Large Deviations for Performance Analysis[1]

The Lunteren Conference, January 1996.

Adam Shwartz

Electrical Engineering
Technion—Israel Institute of Technology
Israel

1 Introduction

In very large, very fast networks, rare events are commonplace. For example, the error rate in modern digital communication networks is of the order of 10^{-9}. That is, one in 10^{-9} transmitted bits is distorted to the extent that it is no longer recognizable. Unquestionably, 10^{-9} is a very small number, so that this is indeed a rare event. However, with transmission rates of 10^9 bits per second, this event will occur, on average, every second!

Moreover, rare events tend to be the critical quantities in the performance of many systems. For example, information is typically transmitted along a communications line, and stored temporarily along the way in buffers. The channel is typically designed so that the average traffic rate is well below capacity. The purpose of the buffer is to allow short peaks in transmission rate without the loss of information. Buffer overflows in such a system mean loss of information, and this then is the limiting factor in the performance of the communications line.

The theory of Large deviations provides a method for computing probabilities of rare events. This is done by providing an expression for the probabilities in terms of a solution to a related variational problem. Hopefully, it would be easier to find the solution to this variational problem, and hence to our original question of the probabilities of rare events. But in fact, we obtain more: as we shall see, from the solution to the variational problem we often obtain a description of the way that the rare event occurred.

The purpose of these lectures is to introduce the novice to the theory of large deviations, as it pertains to performance analysis: that is, to the study of stochastic models arising in computer communications, digital communications systems and computing systems. Some applications will be discussed, in order to illustrate the power and the limitations of this approach.

The view presented here is not the most general. Indeed, there is no "general theory" of large deviations. Rather, the subject is a collection of techniques and approaches, which share only the general questions. Our point of view is reported in full in the book [SW] I wrote with Alan Weiss, of AT&T Bell Laboratories, Murray Hill, NJ. Other points of view will be discussed in §5–6.

The pre-requisites for these lectures is basic probability, and some mathematical maturity. We shall concentrate on intuition and methods, not on detail or proof. We begin with large deviations for scalar random variables, with examples including the Poisson case. We shall then discuss the general formulation of the large deviations problem.

The second lecture will be devoted to the theory of large deviations for the sample paths of a jump-Markov process. We illustrate the idea of large deviations for sample paths through a random-walk example. In the third and final lecture we shall see some examples of applications and of the use

[1] Research supported in part by the fund for the promotion of research at the Technion

of the theory. Time permitting, we shall comment on other large deviations approaches and their application to performance evaluation; these are mentioned in the annotated bibliography.

2 Large Deviations for Random Variables

Let $X_1, X_2, \ldots$ be i.i.d., zero mean random variables, and assume they possess exponential moments, that is, $\mathbb{E}\exp(\theta|X_1|) < \infty$ for all real θ. Then, for $\alpha > 0$

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i > \alpha\right) \to 0 \tag{1}$$

as $n \to \infty$, by the (weak) law of large numbers. But how fast is this convergence? Since for each integer k (such that n/k is an integer)

$$\begin{aligned}
&\mathbb{P}\left(X_1 + X_2 + \ldots + X_n > \alpha \cdot n\right) && (2)\\
&\geq \mathbb{P}\left(X_1 + X_2 + \ldots + X_{n/k} > \alpha \cdot n/k, \ldots, X_{n-n/k} + \ldots + X_n > \alpha \cdot n/k\right) && (3)\\
&= \mathbb{P}\left(X_1 + X_2 + \ldots + X_{n/k} > \alpha \cdot n/k\right)^k, && (4)
\end{aligned}$$

due to independence, we conclude that the convergence rate is at most exponential. Now fix some $\theta > 0$. Then, since the function $e^{\theta y}$ is monotone and positive,

$$\begin{aligned}
\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i > \alpha\right) &= \mathbb{P}\left(\exp\left(\theta\sum_{i=1}^{n} X_i\right) > e^{n\theta\alpha}\right) && (5)\\
&\leq e^{-n\theta\alpha}\mathbb{E}\exp\left(\theta\sum_{i=1}^{n} X_i\right) && (6)\\
&= e^{-n\theta\alpha}\left(\mathbb{E}e^{\theta X_1}\right)^n && (7)
\end{aligned}$$

by Chebychev's inequality and by independence, respectively. So, convergence is indeed at an exponential rate. Since this holds for all positive θ, we can minimize over θ and obtain

$$\log \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i > \alpha\right) \leq -n\sup_{\theta}\left\{\theta\alpha - \log \mathbb{E}e^{\theta X_1}\right\} \tag{8}$$

where the "sup" is over $\theta > 0$. Define the "Rate Function" ℓ as

$$\ell(\alpha) \stackrel{\triangle}{=} \sup_{\theta}\left\{\theta\alpha - \log \mathbb{E}e^{\theta X_1}\right\}, \tag{9}$$

where the "sup" here is over all real θ. By Jensen's inequality,

$$\log \mathbb{E}e^{\theta X_1} \geq \log e^{\mathbb{E}\theta X_1} = \mathbb{E}\theta X_1 = 0 = \log \mathbb{E}e^{0\cdot X_1}, \tag{10}$$

which shows that, since $\alpha > 0$, taking "sup" over $\theta > 0$ or all real θ is equivalent. It turns out that this upper bound is the correct bound: that is,

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i > \alpha\right) \approx e^{-n\ell(\alpha)} \quad \text{for } \alpha > \mathbb{E}X_1. \tag{11}$$

This definition of the rate function ℓ applies to i.i.d. random variables with exponential moments around $\theta = 0$. The function ℓ is positive, convex, and its minimum is 0 at $\mathbb{E}X_1$. To show that the lower bound matches the upper bound, one typically uses a change of measure argument, as follows. Using an exponential change of measure so that under the new probability measure, the mean of X_1 is exactly α, and evaluating just the probability of the sample mean falling near α, we obtain (11). Unlike the upper bound, the lower bound is asymptotic: it holds only as $n \to \infty$.

Now, what is the probability that the sample mean falls into some (open) set? Due to (11), if $b > a > \mathbb{E}X_1$ then

$$\log \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \in (a,b)\right) \approx -n\ell(a) = -n \inf_{\alpha\in(a,b)} \ell(\alpha), \tag{12}$$

since ℓ is increasing for $\alpha > a$. Clearly, this holds not only for intervals, but also to open sets (which are unions of intervals). This leads to the general statement 2.2 of a large deviations theorem.

It was S.R.S. Varadhan who gave large deviations its modern formulation. Let $\mathcal{X}$ be a Polish space and z_n some random variables with values in $\mathcal{X}$. For those who dislike abstract objects, consider $\mathcal{X}$ to be the real n-dimensional Euclidean space.

Definition 2.1 *A real valued function I on $\mathcal{X}$ is called a rate function if I is non-negative and lower semicontinuous, that is, the convergence $y_n \to y$ implies* $\liminf I(y_n) \geq I(y)$. *If in addition the sets $\{x : I(x) \leq \alpha\}$ are compact for all α than I is called a good rate function.*

Definition 2.2 *A sequence z_n satisfies the large deviations principle with rate function I if*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(z_n \in C) \leq -\inf_{x\in C} I(x), \tag{13}$$

for every closed set C, while for every open set G,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(z_n \in C) \geq -\inf_{x\in G} I(x). \tag{14}$$

So, in terms of our previous calculation, if X_i are i.i.d. with exponential moments, then the sample mean $z_n \triangleq n^{-1}\sum_{i=1}^{n} X_i$ satisfies the large deviations principle with rate function ℓ.

Let us now calculate an example of a rate function, using (9). If X_i are Poisson (λ), then

$$\ell(a) = a\left(\log\left(\frac{a}{\lambda}\right) - 1\right) + \lambda \quad \text{if } a \geq 0, \text{ and } \ell(\alpha) = \infty \text{ otherwise.} \tag{15}$$

The situation in the vector case is slightly more complicated, especially conditions and proofs. Let $\langle x, y\rangle$ denote the inner product in $\mathbb{R}^d$. For θ and a in $\mathbb{R}^d$, define

$$M(\theta) \triangleq \mathbb{E}e^{\langle\theta, X_1\rangle} \tag{16}$$

$$\ell(a) \triangleq \sup_\theta \left(\langle\theta, a\rangle - \log M(\theta)\right). \tag{17}$$

Then ℓ is a rate function, and we can write a more explicit expression for the Poisson case, as follows. If $e_j,\ j = 1,\ldots,k$ are vectors in $\mathbb{R}^d$ and $y_{ij},\ j = 1,\ldots k,\ i = 1,2,\ldots$ independent Poisson random variables where y_{ij} has mean λ_j, then the rate function for $z_n \triangleq n^{-1}\sum_{i=1}^{n}\left(\sum_{j=1}^{k} y_{ij}e_j\right)$ is

$$\ell(a) = \sup_\theta \left(\langle\theta, a\rangle - \sum_{j=1}^{k} \lambda_j \left(e^{\langle\theta, e_j\rangle} - 1\right)\right). \tag{18}$$

3 Large Deviations for Processes

Now how did it happen that $\sum_{i=1}^n X_i > n\alpha$?

Claim 1 $\mathbb{P}\left(\sum_{i=1}^{n} X_i > n\alpha\right) \approx \mathbb{P}\left(\sum_{i=1}^{j} X_i > j\alpha,\ \text{all } j \le n\right).$

More precisely, on an exponential scale the probabilities are the same. That is, Large Deviations tells us how the rare event occurred. It occurred through a collusion of all the random variables, rather than through the abnormal behavior of just a few. This is typical of random variables possessing exponential moments, and is not generally true without this existence.

"**Proof.** " Clearly $\ge$ holds. On the other hand, if $\sum_{i=1}^n X_i > n\alpha$ then there is a cyclic permutation for which $\sum_{i=1}^j X_i > j\alpha$, all $j \le n$. Since all n permutations are equally likely,

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i > \alpha\right) \le n \cdot \mathbb{P}\left(\sum_{i=1}^{j} X_i > j\alpha,\ \text{all } j \le n\right). \tag{19}$$

Taking $\frac{1}{n}\log$ on both sides we obtain the $\approx$.

Large Deviations—Jump-Markov processes.

Jump Markov processes are just Poisson processes, at least locally: if e_i, $i = 1, \ldots, k$ are k possible jump directions, then the process jumps in direction e_i with rate $\lambda_i(x)$ (depending on the location x!). This provides a good model of many communications systems. We will be interested in scaled process, indexed by n: the n^{th} processes $z_n(t)$ is a jump process with (local) rates $n \cdot \lambda_i(z)$, for jumps of size e_i/n.

What is the most likely behavior of this processes? By the law of large numbers, we expect the process (for large n) to move in the direction of the average. So, define z_∞ as the solution of

$$\frac{d}{dt} z_\infty(t) = \sum_{i=1}^{k} \lambda_i(z_\infty(t)) e_i. \tag{20}$$

Theorem 3.1 *If* $\log \lambda_i(x)$ *are bounded and Lipschitz continuous, and if* z_∞ *is the unique solution of (20) starting at* x*, then*

$$\mathbb{P}\left(\sup_{0\le t\le T} |z_n(t) - z_\infty(t)| \ge \epsilon \;\middle|\; z_n(0) = x\right) \le C_1 e^{-nC_2(\epsilon)}, \tag{21}$$

where $C_1 > 0$, $C_2 > 0$ *and* C_2 *is linear for* ϵ *large and quadratic for* ϵ *small.*

Now, what is the probability that the process follows a given (deterministic) path $r(t) \ne z_\infty(t)$? Divide time to Δ intervals. By the Markov properly, it suffices to compute, for each m,

$$\mathbb{P}\left(z_n(t) \text{ follows } r(t),\ m\Delta \le t \le (m+1)\Delta \;\middle|\; z_n(m\Delta) \text{ is near } r(m\Delta)\right). \tag{22}$$

To do that, we need $z_n(t)$ to follow the slope $r'(m\Delta) \approx [r((m+1)\Delta) - r(\Delta)]/\Delta$ of r.
Since increments of the process are "locally Poisson," we can represent them as a sum of $n \cdot \Delta$ independent random variables, and the result (18) for random variables implies

$$\mathbb{P}\left(z_n(t) \text{ follows } r(t),\ m\Delta \le t \le (m+1)\Delta \;\middle|\; z_n(m\Delta) \text{ is near } r(m\Delta)\right) \approx e^{-n\cdot\Delta\cdot\ell(r(m\Delta), r'(m\Delta))}. \tag{23}$$

By the Markov property,

$$\begin{aligned}
\mathbb{P}\left(\mathrm{z_n(t)\ follows\ the\ path\ r(t),\ 0 \le t \le T}\right) &\approx \prod_m e^{-n\cdot\Delta\ell(r(m\Delta),r'(m\Delta))} \\
&= \exp\left(-n\sum_m [\ell(r(m\Delta), r'(m\Delta))\cdot\Delta]\right) \\
&\approx \exp\left(-n\int_0^T \ell(r(t), r'(t))\,dt\right).
\end{aligned}$$

So, probabilities of Large Deviations for processes are given by

$$\begin{aligned}
\mathbb{P}\left(\mathrm{z_n(t)\ near\ r(t)}\right) &\approx \exp\left(-nI_0^T(r)\right) &(24)\\
I_0^T(r) &\triangleq \int_0^T \ell(r(t), r'(t))\,dt, &(25)
\end{aligned}$$

and $I_0^T(\cdot)$ is a good rate function. An "event" is a set of possible paths: so

$$\mathbb{P}_{\mathrm{n}}\left(\mathrm{event}\right) \approx \exp\left[-\mathrm{n}\inf_{\mathrm{r\in event}} \mathrm{I_0^T(r)}\right] \tag{26}$$

Moreover, if an event occurs, it does near r^*—the minimizer in (26).

4 Performance Evaluation

How do we use the tools we developed for performance evaluation? The paradigm is roughly

- Write the model, jump rates and jump directions, and determine the scaling,
- Find the most likely behavior of the model using (20),
- Determine the critical rare events in the system,
- Write down the "local rate function" ℓ,
- Compute (analytically or numerically) the value and the minimizer of the variational problem. This gives the probability of the rare event, and the way it happened.
- Justify the steps mathematically, if you can.

Here we sketch two examples for the use of large deviations. The obvious one is

The $M/M/1$ Process

A Poisson arrival stream of rate λ feeds a buffer, while customers are removed from the buffer according to an Exponential (μ) service rate. Therefore, the departure process is Poisson (μ) as long as the buffer is not empty. We assume that the queue is stable: $\lambda < \mu$. So here,

$$\begin{aligned}
e_1 &= +1 \qquad & \lambda_1(x) &= \lambda &(27)\\
e_2 &= -1 \qquad & \lambda_2(x) &= \mu\cdot\mathbf{1}[x>0]. &(28)
\end{aligned}$$

The most likely behavior is a straight line from x to 0, followed by a straight line at level 0.

The rate function is easily calculated, as long as $x > 0$. For $x < 0$ or $x = 0$ but $y < 0$ (so that the processes tries to escape the boundary, which is—no service without customers) we have $\ell(x, y) = \infty$. Denote by z_n the scaled process. According to our formulation, the probability that the buffer fills from 0 to a level $n \cdot a$ before time $n \cdot T$ can be approximated as

$$\mathbb{P}\left(\sup_{0\le t\le T} z_n(t) \ge a \mid z_n(0) = 0\right) \approx \exp\left(-n \cdot \inf\left\{I_0^T(r) : r(0) = 0,\ r(t) = a \text{ for some } t \le T\right\}\right). \tag{29}$$

This is a variational problem that can be solved explicitly. The optimal path is a straight line, and its "speed" and the resulting cost can be obtained in closed form.

The Anick-Mitra-Sondhi model

This process consists of n fluid sources, producing fluid at rate 1 each when they are in the "on" state. The change between "on" and "off" states is as a two-state Markov chain. Sources feed a buffer of size $n \cdot B$, and fluid is removed at rate $n \cdot C$. Assume link capacity exceeds mean demand: that is, $q \triangleq \frac{\lambda}{\lambda+\mu} < C$. Define

$$z_n(t) \triangleq \frac{1}{n} \cdot \text{number of "on" sources} \tag{30}$$

$$b_n(t) \triangleq \frac{1}{n} \cdot \text{buffer contents.} \tag{31}$$

Then z_n is; here $\lambda_1(x) = \lambda \cdot x$ and $\lambda_2(x) = \mu \cdot (1 - x)$. The dynamics of the buffer are

$$\frac{d}{dt} b_n(t) = \begin{cases} z_n(t) - C & \text{if } z_n(t) > C \text{ or } b_n(t) > 0 \\ 0 & \text{if } z_n(t) \le C \text{ and } b_n(t) = 0 \end{cases} \tag{32}$$

Note that time is not scaled here, but that the resulting rates are scaled as they should, since the number of sources increases with our parameter n. Again, we can compute the most likely behavior, which is not a straight line anymore.

The probabilistic question of interest is: how does a buffer fill to level B (overflow)?

Again, we can convert this to a variational problem (although this process does not have smooth rates!). Using the principle of optimality, the problem can be split into two smaller ones:

First make $z_n = C$ optimally (otherwise the buffer cannot start filling).

Then make (area under $z_n(t) - C) = B$ optimally, starting with $z_n(t) = C$.

The first variational problem is (q is the average number of "on" sources):

$$\inf\left\{I_0^T(r) : r(0) = q,\ r(T) = C,\ T > 0\right\}. \tag{33}$$

This is a "free time," standard variational problem. The second one is more complicated:

$$\inf\left\{I_0^T(r) : r(0) = C,\ \int_0^T (r(t) - C)\, dt = B\right\}. \tag{34}$$

We can show that the optimizer is a hyperbolic cosine, and we can obtain explicit asymptotics ($B \to 0$ or $B \to \infty$) of both probabilities and most likely paths. From this we can obtain information about the design of such systems, including control issues: see [SW, Chapter 13].

5 Concluding Remarks

There are many approaches to large deviations, and many results using various approaches to its applications in performance evaluation. A partial list is compiled (and annotated) in the next section. For more on the approach presented here, theory and applications, see [SW].

6 Annotated References

[An] Anantharam, V., "How large delays build up in a $GI/G/1$ queue," *Queueing Systems* **5** pp. 345–368, 1988.

[AMS] Anick, D., D. Mitra, M.M. Sondhi, "Stochastic theory of a data-handling system and multiple sources," *Bell Sys. Tech. J.* **61** pp. 1871–1894, 1982. The origin of the eponymous model of ATM.

[Bu] Bucklew, J.A., *Large Deviations Techniques in Decision, Simulation, and Estimation*, Wiley, New-York, 1990. Interesting applications, mostly to simulations and estimation.

[CKW] Courcoubetis, C., G. Kesidis, A. Ridder, J. Walrand, R. Weber, "Admission control and routing in ATM networks using inferences from measured buffer occupancy," To appear in IEEE Trans. Communications, 1995.

[DZ] Dembo, A., and O. Zeitouni, *Large deviations techniques and applications*, Jones and Bartlett, Boston, 1992. General theory and the main techniques in depth. Applications to information theory.

[DP] Dobrushin, R.L. and E.A. Pechersky, "Large deviations for tandem queuing systems," "Large deviations for random processes with independent increments on infinite intervals," Preprints, 1994.

[DE] Dupuis, P., and R.S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*, J. Wiley, 1995. Treats the discrete time case, using an original approach. This is the most advanced method to deal with discontinuous rates.

[DIS] Dupuis, P., H. Ishii, H.M. Soner, "A viscosity solution approach to the asymptotic analysis of queueing systems," *Ann. Prob.* **18** pp. 226–255, 1990.

EM] Elwalid, A.I., and D. Mitra, "Effective Bandwidth of General Markovian Traffic Sources and Admission Control of High Speed Networks," *IEEE/ACM Trans. Networking* **1** pp. 329–343, 1993.

[FW] Freidlin, M.I., and A.D. Wentzell, *Random Perturbations of Dynamical Systems*, Springer-Verlag, New York, 1984. Mostly diffusions; the origin of exit probabilities via Large Deviations.

[GLM] Gazdzicki, P., I. Lambadaris, R. Mazumdar, "Blocking Probabilities for Large Multirate Erlang Loss Systems," *Adv. Appl. Prob.* **25** pp. 997–1009, 1993.

[Ke] Kelly, F.P., "Effective Bandwidths at Multi-class Queues," *QUESTA* **29** pp. 1474–1481, 1981.

[LO] Lewis, J.T., Neil O'Connell and Raymond Russell, *An Introduction to Large Deviations for Teletraffic Engineers*, Preprint, 1995. This very elementary introduction can be obtained through http://stp01.stp.dias.ie, where you will find other related work.

[SWa] Shwartz, A., and A. Weiss, "Induced rare events: analysis via large deviations and time reversal," *Adv. Appl. Prob.* **25** pp. 667–689, 1993.

[SW] Shwartz, A. and A. Weiss, *Large Deviations for Performance Analysis: queues, communications and computing*, Chapman Hall, 1995.

[Ts] Tsoucas, P., "Rare Events in Series of Queues," *J. Applied Prob.* **29** pp. 168–175, 1992.

[Va] Varadhan, S.R.S., *Large Deviations and Applications*, SIAM, Philadelphia, 1984.

[Ws] Weiss, A., "A new technique for analyzing large traffic systems," *Adv. Appl. Prob.* **18** pp. 506–532, 1986.

[Wn] Wentzell, A.D., *Limit Theorems on Large Deviations for Markov Stochastic Processes*, Kluwer, Dordrecht, Holland, 1990.