Φ_{X_n} and Φ_X are the characteristic functions corresponding to the random vectors X_n and X, respectively, then $\Phi_{X_n}(\underset{\sim}{t}) \to \Phi_X(\underset{\sim}{t})$ for every $\underset{\sim}{t} \in \mathbb{R}^d$.

Finally, let us consider weak convergence and convergence in distribution for random measure. We have already introduced $\mathfrak{M}$, the collection of Radon measures on a metric space S in Section 2.1.. Here we shall follow the same notations that we defined in Section 2.1.. Also we recall that a random measure X is a measurable mapping from a probability space $(\Omega, \mathcal{A}, P)$ into $(\mathfrak{M}, \mathcal{M})$ defined on $\mathbb{R}^d$.

Definition 2.23 Let X, X_1, X_2, . . . be a sequence of random measures. If the sequence of probability measures $\{\mathcal{P}_{X_n}\}$ induced by X_n converges weakly to the probability measure $\mathcal{P}_X$ induced by the limit X, then we say that the sequence $\{X_n\}$ converges weakly to a limit X.

Definition 2.24 For random measures X, X_1, X_2, . . . , if $\lim_{n\to\infty} E[f(X_n)] = E[f(X)]$ for every bounded continuous real-valued function defined on $\mathfrak{M}$, then we say that the sequence $\{X_n\}$ of random measures converges in distribution to a limit X and we denote $X_n \xrightarrow{\mathfrak{D}} X$.

In above, we can know that the functions $f(X_n)$ and $f(X)$ are random variables because $f(X_n)$ and $f(X)$ are functions from Ω to R. In fact the convergence in distribution in this case is the same convergence in distribution as given for random element in Definition 2.16. Now, using the characteristic functional and Laplace functional, we can give the equivalent statements of convergence in distribution.

Theorem 2.25 Let X, X_1, X_2, . . . be random measures. C_{X_n} and C_X are the characteristic functionals corresponding to X_n and X, respectively. L_{X_n} and L_X

are the Laplace functionals corresponding to X_n and X, respectively. Then the following statements are equivalent :

(a) $X_n \overset{\mathfrak{D}}{\to} X$.

(b) $X_n f \to Xf$ in distribution as random variables for every continuous function f from R^d to R with compact support.

(c) C_{X_n} converges to C_X.

(d) L_{X_n} converges to L_X.

2.4. Introduction to Large Deviation Theory

For more complete discussion of theory of large deviations, the reader is referred to Ellis's book [12], Stroock's book [29] and Varadhan's book [30].

The large deviation theory is concerned with the exponential decay rate. We first consider a general framework for such problems. Let E be a complete metric space and $\{P_n : n = 1, 2, \ldots\}$ be a sequence of probability measures on the Borel subsets of E. Suppose $\{P_n\}$ converges weakly to the unit point measure at $x_0 \in E$ as $n \to \infty$. It means that $P_n \Rightarrow \delta_{x_0}$ as $n \to \infty$. If A is any Borel set such that $x_0 \notin \bar{A}$, $P_n(A) \to 0$ as $n \to \infty$. We think that $P_n(A)$ converges to 0 exponentially fast as $n \to \infty$ with an exponential rate depending on the set A.

Here, there is the simplest example in the theory of large deviations. Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with $E[X_1] = m$ and $Var(X_1) > 0$. Let S_n be the n-th partial sums of the random variables, that is, $S_n = X_1 + \ldots + X_n$. If $E[|X_1|]$ is finite, then $\frac{S_n}{n}$ converges to m with probability one by the strong law of large numbers and $\frac{S_n - nm}{\sqrt{n}} \Rightarrow$

$\mathcal{N}(0, \sigma^2)$ by the central limit theorem, where $\mathcal{N}(0, \sigma^2)$ is normally distributed with mean 0 and variance σ^2. These mean that

$$\text{L.L.N. says that } P\left(\frac{S_n}{n} > a\right) \to 0 \text{ as } n \to \infty$$

and

$$\text{C.L.T. says that } P\left(\frac{S_n - nm}{n} > a\right) = P\left(\frac{S_n - nm}{\sqrt{n}} > a\sqrt{n}\right) \to 0$$

as $n \to \infty$ for $a > m$.

Now we know that the event $\{\frac{S_n}{n} > a\}$ is asymptotically "rare" for $a > m$. That is, this event occurs for only finitely many n. But the law of large numbers and the central limit theorem do not give a clue as to how to estimate the desired probability or its rate of convergence to zero as $n \to \infty$.

Large deviation theory shows that if the random variables are exponentially bounded, then the probability converges to zero exponentially fast as $n \to \infty$. The most fascinating aspect of this theory is that the exponential decay rates are computable in terms of entropy functions.

Definition 2.26 Let E be a Polish space with $\mathfrak{B}_E$, the Borel σ-field. If a function $I(\cdot) : E \to [0, \infty]$ satisfies the following conditions :

(a) $I(\cdot)$ is lower semi-continuous,

(b) I has compact level sets, that is, the set $\{x \in E : I(x) \leq a\}$ is compact for every real number $a < \infty$,

then we say that the function $I(\cdot)$ is a rate function.

In addition, let $\{P_\epsilon : \epsilon > 0\}$ be a family of probability measures on

$(E, \mathscr{B}_E)$. We say that a family $\{P_\epsilon : \epsilon > 0\}$ satisfies a large deviation principle with rate function $I(\cdot)$ if the following hold ;

(c) For all closed subsets F in E,

$$\limsup_{\epsilon \to 0} \epsilon \cdot \log P_\epsilon(F) \leq -\inf_{x \in F} I(x),$$

(d) For all open subsets G in E,

$$\liminf_{\epsilon \to 0} \epsilon \cdot \log P_\epsilon(G) \geq -\inf_{x \in G} I(x).$$

Let $X_1, X_2, \ldots, X_n$ be independent random variables with common distribution $F(x)$. Set $S_n = \sum_{k=1}^{n} X_k$ and let P_n be the distribution of $\frac{S_n}{n}$. Now let us consider the moment generating function of S_n as follows :

$$M_n(t) = E[\exp(tS_n)] = \Big\{E[\exp(tX_1)]\Big\}^n \text{ by independence.}$$

The limit of the cumulant generating function is

$$M(t) = \lim_{n \to \infty} \frac{1}{n} \cdot \log M_n(t) = \lim_{n \to \infty} \frac{1}{n} \cdot \log \Big\{E[\exp(tX_1)]\Big\}^n = \log E[e^{tX_1}].$$

Let us define Legendre-Fenchel transformation of M(t) as follows :

$$I(x) = \sup_{t \in R} \Big\{tx - M(t)\Big\} = \sup_{t \in R} \Big\{tx - \log E[e^{tX_1}]\Big\}.$$

Remark 2.27 $I(m) = 0$ whenever $E[X] = m$: For every real number t, $tm - \log M(t) \leq 0$ since $\log E[e^{tX}] \geq E[\log(e^{tX})] = tm$ by Jensen's inequality.

Thus, $I(m) = 0$.

There are some properties of $I(x)$ for $x \in \mathbb{R}$ as follows :

(a) $I(x) \in [0, \infty]$ since $tx - \log M(t) = 0$ when $t = 0$ for every real number x in $\mathbb{R}$.

(b) $I(x)$ is lower semi-continuous as the supremum of continuous functions.

(c) $I(\cdot)$ has compact level sets, that is, for any $a < \infty$, $\{x \in \mathbb{R} : I(x) \le a\}$ is compact : For every real number $a < \infty$, the set $\{x \in \mathbb{R} : I(x) \le a\}$ is closed since $I(x)$ is lower semi-continuous in (b). Moreover, it is also bounded. That is to say, $\{x \in \mathbb{R} : I(x) \le a\} \subseteq [-a - \log M(-1), a + \log M(1)]$ because $I(x) = \sup_{t\in\mathbb{R}} \{tx - \log M(t)\} \le a$ implies that $-a - \log M(1) \le x \le a + \log M(1)$. Thus the set $\{x \in \mathbb{R} : I(x) \le a\}$ is compact by the Heine-Borel theorem.

(d) $I(x)$ is convex : For every x and y in E and $\alpha \in [0, 1]$,

$$I\Big(\alpha x + (1 - \alpha)y\Big) = \sup_{t\in\mathbb{R}} \Big\{\{\alpha x + (1 - \alpha)y\}t - \log M(t)\Big\}$$

$$= \sup_{t\in\mathbb{R}} \Big\{\{\alpha x + (1 - \alpha)y\}t - \log M(t) + \alpha\cdot\log M(t) - \alpha\cdot\log M(t)\Big\}$$

$$= \sup_{t\in\mathbb{R}} \Big\{\alpha\{xt - \log M(t)\} + (1 - \alpha)\{ yt - \log M(t)\}\Big\}$$

$$\le (\alpha)\sup_{t\in\mathbb{R}} \Big\{xt - \log M(t)\Big\} + (1 - \alpha)\sup_{t\in\mathbb{R}} \Big\{yt - \log M(t)\Big\}$$

$$= \alpha\cdot I(x) + (1 - \alpha)\cdot I(y).$$

(e) $I(x)$ is non-increasing on $(-\infty, m]$ and non-decreasing on $[m, \infty)$, where $E[X] = m$: First, to show that $I(x)$ is non-increasing on $[-\infty, m]$; For $x \le y \le m$ and $0 < \alpha < 1$ such that $y = \alpha x + (1 - \alpha)m$,

$$
\begin{aligned}
I(y) &= I\big(\alpha x + (1 - \alpha)m\big) \\
&\leq \alpha \cdot I(x) + (1 - \alpha)\cdot I(m) \text{ by convexity of I} \\
&= \alpha \cdot I(x) \text{ since } I(m) = 0. \\
&\leq I(x).
\end{aligned}
$$

That is, $I(y) \leq I(x)$ for every $x \leq y$ in $(-\infty, m]$. It means that I(x) is non-increasing on $[-\infty, m]$.

To show that I(x) is non-decreasing on $[m, \infty)$;

Similarly as above, for $m \leq u \leq v$ and $0 < \beta < 1$ such that $u = \beta m + (1 - \beta)v$,

$$
\begin{aligned}
I(u) &= I\big(\beta m + (1 - \beta)v\big) \\
&\leq \beta \cdot I(m) + (1 - \beta)\cdot I(v) \text{ by convexity of I} \\
&= (1 - \beta)\cdot I(v) \text{ since } I(m) = 0. \\
&\leq I(v).
\end{aligned}
$$

That is, $I(u) \leq I(v)$ for every $u \leq v$ in $[m, \infty)$. It means that I(x) is non-decreasing on $[m, \infty)$.

<u>Lemma</u> <u>2.29</u> The function $I(\cdot)$ is a rate function. In fact I(x) is non-decreasing on $[m, \infty)$ and non-increasing on $(-\infty, m]$, where $m = E[X]$.

The following theorem due to Cramér is the basic large deviation theorem. It will be extended in this thesis. Its proof may be found in Ellis [12], Strook [29] and Varadhan [30].

<u>Theorem</u> <u>2.30</u> Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables. Let P_n be the distribution of S_n/n, where $S_n = X_1 + \ldots + X_n$. If the moment generating function is finite, then $\{P_n : n = 1, 2, \ldots\}$ satisfies a large deviation principle with rate function $I(\cdot)$.

3. LARGE DEVIATION PROPERTIES FOR A FINITE DIMENSIONAL CASE OF RANDOM MEASURE

3.1. General Definition

We have already defined random measures on a locally compact Polish space S in Section 2.1.. Here, let us recall the special case where $S = \mathbb{R}^d$, $d \geq 1$.

Let $\mathbb{R}^d$ be d-dimensional Euclidean space. Let $\mathfrak{B}^d$ denote the collection of Borel subsets of $\mathbb{R}^d$. Let $\mathfrak{B}_0{}^d$ be the collection of bounded subsets (sets with compact closure) of $\mathfrak{B}^d$. $\mathfrak{M}$ denotes the set of all non-negative measures μ defined on $(\mathbb{R}^d, \mathfrak{B}^d)$ and finite on bounded sets, which are called Radon measures. $\mathcal{M}$ is the σ-field on $\mathfrak{M}$ generated by sets of the form $\{\mu \in \mathfrak{M} : \mu(B) < r\}$ for all $B \in \mathfrak{B}^d$ and $r \geq 0$. Moreover, $\mathfrak{N}$ is the subset of $\mathfrak{M}$ consisting of non-negative integer measures. Thus

$\mu \in N$ if and only if $\mu(B) \in \mathbb{Z}^+ = \{0, 1, 2, \ldots\}$ and $\mu(B) < \infty$ for all $B \in \mathfrak{B}_0{}^d$.

Likewise, $\mathfrak{N}$ has the σ-field $\mathcal{N}$ generated by the sets of the form $\{\mu \in \mathfrak{N} : \mu(B) = k\}$ for $B \in \mathfrak{B}^d$, and $k \in \mathbb{Z}^+$. $\mathfrak{N}$ may be naturally identified with the set of all finite or infinite configurations of points in B without limit points, but which may have multiplicities. Also $\mathfrak{N}$ is a closed subspace of $\mathfrak{M}$. Note that $\mathcal{N} \subseteq \mathcal{M}$ and $\mathcal{N}$ is the restriction of $\mathcal{M}$ to $\mathfrak{N}$.

Ellis has studied large deviation theorems for sums of general random vectors X_n.

Here, we will consider large deviation properties of the finite dimensional distributions of random measures.

Let X be a random measure in $\mathfrak{M}$. Now, we define

$$X_\lambda(A) = X(\lambda A), \text{ where A is a Borel subset of } \mathbb{R}^d \tag{3.1}$$

and

$$f_\lambda(x) = f(\tfrac{x}{\lambda}) \text{ and } X_\lambda(f) = X(f_\lambda). \tag{3.2}$$

Notice that these definitions are consistent in the sense that $X_\lambda(A) = X_\lambda(1_A)$.

Let us consider a random vector $\underset{\sim}{Y}_\lambda = \big(X_\lambda(A_1), \ldots, X_\lambda(A_n)\big)$, where all A_i's are disjoint and bounded Borel subsets of $\mathbb{R}^d$. The moment generating function of the random vector $\underset{\sim}{Y}_\lambda$ is defined by

$$M_\lambda(\underset{\sim}{t}) = E[e^{<\underset{\sim}{t},\, \underset{\sim}{Y}_\lambda>}] = E[\exp\Big(\sum_{i=1}^n t_i X_\lambda(A_i)\Big)] \text{ for all } \underset{\sim}{t} \in \mathbb{R}^n, \tag{3.3}$$

where $\underset{\sim}{t} = (t_1, \ldots, t_n)$.

Assume that

(3.4) (i) $\Phi_\lambda(\underset{\sim}{t}) = \frac{1}{\lambda^d}\cdot\log E[e^{<\underset{\sim}{t},\, \underset{\sim}{Y}_\lambda>}]$ is finite for all $\underset{\sim}{t} \in \mathbb{R}^n$

and

(ii) $\Phi(\underset{\sim}{t}) = \lim_{\lambda\to\infty} \Phi_\lambda(\underset{\sim}{t})$ exists and is finite for all $\underset{\sim}{t} \in \mathbb{R}^n$.

Define the Legendre-Fenchel transformation of the convex function Φ by

$$I(\underset{\sim}{x}) = \sup_{\underset{\sim}{t}\in\mathbb{R}^n}\Big\{<\underset{\sim}{t}, \underset{\sim}{x}> - \Phi(\underset{\sim}{t})\Big\}. \tag{3.5}$$

Definition 3.1 We assume the above properties (i) and (ii) in (3.4). We will say that the function $I(\cdot) : \mathbb{R}^n \to [0, \infty]$ is a rate function if

(a) $0 \le I(\underset{\sim}{x}) \le \infty$ for all $\underset{\sim}{x} \in \mathbb{R}^n$,

(b) $I(\underset{\sim}{x})$ is lower semi-continuous for $\underset{\sim}{x} \in \mathbb{R}^n$,

(c) I has compact level sets, that is, for each $a < \infty$, the set $\{\underset{\sim}{x} \in \mathbb{R}^n : I(\underset{\sim}{x}) \le a\}$ is compact in $\mathbb{R}^n$.

Definition 3.2 We say a random measure X on $\mathbb{R}^d$ satisfies a large deviation principle in the sense of the finite dimensional distributions if for all bounded and disjoint Borel subsets, $A_1, \ldots, A_n$ and $\underset{\sim}{Y}_\lambda$ as above, there is a rate function $I(\cdot) : \mathbb{R}^d \to [0, \infty]$ (where $I(\cdot)$ depends on the sets $A_1, \ldots, A_n$) so that $\frac{\underset{\sim}{Y}_\lambda}{\lambda^d}$ satisfies a large deviation principle with rate function $I(\cdot)$ as defined in (3.5). This means that

(d) For each closed subset $F \subseteq \mathbb{R}^n$,

$$\limsup_{\lambda \to \infty} \frac{1}{\lambda^d} \cdot \log P\left(\frac{\underset{\sim}{Y}_\lambda}{\lambda^d} \in F \right) \le -\inf_{\underset{\sim}{x} \in F} I(\underset{\sim}{x}), \tag{3.6}$$

(e) For each open subset $G \subseteq \mathbb{R}^n$,

$$\liminf_{\lambda \to \infty} \frac{1}{\lambda^d} \cdot \log P\left(\frac{\underset{\sim}{Y}_\lambda}{\lambda^d} \in G \right) \ge -\inf_{\underset{\sim}{x} \in G} I(\underset{\sim}{x}). \tag{3.7}$$

Ellis has proved a large deviation principle for random vectors in his paper [13].

In particular, Ellis' Theorem imlpies the following theorem.

Theorem 3.3 We assume the above hypotheses (i) and (ii) in (3.4). Then the following conclusions hold ;

(a) $I(\underset{\sim}{x})$ is convex, non-negative, lower semi-continuous, and has compact level sets. It means that $I(\cdot)$ is a rate function.

(b) The upper large deviation bound is valid :

$$\limsup_{\lambda \to \infty} \frac{1}{\lambda^d} \cdot \log P\left(\frac{\underset{\sim}{Y}_\lambda}{\lambda^d} \in F\right) \leq -\inf_{\underset{\sim}{x} \in F} I(\underset{\sim}{x}) \text{ for each closed subset } F \text{ in } \mathbb{R}^n.$$

(c) If $\Phi(\underset{\sim}{t})$ is differentiable for all $\underset{\sim}{t} \in \mathbb{R}^n$, then the lower large deviation bound is valid :

$$\liminf_{\lambda \to \infty} \frac{1}{\lambda^d} \cdot \log P\left(\frac{\underset{\sim}{Y}_\lambda}{\lambda^d} \in G\right) \geq -\inf_{\underset{\sim}{x} \in G} I(\underset{\sim}{x}) \text{ for each open subset } G \text{ in } \mathbb{R}^n.$$

Hence, if $\Phi(\underset{\sim}{t})$ is differentiable for all $\underset{\sim}{t} \in \mathbb{R}^n$, then (a), (b), and (c) imply that $\frac{\underset{\sim}{Y}_\lambda}{\lambda^d}$ satisfies a large deviation principle with rate function $I(\cdot)$.

Proof. The above theorem is proved in Ellis's book [12].

Next, we consider the finite dimensional large deviation properties of Poisson random measures, Poisson center cluster random measures and doubly stochastic processes.

3.2. Poisson Random Measure with Intensity $\alpha \geq 0$

Let X be a Poisson point process with intensity $\alpha > 0$. Let A be a bounded Borel subset of $\mathbb{R}^d$. The moment generating function of a random variable $X_\lambda(A)$ is

$$\Phi_\lambda(t) = E[e^{t \cdot X_\lambda(A)}] = e^{\alpha\lambda^d|A|(e^t - 1)}$$ for every $t \in \mathbb{R}$ in the special

case where $n = 1$ from (3.3).

Let $\underset{\sim}{Y}_\lambda = \big(X_\lambda(A_1), \ldots, X_\lambda(A_n)\big)$, where all A_i's are disjoint bounded subsets of $\mathbb{R}^d$. Recall $\Phi(\underset{\sim}{t})$ and $I(\underset{\sim}{x})$ of a random vector $\underset{\sim}{Y}_\lambda$ for a Poisson process X defined in (3.4) and (3.5) as follows :

$$\Phi(\underset{\sim}{t}) = \lim_{\lambda \to \infty} \frac{1}{\lambda^d} \cdot \log \Phi_\lambda(\underset{\sim}{t}) \quad \text{and} \quad I(\underset{\sim}{x}) = \sup_{\underset{\sim}{t} \in \mathbb{R}^n} \left\{ <\underset{\sim}{t}, \underset{\sim}{x}> - \Phi(\underset{\sim}{t}) \right\}.$$

In this case, we will prove that $\dfrac{\underset{\sim}{Y}_\lambda}{\lambda^d}$ satisfies a large deviation principle with rate function $I(\cdot)$.

Remark 3.4 The stationary Poisson point process X with intensity $\alpha > 0$ is characterized by

(a) X(A) is a Poisson random variable with parameter $\alpha \cdot |A|$ for every $A \in \mathfrak{B}^d$.

(b) If $A_1, A_2, \ldots$ are disjoint subsets of $\mathfrak{B}_0{}^d$, then $X(A_1), X(A_2), \ldots$ are independent random variables.

Proposition 3.5 Let X be a Poisson random measure with intensity $\alpha > 0$. Then

$$\Phi(\underset{\sim}{t}) = \alpha\sum_{i=1}^{n}(e^{t_i} - 1)|A_i| \text{ and } I(\underset{\sim}{x}) = \sup_{\underset{\sim}{t}\in\mathbb{R}^n}\Big\{<\underset{\sim}{t}, \underset{\sim}{x}> - \alpha\sum_{i=1}^{n}(e^{t_i} - 1)|A_i|\Big\}.$$

Moreover, $I(\underset{\sim}{x})$ can be rewritten by $I(\underset{\sim}{x}) = \sum_{i=1}^{n}\Big\{x_i\cdot\log\Big(\frac{x_i}{\alpha|A_i|}\Big) - x_i + \alpha|A_i|\Big\}$.

<u>Proof.</u> First let us calculate the moment generating function $M_\lambda(\underset{\sim}{t})$ of $\underset{\sim}{Y}_\lambda$.

$$M_\lambda(\underset{\sim}{t}) = E[e^{<\underset{\sim}{t}, \underset{\sim}{Y}_\lambda>}]$$

$$= E[e^{t_1X_\lambda(A_1) + \ldots + t_nX_\lambda(A_n)}]$$

$$= e^{\alpha\lambda^d\{(e^{t_1} - 1)|A_1| + \ldots + (e^{t_n} - 1)|A_n|\}} \text{ by Remark 3.4.}$$

So,

(3.8) $$\Phi(\underset{\sim}{t}) = \alpha\sum_{i=1}^{n}\Big(e^{t_i} - 1\Big)|A_i|.$$

Also,

(3.9) $$I(\underset{\sim}{x}) = \sup_{\underset{\sim}{t}\in\mathbb{R}^n}\Big\{<\underset{\sim}{t}, \underset{\sim}{x}> - \alpha\sum_{i=1}^{n}(e^{t_i} - 1)|A_i|\Big\}.$$

Now, let $$H_{\underset{\sim}{x}}(\underset{\sim}{t}) = <\underset{\sim}{t}, \underset{\sim}{x}> - \alpha\sum_{i=1}^{n}(e^{t_i} - 1)|A_i|.$$

Differentiating $H_{\underset{\sim}{x}}(\underset{\sim}{t})$ with respect to each t_i and then solving in terms of each x_i ,

$$\frac{\partial H_{\underset{\sim}{x}}}{\partial t_i} = x_i - \alpha\cdot\exp(t_i)|A_i| = 0 \text{ for each } i.$$

(3.10) $$\Rightarrow \quad \frac{x_i}{\alpha|A_i|} = \exp(t_i) \text{ for each } i.$$

Note that each x_i is positive.

So, taking logarithm on both sides in (3.10), we get

(3.11) $$t_i = \log\Big(\frac{x_i}{\alpha|A_i|}\Big) \text{ for each } i.$$

Replugging (3.11) in (3.9), then we get that

$$I(\underset{\sim}{x}) = \sum_{i=1}^{n} x_i \cdot \log\left(\frac{x_i}{\alpha|A_i|}\right) - \sum_{i=1}^{n} x_i + \alpha\sum_{i=1}^{n}|A_i|.$$

<u>Theorem 3.6</u> Let X be a Poisson random measure with intensity $\alpha > 0$. Let $\underset{\sim}{Y}_\lambda = \left(X_\lambda(A_1), \ldots, X_\lambda(A_n)\right)$, where all A_i's are disjoint and bounded subsets of $\mathbb{R}^d$. Then $\frac{\underset{\sim}{Y}_\lambda}{\lambda^d}$ satisfies a large deviation principle with rate function $I(\cdot)$ as follows :

$$I(\underset{\sim}{x}) = \sup_{\underset{\sim}{t}\in\mathbb{R}^n}\left\{<\underset{\sim}{t}, \underset{\sim}{x}> - \alpha\sum_{i=1}^{n}(e^{t_i} - 1)|A_i|\right\} = \sum_{i=1}^{n} x_i \log\left(\frac{x_i}{\alpha|A_i|}\right) - \sum_{i=1}^{n} x_i + \alpha\sum_{i=1}^{n}|A_i|.$$

<u>Proof.</u> (i) If $\underset{\sim}{t} = 0$, that is to say, $\underset{\sim}{t} = (0, \ldots, 0)$ then $<\underset{\sim}{t}, \underset{\sim}{x}> - \Phi(\underset{\sim}{t}) = 0$. Thus, $I(\underset{\sim}{x}) \geq 0$.

(ii) $<\underset{\sim}{t}, \underset{\sim}{x}>$ is continuous for $\underset{\sim}{x}$. Thus $<\underset{\sim}{t}, \underset{\sim}{x}> - \Phi(\underset{\sim}{t})$ is also continuous for $\underset{\sim}{x}$. So, $I(\underset{\sim}{x})$ is lower semi-continuous.

(iii) For any $a \geq 0$, let $K_a = \{\underset{\sim}{x} \in \mathbb{R}^n : I(\underset{\sim}{x}) \leq a\}$. Since $I(\underset{\sim}{x})$ is lower semi-continuous by (ii), $\{\underset{\sim}{x} \in \mathbb{R}^n : I(\underset{\sim}{x}) > a\}$ is open, that is, $\{\underset{\sim}{x} \in \mathbb{R}^n : I(\underset{\sim}{x}) > a\}^C$ is closed. So, K_a is closed.

On the other hand, let $\underset{\sim}{x} \in K_a$. Then $I(\underset{\sim}{x}) \leq a$ implies $<\underset{\sim}{t}, \underset{\sim}{x}> - \Phi(\underset{\sim}{t}) \leq a$, that is, $<\underset{\sim}{t}, \underset{\sim}{x}> \leq a + \Phi(\underset{\sim}{t})$. Take $t_i = (0, \ldots, 0, \pm1, 0, \ldots, 0)$. Then

$$-a + (1 - \tfrac{1}{e})|A| \leq x_i \leq a + (e - 1)|A|.$$

Each x_i is bounded. So, $\underset{\sim}{x} = (x_1, \ldots, x_n)$ is bounded, that is, K_a is bounded. Thus K_a is compact.

Finally, for (iv) and (v), obviously $\Phi(\underset{\sim}{t})$ is differentiable for all $\underset{\sim}{t}$. Therefore, $\frac{\underset{\sim}{Y}_\lambda}{\lambda^d}$ satisfies a large deviation principle with rate function $I(\cdot)$.

3.3. Poisson Center Cluster Process

One class of cluster processes occurs so frequently in applications and is so important in the theory that it warrants special attention. In this class the cluster centers follow a Poisson distribution and the clusters are independent and finite with probability 1. Let X be a Poisson center cluster process with centers U, a Poisson random field with intensity ρ and members V, a random measure with finite expected total mass, $E[V(\mathbb{R}^d)] = \gamma$. We define the moment generating function of $V(\mathbb{R}^d)$ by

$$M_{V(\mathbb{R}^d)}(t) = E[e^{tV(\mathbb{R}^d)}]$$

and assume that $M_{V(\mathbb{R}^d)}(t)$ is finite for each $t \in \mathbb{R}$. The random measure X is defined by superimposing i.i.d. copies of V centered at the occurrences of U. More precisely, let U have occurrences $\{x_i\}$ and let $\{V_i\}$ be i.i.d. random measures independent of U and distributed as V. If A is a bounded Borel subset of $\mathbb{R}^d$, X is defined by

$$X(A) = \sum_{X_i} V_i(A - x_i) = \int_{\mathbb{R}^d} V_X(A - x)U(dx) .$$

We denote X by [U,V] and X is a stationary random measure with intensity $\rho\gamma$. Note that $E[X(A)] = \rho\gamma|A|$.

Define $$X(f) = \int_{\mathbb{R}^d} V_X(T_X f)U(dx) = \sum_i V_i(T_{X_i} f),$$

where T_X is a translation operator, that is, $T_X f(y) = f(x+y)$ and f is a bounded measurable function with compact support defined on $\mathbb{R}^d$.

<u>Theorem</u> <u>3.7</u> The moment generating functional of X is

$$M_X(f) = E[e^{X(f)}] = \exp\Big(\rho\int_{\mathbb{R}^d}\big(M_V(T_xf) - 1\big)dx\Big).$$

<u>Proof.</u> $M_X(f) = E[e^{X(f)}] = E[E[\exp(\sum_i V_i(T_{x_i}f)) \mid U=\{x_i\}]]$

$$= E[\prod_i E[\exp(V_i(T_{x_i}f)) \mid U=\{x_i\}]]$$

$$= E[\exp(\sum_i \log E[\exp(V(T_{x_i}f))]) \mid U=\{x_i\}]$$

$$= E[\exp(\int_{\mathbb{R}^d} \log E[\exp(V(T_xf))]U(dx)],$$

where $M_V(f) = E[e^{V(f)}]$

$$= E[\exp\big\{U\big(\log M_V(T_xf)\big)\big\}]$$

$$= \exp\Big(\rho\int_{\mathbb{R}^d}\big(M_V(T_xf) - 1)\big)dx\Big),$$

where U is a Poisson random measure with intensity ρ.

The last equality makes sense since $\int_{\mathbb{R}^d}\big(M_V(T_xf) - 1\big)dx$

$$= \int_{\mathbb{R}^d} E[\exp\Big(\int_{\mathbb{R}^d} f(x+y)V(dy)\Big) - 1]dx$$

$$= \int_{\mathbb{R}^d} E[\sum_{k=1}^{\infty}\tfrac{1}{k!}\{\int_{\mathbb{R}^d} f(x+y)V(dy)\}^k]dx$$

$$\leq E[\sum_{k=1}^{\infty}\tfrac{1}{k!}c^k\int_{\mathbb{R}^d}\cdots\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} 1_K(x+y_1)\ldots1_K(x+y_k)dxV(dy_1)\ldots V(dy_k)]$$

$$\leq E[\sum_{k=1}^{\infty}\frac{1}{k!}c^k\int_{\mathbb{R}^d}\cdots\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}1_K(x)dxV(dy_1)...V(dy_k)]$$

$$= E[\sum_{k=1}^{\infty}\frac{1}{k!}c^k\int_{\mathbb{R}^d}\cdots\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}|K|V(dy_1)...V(dy_k)]$$

$$= |K|\cdot E[e^{cV(\mathbb{R}^d)} - 1] < \infty.$$

Thus, $$M_X(f) = \exp\left(\rho\int_{\mathbb{R}^d}\left(M_V(T_xf) - 1\right)dx\right).$$

<u>Proposition 3.8</u> $$M_{X_\lambda}(f) = \exp\left(\rho\int_{\mathbb{R}^d}\left(M_V(T_xf_\lambda) - 1\right)dx\right).$$

In the case of a Poisson center cluster random measure, let us consider a finite dimensional distribution, that is to say, a random vector $\underset{\sim}{Y}_\lambda = \left(X_\lambda(A_1), \ldots, X_\lambda(A_n)\right)$, where all A_i's are disjoint and bounded Borel subsets of $\mathbb{R}^d$ for $i = 1, \ldots, n$.

In the case of $n = 1$, $\dfrac{X_\lambda(A)}{\lambda^d}$ satisfies a large deviation principle with rate function $I_X(\cdot)$ as follows :

$$I_X(x) = \sup_{t\in\mathbb{R}}\left\{tx + \rho|A| - \rho|A|M_{V(\mathbb{R}^d)}(t)\right\},$$

where A is a bounded Borel subset of $\mathbb{R}^d$.

Here, we consider the above property to a finite dimensional case of $n \geq 2$. This means that $\underset{\sim}{Y}_\lambda/\lambda^d$ satisfies a large deviation principle with rate function $I_Y(\cdot)$ as follows :

$$I_Y(x) = \sup_{\underset{\sim}{t}\in\mathbb{R}^n} \left\{ <\underset{\sim}{t}, \underset{\sim}{x}> + \sum_{i=1}^{n} \rho|A_i| - \sum_{i=1}^{n} \rho|A_i| M_{V(\mathbb{R}^d)}(\underset{\sim}{t}) \right\}.$$

Now consider $M_{\underset{\sim}{Y}_\lambda}(\underset{\sim}{t}) = E[\exp(<\underset{\sim}{t}, \underset{\sim}{Y}_\lambda>)]$, where $\underset{\sim}{Y}_\lambda = \big(X_\lambda(A_1), \ldots, X_\lambda(A_n)\big)$.

$$M_{\underset{\sim}{Y}_\lambda}(\underset{\sim}{t}) = E[\exp(<\underset{\sim}{t}, \underset{\sim}{Y}_\lambda>)]$$

$$= E[e^{X_\lambda(f)}], \text{ where } f = \sum_{i=1}^{n} t_i 1_{A_i}$$

$$= E[e^{X(f_\lambda)}]$$

$$= \exp\left\{ \rho \int_{\mathbb{R}^d} E\left\{ \exp\Big(\int_{\mathbb{R}^d} f\Big(\frac{x+y}{\lambda}\Big) V(dy) \Big) - 1 \right\} dx \right\}$$

by Proposition 3.8.

Proposition 3.9 $\lim_{\lambda\to\infty} \frac{1}{\lambda^d} \cdot \log M_{\underset{\sim}{Y}_\lambda}(\underset{\sim}{t}) = \rho \int_{\mathbb{R}^d} \big(M_{V(\mathbb{R}^d)}(f) - 1 \big) dx.$

Proof. $\lim_{\lambda\to\infty} \frac{1}{\lambda^d} \cdot \log M_{\underset{\sim}{Y}_\lambda}(\underset{\sim}{t})$

$$= \lim_{\lambda\to\infty} \frac{\rho}{\lambda^d} \int_{\mathbb{R}^d} E\left\{ \exp\Big(\int f\Big(\frac{x+y}{\lambda}\Big) V(dy) \Big) - 1 \right\} dx$$

$$= \lim_{\lambda\to\infty} \rho \int_{\mathbb{R}^d} E\Big[e^{\int_{\mathbb{R}^d} f(x'+y/\lambda) V(dy)} - 1\Big] dx', \text{ letting } x' = x/\lambda$$

$$= \lim_{\lambda\to\infty} \rho \int_{\mathbb{R}^d} E\Big[e^{\int_{\mathbb{R}^d} f(x+y/\lambda) V(dy)} - 1\Big] dx, \text{ letting } x' = x$$

$$= \lim_{\lambda\to\infty} \rho \int_{\mathbb{R}^d} E\Big[\sum_{k=1}^{\infty} \frac{1}{k!} \Big\{ \int_{\mathbb{R}^d} f(x+y/\lambda) V(dy) \Big\}^k \Big] dx$$

$$= \lim_{\lambda \to \infty} \rho \sum_{k=1}^{\infty} \frac{1}{k!} E\Big[\int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} f(x+\tfrac{y_1}{\lambda}) \ldots f(x+\tfrac{y_k}{\lambda}) V(dy_1) \ldots V(dy_k) dx\Big]$$

$$= \lim_{\lambda \to \infty} \rho \sum_{k=1}^{\infty} \frac{1}{k!} E\Big[\int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} f(x+\tfrac{y_1}{\lambda}) \ldots f(x+\tfrac{y_k}{\lambda}) dx V(dy_1) \ldots V(dy_k)\Big]$$

The last term converges as $\lambda \to \infty$ by the Lebesque Dominated Convergence Theorem and Lusin's theorem to

$$= \rho \sum_{k=1}^{\infty} \frac{1}{k!} E\Big[\int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} f(x)^k dx V(dy_1) \ldots V(dy_k)\Big]$$

$$= \rho \int_{\mathbb{R}^d} E\Big[\sum_{k=1}^{\infty} \frac{1}{k!} \{f(x) V(\mathbb{R}^d)\}^k\Big] dx$$

$$= \rho \int_{\mathbb{R}^d} E\Big[e^{f(x)V(\mathbb{R}^d)} - 1\Big] dx$$

$$= \rho \int_{\mathbb{R}^d} \Big(M_{V(\mathbb{R}^d)}(f) - 1\Big) dx.$$

<u>Theorem 3.10</u> Let X be a Poisson center cluster random measure. Let A be a bounded Borel subset of $\mathbb{R}^d$. Then $\frac{X_\lambda(A)}{\lambda^d}$ satisfies a large deviation principle with rate function

$$I_X(x) = \sup_{t \in \mathbb{R}} \Big\{ tx + \rho|A| - \rho|A| M_{V(\mathbb{R}^d)}(t) \Big\}.$$

<u>Theorem 3.11</u> Let X be a Poisson center cluster random measure. Let $\underset{\sim}{Y}_\lambda = \Big(X_\lambda(A_1), \ldots, X_\lambda(A_n)\Big)$, where all A_i's are disjoint and bounded Borel subsets of $\mathbb{R}^d$. Then $\frac{\underset{\sim}{Y}_\lambda}{\lambda^d}$ satisfies a large deviation principle with rate function

$$I_Y(\underset{\sim}{x}) = \sup_{\underset{\sim}{t} \in \mathbb{R}^n} \Big\{ <\underset{\sim}{t}, \underset{\sim}{x}> + \sum_{i=1}^{n} \rho|A_i| - \sum_{i=1}^{n} \rho|A_i| M_{V(\mathbb{R}^d)}(\underset{\sim}{t}) \Big\}.$$

Proof. In Proposition 3.9, take $f = \sum_{i=1}^{n} t_i 1_{A_i}$, where all A_i's are disjoint.

$$\Phi(\underset{\sim}{t}) = \lim_{\lambda\to\infty} \frac{1}{\lambda^d}\cdot\log M_{\underset{\sim}{Y}_\lambda}(\underset{\sim}{t})$$

$$= \rho \int_{\mathbb{R}^d} E[\exp\Big\{\sum_{i=1}^{n} t_i 1_{A_i}(x) V(\mathbb{R}^d)\Big\} - 1]dx$$

$$= \rho \sum_{i=1}^{n} \int_{A_i} E[\exp\big(t_i V(\mathbb{R}^d)\big) - 1]dx$$

$$= \rho \sum_{i=1}^{n} |A_i|\big(M_{V(\mathbb{R}^d)}(t_i) - 1\big).$$

So, $$I_Y(\underset{\sim}{x}) = \sup_{\underset{\sim}{t}\in\mathbb{R}^n}\Big\{<\underset{\sim}{t}, \underset{\sim}{x}> - \rho\sum_{i=1}^{n}|A_i|\cdot E[\exp\big(t_i V(\mathbb{R}^d)\big) - 1]\Big\}.$$

Since $M_{V(\mathbb{R}^d)}(t_i)$ is differentiable for all $t_i \in \mathbb{R}$, $\Phi(\underset{\sim}{t})$ is also differentiable for all $\underset{\sim}{t} \in \mathbb{R}^n$. Therefore, $\underset{\sim}{Y}_\lambda/\lambda^d$ satisfies a large deviation principle with rate function $I_Y(\cdot)$ from Theorem II.6.1., Ellis [12].

3.4. Doubly Stochastic Process

Let Z be a stationary, associated random measure and let X_Z be a doubly stochastic process with environment Z. that is to say, X_Z is Z-conditionally Poisson with intensity measure Z. Of course, the following hypothese are assumed to hold for Z ;

(i) $\Phi_\lambda(\underset{\sim}{t}) = \frac{1}{\lambda^d}\cdot\log E[e^{<\underset{\sim}{t}, \underset{\sim}{Z}_\lambda>}]$ is finite for all $\underset{\sim}{t} \in \mathbb{R}^n$,

where $\underset{\sim}{Z}_\lambda = \big(Z_\lambda(A_1), \ . \ . \ . \ , \ Z_\lambda(A_n)\big)$ and all A_i's are disjoint and bounded subsets of $\mathbb{R}^d$,

and

(ii) $\Phi(\underset{\sim}{t}) = \lim_{\lambda \to \infty} \Phi_\lambda(\underset{\sim}{t})$ exists and is differentiable for all $\underset{\sim}{t} \in \mathbb{R}^n$.

Now, for a doubly stochastic process X_Z, define that

$$\Phi_{X_Z}(\underset{\sim}{t}) = \lim_{\lambda \to \infty} \frac{1}{\lambda^d} \cdot \log E[e^{<\underset{\sim}{t},\, (\underset{\sim}{Y}_Z)_\lambda>}] \text{ exists for } \underset{\sim}{t} \in \mathbb{R}^n,$$

where $(\underset{\sim}{Y}_Z)_\lambda = \big((X_Z)_\lambda(A_1), \ldots, (X_Z)_\lambda(A_n)\big)$ and all A_i's are disjoint and bounded subsets of $\mathbb{R}^d$,

and

$$I(\underset{\sim}{x}) = \sup_{\underset{\sim}{t} \in \mathbb{R}^n} \Big\{<\underset{\sim}{t}, \underset{\sim}{x}> - \Phi_{X_Z}(\underset{\sim}{t})\Big\}.$$

Then we will have the following theorem.

Theorem 3.12 Let X_Z be a doubly stochastic process with environment Z. Let $(\underset{\sim}{Y}_Z)_\lambda$ be a random vector defined as above. Also we assume the above hypotheses (i) and (ii). Then $\frac{(\underset{\sim}{Y}_Z)_\lambda}{\lambda^d}$ satisfies a large deviation principle with rate function

$$I(\underset{\sim}{x}) = \sup_{\underset{\sim}{t} \in \mathbb{R}^n} \Big\{<\underset{\sim}{t}, \underset{\sim}{x}> - \Phi_Z \circ H(\underset{\sim}{t})\Big\},$$

where $H(\underset{\sim}{t}) = (e^{t_1} - 1, \ldots, e^{t_n} - 1)$.

Proof. First calculate the moment generating function of $(\underset{\sim}{Y}_Z)_\lambda$.

$$E[e^{<\underset{\sim}{t},\, (\underset{\sim}{Y}_Z)_\lambda>}] = E[e^{t_1 \cdot (X_Z)_\lambda(A_1) + \cdots + t_n \cdot (X_Z)_\lambda(A_n)}]$$

$$= E[E[e^{X_Z(\sum_{i=1}^{n} t_i 1_{\lambda A_i})} \mid Z]]$$

$$= E[e^{\int (e^f - 1) Z(dx)}], \text{ where } f = \sum_{i=1}^{n} t_i 1_{\lambda A_i}$$

$$= E[e^{Z_\lambda (e^f - 1)}].$$

So, $\Phi_{X_Z}(\underset{\sim}{t}) = \lim_{\lambda \to \infty} \frac{1}{\lambda^d} \cdot \log E[e^{Z_\lambda (e^f - 1)}]$

$$= \lim_{\lambda \to \infty} \frac{1}{\lambda^d} \cdot \log E[e^{(e^{t_1} - 1) Z_\lambda(A_1) + \ldots + (e^{t_n} - 1) Z_\lambda(A_n)}].$$

Since $\Phi_Z(\underset{\sim}{t}) = \lim_{\lambda \to \infty} \frac{1}{\lambda^d} \cdot \log E[e^{t_1 Z_\lambda(A_1) + \ldots + t_n Z_\lambda(A_n)}]$ exists for

every $\underset{\sim}{t} \in \mathbb{R}^n$, $\Phi_{X_Z}(\underset{\sim}{t}) = \Phi_Z(e^t - 1, \ldots, e^t - 1)$ and $\Phi_{X_Z}(\underset{\sim}{t}) = \Phi_Z(H(\underset{\sim}{t})) = \Phi_Z \circ H(\underset{\sim}{t})$.

Moreover, $\Phi_{X_Z}(\underset{\sim}{t})$ is also differentiable for each $\underset{\sim}{t} \in \mathbb{R}^n$ because $H(\underset{\sim}{t})$ is differentiable for every $\underset{\sim}{t} \in \mathbb{R}^n$ and also the composition function of two differentiable functions is differentiable.

Therefore, $\frac{(\underset{\sim}{Y}_Z)_\lambda}{\lambda^d}$ satisfies a large deviation principle with rate function $I(\cdot)$ by Theorem 3.3.

4. LARGE DEVIATION PROPERTIES OF RANDOM MEASURE

4.1. Introduction

Recall that we have already defined a random measure on a locally compact Polish space S in Section 2.1.. Moreover, we have discussed random measures on $\mathbb{R}^d$ in Section 3.1..

From now on, we denote the collection of Radon measures on $\mathbb{R}$ by $M(\mathbb{R})$ (write $M = M(\mathbb{R})$). Here we shall give large deviation properties of random measures with values in $M(\mathbb{R})$. To do this, we will use the restriction of measures on $M(\mathbb{R})$ to $[0, L] \subseteq \mathbb{R}$. We will denote by $M[0, L]$ the collection of Radon measures on $[0, L]$. We will put the topology of weak convergence on $M[0, L]$.

Throughout this Chapter IV, we shall show the upper large deviation bound and the lower large deviation bound about random measures considered as "dual" to an appropriate space of test functions as follows :

Let X be a random measure in M. Let X^L be the random measure in $M[0, L]$ obtained by restricting X to $[0, L]$. We denote $\tilde{X}_\lambda = (X_\lambda)^L$. We will show that $\tilde{X}_\lambda/\lambda$ satisfies a large deviation principle, which will be given in details in Definition 4.4, if for all $L > 0$ we have the following properties :

(*Upper Large Deviation Bound*); For any closed subset F in $M[0, L]$,

$$\limsup_{\lambda\to\infty} \frac{1}{\lambda}\cdot\log P\left(\frac{\tilde{X}_\lambda}{\lambda} \in F\right) \le -I_L(F).$$

(*Lower Large Deviation Bound*); For any open subset G in M[0, L],

$$\liminf_{\lambda\to\infty} \tfrac{1}{\lambda}\cdot\log P\left(\frac{\tilde{X}_\lambda}{\lambda} \in G\right) \geq -I_L(G).$$

We denote $I_L(A) = \inf_{\mu\in A} I_L(\mu)$ for a Borel set A and

$$I_L(\mu) = \sup_{f\in C[0,L]} \left\{\mu(f) - \lim_{\lambda\to\infty} \tfrac{1}{\lambda}\cdot\log E[\exp(\tilde{X}_\lambda(f))]\right\}.$$

Later, we will show that if the above property is satisfied for some $L > 0$, then it is satisfied for every $L' > 0$.

In particular, we shall give some examples which satisfy those properties.

Let $\tilde{X}_\lambda$ $\left(\text{recall } \tilde{X}_\lambda = (X_\lambda)^L\right)$ be a random measure restricted to M[0, L] which will be given by Definition 4.2.

Define

$$\Phi_{\tilde{X}_\lambda}(f) = \tfrac{1}{\lambda}\cdot\log E[e^{\tilde{X}_\lambda(f)}] \text{ for } f \in C[0, L],$$

Here, recall that we define f_λ and X_λ in (3.2) by $f_\lambda(x) = f(\frac{x}{\lambda})$ and $X_\lambda(f) = X(f_\lambda)$.

The following two hypotheses are assumed to hold ;

(4.1)

(a) Each function $\Phi_{\tilde{X}_\lambda}(f)$ is finite for every $f \in C[0, L]$,

(b) $\Phi_{\tilde{X}}(f) = \lim_{\lambda\to\infty} \Phi_{\tilde{X}_\lambda}(f)$ exists for every $f \in C[0, L]$ and is also finite.

Here we shall denote $\Phi(f) := \Phi_{\tilde{X}}(f)$ and $\Phi_\lambda(f) := \Phi_{\tilde{X}_\lambda}(f)$.

Proposition 4.1 The above two functions $\Phi_\lambda(f)$ and $\Phi(f)$ are convex on C[0, L].

Proof. Let f and g in C[0, L] and $0 < t < 1$.

$$\Phi_\lambda\{t\cdot f + (1-t)\cdot g\} = \tfrac{1}{\lambda}\cdot\log E[e^{\tilde{X}_\lambda\{t\cdot f + (1-t)\cdot g\}}]$$

Now applying the Hölder's inequality with $p = \frac{1}{t}$ and $q = \frac{1}{1-t}$, we get the following inequality

$$\leq \tfrac{1}{\lambda}\cdot\log\left\{E[e^{\tilde{X}_\lambda(f)}]\right\}^t\cdot\left\{E[e^{\tilde{X}_\lambda(g)}]\right\}^{(1-t)}$$

$$= t\cdot\Phi_\lambda(f) + (1-t)\cdot\Phi_\lambda(g).$$

Thus, $\Phi_\lambda(f)$ is convex and $\Phi(f)$ is also convex because of preservation of convexity under pointwise limits.

Definition 4.2 M[0, L] is the collection of Radon measures on the restricted [0, L]. That is to say, $M[0, L] = \{\mu : \text{Borel finite measure on } [0, L]\}$ for every $L < \infty$.

Now, define for $\mu \in M[0, L]$

$$(4.2) \qquad I_L(\mu) = \sup_{f\in C[0, L]}\left\{\mu(f) - \Phi(f)\right\}$$

In the future, we shall prove that $I_L(\mu)$ defined in (4.2) is a rate function defined as below.

Definition 4.3 A function $I_L(\cdot)$ on M[0, L] is said to be a rate function if

(a) $I_L(\mu)$ is non-negative,

(b) $I_L(\cdot)$ is lower semi-continuous,

and

(c) I_L has compact level sets, that is, $\{\mu \in M[0, L] : I_L(\mu) < r\}$ is compact for every $r < \infty$.

Recall that we have put the weak topology on M[0, L].

Definition 4.4 Let X be a random measure and $(X_\lambda)^L$ the random measure in M[0, L] obtained by restricting X to [0, L]. Let $\tilde{X}_\lambda = (X_\lambda)^L$. Let P_λ be the distribution of $\frac{\tilde{X}_\lambda}{\lambda}$. It is said that the measures $\{P_\lambda\}$ satisfy a large deviation principle with rate function $I_L(\cdot)$ if

(a) $I_L(\cdot)$ is a rate function,

(b) (Upper bound) ; For each closed subset F in M[0, L],

$$\limsup_{\lambda\to\infty} \frac{1}{\lambda}\cdot\log P_\lambda(F) \le -\inf_{\mu\in F} I_L(\mu),$$

(c) (Lower bound) ; For each open subset G in M[0, L],

$$\liminf_{\lambda\to\infty} \frac{1}{\lambda}\cdot\log P_\lambda(G) \ge -\inf_{\mu\in G} I_L(\mu),$$

where $P_\lambda(A)$ denotes the probability of $\{\frac{\tilde{X}_\lambda}{\lambda} \in A\}$ for any Borel subset A in M[0, L].

4.2. On the Upper Bound for Large Deviation on Random Measure

Throughout this section, we follow the same notations as in the previous sections.

In this section, we obtain a upper bound for the large deviation property for certain classes of random measures. One of the situations we study is as follows. We follow the notation $\tilde{X}_\lambda$ given in Definition 4.1. We obtain the upper bound of random measures of the type :

$$\limsup_{\lambda\to\infty} \frac{1}{\lambda}\cdot\log P\left(\frac{\tilde{X}_\lambda}{\lambda} \in F\right) \le -I_L(F) \text{ for any closed subset F in M[0, L].}$$

Here, recall that $I_L(\mu) = \sup_{f \in C[0,L]} \left\{\mu(f) - \Phi(f)\right\}$ defined in (4.2)

and $$I_L(F) = \inf\left\{I_L(\mu) : \mu \in F\right\}.$$

Let us first check that $I_L(\mu)$ is a rate function.

Theorem 4.5 The hypotheses (4.1) (a)-(b) are assumed to hold. The following conclusions hold ;

(a) $I_L(\mu)$ is non-negative.

(b) $I_L(\mu)$ is a lower semi-continuous convex function of M[0, L].

(c) Given every real $r < \infty$, the set $\left\{\mu \in M[0, L] : I_L(\mu) < r\right\}$ is compact in the weak topology.

Remark 4.6 Recall that a function such as $I_L(\cdot)$ defined on M[0, L] which satisfies (a) – (c) in Theorem 4.5 is a rate function given in Definition 4.3.

Proof of Theorem 4.5 (a) Let us take $f = 0$. Then $\mu(f) - \Phi(f) = 0$. Thus, $I_L(\mu) \geq 0$. This implies that $I_L(\mu) \in [0, \infty]$.

(b) Since $\mu_n \to \mu$ if and only if $\mu_n(f) \to \mu(f)$ for $f \in C[0, L]$, $\mu(f)$ is continuous for μ. So, $\mu(f) - \Phi(f)$ is continuous. Since sup of continuous function is lower semi-continuous, $I_L(\mu)$ is also lower semi-continuous. Obviously $I_L(\mu)$ is convex. Thus, $I_L(\mu)$ is a lower semi-continuous and convex function.

(c) Let $K_r = \left\{\mu \in M[0, L] : I_L(\mu) < r\right\}$. Since $C[0, L]$ is separable, let $f_0, f_1, \ldots$ be countable dense subsets. Take $f_0 = 1_{[0, L]}$.

For any sequence $\{\mu_k\}$ in K_r and f_n in $C[0, L]$, $\mu_k(f_n) \leq r + \Phi(f_n)$. By diagonal procedure, there exists a subsequence $\{\mu_{k_i}\}$ of $\{\mu_k\}$ such that $\lim_{i\to\infty}\int f_n\mu_{k_i}$ exists for $n = 0, 1, \ldots$. Since $\lim_{i\to\infty}\int f_0\mu_{k_i}$ exists, let the value denote by $\mu_{k\infty}([0, L])$. That is, there exists $\delta > 0$ such that for k_∞ big enough, $|\mu_{k\infty}([0, L]) - \lim_{i\to\infty}\int 1_{[0, L]}d\mu_{k_i}| < \delta$. For every continuous function f_n in $C[0, L]$ and $\epsilon > 0$, $\lim_{i\to\infty}\int f_n\mu_{k_i}$ exists and $\| f_n - f \| \leq \epsilon$.

$$\lim_{i\to\infty}\int fd\mu_{k_i} = \lim_{i\to\infty}\int(f - f_n)d\mu_{k_i} + \lim_{i\to\infty}\int f_n\, d\mu_{k_i}$$

$$\leq \epsilon\cdot\lim_{i\to\infty}\int 1_{[0, L]}\, d\mu_{k_i} + \lim_{i\to\infty}\int f_n\, d\mu_{k_i}\,.$$

So,

(1) $$\lim_{i\to\infty}\int fd\mu_{k_i} \leq \epsilon\Big(\mu_{k\infty}([0, L])\Big) + \liminf_{n}\lim_{i\to\infty}\int f_n d\mu_{k_i}.$$

Likewise, replaceing f_n by f,

(2) $$\limsup_{n}\lim_{i\to\infty}\int f_n\, d\mu_{k_i} \leq \epsilon\mu_{k\infty}([0, L]) + \lim_{i\to\infty}\int fd\mu_{k_i}.$$

Letting $\epsilon \to 0$ from (1) and (2), we get

$$\lim_{n} \sup \lim_{i \to \infty} \int f_n d\mu_{k_i} \leq \lim_{i \to \infty} \int f d\mu_{k_i} \leq \lim_{n} \inf \lim_{i \to \infty} \int f_n d\mu_{k_i}.$$

So, the limit exists. That is to say,

$$\lim_{n} \lim_{i \to \infty} \int f_n d\mu_{k_i} = \lim_{i \to \infty} \int f d\mu_{k_i} \text{ and } || f - f_n || \leq \epsilon$$

implies that

$$| \lim_{i \to \infty} \int f_n d\mu_{k_i} - \lim_{i \to \infty} \int f d\mu_{k_i} | \leq \epsilon.$$

Define $\Lambda(f) := \lim_{i \to \infty} \int f d\mu_{k_i}$.

Λ is obviously linear because of the linearity of an integral, i.e.

$$\Lambda(f + g) = \lim_{i \to \infty} \int (\alpha f + g) d\mu_{k_i}$$

$$= \alpha \lim_{i \to \infty} \int f d\mu_{k_i} + \lim_{i \to \infty} \int g d\mu_{k_i}$$

$$= \alpha\Lambda(f) + \Lambda(g) \text{ for } \alpha \in R \text{ and } f, g \in C[0, L].$$

$\Lambda(f)$ defines a non-negative linear functional on C[0, L]. By the Markov-Riesz representation theorem, the limiting linear function is represented as an integral and $\Lambda(f) = \int f d\mu$.

So,

$$\Lambda(f) = \int f d\mu = \lim_{i \to \infty} \int f d\mu_{k_i}.$$

This means that μ_{k_i} converges to μ weakly. Therefore, K_T is compact.

Definition 4.7 It is said that P_λ is large deviation tightness if for each $a < \infty$, there exists a compact set K_a such that $\limsup_{\lambda\to\infty} \frac{1}{\lambda}\cdot\log P\left(\frac{\tilde{X}_\lambda}{\lambda} \in K_a{}^c\right) \le -a$,

where $P\left(\frac{\tilde{X}_\lambda}{\lambda} \in K_a{}^c\right) = P_\lambda(K_a{}^c)$.

Theorem 4.8 If $\frac{\tilde{X}_\lambda}{\lambda}[0, L]$ is large deviation tightness as random variables for some $L > 0$, then $\frac{\tilde{X}_\lambda}{\lambda}$ is also large deviation tightness.

Proof. For every $a \in [0, \infty)$, let $K_a = \left\{\frac{\tilde{X}_\lambda}{\lambda} : \frac{\tilde{X}_\lambda}{\lambda}[0, L] \le a\right\}$. Then K_a is compact in the weak topology by the Banach-Alaoglu theorem. Let P_λ be the distribution of $\frac{\tilde{X}_\lambda}{\lambda}$.

So, $P_\lambda\left(K_a{}^c\right) = P\left(\frac{\tilde{X}_\lambda}{\lambda} \in K_a{}^c\right) = P\left(\frac{\tilde{X}_\lambda}{\lambda} : \frac{\tilde{X}_\lambda}{\lambda}[0, L] > a\right) = P\left(\frac{\tilde{X}_\lambda}{\lambda}[0, L] > a\right)$.

Thus,

$$\limsup_{\lambda\to\infty} \frac{1}{\lambda}\cdot\log P\left(\frac{\tilde{X}_\lambda}{\lambda} \in K_a{}^c\right) = \limsup_{\lambda\to\infty} \frac{1}{\lambda}\cdot\log P\left(\frac{\tilde{X}_\lambda}{\lambda}[0, L] > a\right) \le -a.$$

Therefore $\frac{\tilde{X}_\lambda}{\lambda}$ is large deviation tightness.

Theorem 4.9 Let $\mathbb{D}(\Phi) := \left\{f \in C[0, L] : \Phi(f) < \infty\right\}$. Given $f \in \mathbb{D}(\Phi)$ and α real, define

$$H_+(f,\alpha) = \left\{\mu \in M[0, L] : \mu(f) - \Phi(f) \ge \alpha\right\}$$

and

$$H_-(f,\alpha) = \left\{\mu \in M[0, L] : \mu(f) - \Phi(f) \le \alpha\right\}.$$

Let K be a compact subset of M[0, L].

(a) If $0 < I_L(K) < \infty$, then for any $\epsilon \in (0, I_L(K))$ there exist finitely many points $f_1, \ldots, f_n$ in $\mathbb{D}(\Phi)$ such that

$$K \subseteq \bigcup_{k=1}^{n} H_+\left(f_k, I_L(K) - \epsilon\right)$$

(b) If $I_L(K) = \infty$, then for any $R > 0$ there exist finitely many points $f_1, \ldots, f_n$ in $\mathbb{D}(\Phi)$ such that

$$K \subseteq \bigcup_{k=1}^{n} H_+(f_k, R),$$

where $I_L(K) = \inf_{\mu \in K} I_L(\mu)$.

Proof. Here we prove only the case (a). That is, $0 < I_L(K) < \infty$ since we can do the case (b) similarly as in (a).

Let $K_* = \left\{\mu \in M[0, L] : I_L(\mu) \leq I_L(K) - \epsilon\right\}$.

K_* is closed since $I_L(\mu)$ is lower semi-continuous. In fact K_* is compact by Theorem 4.5 (c).

Our claim is $\quad K_* = \bigcap_{f \in \mathbb{D}(\Phi)} H\text{-}\left(f, I_L(K) - \epsilon\right)$.

To show this claim, let $\mu \in K_*$. Then $I_L(\mu) \leq I_L(\mu) - \epsilon$ by definition. Thus, $\mu(f) - \Phi(f) \leq I_L(K) - \epsilon$ from the definition of $I_L(\mu)$. Hence, $\mu \in H\text{-}\left(f, I_L(K) - \epsilon\right)$ for every $f \in \mathbb{D}(\Phi)$.

Conversely, let $\nu \in H\text{-}\left(f, I_L(K) - \epsilon\right)$ for every $f \in \mathbb{D}(\Phi)$. Thus, $\nu(f) - \Phi(f) \leq I_L(K) - \epsilon$. Taking sup on both sides with respect to $f \in C[0, L]$, we get

$$\sup_{f \in C[0,L]} \{\nu(f) - \Phi(f)\} \le I_L(K) - \epsilon. \text{ That is, } I_L(\nu) \le I_L(K) - \epsilon.$$

This means that $\nu \in K_*$. Therefore our claim is proved.

Note that $K \bigcap K_* = \emptyset$. Since, if not, we can take $\mu \in K_*$ and $\mu \in K$. This implies that $I_L(\mu) \le I_L(K) - \epsilon \le I_L(\mu) - \epsilon$. This shows a contradiction.

So,
$$K \subseteq {K_*}^c = \left\{ \bigcap_{f \in \mathbb{D}(\Phi)} H\text{-}\big(f,\ I_L(K) - \epsilon\big)\right\}^c$$

$$= \bigcup_{f \in \mathbb{D}(\Phi)} \text{int } H_+\big(f, I_L(K) - \epsilon\big).$$

Since $K \subseteq {K_*}^c$, for each $\mu \in K$, there exists an open neighborhood $N(\mu)$ of μ and a point $f \in \mathbb{D}(\Phi)$ such that $N(\mu) \subseteq H_+\big(f, I_L(K) - \epsilon\big)$. Since K is compact, the case (a) is proved.

Theorem 4.10 Let P_λ be the distribution of $\frac{\tilde{X}_\lambda}{\lambda}$.

(a) For any compact subset K of M[0, L],

$$\limsup_{\lambda \to \infty} \frac{1}{\lambda} \cdot \log P_\lambda(K) \le -I_L(K).$$

(b) If P_λ satisfies large deviation tightness, then for any closed subset F of M[0, L]

$$\limsup_{\lambda \to \infty} \frac{1}{\lambda} \cdot \log P_\lambda(F) \le -I_L(F).$$

Proof. (a) In the case of $I_L(K) = 0$, it is obvious.

So, we prove only the case of $0 < I_L(K) < \infty$ because we can prove the case of $I_L(K) = \infty$ similarly by Theorem 4.9 (b).

$$P_\lambda(K) = P\left(\frac{\tilde{X}_\lambda}{\lambda} \in K\right)$$

$$\leq \sum_{k=1}^{n} P\left(\frac{\tilde{X}_\lambda}{\lambda} \in H_+\left(f_k,\ I_L(K) - \epsilon\right)\right)$$

$$= \sum_{k=1}^{n} P\left(\frac{\tilde{X}_\lambda}{\lambda}(f_k) \geq \Phi(f_k) + I_L(K) - \epsilon\right)$$

$$\leq \sum_{k=1}^{n} \frac{E[e^{\tilde{X}_\lambda(f_k)}]}{e^{\lambda\left(\Phi(f_k) + I_L(K) - \epsilon\right)}} \quad \text{by Chebyshev's inequality}$$

Here, applying the fact that $\Phi_\lambda(f) = \frac{1}{\lambda}\cdot\log\ E[e^{\tilde{X}_\lambda(f)}]$ implies $E[e^{\tilde{X}_\lambda(f)}] = \exp\left(\lambda\Phi_\lambda(f)\right)$, then we get the following equality from the last term ;

$$= \sum_{k=1}^{n} e^{\lambda\left\{\Phi_\lambda(f_k) - \Phi(f_k) - I_L(K) + \epsilon\right\}}.$$

Since $\Phi_\lambda(f_k) \to \Phi(f_k)$ as $\lambda \to \infty$ and $\epsilon \in \left(0,\ I_L(K)\right)$ is arbitrary,

$$P_\lambda(K) = P\left(\frac{\tilde{X}_\lambda}{\lambda} \in K\right)$$

$$\leq ne^{-\lambda\cdot I_L(K)}.$$

So, taking the logarithm on both sides, we have that

$$\frac{1}{\lambda}\cdot\log P_\lambda(K) \leq \frac{\log(n)}{\lambda} - I_L(K)$$

Therefore $\limsup_{\lambda\to\infty} \frac{1}{\lambda}\cdot\log P_\lambda(K) \leq -I_L(K)$.

(b) Since P_λ is large deviation tightness, there exists a compact set K for every real $a < \infty$.

$$P_\lambda(F) = P_\lambda(K_a \cap F) + P_\lambda(K_a{}^c \cap F)$$

$$\leq 2\max\left\{P_\lambda(K_a \cap F),\ P_\lambda(K_a{}^c \cap F)\right\}$$

$$\leq 2\max\left\{P_\lambda(K_a \cap F),\ P_\lambda(K_a{}^c)\right\}$$

$$\leq 2\max\left\{P_\lambda(K_a \cap F),\ e^{-\lambda a}\right\},$$

Since $\frac{1}{\lambda}\cdot\log P_\lambda(K_a{}^c) \leq -a$ implies $P_\lambda(K_a{}^c) \leq e^{-\lambda a}$, we get the following inequality from the last term ;

$$\leq 2\max\left\{e^{-\lambda I_L(K_a \cap F)},\ e^{-\lambda a}\right\}.$$

Applying the above property (a) in the last term since $K_a \cap F$ is compact, we have the following inequality

$$\leq 2\max\left\{e^{-\lambda I_L(F)},\ e^{-\lambda a}\right\}.$$

So, the above inequality implies that

$$\frac{1}{\lambda}\cdot\log P_\lambda(F) \leq \frac{\log(2)}{\lambda} + \max\left\{-\ I_L(F),\ -\ a\right\}.$$

Since a is arbitrary, we have that

$$\limsup_{\lambda\to\infty} \frac{1}{\lambda}\cdot\log P_\lambda(F) \leq -I_L(F).$$

Now, we will show that if the large deviation upper bound holds on M[0, L] for some $L > 0$, then such property also holds for every $L' > 0$.

Theorem 4.11 If $\frac{\tilde{X}_\lambda}{\lambda}$ satisfies a large deviation upper bound for some $L > 0$, then $\frac{\tilde{X}_\lambda{}'}{\lambda}$ also satisfies a large deviation upper bound for every $L' > 0$, where we denote $\tilde{X}_\lambda{}' = \left(X_\lambda \right)^{L'}$.

Sketch of Proof. First, let us check the relation between $I_{L'}$ and I_L.

For every $g \in C[0, L']$, there exists $f \in C[0, L]$ such that $g(y) = \tilde{f}(y) = f_{L'/L}(y)$ for every $y \in [0, L']$. Also for every $\nu \in M[0, L']$, there exists $\mu \in M[0, L]$ such that $\tilde{\mu} = \nu = \mu_{L/L'}$. Now, calculating $\Phi_{L'}(g)$ for every $g \in C[0, L']$;

we have the relation, $\Phi_{L'}(g) = \frac{L'}{L}\Phi_L\left(g_{L/L'}\right)$. That is, $\frac{1}{L'}\Phi_{L'}\left(g_{L'}\right) = \frac{1}{L}\Phi_L\left(g_L\right)$.

Secondly, calculating $I_{L'}(\tilde{\mu})$ for every $\tilde{\mu} \in M[0, L']$ by

$$I_{L'}(\tilde{\mu}) \quad = \quad \sup_{\tilde{f} \in C[0, L']} \left\{ \tilde{\mu}(\tilde{f}) - \Phi_{L'}(\tilde{f}) \right\},$$

we have the relation, $I_{L'}(\nu) = \frac{L'}{L} I_L\left(\frac{L}{L'}(\nu)_{L'/L}\right)$ for every $\nu \in M[0, L']$.

Finally, it remains to show that for every closed subset F' in $M[0, L']$,

$$\limsup_{\lambda \to \infty} \frac{1}{\lambda} \cdot \log P\left(\frac{\tilde{X}_\lambda{}'}{\lambda} \in F'\right) \le -I_{L'}(F').$$

Let F' be a closed subset in $M[0, L']$. Let $F'_{L'/L} = \left\{ \nu_{L'/L} : \nu \in F' \right\}$, which is also the closed subset in M[0, L].

$$\limsup_{\lambda\to\infty} \frac{1}{\lambda}\cdot\log P\left(\frac{\tilde{X}_\lambda{}'}{\lambda} \in F'\right) = \frac{L'}{L}\cdot\limsup_{\lambda\to\infty} \frac{1}{\frac{L'}{L}\lambda}\cdot\log P\left(\frac{\left(\tilde{X}_\lambda{}'\right)_{L'/L}}{\lambda\frac{L'}{L}} \in \frac{L}{L'}F'_{L'/L}\right)$$

$$\leq -\frac{L'}{L}\cdot I_L\left(\frac{L}{L'}F'_{L'/L}\right)$$

$$= -I_{L'}\left(F'\right).$$

The last equality holds since

$$I_{L'}\left(F'\right) = \inf\left\{I_{L'}(\nu) : \nu \in F'\right\}$$

$$= \inf\left\{\frac{L'}{L}\cdot I_L\left(\frac{L}{L'}\nu_{L'/L}\right) : \nu \in F'\right\}$$

$$= \frac{L'}{L}\inf\left\{I_L(\mu) : \mu \in \frac{L}{L'}F'_{L'/L}\right\}, \text{ where } \mu = \frac{L}{L'}\nu_{L'/L} \in \frac{L}{L'}F'_{L'/L}$$

$$= \frac{L'}{L}I_L\left(\frac{L}{L'}F'_{L'/L}\right).$$

Therefore $\frac{\tilde{X}_\lambda{}'}{\lambda}$ satisfies a large deviation upper bound for every $L' > 0$.

4.3 Some Examples for the Upper Bound

1. Poisson random measure : Let X be a Poisson random measure with intensity 1 (W.L.O.G., assume intensity $\rho = 1$ by rescaling).

For an arbitrarily real $a < \infty$,

$$\text{let } K_a = \left\{\frac{\tilde{X}_\lambda}{\lambda} : \frac{\tilde{X}_\lambda}{\lambda}[0, L] \leq 2a\right\}.$$

Then K_a is compact. Now let P_λ be the distribution of $\frac{\tilde{X}_\lambda}{\lambda}$.

So,

$$P_\lambda\left(K_a{}^c\right) = P\left(\frac{\tilde{X}_\lambda}{\lambda} \in K_a{}^c\right)$$

$$= P\left(\frac{\tilde{X}_\lambda}{\lambda}[0, L] > 2a\right)$$

$$\leq \frac{\exp\left(\lambda L(e^t - 1)\right)}{\exp(2\lambda at)} \text{ by Markov's inequality}$$

$$= e^{-\lambda\left(2a - L(e - 1)\right)} \text{ replacing } t = 1.$$

Now, taking $a \geq L(e - 1)$, we have that

$$\limsup_{\lambda\to\infty} \tfrac{1}{\lambda}\cdot\log P\left(\frac{\tilde{X}_\lambda}{\lambda}[0, L] > 2a\right) \leq -a.$$

Thus,
$$\limsup_{\lambda\to\infty} \tfrac{1}{\lambda}\cdot\log P_\lambda\left(K_a{}^c\right) \leq -a.$$

This means that P_λ satisfies large deviation tightness by Definition 4.7. Therefore, Theorem 4.10 (b) implies that $\frac{\tilde{X}_\lambda}{\lambda}$ satisfies a large deviation upper bound.

2. Poisson center cluster process : Let U be a Poisson random field with parameter ρ and V a random measure with finite expected total mass, $E[V(\mathbb{R})] = \gamma$. Let U have occurrences $\{u_i\}$ and let $\{v_i\}$ be independent and identically distributed (i.i.d.) random measures independent of U and distributed as V. We denote a Poisson center cluster process by $X = [U, V]$. More precisely, a random measure X is defined by letting the occurrences of U act as centers or initiators and then

superimposing i.i.d. random measure distributed as V but centered at the occurrences of U. For a bounded subset $A \in \mathfrak{B}$, we denote

$$X(A) = \sum_i V_i(A - X_i).$$

From Section 3.3., we have already showed that

$$M_{X_\lambda(A)}(t) = \exp\left(\rho \int E[e^{tV(\lambda A - x)} - 1]dx\right).$$

First we will show that $M_{X_\lambda(A)}(t) \le \exp\left(\lambda\rho|A|\left(M_{V(\mathbb{R})}(t) - 1\right)\right)$.

$$M_{X_\lambda(A)}(t) = e^{\rho\int E[e^{tV(\lambda A - x)} - 1]dx}$$

$$= e^{\rho\lambda\int E[e^{tV(\lambda A - \lambda y)} - 1]dy}$$

$$= e^{\rho\lambda\sum_{k=1}^{\infty}\frac{t^k}{k!} E[\int\left(V(\lambda A - \lambda y)\right)^k]dy}$$

$$= e^{\rho\lambda\sum_{k=1}^{\infty}\frac{t^k}{k!}E[\int\cdots\int 1_{\lambda A - \lambda y}(x_1)\ldots 1_{\lambda A - \lambda y}(x_k)dydV(x_1)\ldots dV(x_k)]}$$

$$= e^{\rho\lambda\sum_{k=1}^{\infty}\frac{t^k}{k!}E[\int\cdots\int 1_{A - \frac{x_1}{\lambda}}(y)\ \ldots\ 1_{A-\frac{x_k}{\lambda}}(y)dy\ dV(x_1)\ \ldots\ dV(x_k)]}$$

$$= e^{\lambda\rho\sum_{k=1}^{\infty}\frac{t^k}{k!}E[\int\cdots\int |(A - \frac{x_1}{\lambda})\ \ldots(A - \frac{x_k}{\lambda})|dV(x_1)\ \ldots dV(x_k)}$$

$$\le e^{\lambda\rho\sum_{k=1}^{\infty}\frac{t^k}{k!}E[\int\cdots\int |A - \frac{x_1}{\lambda}|dV(x_1)\ \ldots dV(x_k)}$$

$$= e^{\lambda\rho\sum_{k=1}^{\infty}\frac{t^k}{k!}E[\int\cdots\int|A|dV(x_1)\ldots dV(x_k)}$$

$$= e^{\lambda\rho|A|E[\sum_{k=1}^{\infty}\frac{t^k}{k!}[V(\mathbb{R})]^k]}$$

$$= e^{\lambda\rho|A|E[e^{tV(\mathbb{R})} - 1]}$$

$$= e^{\lambda\rho|A|[M_{V(\mathbb{R})}(t) - 1]}.$$

Now, let $K_{2a} = \left\{\frac{\tilde{X}_\lambda}{\lambda} : \frac{\tilde{X}_\lambda}{\lambda}[0, L] \leq 2a\right\}$. Let P_λ be the distribution of $\frac{\tilde{X}_\lambda}{\lambda}$.

$$P_\lambda\left({K_{2a}}^c\right) = P\left(\frac{\tilde{X}_\lambda}{\lambda} \in {K_{2a}}^c\right)$$

$$= P\left(\frac{\tilde{X}_\lambda}{\lambda}[0, L] > 2a\right)$$

$$= P\left(\tilde{X}_\lambda[0, L] > 2\lambda a\right)$$

$$\leq \frac{E[\exp\left(\tilde{X}_\lambda[0, L]\right)]}{\exp(2\lambda a)} \text{ by Markov's inequality}$$

$$\leq \frac{\exp\left(\lambda\rho L[M_{V(\mathbb{R})}(1) - 1]\right)}{\exp(2\lambda a)}$$

$$= \exp\left(-\lambda\left(2a - \rho L[M_{V(\mathbb{R})}(1) - 1]\right)\right).$$

Now, taking $a \geq \rho L[M_{V(\mathbb{R})}(1) - 1]$, we have that

$$\limsup_{\lambda\to\infty} \frac{1}{\lambda}\cdot\log P\left(\frac{\tilde{X}_\lambda}{\lambda} \in {K_{2a}}^c\right) \leq -a.$$

Thus, $$\limsup_{\lambda\to\infty} \frac{1}{\lambda}\cdot\log P_\lambda\left({K_{2a}}^c\right) \le -a.$$

This means that P_λ satisfies large deviation tightness by Definition 4.7.

Therefore Theorem 4.10 (b) implies that $\frac{\tilde{X}_\lambda}{\lambda}$ satisfies a large deviation upper bound.

4.4. On the Lower Bound for Large Deviation on Random Measure

Throughout this section, we follow the same notations as in the previous sections.

In this section we obtain a lower bound for the large deviation property for certain classes of random measures. One of the situations we study is as follows. We follow the notation $\tilde{X}_\lambda$ given in Definition 4.4. We obtain the lower bound of random measure of the type ;

$$\liminf_{\lambda\to\infty} \frac{1}{\lambda}\cdot\log P\left(\frac{\tilde{X}_\lambda}{\lambda} \in G\right) \ge -I_L(G) \text{ for any open subset G in M[0, L].}$$

Here, recall that $I_L(\mu) = \sup_{f\in C[0,L]} \left\{\mu(f) - \Phi(f)\right\}$ defined in (4.2)

and $$I_L(G) = \inf\left\{I_L(\mu) : \mu \in G\right\}.$$

This extends some of the works on large deviation by Ellis.

Assumption A : If $\{f_n\}$ is a sequence of uniformly bounded measurable functions on [0, L], which converges μ-almost everywhere to a bounded measurable function f, then $\Phi(f_n) \to \Phi(f)$, that is, Φ is continuous with respect to f.

Assumption B : The lower bound of large deviation holds for in the sense of the finite dimensional distributions as Theorem 3.3 (c) in Section 3.1..

Lemma 4.12 Under the Assumption A, we get

$$\sup_{f \in C[0,L]} \left\{\mu(f) - \Phi(f)\right\} \geq \sup_{\underset{\sim}{t} \in R^n} \left\{\sum_{i=1}^{n} t_i \mu(A_i) - \Phi(\underset{\sim}{t})\right\},$$

where $\Phi(\underset{\sim}{t}) = \Phi\left(\sum_{i=1}^{n} t_i \mu(A_i)\right)$ and $\underset{\sim}{t} = (t_1, \ldots, t_n)$.

To prove the above lemma, it is enough to get the following theorem.

Theorem 4.13 Under the Assumption A, the following property holds :

$$\sup_{f \in C[0,L]} \left\{\mu(f) - \Phi(f)\right\} = \sup_{n,\, A_1, \ldots, A_n} \ \sup_{\underset{\sim}{t} \in R^n} \left\{\sum_{i=1}^{n} t_i \mu(A_i) - \Phi(\underset{\sim}{t})\right\}.$$

Proof. Given $f \in C[0, L]$. We can find uniformly bounded simple functions f_n such that $f_n \to f$. So $\mu(f_n) \to \mu(f)$ and $\Phi(f_n) \to \Phi(f)$ by Assumption A.

So,

$$\mu(f) - \Phi(f) \leq \sup_{n,\, A_1, \ldots, A_n} \ \sup_{\underset{\sim}{t} \in R^n} \left\{\mu(f_n) - \Phi(f_n)\right\}.$$

Taking sup on both sides over all $f \in C[0, L]$, then we have the inequality ;

$$\sup_{f \in C[0,L]} \left\{\mu(f) - \Phi(f)\right\} \leq \sup_{n,\, A_1, \ldots, A_n} \ \sup_{\underset{\sim}{t} \in R^n} \left\{\sum_{i=1}^{n} t_i \mu(A_i) - \Phi(\underset{\sim}{t})\right\}.$$

On the other hand,

given $g = \sum_{i=1}^{n} t_i 1_{A_i}$, we can find uniformly bounded continuous functions g_m such that $g_m \to g$ a.e.. By Bounded Convergence Theorem and Assumption A, we get $\mu(g_m) \to \mu(g)$ and $\Phi(g_m) \to \Phi(g)$.

So, similarly as before,

$$\mu(g_m) - \Phi(g_m) \leq \sup_{f \in C[0,\, L]} \left\{ \mu(f) - \Phi(f) \right\}.$$

Taking sup on both sides over all g_j, that is, over all n, A_1, A_2, . . . , A_n and t_1, t_2, . . . , t_n, then we have the reverse inequality ;

$$\sup_{f \in C[0,\, L]} \left\{ \mu(f) - \Phi(f) \right\} \geq \sup_{n,\, A_1,\, \dots,\, A_n} \; \sup_{\underset{\sim}{t} \in R^n} \left\{ \sum_{i=1}^{n} t_i \mu(A_i) - \Phi(\underset{\sim}{t}) \right\}.$$

So, the above equality is proved.

<u>Definition</u> <u>4.14</u> Let $(T, \mathcal{T})$ be a topological space. A class $\mathfrak{B}$ of open subsets of T, that is, $\mathfrak{B} \subseteq \mathcal{T}$, is a base for the topology $\mathcal{T}$ if and only if

(a) every open set $G \in \mathcal{T}$ is the union of members of $\mathfrak{B}$.

Equivalently, $\mathfrak{B} \subseteq \mathcal{T}$ is a base for $\mathcal{T}$ if and only if

(b) for any point p belonging to an open set G, there exists $B \in \mathfrak{B}$ with $p \in B \subseteq G$.

Recall that the topology of weak convergence is generated by the following class of basic open neighborhoods of $\nu \in M[0, L]$,

$$\mathfrak{B}_1 = \left\{ \mu \in M[0, L] : \left| \int_0^L f_i d\mu - \int_0^L f_i d\nu \right| < \epsilon,\ i = 1, \dots, k \right\},$$

where $f_1, \ldots, f_k$ are elements of $C[0, L]$ and ϵ is positive.

Theorem 4.15 Let $\mathfrak{B}_2 = \Big\{\mu \in M[0, L] : \mu(F_i) < \nu(F_i) + \epsilon,\ 1 \le i \le k$ and $|\mu[0, L] - \nu[0, L]| < \epsilon\Big\}$. Then $\mathfrak{B}_1$ and $\mathfrak{B}_2$ are equivalent.

Proof. First we show that every set in $\mathfrak{B}_1$ contains a set in $\mathfrak{B}_2$. We need consider only a single function f in $C[0, L]$ and further we may assume that $0 < f(x) < 1$ for all x. Choose k so that $\frac{\mu[0, L]}{k} < \epsilon$ and let $F_i = \Big\{x : \frac{i}{k} \le f(x)\Big\}$. Note that

$$(3)\ \sum_{i=1}^{k} \tfrac{i-1}{k}\cdot\mu\Big\{x : \tfrac{i-1}{k} \le f(x) \le \tfrac{i}{k}\Big\} = \sum_{i=1}^{k} \tfrac{i-1}{k}\cdot\Big\{\mu(F_{i-1}) - \mu(F_i)\Big\} = \tfrac{1}{k}\cdot\sum_{i=1}^{k}\mu(F_i).$$

$$(4)\ \sum_{i=1}^{k} \tfrac{i}{k}\cdot\mu\Big\{x : \tfrac{i-1}{k} \le f(x) \le \tfrac{i}{k}\Big\} = \sum_{i=1}^{k} \tfrac{i}{k}\cdot\Big\{\mu(F_{i-1}) - \mu(F_i)\Big\} = \frac{\mu[0, L]}{k} + \tfrac{1}{k}\cdot\sum_{i=1}^{k}\mu(F_i).$$

$$(5)\ \sum_{i=1}^{k} \tfrac{i-1}{k}\cdot\mu\Big\{x : \tfrac{i-1}{k} \le f(x) \le \tfrac{i}{k}\Big\} \le \int f d\mu < \sum_{i=1}^{k} \tfrac{i}{k}\cdot\mu\Big\{x : \tfrac{i-1}{k} \le f(x) < \tfrac{i}{k}\Big\}.$$

From (3), (4), and (5),

$$(4.3) \qquad \tfrac{1}{k}\cdot\sum_{i=1}^{k}\mu(F_i) \le \int f d\mu < \frac{\mu[0, L]}{k} + \tfrac{1}{k}\cdot\sum_{i=1}^{k}\mu(F_i).$$

From (4.3), we have

$$\tfrac{1}{k}\cdot\sum_{i=1}^{k}\nu(F_i) \le \int f d\nu \quad \text{and} \quad \int f d\mu < \epsilon + \tfrac{1}{k}\cdot\sum_{i=1}^{k}\mu(F_i).$$

Since $\mu(F_i) < \nu(F_i) + \epsilon$, we have

$$\int f d\mu < \epsilon + \tfrac{1}{k}\cdot\sum_{i=1}^{k}\mu(F_i) < \epsilon + \tfrac{1}{k}\cdot\sum_{i=1}^{k}\Big(\nu(F_i) + \epsilon\Big)$$

$$< \epsilon + \frac{1}{k}\cdot\sum_{i=1}^{k}\nu(F_i) + \epsilon < 2\epsilon + \int f d\nu.$$

So,

$$(4.4) \qquad \int f d\mu - \int f d\nu < 2\epsilon.$$

On the other hand, taking $1 - f$ instead of f in the formula (4.4) as in the above argument, we have

$$\int(1 - f)d\mu - \int(1 - f)d\nu < 2\epsilon.$$

This means that $\mu[0, L] - \int f d\mu - \nu[0, L] + \int f d\nu < 2\epsilon$. That is,

$$\int f d\nu - \int f d\mu + (\mu[0, L] - \nu[0, L]) < 2\epsilon.$$

So, $\int f d\nu - \int f d\mu < 3\epsilon$. Therefore, we can get that $|\int f d\mu - \int f d\nu| < \epsilon$. This means that we can find within $\mathfrak{B}_1$ a set of the form $\mathfrak{B}_2$. That is, $\mathfrak{B}_1$ contains $\mathfrak{B}_2$.

Now, it remains to show that each set in $\mathfrak{B}_2$ contains a set in $\mathfrak{B}_1$.

Given $|\int f d\mu - \int f d\nu| < \frac{1}{2}\epsilon$ in $\mathfrak{B}_1$ and a closed set F.

Let $$F^{1/n} = \left\{x : \rho(x, F) < \frac{1}{n}\right\}.$$

Since $F^1 \supseteq F^{\frac{1}{2}} \supseteq \ldots$ and $F = \bigcap F^{\frac{1}{n}}$ and $\nu(F^1) < \infty$,

$$\nu(F) = \lim_{n\to\infty} \nu\left(F^{\frac{1}{n}}\right). \text{ That is, } \nu\left(F^{\frac{1}{n}}\right) < \nu(F) + \frac{1}{2}\epsilon.$$

Choosing in $C[0, L]$ a function f with value 1 on F, value 0 outside $F^{\frac{1}{n}}$, and value everywhere contained between 0 and 1, we have

$$|\int fd\mu - \int fd\nu| < \tfrac{1}{2}\epsilon. \quad \text{So} \int fd\mu - \int fd\nu < \tfrac{1}{2}\epsilon.$$

$$\Rightarrow \int_F fd\mu < \int_{F^{\frac{1}{n}}} fd\mu < \int_{F^{\frac{1}{n}}} fd\nu + \tfrac{1}{2}\epsilon.$$

$$\Rightarrow \mu(F) < \int_{F^{\frac{1}{n}}} fd\mu < \nu(F^{\frac{1}{n}}) + \tfrac{1}{2}\epsilon < \nu(F) + \tfrac{1}{2}\epsilon + \tfrac{1}{2}\epsilon.$$

So, $\mu(F) < \nu(F) + \epsilon.$

Finally, taking $f = 1_{[0,\, L]}$, then $|\mu[0, L] - \nu[0, L]| < \epsilon$.

Theorem 4.16 We assume the properties A and B. Then the large deviation lower bound holds. That is, for any open subset G in M[0, L],

$$\liminf_{\lambda\to\infty} \frac{1}{\lambda}\cdot\log P\left(\frac{\tilde{X}_\lambda}{\lambda} \in G\right) \geq -I(G), \text{ where } I(G) = \inf_{\nu\in G} I(\nu).$$

Proof. Let G be an open subset in M[0, L]. Then G is a union of open sets from $\mathfrak{B}_2$, since $\mathfrak{B}_2$ forms a base for the topology generated by weak convergence, where $\mathfrak{B}_2 = \Big\{B$, open in M[0, L] : $B = B(\nu, F_1, \ldots, F_n, \epsilon) = \{\mu \in M[0, L] : \mu(F_i) < \nu(F_i) + \epsilon,\ i = 1, \ldots, n$ and $|\mu[0, L] - \nu[0, L]| < \epsilon$ for Borel sets $F_1 \subseteq F_2 \subseteq \cdots \subseteq F_n \subseteq [0, L]\}\Big\}$.

Let

$$E_1 = F_1$$

(4.5) $$E_k = F_k \backslash F_{k-1}\ ,\ k = 2, \ldots, n$$

$$E_0 = [0, L] \backslash F_n.$$

Then there exists an open set $\mathcal{U} \subseteq [0, L]^{n+1}$ depending on $\nu(F_1), \ldots, \nu(F_n)$, $[0, L]$ and ϵ.

$$B = \Big\{\mu \in M[0, L] : \big(\mu(E_0), \ldots, \mu(E_n)\big) \in \mathcal{U}\Big\}$$

$$=: C(E_0, \ldots, E_n, \mathcal{U}) \text{ (we denote this by C), where all } E_i\text{'s are disjoint.}$$

So we have the relation such that

$$\text{(4.6)} \qquad G = \bigcup_{B \in \mathfrak{B}_2,\, B \subseteq G} B = \bigcup_{C \in \mathfrak{B}_2,\, C \subseteq G} C,$$

where C is obtained from B by letting each F_i disjoint as in (4.5).

Thus,

$$P\Big(\frac{\tilde{X}_\lambda}{\lambda} \in B\Big) = P\Big(\frac{\tilde{X}_\lambda(F_i)}{\lambda} < \nu(F_i) + \epsilon \text{ for } i = 1, \ldots, n \text{ and}$$

$$\nu[0, L] - \epsilon < \frac{\tilde{X}_\lambda[0, L]}{\lambda} < \nu[0, L] + \epsilon\Big)$$

$$= P^{\text{f.d.}}(C),$$

where C depends on $F_1, \ldots, F_n$ and $[0, L]$ which is obtained by making each F_i disjoint as in (4.5).

Since $I(G) = \inf_{\nu \in G} I(\nu)$ by definition, using the relation (4.6) we have

$$-I(G) = -\inf_{\nu \in G} I(\nu) = \sup_{\nu \in G} -I(\nu) = \sup_{B \in \mathfrak{B}_2,\, B \subseteq G} -I(B) = \sup_{C \in \mathfrak{B}_2,\, C \subseteq G} -I(C).$$

Moreover if $\nu \in G$, there exists $C \in \mathfrak{B}_2$ (obtained from B) depending upon a finite dimension which contains ν such that for each $\epsilon > 0$, $-I(G) \leq -I(B) + \epsilon$.

Therefore, $\liminf_{\lambda\to\infty} \frac{1}{\lambda}\cdot\log P\left(\frac{\tilde{X}_\lambda}{\lambda} \in G\right) \geq \liminf_{\lambda\to\infty} \frac{1}{\lambda}\cdot\log P\left(\frac{\tilde{X}_\lambda}{\lambda} \in B\right)$

$$= \liminf_{\lambda\to\infty} \frac{1}{\lambda}\cdot\log P^{\text{ f.d.}}(C)$$

$$\geq -\inf_{\underset{\sim}{t}\in C} I(\underset{\sim}{t}) \qquad \text{by Assumption B}$$

$$\geq -\inf_{\mu\in B} I(\mu) \qquad \text{by Lemma 4.12}$$

$$= -I(B)$$

$$\geq -I(G) - \epsilon.$$

Since ϵ is arbitrary, letting $\epsilon \to 0$, we get

$$\liminf_{\lambda\to\infty} \frac{1}{\lambda}\cdot\log P\left(\frac{\tilde{X}_\lambda}{\lambda} \in G\right) \geq -I(G).$$

Theorem 4.17 If $\frac{\tilde{X}_\lambda{}'}{\lambda}$ satisfies a large deviation lower bound for some $L > 0$, then $\frac{\tilde{X}_\lambda{}'}{\lambda}$ also satisfies a large deviation lower bound for every $L' > 0$,
where we denote $\tilde{X}_\lambda{}' = \left(X_\lambda\right)^{L'}$.

Proof. This can be proved by the same method as the proof of Theorem 4.11.

4.5. Some Examples for the Lower Bound

1. Poisson Random Measure Case : Let X be a Poisson random measure with intensity ρ. Recall the moment generating functional of X is

$$E[\exp(X(f))] = \exp\left(\rho\int_0^L \left(e^{f(x)} - 1\right)\right)dx \text{ for every } f \in C[0, L].$$

Also recall that $\Phi(f) = \lim_{\lambda\to\infty} \frac{1}{\lambda}\cdot\log E[e^{\tilde{X}_\lambda(f)}]$ for every $f \in C[0, L]$.

For a Poisson random measure X, we get $\Phi(f) = \rho\int_0^L\left(e^{f(x)} - 1\right)dx$. Now let us check the condition of the Assumption A.

$$\begin{aligned}\lim_{n\to\infty}\Phi(f_n) &= \lim_{n\to\infty} \rho\int_0^L\left(e^{f_n(x)} - 1\right)dx \\ &= \rho\int_0^L \lim_{n\to\infty}\left(e^{f_n(x)} - 1\right)dx \quad \text{by L.D.C.T.} \\ &= \rho\int_0^L\left(e^{f(x)} - 1\right)dx = \Phi(f).\end{aligned}$$

Therefore $\Phi(f_n) \to \Phi(f)$ as $f_n \to f$ a.e.. The Assumption A in Section 4.4. is satisfied for a Poisson random measure X. Now, for the Assumption B, we have already proved such property in Section 3.2..

2. Poisson center cluster process : Let U be a stationary Poisson process with parameter ρ and V be a point process satisfying $E[V(\mathbb{R})] < \infty$. Let u_i denote the random occurrences of U and take these as the cluster centers. Let the cluster members $V_1, V_2, \ldots$ be point processes which are independent and identically

distributed as V. Then X = [U, V] is a Poisson center cluster process with centers U and clusters V and we denote

$$X = \sum_{u_i \in U} V(A - u_i) \text{ for a bounded Borel set A.}$$

From the previous section, recall that the moment generating functional of a Poisson center cluster process is

$$\begin{aligned}
\Phi(f) &= \lim_{\lambda \to \infty} \frac{1}{\lambda} \cdot \log E[e^{\tilde{X}_\lambda(f)}] \\
&= \lim_{\lambda \to \infty} \frac{1}{\lambda} \cdot \log\left(e^{\rho \int_0^L \left(M_V(T_x f_\lambda) - 1\right) dx} \right) \text{ by Proposition 3.8} \\
&= \rho \int_0^L \left(M_{V([0,\, L])}(f) - 1\right) dx \text{ by using the proof of Proposition 3.9.}
\end{aligned}$$

$$\begin{aligned}
\text{So,} \quad \lim_{n \to \infty} \Phi(f_n) &= \lim_{n \to \infty} \rho \int_0^L \left(M_{V([0,\, L])}(f_n) - 1\right) dx \\
&= \rho \int_0^L \lim_{n \to \infty} \left(E[e^{V([0,\, L])(f_n)}] - 1 \right) dx \text{ by L. D. C. T.} \\
&= \rho \int_0^L \left(E[e^{V([0,\, L])(f)}] - 1 \right) dx = \Phi(f).
\end{aligned}$$

Therefore $\Phi(f_n) \to \Phi(f)$ as $f_n \to f$ a.e.. The Assumption A is satisfied for a Poisson center cluster process. For the Assumption B in Section 4.4., we have already proved such property in Section 3.3..

BIBLIOGRAPHY

[1] Barlow, R.E. and F. Proschen : Statistical Theory of Reliability and Life Testing. Holt, Reinhart and Winston, Inc., (1981).

[2] Billingsley, P. : Probability and Measure. Wiley, New York, (1979).

[3] Billingsley, P. : Convergence of Probability Measures. Wiley, New York, (1968).

[4] Burton, R.M. and E. Waymire : The central limit problem for infinitely divisible random measure, in Dependence in Probability and Statistics. 383 - 395, ed. M. Taqqu and E. Eberlin. Birkhauser Press, Boston, (1986).

[5] Burton, R.M. and E. Waymire : Classical scaling limits for associated random measure. Ann. of Prob. 13(4), 1267 - 1278, (1985).

[6] Burton, R.M. and E. Waymire : Scaling limits for random fields. J. Mult. Analysis, vol. 15, no. 2, 237 - 251, (1984).

[7] Deuschel J. and D.W. Stroock : Large Deviations. Academic press, Inc. (1989).

[8] Daley, D.J. and D. Vere-Jones : An Introduction to the Thery of Point Processes. Springer-Verlag, New York Berlin Heidelberg London Paris Tokyo, (1988).

[9] Daley, D.J. and D. Vere-Jones : A summery of the theory of point processes, in Stochastic Point Processes. 299 - 353, ed. Lewis. Wiley, New York, (1972).

[10] Donsker, M.D. and Varadhan, S.R. : Asymptotic evaluation of certain Markov process expectations for large time. III, Comm. Pure Appl. Math. Vol. 29, 389 - 461, (1976).

[11] Donsker, M.D. and Varadhan, S.R. : Asymptotic evaluation of certain Markov process expectations for large time. II, Comm. Pure Appl. Math. Vol. 28, 1 - 47 (1975a), 279 - 301 (1975b).

[12] Ellis, R. : Entropy, Large Deviations and Statistical Mechanics. Berlin Heidelberg New-York, Springer, (1985).

[13] Ellis, R. : Large deviations for a general class of dependent random vectors. Ann. Probab. 12, 1 - 12, (1984).

[14] Feller, W. : An Introduction to Probability Theory and Its Applications. Vol. I, II. Wiley, New York, (1971).

[15] Fisher, L. : A survey of the mathematical theory of multidimensional point processes in Stochastic Point Processes, 468 - 513, ed. Lewis. Wiley, New York, (1972).

[16] Gärtner, J. : On large deviations from the invariant measure. Theor. Probab. Appl. 22, 24 - 39, (1977).

[17] Groeneboom, P. : Large Deviations and Asymptotic Efficiencies, Mathematisch Centrum Amsterdam, (1980).

[18] Gupta, V.K. and E. Waymire : The mathematical structure of rainfall, part II, a review of the theory of point processes. Water Resourcess Research 17(5), (19), 1275 - 1285, (1981).

[19] Kallenberg, O. : Random Measures. Academic Press, New York, (1983).

[20] Kallenberg, O. : On symmetrically distributed random measures. Trans. Am. Math. Soc. (202), 105 - 121, (1975).

[21] Karr, A.F. : Point Processes and Theory Statistical Inference. Marcel Dekker, Inc., (1986).

[22] Kingman, J.F.C. : On doubly stochastic Poisson processes. Proc. Camb. Phil. Soc., 60, (1964), 923 - 930.

[23] Lanford, O.E. : Entropy and Equilibrium States in Classical Statistical Mechanics. (Lect. Notes Phys. Vol. 20), Berlin Heidelberg New York, Springer (1973).

[24] Resnick, Sidney I. : Extreme Values, Regular Variation, and Point Processes. Springer-Verlag, New York Berlin Heidelberg London Paris Tokyo, (1987).

[25] Rockafflar, R.T. : Convex Analysis. Princeton Univ. Press, Princeton, N. J., (1970).

[26] Rudin, W. : Real and Complex Analysis. 2nd edition. Mcgraw-Hill, New York, (1974).

[27] Rudin, W. : Functional Analysis. Mcgraw-Hill, New York, (1973).

[28] Steinebach, J. : Convergence rates of large deviation probabilities in the multi-dimensional case. Ann. Probab. 6, 751 - 759, (1978).

[29] Stroock, D.W. : An Introduction to the Theory of Large Deviations. Berlin - Heidelberg - New-Yock - Tokyo, Springer, (1984).

[30] Varadhan, S.R.S. : Large Deviations and Applications, (CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 46)Philadelphia : Society for Industrial and Applied Mathematics, (1984).

[31] Ventsel, A.D. : Limit theorems on large deviations for random processes, Theor. Probab. Applications. 18, 779 - 784, (1974).

[32] Westcott, M. : The probability generating functional. J. Austral. Math. Soc. (14), (1972), 448 - 466.