Large Deviations

Jean-Dominique Deuschel

Daniel W. Stroock

Large Deviations

Jean-Dominique Deuschel
Department of Mathematics
Cornell University
Ithaca, New York

Daniel W. Stroock
Department of Mathematics
Massachusetts Institute of Technology
Cambridge, Massachusetts

ACADEMIC PRESS, INC.
Harcourt Brace Jovanovich, Publishers
Boston San Diego New York
Berkeley London Sydney
Tokyo Toronto

ACADEMIC PRESS, INC.
1250 Sixth Avenue, San Diego, CA 92101

United Kingdom Edition published by
ACADEMIC PRESS INC. (LONDON) LTD.
24-28 Oval Road, London NW1 7DX

Library of Congress Cataloging-in-Publication Data

Deuschel, Jean-Dominique, Date–
Large deviations / Jean-Dominique Deuschel, Daniel W. Stroock.
p. cm.—(Pure and applied mathematics; v. 137)
Rev. ed. of: An introduction to the theory of large deviations / D.W. Stroock. c1984.
Bibliography: p.
Includes index.
ISBN 0-12-213150-9
1. Large deviations. I. Stroock, Daniel W. Introduction to the theory of large deviations. II. Title. III. Series: Pure and applied mathematics (Academic Press); 137.
QA3.P8 vol. 137
[QA273.67]
510 s—dc19 89-397
[519.5'34] CIP

Printed in the United States of America
89 90 91 92 9 8 7 6 5 4 3 2 1

Preface

The title of this book to the contrary not withstanding, there is no more a "theory" of large deviations than there is a "theory" of partial differential equations; and what passes for the "theory" is, in reality, little more than a grab-bag of techniques which have been successfully applied to special situations and are therefore worth trying in sufficiently closely related settings. Thus, even though the title implies that a master key is contained herein, the reader will discover that reading this book prepares him to analyze large deviations in the same sense as the manual for his computer prepared him to write his first program; that is, hardly at all! In spite of the preceding admission, we have written this book in the belief that even (and, perhaps, particularly) when a field possesses no "CAUCHY integral formula," a useful purpose can be served by a book which surveys a few outstanding successes and attempts to codify some of the principles on which those successes are based. In the present case, the examples of success are plentiful but the underlying principles are few and somewhat illusive. We hope that the brief synopsis given below will help the reader spot and understand these few principles, at least in so far as we have recognized and understood them ourselves.

After attempting, in Section 1.1, a heuristic explanation of the ideas on which the theory of large deviations rests, the remainder of Chapter I is devoted to a detailed account of two basic examples. The first of these, which is the content of Section 1.2, is CRAMÈR's renowned theorem on the large deviations of the CESÀRO means of independent $\mathbb{R}$-valued random variables from the Law of Large Numbers. In order to emphasize, as soon as possible, that large deviations can be successfully analyzed even in an infinite dimensional context, for our second example we have chosen

SCHILDER's Theorem for re-scaled WIENER's measure. The derivation is carried out in Section 1.3, and applications to first STRASSEN's Law of the Iterated Logarithm and second to the estimates of VENTCEL and FREIDLIN are given in Section 1.4. In connection with the VENTCEL-FREIDLIN estimates, we have assumed that the reader is familiar with the elements of ITÔ's theory of stochastic differential equations; however, because the rest of the book relies on neither the contents of Section 1.4 nor a knowledge of ITÔ's calculus, readers who are not acquainted with the quirks of stochastic integration need not (on that account) be too concerned about what lies ahead.

Armed with the examples from Chapter I, we turn in Chapter II to the formulation of two of the guiding principles on which the rest of the book is more or less based. The first of these is contained in Lemma 2.1.4 which provides a reasonably general statement of the "covariant" nature of large deviations results under mappings which are sufficiently continuous. (The treatment given in Section 1.4 of the VENTCEL-FREIDLIN estimates should be ample evidence of the potential power of this principle.) In order to formulate the second general principle set forth in this chapter, we start in Section 2.1 with VARADHAN's version of the LAPLACE asymptotic formula (cf. Theorem 2.1.10) and combine this in Section 2.2 with a little elementary convex analysis to arrive at the conclusion (drawn in Theorem 2.2.21) that when large deviations are governed by a convex rate function then that rate function must be the LEGENDRE transform of the logarithmic moment generating function. Since, as we saw in Chapter I, the rate functions produced in both CRAMÈR's and SCHILDER's Theorems are in fact LEGENDRE transforms of the corresponding logarithmic moment generating functions, this observation leads one to guess that there may be circumstances in which the easiest approach to large deviation results will consist of two steps: one being an abstract existential proof that the large deviations are governed by a convex rate function and the second being the "computation" of a LEGENDRE transform. (Such a procedure is reminiscent of the time-honored technique to describe the solution to a partial differential equation by first invoking some abstract existence principle and only then trying to actually say something concrete about its properties.)

The contents of Chapters III and IV may be viewed as a sequence of examples to which the principles developed in Chapter II can be applied. In Chapter III, all the examples concern partial sums of independent random variables. After introducing, in Section 3.1, a general argument (cf. Theorem 3.1.6 and its Corollary 3.1.7) for carrying out an abstract existential proof that large deviation results for such sums are governed by convex rate functions, we return in the rest of the chapter to CRAMÈR's Theorem: this time in its full glory as a statement about random variables taking values either in a space of probability measures or in a BANACH space. Thus, Section 3.2 contains a proof of SANOV's Theorem (cf. Theorem 3.2.17) for empirical distributions; and Section 3.3 is devoted to the BANACH space version of CRAMÈR's Theorem. (In connection with the derivation of these results, we introduce in Lemma 3.2.7 a somewhat technical mini-principle which turns out to play an important role throughout the rest of the book.) Finally, in Section 3.4, we show that SCHILDER's Theorem is a special case of the BANACH space statement of CRAMÈR's Theorem and, in fact, that a SCHILDER-like result can be proved for general GAUSSian measures.

As we said before, Chapter IV is again an application of the principles laid down in Chapter 2. In particular, we now take up the study of SANOV-type theorems for MARKOV processes which do not necessarily have independent increments. In order to make the development here mimic the one in Chapter III, we impose extremely strong hypotheses to guarantee that the processes with which we are dealing possess ergodic properties which are nearly as good as those possessed by processes with independent increments. As a result, basically the same ideas as those in Chapter III apply to nice additive functionals of such processes and allow us to prove (cf. Theorems 4.1.14 and 4.2.16) that these functionals have large deviations which are governed by a convex rate function. In particular, after identifying the rate functions involved, we use these considerations to obtain a variant of the original DONSKER-VARADHAN theory for the large deviations of the normalized occupation time distribution (i.e. the empirical distribution of the position) of a MARKOV process (cf. Theorems 4.1.43 and 4.2.43). Because it is technically the simpler, we do MARKOV chains (i.e., MARKOV processes with a discrete time-parameter) in Section 4.1 and move to the continuous-time setting in Section 4.2; and in Section 4.4 we show how, under the hypotheses used in Sections 4.1 and 4.2, one can realize the large deviation theory for the empirical distribution of the whole process as the projective limit of the theory for the position. Section 4.3, which is somewhat a digression from the main theme and should probably be skipped on first reading, contains DONSKER and VARADHAN's analysis of the WIENER sausage problem.

To some extent, Chapter V represents to retreat from the pattern set in Chapters III and IV and a return to the more "hands-on" approach of Chapter I. Thus, just as in Chapter I, the approach in Chapter V is to first get an upper bound, basically as an application CHEBYSHEV's inequality; then a lower bound via ergodic considerations; and finally a reconciliation the two. A rather general treatment of the upper bound is given in Section 5.1, where, in Theorem 5.1.6 and Corollary 5.1.11, we sharpen results obtained earlier in Theorem 2.2.4. In preparation for the derivation of the lower bound, we digress in Section 5.2 and give a brief resumé of a few more or less familiar results from ergodic theory. As a first application of these considerations, we present in Section 5.3 a very general large deviation result for the empirical distribution of the position of a symmetric MARKOV process (cf. Theorem 5.3.10). Our second application is the content of Section 5.4, where we prove CHIYONOBU and KUSUOKA's recent theorem about the process level large deviations of a (not necessarily MARKOV) hypermixing process (cf. Theorem 5.4.27); and, in Section 5.5, we discuss the hypermixing property for processes which are ϵ-MARKOV.

The motivation behind Chapter V has been our desire to get away from the extremely strong ergodic assumptions on which the techniques in Chapters III and IV depend and to replace them with assumptions which have a better chance of holding in either non-compact or infinite dimensional situations. In order to test and compare the scope of the various techniques which are contained in Chapters IV and V, we describe in Chapter VI some analytic results with which one can see, at least in the context of diffusion processes, the relative position of these results as measured on the scale of elliptic coercivity.

The contents of Chapters I through IV constitute a reasonably thorough introduction to the basic ideas of the theory and more or less record lectures given by the second author during the fall of 1987. Thus, we consider these four chapters as a suitable package on which to base a semester length course for advanced graduate students with a strong background in analysis and some knowledge of probability theory. In this connection, we point out that each section ends with a large selection of exercises. Although some of these exercises are quite routine and do not require any particular ingenuity on the part of the student, others are more demanding. Indeed, we have not hesitated to include in the exercises a good deal of important material. In particular, it is only in the exercises that one can find most of the applications.

Finally, a word about the history of this book may be in order. In 1983, the second author gave a course, at the University of Colorado, in which he taught himself and one or two others something about the modern theory of large deviations. Having expended considerable effort on the task, he decided to set down everything which he then knew about the subject in a little book [**101**]. That was five years ago. In the intervening years, both the subject as well as his understanding of it have grown; and, with the aid and comfort provided by a fellow sufferer, he took on the more ambitious project of basing a full blown exposition on the course which he gave in fall of 1987 at M.I.T. Thus, the present book is a great deal longer: both because it contains more material and because the exposition is more detailed. Unfortunately, in the process of removing some of the more glaring imperfections and omissions in [**101**], we are confident that we have introduced a sufficient number of new flaws to keep our readers somewhat annoyed and, occasionally, thoroughly confounded. However, the responsibility for these flaws is entirely ours and not that of the ever patient students in 18.158, who struggled with the class notes out of which this final version evolved. In particular, we take this opportunity to thank STEVE FROMM for goading us into addressing several of the more perplexing inanities in those class notes. Also, we are indebted to MICHAEL SHARPE who saved us many harrowing hours manipulating TEX into doing our bidding (cf. the similarity between the format, if not the content, of the present volume and volume # 133 in the same series); and, last but not least, it is a pleasure for us to thank our typist for Eir beautiful work.

Cambridge, MA
December 31, 1988

Contents

I Some Examples

1.1 The General Idea

Let E be a **Polish space** (i.e.. a complete, separable metric space) and suppose that $\{\mu_\epsilon : \epsilon > 0\}$ is a family of probability measures on E with the property that $\mu_\epsilon \Longrightarrow \delta_p$ as $\epsilon \to 0$ for some $p \in E$ (i.e., μ_ϵ tends weakly to the point mass δ_p). Then, for each open set $U \ni p$, we have that $\mu_\epsilon(U^c) \longrightarrow 0$; and so we can reasonably say that, as $\epsilon \to 0$, the measures μ_ϵ "see p as being typical." Equivalently, one can say that events $\Gamma \subseteq E$ lying outside of a neighborhood of p describe increasingly "deviant" behavior. What is often an important and interesting problem is the determination of just how "deviant" a particular event is. That is, given an event Γ for which $p \notin \overline{\Gamma}$, one wants to know the rate at which $\mu_\epsilon(\Gamma)$ is tending to 0. In general, a detailed answer to this question is seldom available. However, if one restricts ones attention to events which are "very deviant" in the sense that $\mu_\epsilon(\Gamma)$ goes to zero exponentially fast and if one only asks about the exponential rate, then one has a much better chance of finding a solution and one is studying the *large deviations* of the family $\{\mu_\epsilon : \epsilon > 0\}$. In order to understand why the analysis of large deviations ought to be relatively easy and what one should expect such an analysis to yield, consider the case in which all of the measures μ_ϵ are absolutely continuous with respect to some fixed reference measure m. Since $\mu_\epsilon \Longrightarrow \delta_p$, it is reasonable to suppose that

$$\frac{d\mu_\epsilon}{dm} = g_\epsilon \exp[-I/\epsilon]$$

where $\epsilon \log g_\epsilon \longrightarrow 0$ uniformly fast as $\epsilon \to 0$ and I is a non-negative function which vanishes only at the point p. One then has, for any Γ with $m(\Gamma) < \infty$,

$$\epsilon \log\big(\mu_\epsilon(\Gamma)\big) = \log\left(\int_\Gamma g_\epsilon \cdot \exp\big[-I/\epsilon\big]\, dm\right)^\epsilon$$
$$= \log\left(\int_\Gamma \exp\big[-I/\epsilon\big]\, dm\right)^\epsilon + o(1);$$

and so (since $m(\Gamma) < \infty$)

$$\left(\int_\Gamma \exp\big[-I/\epsilon\big]\, dm\right)^\epsilon \longrightarrow -\text{ess.}\sup\Big\{\exp\big[-I(q)\big] : q \in \Gamma\Big\}$$

as $\epsilon \to 0$. (The "essential" here refers to the measure m.)

Hence, in the situation described above, we have, at least when $m(\Gamma) < \infty$:

(1.1.1) $$\lim_{\epsilon \to 0} \epsilon \log \mu_\epsilon(\Gamma) = -\text{ess.}\inf\big\{I(q) : q \in \Gamma\big\}.$$

In particular, the factor g_ϵ plays no role in the analysis of large deviations; and it is this fact which accounts for the relative simplicity of this sort of analysis. Moreover, it is often easy to extend (1.1.1) to cover all Γ's. For instance, such an extension can certainly be made if one knows that for each $L > 0$ there is a Γ_L such that

(1.1.2) $$m(\Gamma_L) < \infty \quad \text{and} \quad \overline{\lim_{\epsilon \to 0}}\, \epsilon \log\big(\mu_\epsilon(\Gamma_L^c)\big) \le -L.$$

In particular, we see that if $E = \mathbb{R}^d$, $\lambda_{\mathbb{R}^d}$ is LEBESGUE's measure on $\mathbb{R}^d$, and

(1.1.3) $$\gamma_\epsilon(dq) = (2\pi\epsilon)^{-d/2} \exp\left[-\frac{|q|^2}{2\epsilon}\right] \lambda_{\mathbb{R}^d}(dq),$$

then

(1.1.4) $$\lim_{\epsilon \to 0} \epsilon \log\big(\gamma_\epsilon(\Gamma)\big) = -\text{ess.}\inf\{|q|^2/2 : q \in \Gamma\}$$

for all measurable Γ in $\mathbb{R}^d$.

Although the preceding gives some insight into the phenomena of large deviations, it relies entirely on the existence of the reference measure m and therefore does not apply to many situations of interest (e.g., it will nearly never apply when E is an infinite dimensional space). When there is no reference measure, it is clear that (1.1.1) has got to be replaced by an expression in which m does not appear. Taking a hint from the theory of weak convergence, one is tempted to guess that a reasonable replacement for (1.1.1) in more general situations is the statement that there exists a function $I : E \longrightarrow [0,\infty]$ with the property that

$$(1.1.5) \qquad -\inf_{\Gamma^\circ} I \le \underline{\lim}_{\epsilon\to 0} \epsilon \log\big(\mu_\epsilon(\Gamma)\big) \le \overline{\lim}_{\epsilon\to 0} \epsilon \log\big(\mu_\epsilon(\Gamma)\big) \le -\inf_{\overline{\Gamma}} I.$$

For instance, it is easy to pass from (1.1.4) to (1.1.5) with $\mu_\epsilon = \gamma_\epsilon$ and $I(q) = |q|^2/2$.

With the preceding in mind, we will adopt the attitude that the study of large deviations for $\{\mu_\epsilon : \epsilon > 0\}$ centers around the identification of an appropriate I for which (1.1.5) holds. Before attempting to lay out a general strategy, we will begin by presenting two classical cases in which such a program can be successfully carried to completion.

1.1.6 Exercise.

Let $E = [0,\infty)$ and define

$$\mu_\epsilon(dq) = \frac{1}{\epsilon}\exp\big[-q/\epsilon\big]\,\lambda_{[0,\infty)}(dq)$$

for $\epsilon \in (0,\infty)$. Show that (1.1.5) holds with $I(q) = q$, $q \in [0,\infty)$.

1.2 The Classical Cramèr Theorem

Let μ be a probability measure on $\mathbb{R}$ and, for $n \ge 1$, let μ^n on $\mathbb{R}^n$ denote the n-fold tensor product of μ with itself. Next, let μ_n on $\mathbb{R}$ denote the distribution of $\mathbf{x} \in \mathbb{R}^n \longmapsto \frac{1}{n}\sum_1^n x_i$ under μ^n. Assuming that $\int_{\mathbb{R}} |x|\,\mu(dx) < \infty$, the weak law of large numbers says that $\mu_n \Longrightarrow \delta_p$, where $p = \int x\,\mu(dx)$. Thus, $\{\mu_n : n \ge 1\}$ is a candidate for a theory of large deviations (take $\mu_\epsilon = \mu_n$ for $n - 1 < 1/\epsilon \le n$ in order to make the notation here conform with that in Section 1.1). Moreover, in the case when $\mu(dx) = \gamma_1(dx)$ (cf. (1.1.3) and take the d there to be 1), we have that $\mu_n = \gamma_{1/n}$. Hence, at least for this special case, we know the theory of large deviations. Namely, we know that we can take $I(x) = |x|^2/2$. The purpose of the present section is to find the large deviation theory for other choices of μ.

We begin our program by introducing the **logarithmic moment generating function**

$$(1.2.1) \qquad \Lambda_\mu(\lambda) \equiv \log\left(\int_{\mathbb{R}} \exp\big[\lambda q\big]\,\mu(dq)\right), \quad \lambda \in \mathbb{R}.$$

Note that $\lambda \in \mathbb{R} \longmapsto \Lambda_\mu(\lambda) \in [0,\infty]$ is a lower semi-continuous convex function. Indeed, by truncation, it is easy to write Λ_μ as the non-decreasing limit of smooth functions, and the convexity of Λ_μ follows from HÖLDER's inequality. Next, let Λ_μ^* be the **Legendre transform** of Λ_μ:

$$\Lambda_\mu^*(x) \equiv \sup\{\lambda x - \Lambda_\mu(\lambda) : \lambda \in \mathbf{R}\}, \quad x \in \mathbf{R}. \tag{1.2.2}$$

Note that, by its definition as the point-wise supremum of linear functions, Λ_μ^* is necessarily lower semi-continuous and convex. In order to develop some feeling for the relationship between Λ_μ, Λ_μ^*, and μ, we present the following elementary lemma.

1.2.3 Lemma. *Let μ be a probability measure on $\mathbf{R}$. Then $\Lambda_\mu^* \geq 0$. Moreover:*

(i) *If $\int_{\mathbf{R}} |x|\,\mu(dx) < \infty$ and $p = \int_{\mathbf{R}} x\,\mu(dx)$, then $\Lambda_\mu^*(p) = 0$, Λ_μ^* is non-decreasing on $[p, \infty)$ and non-increasing on $(-\infty, p]$. In addition, for $q \geq p$, $\Lambda_\mu^*(q) = \sup\{\lambda q - \Lambda_\mu(\lambda) : \lambda \geq 0\}$ and $\mu([q,\infty)) \leq \exp[-\Lambda_\mu^*(q)]$; and, for $q \leq p$, $\Lambda_\mu^*(q) = \sup\{\lambda q - \Lambda_\mu(\lambda) : \lambda \leq 0\}$ and $\mu((-\infty, q]) \leq \exp[-\Lambda_\mu^*(q)]$.*

(ii) *If $\Lambda_\mu(\lambda) < \infty$ for all λ's in a neighborhood of 0, then $\Lambda_\mu^*(x) \longrightarrow \infty$ as $|x| \to \infty$.*

(iii) *If $\Lambda_\mu(\lambda) < \infty$ for all $\lambda \in \mathbf{R}$, then $\Lambda_\mu \in C^\infty(\mathbf{R})$ and $\Lambda_\mu^*(x)/|x| \longrightarrow \infty$ as $|x| \to \infty$.*

PROOF: We begin by noting that, since $\lambda x - \Lambda_\mu(\lambda) = 0$ for $\lambda = 0$ and every $x \in \mathbf{R}$, $\Lambda_\mu^*(x) \geq 0$.

Now suppose that $\int_{\mathbf{R}} |x|\,\mu(dx) < \infty$ and set $p = \int_{\mathbf{R}} x\,\mu(dx)$. To see that $\Lambda_\mu^*(p) = 0$, we use JENSEN's inequality to obtain

$$\Lambda_\mu(\lambda) \geq \lambda p \quad \text{for all } \lambda \in \mathbf{R}. \tag{1.2.4}$$

In particular, this shows that $\lambda p - \Lambda_\mu(\lambda) \leq 0$ for all $\lambda \in \mathbf{R}$ and so $\Lambda_\mu^*(p) \leq 0$. Since Λ_μ^* is non-negative and convex, this proves that $\Lambda_\mu^*(p) = 0$, Λ_μ^* is non-decreasing on $[p, \infty)$, and Λ_μ^* is non-increasing on $(-\infty, p]$. To complete the proof of i), we first note that, as a consequence of (1.2.4), if $q \geq p$ then $\Lambda_\mu^*(q) = \sup\{\lambda q - \Lambda_\mu(\lambda) : \lambda \geq 0\}$ and if $q \leq p$ then $\Lambda_\mu^*(q) = \sup\{\lambda q - \Lambda_\mu(\lambda) : \lambda \leq 0\}$. Hence, if $q \geq p$, then, since (by CHEBYCHEV's inequality)

$$\mu([q,\infty)) \leq \exp[-(\lambda q - \Lambda_\mu(\lambda))], \quad \lambda \geq 0,$$

we see that

$$\mu([q,\infty)) \leq \exp[-\Lambda_\mu^*(q)].$$

Similarly, if $q \leq p$, then

$$\mu((-\infty, q]) \leq \exp[-\Lambda_\mu^*(q)].$$

We next turn to the proof of (ii) and (iii). To this end, note that if $\lambda > 0$ ($\lambda < 0$) and $\Lambda_\mu(\lambda) < \infty$, then $\overline{\lim}_{x\to\infty} \Lambda_\mu^*(x)/x \geq \lambda$ ($\underline{\lim}_{x\to-\infty} \Lambda_\mu^*(x)/x \leq -\lambda$). Hence, the only assertion left to be proved is that $\Lambda_\mu \in C^\infty(\mathbf{R})$ if $\Lambda_\mu(\lambda) < \infty$ for all λ. But, by TAYLOR's Theorem and the LEBESGUE Dominated Convergence Theorem, it is easy to check that $\lambda \in (-\delta, \delta) \longmapsto \Lambda_\mu(\lambda)$ is, in fact, real-analytic as long as $\Lambda_\mu(\pm\delta) < \infty$. ∎

As a consequence of part (i) of Lemma 1.2.3 we have the following.

1.2.5 Lemma. *If $\int_{\mathbf{R}} |x|\,\mu(dx) < \infty$ then for every closed set $F \subseteq \mathbf{R}$*

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log\big(\mu_n(F)\big) \leq -\inf_F \Lambda_\mu^*.$$

PROOF: Let $p = \int_{\mathbf{R}} x\,\mu(dx)$ and note that $\int_{\mathbf{R}} |x|\,\mu_n(dx) \leq \int_{\mathbf{R}} |x|\,\mu(dx) < \infty$ and $\int_{\mathbf{R}} x\,\mu_n(dx) = p$ for all $n \geq 1$. Next, observe that if $\Lambda_n = \Lambda_{\mu_n}$, then $\Lambda_n(\lambda) = n\Lambda_\mu(\lambda/n)$, and therefore that $\Lambda_n^* = n\Lambda_\mu^*$. Now suppose that $q \geq p$ ($q \leq p$). Then, by (i) applied to μ_n we see that $\mu_n([q,\infty)) \leq \exp[-n\Lambda_\mu^*(q)]$ ($\mu_n((-\infty,q]) \leq \exp[-n\Lambda_\mu^*(q)]$). Since Λ_μ^* is non-decreasing (non-increasing) on $[p,\infty)$ (on $(-\infty,p]$), this proves the result when either $F \subseteq [p,\infty)$ or $F \subseteq (-\infty,p]$. On the other hand, if both $F \cap [p,\infty) \neq \emptyset$ and $F \cap (-\infty,p] \neq \emptyset$, let $q_+ = \inf\{x \geq p : x \in F\}$ and $q_- = \sup\{x \leq p : x \in F\}$. Then

$$\mu_n(F) \leq \exp\big[-n\Lambda_\mu^*(q_-)\big] + \exp\big[-n\Lambda_\mu^*(q_+)\big] \leq 2\exp\big[-n\inf_F \Lambda_\mu^*\big],$$

and so the result holds in this case also. ∎

1.2.6 Theorem. (CRAMÈR) *Assume that $\Lambda_\mu(\lambda) < \infty$ for every $\lambda \in \mathbf{R}$. Then for every measurable $\Gamma \subseteq \mathbf{R}$ one has that*

$$-\inf_{\Gamma^\circ} \Lambda_\mu^* \leq \underline{\lim_{n\to\infty}} \frac{1}{n} \log\big(\mu_n(\Gamma)\big) \leq \overline{\lim_{n\to\infty}} \frac{1}{n} \log\big(\mu_n(\Gamma)\big) \leq -\inf_{\overline{\Gamma}} \Lambda_\mu^*.$$

(*We adopt here, and throughout, the convention that the infimum over the null set is $+\infty$.*)

PROOF: In view of Lemma 1.2.5, we need only show that if $q \in \mathbf{R}$ and $\delta > 0$,

$$\underline{\lim_{n\to\infty}} \frac{1}{n} \log\big[\mu_n((q-\delta, q+\delta))\big] \geq -\Lambda_\mu^*(q). \tag{1.2.7}$$

In proving (1.2.7), we first suppose that there is a $\lambda \in \mathbf{R}$ for which $\Lambda_\mu^*(q) = \lambda q - \Lambda_\mu(\lambda)$. Consider the probability measure

$$\tilde{\mu}(dx) = \frac{\exp[\lambda x]}{\exp[\Lambda_\mu(\lambda)]}\mu(dx),$$

and define the measures $\tilde{\mu}_n$ accordingly. Note that $\int_{\mathbb{R}} |x|\,\tilde{\mu}(dx) < \infty$ and that

$$\int_{\mathbb{R}} x\,\tilde{\mu}(dx) = \frac{1}{\exp[\Lambda_\mu(\lambda)]}\int_{\mathbb{R}} x\exp[\lambda x]\mu(dx) = \frac{d}{dt}\Lambda_\mu(t)\Big|_{t=\lambda}.$$

At the same time, note that $\frac{d}{dt}\big(tq - \Lambda_\mu(t)\big)\big|_{t=\lambda} = 0$ since $t \in \mathbb{R} \longmapsto tq - \Lambda_\mu(t)$ achieves its maximum value at λ. Combining these, we conclude that $q = \int_{\mathbb{R}} x\,\tilde{\mu}(dx)$ and therefore (by the Weak Law of Large Numbers) that $\tilde{\mu}_n((q-\delta, q+\delta)) \longrightarrow 1$ as $n \to \infty$. Assuming that $\lambda \geq 0$, note that

$$\begin{aligned}\mu_n((q-\delta,q+\delta)) &= \mu^n(\Delta_n)\\ &\geq \exp[-n\lambda(q+\delta)]\int_{\Delta_n}\exp\left[\lambda\sum_1^n y_k\right]\mu^n(dy)\\ &= \exp\Big[-n\big(\lambda(q+\delta) - \Lambda_\mu(\lambda)\big)\Big]\tilde{\mu}_n\big((q-\delta,q+\delta)\big),\end{aligned}$$

where

$$\Delta_n = \left\{\mathbf{y} \in \mathbb{R}^n : \left|\frac{1}{n}\sum_1^n y_k - q\right| < \delta\right\}.$$

From this and the preceding comments, we conclude that

$$\varliminf_{n\to\infty}\frac{1}{n}\log\Big[\mu_n((q-\delta,q+\delta))\Big] \geq -\Lambda_\mu^*(q) - \lambda\delta$$

for every $\delta > 0$. Since the left hand side of the above is clearly non-decreasing as a function of $\delta > 0$, we have now proved (1.2.7) for the case when there is a $\lambda \geq 0$ for which $\Lambda_\mu^*(q) = \lambda q - \Lambda_\mu(\lambda)$. Clearly, the same argument (with $q - \delta$ replacing $q + \delta$) works when $\Lambda_\mu^*(q) = \lambda q - \Lambda_\mu(\lambda)$ for some $\lambda \leq 0$.

We must now handle the case in which $\Lambda_\mu^*(q) > \lambda q - \Lambda_\mu(\lambda)$ for all $\lambda \in \mathbb{R}$. If $q \geq \int_{\mathbb{R}} x\mu(dx)$, then (cf. (i) of Lemma 1.2.3) there exists a sequence $\lambda_\ell \nearrow \infty$ such that $\lambda_\ell q - \Lambda_\mu(\lambda_\ell) \nearrow \Lambda_\mu^*(q)$. Since it is clear that

$$\int_{(-\infty,q)}\exp[\lambda_\ell\cdot(x-q)]\,\mu(dx) \longrightarrow 0,$$

we have that

$$\int_{[q,\infty)}\exp[\lambda_\ell\cdot(x-q)]\,\mu(dx) \longrightarrow \exp[-\Lambda_\mu^*(q)].$$

But this is possible only if $\mu((q,\infty)) = 0$ and $\mu(\{q\}) = \exp[-\Lambda_\mu^*(q)]$. Hence, $\mu^n(\{(q,\ldots,q)\}) = \exp[-n\Lambda_\mu^*(q)]$, and so $\mu_n(\{q\}) \geq \exp[-n\Lambda_\mu^*(q)]$. Clearly, this implies (1.2.7) holds for every $\delta > 0$. An analogous argument can be used in the case when $q \leq \int_{\mathbb{R}} x\,\mu(dx)$. ∎

1.2.8 Remark.

The reader should take note of the structure of the preceding line of reasoning. Namely, the upper bound comes from optimizing over a family of CHEBYCHEV inequalities; while the lower bound comes from introducing a RADON–NIKODYM factor in order to make what was originally "deviant" behavior look like typical behavior. This pattern of proof is one of the two most powerful tools in the theory of large deviations. In particular, it will be used in the next section as well as Sections 5.3 and 5.4.

1.2.9 Exercise.

Assuming that $\int_{\mathbb{R}} |x|\, \mu(dx) < \infty$, show that

$$\int_{\mathbb{R}} \exp\left[\alpha \Lambda_\mu^*(x)\right] \mu(dx) \le \frac{2}{1-\alpha}, \quad \alpha \in (0,1). \tag{1.2.10}$$

Hint: Set $p = \int_{\mathbb{R}} x\, \mu(dx)$ and show that if $\Lambda_\mu^*(q) < \infty$, then

$$\int_{[p,q]} \exp\left[\alpha \Lambda_\mu^*(x)\right] \mu(dx) \le \frac{1}{1-\alpha} \quad \text{or} \quad \int_{[q,p]} \exp\left[\alpha \Lambda_\mu^*(x)\right] \mu(dx) \le \frac{1}{1-\alpha}$$

according to whether $q \ge p$ or $q \le p$.

1.2.11 Exercise.

(i) Show that for every $p \in \mathbb{R}$: $\Lambda_{\mu_p}^*(x) = \Lambda_\mu^*(x-p)$, $x \in \mathbb{R}$, where $\mu_p \equiv \delta_p * \mu$ and we use $\nu * \mu$ to denote the convolution of ν with μ.

(ii) If $\mu = \alpha\delta_a + (1-\alpha)\delta_b$, where $a < b$ and $\alpha \in (0,1)$, show that

$$\Lambda_\mu^*(x) = \begin{cases} \infty & \text{for } x \notin [a,b] \\ \frac{x-a}{b-a} \log \frac{x-a}{(1-\alpha)(b-a)} + \frac{b-x}{b-a} \log \frac{b-x}{\alpha(b-a)} & \text{for } x \in [a,b], \end{cases}$$

where $0 \log 0 \equiv 0$.

(iii) If $\mu(dx) = \chi_{[0,\infty)}(x) e^{-x}\, dx$, show that

$$\Lambda_\mu^*(x) = \begin{cases} \infty & \text{for } x \le 0 \\ x - 1 - \log x & \text{for } x > 0. \end{cases}$$

(iv) If $\mu(dx) = (2\pi\sigma^2)^{-1/2} \exp\left[-(x-a)^2/2\sigma^2\right] dx$, where $a \in \mathbb{R}$ and $\sigma > 0$, show that

$$\Lambda_\mu^*(x) = \frac{(x-a)^2}{2\sigma^2}.$$

1.3 Schilder's Theorem

In this section we give an example of a large deviation result for a certain family of measures on an infinite dimensional space.

Let $d \in \mathbb{Z}^+$ be given and set

$$\Theta = \left\{\theta \in C\big([0,\infty);\mathbf{R}^d\big) : \theta(0) = 0 \quad \text{and} \quad \lim_{t\to\infty} \frac{|\theta(t)|}{t} = 0\right\}. \tag{1.3.1}$$

For $\theta \in \Theta$ define

$$\|\theta\|_\Theta = \sup_{t\geq 0} \frac{|\theta(t)|}{1+t},$$

and observe that $(\Theta, \|\cdot\|_\Theta)$ is a separable real BANACH space. In order to represent the dual Θ^* of Θ, note that Θ is naturally isometric to the space of continuous paths on $[0,\infty)$ which vanish at 0 and at ∞ (namely, map θ to the path $t \longmapsto (1+t)^{-1}\theta(t)$); and use this isometry to identify Θ^* with the space of $\mathbf{R}^d$-valued, BOREL measures λ on $[0,\infty)$ with the properties that $\lambda(\{0\}) = 0$ and $\int_{[0,\infty)}(1+t)\,|\lambda|(dt) < \infty$, where $|\lambda|$ denotes the variation measure associated with λ. With this identification, the duality relation ${}_{\Theta^*}\langle\lambda,\theta\rangle_\Theta$ is given by $\int_{[0,\infty)}\theta(t)\cdot\lambda(dt)$ (the "$\cdot$" here standing for the ordinary inner product in $\mathbf{R}^d$) and $\|\lambda\|_{\Theta^*} = \int_{[0,\infty)}(1+t)\,|\lambda|(dt)$.

Let $\mathcal{B} = \mathcal{B}_\Theta$ denote the BOREL field over Θ; and, for $t \geq 0$, let $\mathcal{B}_t$ denote the smallest σ-algebra over Θ with respect to which all of the maps $\theta \longmapsto \theta(s)$, $s \in [0,t]$, are measurable. As is easy to check, $\mathcal{B} = \sigma\big(\bigcup_{t\geq 0}\mathcal{B}_t\big)$. The following remarkable existence theorem is due to N. WIENER [112]. We have added a few small embellishments to WIENER's original statement.

1.3.2 Theorem. (WIENER) *There is a unique probability measure $\mathcal{W}$ on $(\Theta,\mathcal{B})$ with the property that*

$$\int_\Theta \exp\Big[\sqrt{-1}\;{}_{\Theta^*}\langle\lambda,\theta\rangle_\Theta\Big]\,\mathcal{W}(d\theta) = \exp\big[-\Lambda_{\mathcal{W}}(\lambda)\big], \quad \lambda \in \Theta^*, \tag{1.3.3}$$

where

$$\Lambda_{\mathcal{W}}(\lambda) \equiv \frac{1}{2}\int_{[0,\infty)^2} s\wedge t\;\lambda(ds)\cdot\lambda(dt). \tag{1.3.4}$$

Moreover, if P is a probability measure on $(\Theta,\mathcal{B})$, then $P = \mathcal{W}$ if and only if any one of the following holds:

(i) *For all $0 \leq s < t$, the random variable $\theta \longmapsto \theta(t) - \theta(s)$ under P is independent of $\mathcal{B}_s$ and is GAUSSian with mean 0 and covariance $(t-s)I_{\mathbf{R}^d}$.*