LARGE DEVIATIONS

S. R. S. VARADHAN

Date: Spring, 2010.

1. Introduction

The theory of large deviations deals with rates at which probabilities of certain events decay as a natural parameter in the problem varies. It is best to think of a specific example to clarify the idea. Let us suppose that we have n independent and identically distributed random variables $\{X_i\}$ having mean zero and variance one with a common distribution μ. The distribution of $\overline{X}_n = \frac{X_1+\cdots+X_n}{n}$ converges to the degenerate distribution at 0, while $Z_n = \frac{X_1+\cdots+X_n}{\sqrt{n}}$ has a limiting normal distribution according to the Central Limit Theorem. In particular

$$\lim_{n\to\infty} P[Z_n \geq \ell] = \frac{1}{\sqrt{2\pi}} \int_\ell^\infty e^{-\frac{x^2}{2}} dx \tag{1.1}$$

While the convergence is uniform in ℓ it does not say much if ℓ is large. A natural question to ask is does the ratio

$$\frac{P[Z_n \geq \ell]}{\frac{1}{\sqrt{2\pi}} \int_\ell^\infty e^{-\frac{x^2}{2}} dx}$$

tend to 1 even when $\ell \to \infty$? It depends on how rapidly ℓ is getting large. If $\ell << \sqrt{n}$ it holds under suitable conditions. But if $\ell \simeq \sqrt{n}$ it clearly does not. For instance if $\{X_i\}$ are bounded by a constant C, and $\ell > C\sqrt{n}$, $P[Z_n \geq \ell] = 0$ while the Gaussian probability is not. Large deviations are when $\ell \simeq \sqrt{n}$. When $\ell << \sqrt{n}$ they are referred to as "moderate deviations". While moderate deviations are refinements of the Central Limit Theorem, large deviations are different. It is better to think of them as estimating the probability

$$p_n(\ell) = P[X_1 + \cdots + X_n \geq n\ell] = P[\overline{X}_n \geq \ell]$$

It is expected that the probability decays exponentially. In the Gaussian case it is clear that

$$p_n(\ell) = \frac{\sqrt{n}}{\sqrt{2\pi}} \int_\ell^\infty e^{-\frac{nx^2}{2}} dx = e^{-\frac{n\ell^2}{2} + o(n)} = e^{-n\, I(\ell) + o(n)}$$

The function $I(\ell) = \frac{\ell^2}{2}$ reflects the fact that the the common distribution of $\{X_i\}$ was Gaussian to begin with.

Following Cramer, let us take for the common distribution of our random variables $X_1, X_2, \cdots, X_n$ an arbitrary distribution μ. We shall try to estimate $P[\overline{X}_n \in A]$ where $A = \{x : x \geq \ell\}$ for some $\ell > m$, m being the mean of the distribution μ. Under suitable assumptions $\mu_n(A)$ should again decay exponentially rapidly in n as $n \to \infty$, and our goal is to find the precise exponential constant. In other words we want to calculate

$$\lim_{n\to\infty} \frac{1}{n} \log \mu_n(A) = -I_\mu(\ell)$$

as explicitly as possible. The attractive feature of the large deviation theory is that such objects can be readily computed. If we denote the sum by $S_n = \sum_{i=1}^n X_i$ then

$$\mu_n(A) = \Pr[S_n \geq n\ell]$$

can be estimated by the standard Tchebechev type estimate

$$\mu_n(A) \leq e^{-\theta\, n\, \ell} E\{\exp[\theta S_n]\}$$

with any $\theta > 0$. Denoting the moment generating function of the underlying distribution μ by

$$M(\theta) = \int e^{\theta x} d\mu$$

and its logarithm by

$$\psi(\theta) = \log M(\theta)$$

we obtain the obvious inequality

$$\mu_n(A) \leq e^{-\theta\, n\, \ell}[M(\theta)]^n$$

and by taking logarithms on both sides and dividing by n

$$\frac{1}{n} \log \mu_n(A) \leq -\theta\, \ell + \psi(\theta).$$

Since the above inequality is valid for every $\theta \geq 0$ we should optimize and get

$$\frac{1}{n} \log \mu_n(A) \leq - \sup_{\theta \geq 0} [\theta\, \ell - \psi(\theta)].$$

By an application of Jensen's inequality we can see that if $\theta < 0$, then $\psi(\theta) \geq m\theta \geq \ell\theta$. Because 0 is always a trivial lower bound replacing $\sup_{\theta \geq 0} [\theta\, \ell - \psi(\theta)]$ by $\sup_\theta [\theta\, \ell - \psi(\theta)]$ does not increase its value. Therefore we have

$$\frac{1}{n} \log \mu_n(A) \leq - \sup_{\theta} [\theta\, \ell - \psi(\theta)]. \tag{1.2}$$

It is in fact more convenient to introduce the conjugate function

$$h(\ell) = \sup_{\theta} [\theta\, \ell - \psi(\theta)] \tag{1.3}$$

which is seen to be a nonnegative convex function of ℓ with a minimum value of 0 at the mean $\ell = m$. From the convexity we see that $h(\ell)$ is non-increasing for $\ell \leq m$ and nondecreasing for $\ell \geq m$. We can now rewrite our upper bound as

$$\frac{1}{n} \log \mu_n(A) \leq -h(\ell)$$

for $\ell \geq m$. A similar statement is valid for the sets of the form $A = \{x : x \leq \ell\}$ with $\ell \leq m$. We shall now prove that the upper bounds that we obtained are optimal. To do this we need effective lower bounds. Let us first assume that $\ell = m$. In this case we know by central limit theorem that

$$\lim_{n \to \infty} \mu_n(A) = \frac{1}{2}$$

and clearly it follows that

$$\liminf_{n \to \infty} \frac{1}{n} \log \mu_n(A) \geq 0 = -h(m).$$

To do the general case let us assume that the distribution μ does not live on any proper subinterval of $(-\infty, \infty)$. Then $\psi(\cdot)$ grows super-linearly at ∞ and the supremum in $\sup_\theta[\theta\ell - \psi(\theta)]$ is attained at some value θ_0 for θ. Equating the derivative to 0,

$$\ell = \psi'(\theta_0) = \frac{M'(\theta_0)}{M(\theta_0)} = \frac{1}{M(\theta_0)} \int e^{\theta_0 x} d\mu(x). \tag{1.4}$$

If we now define a new probability distribution μ' by the relation

$$\frac{d\mu'}{d\mu}(x) = \frac{e^{\theta_0 x}}{M(\theta_0)} \tag{1.5}$$

then μ' has ℓ for its expected value. If we denote be μ'_n the distribution of the mean of the n random variables under the assumption that their common distribution is μ' rather than μ, an elemenatary calculation yields

$$\mu_n(A) = \int_A [M(\theta_0)e^{-\theta_0 x}]^n d\mu'_n(x).$$

To get a lower bound let us replace A by $A_\delta = \{x : \ell \leq x \leq \ell + \delta\}$ for some $\delta > 0$. Then

$$\mu_n(A) \geq \mu_n(A_\delta) \geq [M(\theta_0)e^{-\theta_0(\ell+\delta)}]^n \mu'_n(A_\delta).$$

Again applying the central limit theorem, but now to μ'_n,we see that $\mu'_n(A_\delta) \to \frac{1}{2}$, and taking logarithms we get

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(A) \geq \psi(\theta_0) - \theta_0(\ell + \delta).$$

Since $\delta > 0$ was arbitrary we can let it go to 0 and obtain

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(A) \geq \psi(\theta_0) - \theta_0\, \ell = -h(\ell).$$

A similar proof works for intervals of the form $A = \{x : x \leq \ell\}$ as well. We have therefore calculated the precise exponential decay rate of the relevant probabilities.

Remark 1.1. This way of computing the probability of a rare event by changing the underlying model to one where the rare event is no longer rare and using the explicit form of the Radon-Nikodym derivative to estimate the probabiliy goes back to Cramér-Lundberg in the 1930's. The problem is to estimate

$$P[\sup_t X(t) \geq \ell] = q(\ell) \tag{1.6}$$

where $X(t) = \sum_{\tau_i \leq t} \xi_i - m\,t$ is all the claims paid up to time t less the premiums collected reflecting the net total loss up to time t. ℓ is the initial capital of an insurance company. The claims follow a compound Poisson process with the Lévy-Khinitchine representation

$$\log E[\exp[\theta\, X(t)]] = t\, \log \psi(\theta) = t\, \lambda \int (e^{\theta\, x} - 1) dF(x) - m\, \theta\, t \tag{1.7}$$

λ is the claims rate, F is the distribution of the individual claim amounts and m is the rate at which premium is being collected. It is assumed that the company is profitable, i.e. $\lambda \int x\, dF(x) < m$. If ℓ is large the probability of ever running out of cash should be small.

The problem is to estimate $q(\ell)$ for large ℓ. First consider the imbedded random walk, which is the net out flow of cash after each claim is paid. $S_{n+1} = S_n + \xi_{n+1} - \tau_{n+1} m = S_n + Y_{n+1}$. ξ_{n+1} is the amount of the new claim and τ_{n+1} is the gap between the claims during which a premium of $m\,\tau_{n+1}$ was collected.The idea of "tilting" is to consider $\theta_0 > 0$ so that

$$E[e^{\theta_0 Y}] = E[e^{\theta_0(\xi - m\,\tau)}] = E[e^{\theta_0 \xi}]E[e^{-m\theta_0 \tau}] = 1$$

Solve for

$$M(\theta_0) = \frac{\lambda + m\theta_0}{\lambda}$$

Such $\theta_0 \neq 0$ exists because $E[Y] < 0$. Tilt so that the (ξ, τ) is now distributed as $e^{\theta_0 y} e^{-\theta_0\,\tau} dF(y) e^{-\lambda\tau} d\tau$ The tilted process Q will now have net positive outflow. We have made the claims more frequent and more expensive, while keeping the premium the same. It will run out of cash. Let σ_ℓ be the overshoot, i.e. shortfall when that happens. For the tilted process $E[e^{-\theta_0 Y}] = 1$.

$$q(\ell) = E^Q[e^{-\theta_0(\ell + \sigma_\ell)}] = e^{-\theta_0\,\ell} E^Q[e^{-\theta_0 \sigma_\ell}]$$

$E^Q[e^{-\theta_0 \sigma_\ell}]$ has a limit as $\ell \to \infty$ that is calculated by the renewal theorem.

Remark 1.2. If the underlying distribution were the standard Gaussian then we see that $h(\ell) = \frac{\ell^2}{2}$.

Remark 1.3. One should really think of $h(\ell)$ as giving the "local" decay rate of the probability at or near "ℓ". Because $a^n + b^n \leq 2[\max(a, b)]^n$ and the factor 2 leaves no trace when we take logarithms and divide by n, the global decay rate is really the worst local decay rate. The correct form of the result should state

$$\lim_{n\to\infty} \frac{1}{n} \log \mu_n(A) = -\inf_{x\in A} h(x).$$

Using the monotonicity of h on either side of m one can see that this is indeed correct for sets of the form $A = (-\infty, \ell)$ or (ℓ, ∞). In fact the upper bound

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(A) \leq -\inf_{x\in A} h(x)$$

is easily seen to be valid for more or less any set A. Just consider the smallest set of the form $(-\infty, a] \cup [b, \infty)$, with $a \leq m \leq b$ that contains A. While a and b may not be in A they clearly belong to $\bar{A}$, the closure of A. If we use the continuity of h we see that A can be arbitrary. If we do not want to use the continuity of h then it is surely true for closed sets. The lower bound is more of a problem. In the Gaussian case if A is a set of Lebesgue measure zero, then $\mu_n(A) = 0$ for all n and the lower bound clearly does not hold in this case. Or if we take the coin tossing or the Binomial case, if A contains only irrationals, we are again out of luck. We can see that the proof we gave works if the point ℓ is an interior point of A. Therefore it is natural to assume for the lower bound that A is an open set just as it is natural to assume that A is closed for the upper bound.

Exercise 1.
For the coin tossing where μ is the discrete distribution with masses p and $q = 1 - p$ at 1 and 0 respectively calculate explicitly the function $h(x)$. In this case our proof of the lower bound may not be valid because μ lives on a bounded interval. Provide two proofs for the lower bound, one by explicit combinatorial calculation and Stirling's formula and the other by an argument that extends our proof to the general case.

Exercise 2.
Extend the result to the case where $X_1, X_2, \cdots, X_n$ take values in R^d with a common distribution μ on R^d. Tchebechev will give estimates of probabilities for halfspaces and for a ball we can try to get the halfspace that gives the optimal estimate. If the ball is small this is OK. A compact set is then covered by a finite number of small balls and a closed set is approximated by closed bounded sets by cutting it off. This will give the upper bound for closed sets and the lower bound for open sets presents no additional difficulties in the multidimensional case.

Exercise 3.
Let us look at the special case of μ on R^d that has probabilities $\pi_1, \pi_2, \cdots, \pi_d$, (with $\sum_i \pi_i = 1$), respectively at the unit vectors in the coordinate directions $e_1, e_2, \cdots, e_d$. Calculate explicitly the rate function $h(x)$ in this case. Is there a way of recovering from this example the one dimensional result for a general discrete μ with d mass points?

Exercise 4.
If we replace the compound Poisson process by Brownian motion with a positive drift $x(t) = \beta(t) + mt$, $m > 0$, then

$$q(\ell) = P[\inf_{t \geq 0} x(t) \leq -\ell] = e^{-2m\ell}.$$

References:

1. H. Cramér; Collective Risk Theory, Stockholm, 1955
2. H. Cramér; Sur un nouveau théorème limite de la théorie des probabilités; Actualités Scientifiques, Paris, 1938

2. General Principles

In this section we will develop certain basic principles that govern the theory of large deviations. We will be working for the most part on Polish (i.e. complete separable metric) spaces. Let X be such a space. A function $I(\cdot) : X \to [0, \infty]$, will be called a (proper) rate function if it is lower semicontinuous and the level sets $K_\ell = \{x : I(x) \leq \ell\}$ are compact in X for every $\ell < \infty$. Let P_n be a sequence of probability distributions on X. We say that P_n satiesfies the large deviation principle on X with rate function $I(\cdot)$ if the following two statements hold. For every closed set $C \subset X$,

$$\limsup_{n \to \infty} \frac{1}{n} \log P_n(C) \leq - \inf_{x \in C} I(x)$$

and for every open set $G \subset X$,

$$\liminf_{n\to\infty} \frac{1}{n} \log P_n(G) \geq -\inf_{x\in G} I(x)$$

Remark 2.1. The value of $+\infty$ is allowed for the function $I(\cdot)$. All the infima are effectively taken only on the set where $I(\cdot) < \infty$.

Remark 2.2. The examples of the previous section are instances where the LDP is valid. One should check that the assumption of compact level sets holds in those cases, and this verification is left as an exercise.

Remark 2.3. Under our assumptions the infimum $\inf_x I(x)$ is clearly attained and since $P_n(X) = 1$, this infimum has to be 0. If we define $K_0 = \{x : I(x) = 0\}$, then K_0 is a compact set and the sequence P_n is tight with any limit point Q of the sequence satisfying $Q(K_0) = 1$. In particular if K_0 is a single point x_0 of X, then $P_n \Rightarrow \delta_{x_0}$, i.e. P_n converges weakly to the distribution degenerate at the point x_0 of X.

There are certain general properties that are easy to prove.

Theorem 2.4. *Suppose P_n is a sequence that satisfies an LDP on X with respect to a proper rate function $I(\cdot)$. Then for any $\ell < \infty$ there exists a compact set $D^\ell \subset X$ such that*

$$P_n(D^\ell) \geq 1 - e^{-n\,\ell}$$

for every n.

Proof. Let us consider the compact level set $A^\ell = \{x : I(x) \leq \ell+2\}$. From the compactness of A^ℓ it follows that for each k, it can be covered by a finite number of open balls of radius $\frac{1}{k}$. The union, which we will denote by U_k, is an open set and $I(x) \geq (\ell+2)$ on the closed set U_k^c. By the LDP, for every k,

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(A_k^c) \leq -(\ell+2).$$

In particular there is an $n_0 = n_0(k)$ such that for $n \geq n_0$

$$P_n(A_k^c) \leq e^{-n\,(\ell+1)}.$$

We can assume without loss of generality that $n_0(k) \geq k$ for every $k \geq 1$. We then have for $n \geq n_0$

$$P_n(A_k^c) \leq e^{-k} e^{-n\,\ell}.$$

For $j = 1, 2, \cdots, n_0(k)$, we can find compact sets $B_{k,1}, B_{k,2}, \cdots, B_{k,n_0}$ such that $P_j(B_{k,j}^c) \leq e^{-k} e^{-j\,\ell}$ for every j. Let us define $E = \cap_k [A_k \cup (\cup_j B_{k,j})]$. E is clearly totally bounded and

$$P_n(E^c) \leq \big[\sum_{k\geq 1} e^{-k}\big] e^{-n\,\ell} \leq e^{-n\,\ell}$$

We are done. ∎

Remark 2.5. The conclusion of the theorem which is similar in spirit to Prohorov's tightness condition in the context of weak convergence will be caled superexponential tightness.

There are some elementary relationships where the validity of LDP in one context implies the same in other related situations. We shall develop a couple of them.

Theorem 2.6. *Suppose P_n and Q_n are two sequences on two spaces X and Y satisfying the LDP with rate functions $I(\cdot)$ and $J(\cdot)$ respectively. Then the sequence of product measures $R_n = P_n \times Q_n$ on $X \times Y$ satisfies an LDP with the rate function $K(x, y) = I(x) + J(y)$.*

Proof. The proof is typical of large deviation arguments and we will go through it once for the record.

Step 1. Let us pick a point $z = (x, y)$ in $Z = X \times Y$. Let $\epsilon > 0$ be given. We wish to show that there exists an open set $N = N_{z,\epsilon}$ containing z such that

$$\limsup_{n\to\infty} \frac{1}{n} \log R_n(N) \le -K(z) + \epsilon.$$

Let us find an open set U_1 in X such that $I(x') \ge I(x) - \frac{1}{2}\epsilon$ for all $x \in U_1$. This is possible by the lower semicontinuity of $I(\cdot)$. By general separation theorems in a metric space we can find an open set U_2 such that $x \in U_2 \subset \bar{U}_2 \subset U_1$. By the LDP of the sequence P_n

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(U_2) \le \limsup_{n\to\infty} \frac{1}{n} \log P_n(\bar{U}_2) \le -\inf_{x'\in\bar{U}_2} I(x') \le -I(x) + \frac{1}{2}\epsilon.$$

We can repeat a similar construction around y in Y to get

$$\limsup_{n\to\infty} \frac{1}{n} \log Q_n(V_2) \le \limsup_{n\to\infty} \frac{1}{n} \log P_n(\bar{V}_2) \le -\inf_{y'\in\bar{V}_2} J(y') \le -J(y) + \frac{1}{2}\epsilon$$

for some open sets V_1, V_2 with $y \in V_2 \subset \bar{V}_2 \subset V_1$. If we take $N = U_2 \times V_2$ as the neighborhood of $z = (x, y)$ we are done.

Step 2. Let $D \subset Z = X \times Y$ be a compact set. Let $\epsilon > 0$ be given. We will show that for some neighborhood D_ϵ of D

$$\limsup_{n\to\infty} \frac{1}{n} \log R_n(D_\epsilon) \le -\inf_{z\in D} K(z) + \epsilon.$$

We know from step 1. that for each $z \in D$ there is a neighborhood N_z such that

$$\limsup_{n\to\infty} \frac{1}{n} \log R_n(N_z) \le -K(z) + \epsilon \le -\inf_{z\in D} K(z) + \epsilon.$$

From the open covering $\{N_z\}$ of D we extract a finite subcover $\{N_j : 1 \le j \le k\}$ and if we take $D_\epsilon = \cup_j N_j$, then

$$R_n(D_\epsilon) \le \sum_j R_n(N_j) \le k \sup_j R_n(N_j).$$

Since k leaves no trace after taking logarithms, dividing by n and passing to the limit we get

$$\limsup_{n\to\infty}\frac{1}{n}\log R_n(D_\epsilon)\le -\inf_{z\in D}K(z)+\epsilon.$$

In particular because $\epsilon>0$ is arbitrary we get

$$\limsup_{n\to\infty}\frac{1}{n}\log R_n(D)\le -\inf_{z\in D}K(z).$$

Step 3. From the superexponential tightness, for any given ℓ there are compact sets A^ℓ and B^ℓ in X and Y respectively such that $P_n([A^\ell]^c)\le e^{-n\,\ell}$ and $P_n([B^\ell]^c)\le e^{-n\,\ell}$. If we define the compact set $C^\ell\subset Z$ by $C^\ell=A^\ell\times B^\ell$ then $R_n([C^\ell]^c)\le 2e^{-n\,\ell}$. We can complete the proof of the theorem by writing any closed set C as the union $C=[C\cap C^\ell]\cup[C\cap(C^\ell)^c]$. An easy calculation yields

$$\limsup_{n\to\infty}\frac{1}{n}\log R_n(C)\le\max(-\inf_{z\in C^\ell}K(z),-\ell)\le\max(-\inf_{z\in C}K(z),-\ell)$$

and we can let $\ell\to\infty$ to obtain the upper bound for arbitrary closed sets.

Step 4. The lower bound a much simpler. We need to prove only that if $z\in Z$ is arbitrary and N is a neighborhood of z in Z, then

$$\liminf_{n\to\infty}\frac{1}{n}\log R_n(N)\ge -I(z).$$

Since any neighborhood of z contains a product $U\times V$ of neighborhoods U and V the lower bound for R_n follows easily from the lower bounds in the LDP for P_n and Q_n. □

The next result has to do with the behavior of LDP under continuous mapping from one space to another and is referred to as the "Contraction Principle"

Theorem 2.7. *If P_n satisfies an LDP on X with a rate function $I(\cdot)$, and F is a continuous mapping from the Polish spacs X to another Polish space Y, then the family $Q_n=P_nF^{-1}$ satisfies an LDP on Y with a rate function $J(\cdot)$ given by*

$$J(y)=\inf_{x:F(x)=y}I(x).$$

Proof. Let $C\subset Y$ be closed. Then $D=F^{-1}C=\{y:F(x)\in C\}$ is a closed subset of X.

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log Q_n(C)&=\limsup_{n\to\infty}\frac{1}{n}\log P_n(D)\\&\le -\inf_{x\in D}I(x)\\&= -\inf_{y\in C}\inf_{x:F(x)=y}I(x)\\&= -\inf_{y\in C}J(y)\end{aligned}$$

The lower bound is proved just as easily from the definition

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{n}\log Q_n(U) &= \liminf_{n\to\infty} \frac{1}{n}\log P_n(V)\\ &\geq -\inf_{x\in V} I(x)\\ &= -\inf_{y\in U}\inf_{x:F(x)=y} I(x)\\ &= -\inf_{y\in U} J(y)\end{aligned}$$

where $V = F^{-1}U \subset X$ is open corresponding to the open set $U \subset Y$. □

If we want to consider the behavior of integrals of the form

$$a_n = \int_0^1 e^{nF(x)} d\mu_n(x)$$

where $d\mu_n(x) = e^{-nI(x)}dx$ It is clear that the contribution comes mostly from the point where $F(x) - I(x)$ achieves its maximum. In fact under mild conditions it is easy to see that

$$\lim_{n\to\infty} \frac{1}{n}\log a_n = \sup_x [F(x) - I(x)].$$

This essentially remains true under LDP.

Theorem 2.8. *Assume that P_n satisfies an LDP with rate function $I(\cdot)$ on X. Suppose that $F(\cdot)$ is a bounded continuous function on X, then*

$$\lim_{n\to\infty} \frac{1}{n}\log a_n = \sup_x [F(x) - I(x)]$$

where

$$a_n = \int_X e^{nF(x)} dP_n(x).$$

Proof. The basic idea of the proof is that the sum of a finite number of exponentials grows like the largest of them and if they are all nonnegative then there is no chance of any cancellation.

Step 1. Let $\epsilon > 0$ be given, For every $x \in X$, there is a neighborhood N_x such that, $F(x') \leq F(x) + \epsilon$ on N_x and $I(x') \geq I(x) - \epsilon$ on the closure $\bar{N}_x$ of N_x. We have here used the lower semicontinuity of $I(\cdot)$ and the continuity, (in fact only the upper semicontinuity) of $F(\cdot)$. If we denote by

$$a_n(A) = \int_A e^{nF(x)} dP_n(x)$$

we have shown that for any $x \in X$, there is a neighborhood N_x of x, such that,

$$a_n(N_x) \leq e^{n[F(x)+\epsilon]} P_n(N_x) \leq e^{n[F(x)+\epsilon]} P_n(\bar{N}_x)$$

and by LDP

$$\limsup_{n\to\infty} \frac{1}{n} \log a_n(N_x) \le [F(x) - I(x)] + 2\epsilon.$$

Step 2. By a standard compactness argument for any compact $K \subset X$,

$$\limsup_{n\to\infty} \frac{1}{n} \log a_n(K) \le \sup_x [F(x) - I(x)] + 2\epsilon.$$

Step 3. As for the contribution from the complement of a compact set by superexponential tightness for any $\ell < \infty$ there exist a compact set K_ℓ such that

$$P_n(K_\ell^c) \le e^{-n\,\ell}$$

and if F is bounded by a constant M, the contribution from K_ℓ^c can be estimated by

$$\limsup_{n\to\infty} \frac{1}{n} \log a_n(K_\ell^c) \le [M - \ell].$$

Putting the two pieces together we get

$$\limsup_{n\to\infty} \frac{1}{n} \log a_n = \limsup_{n\to\infty} \frac{1}{n} \log a_n(X) \le \max(M - \ell, \sup_x [F(x) - I(x)] + 2\epsilon).$$

If we let $\epsilon \to 0$ and $\ell \to \infty$ we are done.

Step 4. The lower bound is elementary. Let $x \in X$ be arbitrary. By the continuity of $F(\cdot)$, (in fact lower semicontinuity at this time) $F(x') \ge F(x) - \epsilon$ in some neighborhood N_x of x, and

$$a_n = a_n(X) \ge a_n(N_x) \ge e^{n[F(x)-\epsilon]} P_n(N_x).$$

By taking logarithms, dividing by n and letting $n \to \infty$ we get

$$\liminf_{n\to\infty} \frac{1}{n} \log a_n \ge [F(x) - I(x)] - \epsilon.$$

We let $\epsilon \to 0$, and take the supremum over the points of $x \in X$ to get our result. □

Remark 2.9. We have used only the upper semicontinuity of F for the upper bound and the lower semicontinuity of F for the lower bound. It is often necessary to weaken the regularity assumptions on F and this has to be done with care. More on this later when we start applying the theory to specific circumstances.

Remark 2.10. The boundedness of F is only needed to derive the upper bound. The lower bound is purely local. For unbounded F we could surely try our luck with truncation and try to estimate the errors. With some control this can be done.

Finally we end the section by transfering the LDP from P_n to Q_n where Q_n is defined by

$$Q_n(A) = \frac{\int_A e^{nF(x)} dP_n(x)}{\int_X e^{nF(x)} dP_n(x)}$$

for Borel subsets $A \subset X$.

Theorem 2.11. *If P_n satisfies an LDP with rate function $I(\cdot)$ and F is a bounded continuous function on X, then Q_n defined above satisfies an LDP on X as well with the new rate function $J(\cdot)$ given by $J(x) = \sup_{x\in X}[F(x) - I(x)] - [F(x) - I(x)]$.*

Proof. The proof is essentially repeating the arguments of Theorem 4. Let us remark that in the notation of Theorem 4, along the lines of the proof there one can establish

$$\limsup_{n\to\infty} \frac{1}{n} \log a_n(C) \leq \sup_{x\in C}[F(x) - I(x)].$$

Now,

$$\log Q_n(C) = \log a_n(C) - \log a_n$$

and we get

$$\limsup_{n\to\infty} \frac{1}{n} \log Q_n(C) \leq \sup_{x\in C}[F(x) - I(x)] - \sup_{x\in X}[F(x) - I(x)] = -\inf_{x\in C} J(x)$$

This gives us the upper bound and the lower bound is just as easy and is left as an exercise. ☐