



Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Random Sampling and Large Deviations

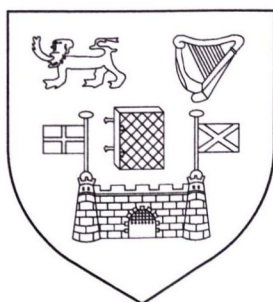
by

Brian McGurk

A thesis submitted to
the University of Dublin
for the degree of

Doctor of Philosophy

School of Mathematics,
University of Dublin, Trinity College



September, 2000

Declaration

This thesis has not been submitted as an exercise for a degree at any other University. Except where otherwise stated, the work described herein has been carried out by the author alone. This thesis may be borrowed or copied upon request with the permission of the Librarian, University of Dublin, Trinity College. The copyright belongs jointly to the University of Dublin and Brian McGurk.

Signature of Author.....*Brian McGurk*.....

Brian McGurk
29 September, 2000

Abstract

In this thesis, we are concerned with the effect of randomly sampling a stochastic process. We consider two stochastic processes: the *underlying* process, $\{X_t\}_{t \in \tau}$, and the *observation* process $\{T_n\}_{n \in \mathbb{N}}$, a strictly increasing process taking values in τ . The process of interest is generated by sampling the underlying process at the times specified by the observation process,

$$Y_n := X_{T_n}.$$

We call this transformation *random sampling* and refer to $\{Y_n\}$ as the *observed*, or *sample*, process.

In particular, we are interested in how the large deviations properties of the process are affected by this operation and so we only consider those that satisfy an LDP. In the case where the process counting the number of observations is Poisson, a formula can be obtained for the new rate function in terms of the rate function for the underlying process. Using order theoretic ideas, we can, in certain cases, characterise the inverse to this formula and partially determine the rate function of the underlying process, given only that of the observed process. Applying this formula when the underlying process is Markov, we derive a novel expression for the 2-state Markov rate function.

In the case where the observation process is not Poisson, we can establish general conditions that ensure the observed process will obey an LDP. For a large class of observation processes, we show how certain conditions on the underlying process, sufficient for an LDP, are preserved by the random sampling transformation.

Acknowledgements

Many people deserve a share of my thanks for their help and support in the course of my research.

John Lewis has been a constant source of encouragement over my time in DIAS. Apart from the practical matters of maintaining the Applied Probability Group, on which he has worked ceaselessly, I have benefitted greatly from his wide-ranging mathematical experience. Most of all, on those days when the proof doesn't work, I have relied on John's confidence when mine has evaporated.

Raymond Russell has displayed unflagging patience under a barrage of half-baked ideas and conjectures. The adjacency of his office was certainly to blame, but not so much as his willingness to listen and sound advice. Of course, he is only one among many to have heard me discourse on the technical minutiae of lemmas. I point to Fergal Toomey, Ken Duffy, Cormac Walsh, Mark Dukes and Paul Watts, whose help I gratefully acknowledge; they have provided much advice and many interesting conversations on all things mathematical. I should single out Ken, Paul and Cormac for their proof-reading of this thesis, which has reduced the number of mistakes therein.

My time in DIAS was made much easier by the smooth running of the administration; by Anne Goldsmith's assistance with library resources and by Margaret Matthews for all her help with travel arrangements. Ian Dowse has provided an impeccably maintained computer system; I am very grateful for that and for what he has taught me of its workings.

While I have worked on this thesis, and in particular over the final few months, Lisa Carey has been very supportive and understanding. She has played a large part in making the last three years a happy and productive time for me. She has also helped me with my words.

Finally, I would like to thank my parents for the love and support they have provided me with throughout my life. More so than anyone else, they are responsible for my being able to do this research, and that is but one of my many debts to them.

This work was supported by the European Commission under ESPRIT LTR project 20113, 'Measure', and by a grant from Microsoft Research, Cambridge.

Contents

1	Introduction	1
1.1	What are large deviation asymptotics?	1
1.2	What we talk about when we talk about random sampling	4
2	Large deviations	7
2.1	Semi-continuity, set functions and level sets	8
2.2	The vague and narrow path to the LDP	10
2.3	Varadhan's theorem	12
2.4	Stochastic processes	21
3	Markov underlying processes	27
3.1	Randomly sampled Markov processes	27
3.2	The rate function of a Markov chain	32
4	Poisson sampling : I	38
4.1	A rough sketch	38
4.2	The details	40
5	Poisson sampling : II	59
5.1	Galois connections	59
5.2	Recovering the underlying rate function	62
6	Non-Poisson sampling	72
6.1	Correlated processes and large deviations	75
6.2	Stationarity and correlation structure after sampling	77
	Notation	85
	Bibliography	87

Chapter 1

Introduction

We are interested in considering how certain asymptotic properties of a stochastic process are affected by considering only the values assumed by that process at a random subset of its index set. Specifically, we have in mind a second stochastic process which determines a strictly increasing sequence of times at which the first process is observed. We wish to determine how the two processes interact in generating the observed process and how much information it is possible to regain about the asymptotics of the underlying process.

In particular, we are concerned with looking at the large deviation asymptotics of the process. In Chapter 2, we provide a technical summary of the aspects of the theory which are relevant to our investigations. In the following section, we provide a context for the abstract treatment with a heuristic discussion of a concrete example.

1.1 What are large deviation asymptotics?

The theory of large deviations grew from a study of the probabilities of rare events in widely different areas, most famously risk theory and thermodynamics. It is concerned with rare events whose probabilities decay due to some inherent scaling. To provide a concrete example, consider the much beloved coin-tossing experiment. Out of n tosses of an unbiased coin, we expect that roughly half will come up heads. In fact, the weak law of large numbers tells us that the probability of *unusual* behaviour becomes arbitrarily small as we toss more and more coins. To be precise, if we write the proportion of heads after n tosses as M_n ,

then the weak law of large numbers tells us that for an arbitrarily small ϵ ,

$$\lim_{n \rightarrow \infty} P\left(\left|M_n - \frac{1}{2}\right| > \epsilon\right) = 0.$$

Large deviation theory tells us how this probability decays when examined on an exponential scale.

An obvious question to ask is why we expect decay on an exponential scale to be of interest. The answer to this lies in the fact that we assume that successive coin tosses are independent and statistically identical. Hence, any particular sequence of heads and tails is equally likely. As a consequence of this, the probability of any event which depends only on the aggregate behaviour, such as our example where we look at the proportion of heads, will be determined by the number of different sequences which give rise to that behaviour. This means that the probability of getting an unusually high number of heads, after many tosses, will be dominated by those sequences where the number of *extra* heads have been more or less evenly distributed throughout the course of the experiment. This is the case simply because there are a far greater number of these sequences than with the extra heads arriving in a clump. It is not hard to see that as the number of tosses grows, so too does the domination of the probability by those sequences where the heads are most evenly distributed. Effectively, as n becomes larger, the deviation from the expected behaviour is shared as equally as is possible over all the coin tosses. Since we are thus expecting each toss to behave in a similarly deviant fashion, and the tosses are independent of each other, we would expect that the probability would roughly factor into n equal contributions. In other words, where x is some number greater than $\frac{1}{2}$, we might hope to prove the existence of the limit

$$\lim_{n \rightarrow \infty} P(M_n \approx x)^{\frac{1}{n}}.$$

To be more specific, a typical large deviation statement is of the form

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P(M_n \approx x) = -I(x). \quad (1.1)$$

$I(x)$ is called the *rate function* because it describes the rate of decay of the probability, which depends on how far the behaviour is from that which is expected. To emphasise this,

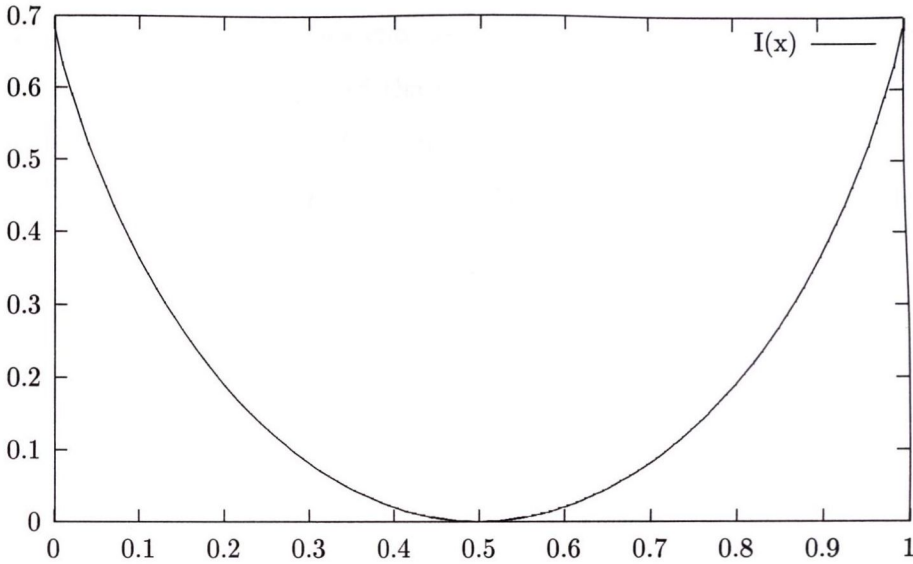


Figure 1-1: The rate function for coin tossing

we sometimes write equation (1.1) using the following notation,

$$P(M_n \approx x) \asymp e^{-nI(x)}.$$

Obviously, the rate function has to be non-negative everywhere, and it reaches zero at the mean because the probability of this typical behaviour tends to 1. In the case of our coin-tossing example, a graph of the rate function is shown in Figure 1.1.

One can think of the rate function as associating a probabilistic *cost* to a specified deviation and only mean behaviour has no cost. The rate function is a measure of the limiting cost per unit time associated of the behaviour in question. Using this analogy, the heuristic argument above can be summarised by saying that, as the time interval in question grows, the cheapest way that the process can achieve a specified deviation from mean behaviour is by spreading that deviation evenly over the interval. This is reflected in the convex shape of I , which is a common, though not essential, feature of rate functions. For example, we consider a deviation of 0.7 in the coin-tossing experiment. Convexity derives from the fact that it is cheaper for the coin to turn up heads with a frequency of 0.7 for the duration of the experiment, than for the coin to turn heads with a frequency of 0.6

half the time and with a frequency of 0.8 the rest of the time.

In a more general situation, we can consider arbitrary stochastic processes and ask whether some large scale property of the process, such as its empirical average, satisfies a large deviation principle (LDP). In such cases, the heuristic argument above can break down. For example, *heavy-tailed* processes, where, for each individual random variable, the probability of unusual behaviour does not decay quickly, may satisfy LDPs on hyper-exponential scales. In such cases, the deviation can be caused by the action of a single random variable, rather than being spread over the entire history of the process, and the rate function will not necessarily be convex.

The precise definition of an abstract LDP is somewhat technical and is covered in Chapter 2. In the general case, instead of a sequence of laws of empirical averages, as for coin-tossing, we look at an arbitrary sequence of measures and examine the logarithmic asymptotics of the sequence on certain sets, against some specified scale. Roughly speaking, the LDP holds when

$$\lim_{n \rightarrow \infty} \frac{1}{V_n} \log P(M_n \in B) = - \inf_{x \in B} I(x),$$

for B in some suitably defined class of sets.

1.2 What we talk about when we talk about random sampling

We have in mind a situation where a stochastic process can only be observed at certain times, which are governed by some other random process. Thus the observer has only partial information about the behaviour of the underlying process. As a concrete, if somewhat artificial, example, consider the transit time along some city bus route, subject to unpredictable variations, as the underlying stochastic process. Consider also an individual commuter who, subject to random whims, may choose alternatives to the public transport provided. We could describe this situation as a random sampling of the transit time process. The process consisting of the commuter's transit times on public transport is obviously related to the underlying process, but there is also an effect due to his random decision process.

The process under observation is referred to as the *underlying* process, and in continuous time, is usually denoted $\{X_t\}_{t \in \mathbb{R}_+}$. The times at which $\{X_t\}$ is sampled are determined by

the *observation* process, and the time of the n^{th} observation is usually denoted T_n . N_t typically refers to the number of observations before time t . In that case, the *observed* process, consisting of the sequence of samples, is defined by

$$Y_n := X_{T_n}.$$

We are interested in finding out how the random sampling transformation affects the large deviations properties of the underlying process, and, if possible, relating the rate functions of $\{X_t\}$ and $\{T_n\}$ to that of $\{Y_n\}$. We would expect that the asymptotic properties of the process would be changed by this operation, since the number of ways in which an event can happen is greatly increased by the addition of random time intervals between observations. For instance, to achieve some deviation of the observed process, either the underlying process could behave in a similarly deviant fashion and the observations be typically distributed, or the underlying process could behave in a more typical manner and the deviation be attributable to an unusual sequence of observations. Hence, to use the cost analogy, we can see there should be some trade-off of the cost of atypical behaviour of the underlying process against that of atypical behaviour of the observation process.

We start by considering a concrete example where the underlying process is a Markov process and the times between observations are independent. In this case, the resulting process can be completely described by a Markov chain, the large deviation properties of which are well-known. When the state space of the Markov process consists of only two values, and the inter-observation times are exponentially distributed, we can identify the Markov chain precisely. This we describe in Chapter 3, where we use this result to come up with a new, if rather unwieldy, formula for the rate function of a 2-state Markov chain.

In so doing, we rely on a formula which is at the centre of Chapter 4 where we relax the conditions on the underlying process while keeping the Poisson structure of the observation process. In this case, we can derive a formula which describes the rate function for the observed process in terms of the underlying one.

Random sampling is related to sampling without replacement, the large deviations of which is considered by Dembo and Zeitouni in [4, Section 7.2]. They consider a deterministic underlying sequence, $\{x_n\}_{n \in \mathbb{N}}$, which is sampled according to a random subsequence drawn from its index set, $\mathbb{N}$. The random subsequence is such that for each n , there are n samples

drawn from the first $m(n)$ elements, where

$$\lim_{n \rightarrow \infty} \frac{n}{m(n)} = \beta, \quad \text{for } \beta \in (0, 1).$$

Its distribution is such that every distinct set of n samples in $\{1, \dots, m\}$ is equally likely. When the empirical average of the underlying deterministic sequence converges, they show that the resulting stochastic process obeys an LDP and they provide a formula for the rate function. They generalise this analysis to consider empirical measures.

For the most part, we restrict our intentions to continuous time processes. When the arguments used in Chapter 4 are converted to the discrete time case, the observation process we choose has geometrically distributed inter-observation times, this being the closest analogue of the Poisson process. This process conforms to the conditions for the LDP derived by Dembo and Zeitouni, and, in that case, the rate function they derive occurs as an intermediate step in the derivation of the rate function in the discrete time random sampling transformation.

Under conditions on the continuity of the rate function, we can affect a partial inversion of the formula derived in Chapter 4. Viewing that formula as a mapping from the space of rate functions to itself, the pseudo-inversion characterises the pre-image. This derivation depends on some elementary results from order theory and is detailed in Chapter 5. Finally, we examine the case where the inter-observation times are not exponentially distributed. In this case, we have not been able to find a formula for the rate function of the observed process but we can nonetheless prove that it satisfies an LDP, and we consider sufficient conditions in Chapter 6. We approach this problem by imposing conditions on the underlying process which guarantee that it satisfies an LDP, and then showing that the observed process obeys a similar condition. These conditions effectively limit the strength of the process's correlation structure.

Chapter 2

Large deviations

The aim of this chapter is to provide an abstract introduction to some of the technical aspects of the theory of large deviations and to define some concepts that we will use throughout the remaining chapters. We use the framework described and developed by Lewis and Pfister in [7]. A minor, but potentially confusing, difference is that in our discussion of the LDP, the rate function is defined to have a different sign to that of the rate function in [7]. We do this because, outside of this chapter, we are only interested in the large deviations of probability measures, whose logarithm is always non-positive. For this reason, we follow the convention of the probability literature and choose to deal with non-negative rate functions. This implies that the rate function be lower, rather than upper, semi-continuous.

In Section 2.1, we define some of the terminology we will use and state some simple results regarding the properties of semi-continuous functions and their relation to set functions. In the following section we describe the incremental approach to proving an LDP which is explained in some detail by Lewis and Pfister in [7]. In Section 2.3, we investigate the general form of Varadhan's theorems, originally proved by Varadhan in [10]. We prove these theorems, in a more restricted setting, within the framework of [7]. In Section 2.4, we review some issues relating to stochastic processes and the LDPs deriving from them.

The material in Sections 2.1 and 2.2 is not original, being merely a summary of [7], but in Section 2.3, where proofs are provided either the result or the method of proof is believed to be novel.

2.1 Semi-continuity, set functions and level sets

In advance of discussing the definition of an LDP, it is helpful to gather together some facts about semi-continuous functions, set functions and the links between them. Let X be a Hausdorff topological space, and let G denote a generic member of $\mathcal{G}$, the collection of open sets.

For a function $f : X \rightarrow \overline{\mathbb{R}}$, its *level sets* are those of the form

$$\{x : f(x) \leq a\},$$

for some finite a . We say that f is lower semi-continuous (lsc) iff all its level sets are closed. f is upper semi-continuous (usc) if $-f$ is lsc, or, in other words, if all sets of the form $\{x : f(x) \geq a\}$ are closed, for finite a .

A set function is a function from a class of subsets of X to $\overline{\mathbb{R}}$. We define two operations, which transform set functions into real-valued functions on X and vice versa. We only concern ourselves with set functions whose domain includes $\mathcal{G}$.

Definition 1 For a set function c , the inf-derivative of c is

$$^{\wedge}c(x) := \inf_{G \ni x} c[G] = \{c[G] : G \in \mathcal{G}, x \in G\}.$$

For a function $f : X \rightarrow \overline{\mathbb{R}}$, the sup-integral of f is

$$^{\vee}f[G] := \sup_{x \in G} f(x).$$

Using this terminology, for an arbitrary function we define semi-continuous regularisations.

Definition 2 For an arbitrary function $f : X \rightarrow \overline{\mathbb{R}}$, we define the usc regularisation by

$$f^{\diamond} := {}^{\wedge\vee}f,$$

and the lsc regularisation by

$$f_{\diamond} := -(-f)^{\diamond}.$$

It is easy to see, directly from the definitions, that $f_{\diamond} \leq f \leq f^{\diamond}$.

Lemma 2.1 *f is usc if and only if $f^\diamond = f$.*

Proof. For any a , we define $G_a := \{x : f(x) < a\}$.

Assume first that f is usc. Since $f^\diamond \geq f$ automatically, we need only prove the opposite inequality. Fix x , and take any $a > f(x)$. Since f is usc, $G_a \in \mathcal{G}$, and of course $\vee f[G_a] \leq a$. Therefore, because $G_a \ni x$,

$$f^\diamond(x) = \inf_{G \ni x} \sup_{y \in G} f(y) \leq a.$$

Since this holds for all $a > f(x)$, we can conclude that $f^\diamond \leq f$.

On the other hand, assuming that $f^\diamond = f$, we want to show that G_a is open.

$$\begin{aligned} G_a &= \left\{ x : \inf_{G \ni x} \sup_{y \in G} f(y) < a \right\} \\ &= \left\{ x : \exists G \ni x \text{ such that } G \text{ is open and } G \subset G_a \right\}. \end{aligned}$$

Thus G_a is equal to its interior and so is open. Hence f is usc. □

For an arbitrary function, f , it can be shown, by a similar argument to that used in proving the second part of the previous lemma, that $\{x : f^\diamond(x) < a\}$ is open for all finite a . Hence, $f^\diamond$ is always usc. Of course, parallel statements hold true for the lsc regularisation: $f_\diamond$ is lsc for any f , and f itself is lsc if and only if $f = f_\diamond$.

Using this characterisation of upper semi-continuity, we can show that the inf-derivative of c , a set function, is a usc function by showing that for any x ,

$$(\wedge c)^\diamond(x) \leq \wedge c(x),$$

and this follows from the fact that

$$\sup_{y \in G} \inf_{G' \ni y} c[G'] \leq c[G], \quad \forall G \ni x$$

by taking infima over all $G \ni x$.

2.2 The vague and narrow path to the LDP

In later chapters, we will only be concerned with LDPs on $\mathbb{R}^d$, but for the present we wish to define a more general setting in which to prove the extension to Varadhan's theorem.

We consider a sequence of measures $\{\mathbb{M}_n\}_{n \geq 1}$ on a Hausdorff topological space X , equipped with a Borel structure. We denote the collection of closed sets by $\mathcal{F}$, and the compact sets by $\mathcal{K}$. We let $\{V_n\}_{n \geq 1}$ denote a scale, an increasing sequence of positive real numbers, diverging to $+\infty$. We are interested in examining the exponential asymptotic properties of these measures on this scale. To this end, we make the following definitions

$$\begin{aligned} m_n[B] &:= \frac{1}{V_n} \log \mathbb{M}_n[B]; \\ \overline{m}[B] &:= \limsup_{n \rightarrow \infty} m_n[B]; \\ \underline{m}[B] &:= \liminf_{n \rightarrow \infty} m_n[B]. \end{aligned}$$

Using these definitions, we can state the abstract LDP as follows.

Definition 3 $\{M_n\}$ satisfies an LDP, on the scale $\{V_n\}$, with rate function $I : X \rightarrow \overline{\mathbb{R}}$, if the following conditions are satisfied

(LDP1) I is lower semi-continuous;

(LDP2) I has compact level-sets;

(LDP3) $-\inf_{x \in G} I(x) \leq \underline{m}[G], \quad \forall G \in \mathcal{G};$

(LDP4) $\overline{m}[F] \leq -\inf_{x \in F} I(x), \quad \forall F \in \mathcal{F}.$

There is a general approach to showing the existence of a rate function for a given sequence $\{(\mathbb{M}_n, V_n)\}$ which relies on finding a function to satisfy a weaker set of conditions, and then using topological arguments to show that this function qualifies as a rate function. We make these weaker conditions precise in the following definition.

Definition 4 $\{\mathbb{M}_n\}$ satisfies a vague LDP on the scale $\{V_n\}$ with rate function I , if I satisfies LDP1, LDP3 and

(LDP4') $\overline{m}[K] \leq -\inf_{x \in K} I(x), \quad \forall K \in \mathcal{K}.$

In order to establish some useful conditions under which a vague LDP is guaranteed, we define two new functions,

$$\begin{aligned}\underline{\mu}(x) &:= \inf_{G \ni x} \underline{m}[G] , \\ \overline{\mu}(x) &:= \inf_{G \ni x} \overline{m}[G] ,\end{aligned}$$

called, respectively, the upper and lower deviation functions. This pair of functions can be used to show the existence of the function I , satisfying the vague LDP.

The inf-derivative of a set function is necessarily upper semi-continuous, which means that $-\underline{\mu}$ and $-\overline{\mu}$ both satisfy LDP1. We also note that the definition of $\underline{\mu}$ implies

$$\sup_{x \in G} \underline{\mu}(x) \leq \underline{m}[G] , \forall G \in \mathcal{G} ,$$

and it is proved in [7, Lemma 2.5], via the principle of the largest term, that

$$\overline{m}[K] \leq \sup_{x \in K} \overline{\mu}(x) , \forall K \in \mathcal{K} .$$

Hence, if $\underline{\mu}$ and $\overline{\mu}$ are equal, then the conditions for the vague LDP are met by the negative of that function. In this case, we define the Ruelle-Lanford (RL) function,

$$\mu(x) := \underline{\mu}(x) = \overline{\mu}(x) .$$

Of course, this merely reduces the problem of proving the vague LDP to showing the existence of the RL function. However, there is a fact to note about the definition of the upper and lower deviation functions which helps us to do this. If $\mathcal{B}$ is a base of Borel neighbourhoods, then for all $G \in \mathcal{G}$ there exists $B \in \mathcal{B}$ such that $B \subset G$. Since both $\overline{m}$ and $\underline{m}$ are increasing set functions, this means that (using B to denote a generic element of $\mathcal{B}$)

$$\begin{aligned}\underline{\mu}(x) &= \inf_{G \ni x} \underline{m}[G] = \inf_{B \ni x} \underline{m}[B] , \\ \overline{\mu}(x) &= \inf_{G \ni x} \overline{m}[G] = \inf_{B \ni x} \overline{m}[B] .\end{aligned}$$

Hence, if we can show that for some such base, $\mathcal{B}$,

$$\underline{m}[B] = \overline{m}[B] , \quad \forall B \in \mathcal{B} ,$$

then we can conclude that the RL function exists. At several points, we will use this technique in the case where X is $\mathbb{R}^d$, where the open balls are a base for the usual topology.

In certain cases, it is possible to promote a vague LDP to a full LDP. This involves two steps, the first is changing the scope of the upper bound by showing that the rate function satisfies LDP4, rather than just LDP4'; the second is proving that the level sets of the rate function are compact. With the strengthening of the upper bound from compact sets to closed sets, we get what Lewis and Pfister term a *narrow* LDP. Some authors, such as Dembo and Zeitouni in [4], define an LDP to be what we term a narrow LDP; they then use the term *good* to describe a rate function with compact level sets.

If all closed sets are compact, then LDP4' implies LDP4, and since the level sets of a lower semi-continuous function are closed, by definition, we can move immediately from a vague to a full LDP. In most cases that will concern us in this thesis, X itself is compact, in which case all closed sets are also compact. Therefore, in such cases, if we choose to prove the LDP via showing the existence of the RL function, then that is in fact all that we need to prove since its existence guarantees a vague LDP which leads automatically to the full LDP. Lewis and Pfister show that, even if the space is not compact, the *exponential tightness* condition guarantees that a sequence satisfying a vague LDP will obey a full LDP.

Definition 5 A sequence of measures, $\{\mathbb{M}_n\}$, is *exponentially tight on the scale* $\{V_n\}$ if there exists a sequence, $\{K_n\}$, of compact subsets of X such that

$$\lim_{n \rightarrow \infty} \overline{m}[X \setminus K_n] = -\infty.$$

In [7, Section 5], they show that if a sequence of measures obeys a vague LDP and is exponentially tight on the same scale, then the sequence obeys a narrow LDP with the same rate function. In the case where X is regular, this rate function is unique, and so is equal to the RL function whose level sets are compact, which completes the conditions for the full LDP.

2.3 Varadhan's theorem

We now investigate a method of using an LDP for one sequence of measures to prove an LDP for a related sequence. Throughout this section, we assume the following:

Assumption 2.1 $\{\mathbb{M}_n\}_{n \geq 1}$ is a sequence of locally finite measures on a Hausdorff space X , which obey an LDP on the scale $\{V_n\}$ with rate function $I = -\mu$, where μ is the RL function for the sequence.

We are interested in considering the large deviations of a sequence of exponentially tilted measures, $\{\mathbb{M}_n^{g_n}\}_{n \geq 1}$, where we define

$$\mathbb{M}_n^{g_n}[B] := \int_B e^{V_n g_n(x)} \mathbb{M}_n[dx].$$

With suitable conditions on the sequence of functions $\{g_n\}$, we can show that $\{(\mathbb{M}_n^{g_n}, V_n)\}$ satisfies an LDP. In keeping with the notation of the previous section, we define the following functions.

$$\begin{aligned} m_n^{g_n}[B] &:= \frac{1}{V_n} \log \mathbb{M}_n^{g_n}[B] \\ \underline{m}^g[B] &:= \liminf_{n \rightarrow \infty} m_n^{g_n}[B], \quad \underline{\mu}^g(x) := \inf_{G \ni x} \underline{m}^g[G], \\ \overline{m}^g[B] &:= \limsup_{n \rightarrow \infty} m_n^{g_n}[B], \quad \overline{\mu}^g(x) := \inf_{G \ni x} \overline{m}^g[G]. \end{aligned}$$

The result known as Varadhan's theorem concerns the exponential asymptotics of the tilted measure applied to the whole space, $\mathbb{M}_n^{g_n}[X]$.

Varadhan's approach

In Varadhan's original paper, [10], he imposes certain convergence criteria on the sequence of functions which ensure that the tilted measures, applied to the whole space, have the appropriate exponential asymptotics. In order to obtain the upper bound, he also imposes a tail condition on the tilted measures which ensures that the functions $\{g_n\}$ do not grow too quickly. This condition is trivially satisfied by a sequence of functions which are uniformly bounded.

To be more explicit, for a sequence of functions, $\{g_n\}$, and a function g , he defines two convergence-like conditions at each $x_0 \in X$:

V1 $\forall \epsilon > 0$, there exists a neighbourhood, $N \ni x_0$, such that for all sufficiently large n and all $x \in N$,

$$g_n(x) \leq g(x_0) + \epsilon;$$

V2 $\forall \epsilon > 0$, there exists a neighbourhood, $N \ni x_0$, such that for all sufficiently large n and all $x \in N$,

$$g_n(x) \geq g(x_0) - \epsilon;$$

and the tail condition,

$$\mathbf{V3} \quad \lim_{L \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{V_n} \log \mathbb{M}_n^{g_n} [g_n^{-1}[L, \infty)] = -\infty.$$

Using these conditions, Varadhan proves the result in three stages. First, he proves the upper bound.

Theorem 2.2 *If condition V1 holds at every x for the function $g(x)$, and V3 is satisfied, then*

$$\overline{m}^g[X] \leq \sup_{x \in X} \{g(x) - I(x)\}.$$

Next, he defines the set X_0 to be the set of points in X where condition V2 holds and this gives the following lower bound.

Theorem 2.3

$$\underline{m}^g[X] \geq \sup_{x \in X_0} \{g(x) - I(x)\}.$$

These are then combined to give the result known as Varadhan's theorem.

Theorem 2.4 *If condition V3 holds and there exists a function $g : X \rightarrow \overline{\mathbb{R}}$ satisfying condition V1 at every x such that*

$$\sup_{x \in X_0} \{g(x) - I(x)\} = \sup_{x \in X} \{g(x) - I(x)\},$$

then

$$\underline{m}^g[X] = \overline{m}^g[X] = \sup_{x \in X} \{g(x) - I(x)\}.$$

The Lewis-Pfister approach

In the more common statement of Varadhan's theorem, the sequence of tilted measures is obtained from the original sequence by means of a single function,

$$\mathbb{M}_n^{g_n} = \mathbb{M}_n^g, \quad (2.1)$$

where $g : X \rightarrow \mathbb{R}$ is some locally bounded function. This is the situation considered by Lewis and Pfister in [7]. In the case where g is continuous, they show the existence of, and identify, the RL function

$$\mu^g = \mu + g, \quad (2.2)$$

and conclude that the sequence satisfies a vague LDP. As explained in the previous section, when the space is compact, a vague LDP is equivalent to a full LDP, and so they get the following theorem.

Theorem 2.5 *If X is compact then, for a locally bounded, continuous function, $g : X \rightarrow \mathbb{R}$,*

$$\overline{m}^g[X] = \underline{m}^g[X] = \sup_{x \in X} \{g(x) - I(x)\}.$$

They also show that for an arbitrary Hausdorff space, this vague LDP can be promoted to a full LDP by using Varadhan's condition, V3.

In the remainder of this section, we wish to expand the approach taken by Lewis and Pfister to cover the situation where the measures are tilted by a sequence of functions. We also explore equivalent conditions to V1 and V2.

A central lemma in the investigation of tilted measures, defined as in equation (2.1), in [7, section 2], states that

$$\begin{aligned} \underline{\mu}(x) + g_\diamond(x) &\leq \overline{\mu}^g(x) \leq \overline{\mu}(x) + g^\diamond(x), \\ \underline{\mu}(x) + g_\diamond(x) &\leq \underline{\mu}^g(x) \leq \overline{\mu}(x) + g^\diamond(x). \end{aligned}$$

If g is continuous, then $g_\diamond = g^\diamond$, and

$$\begin{aligned} \overline{\mu}(x) + g(x) &= \overline{\mu}^g \\ \underline{\mu}(x) + g(x) &= \underline{\mu}^g, \end{aligned}$$

and the existence of μ gives the vague LDP for $\{(\mathbb{M}_n^g, V_n)\}$ with the rate function given in equation (2.2).

We can prove a similar lemma in the more general setting, but we require more than just continuity of the limit function to attain equality. However, before imposing any conditions on $\{g_n\}$, and without needing to assume that $\underline{\mu} = \bar{\mu}$, we can prove the following:

Lemma 2.6 *Let g_n be locally bounded and measurable for all n . Then the upper and lower deviation functions for $(\mathbb{M}_n^g, V_n)$ are related to those for $(\mathbb{M}_n, V_n)$ by*

$$\begin{aligned}\bar{\mu}^g(x) &\leq \bar{\mu}(x) + \inf_{G \ni x} \limsup_{n \rightarrow \infty} \sup_{y \in G} g_n(y), \\ \underline{\mu}^g(x) &\geq \underline{\mu}(x) + \sup_{G \ni x} \liminf_{n \rightarrow \infty} \inf_{y \in G} g_n(y).\end{aligned}$$

Proof. The proofs of the inequalities are almost identical, so we will just prove the first. For any open set G ,

$$m_n^g[G] \leq m_n[G] + \sup_{y \in G} g_n(y).$$

In fact, for $G \supset G' \ni x$,

$$\begin{aligned}m_n^g[G'] &\leq m_n[G'] + \sup_{y \in G} g_n(y), \\ \bar{m}^g[G'] &\leq \bar{m}[G'] + \limsup_{n \rightarrow \infty} \sup_{y \in G} g_n(y).\end{aligned}$$

Taking the infimum over G' and the supremum over G ,

$$\bar{\mu}^g(x) \leq \bar{\mu}(x) + \inf_{G \ni x} \limsup_{n \rightarrow \infty} \sup_{y \in G} g_n(y).$$

□

Because of the LDP we have assumed for the sequence $\{(\mathbb{M}_n, V_n)\}$, we know that the RL function exists; $\mu = \underline{\mu} = \bar{\mu}$. In order to deduce, from this lemma, an LDP for the tilted measures, we need to find conditions under which the inequalities in Lemma 2.6 collapse down to an equality. We first define two operations on the sequence $\{g_n\}$.

$$(Sg)(x) := \sup_{G \ni x} \liminf_{n \rightarrow \infty} \inf_{y \in G} g_n(y),$$

$$(\mathcal{L}g)(x) := \inf_{G \ni x} \limsup_{n \rightarrow \infty} \sup_{y \in G} g_n(y).$$

Lemma 2.6 then implies

$$\mu + (\mathcal{S}g) \leq \underline{\mu}^g \leq \bar{\mu}^g \leq \mu + (\mathcal{L}g),$$

and hence equality of $(\mathcal{S}g)(x)$ and $(\mathcal{L}g)(x)$ implies existence of the RL function at x for the tilted sequence. Not surprisingly, it turns out that the conditions V1 and V2 are sufficient to demonstrate the equality of $\mathcal{L}g$ and $\mathcal{S}g$, and in fact, they are also necessary.

Lemma 2.7

$$(\mathcal{L}g)(x) \leq g(x) \iff V1 \text{ holds at } x \text{ for } g(x).$$

$$(\mathcal{S}g)(x) \geq g(x) \iff V2 \text{ holds at } x \text{ for } g(x).$$

Proof. We check the details for the first equivalence by simply following the definitions.

$$(\mathcal{L}g)(x) \leq g(x)$$

$$\iff \inf_{G \ni x} \limsup_{n \rightarrow \infty} \sup_{y \in G} g_n(y) \leq g(x),$$

$$\iff \forall \epsilon > 0, \exists G_\epsilon \ni x \text{ s.t. } \limsup_{n \rightarrow \infty} \sup_{y \in G_\epsilon} g_n(y) \leq g(x) + \epsilon,$$

$$\iff \forall \epsilon > 0, \exists G_\epsilon \ni x \text{ s.t. } \forall \epsilon' > 0, \exists N_{(\epsilon, \epsilon')} \text{ s.t. } \sup_{n > N_{(\epsilon, \epsilon')}} \sup_{y \in G_\epsilon} g_n(y) \leq g(x) + \epsilon + \epsilon',$$

$$\iff \forall \delta > 0, \exists G_\delta \ni x, N_\delta \text{ s.t. } \sup_{n > N_\delta} \sup_{y \in G_\delta} g_n(y) \leq g(x) + \delta,$$

$$\iff V1 \text{ holds at } x \text{ for } g(x)$$

The proof of the second equivalence is identical. □

In the light of Lemma 2.7, it is easy to see the truth of the remark before Lemma 2.6 that continuity of the limit function is not sufficient to strengthen the inequalities to equalities. Varadhan's conditions imply a mixture of continuity of g , and a local uniformity of convergence of g_n to g . Hence, if we consider continuous functions converging only point-wise to a continuous limit, we can see that Varadhan's condition will not be met. For example,

consider

$$g_n(x) = \begin{cases} 0 & : x \leq -\frac{2}{n} \text{ or } x \geq 0 \\ nx + 2 & : -\frac{2}{n} \leq x \leq -\frac{1}{n} \\ -nx & : -\frac{1}{n} \leq x \leq 0 \end{cases}$$

The limit of $g_n(x)$ is zero for all x , but there does not exist $G \ni x$ for which $\sup_G g_n$ is not ultimately 1.

If the conditions V1 and V2 both hold at a point $x \in X$ with value $g(x)$, then we say that condition V is satisfied, where the condition is stated as follows.

$$(V) \quad \forall \epsilon > 0, \text{ there exists } G \ni x \text{ and } N \text{ such that } |g_n(y) - g(x)| < \epsilon, \forall n > N.$$

So, if we assume that condition V holds true at a point x , we know from Lemmas 2.6 and 2.7 that the RL function exists there and is given by

$$\mu^g(x) = \mu(x) + g(x),$$

and if condition V is met at every $x \in X$, we know that the tilted measures obey a vague LDP. If X is compact then, as explained before, Varadhan's theorem is immediate; since we know that X is also open, we can use the equal bounds provided by the upper and lower deviation functions.

$$\sup_{x \in X} \mu^g(x) \leq \underline{m}^g[X] \leq \overline{m}^g[X] \leq \sup_{x \in X} \mu^g(x),$$

Using the observation in Lemma 2.9, this can be summarised in the following theorem.

Theorem 2.8 *If X is compact, then for a sequence of locally bounded functions $g_n : X \rightarrow \mathbb{R}$, which satisfy condition V at all $x \in X$*

$$\lim_{n \rightarrow \infty} \frac{1}{V_n} \log \int_X e^{V_n g(x)} \mathbb{M}_n[dx] = \sup_{x \in X} \{g(x) - I(x)\},$$

where $g(x) = \lim_{n \rightarrow \infty} g_n(x)$.

When the space is not compact, it is more difficult to apply the style of argument used by Lewis and Pfister to the case of the measures tilted by a sequence of functions.

Locally uniform convergence

We now wish to detail some implications, and provide alternative formulations, of condition V.

When V holds at a point x , we know that given $\epsilon > 0$, there exists an open neighbourhood $G \ni x$ such that $|g_n(y) - g(x)| < \epsilon$ for all $y \in G$. Since all such G certainly contain x , we obtain the following natural result.

Lemma 2.9 *If V holds at x for the value $g(x)$, then $\lim_{n \rightarrow \infty} g_n(x)$ exists and is equal to $g(x)$.*

In order to make condition V easier to check we prove a couple of theorems about the convergence of sequences of functions. First of all we isolate the convergence aspect of Varadhan's condition.

Definition 6 *The convergence of a sequence of functions, $\{g_n\}$, to the function, g , is locally uniform iff given $\epsilon > 0$ and $x \in X$, there exists a neighbourhood of $G \ni x$, and an N such that,*

$$|g_n(y) - g(y)| < \epsilon, \quad \forall y \in G, n > N.$$

Theorem 2.10 *The convergence of g_n to g is locally uniform and g is continuous if and only if condition V holds at every $x \in X$.*

Proof. Considering the “if” part first, we start by proving the continuity of g . By Lemma 2.9, we know that g is the pointwise limit of the sequence $\{g_n\}$. For any $\epsilon > 0$, condition V tells us that there exists $G \ni x$ and number N such that

$$\begin{aligned} |g_n(y) - g(x)| &< \epsilon, & \forall y \in G, n > N, \\ \Rightarrow \left| \lim_{n \rightarrow \infty} g_n(y) - g(x) \right| &< \epsilon, & \forall y \in G, \\ |g(y) - g(x)| &< \epsilon, & \forall y \in G. \end{aligned}$$

Hence g is continuous.

Now we wish to show that the convergence is locally uniform. Consider $\epsilon > 0$. Since g is continuous there exists a set $G_1 \ni x$ such that

$$|g(y) - g(x)| < \frac{\epsilon}{2}, \quad \forall y \in G_1,$$

and, by condition V, there exists a set $G_2 \ni x$, and a number N such that

$$|g_n(y) - g(x)| < \frac{\epsilon}{2}, \quad \forall y \in G_2, n > N.$$

Hence, as required

$$|g_n(y) - g(y)| < \epsilon, \quad \forall y \in G_1 \cap G_2, n > N.$$

The proof of the “only if” part is almost identical. Again, take $\epsilon > 0$. Since g is continuous, and the convergence of g_n to g is locally uniform, we know there exist open neighbourhoods of x , G_1 and G_2 and a number N such that

$$\begin{aligned} |g_n(y) - g(y)| < \epsilon, \quad |g(y) - g(x)| < \epsilon, \quad \forall y \in G_1 \cap G_2, n > N, \\ \Rightarrow |g_n(y) - g(x)| < 2\epsilon, \quad \forall y \in G_1 \cap G_2, n > N, \end{aligned}$$

which shows that condition V is satisfied. $\square$

For characterising the convergence of sequences of functions, uniform convergence on compact subsets is a more common criterion than locally uniform convergence. The following theorem shows that it is, in fact, sufficient, and in the case of locally compact spaces, necessary.

Theorem 2.11 *If the convergence to g , of the sequence of functions $\{g_n\}$, is locally uniform, then it is uniform on compact subsets of X .*

In a locally compact space, if the sequence converges uniformly to g on compact subsets, then the convergence is locally uniform.

Proof. Assume first that the convergence is locally uniform. Then, given a compact subset K , and $\epsilon > 0$, then for every $x \in K$, we can find a neighbourhood $G_x \ni x$, and a number N_x , such that

$$|g_n(y) - g(y)| < \epsilon, \quad \forall y \in G_x, n > N_x.$$

Since these neighbourhoods cover K and K is compact, there exists a finite collection

$\{G_{x_1}, \dots, G_{x_n}\}$, whose union contains K . Hence

$$|g_n(y) - g(y)| < \epsilon, \quad \forall y \in K, n > \max \{N_{x_1}, \dots, N_{x_n}\}.$$

On the other hand, if X is locally compact then every point, x , has a compact neighbourhood, say K_x . Since convergence is compact on this subset and it contains an open neighbourhood, we can conclude that condition V holds at this point. $\square$

2.4 Stochastic processes

The LDPs which will concern us in the following chapters derive from either the empirical averages, or empirical measures, of stochastic processes.

Empirical Averages

For a stochastic process, $\{Y_n\}$, taking values in Σ , we might be interested in the large deviation properties of the sequence of measures, $\{\mathbb{M}_n\}$, defined by

$$\mathbb{M}_n[B] := P\left(\frac{Y_1 + \dots + Y_n}{n} \in B\right), \text{ for all Borel sets } B.$$

These measures are the laws of the empirical means. We will usually only consider the empirical average of a 2-state process. In that case, the average at time n tells us the number of those n random variables in each of the states; for instance, if $\Sigma = \{0, 1\}$, then the partial sum, $Y_1 + \dots + Y_n$ is the number of times the process has been in state 1.

Empirical Measures

For a process where the cardinality of the state space is greater than two, the partial sum is likely to hide the percentage of the total time spent in each state. To capture the same type of information about such a process, we construct the empirical measure. The empirical measure of a discrete-time process, $\{Y_n\}$, is a measure on Σ , defined by

$$\mu_n[A] := \frac{1}{n} \sum_{i \leq n} \delta_{Y_i}[A], \quad \forall A \subset \Sigma.$$

In order to examine the large deviations of the laws of these measures, we need to specify a topology on the space of measures, $\mathcal{M}(\Sigma)$. In general, this space can have many topologies, but we will only concern ourselves with processes taking values in a finite set. We let the cardinality of Σ be M and, to simplify notation, we take Σ to be the set $\{1, \dots, M\}$. In this case, each measure can be represented as a sum of Dirac measures:

$$\mu = \frac{1}{n} \sum_{i \leq M} \mu[\{i\}] \delta_i.$$

In this way, to each measure, we can associate a vector in $\mathbb{R}_+^M$. We define the set of all probability vectors

$$\Gamma = \{\mathbf{y} \in \mathbb{R}_+^M : y_1 + \dots + y_M = 1\},$$

and the natural mapping to this set from the space of measures,

$$\begin{aligned} \mathcal{C} &: \mathcal{M}(\Sigma) \rightarrow \Gamma, \\ \mathcal{C} &: \mu \mapsto (\mu[\{1\}], \dots, \mu[\{M\}]). \end{aligned}$$

We will refer to the components of the image vector as the *coordinates* of the measure. We take Γ to be equipped with a topology which is the relativisation of the usual topology on $\mathbb{R}^M$. Since $\mathcal{C}$ is bijective, we will talk about a measure and its coordinates interchangeably. In particular, we will refer to an LDP on Γ , for the coordinates of the empirical measure, as an LDP on $\mathcal{M}(\Sigma)$ for the empirical measures.

Relative entropy and accessibility

For the purposes of defining some further notation, we consider a Σ -valued stochastic process $\{Y_n\}_{n \geq 1}$, consisting of independent random variables sharing a common distribution on Σ ; we fix a probability vector $\mathbf{v} \in \Gamma$ to describe this distribution,

$$P(Y_n = i) = v_i, \quad \forall i \in \Sigma, \quad \forall n.$$

Let μ_n be the empirical measure of the process at time n , and define $\mathbf{U}^{(n)} := \mathcal{C}(\mu_n)$ so that

$$U_i^{(n)} = \frac{1}{n} \sum_{j=1}^n 1_{\{Y_j=i\}}(\omega).$$

Of course, the components of this vector must be of the form $\frac{k}{n}$ for $k \in \{0, 1, \dots, n\}$, and so we define the following finite subset of Γ , which contains the attainable vectors at time n ,

$$\Gamma_n = \{\mathbf{q} \in \Gamma : nq_i \in \mathbb{N} \text{ for all } 1 \leq i \leq M\}.$$

Thus, we have that for $\mathbf{y} \in \Gamma \setminus \Gamma_n$, $P(\mathbf{U}^{(n)} = \mathbf{y}) = 0$, and it is easy to see that for $\mathbf{q} \in \Gamma_n$,

$$P(\mathbf{U}^{(n)} = \mathbf{q}) = \frac{n!}{(nq_1)! \dots (nq_M)!} (v_1)^{nq_1} \dots (v_M)^{nq_M}.$$

This function occurs frequently in Chapter 4, and so, with exponential asymptotics in mind, we define, for $\mathbf{q} \in \Gamma_n$,

$$h_n(\mathbf{q}, \mathbf{v}) := \frac{1}{n} \log P(\mathbf{U}^{(n)} = \mathbf{q}) = \frac{1}{n} \log \frac{n!}{(nq_1)! \dots (nq_M)!} - H(\mathbf{q} | \mathbf{v}) - H(\mathbf{q}), \quad (2.3)$$

where $H(\mathbf{u} | \mathbf{v})$, the *relative entropy* of $\mathbf{u}$ with respect to $\mathbf{v}$, is defined for $\mathbf{u} \in \Gamma$ by

$$H(\mathbf{u} | \mathbf{v}) := \sum_{i=1}^M u_i \log \frac{u_i}{v_i},$$

and $H(\mathbf{u})$, the *entropy* of $\mathbf{u} \in \Gamma$ is given by

$$H(\mathbf{u}) := - \sum_{i=1}^M u_i \log u_i.$$

The relative entropy, $H(\cdot | \mathbf{v})$, is the rate function for the laws of the empirical measures of $\{Y_n\}$. We also extend h_n to a set function by the following definition. For B , a Borel subset of Γ , we define

$$h_n(B, \mathbf{v}) := \frac{1}{n} \log \left(\sum_{\mathbf{q} \in B \cap \Gamma_n} e^{nh_n(\mathbf{q}, \mathbf{v})} \right),$$

where, of course,

$$P\left(\mathbf{U}^{(n)} \in B\right) = e^{nh_n(B, \mathbf{v})}.$$

Lastly, we wish to define the set of probability vectors from which a given vector is *accessible*. This definition helps to simplify some technical details of the proofs in Chapter 4. For a process such as $\{Y_n\}$, it is possible to choose $\mathbf{v}$, the vector describing the distribution, to be on the boundary of Γ . In that case, the probability of visiting certain states will be zero. We say that a state is accessible from $\mathbf{v}$ if it has a non-zero probability of being visited. A support for a probability measure is any set whose measure is 1. Similarly, we define the support, $S(\mathbf{v})$, of a probability vector, $\mathbf{v}$, to be

$$S(\mathbf{v}) := \{i : 1 \leq i \leq M, v_i > 0\},$$

and using this, the set of vectors from which $\mathbf{x}$ is accessible is given by

$$\mathcal{A}(\mathbf{x}) := \{\mathbf{v} \in \Gamma : S(\mathbf{x}) \subset S(\mathbf{v})\}.$$

We have, therefore, that for $\mathbf{v} \in \mathcal{A}(\mathbf{x})$, $v_i = 0 \Rightarrow x_i = 0$.

Continuous time processes and measurability

Although the process resulting from the random sampling transformation is necessarily a discrete time process, we will also deal with continuous time processes. The underlying process in many cases will be indexed by $\mathbb{R}_+$, rather than $\mathbb{N}$. The statement of the LDP in continuous time is almost identical to that in discrete time. For example, we can define the empirical average of a process $\{X_t\}_{t \in \mathbb{R}_+}$ by

$$Z_t = \frac{1}{t} \int_0^t X_s ds, \tag{2.4}$$

and then look at an LDP for the sequence $\{\mathcal{L}_{(Z_t)}\}$ on the scale t .

The discussion of LDPs earlier in this chapter goes through analogously when the sequence of measures is indexed by $\mathbb{R}$ rather than $\mathbb{N}$. However, there is a potential difficulty in the definition of quantities such as the integral in equation (2.4). In order to guarantee that functions such as Z_t are measurable, we need to impose an additional measurability constraint on the process $\{X_t\}$. In [5], Doob defines measurability of a stochastic process

in the following way:

Definition 7 A stochastic process $\{X_t\}_{t \in \mathbb{R}}$, defined on the probability triple $(\Omega, \mathcal{F}, \lambda)$, is measurable iff the function $X : \Omega \times \mathbb{R} \rightarrow \mathbb{R}$, where $X(\omega, t) = X_t(\omega)$, is measurable with respect to the product space $(\Omega \times \mathbb{R}, \mathcal{F} \times \mathcal{B}, \lambda \times \text{Leb})$.

Doob discusses stochastic convergence properties that guarantee measurability, and shows how measurability of a stochastic process ensures that integrals of the process, such as in equation (2.4), are measurable.

Throughout the remaining chapters, we will assume that all continuous time processes are measurable.

The Poisson process

In this section, we wish to gather a few facts about the Poisson process, which we will have need of in Chapter 4.

Consider a Poisson process with rate α , $\{N_t\}_{t \in \mathbb{R}_+}$, defined on the probability triple $(\Omega, \mathcal{F}, P)$. We will be principally interested in the related process, $\{T_n\}_{n \in \mathbb{N}}$, which defines the sequence of jump times,

$$T_n = \sup\{t : N_t < n\}.$$

Since N_t is Poisson, $\{T_{n+1} - T_n\}$ are independent and are all exponentially distributed with parameter α . Using this simple structure, it is not difficult to see that the joint law of $\{T_1, \dots, T_n\}$ has a density given by

$$f_{(T_1, \dots, T_n)}(t_1, \dots, t_n) = \begin{cases} \alpha^n e^{-\alpha t_n} & : \text{ if } t_1 \leq \dots \leq t_n \\ 0 & : \text{ otherwise} \end{cases}. \quad (2.5)$$

Of course, by integrating over the first $n - 1$ variables, we see that the law of T_n has the density

$$f_{(T_n)}(t) = \frac{\alpha^n t^{n-1}}{(n-1)!} e^{-\alpha t}. \quad (2.6)$$

We also make use of the fact that the laws of T_n/n obey an empirical average LDP on the scale n . This can be easily seen from Cramér's theorem (see, for example, [4, Chapter 2]). Since T_n is a sum of iid random variables, its scaled cumulant generating function is

given by

$$\begin{aligned}\lambda(\theta) &= \log E e^{\theta T_1}, \\ &= \begin{cases} \log \left(\frac{\alpha}{\alpha - \theta} \right) & , \quad \theta < \alpha \\ +\infty & , \quad \theta \geq \alpha \end{cases}.\end{aligned}\tag{2.7}$$

Hence, we can conclude that $\{\mathcal{L}_{(T_n/n)}\}$ satisfy an LDP with rate function I_T , the Legendre-Fenchel transform of λ ,

$$\begin{aligned}I_T(t) &= \sup_{\theta} \{t\theta - \lambda(\theta)\}, \\ &= \alpha t - \log(\alpha t) - 1.\end{aligned}\tag{2.8}$$

The scaled cumulant generating function

We made use in the previous section of the scaled cumulant generating function (sCGF); this is a very useful tool in the theory of large deviations. In general, for a discrete time process $\{X_n\}$, the sCGF is defined by

$$\lambda(\theta) := \lim_{n \rightarrow \infty} \frac{1}{n} \log E \exp \left(\theta (X_1 + \dots + X_n) \right),$$

where this limit exists. The usefulness of the sCGF stems from the fact that existence of the rate function implies, by Varadhan's theorem, existence of the sCGF, and furthermore that λ is the Legendre-Fenchel transform of I . If, in addition, I is known to be convex, then there is a duality and I is the Legendre transform of λ . This is often the easiest way to calculate the rate function.

Chapter 3

Markov underlying processes

For this initial investigation of random sampling, we will restrict our attention to a specific class of processes. If we examine random sampling in discrete time, we have the option of choosing an uncorrelated underlying process, for instance a Bernoulli process. However, for any choice of observation process (with strictly increasing observation times), it is not hard to see that the observed process is identical in distribution to the underlying process. So, in order to have a non-trivial transformation to examine, we work with a correlated process and choose a Markov process as a relatively simple example.

In fact, we choose to work in continuous time since, despite some additional technical considerations, the calculations are more tractable. We will use a Poisson process to determine the sequence of observations and restrict the underlying process to lie in the class of finite state Markov processes. An almost identical analysis can be carried out in discrete time if we replace the continuous time Markov process by a Markov chain and the observation process by a counting process whose inter-arrival times are geometrically distributed.

3.1 Randomly sampled Markov processes

Let $(\Omega, \mathcal{F}, P)$ be a probability triple. With respect to this triple, we denote the time of the n^{th} observation by the random variable T_n and the number of observations before time t by N_t , where

$$N_t := \sup \{n : T_n \leq t\} .$$

We also define the inter-observation process, $\{W_n\}_{n \in \mathbb{N}}$, by

$$W_n := T_n - T_{n-1},$$

where $T_0 := 0$.

Assumption 3.1 *The inter-observation times $\{W_n\}$ are independent and share a common distribution with W , an almost surely bounded random variable taking values in $\mathbb{R}_+$.*

We denote the underlying process by $\{X_t\}_{t \in \mathbb{R}_+}$, and its countable state space by Σ . For convenience we take Σ to be a copy of $\mathbb{N}$. We define $\mathcal{F}_{=t} := \sigma(X_t)$ and

$$\mathcal{F}_{\leq t} := \sigma(X_s : s \leq t).$$

Assumption 3.2 *$\{X_t\}$ is a time-homogeneous Markov process with transition function*

$$F(i, j; t) := P(X_{s+t} = j \mid X_s = i), \quad \forall s \geq 0.$$

The sample process is then given by

$$Y_n := X_{T_n}.$$

We are interested in finding out how the correlation structure of $\{X_t\}$ is related to that of $\{Y_n\}$. As shown in the following theorem, the Markov character is preserved by the sampling.

Theorem 3.1 *$\{Y_n\}_{n \in \mathbb{N}}$ is a Markov chain with transition matrix*

$$f_{ij} := E F(i, j; W).$$

Proof. We wish to show that for any $i_{n+1} \in \Sigma$,

$$P(Y_{n+1} = i_{n+1} \mid Y_0, \dots, Y_n) = P(Y_{n+1} = i_{n+1} \mid Y_n).$$

$\sigma(Y_0, \dots, Y_n)$ is generated by the family $\mathcal{P}$ consisting of all sets of the form

$$\{\omega : Y_0(\omega) = i_0, \dots, Y_n(\omega) = i_n\}.$$

$\mathcal{P}$ is a partition of Ω , and is countable due to the countability of Σ . Therefore, we can write

$$P(Y_{n+1} = i_{n+1} \mid Y_0, \dots, Y_n)(\omega) = \sum_{\{B \in \mathcal{P} : P(B) > 0\}} \frac{P(\{Y_{n+1} = i_{n+1}\} \cap B)}{P(B)} 1_B(\omega), \quad (3.1)$$

and

$$P(Y_{n+1} = i_{n+1} \mid Y_n)(\omega) = \sum_{\{k \in \mathbb{N} : P(Y_n = k) > 0\}} \frac{P(Y_{n+1} = i_{n+1}, Y_n = k)}{P(Y_n = k)} 1_{\{Y_n = k\}}(\omega). \quad (3.2)$$

We can calculate these functions explicitly. For an arbitrary set

$$B = \{Y_0 = i_0, \dots, Y_n = i_n\} \in \mathcal{P},$$

we can use Theorem 4.1 to write

$$\begin{aligned} & P(\{Y_{n+1} = i_{n+1}\} \cap B) \\ &= P(Y_0 = i_0, \dots, Y_{n+1} = i_{n+1}) \\ &= \int_{0 \leq t_1 \leq \dots \leq t_{n+1}} \mathcal{L}_{(T_1, \dots, T_{n+1})}(dt_1 \times \dots \times dt_{n+1}) P(X_0 = i_0, X_{t_1} = i_1, \dots, X_{t_{n+1}} = i_{n+1}). \end{aligned}$$

Now, since $\{X_t\}$ is a Markov process, we can rewrite the integrand as

$$\begin{aligned} & P(X_0 = i_0, X_{t_1} = i_1, \dots, X_{t_{n+1}} = i_{n+1}) \\ &= P(X_0 = i_0, \dots, X_{t_n} = i_n) F(i_n, i_{n+1}; t_{n+1} - t_n). \end{aligned}$$

This expression only depends on t_{n+1} through the difference $(t_{n+1} - t_n)$ and so, since $\{T_n\}$ has independent increments, we can split the integral into a product

$$\begin{aligned}
 & P(\{Y_{n+1} = i_{n+1}\} \cap B) \\
 &= \int_{0 \leq t_1 \leq \dots \leq t_n} \mathcal{L}_{(T_1, \dots, T_n)}(dt_1 \times \dots \times dt_n) P(X_0 = i_0, X_{t_1} = i_1, \dots, X_{t_n} = i_n) \\
 &\quad \times \int_{w=0}^{\infty} \mathcal{L}_{(W)}(dw) F(i_n, i_{n+1}; w) \\
 &= P(Y_0 = i_0, \dots, Y_n = i_n) E\left(F(i_n, i_{n+1}; W)\right) \\
 &= P(B) E\left(F(i_n, i_{n+1}; W)\right).
 \end{aligned}$$

By summing over $i_j \in \Sigma$, for each $j \leq n-1$, and taking $i_n = k$, we can also conclude that

$$P(Y_{n+1} = i_{n+1}, Y_n = k) = P(Y_n = k) E\left(F(i_n, i_{n+1}; W)\right).$$

Hence, using the equations (3.1,3.2), we can see that

$$P(Y_{n+1} = i | Y_n) = P(Y_{n+1} = i | Y_0, \dots, Y_n) = \sum_{j \in \Sigma} E\left(F(j, i; W)\right) 1_{\{Y_n=j\}}.$$

□

In order to further our investigation, we now choose to be more specific about the underlying process.

Assumption 3.3 $\{X_t\}$ is a stationary 2-state process, taking values in $\{0, 1\}$, with transition function

$$F(i, j; t) = e^{tA} := \sum_{n \geq 0} \frac{t^n}{n!} A^n,$$

where, with $\lambda_0, \lambda_1 > 0$,

$$A := \begin{pmatrix} -\lambda_0 & \lambda_0 \\ \lambda_1 & -\lambda_1 \end{pmatrix}.$$

From the condition that $\{X_t\}$ be stationary, we know that its starting distribution must be the invariant distribution of its transition function, which is defined by

$$P(X_0 = 1) = \frac{\lambda_0}{\lambda_0 + \lambda_1}.$$

Starting from this distribution, the process then consists of a sequence of alternate sojourn times in states 0 and 1, whose lengths are exponentially distributed with parameters λ_0, λ_1 respectively.

In the light of this further assumption, $\{Y_n\}$, being a stationary 2-state Markov chain, is characterised by the two transition probabilities

$$a := P(Y_{n+1} = 1 | Y_n = 0) ,$$

$$d := P(Y_{n+1} = 0 | Y_n = 1) .$$

By applying Theorem 3.1, we can calculate these probabilities in the case where the observation process is Poisson.

Theorem 3.2 *If W is exponentially distributed with parameter α then*

$$a = \frac{\lambda_0}{\lambda_0 + \lambda_1 + \alpha} \quad \text{and} \quad d = \frac{\lambda_1}{\lambda_0 + \lambda_1 + \alpha} .$$

Proof. We first calculate a . Using Theorem 3.1, we can write

$$a = E[e^{AW}]_{01} = \int_0^\infty [e^{tA}]_{01} \alpha e^{-\alpha t} dt . \quad (3.3)$$

We first need to evaluate e^{tA} . Because of the simple form of A , it is easy to see that

$$\begin{aligned} A^2 &= (-\lambda_0 - \lambda_1)A \\ \Rightarrow A^n &= (-\lambda_0 - \lambda_1)^{n-1}A, \end{aligned}$$

and so,

$$\begin{aligned} e^{tA} &= \sum_{n=0}^\infty \frac{(tA)^n}{n!} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{e^{-(\lambda_0+\lambda_1)t} - 1}{\lambda_0 + \lambda_1} \begin{pmatrix} -\lambda_0 & \lambda_0 \\ \lambda_1 & -\lambda_1 \end{pmatrix}, \end{aligned}$$

Substituting this in equation (3.3) gives

$$a = \int_0^\infty \frac{\lambda_0}{\lambda_0 + \lambda_1} (1 - e^{-(\lambda_0+\lambda_1)t}) \alpha e^{-\alpha t} dt$$

$$= \frac{\lambda_0}{\lambda_0 + \lambda_1 + \alpha},$$

The derivation for d is almost identical. $\square$

We now wish to examine the large deviations of $\{Y_n\}$. Dembo and Zeitouni show, in [4, Theorem 3.1.2], that a finite-state Markov chain with irreducible transition matrix will obey an LDP. Furthermore, the rate function, J , is given by a Legendre-Fenchel transform,

$$J(y) = \sup_{\theta \in \mathbb{R}} \{y\theta - \lambda(\theta)\}. \quad (3.4)$$

$\lambda(\theta)$, the sCGF, is the logarithm of the largest eigenvalue of the twisted transition matrix, which is obtained from the transition matrix by multiplying the elements of row k by $\exp(\theta k)$. Having restricted to the 2-state case we have explicit formulae for the transition rates and the matrix in question is given by

$$\begin{pmatrix} 1-a & ae^\theta \\ d & (1-d)e^\theta \end{pmatrix}.$$

This matrix is obviously irreducible since λ_0 , λ_1 and μ are all finite and greater than zero. Solving the characteristic equation for this matrix, we see that

$$\lambda(\theta) = \log \left[\frac{(1-a) + (1-d)e^\theta + \sqrt{[(1-a) + (1-d)e^\theta]^2 - 4(1-a-d)e^\theta}}{2} \right].$$

To obtain an explicit form for J , we need to complete the optimisation in equation (3.4). Unfortunately, due to the form of λ , this calculation reduces to the solution of a quartic equation in e^θ for the optimising value of θ .

3.2 The rate function of a Markov chain

Taking a more abstract approach to the random sampling problem provides a novel expression for the rate function of the empirical mean LDP of a 2-state Markov chain. This approach relies on a general result, Theorem 4.5, concerning the large deviations of random sampling. If the observation process is Poisson and the underlying process satisfies an em-

pirical measure LDP with rate function I , then the observed process will satisfy an LDP with a rate-function that can be expressed in terms of I .

As explained in Section 2.4, in the case of a 2-state process the empirical mean LDP is identical to the empirical measure LDP. We use $I(x)$ to denote the rate function for the empirical mean of $\{X_t\}$. Applying Theorem 4.5, we obtain an alternative expression for the rate function for $\{Y_n\}$,

$$J(y) = \inf_{0 \leq x \leq 1} \left\{ H(y|x) + \log \left(\frac{I(x)}{\alpha} + 1 \right) \right\}, \quad (3.5)$$

where,

$$H(y|x) = y \log \frac{y}{x} + (1-y) \log \frac{1-y}{1-x}.$$

The Markov fluid rate function

In order to complete the optimisation in equation (3.5), we need an expression for the $I(x)$. This can be calculated by means of the Legendre-Fenchel transformation once the scaled cumulant generating function for the 2-state Markov process is known. In [6, Section 2.3.2], Kelly provides an expression for the effective bandwidth function for a process identical to $\{X_t\}$, assuming that in Kelly's model we set $h = 1, \lambda = \lambda_0, \mu = \lambda_1$. Kelly defines the effective bandwidth function, α , for a process whose partial sum up to time t is denoted Z_t , to be

$$\alpha(s, t) = \frac{1}{st} \log E [\exp (sZ_t)],$$

from which it follows that the sCGF for such a process can be written as

$$\mu(\theta) = \theta \lim_{t \rightarrow \infty} \alpha(\theta, t).$$

Using the expression provided for this limit by Kelly, we find that the sCGF for $\{X_t\}$ is given by

$$\mu(\theta) = \frac{1}{2} \left(\theta - \lambda_1 - \lambda_0 + \sqrt{(\theta - \lambda_1 + \lambda_0)^2 + 4\lambda_0\lambda_1} \right).$$

Calculation of the Legendre-Fenchel transform of this function is quite straightforward and a small amount of analysis yields an explicit form for $I(x)$.

$$\begin{aligned} I(x) &= \sup_{\theta \in \mathbb{R}} \{x\theta - \mu(\theta)\} \\ &= \lambda_1 x + \lambda_0(1-x) - 2\sqrt{\lambda_0 \lambda_1 x(1-x)} \end{aligned}$$

The 2-state chain rate function

Using this expression in equation (3.5), it is possible to perform the optimisation analytically. In Chapter 5, we prove results useful in calculating the infimum. In particular, Theorem 5.2 shows how the infimum is attained. To check that $I(x)$ satisfies the conditions of the theorem, we note first that I is continuous. Secondly, it is easy to see that I must be positive because if

$$I(x) < 0$$

then, since $\lambda_1 x$ and $\lambda_0(1-x)$ are positive,

$$\lambda_1 x - 2\sqrt{\lambda_1 x \lambda_0(1-x)} + \lambda_0(1-x) < 0,$$

which implies that

$$\left(\sqrt{\lambda_1 x} - \sqrt{\lambda_0(1-x)}\right)^2 < 0,$$

and this is obviously false. Finally, the convexity of I is ensured by considering its second derivative

$$I''(x) = \sqrt{\frac{\lambda_0 \lambda_1}{(x(1-x))}},$$

which is strictly positive.

Therefore, we can conclude that for $y \in (0, 1)$, the infimum in equation (3.5) is attained at the point x which is the unique solution of the equation

$$\frac{x-y}{x(1-x)} - \frac{I'(x)}{I(x)+\alpha} = 0. \quad (3.6)$$

After substitution of explicit forms for I and I' this equation can be reduced to the form

$$Ax - B = C\sqrt{x(1-x)}, \quad (3.7)$$

where

$$A = \alpha + \lambda_0 y + \lambda_1(1-y), \quad B = y(\alpha + \lambda_0), \quad C = \sqrt{\lambda_0 \lambda_1}(1-2y). \quad (3.8)$$

Squaring both sides of this equation yields a quadratic equation in x whose solutions are

$$x_{\pm} = \frac{2AB + C^2 \pm C\sqrt{4AB + C^2 - 4B^2}}{2(A^2 + C^2)}. \quad (3.9)$$

Of these two roots, one solves equation (3.7) and the other solves the equation

$$Ax - B = -C\sqrt{x(1-x)}. \quad (3.10)$$

We denote the solution of equation (3.10) by x_2 and the solution of equation (3.7) by x_1 and can conclude, since the function $Ax - B$ is increasing and the function $C\sqrt{x(1-x)}$ is positive, that $x_2 \leq x_1$. The sign of C determines which of the two roots, $x_{\pm}$, is larger; this in turn depends upon the sign of $(1-2y)$. Hence, the solution of equation (3.6) must be

$$x(y) := \begin{cases} x_+ & : y \leq \frac{1}{2} \\ x_- & : y > \frac{1}{2} \end{cases} \quad (3.11)$$

and so, for $y \in (0, 1)$,

$$J(y) = H(y|x(y)) + \log \left(\frac{I(x(y))}{\alpha} + 1 \right).$$

For $y = 0$ or $y = 1$, Theorem 5.2 tells us that the infimum will be attained on the interior of $[0, 1]$ if, respectively,

$$\lim_{x \rightarrow 0} \frac{I'(x)}{I(x) + \alpha} < -1, \quad \text{or} \quad \lim_{x \rightarrow 1} \frac{I'(x)}{I(x) + \alpha} > 1.$$

I is finite at both endpoints and it is clear from the form of I'

$$I'(x) = \lambda_1 - \lambda_0 - \sqrt{\lambda_0 \lambda_1} \frac{1 - 2x}{\sqrt{x(1-x)}},$$

that it diverges to $-\infty$ at 0, and to $+\infty$ at 1. The infimum being attained on the interior for both $y = 0$ and $y = 1$, Theorem 5.2 shows that the point at which the infimum occurs can be found, in both cases, by differentiation. Running through some analysis, which is similar to, but slightly simpler than, that for $y \in (0, 1)$, yields the expressions

$$x(0) = \frac{\lambda_0 \lambda_1}{(\lambda_1 + \alpha)^2 + \lambda_0 \lambda_1}, \quad x(1) = \frac{(\lambda_0 + \alpha)^2}{(\lambda_0 + \alpha)^2 + \lambda_0 \lambda_1},$$

which are the values of x_+ , and x_- respectively, with $y = 0$ or $y = 1$ in the definitions of A , B and C .

In summary, we can state that for $y \in [0, 1]$, with A , B and C defined as in equation (3.8) and $x(y)$ defined as in equations (3.11, 3.9),

$$J(y) = H(y | x(y)) + \log \left(\frac{I(x(y))}{\alpha} + 1 \right). \quad (3.12)$$

We can use this formula to provide an explicit, albeit complicated, expression for the rate function of any positively correlated Markov chain. For a Markov chain with transition probabilities a and d , we define the burstiness,

$$b := 1 - a - d.$$

This parameter measures the strength of the correlation structure, through the relation,

$$\begin{aligned} \text{corr}(X_0, X_n) &:= \frac{\text{cov}(X_0, X_n)}{\sqrt{\text{var}(X_0) \text{var}(X_n)}} \\ &= b^n. \end{aligned}$$

Since a and d are probabilities, b can, in general, take values between -1 and 1 . A Markov chain is said to be positively correlated when the burstiness is positive.

By Theorem 3.2 we know that a 2-state Markov process sampled according to a Poisson

observation process results in a 2-state Markov chain. This chain will necessarily have a positive burstiness since, using the notation of the previous section, we can see that

$$b = \frac{\alpha}{\lambda_0 + \lambda_1 + \alpha}.$$

This is to be expected since the underlying process cannot be negatively correlated. We can turn this around and represent any positively correlated Markov chain by a randomly sampled continuous time Markov process. If we are given a chain with transition probabilities a and d , then for some $\alpha > 0$, we define the transition rates

$$\lambda_0 = \frac{a\alpha}{1-a-d}, \quad \lambda_1 = \frac{d\alpha}{1-a-d},$$

α is a free parameter which determines the time scale on which the Markov process will change state. The rate function for the chain is then given by equation (3.12).

Chapter 4

Poisson sampling : I

In this chapter, we consider random sampling where the underlying process does not have to be Markovian. Instead, we ask only that it satisfy an LDP and require that the observation process be Poisson. Due to the specific structure of the Poisson distribution, we can show that the rate function for the observed process is related to that of the underlying process by means of an explicit formula involving the relative entropy function, $H(\cdot | \cdot)$.

4.1 A rough sketch

In this section, we wish to provide an outline of the proof of the main result, Theorem 4.5. To illustrate roughly how the formula is derived, we consider $\{X_t\}$, a 2-state process in continuous time. We take the state space to be $\{0, 1\}$ and assume an empirical average LDP. We denote the Poisson process by $\{N_t\}$, and the corresponding sequence of jump times by $\{T_n\}$, where $T_n = \sup\{t : N_t < n\}$. A well known property of the Poisson process tells us that, conditioned on the event $\{N_t = n\}$, and up to a re-ordering, the variables $T_1, \dots, T_n$ are independent and uniformly distributed in $[0, t]$. We are interested in calculating a quantity of the form

$$P\left(\frac{X_{T_1} + \dots + X_{T_n}}{n} \approx y\right), \quad (4.1)$$

and so the question of re-ordering need not concern us. We examine this probability conditioned on the event $\{T_{n+1} \approx t\}$ and also on the empirical average of the process up to

time T_{n+1} being x . Roughly speaking, the former ensures that the observation times are independent and uniformly distributed in $[0, t]$; the latter ensures that the process has spent a fraction x of its time in state 1. Hence, for $1 \leq i \leq n$, $P(X_{T_i} = 1)$ is just x . Therefore, we would expect this conditional probability to behave like a Bernoulli process; in other words,

$$P\left(\frac{X_{T_1} + \dots + X_{T_n}}{n} \approx y \mid T_{n+1} \approx t, \frac{1}{T_{n+1}} \int_0^{T_{n+1}} X_s ds \approx x\right) = \binom{n}{ny} x^{ny} (1-x)^{n(1-y)}.$$

It is easy to see, by Stirling's formula, that

$$b_n(y, x) := \frac{1}{n} \log \left[\binom{n}{ny} x^{ny} (1-x)^{n(1-y)} \right] \rightarrow H(y|x),$$

where the relative entropy reduces, in the 2-state case, to

$$H(y|x) = y \log \frac{y}{x} + (1-y) \log \frac{1-y}{1-x}.$$

To regain the original probability in the expression (4.1), we need to integrate over all possible values of T_{n+1} and $X_{T_{n+1}}$. This leads to the expression

$$\begin{aligned} P\left(\frac{X_{T_1} + \dots + X_{T_n}}{n} \approx y\right) &= \int_{0 \leq t} \int_{0 \leq x \leq 1} e^{nb_n(y,x)} \mathcal{L}_{(X_t)}[dx] \mathcal{L}_{(T_{n+1})}[dt], \\ &= \int_{0 \leq x \leq 1} e^{nb_n(y,x)} \mathcal{L}_{(X_{T_{n+1}})}[dx]. \end{aligned}$$

The random time-change, defined by Russell in [9], tells us that the measures $\{\mathcal{L}_{(X_{T_{n+1}})}\}$ obey an LDP and provides a formula for the rate function. We can also show that the convergence of the functions $\{b_n(y, \cdot)\}$ to $-H(y|\cdot)$ is locally uniform. Hence, we can apply the version of Varadhan's theorem extended to sequences of functions. Finally, we arrive at the expression

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P\left(\frac{X_{T_1} + \dots + X_{T_n}}{n} \in B\right) = - \inf_{0 \leq x \leq 1} \left\{ H(y|x) + \log \left(\frac{I(x)}{\alpha} + 1 \right) \right\}.$$

To convert this heuristic argument into a proof, we need to make certain aspects more precise; while doing so, we also allow $\{X_t\}$ to be any process with a finite state space. We replace undefined expressions of the form $\{Y \approx y\}$ by $\{Y \in B\}$, for some suitably small

neighbourhood $B \ni y$. We also need to prove that the convergence of the multi-dimensional analogue of $\{b_n\}$ to $H(\cdot | \cdot)$ is locally uniform. In fact, as a consequence of the introduction of neighbourhoods, this convergence must be established for related set functions, making the proof of Lemma 4.4 quite technical. The main work in the proof of the central result, Theorem 4.5, lies in showing that conditional probability exhibits the required Bernoulli-like behaviour.

4.2 The details

In proving the central result of this chapter, we need to be able to treat separately the effects of the randomness due to the underlying process and that due to the observation process. The following simple theorem allows us to examine the effect of conditioning on particular events determined by the observation process.

Theorem 4.1 *Let $\{X_t\}_{t \in \mathbb{R}^+}$ be a real valued measurable stochastic process, and let $\{T_n\}_{n \in \mathbb{N}}$ be a sequence of positive random variables. If the two processes are independent and f is a real valued measurable function on $\mathbb{R}^n$, then*

$$\begin{aligned} & \mathbb{P}\left(f(X_{T_1}, \dots, X_{T_n}) \in A, (T_1, \dots, T_n) \in B\right) \\ &= \int_{(t_1, \dots, t_n) \in B} \mathbb{P}\left(f(X_{t_1}, \dots, X_{t_n}) \in A\right) \mathcal{L}_{(T_1, \dots, T_n)}(dt_1 \times \dots \times dt_n), \end{aligned}$$

where $A \subset \mathbb{R}$ and $B \subset \mathbb{R}^n$ are Borel and $\mathcal{L}_{(T_1, \dots, T_n)}$ is the joint law of $T_1, \dots, T_n$.

Proof. Let $(\Omega, \mathcal{F}, P)$ be the underlying probability space. For $\omega \in \Omega$, let $X_\bullet(\omega) \in \mathcal{S}_1$ denote a sample path of the process, where $\mathcal{S}_1 = \mathbb{R}^+$ is the space of all sample paths, and note that $\{T_i\}_{1 \leq i \leq n}$ is a random vector taking values in $\mathcal{S}_2 = \mathbb{R}_+^n$. Since $X_\bullet$ is independent of $\{T_n\}$, we can take $\Omega = \mathcal{S}_1 \times \mathcal{S}_2$, $P = \mathcal{L}_{(X_\bullet)} \times \mathcal{L}_{(T_1, \dots, T_n)}$ and $\mathcal{F} = \mathcal{B}^{\mathbb{R}^+} \times \mathcal{B}^n$, where $\mathcal{B}$ is the Borel σ -algebra on $\mathbb{R}$.

We can use the product structure to verify the result, once we have checked the measurability of the function

$$F : \mathcal{S}_1 \times \mathcal{S}_2 \longrightarrow \mathbb{R}^{2n}$$

where,

$$F(\phi(\cdot), t_1, \dots, t_n) = (\phi(t_1), \dots, \phi(t_n), t_1, \dots, t_n).$$

It is easily shown that a function, f , from $(\Omega, \mathcal{F})$ to $(\Omega', \mathcal{F}')$ is measurable if there exists $\mathcal{B}$ such that $\mathcal{F} = \sigma(\mathcal{B})$ and $f^{-1}B \in \mathcal{F}$ for all $B \in \mathcal{B}$. Hence F will be measurable if, for any Borel sets in $\mathbb{R}$, $\{A_i\}_{i \leq n}, \{B_i\}_{i \leq n}$, the set

$$\{(\phi, t_1, \dots, t_n) : \phi(t_i) \in A_i, t_i \in B_i, 1 \leq i \leq n\}$$

is measurable with respect to the product σ -algebra on $\mathcal{S}_1 \times \mathcal{S}_2$. But this set is just

$$\bigcap_{1 \leq i \leq n} \{(\phi, t_1, \dots, t_n) : \phi(t_i) \in A_i, t_i \in B_i\}. \quad (4.2)$$

Now, since $\{X_t\}$ is measurable, the set

$$\{(X_\bullet, t) : X_t \in A\}$$

is measurable; it follows that each of the sets in 4.2 is measurable, and hence so is the intersection.

Applying this to the event in question, and using the product structure we obtain:

$$\begin{aligned} & P\left(f(X_{T_1}, \dots, X_{T_n}) \in A, (T_1, \dots, T_n) \in B\right) \\ &= (\mathcal{L}_{(X_\bullet)} \times \mathcal{L}_{(T_1, \dots, T_n)}) \left[F^{-1}(f^{-1}A \times B)\right] \\ &= \int_{\mathbb{R}^n} \mathcal{L}_{(T_1, \dots, T_n)}(dt_1 \times \dots \times dt_n) \mathcal{L}_{(X_\bullet)} [\{\phi : F(\phi(\cdot), t_1, \dots, t_n) \in f^{-1}A \times B\}] \\ &= \int_B \mathcal{L}_{(T_1, \dots, T_n)}(dt_1 \times \dots \times dt_n) \mathcal{L}_{(X_\bullet)} [\{\phi : f(\phi(t_1), \dots, \phi(t_n)) \in A\}] \\ &= \int_{(t_1, \dots, t_n) \in B} P(f(X_{t_1}, \dots, X_{t_n}) \in A) \mathcal{L}_{(T_1, \dots, T_n)}(dt_1 \times \dots \times dt_n) \end{aligned}$$

which is the required result. $\square$

We now describe the precise setting in which we shall consider random sampling. Let $(\Omega, \mathcal{F}, P)$ be a probability triple. Our underlying process, $\{X_t\}_{t \in \mathbb{R}_+}$, takes values in $\Sigma = \{1, \dots, M\}$. The observation process, $\{T_n\}_{n \geq 1}$, takes values in $(0, \infty)$ and is strictly increasing; we define

$$N_t := \sup\{n : T_n \leq t\},$$

as the number of observations before time t . The observed process, $\{Y_n\}_{n \geq 1}$, is then given by

$$Y_n = X_{T_n}.$$

Assumption 4.1 *The processes $\{X_t\}$ and $\{T_n\}$ are independent of each other.*

We define ν_t to be the empirical measure of $\{X_t\}$ at time t ,

$$\nu_t := \sum_{i=1}^M V_i^{(t)} \delta_i,$$

where

$$V_i^{(t)}(\omega) := \frac{1}{t} \int_0^t 1_{\{X_s=i\}}(\omega) ds$$

denotes the fraction of the time before t spent in state i . As discussed in Section 2.4, we talk interchangeably of an LDP for the empirical measure and an LDP for the vector $\mathbf{V}^{(t)}$ on Γ . For the remainder of this chapter we assume that the underlying process obeys an empirical measure LDP.

Assumption 4.2 *The laws of $\mathbf{V}^{(t)}$ obey an LDP on the scale t , with rate function I_X .*

As mentioned in the heuristic argument given at the start of this chapter, the large deviation properties of the process obtained by examining $\mathbf{V}^{(t)}$ at the sampling times $\{T_n\}$ are described by the random time-change investigated by Russell in [9]. In his treatment, $\{X_t\}$ and $\{N_t\}$ are defined in a more general setting and do not have to be independent of each other. In our case, the mutual independence of the two processes allows for a simple and novel proof of the result. We require the following lemma, which is used in proving a result akin to locally uniform convergence.

Lemma 4.2 *If a function $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ satisfies*

$$\liminf_{x \rightarrow \infty} f(x) \geq \alpha,$$

then for $t > 0$

$$\sup_{G \ni t} \liminf_{x \rightarrow \infty} \inf_{s \in G} sf(xs) \geq t\alpha.$$

Proof. By assumption, given $\epsilon > 0$, there exists $B(\epsilon)$ such that

$$f(x) > \alpha - \epsilon, \quad \forall x > B(\epsilon).$$

Consider $t > 0$ and $\epsilon > 0$. Since

$$\liminf_{x \rightarrow \infty} \inf_{s \in G} sf(xs)$$

is a decreasing set function, we can choose to only consider intervals containing t whose infimum is greater than zero (any neighbourhood of t will contain such a set). We let $G = (t_1, t_2) \ni t$, where $t_1 > 0$ and define $B'(\epsilon) = B(\epsilon/t_2)/t_1$. Then, for any $x > B'(\epsilon)$ and $s \in G$,

$$\begin{aligned} xs &> B\left(\frac{\epsilon}{t_2}\right), \\ sf(xs) &> s\left(\alpha - \frac{\epsilon}{t_2}\right), \\ \Rightarrow \inf_{s \in G} sf(xs) &\geq \inf_{s \in G} \left\{ s\left(\alpha - \frac{\epsilon}{t_2}\right) \right\}, \quad \forall x > B'(\epsilon), \\ \liminf_{x \rightarrow \infty} \inf_{s \in G} sf(xs) &\geq \inf_{s \in G} \left\{ s\left(\alpha - \frac{\epsilon}{t_2}\right) \right\}. \end{aligned}$$

If $\alpha \leq 0$, then $\alpha - \epsilon/t_2 \leq 0$ and

$$\inf_{s \in G} \left\{ s\left(\alpha - \frac{\epsilon}{t_2}\right) \right\} = (\sup G) \left(\alpha - \frac{\epsilon}{t_2}\right) = t_2\alpha - \epsilon.$$

Therefore,

$$\liminf_{x \rightarrow \infty} \inf_{s \in G} sf(xs) \geq t_2\alpha - \epsilon,$$

and since this holds for any $\epsilon > 0$, we have that

$$\liminf_{x \rightarrow \infty} \inf_{s \in G} sf(xs) \geq t_2\alpha,$$

and since $\alpha < 0$

$$\sup_{G \ni t} \liminf_{x \rightarrow \infty} \inf_{s \in G} sf(xs) \geq t\alpha.$$

On the other hand, if $\alpha > 0$, then we can choose $\epsilon < \alpha t_2$. In that case,

$$\inf_{s \in G} \left\{ s \left(\alpha - \frac{\epsilon}{t_2} \right) \right\} = (\inf G) \left(\alpha - \frac{\epsilon}{t_2} \right) = t_1 \alpha - \frac{t_1}{t_2} \epsilon.$$

Hence,

$$\liminf_{x \rightarrow \infty} \inf_{s \in G} sf(xs) \geq t_1 \alpha - \frac{t_1}{t_2} \epsilon,$$

and since this holds for any positive $\epsilon < \alpha t_2$, we have that

$$\liminf_{x \rightarrow \infty} \inf_{s \in G} sf(xs) \geq t_1 \alpha.$$

Since $\alpha > 0$

$$\sup_{G \ni t} \liminf_{x \rightarrow \infty} \inf_{s \in G} sf(xs) \geq t \alpha.$$

□

The next theorem is the statement of the random time-change, in our simplified setting. The proof relies on applying Lemma 4.2 to obtain large deviation bounds for open and closed sets, and then showing that LDP1 and LDP2 are also satisfied.

Theorem 4.3 *If the sequence $\{\mathcal{L}_{(T_n/n)}\}_{n \geq 1}$ obeys an LDP on the scale n with rate function I_T , then the sequence $\{\mathcal{L}_{(\mathbf{V}^{T_n})}\}_{n \geq 1}$ obeys an LDP on the scale n with rate function*

$$J(\mathbf{v}, t) = \inf_{t \in \mathbb{R}_+} \{t I_X(\mathbf{v}) + I_T(t)\}.$$

Proof. We are interested in the large deviations of the laws of $\{\mathbf{V}^{(T_n)}\}$. Using Theorem 4.1 we can express $\{\mathcal{L}_{(\mathbf{V}^{T_n})}\}_{n \geq 1}$ as a tilted form of $\{\mathcal{L}_{(T_n/n)}\}_{n \geq 1}$. Then Varadhan's theorem can be used to calculate upper and lower bounds on the exponential asymptotics of these laws.

Let B be a measurable set in $\mathbb{R}^M$. Using Theorem 4.1, we can write

$$\begin{aligned} \mathbb{P}(\mathbf{V}^{(T_n)} \in B) &= \int_{t>0} \mathbb{P}(\mathbf{V}^{(t)} \in B) \mathcal{L}_{(T_n)}[dt] \\ &= \int_{t>0} \mathbb{P}(\mathbf{V}^{(nt)} \in B) \mathcal{L}_{(\frac{T_n}{n})}[dt]. \end{aligned} \tag{4.3}$$

We define the function $h : \mathbb{R}_+^2 \rightarrow \mathbb{R}$ by

$$(x, t) \mapsto h_x(t) := \frac{1}{x} \log P\left(\mathbf{V}^{(xt)} \in B\right).$$

Using this function, we can write equation (4.3) as

$$P\left(\mathbf{V}^{(T_n)} \in B\right) = \int_{t>0} e^{nh_n(t)} \mathcal{L}_{\left(\frac{T_n}{n}\right)}[dt].$$

We note first that

$$h_x(t) = th_{xt}(1). \quad (4.4)$$

Using x to denote a real number and n to denote a natural number, we then consider the bounds for open and closed sets separately.

- Taking $B = G \in \mathcal{G}$, the LDP for the underlying process tells us that

$$\liminf_x h_x(1) \geq - \inf_{\mathbf{v} \in G} I_X(\mathbf{v}).$$

Given this lower bound, we can conclude from Lemma 4.2 that for all $t > 0$,

$$\sup_{T \ni t} \liminf_{x \rightarrow \infty} \inf_{s \in T} sh_{xs}(1) \geq -t \inf_{\mathbf{v} \in G} I_X(\mathbf{v}),$$

which implies

$$\sup_{T \ni t} \liminf_{n \rightarrow \infty} \inf_{s \in T} sh_{ns}(1) \geq -t \inf_{\mathbf{v} \in G} I_X(\mathbf{v}),$$

and by equation (4.4) we get

$$\sup_{T \ni t} \liminf_{n \rightarrow \infty} \inf_{s \in T} h_n(s) \geq -t \inf_{\mathbf{v} \in G} I_X(\mathbf{v}). \quad (4.5)$$

By Theorem 2.7, equation (4.5) guarantees that, at every $t > 0$, the sequence $\{h_n(\cdot)\}$ satisfies V2, Varadhan's convergence condition for the lower bound. Since we have assumed that $\{\mathcal{L}_{(T_n/n)}\}$ obeys an LDP we can conclude, by Theorem 2.3, that

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \log P\left(\mathbf{V}^{(T_n)} \in G\right) &\geq \sup_{t>0} \left\{ -t \inf_{\mathbf{v} \in G} I_X(\mathbf{v}) - I_T(t) \right\} \\ &= - \inf_{\mathbf{v} \in G} \inf_{t>0} \{tI_X(\mathbf{v}) + I_T(t)\}. \end{aligned}$$

- Similarly, taking $B = F \in \mathcal{F}$, the LDP for the underlying process tells us that

$$\limsup_x h_x(1) \leq - \inf_{\mathbf{v} \in F} I_X(\mathbf{v}).$$

We can convert this into a lower bound of the form

$$\liminf_x (-h_x(1)) \geq \inf_{\mathbf{v} \in F} I_X(\mathbf{v}),$$

and, by a similar argument to that just used for the lower bound, we can conclude that for all $t > 0$,

$$\begin{aligned} \sup_{T \geq t} \liminf_{n \rightarrow \infty} \inf_{s \in T} (-h_n(s)) &\geq t \inf_{\mathbf{v} \in F} I_X(\mathbf{v}), \\ \inf_{T \geq t} \limsup_{n \rightarrow \infty} \sup_{s \in T} h_n(s) &\leq -t \inf_{\mathbf{v} \in F} I_X(\mathbf{v}). \end{aligned} \quad (4.6)$$

By Theorem 2.7, equation (4.6) guarantees that the sequence, $\{h_n\}$, satisfies V1, Varadhan's convergence condition for the upper bound. Obviously, h_n is bounded above by 0 for all n . Therefore, we can conclude by Theorem 2.2 that for all $F \in \mathcal{F}$,

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log P(\mathbf{V}^{(T_n)} \in F) &\leq \sup_{t > 0} \left\{ -t \inf_{\mathbf{v} \in F} I_X(\mathbf{v}) - I_T(t) \right\} \\ &= - \inf_{\mathbf{v} \in F} \inf_{t > 0} \{ t I_X(\mathbf{v}) + I_T(t) \}. \end{aligned}$$

Therefore, the function

$$J(\mathbf{v}) := \inf_{t > 0} \{ t I_X(\mathbf{v}) + I_T(t) \} \quad (4.7)$$

is a candidate for a rate function. It provides the appropriate upper and lower bounds, meaning that LDP3 and LDP4 are satisfied. It remains to check that it satisfies LDP1 and LDP2. However, the space Γ is compact, and so, as noted in Section 2.2, LDP1 implies LDP2. Hence, to conclude that the laws of $\mathbf{V}^{(T_n)}$ satisfy an LDP with rate function J we need only check that J is lower semi-continuous.

J is lsc if it is equal to its lsc regularisation, i.e. if

$$J(\mathbf{v}) = J_\diamond(\mathbf{v}) := \sup_{G \ni \mathbf{v}} \inf_{\mathbf{u} \in G} J(\mathbf{u}).$$

First we consider a point $\mathbf{v}$ where $J(\mathbf{v}) = \infty$. Since I_T is the rate function of a sequence of probability laws, $\inf\{I_T(t) : t > 0\} = 0$; hence, I_T must be finite somewhere. So, by the definition of J in equation (4.7), we can see that

$$J(\mathbf{v}) = \infty \quad \Longleftrightarrow \quad I(\mathbf{v}) = \infty.$$

Therefore, by the lower semi-continuity of I_X , there exists $G \ni \mathbf{v}$ such that $I_X(\mathbf{u}) = \infty$ for all $\mathbf{u} \in G$, and this implies $J_\diamond(\mathbf{v}) = J(\mathbf{v})$.

We now consider $J(\mathbf{v}) < \infty$. If $I_X(\mathbf{v}) = 0$ then $J(\mathbf{v}) = 0$. But, since I_X and I_T are non-negative, J is non-negative, and hence, so too is $J_\diamond$. Therefore,

$$I_X(\mathbf{v}) = 0 \quad \Rightarrow \quad 0 \leq J_\diamond(\mathbf{v}) \leq J(\mathbf{v}) \leq 0 \quad \Rightarrow \quad J \text{ is lsc at } \mathbf{v}.$$

The only remaining points in Γ to consider are those where J is finite and $I_X > 0$. Let $\mathbf{v}$ be such a point. Since I_X is lsc, we know that for any ϵ , there exists a set $G(\epsilon) \ni \mathbf{v}$ for which

$$I_X(\mathbf{v}) - \epsilon \leq \inf_{\mathbf{u} \in G(\epsilon)} I_X(\mathbf{u}) \leq I_X(\mathbf{v}).$$

From this, we can conclude by equation (4.7) that

$$\inf_{t>0} \{I_T(t) + t(I_X(\mathbf{v}) - \epsilon)\} \leq \inf_{\mathbf{u} \in G(\epsilon)} J(\mathbf{u}) \leq J(\mathbf{v}). \quad (4.8)$$

Since $I_X(\mathbf{v}) > 0$, we can choose $\epsilon < I_X(\mathbf{v})$, and hence

$$I_T(t) + t(I_X(\mathbf{v}) - \epsilon) \geq I_T(t).$$

Therefore, for any b ,

$$\{t : I_T(t) + t(I_X(\mathbf{v}) - \epsilon) \leq b\} \subset \{t : I_T(t) \leq b\}.$$

I_T has compact level sets and so there must exist some finite $L(b)$ such that

$$\{t : I_T(t) + t(I_X(\mathbf{v}) - \epsilon) \leq b\} \subset [0, L(b)].$$

Because $J(\mathbf{v})$ is finite, we can choose $b > J(\mathbf{v})$, and because of equation (4.8),

$$\begin{aligned} \inf_{t>0} \{I_T(t) + t(I_X(\mathbf{v}) - \epsilon)\} &= \inf_{t \in [0, L(b)]} \{I_T(t) + t(I_X(\mathbf{v}) - \epsilon)\} \\ &\geq \inf_{t>0} \{I_T(t) + tI_X(\mathbf{v})\} - L(b)\epsilon. \end{aligned}$$

Therefore, using equation (4.8),

$$J(\mathbf{v}) - L(b)\epsilon \leq \inf_{\mathbf{u} \in G(\epsilon)} J(\mathbf{u}) \leq J(\mathbf{v}).$$

So, for $\epsilon > 0$,

$$\begin{aligned} J(\mathbf{v}) - \epsilon &\leq \inf \left\{ J(\mathbf{u}) : \mathbf{u} \in G\left(\frac{\epsilon}{L(b)}\right) \right\} \leq J(\mathbf{v}), \\ J(\mathbf{v}) - \epsilon &\leq \sup_{G \ni \mathbf{v}} \inf_{\mathbf{u} \in G} J(\mathbf{u}) \leq J(\mathbf{v}). \end{aligned}$$

Since ϵ is arbitrary we can conclude that $J_\diamond(\mathbf{v}) = J(\mathbf{v})$. $\square$

The last result we need to prove, before we address the main theorem, concerns the convergence of the functions $\{h_n(\mathbf{x}, \cdot)\}$ to the relative entropy function. As indicated in the sketch of the derivation of the random sampling formula at the start of this chapter, these functions occur in the application of Varadhan's theorem. Hence we need to identify the limit function and ensure that the convergence is locally uniform. As we shall see in the proof of Theorem 4.5, we need to consider these functions extended to set functions, as defined in Section 2.4.

Lemma 4.4 *For any $\mathbf{x} \in \Gamma$, there exists $r > 0$ such that $B(\mathbf{x}, r) \subset \mathcal{A}(\mathbf{x})$ and*

$$\lim_{n \rightarrow \infty} h_n(B(\mathbf{x}, r), \mathbf{v}) = - \inf_{\mathbf{y} \in B(\mathbf{x}, \epsilon)} H(\mathbf{y}|\mathbf{v}),$$

and this convergence is locally uniform over $\mathbf{v} \in \mathcal{A}(\mathbf{x})$.

Proof. First we choose an arbitrary $x \in \Gamma$. We then choose $r < \min\{x_i : i \in S(\mathbf{x})\}$ and this ensures that $\overline{B(\mathbf{x}, r)} \subset \mathcal{A}(\mathbf{x})$. We then take any open ball $B \ni x$ contained in $B(\mathbf{x}, r)$.

We establish the convergence in three stages.

First we use an argument known as the principle of the largest term. Remembering the definition of $h_n(B, \cdot)$,

$$h_n(B, \mathbf{v}) = \frac{1}{n} \log \left(\sum_{\mathbf{q} \in B \cap \Gamma_n} e^{nh_n(\mathbf{q}, \mathbf{v})} \right),$$

we note that

$$\sup_{\mathbf{q} \in B \cap \Gamma_n} h_n(\mathbf{q}, \mathbf{v}) \leq h_n(B, \mathbf{v}) \leq \sup_{\mathbf{q} \in B \cap \Gamma_n} h_n(\mathbf{q}, \mathbf{v}) + \frac{1}{n} \log |B \cap \Gamma_n|,$$

Since $|B \cap \Gamma_n|$ grows algebraically, we see that

$$\begin{aligned} \frac{1}{n} \log |B \cap \Gamma_n| &\rightarrow 0 \\ \Rightarrow \lim_{n \rightarrow \infty} h_n(B, \mathbf{v}) &= \lim_{n \rightarrow \infty} \sup_{\mathbf{q} \in B} h_n(\mathbf{q}, \mathbf{v}). \end{aligned}$$

Obviously, this convergence is independent of $\mathbf{v}$.

Next we use a crude version of Stirling's formula, based on a derivation in [1]:

$$\begin{aligned} 0 &\leq (n \log n - n + 1) - \left(\log n! - \frac{1}{2} \log n \right) \leq \log 2, \\ &\Rightarrow \left| \frac{1}{n} \log n! - \log n + 1 \right| \leq \frac{1 + \frac{1}{2} \log n}{n}. \end{aligned} \quad (4.9)$$

Now, from the definition of h_n , equation (2.3), we see that

$$h_n(\mathbf{q}, \mathbf{v}) + H(\mathbf{q} | \mathbf{v}) = \frac{1}{n} \log \frac{n!}{(nq_1)! \dots (nq_M)!} + \sum_{1 \leq i \leq M} q_i \log q_i.$$

Using the inequality, (4.9), and adhering to the conventions that $0 \log 0 = 0$ and $0! = 1$, we can see that for $\mathbf{q} \in \Gamma_n$,

$$\begin{aligned} \left| \frac{1}{n} \log(nq_1)! - q_1 \log(nq_1) + q_1 \right| &\geq \frac{1 + \frac{1}{2} \log n}{n}, \\ \Rightarrow \left| \frac{1}{n} \log \frac{n!}{(nq_1)! \dots (nq_M)!} + \sum_{1 \leq i \leq M} q_i \log q_i \right| &\geq (M+1) \frac{1 + \frac{1}{2} \log n}{n}, \\ \Rightarrow \lim_{n \rightarrow \infty} \left| h_n(\mathbf{q}, \mathbf{v}) + H(\mathbf{q} | \mathbf{v}) \right| &= 0. \end{aligned}$$

This convergence is independent of $\mathbf{q}$, so we can conclude that

$$\begin{aligned} \lim_{n \rightarrow \infty} \sup_{\mathbf{q} \in B \cap \Gamma_n} h_n(\mathbf{q}, \mathbf{v}) &= \lim_{n \rightarrow \infty} \sup_{\mathbf{q} \in B \cap \Gamma_n} (-H(\mathbf{q}|\mathbf{v})) , \\ &= - \lim_{n \rightarrow \infty} \inf_{\mathbf{q} \in B \cap \Gamma_n} H(\mathbf{q}|\mathbf{v}) . \end{aligned}$$

Finally we need to show that

$$\lim_{n \rightarrow \infty} \inf_{\mathbf{q} \in B \cap \Gamma_n} H(\mathbf{q}|\mathbf{v}) = \inf_{\mathbf{y} \in B} H(\mathbf{y}|\mathbf{v})$$

and that this convergence is locally uniform over $\mathbf{v} \in \mathcal{A}(\mathbf{x})$. In other words, given an arbitrary such $\mathbf{v}$ and $\epsilon > 0$, we wish to show the existence of a radius δ and an N such that, for $C := \{\mathbf{u} : |\mathbf{u} - \mathbf{v}| < \delta\}$,

$$\inf_{\mathbf{q} \in B \cap \Gamma_n} H(\mathbf{q}|\mathbf{u}) - \inf_{\mathbf{y} \in B} H(\mathbf{y}|\mathbf{u}) < \epsilon, \quad \forall \mathbf{u} \in C, n > N.$$

We choose δ such that $C \subset \mathcal{A}(\mathbf{x})$ and fix an arbitrary $\mathbf{u} \in C$. We note that the infimum, $\inf_{\mathbf{z} \in B} H(\mathbf{z}|\mathbf{v})$, will be attained at some point, in the closure of B , which depends on the choice of $\mathbf{u}$; we denote this point by $\mathbf{y} \in \mathcal{A}(\mathbf{x})$. We define I to be the complement in Σ , of $S(\mathbf{v})$, the support of $\mathbf{v}$; and so $i \in I$ iff $v_i = 0$. We can choose a sequence of points $\{\mathbf{q}(n)\}$, where $\mathbf{q}(n) \in B \cap \Gamma_n$, which converges to $\mathbf{y}$ and satisfies

$$q_i(n) = \frac{\lfloor n y_i \rfloor}{n}, \quad \forall i \in I \quad \text{and} \quad |q_i(n) - y_i| < \frac{M}{n}, \quad \text{for } 1 \leq i \leq M.$$

We wish to compare $H(\cdot|\mathbf{u})$ on this sequence of points to $H(\mathbf{y}|\mathbf{u})$:

$$H(\mathbf{q}(n)|\mathbf{u}) - H(\mathbf{y}|\mathbf{u}) = \sum_{1 \leq i \leq M} \left(q_i(n) \log \frac{q_i(n)}{u_i} - y_i \log \frac{y_i}{u_i} \right)$$

We examine this sum on a component-wise basis, treating those components in the support of $\mathbf{v}$ differently from the rest.

We are interested in the function

$$f(y, v) = y \log \frac{y}{v},$$

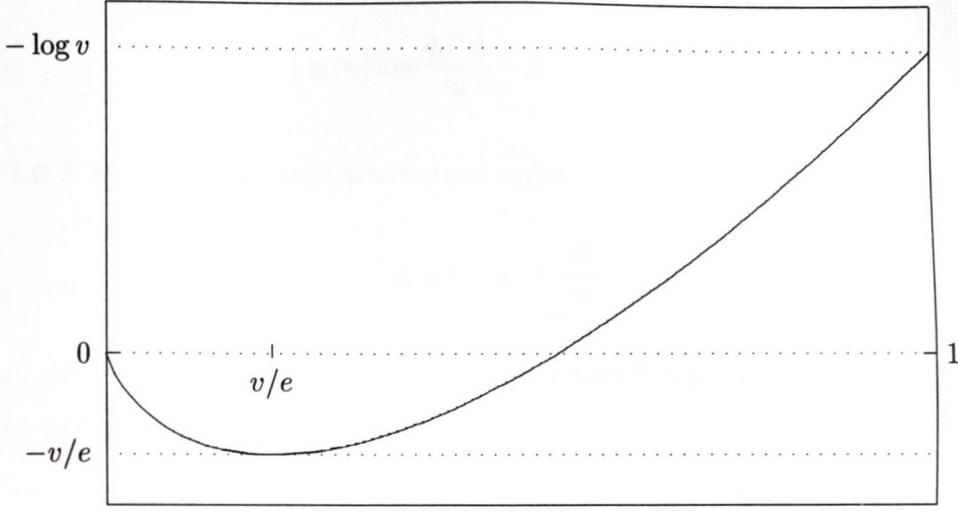


Figure 4-1: A plot of $y \log(y/v)$ for a typical v

which is convex and decreases from 0 at $y = 0$ to a minimum value of $-\frac{v}{e}$ at $\frac{v}{e}$, as shown in Figure 4.2. Hence,

$$0 \leq x < y \leq 1 \Rightarrow f(x, v) - f(y, v) < \frac{v}{e}.$$

First consider $i \in I$. By construction, we must have $q_i(n) \leq y_i$. If $u_i > 0$, then

$$f(q_i(n), u_i) - f(y_i, u_i) < \frac{u_i}{e}.$$

However, for $i \in I$ it is possible that $u_i = 0$. $\mathbf{u} \in \mathcal{A}(\mathbf{x})$ means that $H(\mathbf{x}|\mathbf{u}) < \infty$ and so, by definition of $\mathbf{y}$,

$$H(\mathbf{y}|\mathbf{u}) < \infty.$$

If $u_i = 0$, the definition of H then implies that $y_i = 0$ and then $q_i(n) \leq y_i$ implies that $q_i = 0$. In that case, taking $0 \log \frac{0}{0} = 0$, we get

$$u_i = 0 \Rightarrow f(q_i(n), u_i) - f(y_i, u_i) = 0.$$

Using the fact that $v_i = 0$ implies $u_i < \delta$, we get

$$\sum_{i \in I} \left(q_i(n) \log \frac{q_i(n)}{u_i} - y_i \log \frac{y_i}{u_i} \right) < \frac{|I|\delta}{e}. \quad (4.10)$$

For $i \notin I$, we have less information about $q_i(n)$ in relation to y_i ; we only know that

$$|q_i(n) - y_i| < \frac{M}{n}.$$

However, since $v_i > 0$ we can ensure $u_i > 0$ by choosing δ appropriately. The steepest parts of the $f(\cdot, u)$ are at 0 and 1, so the larger of

$$f(0, u) - f\left(\frac{M}{n}, u\right) = \frac{M}{n} \log \frac{nu}{M},$$

and

$$f(1, u) - f\left(1 - \frac{M}{n}, u\right) = -\frac{M \log u}{n} - \left(1 - \frac{M}{n}\right) \log \left(1 - \frac{M}{n}\right),$$

gives us a bound on $|f(x, u) - f(y, u)|$ when $|x - y| < \frac{M}{n}$. Now, noting that we can choose δ sufficiently small so that $\underline{u} := \min_{i \notin I} u_i > 0$, we have the upper bound

$$\begin{aligned} \sum_{i \notin I} \left(q_i(n) \log \frac{q_i(n)}{u_i} - y_i \log \frac{y_i}{u_i} \right) < \\ (M - |I|) \max \left\{ f\left(\frac{M}{n}, 1\right), \frac{-M \log \underline{u}}{n} - f\left(1 - \frac{M}{n}, 1\right) \right\}, \end{aligned}$$

which is uniform over $\mathbf{u} \in C$. Since $f(\cdot, 1)$ is continuous and zero at both 0 and 1, it is clear that we can choose N such that

$$\sum_{i \notin I} \left(q_i(n) \log \frac{q_i(n)}{u_i} - y_i \log \frac{y_i}{u_i} \right) < \frac{\epsilon}{2} \quad \forall n > N.$$

If we also ensure that $\delta < \frac{e\epsilon}{2|I|}$, we know by the earlier bound, 4.10, that

$$H(\mathbf{q}(n)|\mathbf{u}) - H(\mathbf{y}|\mathbf{u}) < \epsilon, \quad \forall \mathbf{u} \in C, n > N.$$

Hence

$$\begin{aligned} \lim_{n \rightarrow \infty} \inf_{\mathbf{q} \in B \cap \Gamma_n} H(\mathbf{q}|\mathbf{v}) &= \inf_{\mathbf{y} \in B} H(\mathbf{y}|\mathbf{v}), \\ \Rightarrow \lim_{n \rightarrow \infty} h_n(B(\mathbf{x}, r), \mathbf{v}) &= - \inf_{\mathbf{y} \in B} H(\mathbf{y}|\mathbf{v}), \end{aligned}$$

and this convergence is locally uniform in $\mathbf{v} \in \mathcal{A}(\mathbf{x})$.

□

Finally, we are ready to prove the main result. We define the empirical measure induced by the random sampling to be

$$\mu_n := \sum_{1 \leq i \leq M} U_i^{(n)} \delta_i,$$

where

$$U_i^{(n)}(\omega) := \frac{1}{n} \sum_{j=1}^n 1_{\{X_{T_j}=i\}}(\omega)$$

is the fraction of the samples observed in state i . We wish to prove an LDP for the vector $\mathbf{U}^{(n)}$ and find its rate function. To do so, we must impose an additional constraint on $\{T_n\}$.

Theorem 4.5 *If $\{N_t\}$ is a Poisson process with rate α , then the sequence $\{\mathcal{L}_{(\mathbf{U}^{(n)})}\}$ obeys an LDP on the scale n , with rate function $J_\diamond$, the lsc regularisation of*

$$J(\mathbf{u}) := \inf_{\mathbf{v} \in \Gamma} \left\{ H(\mathbf{u}|\mathbf{v}) + \log \left(\frac{I_X(\mathbf{v})}{\alpha} + 1 \right) \right\}.$$

Proof. We prove this result by using Varadhan's theorem to demonstrate the equality of the upper and lower deviation functions. Due to the compactness of Γ , it is easy to extend this to a full LDP.

Following the notation defined in Section 2.4, we choose an arbitrary $\mathbf{x} \in \Gamma$, and by virtue of Lemma 4.4 we know there exists an open ball B , centred on x , such that $B \subset \mathcal{A}(\mathbf{x})$. We wish to show the existence of, and calculate, the limit

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P(\mathbf{U}^{(n)} \in B).$$

Since the vector $\mathbf{U}^{(n)}$ can only take values in Γ_n (using the notation of Lemma 4.4) we will need to obtain estimates on quantities of the form, $P(\mathbf{U}^{(n)} = \mathbf{q})$, for $\mathbf{q} \in B \cap \Gamma_n$. A useful step is to consider the distribution of $\mathbf{U}^{(n)}$ in conjunction with $(T_1, \dots, T_{n+1})$. In anticipation of fixing the sampling times, we confuse notation slightly and define a random vector $\mathbf{U}^{(n)}(t_1, \dots, t_n)$ to be the components of the empirical measure obtained from the process $(X_{t_1}, \dots, X_{t_n})$, where the $(t_1, \dots, t_n)$ are understood to no longer be random. We can then use Theorem 4.1 to write

$$P(\mathbf{U}^{(n)} = \mathbf{q}) = \int_{t_{n+1} \in \mathbb{R}} \dots \int_{t_1 \in \mathbb{R}} P(\mathbf{U}^{(n)}(t_1, \dots, t_n) = \mathbf{q}) \mathcal{L}_{(T_1, \dots, T_{n+1})}(dt_1 \times \dots \times dt_{n+1}).$$

Using equation (2.5), the particular joint density function for $(T_1, \dots, T_n)$, which arises from their exponentially distributed inter-sample times, we find

$$P(\mathbf{U}^{(n)} = \mathbf{q}) = \int_{t_{n+1} \in \mathbb{R}} \int_{t_n=0}^{t_{n+1}} \dots \int_{t_1=0}^{t_2} P(\mathbf{U}^{(n)}(t_1, \dots, t_n) = \mathbf{q}) \alpha^{n+1} e^{-\alpha t_{n+1}} dt_1 \dots dt_{n+1}.$$

Concentrating initially on the integration over $(t_1, \dots, t_n)$, we note that the integrand is symmetric in its remaining arguments. For such an integral, we have the well known result that

$$\begin{aligned} & \int_{t_n=0}^{t_{n+1}} \dots \int_{t_1=0}^{t_2} P(\mathbf{U}^{(n)}(t_1, \dots, t_n) = \mathbf{q}) dt_1 \dots dt_n \\ &= \frac{1}{n!} \int_{t_n=0}^{t_{n+1}} \int_{t_{n-1}=0}^{t_{n+1}} \dots \int_{t_1=0}^{t_{n+1}} P(\mathbf{U}^{(n)}(t_1, \dots, t_n) = \mathbf{q}) dt_1 \dots dt_n, \\ &= \frac{1}{n!} \int_{t_n=0}^{t_{n+1}} \int_{t_{n-1}=0}^{t_{n+1}} \dots \int_{t_1=0}^{t_{n+1}} \int_{\Omega} 1_{\{\mathbf{U}^{(n)}(t_1, \dots, t_n) = \mathbf{q}\}}(\omega) P[d\omega] dt_1 \dots dt_n. \end{aligned}$$

Using the function F , defined in Theorem 4.1, it is easily shown that the integrand is measurable as a joint function of ω and $(t_1, \dots, t_n)$. We can apply Fubini's theorem to change the order of integration and obtain:

$$\int_{t_n=0}^{t_{n+1}} \cdots \int_{t_1=0}^{t_2} P\left(\mathbf{U}^{(n)}(t_1, \dots, t_n) = \mathbf{q}\right) dt_1 \dots dt_n \quad (4.11)$$

$$= \frac{1}{n!} \int_{\Omega} \left(\int_{t_n=0}^{t_{n+1}} \cdots \int_{t_1=0}^{t_{n+1}} 1_{\{\mathbf{U}^{(n)}(t_1, \dots, t_n) = \mathbf{q}\}}(\omega) dt_1 \dots dt_n \right) P[d\omega]. \quad (4.12)$$

For each ω , the integral over $t_1, \dots, t_n$ can now be easily done by remembering that the vector $\mathbf{V}^{(t_{n+1})}(\omega)$ represents the fractions of time spent in each state up to time t_{n+1} . First of all, we note that if $\mathbf{V}^{(t_{n+1})}(\omega) \notin \mathcal{A}(\mathbf{q}) = \mathcal{A}(\mathbf{x})$, the integral is zero. Otherwise, the indicator function selects those combinations of values of the observation times that result in the number of observations of state i being nq_i . Having removed the interdependence of the ranges, the integral is symmetric with respect to the order in which the states are observed, and so we can choose one such ordering and multiply by the number of distinct permutations. To this end, we define the event

$$\begin{aligned} C = \{ & \omega : X_{t_i}(\omega) = 1, 1 \leq i \leq nq_1, \\ & X_{t_i}(\omega) = 2, nq_1 < i \leq nq_1 + nq_2, \\ & \dots, \\ & X_{t_i}(\omega) = M, n \sum_{j=1}^{M-1} q_j < i \leq n \}, \end{aligned}$$

which specifies the ordering where the observations of the first state precede those of the second state and so forth. With the aid of the combinatorial factor that counts the number of ways in which these observations could be re-ordered, we can rewrite the integral as

$$\begin{aligned} & \int_{t_n=0}^{t_{n+1}} \cdots \int_{t_1=0}^{t_{n+1}} 1_{\{\mathbf{U}^{(n)}(t_1, \dots, t_n) = \mathbf{q}\}}(\omega) dt_1 \dots dt_n \\ &= \frac{n!}{(nq_1)! \dots (nq_M)!} \int_{t_n=0}^{t_{n+1}} \cdots \int_{t_1=0}^{t_{n+1}} 1_C(\omega) dt_1 \dots dt_n, \end{aligned} \quad (4.13)$$

Due to the simple structure of the set C , 1_C can be written as an n -fold product of the form

$$1_C = \prod_{i=1}^n 1_{\{X_{t_i} = k_i\}},$$

where $k_i := \min\{m : nq_1 + \dots + nq_m \geq i\}$. Since each function can be integrated separately and for each $i \leq n$,

$$\int_0^{t_{n+1}} 1_{\{X_{t_i}=k_i\}}(\omega) dt_i = t_{n+1} V_{k_i}^{t_{n+1}}(\omega),$$

we have that equation (4.13) reduces to

$$\begin{aligned} & \frac{n!}{(nq_1)! \dots (nq_M)!} \prod_{\{i \leq M : q_i > 0\}} \left(t_{n+1} V_i^{(t_{n+1})}(\omega) \right)^{nq_i} \\ &= 1_{\{\mathbf{V}^{(t_{n+1})} \in \mathcal{A}(\mathbf{x})\}}(\omega) t_{n+1}^n \frac{n!}{(nq_1)! \dots (nq_M)!} \exp \left(-nH \left(\mathbf{q} \middle| \mathbf{V}^{(t_{n+1})} \right) - nH(\mathbf{q}) \right) \\ &= 1_{\{\mathbf{V}^{(t_{n+1})} \in \mathcal{A}(\mathbf{x})\}}(\omega) t_{n+1}^n \exp \left(nh_n \left(q, \mathbf{V}^{(t_{n+1})}(\omega) \right) \right) \end{aligned}$$

Therefore, returning to equation (4.12),

$$\begin{aligned} & \int_{t_n=0}^{t_{n+1}} \dots \int_{t_1=0}^{t_2} P \left(\mathbf{U}^{(n)}(t_1, \dots, t_n) = \mathbf{q} \right) dt_1 \dots dt_{n+1} \\ &= \frac{t_{n+1}^n}{n!} \int_{\Omega} 1_{\{\mathbf{V}^{(t_{n+1})} \in \mathcal{A}(\mathbf{x})\}}(\omega) \exp \left(nh_n \left(q, \mathbf{V}^{(t_{n+1})}(\omega) \right) \right) P[d\omega] \end{aligned}$$

which, by a change of variables,

$$= \frac{t_{n+1}^n}{n!} \int_{\mathbf{v} \in \mathcal{A}(\mathbf{x})} e^{nh_n(q, \mathbf{v})} \mathcal{L}_{(\mathbf{V}^{(t_{n+1})})}[d\mathbf{v}].$$

Using the notation of Lemma 4.4, we return to equation (4.11). Relabelling t_{n+1} at t , we get

$$P \left(\mathbf{U}^{(n)} = \mathbf{q} \right) = \int_{t \in \mathbb{R}} \int_{\mathbf{v} \in \mathcal{A}(\mathbf{x})} e^{nh_n(q, \mathbf{v})} \mathcal{L}_{(\mathbf{V}^{(t)})}[d\mathbf{v}] \frac{\alpha^{n+1} t^n}{n!} e^{-\alpha t} dt.$$

Now, noting that the probability density function, see equation (2.6), for T_{n+1} is given by

$$f_{(T_{n+1})}(t) = \frac{\alpha^{n+1} t^n}{n!} e^{-\alpha t},$$

we can write

$$\begin{aligned} P \left(\mathbf{U}^{(n)} = \mathbf{q} \right) &= \int_{t \in \mathbb{R}} \int_{\mathbf{v} \in \mathcal{A}(\mathbf{x})} e^{nh_n(q, \mathbf{v})} \mathcal{L}_{(\mathbf{V}^{(t)})}[d\mathbf{v}] \mathcal{L}_{(T_{n+1})}[dt] \\ &= \int_{\mathbf{v} \in \mathcal{A}(\mathbf{x})} e^{nh_n(q, \mathbf{v})} \mathcal{L}_{(\mathbf{V}^{(T_{n+1})})}[d\mathbf{v}], \end{aligned}$$

where the last step relies again on Theorem 4.1. Returning to the open ball B ,

$$\begin{aligned} P(\mathbf{U}^{(n)} \in B) &= \sum_{\mathbf{q} \in B \cap \Gamma_n} P(\mathbf{U}^{(n)} = \mathbf{q}) \\ &= \int_{\mathbf{v} \in \mathcal{A}(\mathbf{x})} e^{nh_n(B, \mathbf{v})} \mathcal{L}_{(\mathbf{v}^{(T_{n+1})})}[d\mathbf{v}]. \end{aligned}$$

We know by Lemma 4.4 that $h_n(B, \mathbf{v})$ converges to $\sup_{\mathbf{u} \in B} \{-H(\mathbf{u}|\mathbf{v})\}$ and that the convergence is locally uniform over $\mathbf{v} \in \mathcal{A}(\mathbf{x})$. Since we have assumed that $\{N_t\}$ is Poisson, we know that $\{\mathcal{L}_{(T_n/n)}\}$ obeys an LDP and so we can use Theorem 4.3 and Varadhan's theorem to conclude that

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \log P(\mathbf{U}^{(n)} \in B) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \int_{\mathbf{v} \in \mathcal{A}(\mathbf{x})} e^{nh_n(B, \mathbf{v})} \mathcal{L}_{(\mathbf{v}^{(T_{n+1})})}[d\mathbf{v}] \\ &= \sup_{\mathbf{v} \in \mathcal{A}(\mathbf{x})} \left\{ - \inf_{\mathbf{u} \in B} H(\mathbf{u}|\mathbf{v}) - \inf_{t \in \mathbb{R}_+} \{tI_X(\mathbf{v}) + I_T(t)\} \right\}. \end{aligned}$$

Now, since we have chosen the observation process to be Poisson, equation (2.8) gives an explicit form for I_T . We can complete the infimisation over t by differentiation to obtain

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P(\mathbf{U}^{(n)} \in B) = - \inf_{\mathbf{u} \in B} \inf_{\mathbf{v} \in \mathcal{A}(\mathbf{u})} \left\{ H(\mathbf{u}|\mathbf{v}) + \log \left(\frac{I_X(\mathbf{v})}{\alpha} + 1 \right) \right\}.$$

Since $H(\mathbf{u}|\mathbf{v})$ is infinite for $\mathbf{v} \notin \mathcal{A}(\mathbf{u})$ we can define

$$J(\mathbf{u}) := \inf_{\mathbf{v} \in \Gamma} \left\{ H(\mathbf{u}|\mathbf{v}) + \log \left(\frac{I_X(\mathbf{v})}{\alpha} + 1 \right) \right\},$$

and write

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P(\mathbf{U}^{(n)} \in B) = - \inf_{\mathbf{u} \in B} J(\mathbf{u}).$$

Hence the large deviation limit exists for sufficiently small open balls around any point in Γ . Since the open balls form a base for the neighbourhood system in the usual topology on $\mathbb{R}^M$, and hence in the induced topology on Γ , we can see by the standard argument that the upper and lower deviation functions are equal. Using the notation of Chapter 2, we can write their joint value as

$$\mu(\mathbf{x}) = \inf_{G \ni \mathbf{x}} \overline{m}[G] = \inf_{G \ni \mathbf{x}} \underline{m}[G]$$

$$\begin{aligned}
&= \inf_{G \ni \mathbf{x}} \left\{ - \inf_{\mathbf{u} \in G} J(\mathbf{u}) \right\} \\
&= -J_{\diamond}(\mathbf{x}).
\end{aligned}$$

$J_{\diamond}$ is lower semi-continuous by construction. Since Γ is compact, we have, therefore, that the sequence $\{\mathcal{L}_{(\mathbf{U}^{(n)})}\}$ satisfies an LDP on the scale n with rate function $J_{\diamond}$. □

We remark that the function J must be upper semi-continuous. Indeed, for each $\mathbf{v}$ the function

$$j_{\mathbf{v}}(\mathbf{u}) := H(\mathbf{u}|\mathbf{v}) + \log \left(\frac{I_X(\mathbf{v})}{\alpha} + 1 \right)$$

is continuous for all $\mathbf{v}$. By the definition of J , we have that for any b ,

$$\{\mathbf{u} : J(\mathbf{u}) < b\} = \bigcup_{\mathbf{v}} \{\mathbf{u} : j_{\mathbf{v}}(\mathbf{u}) < b\}.$$

Since all the sets in this union are open, by continuity of $j_{\mathbf{v}}$, we have that $\{J < b\}$ is open. Hence, J is upper semi-continuous.

Chapter 5

Poisson sampling : II

In this chapter, we will consider what information we can regain about the underlying process based on Theorem 4.5, knowing only the rate function of the sample process. It turns out that in the case where the underlying process has a C^2 , convex rate function, we can deduce its shape around the mean, based on the knowing the rate function for the sample process.

5.1 Galois connections

Consider Q as the set of functions from Γ to the positive real line, and let P be a copy of Q , so that

$$P = Q = \mathbb{R}_+^\Gamma.$$

The random sampling transformation, derived in Theorem 4.5, can be viewed as a functional from Q to P given by

$$(fI)(y) = \inf_{\mathbf{x} \in \Gamma} \left\{ H(\mathbf{y} | \mathbf{x}) + \log \left(\frac{I(\mathbf{x})}{\alpha} + 1 \right) \right\}, \quad \forall \mathbf{y} \in \Gamma.$$

We are interested in examining the structure of this transformation, with a view to finding a characterisation of the pre-image. The methodology of Galois connections, a concept used in order theory, proves useful to us here. A Galois connection is a pair of functions between two partially ordered sets (which are often termed posets), that satisfies certain of the properties of inverse functions. We will show that the functional f can be viewed as

half of a Galois connection. The set Q , and hence P , can be considered posets under the product partial order, by which we mean that for all $I_1, I_2 \in Q$,

$$I_1 \leq I_2 \iff I_1(x) \leq I_2(x) \forall x \in \Gamma.$$

In [9, Chapter 4], Russell provides an introduction to order theory and to the theory of Galois connections in particular. For our purposes here, we just want to indicate the basic structure. We start with the definition of a Galois connection. A pair of functions (f, g) between two posets P and Q ,

$$g : P \rightrightarrows Q : f,$$

is termed a Galois connection if

$$g(p) \leq q \iff p \leq f(q).$$

The functions are also referred to as adjoints, with f being the upper adjoint and g being the lower adjoint. Knowing that two functions form a Galois connection defines each in terms of the other, by the following formulae

$$g(p) = \inf\{q \in Q : p \leq f(q)\}, \quad (5.1)$$

$$f(q) = \sup\{p \in P : g(p) \leq q\}. \quad (5.2)$$

Although the functions in a Galois connection are not inverse to each other, they can be used to identify a related bijection. When the domain of each function is restricted to the image of the other,

$$g|_{f(Q)} : f(Q) \rightarrow g(P),$$

$$f|_{g(P)} : g(P) \rightarrow f(Q),$$

the resulting pair forms an isomorphism:

$$g|_{f(Q)}^{-1} = f|_{g(P)}.$$

Returning to the case in question, our posets P and Q are copies of the set of positive functions from Γ to $\overline{\mathbb{R}}_+$. The function f takes $I \in Q$ to $J \in P$ by the transformation

$$J(\mathbf{y}) = (fI)(\mathbf{y}) = \inf_{\mathbf{x} \in \Gamma} \left\{ H(\mathbf{x}|\mathbf{y}) + L(I(\mathbf{x})) \right\},$$

where

$$L(z) = \log \left(\frac{z}{\alpha} + 1 \right).$$

We are interested in determining whether there is a function, g , from P to Q such that (f, g) form a Galois connection. We assume that there is such a function, write it in terms of f using equation (5.1) and then prove that the two functions are adjoint. We first note that the inverse to L is

$$L^{-1}(z) = \alpha(e^z - 1).$$

Theorem 5.1 *If we let*

$$gJ := \inf\{I \in Q : J \leq fI\},$$

then

$$(gJ)(x) = L^{-1} \left(\sup_{\mathbf{y} \in \Gamma} \{J(\mathbf{y}) - H(\mathbf{y}|\mathbf{x})\} \right),$$

and (f, g) form a Galois connection.

Proof. We first establish the form of g :

$$\begin{aligned} gJ &= \inf\{I : J \leq fI\} \\ &= \inf \left\{ I : J(\mathbf{y}) \leq \inf_{\mathbf{x} \in \Gamma} \left\{ H(\mathbf{y}|\mathbf{x}) + L(I(\mathbf{x})) \right\}, \quad \forall \mathbf{y} \in \Gamma \right\} \\ &= \inf \left\{ I : J(\mathbf{y}) \leq H(\mathbf{y}|\mathbf{x}) + L(I(\mathbf{x})), \quad \forall \mathbf{x}, \mathbf{y} \in \Gamma \right\} \\ &= \inf \left\{ I : I(x) \geq L^{-1} \left(J(\mathbf{y}) - H(\mathbf{y}|\mathbf{x}) \right), \quad \forall \mathbf{x}, \mathbf{y} \in \Gamma \right\} \\ &= \inf \left\{ I : I(x) \geq \sup_{\mathbf{y} \in \Gamma} L^{-1} \left(J(\mathbf{y}) - H(\mathbf{y}|\mathbf{x}) \right), \quad \forall \mathbf{x} \in \Gamma \right\} \\ &= \sup_{\mathbf{y} \in \Gamma} L^{-1} \left(J(\mathbf{y}) - H(\mathbf{y}|\mathbf{x}) \right). \end{aligned}$$

We can now show that (f, g) form a Galois connection via the following chain of equiv-

alent inequalities:

$$\begin{aligned}
 gJ &\leq I, \\
 (gJ)(\mathbf{x}) &\leq I(\mathbf{x}), \quad \forall \mathbf{x} \in \Gamma, \\
 \sup_{\mathbf{y} \in \Gamma} L^{-1} \left(J(\mathbf{y}) - H(\mathbf{y}|\mathbf{x}) \right) &\leq I(\mathbf{x}), \quad \forall \mathbf{x} \in \Gamma, \\
 \sup_{\mathbf{y} \in \Gamma} \left\{ J(\mathbf{y}) - H(\mathbf{y}|\mathbf{x}) \right\} &\leq L(I(\mathbf{x})), \quad \forall \mathbf{x} \in \Gamma, \\
 J(\mathbf{y}) - H(\mathbf{y}|\mathbf{x}) &\leq L(I(\mathbf{x})), \quad \forall \mathbf{y}, \mathbf{x} \in \Gamma, \\
 J(\mathbf{y}) &\leq \inf_{\mathbf{x} \in \Gamma} \left\{ H(\mathbf{y}|\mathbf{x}) + L(I(\mathbf{x})) \right\}, \quad \forall \mathbf{y} \in \Gamma, \\
 J &\leq fI.
 \end{aligned}$$

Thus we have established that

$$gJ \leq I \iff J \leq fI.$$

□

5.2 Recovering the underlying rate function

We continue this analysis in the 2-state case, where the calculus is more tractable. As explained in Section 2.4, when the state space is confined to $\{0, 1\}$, the empirical average contains the same information as the empirical measure. In this case, the vectors $\mathbf{x}, \mathbf{y} \in \Gamma$ reduce to two numbers $x, y \in [0, 1]$, which makes it possible to perform the optimisations in the definitions of g and f . We rewrite the relevant function as follows:

$$\begin{aligned}
 B(y, x) := H(\mathbf{y}|\mathbf{x}) &= y \log \frac{y}{x} + (1-y) \log \frac{1-y}{1-x}; \\
 (fI)(y) &= \inf_{0 \leq x \leq 1} \left\{ B(y, x) + L(I(x)) \right\}; \tag{5.3}
 \end{aligned}$$

$$(gJ)(x) = L^{-1} \left(\sup_{0 \leq y \leq 1} \{ J(y) - B(y, x) \} \right). \tag{5.4}$$

Before examining the effect of composing g with f , we need to gather some information

about f . This involves a lot of differentiation, so we first collect some useful functions :

$$B(\xi_1, \xi_2) = \xi_1 \log \left(\frac{\xi_1}{\xi_2} \right) + (1 - \xi_1) \log \left(\frac{1 - \xi_1}{1 - \xi_2} \right) ; \quad (5.5)$$

$$\frac{\partial B}{\partial \xi_1} = \log \frac{\xi_1(1 - \xi_2)}{\xi_2(1 - \xi_1)} ; \quad (5.6)$$

$$\frac{\partial B}{\partial \xi_2} = \frac{\xi_2 - \xi_1}{\xi_2(1 - \xi_2)} . \quad (5.7)$$

We will investigate over the next few theorems how the infimum in equation (5.3) is attained and what information can be recovered from fI about I . In order to do this analysis, we need to assume I is convex and that it satisfies some differentiability conditions.

Assumption 5.1 *I is a strictly convex and twice-differentiable rate function, which is finite on the interior of the unit interval.*

I is a rate-function and so its infimum over $[0, 1]$ must be 0. Since this set is compact, this infimum must be attained and we denote the point at which it is attained by m . We know that this point is unique because I is strictly convex. Of course, I being a rate function also implies that it is positive. We note that since I and I' are both differentiable, they must both be continuous. Finally, we remark that we have $\lim_{x \rightarrow 0} I'(x) < \infty$ because otherwise I is infinite everywhere in $(0, 1)$ since I' is increasing in x ; similarly, we must have that $\lim_{x \rightarrow 1} I'(x) > -\infty$.

We denote by h_y the function that we are interested in infimising.

$$h_y(x) := B(y, x) + \log \left(\frac{I(x)}{\alpha} + 1 \right) . \quad (5.8)$$

Since $[0, 1]$ is compact, this infimum must be attained; the next theorem explains how this happens.

Theorem 5.2 *For all $y \in [0, 1]$, the infimum of $h_y(x)$, over $x \in [0, 1]$, is attained at a unique point $\chi(y)$, defined as follows:*

- For $y \in (0, 1)$, $\chi(y)$ is the unique solution of $h'_y(x) = 0$.
- $\lim_{x \rightarrow 0} \frac{-I'(x)}{I(x) + \alpha} \leq 1 \iff \chi(0) = 0$.
Otherwise, $\chi(0)$ is the unique solution of the equation $h'_0(x) = 0$.

- $\lim_{x \rightarrow 1} \frac{I'(x)}{I(x) + \alpha} \leq 1 \iff \chi(1) = 1.$

Otherwise, $\chi(1)$ is the unique solution of the equation $h'_1(x) = 0$.

Proof. We first wish to show that if h'_y has a root, then that root must be unique. We note that the condition for x to be a root is that

$$h'_y(x) = \frac{x - y}{x(1 - x)} + \frac{I'(x)}{I(x) + \alpha} = 0. \quad (5.9)$$

Now the second derivative is given by

$$h''_y(x) = \frac{x^2 + y(1 - 2x)}{x^2(1 - x)^2} + \frac{I''(x)}{I(x) + \alpha} - \frac{I'(x)^2}{(I(x) + \alpha)^2}. \quad (5.10)$$

At any point, x_0 , that satisfies equation (5.9), we must have that

$$h''_y(x_0) = \frac{y(1 - y)}{x_0^2(1 - x_0)^2} + \frac{I''(x_0)}{I(x_0) + \alpha}.$$

Since $y \in [0, 1]$ and I is strictly convex, we can see that h''_y must be positive at x_0 . Therefore, h'_y is strictly increasing at any point where it is zero. The continuity of I' ensures continuity of h'_y and so we can conclude that h'_y can only reach zero once if it does so at all.

The question of where the infimum is attained is answered by examining h'_y at 0 and 1. If h_y is decreasing at 0 and increasing at 1, then it must attain its infimum on the interior and h'_y must be zero at that point. We know that h'_y can reach 0 at most once, so $\chi(y)$ (which we will sometimes denote by χ for convenience) is defined uniquely by the equation

$$\frac{\chi - y}{\chi(1 - \chi)} + \frac{I'(\chi)}{I(\chi) + \alpha} = 0. \quad (5.11)$$

It remains to check the behaviour of h'_y at the endpoints of the interval.

For $y \in (0, 1)$, we note that

$$\lim_{x \rightarrow 0} \frac{x - y}{x(1 - x)} = -\infty,$$

and, as remarked after assumption 5.1,

$$\lim_{x \rightarrow 0} I'(x) < \infty.$$

Since I is positive, we conclude that $\lim_{x \rightarrow 0} h'_y(x)$ is negative. Similarly, we can show that $\lim_{x \rightarrow 1} h'_y(x)$ is positive. Therefore, for $y \in (0, 1)$, we see that $\chi(y)$ is defined by equation (5.11).

For y equal to 0 or 1, the situation is a little more complicated as the value of the derivative at the endpoints depends on the steepness of I . Taking $y = 0$ in equation (5.9), we note that

$$\lim_{x \rightarrow 0} h'_0(x) = 1 + \lim_{x \rightarrow 0} \frac{I'(x)}{I(x) + \alpha} \quad \text{and} \quad \lim_{x \rightarrow 1} h'_0(x) = \infty.$$

If $\lim_{x \rightarrow 0} \frac{-I'(x)}{I(x) + \alpha} > 1$, then h_0 is decreasing at 0, and so, as before, $\chi(y)$ is defined by equation (5.11), with $y = 0$. Otherwise, h_0 is increasing at 0. Since, for any y , $h'_y(x_0) = 0$ implies that h'_y is increasing at x_0 , we can see that if h_0 is increasing at 0, then $\chi(0) = 0$.

A similar argument holds true for $y = 1$. □

We can use the function $\chi(y)$ defined in the preceding theorem to help analyse the properties of $J := fI$. First we prove a simple lemma concerning χ .

Lemma 5.3 *The function χ is differentiable and is strictly increasing.*

Proof. By applying the implicit function theorem to h'_y (noting that for all x such that $h'_y(x) = 0$, $h''_y(x) > 0$), we can conclude that χ has a continuous derivative on $(0, 1)$.

To examine $\chi'(y)$, we differentiate equation (5.11) with respect to y to obtain

$$\chi' \left[\frac{y(1-y)}{\chi^2(1-\chi)^2} + \frac{I''(\chi)}{I(\chi) + \alpha} \right] = \frac{1}{\chi(1-\chi)}, \quad (5.12)$$

where we have used equation (5.11) again to simplify the expression after differentiation.

$$\chi' [h''_y(\chi)] = \frac{1}{\chi(1-\chi)}.$$

We have demonstrated in the proof of Theorem 5.2 that $h''_y(\chi(y)) > 0$, and $\chi \in (0, 1)$ implies that the right hand side is positive. Hence, we can conclude that $\chi'(y)$ must be positive. □

Theorem 5.4 *The function*

$$J(y) := \inf_{0 \leq x \leq 1} \left\{ B(y, x) + \log \left(\frac{I(x)}{\alpha} + 1 \right) \right\}$$

is convex.

Proof. We know by Theorem 5.2 and Lemma 5.3 that

$$J(y) = h_y(\chi(y)),$$

where χ is differentiable on $(0, 1)$. Since, J is a composition of differentiable functions we can calculate the second derivative and check its positivity.

$$J(y) = B(y, \chi(y)) + \log \left(\frac{I(\chi(y))}{\alpha} + 1 \right).$$

So, using the functions defined in equations (5.6, 5.7),

$$\begin{aligned} J'(y) &= \frac{\partial B}{\partial \xi_1}(y, \chi) + \frac{\partial B}{\partial \xi_2}(y, \chi) \chi' + \frac{I'(\chi)}{I(\chi) + \alpha} \chi', \\ &= \frac{\partial B}{\partial \xi_1}(y, \chi). \end{aligned}$$

since the last two terms cancel by equation (5.11). Hence,

$$J''(y) = \frac{1}{y(1-y)} - \frac{\chi'}{\chi(1-\chi)},$$

which implies that

$$J''(y) \geq 0 \iff \chi' \frac{y(1-y)}{\chi(1-\chi)} \leq 1.$$

From equation (5.12), we know that

$$\chi' \frac{y(1-y)}{\chi(1-\chi)} = 1 - \chi' \frac{I''(\chi)\chi(1-\chi)}{I(\chi) + \alpha}.$$

Hence,

$$J''(y) \geq 0 \iff \chi' \frac{I''(\chi)\chi(1-\chi)}{I(\chi) + \alpha} \geq 0.$$

From Lemma 5.3, we know that χ' positive, and so convexity of I guarantees that J is convex. $\square$

When (f, g) form a Galois connection it is easily shown, see [9], that

$$g \circ f \leq id_Q.$$

From the point of view of recovering information about the original rate function, it is

interesting to examine which function in Q this combined transformation picks out. Suppose $J = fI$, so that

$$J(y) = \inf_{0 \leq x \leq 1} \left\{ B(y, x) + \log \left(\frac{I(x)}{\alpha} + 1 \right) \right\}.$$

We wish to examine

$${}^\circ I(z) := (g \circ f)I = L^{-1} \left(\sup_{0 \leq y \leq 1} \{J(y) - B(y, z)\} \right)$$

Because of the structure of Galois connections, we know that $g|_{f(Q)}$ and $f|_{g(P)}$ are inverse functions and hence

$$f({}^\circ I) = f(I).$$

Before calculating the form of ${}^\circ I$, we need to prove a result about the function which is to be optimised in the definition of g .

Lemma 5.5 *If $J = fI$ then*

$$a_z(y) := J(y) - B(y, z)$$

is strictly concave on $[0, 1]$.

Proof. From Theorem 5.2, we know

$$a_z(y) = B(y, \chi) + L(I(\chi)) - B(y, z).$$

Differentiating, we get

$$a'_z(y) = \frac{\partial B}{\partial \xi_1}(y, \chi) + \frac{\partial B}{\partial \xi_2}(y, \chi)\chi' + \frac{I'(\chi)}{I(\chi) + \alpha}\chi' - \frac{\partial B}{\partial \xi_1}(y, z).$$

Using equation (5.11) once more, to cancel the middle two terms, this reduces to

$$a'_z(y) = \log \frac{(1 - \chi)z}{\chi(1 - z)},$$

which implies that

$$a''_z(y) = -\frac{\chi'}{\chi(1 - \chi)}$$

But Lemma 5.3 states that the derivative of x is positive, thus guaranteeing the concavity of a_z . □

Finally, we can calculate the form of ${}^\circ I$, by using the concave structure of a_z to determine

where the supremum occurs.

Theorem 5.6 *The function ${}^{\circ}I$ has the following structure:*

- For $\chi(0) < z < \chi(1)$, ${}^{\circ}I(z) = I(z)$.
- If $\chi(0) \neq 0$, then for $z \leq \chi(0)$, ${}^{\circ}I(z) = I(\chi(0)) + I'(\chi(0))(z - \chi(0))$.
- If $\chi(1) \neq 1$, then for $z \geq \chi(1)$, ${}^{\circ}I(z) = I(\chi(1)) + I'(\chi(1))(z - \chi(1))$.

Proof. The function ${}^{\circ}I$ is defined by the equation

$${}^{\circ}I(z) = \sup_{0 \leq y \leq 1} a_z(y).$$

We know by Lemma 5.5 that $a_z(y)$ is a concave function of y for all z . Since $[0, 1]$ is compact, the supremum is attained and we can determine where by examining the derivative of a_z at the endpoints. If $a'_z(0) \leq 0$, the supremum is attained at 0, since the derivative is a decreasing function of y . Similarly, if $a'_z(1) \geq 0$ the supremum is attained at 1. Otherwise, the supremum is attained on the interior of the unit interval at the solution of

$$a'_z(y) = 0.$$

From the proof of Lemma 5.5, we know that

$$a'_z(y) = \log \frac{z(1 - \chi(y))}{\chi(y)(1 - z)}.$$

Hence,

$$\begin{aligned} a'_z(0) > 0 &\iff z > \chi(0) \\ a'_z(1) < 0 &\iff z < \chi(1), \end{aligned}$$

and

$$a'_z(y) = 0 \implies z = \chi(y)$$

Since χ has a positive derivative everywhere, it must be continuous and strictly increasing

and hence, invertible. Therefore, for $\chi(0) < z < \chi(1)$, defining $\gamma = \chi^{-1}(z)$, we have that

$$\begin{aligned} L(\circ I(z)) &= a_z(\gamma) \\ &= B(\gamma, \chi(\gamma)) + L(I(\chi(\gamma))) - B(\gamma, z) \\ &= L(I(z)). \end{aligned}$$

Hence

$$\circ I(z) = I(z),$$

since L is invertible.

For $0 \leq z \leq \chi(0)$, we have that

$$\begin{aligned} L(\circ I(z)) &= a_z(0) \\ &= B(0, \chi(0)) + L(I(\chi(0))) - B(0, z) \\ \log\left(\frac{I(z)}{\alpha} + 1\right) &= \log\frac{1-z}{1-\chi(0)} + \log\left(\frac{I(\chi(0))}{\alpha} + 1\right) \\ \circ I(z) + \alpha &= \frac{1-\chi(0)+\chi(0)-z}{1-\chi(0)}(I(\chi(0)) + \alpha), \\ \circ I(z) &= I(\chi(0)) - \frac{I(\chi(0)) + \alpha}{1-\chi(0)}(z - \chi(0)) \\ &= I(\chi(0)) + I'(\chi(0))(z - \chi(0)), \end{aligned}$$

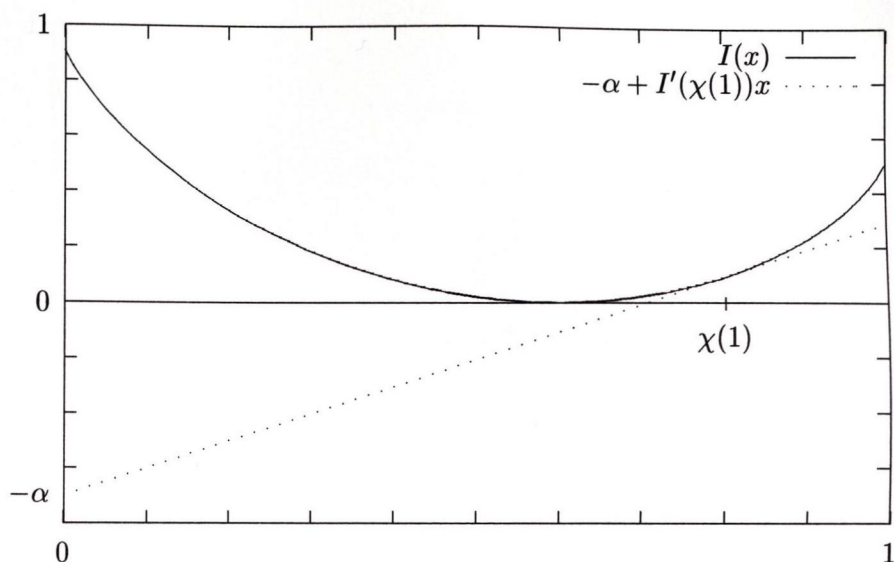
where the last step uses equation (5.11) with $y = 0$.

Similarly, for $\chi(1) \leq z \leq 1$,

$$\begin{aligned} L(\circ I(z)) &= a_z(1), \\ \circ I(z) + \alpha &= \frac{z}{\chi(1)}(I(\chi(1)) + \alpha) \\ &= I(\chi(1)) + \alpha + \frac{I(\chi(1)) + \alpha}{\chi(1)}(z - \chi(1)), \end{aligned}$$

and using equation (5.11), we get

$$\circ I(z) = I(\chi(1)) + I'(\chi(1))(z - \chi(1)).$$

Figure 5-1: A geometrical definition of $\chi(1)$

So, in conclusion,

$${}^{\circ}I(z) = \begin{cases} I(\chi(0)) + I'(\chi(0))(z - \chi(0)) & : 0 \leq z \leq \chi(0) \\ I(z) & : \chi(0) \leq z \leq \chi(1) \\ I(\chi(1)) + I'(\chi(1))(z - \chi(1)) & : \chi(1) \leq z \leq 1 \end{cases}$$

□

We can interpret the points $\chi(0)$ and $\chi(1)$ geometrically. Figure 5.2 shows the graph of a typical function, I , satisfying our requirements. For a given value of α , the corresponding value of $\chi(1)$ is the point at which a tangent to I will pass through vertical axis at $-\alpha$. There is a corresponding construction on the other side to define $\chi(0)$. The condition

$$\lim_{x \rightarrow 1} \frac{I'(x)}{I(x) + \alpha} > 1$$

ensures that there exists a point at which the tangent to I will pass through the point $(0, -\alpha)$.

Figure 5.2 shows the shape of ${}^{\circ}I$; between $\chi(0)$ and $\chi(1)$, ${}^{\circ}I$ agrees with I , and outside this region assumes the values of the dotted tangent lines. Thus ${}^{\circ}I$ is the infimum (in

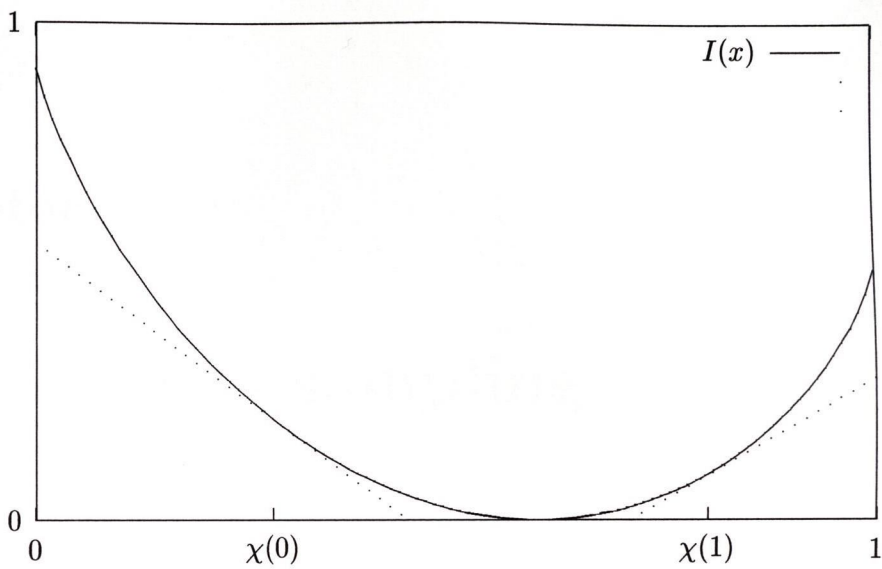


Figure 5-2: $I(x)$ with tangents at $\chi(0)$ and $\chi(1)$

the sense of the product partial order) over all convex functions which agree with I on $(\chi(0), \chi(1))$.

Chapter 6

Non-Poisson sampling

In this chapter we wish to consider the case where the sampling process is not Poisson. In Chapter 4, we were able to find a formula for the rate function of the observed process, requiring only that the underlying process obey an LDP. However, the derivation of this formula relied heavily on the particular form of the Poisson distribution. With fewer restrictions on the observation process, in Chapter 3 we were able to characterise the observed process, but this was subject to a Markovian assumption on the underlying process.

In the general case, without precise information on either the underlying or observation processes, it seems unlikely that a formula for the rate function of the observed process could be found. However, we are able to establish general conditions on the underlying process which ensure that the sample process still satisfies an LDP. The difficulty in only assuming an LDP of the underlying process is that we lose the non-asymptotic properties, while in general, the process formed of the successive samples depends in a finely detailed fashion on the interaction between the observation and underlying processes. To get around this difficulty, we would like to be able to specify some non-asymptotic properties of the underlying process which would hold if and only if the process obeyed an LDP.

In general, there is no such characterisation of the class of processes which satisfy an LDP. However, we can specify sufficient conditions; in fact, any process which is stationary and whose correlation structure is sufficiently weak will be a member of this class. These two conditions generalise, in a natural way, the simplest class of processes for which an LDP can be proved, namely processes which are composed of identically distributed and independent random variables.

We will discuss these conditions in the abstract before applying them to the random sampling transformation, but first we need to describe our setting and notation.

The setting

We choose to work in discrete time to reduce the technical complications.

Let $(\Omega, \mathcal{F}, \lambda)$ be a probability triple. We consider an underlying process $\{X_n\}_{n \in \mathbb{N}}$ taking values in K , a compact subset of $\mathbb{R}^d$, and use $\mathcal{K}$ to denote the Borel σ -algebra on K . For concreteness, we take $\Omega := K^{\mathbb{N}}$,

$$X_n : \Omega \rightarrow K : \omega \mapsto \omega_n,$$

and

$$\mathcal{F} := \sigma(X_1, X_2, \dots) = \mathcal{K}^{\mathbb{N}}.$$

We also define, for $m < n$,

$$\mathcal{F}_m^n := \sigma(X_{m+1}, \dots, X_n) \subset \mathcal{F},$$

and denote $\mathcal{F}_0^n$ by $\mathcal{F}^n$.

We define the inter-observation-time process, $\{W_n\}_{n \in \mathbb{N}}$, on the probability triple $(\Delta, \mathcal{G}, M)$. Again, for the sake of concreteness, we take $\Delta = \mathbb{N}^{\mathbb{N}}$,

$$W_n : \Delta \rightarrow \mathbb{N} : \delta \mapsto \delta_n,$$

and $\mathcal{G} = \sigma(W_1, W_2, \dots)$. We then define the observation process $\{T_n\}_{n \in \mathbb{N}}$ by

$$T_n := W_1 + \dots + W_n.$$

Assumption 6.1 *$\{W_n\}$ is a sequence of independent and identically distributed random variables which are almost surely greater than zero.*

Thus, we are assuming that M is a countable product of identical copies of a probability measure on the strictly positive integers. The random sequence $(T_1, T_2, \dots)$ also takes values

in Δ , and we can define the law of this sequence, $\mathcal{L}$, another probability measures on Δ ,

$$\mathcal{L}[B] = M \left[(T_1, T_2, \dots) \in B \right].$$

We will often use $t(\delta)$ to denote the increasing sequence $(\delta_1, \delta_1 + \delta_2, \dots)$, or $t_n(\delta)$ to denote $\delta_1 + \dots + \delta_n$.

We define the random sampling transformation by the function

$$S : \Omega \times \Delta \longrightarrow \Omega,$$

$$S(\omega, \delta) = (\omega_{\delta_1}, \omega_{\delta_1 + \delta_2}, \dots).$$

It is easy to see that S is measurable. Since $\mathcal{F}$ is generated by $\{X_n^{-1}B : B \in \mathcal{K}, n \in \mathbb{N}\}$, it is sufficient to check that the pre-image of each such set is in $\mathcal{F} \times \mathcal{G}$. For any $B \in \mathcal{K}$

$$\begin{aligned} S^{-1}(X_n^{-1}B) &= \bigcup_{\delta \in \Delta} \{\omega : S(\omega, \delta) \in B\} \times \{\delta\} \\ &= \bigcup_{\delta \in \Delta} \{X_{t_n(\delta)} \in B\} \times \{\delta\} \end{aligned}$$

Each set is obviously measurable, and hence so is their union.

We then define the measure

$$\mu = (\lambda \times M) \circ S^{-1}.$$

Using this measure, the sequence of samples, $Y_n = X_{T_n}$, is a stochastic process with respect to $(\Omega, \mathcal{F}, \mu)$, where

$$Y_n(\omega) = \omega_n.$$

We define stationarity in the usual way. For positive n , we define a shift operator by

$$\theta_n : \Omega \longrightarrow \Omega$$

$$\theta_n(\omega) = (\omega_{n+1}, \omega_{n+2}, \dots).$$

We will sometimes use θ to denote θ_1 . We say that a measure M on Ω is stationary if for

any $B \in \mathcal{F}$,

$$\mathbb{M}[\theta^{-1}B] = \mathbb{M}[B].$$

In the case where the law of the sample paths of a process is stationary, it is usual to refer to the process as being stationary.

It will prove useful to have a family of projection operators which select a subsequence from each member of Ω . For $0 \leq m < n \leq \infty$, we define

$$\begin{aligned} \pi_m^n &: \Omega \rightarrow K^{n-m}, \\ \pi_m^n &: \omega \mapsto (\omega_{m+1}, \dots, \omega_n). \end{aligned}$$

As for the σ -algebras, we use π^n to denote π_0^n . To complement this operator, we define a concatenation operator which we use to append a sequence (usually infinite) to the end of a finite sequence of the same type. For example, if $x \in \pi^n\Omega$ and $y \in \Omega$, then we define

$$x \cdot y = (x_1, \dots, x_n, y_1, y_2, \dots).$$

6.1 Correlated processes and large deviations

In order to ensure that a process obeys an LDP, in addition to stationarity we need to impose a limit on the strength of the correlations between the constituent random variables of the process. This restriction ensures that the behaviour of the process does not diverge sufficiently from that of an independent process to prevent it from obeying an LDP.

We can make this restriction precise in many ways, but most involve constructing a new measure, with respect to which certain classes of events are independent, and then comparing this to the original measure. Because we are interested in asymptotic results, the difference between the two measures will then be required to decay in some suitable fashion. Obviously, there are many procedures for imposing these restrictions; the constructed measure can be defined in terms of many different types of events, the distance between the two measures can be measured in many ways and this distance can be required to decay on any scale at any rate.

Mixing conditions

Many of these conditions are grouped together under the general heading *mixing conditions*. Their definitions typically involve constructing a sequence of measures, indexed by n , with respect to which any σ -algebras that are separated by a distance n (that is, of the form $\mathcal{F}_k^{m+k}$ and $\mathcal{F}_{m+k+n}^{m+k+n+l}$) are independent. There are then a variety of ways of defining the difference between the original measure and a member of this sequence. To ensure that an LDP holds, this difference is required to disappear, in an appropriately quick fashion, as n is increased. Bradley, in [2], reviews several different ways of comparing measures and develops relations between these different methods. These difference functions are denoted by, for example, α_n , or Φ_n , and a process for which the difference decays to zero with n is referred to as α -mixing, or Φ -mixing, respectively. Dembo and Bryc, in [3], state conditions on the rates of decay of several of such differences which guarantee that the process satisfies an LDP. These conditions on mixing rates all imply a more general condition, which they refer to as condition (S). We will refer to a process which satisfies this condition as S-mixing. Assuming the notation of the previous section, we can state the condition as follows:

Definition 8 $\{X_n\}$ is S-mixing iff for all $C < \infty$, there exists a non-decreasing sequence in $\mathbb{N}$, $\{l(n)\}$, for which

$$\sum_{n=1}^{\infty} \frac{l(n)}{n(n+1)} < \infty,$$

and such that, for all $k_1, k_2 \in \mathbb{N}$,

$$\begin{aligned} \sup \left\{ \lambda(A)\lambda(B) - e^{l(n)}\lambda(A \cap B) : A \in \mathcal{F}_0^{k_1}, B \in \mathcal{F}_{k_1+l(n)}^{k_1+l(n)+k_2} \right\} &\leq e^{-nC} \text{ (S1)}, \\ \sup \left\{ \lambda(A \cap B) - e^{l(n)}\lambda(A)\lambda(B) : A \in \mathcal{F}_0^{k_1}, B \in \mathcal{F}_{k_1+l(n)}^{k_1+l(n)+k_2} \right\} &\leq e^{-nC} \text{ (S1)}. \end{aligned}$$

We will say that λ is S-mixing if there is no ambiguity. Dembo and Bryc prove that any stationary and S-mixing process, taking values in a Polish space, will satisfy an empirical measure LDP in the τ -topology.

Weak dependence

There is an alternative approach to ensuring that a process have a sufficiently weak correlation structure. This condition, termed *weak dependence*, was first advanced in the statistical

mechanics setting, by Lewis, Pfister and Sullivan, in [8]. The notion was relaxed slightly and applied to more general stochastic processes by Russell in [9]. Using this condition he establishes a sample path LDP for stationary processes in continuous and discrete time.

In our more narrow setting, we use the following simplified definition of weak dependence.

Definition 9 For a measure $\mathbb{M}$ on Ω , we define the sequence

$$C_n := \sup_{m \geq 0} \sup_{A \in \mathcal{F}^n, B \in \mathcal{F}_n^{n+m}} \left| \log \frac{\mathbb{M}(A) \mathbb{M}(B)}{\mathbb{M}(A \cap B)} \right|.$$

$\mathbb{M}$ is weakly dependent if

$$\lim_{n \rightarrow \infty} \frac{C_n}{n} = 0.$$

The weak dependence condition is so defined in order to allow the the LDP to be proved using the sub-additivity argument; this is the technique used in both [9] and [8].

6.2 Stationarity and correlation structure after sampling

In this section, we wish to investigate whether properties of a process which guarantee an LDP are conserved by the random sampling transformation. We first turn our attention to stationarity.

Theorem 6.1 If λ is stationary, then μ is stationary.

Proof. It is well known (see, for example, [1]) that probability measures which agree on a π -system, $\mathcal{A}$, must be equal on the σ -algebra generated by $\mathcal{A}$. Therefore, if $\mathcal{A}$ is a π -system and $\mathcal{F} = \sigma(\mathcal{A})$, then we can establish stationarity of μ by checking that

$$\mu[\theta_n^{-1} A] = \mu[A] \text{ for all } A \in \mathcal{A}.$$

It is easy to see that

$$\mathcal{A} := \bigcup_{n \in \mathbb{N}} \left\{ A_1 \times \dots \times A_n \times K \times K \times \dots : A_i \in \mathcal{K} \text{ for all } i \leq n \right\}$$

is a π -system which generates $\mathcal{F}$. If A denotes a generic element of $\mathcal{A}$

$$A = A_1 \times \dots \times A_m \times K \times K \times \dots,$$

then

$$\begin{aligned}
 \mu[\theta_n^{-1}A] &= (\lambda \times M) \circ S^{-1} \circ \theta_n^{-1}A \\
 &= (\lambda \times M)[\{(\omega, \delta) : \omega_{t_{n+i}} \in A_i, 1 \leq i \leq m\}] \\
 &= \sum_{d \geq n} \sum_{d_1 \geq 1} \dots \sum_{d_m \geq 1} \lambda[\omega_{d+d_1+\dots+d_i} \in A_i, 1 \leq i \leq m] M[t_n = d, \delta_{n+i} = d_i, 1 \leq i \leq m]
 \end{aligned}$$

by stationarity of λ ,

$$= \sum_{d_1 \geq 1} \dots \sum_{d_m \geq 1} \lambda[\omega_{d_1+\dots+d_i} \in A_i, 1 \leq i \leq m] \sum_{d \geq n} M[t_n = d, \delta_{n+i} = d_i, 1 \leq i \leq m],$$

since M is a product measure,

$$\begin{aligned}
 &= \sum_{d_1 \geq 1} \dots \sum_{d_m \geq 1} \lambda[\omega_{d_1+\dots+d_i} \in A_i, 1 \leq i \leq m] M[\delta_{n+i} = d_i, 1 \leq i \leq m] \sum_{d \geq n} M[t_n = d] \\
 &= \sum_{d_1 \geq 1} \dots \sum_{d_m \geq 1} \lambda[\omega_{d_1+\dots+d_i} \in A_i, 1 \leq i \leq m] M[\delta_{n+i} = d_i, 1 \leq i \leq m]
 \end{aligned}$$

by stationarity of M ,

$$\begin{aligned}
 &= \sum_{d_1 \geq 1} \dots \sum_{d_m \geq 1} \lambda[\omega_{d_1+\dots+d_i} \in A_i, 1 \leq i \leq m] M[\delta_i = d_i, 1 \leq i \leq m] \\
 &= (\lambda \times M)[\{(\omega, \delta) : \omega_{t_i} \in A_i, 1 \leq i \leq m\}] \\
 &= (\lambda \times M) \circ S^{-1}[A] \\
 &= \mu[A]
 \end{aligned}$$

Hence, μ is stationary. □

We next want to investigate whether μ satisfies the same mixing and dependence conditions as λ . First we need to prove a lemma which helps to relate the correlation structure of μ to that of λ . For a fixed sequence of sample times $t \in \Delta$, $S(t)$ denotes the random sequence $(X_{t_1}, X_{t_2}, \dots)$.

Lemma 6.2 *If λ is stationary and $A \in \mathcal{F}^n$ and $B \in \mathcal{F}_{n+k}^{n+k+l}$ for some finite n, k, l then*

$$\mu[A]\mu[B] = \int_{t \in \Delta} \lambda[S(t) \in A] \lambda[S(t) \in B] \mathcal{L}[dt].$$

Proof. By definition,

$$\begin{aligned}\mu[A]\mu[B] &= \int_{t \in \Delta} \lambda[S(t) \in A] \mathcal{L}[dt] \int_{\gamma \in \Delta} \lambda[S(\gamma) \in B] \mathcal{L}[d\gamma] \\ &= \int_{\delta \in \Delta} \int_{\gamma \in \Delta} \lambda[S(t(\delta)) \in A] \lambda[S(t(\gamma)) \in B] M[d\delta] M[d\gamma].\end{aligned}$$

Now, since $A \in \mathcal{F}^n$, we note that

$$\begin{aligned}\{\omega : S(\omega, t(\delta)) \in A\} &= \{\omega : S(\omega, t(\pi^n \delta \cdot \pi_n^\infty \gamma)) \in A\}. \\ \Rightarrow \lambda[S(t(\delta)) \in A] &= \lambda[S(t(\pi^n \delta \cdot \pi_n^\infty \gamma)) \in A].\end{aligned}$$

We can make a similar argument, using stationarity, for the measure of the second set.

$$\begin{aligned}\{S(t(\gamma)) \in B\} &= \{\omega : (\omega_{t_1(\gamma)}, \omega_{t_2(\gamma)}, \dots) \in B\} \\ &= \left\{ \omega : (\omega_{t_{n+k+1}(\gamma)}, \dots, \omega_{t_{n+k+l}(\gamma)}) \in \pi_{n+k}^{n+k+l} B \right\},\end{aligned}$$

since $B \in \mathcal{F}_{n+k}^{n+k+l}$. We note that for all $i > 0$,

$$t_{n+k+i}(\pi^n \delta \cdot \pi_n^\infty \gamma) = t_{n+k+i}(\gamma) + t_n(\delta) - t_n(\gamma).$$

Writing $d = t_n(\delta) - t_n(\gamma)$, it is clear that :

for $d \geq 0$,

$$\theta_d \left\{ S(t(\pi^n \delta \cdot \pi_n^\infty \gamma)) \in B \right\} = \left\{ S(t(\gamma)) \in B \right\},$$

and for $d < 0$,

$$\left\{ S(t(\pi^n \delta \cdot \pi_n^\infty \gamma)) \in B \right\} = \theta_{(-d)} \left\{ S(t(\gamma)) \in B \right\}.$$

In either case, since λ is stationary, we can conclude that

$$\lambda[S(t(\gamma)) \in B] = \lambda[S(t(\pi^n \delta \cdot \pi_n^\infty \gamma)) \in B].$$

Hence,

$$\begin{aligned}\mu[A]\mu[B] &= \int_{\delta \in \Delta} \int_{\gamma \in \Delta} \lambda[S(t(\delta)) \in A] \lambda[S(t(\gamma)) \in B] M[d\gamma] M[d\delta] \\ &= \int_{\delta \in \Delta} \int_{\gamma \in \Delta} \lambda[S(t(\pi^n \delta \cdot \pi_n^\infty \gamma)) \in A] \lambda[S(t(\pi^n \delta \cdot \pi_n^\infty \gamma)) \in B] M[d\gamma] M[d\delta],\end{aligned}$$

and using the simple product structure of M ,

$$\begin{aligned}&= \int_{\delta \in \Delta} \lambda[S(t(\delta)) \in A] \lambda[S(t(\delta)) \in B] M[d\delta] \\ &= \int_{t \in \Delta} \lambda[S(t) \in A] \lambda[S(t) \in B] \mathcal{L}[dt].\end{aligned}$$

□

The following lemma is a simple result required in the proof of the weak dependence of μ .

Lemma 6.3 *Let c_n be a positive sequence such that*

$$\lim_{n \rightarrow \infty} \frac{c_n}{n} = 0.$$

Then, for any integer $B \geq 1$,

$$\begin{aligned}a_n &= \sup_{n \leq m \leq nB} c_m \\ \Rightarrow \lim_{n \rightarrow \infty} \frac{a_n}{n} &= 0.\end{aligned}$$

Proof.

Since $c_n/n \rightarrow 0$, there exists $N(\epsilon)$ such that

$$c_n < n\epsilon, \quad \forall n > N(\epsilon).$$

Therefore, for any $n > N(\epsilon)$,

$$\begin{aligned}\frac{c_m}{nB} &< \epsilon, \quad \forall m \text{ s.t. } n \leq m \leq nB, \\ \frac{a_n}{n} &< B\epsilon.\end{aligned}$$

So,

$$\frac{a_n}{n} < \epsilon, \quad \forall n > N\left(\frac{\epsilon}{B}\right)$$

Hence,

$$\lim_{n \rightarrow \infty} \frac{a_n}{n} = 0$$

□

We can now prove that the random sampling transformation preserves the weak dependence property, assuming λ is stationary and that the inter-sample times are uniformly bounded.

Theorem 6.4 *If λ is stationary and weakly dependent and the process $\{W_n\}$ is uniformly bounded, then μ is weakly dependent.*

Proof. We denote the uniform bound on $\{W_n\}$ by K . Since λ is weakly dependent, we know that $C_n/n \rightarrow 0$ where

$$C_n = \sup_{A \in \mathcal{F}^n, B \in \mathcal{F}_n^\infty} \left| \log \frac{\lambda(A \cap B)}{\lambda(A)\lambda(B)} \right|.$$

We wish to show that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \int_A \exp(-C_{t_n}) \lambda(A_t \cap B_t) \mathcal{L}[dt] \leq \mu(A) \mu(B) \leq \int_{t \in \Delta} \exp(C_{t_n}) \lambda(A_t \cap B_t) \mathcal{L}[dt].$$

Let $A \in \mathcal{F}^n$ and $B \in \mathcal{F}_n^\infty$. We know from Lemma 6.2 that

$$\mu(A) \mu(B) = \int_{t \in \Delta} \lambda(S(t) \in A) \lambda(S(t) \in B) \mathcal{L}[dt].$$

Hence, noting that

$$\begin{aligned} A_t &:= \{\omega : S(\omega, t) \in A\} \subset \mathcal{F}^{t_n}, \\ B_t &:= \{\omega : S(\omega, t) \in B\} \subset \mathcal{F}_{t_n}^\infty, \end{aligned}$$

we can conclude that

$$\int_{t \in \Delta} \exp(-C_{t_n}) \lambda(A_t \cap B_t) \mathcal{L}[dt] \leq \mu(A) \mu(B) \leq \int_{t \in \Delta} \exp(C_{t_n}) \lambda(A_t \cap B_t) \mathcal{L}[dt]. \quad (6.1)$$

We can, of course, write down a similar expression for $\mu(A \cap B)$; namely,

$$\begin{aligned}\mu(A \cap B) &= \int_{t \in \Delta} \lambda(S(t) \in A \cap B) \mathcal{L}(dt) \\ &= \int_{t \in \Delta} \lambda(A_t \cap B_t) \mathcal{L}(dt) .\end{aligned}$$

Defining $g(t) = \lambda(A_t \cap B_t)$, we can write the upper bound of inequality 6.1 as

$$\mu(A)\mu(B) \leq \int_{t \in \Delta} \exp(C_{t_n}) g(t) \mathcal{L}(dt)$$

since $W_n < K$,

$$\begin{aligned}&= \sum_{m=n}^{nK} \int_{\{t: t_n=m\}} \exp(C_m) g(t) \mathcal{L}(dt) , \\ &\leq \exp\left(\sup_{n \leq m \leq nK} C_m\right) \int_{t \in \Delta} g(t) \mathcal{L}(dt) \\ &= \exp\left(\sup_{n \leq m \leq nK} C_m\right) \mu(A \cap B) , \\ - \sup_{n \leq m \leq nK} C_m &\leq \log \frac{\mu(A \cap B)}{\mu(A)\mu(B)} .\end{aligned}$$

We can perform a similar manipulation with the lower bound of inequality 6.1 to obtain

$$\log \frac{\mu(A \cap B)}{\mu[A]\mu[B]} \leq \sup_{n \leq m \leq nK} C_m .$$

Combining these bounds, we find

$$\frac{1}{n} \left| \log \frac{\mu(A \cap B)}{\mu(A)\mu(B)} \right| \leq \frac{1}{n} \sup_{n \leq m \leq nK} C_m .$$

Now, since this bound does not depend on the choice of A or B , and Lemma 6.3 implies that the bound vanishes in the limit, we can conclude that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sup_{A \in \mathcal{F}^n, B \in \mathcal{F}_n^\infty} \left| \log \frac{\mu(A \cap B)}{\mu(A)\mu(B)} \right| = 0 .$$

□

Alternatively, we can also show that the random sampling transformation preserves the S-mixing condition, assuming that λ is stationary.

Theorem 6.5 *If λ is stationary and S-mixing, then so is μ .*

Proof. We wish to show that μ is S-mixing. We begin by choosing an arbitrary $C < \infty$. Since λ is S-mixing, we know that there exists a suitable sequence $\{l(n)\}$, so that (S1,S2) hold for λ . We shall show that μ also satisfies S1 and S2 for this sequence $\{l(n)\}$.

For arbitrary values of k_1, k_2, n we choose any $A \in \mathcal{F}^{k_1}$ and $B \in \mathcal{F}_{k_1+l(n)}^{k_1+l(n)+k_2}$, and using Lemma 6.2 we note that

$$\mu(A)\mu(B) - e^{l(n)}\mu(A \cap B) = \int_{t \in \Delta} \left(\lambda[S(t) \in A] \lambda[S(t) \in B] - e^{l(n)} \lambda[S(t) \in A \cap B] \right) \mathcal{L}(dt).$$

Now, by the definition of S we can see that

$$\begin{aligned} \{\omega : S(\omega, t) \in A\} &\in \mathcal{F}^{t_{k_1}} \\ \{\omega : S(\omega, t) \in B\} &\in \mathcal{F}_{t_{k_1+l(n)}}^{t_{k_1+l(n)+k_2}} \subset \mathcal{F}_{t_{k_1+l(n)}}^{t_{k_1+l(n)+k_3}}, \text{ for some } k_3 < \infty, \end{aligned}$$

where the last set inclusion follows since the δ_n are non-zero and almost surely finite. Since λ is S-mixing we can therefore conclude that

$$\mu(A)\mu(B) - e^{l(n)}\mu(A \cap B) \leq \int_{t \in \Delta} e^{-nC} \mathcal{L}(dt) = e^{-nC}.$$

This bound does not depend on the choice of A or B , and so we have the required uniform bound:

$$\sup \left\{ \mu(A)\mu(B) - e^{l(n)}\mu(A \cap B) : A \in \mathcal{F}^{k_1}, B \in \mathcal{F}_{k_1+l(n)}^{k_1+l(n)+k_2}, k_1, k_2 \in \mathbb{Z}_+ \right\} \leq e^{-nC}.$$

An almost identical argument leads to the second bound:

$$\sup \left\{ \mu(A \cap B) - e^{l(n)}\mu(A)\mu(B) : A \in \mathcal{F}^{k_1}, B \in \mathcal{F}_{k_1+l(n)}^{k_1+l(n)+k_2}, k_1, k_2 \in \mathbb{Z}_+ \right\} \leq e^{-nC}.$$

Since the choice of C was arbitrary, these two bounds together show that $\{Y_n\}$ is S-mixing.

□

Therefore we can conclude that if the underlying process is stationary and the observation process satisfies Assumption 6.1, then

- S-mixing of the underlying process implies an LDP for the sample process
- Weak dependence of the underlying process, and a uniform bound on the inter-observation times, implies an LDP for the sample process

Notation

Notation

$$\begin{aligned}\overline{\mathbb{R}} &= \mathbb{R} \cup \{+\infty\} \cup \{-\infty\} \\ \overline{\mathbb{R}}_+ &= \{x \in \overline{\mathbb{R}} : x \geq 0\}\end{aligned}$$

For a measures space $(\Omega, \mathcal{F})$, we define, for $x \in \Omega$, the measure

$$\delta_x[A] = \begin{cases} 1 & : x \in A \\ 0 & : x \notin A \end{cases},$$

and define a function, 1_A , on Ω by $1_A(x) = \delta_x[A]$.

For a topological space X , we define

$$\begin{aligned}\mathcal{G} &\quad \text{the set of all open sets} \\ \mathcal{F} &\quad \text{the set of all closed sets} \\ \mathcal{K} &\quad \text{the set of all compact sets}\end{aligned}$$

For a random variable X , we denote its law by $\mathcal{L}_{(X)}$.

Bibliography

- [1] Patrick Billingsley. *Probability and Measure*. Wiley series in probability and mathematical statistics. John Wiley and Sons, New York, first edition, 1979.
- [2] R.C. Bradley. Basic properties of strong mixing conditions. In E. Eberlein and M. S. Taqqu, editors, *Dependence in Probability and Statistics*, Progress in Probability and Statistics, pages 165–192. Birkhauser, Boston, U.S.A., 1986.
- [3] W. Bryc and A. Dembo. Large deviations and strong mixing. *Ann. Inst. H. Poincaré Probab. Stat.*, 32:549–569, 1996.
- [4] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*, volume 38 of *Applications of Mathematics*. Springer-Verlag, New York, second edition, 1998.
- [5] J.L. Doob. *Stochastic Processes*. Wiley Publications in Statistics. John Wiley and Sons, New York, 1953.
- [6] F. Kelly. Notes on effective bandwidth. Technical report, Cambridge Statistical Laboratory, Centre for Mathematical Sciences, Wilberforce Road, Cambridge, CB3 0WB, U.K., 1996.
- [7] J.T. Lewis and C.E. Pfister. Thermodynamic probability theory: some aspects of large deviations. *Russian Mathematical Surveys*, 50:2:47–83, 1995.
- [8] J.T. Lewis, C.E. Pfister, and W.G. Sullivan. Entropy, concentration of probability, and conditional limit theorems. *Markov Processes and Related Fields*, 1:319–386, 1995.
- [9] Raymond Russell. The large deviations of random time changes. Ph. D. thesis, 1998.

- [10] S.R.S. Varadhan. Asymptotic probabilities and differential equations. *Comm. Pure Appl. Math.*, 19:261–286, 1966.