§ 1.8. Large deviation and some exponential inequalities.

Theory of large deviation (Varadhan, 1984), concerning chance of rare events that are usually of exponential decay, constitutes a major development in probability theory in the past few decades. Its original idea may be traced back to the Laplace principle in mathematics: for any Borel set $B \in R^d$ and measurable function $g(\cdot)$,

$$\lim_{t\to\infty} \frac{1}{t} \log \int_B e^{-tg(x)}dx = -\mathrm{essinf}_{x\in B}\, g(x), \qquad \lim_{t\to\infty} \frac{1}{t} \log \int_B e^{tg(x)}dx = \mathrm{esssup}_{x\in B}\, g(x),$$

should the integral be finite. The classical form of large deviation in terms of iid random variables is due to Harald Cramer(1938). We present a brief account in the simplest form.

(i). *Large deviation for iid r.v.s.*

EXAMPLE 1.12 (VALUE-AT-RISK) Suppose a portfolio worths $W_0 = 1$ million dollar at inception. Assume the returns of the i-th trading period are X_i, which are iid. Then the portfolio worths $W_n = \prod_{i=1}^n X_i$ at the end of n-th trading period. The so-called value-at-risk, VaR, as a measurement of risk of the portfolio is defined as follows: the n trading period p-percentage VaR of the portfolio is $c_n > 0$ such that

$$P(W_0 - W_n > c_n) = P(W_n < 1 - c_n) = p.$$

In other words, c_n is the amount that the portfolio may lose as much as or more with chance p. In financial industry, p is commonly set to be small, as for example 5% or 1%.

Consider a standard setup with X_i being log-normal, i.e., $\log X_i \sim N(\mu, \sigma^2)$. The critical fact that we shall use in this example is, for any $x < \mu$ such that $n(x-\mu)^2$ is large,

$$\frac{1}{n} \log P\Big(\frac{\sum_{i=1}^n \log(X_i)}{n} < x\Big) = \frac{1}{n} \log P(N(\mu, \sigma^2/n) < x) = \frac{1}{n} \log \Phi(\sqrt{n}(x-\mu)/\sigma) \approx -\frac{(x-\mu)^2}{2\sigma^2},$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the cdf and density of $N(0,1)$, since $\Phi(s) \approx \phi(s)/|s|$ for $s \to -\infty$. Then, for small p, we have

$$\begin{aligned} & \frac{1}{n}\log p = \frac{1}{n}\log P(W_n < 1-c_n) = \frac{1}{n}\log P(\log W_n < \log(1-c_n)) \\ = \;& \frac{1}{n}\log P\Big(\frac{\sum_{i=1}^n \log(X_i)}{n} < \frac{\log(1-c_n)}{n}\Big) \\ \approx \;& -\frac{[(1/n)\log(1-c_n) - \mu]^2}{2\sigma^2} \end{aligned}$$

Suppose $(1/n)\log(1-c_n) \approx a < \min(0, \mu)$ with $n(a-\mu)^2$ large. Then $p \approx e^{-nq}$ where

$$q = \frac{(a-\mu)^2}{2\sigma^2} \qquad \text{and} \qquad a = \mu - \sigma\sqrt{2q} < 0.$$

In other words, the portfolio may shrink at an average compound rate of $|a|$ for n periods with chance $p \approx e^{-nq}$. For example, suppose $\mu = 0$, $\sigma = 1$, $q = 1/2$, $n = 6$. Then, $a = -1$, the portfolio may shrink to e^{-6} with approximate chance e^{-3}.

We note that $(x-\mu)^2/2\sigma^2$ is the so-called rate function, or Cramer or entropy function, for $N(\mu, \sigma^2)$. The above calculation takes advantage of the log-normality assumption. With general population distribution of X_i, the limiting relation between c_n and p is answered by the theorem of large deviation. □

EXAMPLE 1.13 (CRAMER'S ACTUARIAL PROBLEM) Suppose n clients have each paid a premium of c dollars for life insurance over a period of time. Assume the claims are iid nonnegative random variables $X_1, ..., X_n$. Suppose the total premium nc is all the insurance company has to pay out the claims. The chance that the insurance company bankrupts is

$$P(\sum_{i=1}^n X_i > cn) = P(S_n/n > c) = P(S_n/n - \mu > c - \mu),$$

where μ is the common mean of X_i. By weak law of large numbers, the chance is close to 1 if $c < \mu$. Normally, the insurance company sets the premium $c > \mu$. The chance of bankruptcy is close to 0. Since bankruptcy is life-and-death issue for the company, it is critical to have a precise estimation of the chance. By the following Cramer's theorem (Theorem 1.9), under suitable conditions,

$$P(\sum_{i=1}^n X_i > cn) = P(S_n/n > c) \approx e^{-nI(c)} \qquad \text{for } c > \mu \text{ and large } n$$

where $I(x) = \sup_t [xt - \log \varphi(t)]$ and φ is the moment generating function of X_i defined as follows. $\square$

Definition For a r.v. X (or its distribution function), its *moment generation function* is $\varphi(t) = E(e^{tX})$, $t \in (-\infty, \infty)$.

Note that $\varphi(0) = 1$ for any r.v. or distribution but the moment generating functions are not necessarily finite everywhere. Should $\varphi(\cdot)$ be finite in a neighborhood of 0, the k-th derivative of φ at 0 is the k-th moment of X_i, i.e.,

$$\varphi^{(k)}(0) = E(X^k),$$

This explains why it is called moment generating function.

The following are some moment generating functions for commonly used distributions:

Binomial $B(n,p)$: $(1-p+pe^t)^n$; Normal $N(\mu, \sigma^2)$: $\varphi(t) = e^{t\mu + \sigma^2 t^2/2}$;
Poisson $\mathcal{P}(\lambda)$: $\varphi(t) = e^{-\lambda + \lambda e^t}$; Exponential $\mathcal{E}(\lambda)$: $\varphi(t) = \lambda/(\lambda - t)$ for $t < \lambda$.

Lemma Suppose a r.v. X has finite moment generating function $\varphi(\cdot)$ on $(-\infty, \infty)$. Then the rate function

$$I(x) = \sup_t (tx - \log \varphi(t)), \qquad x \in (-\infty, \infty)$$

is a convex function with minimum 0 at $x = E(X)$.

We omit the proof. The essential part is that $tx - \log \varphi(t)$ is concave in t, since $\log \varphi(\cdot)$ is convex.

Theorem 1.9 (Cramer's Theorem) Suppose $X, X_1, X_2, ...$ are iid with mean μ and finite moment generating function $\varphi(\cdot)$ on $(-\infty, \infty)$. Then,

$$\frac{1}{n} \log P(S_n/n > x) \to -I(x) \quad \text{for } x > \mu; \qquad \text{and} \qquad \frac{1}{n} \log P(S_n/n < x) \to -I(x) \quad \text{for } x < \mu.$$

Proof The proof uses the moment generating function and Chebyshev/Markov inequality. First, for $x > \mu$ and $t > 0$,

$$P(S_n/n > x) = P(S_n > nx) \le e^{-nx} E(e^{tS_n}) = e^{-ntx} \varphi(t)^n = e^{-n(tx - \log \varphi(t))}$$

Therefore,

$$\frac{1}{n} \log P(S_n/n > x) \le -\sup_{t \ge 0} (tx - \log \varphi(t)) = -\sup_{t \in (-\infty, \infty)} (tx - \log \varphi(t)), \qquad \text{for } x > \mu.$$

Let F be the common distribution function of X_i. For a fixed $t > 0$, let X_i^* be iid with common cdf

$$F^*(s) = \frac{1}{\varphi(t)} \int_{-\infty}^s e^{-ta} dF(a)$$

Then, for any $y > x$ and $t > 0$,

$$\begin{aligned}
P(S_n/n > x) &\geq P(ny > S_n > nx) \geq e^{-nty} E(e^{tS_n} 1_{\{ny>S_n>nx\}}) \\
&= e^{-nty} \int \cdots \int 1_{\{ny>\sum_{i=1}^n x_i>nx\}} e^{tx_1} dF(x_1) \cdots e^{tx_n} dF(x_n) \\
&= e^{-nty} \varphi(t)^n \int \cdots \int 1_{\{ny>\sum_{i=1}^n x_i>nx\}} \frac{e^{tx_1}}{\varphi(t)} dF(x_1) \cdots \frac{e^{tx_n}}{\varphi(t)} dF(x_n) \\
&= e^{-n(ty-\log\varphi(t))} \int \cdots \int 1_{\{ny>\sum_{i=1}^n x_i>nx\}} dF^*(x_1) \cdots dF^*(x_n) \\
&= e^{-n(ty-\log\varphi(t))} P(y > \sum_{i=1}^n X_i^*/n > x)
\end{aligned}$$

Choose t such that $E(X_i^*) \in (x, y)$. Then, by WLLN,

$$\liminf_n \frac{1}{n} \log P(S_n/n > x) \geq -(ty - \log\varphi(t)).$$

Choose t such that $y \downarrow x$, then $t \to t_0$ where t_0 is such that $t_0 x - \log\varphi(t_0) = \sup_t (tx - \log\varphi(t))$. As a result,

$$\liminf_n \frac{1}{n} \log P(S_n/n > x) \geq -(t_0 x - \log\varphi(t_0)) = -\sup_t (tx - \log\varphi(t)) = -I(x).$$

The first inequality of this theorem is proved. The other inequality for $x < \mu$ can be shown analogously. $\square$

(ii) *Some exponential inequalities.*

The above large deviation results require a common distribution of the r.v.s. If the r.v.s are independent but not necessarily identically distributed, generalization of large deviation is not that simple. However, some exponential inequalities, which can be relatively easily derived and be readily generalized to U-statistics or martingales, are often useful.

Theorem 1.10 (Bernstein's inequality) Suppose $X_n, n \geq 1$, are independent with mean 0 and variance σ_n^2, satisfying

$$E(|X_n|^k) \leq \frac{k!}{2} \sigma_n^2 c^{k-2}, \qquad k \geq 2$$

for some constant $c > 0$. Then, for all $x > 0$,

$$P(S_n/n > x) \leq e^{-nx^2/[2(s_n^2/n+cx)]},$$

where $s_n^2 = \sum_{j=1}^n \sigma_j^2$.

Proof. The proof again uses the moment generating function. Write, for $|t| < 1/c$,

$$\begin{aligned}
E(e^{tX_n}) &\leq 1 + E(tX_n) + \sum_{j=2}^\infty E(|tX_n|^j)/j! \\
&\leq 1 + \frac{t^2\sigma_n^2}{2}(1 + |t|c + t^2c^2 + |t|^3c^3 + \cdots) = 1 + \frac{t^2\sigma_n^2}{2}\frac{1}{1-|t|c} \\
&\leq e^{t^2\sigma_n^2/(2-2c|t|)}.
\end{aligned}$$

Apply Chebyshev's inequality, for $0 < t < 1/c$,

$$P(S_n/n > x) = P(e^{tS_n} > e^{tnx}) \leq e^{-tnx} E(e^{tS_n}) \leq e^{-tnx} \prod_{i=1}^n E(e^{tX_i})$$

$$\leq \; e^{-tnx} e^{\sum_{i=1}^n t^2\sigma_i^2/(2-2ct)} = e^{-tnx+t^2s_n^2/(2-2ct)}$$

Choose $t = nx/(s_n^2 + cnx)$. Bernstein's inequality follows. $\square$

REMARK. A little sharper inequality, called Bennett's inequality, can be obtained by choosing t in the above proof to minimize $-tnx + t^2 s_n^2/(2-2ct)$:

$$P(S_n/n > x) \leq e^{-nx^2/[s_n^2/n(1+\sqrt{1+2cx/s_n^2})+cx]}.$$

Corollary Suppose $X_n, n \geq 1$, are independent with mean 0 and $P(|X_n| \leq c) = 1$ for $c > 0$ and all $n \geq 1$. Then, for $0 < x < c$,

$$P(S_n/n > x) \leq e^{-nx^2/4c^2} \geq P(S_n/n < -x).$$

The proof of this corollary is straightfoward, by observing that $s_n^2/n \leq c$.

An important implication of the above corollary is that for uniformly bounded random variables X_i with mean 0,

$$\limsup_n \frac{|S_n/n|}{\sqrt{(\log n)/n}} < \infty \qquad a.s.$$

This can be proved by citing Borel-Cantelli lemma and choosing, for large n, $x = C(\log n)^\delta/n$ for a large C in the inequality in the above corollary. Notice that the same convergence was also shown in Section 1.7 for iid r.v.s.

The above inequality is an essential building block in a technique, called *empirical approximation*, to prove uniform convergence of random functions. We illustrate it with the following example.

EXAMPLE 1.14 Let $X_1, ..., X_n, ...$ be iid with common cdf $F(\cdot)$ and empirical distribution $F_n(\cdot)$, i.e., $F_n(t) = \sum_{i=1}^n 1_{\{X_i \leq t\}}/n$. Then,

$$\limsup_n \sup_t \frac{|F_n(t) - F(t)|}{\sqrt{(\log n)/n}} \leq 4 \qquad a.s..$$

Proof Without loss of generality, assume $F(\cdot)$ is continuous. The above corollary implies, for all t and $n \geq 1$,

$$P(|F_n(t) - F(t)| > 4\sqrt{(\log n)/n}) \leq 2e^{-4\log n} = 2n^{-4}.$$

Let $t_0 < t_1 < ... < t_{n^2}$ be such that $F(t_k) - F(t_{k-1}) = n^{-2}$. Then,

$$\begin{aligned} & \sum_{n=1}^\infty P(\sup_{1\leq j\leq n^2} |F_n(t_j) - F(t_j)| > 4\sqrt{(\log n)/n}) \\ \leq \quad & \sum_{n=1}^\infty \sum_{j=1}^{n^2} P(|F_n(t_j) - F(t_j)| > 4\sqrt{(\log n)/n}) \leq \sum_{n=1}^\infty 2n^2 n^{-4} < \infty. \end{aligned}$$

By Borel-Cantelli lemma,

$$\limsup_n \frac{\sup_{1\leq j\leq n^2} |F_n(t_j) - F(t_j)|}{4\sqrt{(\log n)/n}} \leq 1 \qquad a.s.$$

It then follows from the monotonicity of $F_n(\cdot)$ and $F(\cdot)$ and the fact $F(t_k) - F(t_{k-1}) = n^{-2}$ that

$$\limsup_n \frac{\sup_t |F_n(t) - F(t)|}{\sqrt{(\log n)/n}} \leq 4 \qquad a.s.$$

$\square$

REMARK. The actual convergence rate of this example is still the law of iterated logarithm.