Advanced Stochastic Processes.

David Gamarnik

LECTURE 4

Large Deviations theory continued

Outline of Lecture

- Large deviations lower bound.
- Examples.
- Bankruptcy problem.

4.1. Large deviations lower bound

We have established an upper bound on the probability of large deviations

$$\mathbb{P}(\frac{\sum_{1\le i\le n} X_i}{n} > a) \le e^{-l(a)n},$$

where $l(a) = \sup_\theta(a\theta - \log M(\theta)) = a\theta^* - \log M(\theta^*)$ is the Legendre transform corresponding to the distribution of random variable X_1. Perhaps the most surprising result of the theory of large deviations is that this upper bound is tight. The following theorem states that. The course packet contains the proof on pages 15-17. It involves an important and useful technique called *change of measures.*

Theorem 4.1. *Given an i.i.d. sequence* $X_1, \ldots, X_n, \ldots$, *suppose* $\mathbb{E}[e^{\theta X_1}] < \infty$ *for all* θ *in some interval* $[0, \theta_0)$. *Let* $a > \mathbb{E}[X_1]$. *Then for every* $\epsilon > 0$ *there exists a sufficiently large* n_0 *such that for all* $n > n_0$

$$\mathbb{P}(\frac{\sum_{1\le i\le n} X_i}{n} > a) \ge e^{-(l(a)-\epsilon)n},$$

where $l(a) = a\theta^* - \log M(\theta^*)$ *is the Legendre transform corresponding to the distribution of* X_1.

Theorem 3.4 from the previous lecture and Theorem 4.1 can be combined into the following result.

Theorem 4.2. *Given an i.i.d. sequence* $X_1, \ldots, X_n, \ldots$, *suppose* $\mathbb{E}[e^{\theta X_1}] < \infty$ *for* $\theta \in [0, \theta_0)$. *Let* $a > \mathbb{E}[X_1]$. *Then*

$$\lim_{n\to\infty} \frac{\log \mathbb{P}(\frac{\sum_{1\le i\le n} X_i}{n} > a)}{n} = -l(a).$$

Equivalently, we say that the probability of large deviations satisfies $\mathbb{P}(\frac{\sum_{1\le i\le n} X_i}{n} > a) = e^{-l(a)n+o(n)}$ when n is large.

4.2. Examples

Let us go over the examples of some distributions and compute their corresponding Legendre transforms.

- **Exponential distribution with parameter** λ**.** Recall that $M(\theta) = \lambda/(\lambda - \theta)$ when $\theta < \lambda$ and $M(\theta) = \infty$ otherwise. Therefore when $\theta < \lambda$

$$\begin{aligned} l(a) &= \sup_\theta (a\theta - \log \frac{\lambda}{\lambda - \theta}) \\ &= \sup_\theta (a\theta - \log \lambda + \log(\lambda - \theta)). \end{aligned}$$

Setting the derivative of $g(\theta) = a\theta - \log \lambda + \log(\lambda - \theta)$ equal to zero we obtain the equation $a - 1/(\lambda - \theta) = 0$ which has the unique solution $\theta^* = \lambda - 1/a$. Therefore

$$\begin{aligned} l(a) &= a(\lambda - 1/a) - \log \lambda + \log(\lambda - \lambda + 1/a) \\ &= a\lambda - 1 - \log \lambda + \log(1/a) \\ &= a\lambda - 1 - \log \lambda - \log a. \end{aligned}$$

Let us see the meaning of this expression. First note that the function $l(a)$ is indeed convex: For large deviations regime we need to consider $a > 1/\lambda$. When $a = 1/\lambda$, we obtain $l(a) = 0$ giving $e^{-l(a)n} = 1$ - useless bound. However, when a is bigger than $1/\lambda$, $l(a)$ becomes bigger than zero (this involves a tradeoff between increasing the linear function $a\lambda$ and decreasing the function $\log(1/a)$, but one can show that the first one "beats"), and the bound becomes meaningful.

The large deviations bound then tells us that when $a > 1/\lambda$

$$\mathbb{P}(\frac{\sum_{1\le i\le n} X_i}{n} > a) \approx e^{-(a\lambda - 1 - \log\lambda - \log a)n}.$$

Say $\lambda = 1$ and $a = 1.2$. Then the approximation gives us $\approx e^{-(.2-\log 1.2)n}$. On the other hand recall that the process $X_1, X_1 + X_2, \ldots, X_1 + X_2 \cdots + X_n, \ldots$ is a Poisson process with parameter $\lambda = 1$. Therefore we can compute the probability $\mathbb{P}(\sum_{1\le i\le n} X_i > 1.2n)$ exactly: it is the probability that the Poisson process has at most $n - 1$ events before time $1.2n$. Thus

$$\mathbb{P}(\frac{\sum_{1\le i\le n} X_i}{n} > 1.2) = \mathbb{P}(\sum_{1\le i\le n} X_i > 1.2n) = \sum_{0\le k\le n-1} \frac{(1.2n)^k}{k!} e^{-1.2n}.$$

It is not at all clear how revealing this expression is. In hindsight, we know that it is approximately $e^{-(.2-\log 1.2)n}$, obtained via large deviations theory.

- **Standard Normal distribution.** Recall that $M(\theta) = e^{\frac{\theta^2}{2}}$ when X_1 has the standard Normal distribution. The expected value $\mu = 0$. Thus we fix $a > 0$ and obtain

$$l(a) = \sup_\theta (a\theta - \frac{\theta^2}{2}) = \frac{a^2}{2},$$

achieved at $\theta^* = a$. Again we see that $l(a)$ is (as it should be) a convex function of a. Thus for $a > 0$, the large deviations theory predicts that

$$\mathbb{P}(\frac{\sum_{1\le i\le n} X_i}{n} > a) \approx e^{-\frac{a^2}{2}n}.$$

Again we could compute this probability directly. We know that $\frac{\sum_{1\le i\le n} X_i}{n}$ is distributed as a Normal random variable with mean zero and variance $1/n$. Thus

$$\mathbb{P}(\frac{\sum_{1\le i\le n} X_i}{n} > a) = \frac{\sqrt{n}}{\sqrt{2\pi}} \int_a^\infty e^{-\frac{t^2 n}{2}} dt.$$

After a little bit of technical work one could show that this integral is "dominated" by its part around a, namely, $\int_a^{a+\epsilon} \cdot$, which is further approximated by the value of the function itself at a, namely $\frac{\sqrt{n}}{\sqrt{2\pi}} e^{-\frac{a^2}{2}n}$. This is consistent with the value given by the large deviations theory. Simply the lower order magnitude term $\frac{\sqrt{n}}{\sqrt{2\pi}}$ disappears in the approximation on the log scale.

- **Poisson distribution.** Suppose X has a Poisson distribution with parameter λ. Recall that in this case $M(\theta) = e^{e^\theta \lambda - \lambda}$. Then

$$l(a) = \sup_\theta (a\theta - (e^\theta \lambda - \lambda)).$$

Setting derivative to zero we obtain $\theta^* = \log(a/\lambda)$ and $l(a) = a\log(a/\lambda) - (a - \lambda)$.

In this case as well we can compute the large deviations probability explicitly. The sum $X_1 + \cdots + X_n$ of Poisson random variables is also a Poisson random variable with parameter λn. Therefore

$$\mathbb{P}(\sum_{1\le i\le n} X_i > an) = \sum_{m > an} \frac{(\lambda n)^m}{m!} e^{-\lambda n}.$$

But again it is hard to infer a more explicit rate of decay using this expression

4.3. Applications

4.3.1. Insurance bankruptcy problem

Imagine an insurance company which has capital x at time $t = 0$. During periods $t = 1, 2, \ldots$ it receives revenues $X_t \geq 0$ from premiums and pays claims in the amount of $Y_t \geq 0$. The wealth of the company at time t is then $S_t \triangleq x + \sum_{1\leq i\leq t} X_i - \sum_{1\leq i\leq t} Y_t$. The company is declared bankrupt if at some point $S_t \leq 0$ (no borrowing is allowed). Assume $X_1, \ldots, X_t, \ldots$ is i.i.d. and $Y_1, \ldots, Y_t, \ldots$ is also i.i.d. with expected values $\mathbb{E}[X_1], \mathbb{E}[Y_1]$, respectively and also that the two processes are independent of each other. How likely is it that the bankruptcy occurs? Namely, what is $\mathbb{P}(\inf_{t\geq 0} S_t < 0)$?

This problem is also known as Gambler's ruin problem. It can be solved exactly when X_t, Y_t take values $0, 1$. In more general case solution becomes quite messy. Here we use large deviations type bounds to obtain simple bounds on the bankruptcy probability. But first we rule out some cases.

Suppose $\mathbb{E}[X_1] < \mathbb{E}[Y_1]$. The SLLN says that $\frac{\sum_{1\leq i\leq t} X_i - Y_i}{t}$ converges almost surely to $\mathbb{E}[X_1] - \mathbb{E}[Y_1]$. Therefore, almost surely, the sum $\sum_{1\leq i\leq t} X_i - Y_i$ becomes smaller than any fixed value including starting capital x. Therefore $\inf_{t\geq 0} S_t = -\infty$ almost surely. We conclude that the bankruptcy probability is $\mathbb{P}(\inf_{t\geq 0} S_t < 0) = 1$

The case $\mathbb{E}[X_1] = \mathbb{E}[Y_1]$ will be analyzed later when we study Brownian motion.

Now we focus on the case of interest: $\mathbb{E}[X_1] > \mathbb{E}[Y_1]$. Our goal is to show that, provided the corresponding exponential moment generating functions $\mathbb{E}[e^{\theta X_1}], \mathbb{E}[e^{\theta Y_1}]$ are finite in some common interval $[0, \theta_0)$, the bankruptcy probability is order e^{-cx} for some constant $c > 0$. Namely, it is exponentially unlikely as a function of the starting capital x.

For simplicity introduce $Z_t = Y_t - X_t$. Then $Z_t, t \geq 0$ is an i.i.d. sequence with $\mu \triangleq \mathbb{E}[Z_1] < 0$. Note that whenever $\theta \in [0, \theta_0)$, we have $M(\theta) \triangleq \mathbb{E}[e^{\theta Z_1}] = \mathbb{E}[e^{\theta Y_1}]\mathbb{E}[e^{-\theta X_1}] \leq \mathbb{E}[e^{\theta Y_1}] < \infty$. Here we use the fact $e^{-\theta X_1} \leq 1$. Now we fix t and consider

$$\begin{aligned}\mathbb{P}(S_t < 0) &= \mathbb{P}(x - \sum_{1\leq i\leq t} Z_i < 0) \\ &= \mathbb{P}(\sum_{1\leq i\leq t} Z_i > x).\end{aligned}$$

We let $a_t = x/t$. We have $a_t > 0 > \mu \; \forall \, t$. Then the Chernoff bound (Theorem 4.2) tells us that this probability is

$$\leq e^{-l(a_t)t},$$

where $l(a_t)$ is the Legendre transform $l(a_t) = \sup_\theta(a_t\theta - \log M(\theta))$. Now since $\mu < 0$, then by Proposition 2 from Lecture 3, there exists $\theta^* \in [0, \theta_0)$ such that $l(0) = 0\theta^* - \log M(\theta^*) = -\log M(\theta^*) > 0$. Using this value $\theta = \theta^*$ we obtain $l(a_t) \geq a_t\theta^* - \log M(\theta^*) \; \forall \, t$ implying for all t,

$$\mathbb{P}(S_t < 0) \leq e^{-a_t\theta^* t + \log M(\theta^*)t} = e^{-\theta^* x + \log M(\theta^*)t}.$$

Therefore

$$\begin{aligned}
\mathbb{P}(\inf_{t\geq 0} S_t < 0) &\leq \sum_{t\geq 0} \mathbb{P}(S_t < 0) \\
&\leq \sum_{t\geq 0} e^{-\theta^* x + \log M(\theta^*) t} \\
&= \sum_{t\geq 0} e^{-\theta^* x} M(\theta^*)^t \\
&= \frac{e^{-\theta^* x}}{1 - M(\theta^*)},
\end{aligned}$$

(which probability law are we using in the first inequality?) where the finiteness of the ratio is guaranteed since $\log M(\theta^*) < 0$. We see that indeed the bankruptcy probability decays at least with the rate $e^{-\theta^* x}$. This is not necessarily the optimum rate as we did not optimize over θ (did we?) and moreover we did not prove a matching lower bound. Usually, in applications as in this one, the upper bound is more important.

4.4. Additional reading materials

- Course packet Sections 1.2-1.5 .
- Dembo and Zeitouni [**1**] Section 2.1 and 2.2 . The large deviations theorems in our lecture notes are called Cramer's bounds in this book. The first section discusses the method of types and gives a information-theoretic take on large deviations theory.

BIBLIOGRAPHY

1. A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, Springer, 1998.