The Mathematical Theory of Moment Generating Functions

Xi Geng

Abstract

In elementary probability theory, we use the moment generating function to compute moments, identify distributions, study convergence in distribution etc. Over there we emphasise the application of methods and do not pursue full mathematical rigor. In advanced probability theory, we work with characteristic functions and develop the entire theory relating the characteristic function with the study of distributional properties. The aim of these notes is to develop the bridge connecting these two approaches in a precise mathematical way. The approach we adopt here relies on complex analysis. We recall the necessary tools in the appendix. We also assume familiarity with characteristic functions.

Contents

1 The moment generating function and its basic properties

The moment generating function is mainly used to study distributional properties of a random variable. It is defined as follows.

Definition 1.1. Let X be a random variable. The *moment generating function* (*m.g.f.*) of X is defined by

$$M_X(t) \triangleq \mathbb{E}[e^{tX}], \quad t \in T,$$

where T is the domain of its definition consisting of those t's at which $\mathbb{E}[e^{tX}] < \infty$.

From the definition, the m.g.f. is always positive. It is clear that $0 \in T$ and $M_X(0) = 1$. Moreover, if $t > 0$ is in T, then $[0, t] \subseteq T$. Indeed, for any $s \in [0, t]$, we have

$$\begin{aligned} \mathbb{E}[e^{sX}] &= \mathbb{E}[e^{sX}\mathbf{1}_{\{X\geqslant 0\}}] + \mathbb{E}[e^{sX}\mathbf{1}_{\{X<0\}}] \\ &\leqslant \mathbb{E}[e^{tX}\mathbf{1}_{\{X\geqslant 0\}}] + \mathbb{P}(X < 0) \\ &\leqslant M_X(t) + 1. \end{aligned}$$

Similarly, if $t < 0$ is in T, then $[t, 0] \subseteq T$. This shows that the domain of definition for an m.g.f. is always an interval. This interval could be degenerate (i.e. $T = \{0\}$), finite or infinite, but in general there is no implication on the openness/closedness at the endpoints. A simple example where the m.g.f. is defined only at $t = 0$ is the Cauchy distribution. It is a good exercise to provide one example for each possible case of T. We give one such example where $T = (-1, 1]$.

Example 1.1. Let X be a random variable with p.d.f.

$$f_X(x) = Ce^{-|x|}\left(|x|^{-2}\mathbf{1}_{\{x>1\}} + \mathbf{1}_{\{x<0\}}\right),$$

where C is a normalising constant so that $\int_{\mathbb{R}} f_X(x)dx = 1$. The m.g.f. of X is

$$M_X(t) = C \cdot \left(\int_{-\infty}^{0} e^{(t+1)x}dx + \int_1^\infty x^{-2}e^{(t-1)x}dx\right).$$

It is clear that $M_X(t) < \infty$ if and only if $t \in (-1, 1]$.

The first application of m.g.f. is the computation of moments, which also justifies its name.

Theorem 1.1. *Suppose that $M_X(t)$ is well defined defined on a neighbourhood of the origin, say $(-\delta, \delta)$. Then X has finite absolute moments of all orders. Moreover, $M_X(t)$ admits an absolutely convergent Taylor expansion*

$$M_X(t) = \sum_{n=0}^{\infty} \frac{\mathbb{E}[X^n]}{n!} t^n, \quad t \in (-\delta, \delta). \tag{1.1}$$

In particular, $\mathbb{E}[X^n] = M_X^{(n)}(0)$ for all $n \geqslant 0$.

Proof. Let $t \in (0, \delta)$. From the simple inequality

$$e^{t|x|} \leqslant e^{tx} + e^{-tx}$$

and the fact that $\pm t \in T$, we have

$$\mathbb{E}[e^{t|X|}] \leqslant M_X(t) + M_X(-t) < \infty.$$

This shows that

$$\sum_{n=0}^{\infty} \frac{\mathbb{E}[|X|^n]}{n!} t^n = \mathbb{E}[e^{t|X|}] < \infty,$$

and in particular $\mathbb{E}[|X|^n] < \infty$ for all n. In addition, the series $\sum_{n=0}^{\infty} \frac{\mathbb{E}[X^n]}{n!} s^n$ is absolutely convergent on $[-t, t]$. If we write

$$S_n \triangleq \sum_{k=0}^{n} \frac{X^k}{k!} t^k,$$

then

$$|M_X(t) - \mathbb{E}[S_n]| \leqslant \sum_{k=n+1}^{\infty} \frac{\mathbb{E}[|X|^k]}{k!} t^k \to 0, \text{ as } n \to \infty.$$

Therefore, $M_X(t)$ admits the expansion (1.1). The last claim follows from general properties of power series. $\square$

Remark 1.1. Using essentially the same argument, one can show the following property: if t is an interior point of T, then

$$\mathbb{E}[|X|^k e^{tX}] < \infty$$

for any $k \in \mathbb{N}$. In fact, if we choose $\eta > 0$ such that $(t - \eta, t + \eta) \subseteq T$, then we have

$$|X|^k e^{tX} \leqslant e^{\eta|X|} e^{tX} \leqslant e^{(t+\eta)X} + e^{(t-\eta)X}$$

when $|X|$ is greater than some constant R depending on k and η. Note that by the choice of η the right hand side of the above inequality is integrable.

The analyticity property at $t = 0$ given by Theorem 1.1 can be extended to the more general complex setting below. This extension will allow us to complexify the m.g.f. and hence relate our study to well known properties of the characteristic function.

Lemma 1.1. *Suppose that $M_X(t)$ is well defined on some open interval $I = (a, b)$. Then the complex function*

$$\Phi_X(z) \triangleq \mathbb{E}[e^{zX}]$$

is well defined and holomorphic in the strip

$$S_I \triangleq \{z = t + is : t \in I, s \in \mathbb{R}\}$$

over the complex plane.

Proof. The well definedness of $\Phi(z)$ is obvious:

$$|\Phi_X(z)| = |\mathbb{E}[e^{tX+isX}]| \leqslant \mathbb{E}[e^{tX} \cdot |e^{isX}|] = M_X(t)$$

for $z = t + is \in S_I$. Now for given $z_0 = t_0 + is_0 \in S_I$, we show that $\Phi_X(z)$ is differentiable at z_0 with

$$\Phi_X'(z_0) = \mathbb{E}[Xe^{z_0X}].$$

To this end, note that

$$\begin{aligned}
&\frac{\Phi_X(z) - \Phi_X(z_0)}{z - z_0} - \mathbb{E}[Xe^{z_0X}] \\
&= \mathbb{E}\left[e^{z_0X}\left(\frac{e^{(z-z_0)X} - 1 - (z - z_0)X}{z - z_0}\right)\right].
\end{aligned}$$

When $|z - z_0| < \eta$, we have

$$\begin{aligned}
&\left|e^{z_0X}\left(\frac{e^{(z-z_0)X} - 1 - (z - z_0)X}{z - z_0}\right)\right| \\
&= \left|e^{z_0X} \cdot \sum_{n=2}^{\infty} \frac{(z - z_0)^{n-1}X^n}{n!}\right| \\
&\leqslant e^{t_0X} \cdot |X| \cdot \sum_{n=2}^{\infty} \frac{|\eta X|^{n-1}}{(n-1)!} \\
&\leqslant |X| \cdot e^{t_0X} \cdot e^{\eta|X|} \\
&\leqslant |X| \cdot \left(e^{(t_0+\eta)X} + e^{(t_0-\eta)X}\right).
\end{aligned}$$

If we choose η so that $t_0 \pm \eta \in T$, from Remark 1.1 we see that the right hand side of the above inequality is integrable. According to the dominated convergence theorem,

$$\lim_{z \to z_0} \frac{\Phi_X(z) - \Phi_X(z_0)}{z - z_0} = \mathbb{E}[Xe^{z_0 X}].$$

Since $z_0 \in S_I$ is arbitrary, we conclude that $\Phi_X(z)$ is holomorphic in S_I. $\square$

Remark 1.2. In the strip S_I, the function $\Phi_X(z)$ is infinitely differentiable and we can differentiate inside the expectation:

$$\Phi_X^{(n)}(z) = \frac{d^n}{dz^n}\mathbb{E}[e^{zX}] = \mathbb{E}[X^n e^{zX}].$$

When restricted back to I, this is also true for the m.g.f. $M_X(t)$.

Another useful property of the m.g.f. is the following so-called *convolution theorem*.

Theorem 1.2. *Let $X_1, \cdots, X_n$ be independent random variables whose m.g.f.'s are well defined on some interval I. Then the m.g.f. of $S_n \triangleq X_1 + \cdots + X_n$ is well defined on I and it is given by*

$$M_{S_n}(t) = M_{X_1}(t) \cdots M_{X_n}(t)$$

for $t \in I$.

Proof. By definition and independence,

$$M_{S_n}(t) = \mathbb{E}[e^{t(X_1 + \cdots + X_n)}] = \mathbb{E}[e^{tX_1}] \cdots \mathbb{E}[e^{tX_n}] = M_{X_1}(t) \cdots M_{X_n}(t).$$

$\square$

Example 1.2. Let $X_1, \cdots, X_n$ be i.i.d. standard normal random variables. The m.g.f. of X_i^2 is given by

$$\begin{aligned} M_{X_i^2}(t) &= \int_{-\infty}^{\infty} e^{tx^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\cdot(1-2t)^{-1}}} dx \\ &= (1-2t)^{-1/2}, \end{aligned}$$

provided $t < 1/2$. From the convolution theorem,

$$M_{X_1^2+\cdots+X_n^2}(t) = (1-2t)^{-n/2}, \quad t < \frac{1}{2}.$$

This is precisely the m.g.f. of the Gamma distribution with parameters $(n/2, 1/2)$. From the uniqueness theorem which we will prove soon (cf. Theorem 2.1 in Section 2), we conclude that

$$X_1^2 + \cdots + X_n^2 \stackrel{d}{=} \gamma(n/2, 1/2).$$

2 The uniqueness theorem

One of the most important results about m.g.f. is the fact that it uniquely determines the distribution. This is the content of the following result.

Theorem 2.1. *Suppose that X and Y are two random variables whose m.g.f.'s are well defined and equal in some neighbourhood of $t = 0$. Then $X \stackrel{d}{=} Y$.*

Proof. Suppose that the m.g.f.'s of X, Y are well defined and equal in $(-\delta, \delta)$ for some $\delta > 0$. Define the complex function

$$\Phi_X(z) \triangleq \mathbb{E}[e^{zX}]$$

for $z \in S \triangleq (-\delta, \delta) \times i\mathbb{R}$ over the complex plane, and similarly for Y. According to Lemma 1.1, we know that $\Phi_X(z)$ and and $\Phi_Y(z)$ are both holomorphic in S. Since they coincide on the real interval $(-\delta, \delta) \times \{0\}$, by the Identity Theorem (cf. Theorem A.2 in the appendix) we know that $\Phi_X = \Phi_Y$ in S. In particular, their restrictions on the imaginary axis, which gives the corresponding characteristic functions, are equal. It follows from the uniqueness theorem for characteristic functions that X and Y must be equal in distribution. $\square$

From the Taylor expansion (1.1) we see that the m.g.f. of X is uniquely determined by its moments, provided the m.g.f. is well defined in some neighbourhood of the origin. Together with the uniqueness theorem, we know that in this case the moments of X uniquely determines its distribution.

However, in general moments can still exist even if the m.g.f. is not well defined or is only defined on one side of the origin. It is interesting to remark that, in such a situation the moments may fail to determine the distribution. In what follows, we give two examples of a family of distinct distributions which share the same moments of all orders. Both of them are quite enlightening. The general principle of constructing such examples, as a necessary condition, is to make the p.d.f. decay slower than exponential to fail the existence of its m.g.f.

Example 2.1. Let $X = e^Z$ be a standard log-normal random variable whose p.d.f. is given by

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi}x} e^{-(\ln x)^2/2}, & x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Note that the moment $\mathbb{E}[X^n]$ exists for all n, while the m.g.f. of X is defined only for $t \leqslant 0$. Consider the family of p.d.f.'s $\{f_\varepsilon : \varepsilon \in [-1, 1]\}$ defined by

$$f_\varepsilon(x) \triangleq \begin{cases} f(x)\left(1 + \varepsilon \sin(2\pi \ln x)\right), & x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Apparently $f_\varepsilon \geqslant 0$. We now show that

$$\int_0^\infty x^n f(x) \sin(2\pi \ln x) dx = 0$$

for every $n \geqslant 0$, which then clearly implies that f_ε is a p.d.f. and its moments coincide with the ones for X. Indeed,

$$\begin{aligned}
&\int_0^\infty x^n f(x) \sin(2\pi \ln x) dx \\
&= \frac{1}{\sqrt{2\pi}} \int_0^\infty x^{n-1} e^{-(\ln x)^2/2} \sin(2\pi \ln x) dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{nu - u^2/2} \sin(2\pi u) du \quad (u \triangleq \ln x) \\
&= \frac{e^{n^2/2}}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-(u-n)^2/2} \sin(2\pi u) du \\
&= \frac{e^{n^2/2}}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-v^2/2} \sin(2\pi v) dv \quad (v \triangleq u - n) \\
&= 0,
\end{aligned}$$

since the integrand in the last integral is an odd function.

Example 2.2. Let X be a non-negative random variable whose p.d.f. is given by

$$f(x) = \begin{cases} Ce^{-\alpha x^\gamma}, & x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Here $\alpha > 0$, $\gamma \in (0, 1/2)$ and C is a normalising constant so that f has total integral one. Again we know that X has finite moments of all orders while its m.g.f. is only defined for $t \leqslant 0$.

To motivate the construction of a family of distinct p.d.f.'s sharing the same moments, we start by observing the following analytic identity:

$$\int_0^\infty u^{p-1}e^{-zu}du = \frac{\Gamma(p)}{z^p} \tag{2.1}$$

for any $p>0$ and $z\in\mathbb{C}$ with $\mathrm{Re}(z)>0$. The power function z^p on the right hand side of (2.1) is understood as

$$z^p \triangleq \exp\left(p\left(\ln|z| + i\arg z\right)\right)$$

with $\arg z\in(-\pi/2,\pi/2)$. When z is a positive real number, (2.1) is merely the usual definition of the Gamma function. The complex case can be justified by using the Identity Theorem after observing that both sides define holomorphic functions on $\{z\in\mathbb{C}:\mathrm{Re}(z)>0\}$.

Now for given $n\geqslant 0$, we choose $p\triangleq(n+1)/\gamma$ and $z=\alpha+i\beta$ where β is to be determined. Then the identity (2.1) becomes

$$\begin{aligned}
&\int_0^\infty u^{\frac{n+1}{\gamma}-1}e^{-(\alpha+i\beta)u}du\\
&=\gamma\int_0^\infty x^n e^{-(\alpha+i\beta)x^\gamma}dx \quad (u\triangleq x^\gamma)\\
&=\gamma\int_0^\infty x^n e^{-\alpha x^\gamma}\cos(\beta x^\gamma)dx - i\gamma\int_0^\infty x^n e^{-\alpha x^\gamma}\sin(\beta x^\gamma)dx\\
&=\frac{\Gamma((n+1)/\gamma)}{(\alpha+i\beta)^{(n+1)/\gamma}}.
\end{aligned}$$

The idea is to choose β independent of n, such that

$$\frac{\Gamma((n+1)/\gamma)}{(\alpha+i\beta)^{(n+1)/\gamma}}\in\mathbb{R}.$$

If this is possible, it will imply that

$$\int_0^\infty x^n e^{-\alpha x^\gamma}\sin(\beta x^\gamma)dx = 0 \quad \text{for all } n.$$

In particular, if we define a family $\{f_\varepsilon:\varepsilon\in[-1,1]\}$ of distinct p.d.f.'s by

$$f_\varepsilon(x)\triangleq\begin{cases} f(x)\cdot\left(1+\varepsilon\sin(\beta x^\gamma)\right), & x>0,\\ 0, \text{otherwise}, \end{cases}$$

then f_ε is a legal p.d.f. sharing the same moments with X of all orders.

To choose β with the desired property, note that

$$(\alpha+i\beta)^{\frac{n+1}{\gamma}} = \rho^{\frac{n+1}{\gamma}} e^{i\cdot\frac{n+1}{\gamma}\theta},$$

where we write $\alpha+i\beta=\rho e^{i\theta}$ with $\rho>0$ and $\theta\in(-\pi/2,\pi/2)$ since $\alpha>0$. To expect that

$$(\alpha+i\beta)^{\frac{n+1}{\gamma}}\in\mathbb{R},$$

we need $\frac{n+1}{\gamma}\theta\in\mathbb{Z}\cdot\pi$ for every n, i.e. $\frac{\theta}{\gamma\pi}\in\mathbb{Z}$. But we know that

$$-\frac{1}{2\gamma}<\frac{\theta}{\gamma\pi}<\frac{1}{2\gamma}.$$

Since $\gamma\in(0,1/2)$ by the assumption, we can choose $\theta=\gamma\pi$. Equivalently, $\beta\triangleq\alpha\tan(\gamma\pi)$ does the job.

It worths pointing out that, the assumption of $\gamma\in(0,1/2)$ is critical here. In fact, when $\gamma\geqslant 1/2$, it is possible to show that there does not exist another distribution whose moments coincide with the ones for X. In other words, in this case the moments uniquely determine the distribution of X. Of course this is only surprising when $\gamma\in[1/2,1)$, since the m.g.f. is still only defined for $t\leqslant 0$. The study of this problem falls into the topic of the *moment problem*, in which general criteria on determinacy/indeterminacy are established. We refer the reader to Durrett [3] for more details.

3 The convergence theorem

The m.g.f. is also an effective tool for studying convergence in distribution. The following *convergence theorem*, which was originally due to Curtiss [2], shows that pointwise convergence for the m.g.f.'s is a sufficient (but *not* necessary) condition for convergence in distribution.

Theorem 3.1. *Let X_n be a sequence of random variables whose m.g.f.'s $M_n(t)$ are well defined in some common neighbourhood $(-\delta,\delta)$ of the origin. Suppose that $M_n(t)$ converges to some function $M(t)$ pointwisely for every $t\in(-\delta,\delta)$. Then there exists a unique random variable X, such that $X_n\to X$ in distribution, and $M(t)$ is the m.g.f. of X on $(-\delta,\delta)$.*

Proof. As before, we consider the complexified m.g.f.'s

$$\Phi_n(z)\triangleq\mathbb{E}[e^{zX_n}]$$

defined on the strip $S \triangleq (-\delta,\delta)\times i\mathbb{R}$. We first claim that, the family $\{\Phi_n\}$ is uniformly bounded over compact subsets of S. Indeed, let K be a compact subset of S. Apparently, there exists $0<\delta'<\delta$, such that $K\subseteq[-\delta',\delta']\times i\mathbb{R}$. For any $z=t+is\in K$, we have

$$|\Phi_n(z)|\leqslant \mathbb{E}[e^{|t|\cdot|X_n|}]\leqslant \mathbb{E}[e^{\delta'|X_n|}]\leqslant M_n(\delta')+M_n(-\delta').$$

Since the right hand side is convergent as $n\to\infty$, it has to be bounded in n. Therefore, the family $\{\Phi_n\}$ is uniformly bounded on K.

In addition, by the assumption we know that $\Phi_n(z)$ is convergent pointwisely on the real interval $(-\delta,\delta)\times\{0\}$, which apparently has a limit point inside the domain S. According to Vitali's theorem (cf. Theorem in the appendix), we conclude that there exists a holomorphic function $\Phi(z)$ on S, such that Φ_n converges to Φ uniformly on compact subsets of S. By restricting on the imaginary axis, we see that the characteristic function of X_n, which is $\Phi_n(it)$ ($t\in\mathbb{R}$), converges pointwisely to $\Phi(it)$. Since Φ is continuous, according to Lévy's continuity theorem for characteristic functions, there exists a unique random variable X such that $X_n\to X$ in distribution, and $\Phi(it)$ ($t\in\mathbb{R}$) is the characteristic function of X.

Next, we show that the m.g.f. $M_X(t)$ of X is well defined for $t\in(-\delta,\delta)$. Indeed, for fix $t\in(-\delta,\delta)$ and any $A>0$, let $\varphi(x)\in C(\mathbb{R}^1)$ be such that

$$\begin{cases}\varphi(x)=1, & x\in[-A,A],\\ \varphi(x)=0, & x\in[-2A,2A]^c,\\ 0\leqslant\varphi\leqslant 1.\end{cases}\tag{3.1}$$

It follows that

$$\begin{aligned}\int_{-A}^{A}e^{tx}dF_X &\leqslant \int_{-\infty}^{\infty}\varphi(x)e^{tx}dF_X\\ &\leqslant \left|\int_{-\infty}^{\infty}\varphi(x)e^{tx}dF_X-\int_{-\infty}^{\infty}\varphi(x)e^{tx}dF_{X_n}\right|\\ &\quad+\int_{-\infty}^{\infty}\varphi(x)e^{tx}dF_{X_n}\\ &\leqslant \left|\int_{-\infty}^{\infty}\varphi(x)e^{tx}dF_X-\int_{-\infty}^{\infty}\varphi(x)e^{tx}dF_{X_n}\right|+\sup_{n\geqslant1}M_n(t),\end{aligned}$$

for all n. Since $X_n\to X$ in distribution and $\varphi(x)e^{tx}\in C_c(\mathbb{R})$, by letting $n\to\infty$ we obtain that

$$\int_{-A}^{A}e^{tx}dF_X\leqslant\sup_{n\geqslant1}M_n(t).$$

Since A is arbitrary, we have

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} dF_X \leqslant \sup_{n \geqslant 1} M_n(t) < \infty.$$

Finally, if we consider the complexified m.g.f. $\tilde{\Phi}_X(z) \triangleq \mathbb{E}[e^{zX}]$ on S, we know that as the characteristic function

$$\tilde{\Phi}_X(it) = \Phi(it), \quad t \in \mathbb{R}.$$

By the Identity Theorem, we conclude that $\tilde{\Phi}_X = \Phi$ on S. In particular,

$$M_X(t) = \Phi(t) = \lim_{n\to\infty} \Phi_n(t) = \lim_{n\to\infty} M_n(t) = M(t)$$

for $t \in (-\delta, \delta)$. Therefore, $M(t)$ is the m.g.f. of X on $(-\delta, \delta)$. □

It should be pointed out that, unlike the situation for the characteristic function, the converse of Theorem 3.1 is not true in general. The following example gives a sequence of random variables which is weakly convergent, but the corresponding sequence of m.g.f.'s do not converge at any $t \neq 0$ even though they are all well defined on $\mathbb{R}$. This is not too surprising, since e^{tx} is unbounded and thus not a legal test function in the notion of weak convergence.

Example 3.1. Let X_n be a sequence of random variables with p.d.f.

$$f_n(x) = \begin{cases} \frac{C_n nx}{1+n^2x^2}, & x \in (-n, n), \\ 0, & \text{otherwise}, \end{cases}$$

where $C_n = (2\arctan n^2)^{-1}$ is the normalising constant. The c.d.f. of X_n is clearly given by

$$F_n(x) = \begin{cases} 0, & x < -n, \\ \frac{1}{2} + \frac{\arctan nx}{2\arctan n^2}, & -n \leqslant x < n, \\ 1 & x \geqslant n. \end{cases}$$

It follows that

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 0, & x < 0, \\ 1, & x > 0. \end{cases}$$

Therefore, $X_n \to X \equiv 0$ in distribution. On the other hand, the m.g.f. of X_n is given by

$$\begin{aligned} M_n(t) &= C_n \int_{-n}^{n} e^{tx} \cdot \frac{ndx}{1+n^2x^2} \\ &\geqslant C_n \cdot \int_0^n e^{|t|x} \frac{ndx}{1+n^2x^2} \\ &= C_n \cdot \int_0^{n^2} e^{|t|\cdot y/n} \frac{dy}{1+y^2} \\ &\geqslant C_n \cdot \int_0^{n^2} \frac{|t|^3y^3}{3!n^3} \frac{dy}{(1+y^2)}. \end{aligned}$$

A moment's thought shows that the integral

$$\int_0^{n^2} \frac{y^3dy}{1+y^2}$$

grows with rate n^4 as $n \to \infty$. Therefore, for any $t \neq 0$, $M_n(t) \to \infty$. This example shows that, convergence in distribution in general does not imply pointwise convergence for the m.g.f.'s, even when all the m.g.f.'s are well defined on $\mathbb{R}$.

We need some boundedness condition to expect the converse of Theorem 3.1. The following result, which was due to Mukherjea-Rao-Suen [5] and Chareka [1], gives such a condition and also partly generalises the convergence theorem.

Theorem 3.2. *Let X_n be a sequence of random variables with m.g.f. $M_n(t)$ and let X be a random variable with m.g.f. $M(t)$. Suppose that all those $M_n(t)$'s and $M(t)$ are well defined on a common interval $I = (a, b)$. Then $M_n(t)$ converges pointwisely to $M(t)$ if and only if:*

(i) X_n converges to X in distribution;
(ii) for every $t \in I$, $\sup_n M_n(t) < \infty$.

Proof. Sufficiency. Let $t \in I$ and pick $\eta > 0$ so that $t \pm \eta \in I$. For any $A > 0$ we define the bump function $\varphi(x) \in C_c([-2A, 2A])$ as in (3.1). It follows that

$$\begin{aligned}
&\left|\mathbb{E}[e^{tX_n}] - \mathbb{E}[e^{tX}]\right| \\
&\leqslant \left|\mathbb{E}[(1-\varphi(X_n))e^{tX_n}]\right| + \left|\mathbb{E}[(1-\varphi(X))e^{tX}]\right| \\
&\quad + \left|\mathbb{E}[\varphi(X_n)e^{tX_n}] - \mathbb{E}[\varphi(X)e^{tX}]\right| \\
&\leqslant \mathbb{E}[e^{tX_n}\mathbf{1}_{\{|X_n|>A\}}] + \mathbb{E}[e^{tX}\mathbf{1}_{\{|X|>A\}}] + \left|\mathbb{E}[\varphi(X_n)e^{tX_n}] - \mathbb{E}[\varphi(X)e^{tX}]\right| \\
&\leqslant e^{-\eta A}\mathbb{E}[e^{tX_n+\eta|X_n|}] + \mathbb{E}[e^{tX}\mathbf{1}_{\{|X|>A\}}] + \left|\mathbb{E}[\varphi(X_n)e^{tX_n}] - \mathbb{E}[\varphi(X)e^{tX}]\right| \\
&\leqslant e^{-\eta A} \cdot \sup_{m\geqslant 1}\left(M_m(t+\eta) + M_m(t-\eta)\right) + \mathbb{E}[e^{tX}\mathbf{1}_{\{|X|>A\}}] \\
&\quad + \left|\mathbb{E}[\varphi(X_n)e^{tX_n}] - \mathbb{E}[\varphi(X)e^{tX}]\right|.
\end{aligned}$$

Since $\varphi(x)e^{tx} \in C_b(\mathbb{R})$, when $n \to \infty$ the last term on the right hand side tends to zero, and we obtain that

$$\begin{aligned}
&\limsup_{n\to\infty}\left|\mathbb{E}[e^{tX_n}] - \mathbb{E}[e^{tX}]\right| \\
&\leqslant e^{-\eta A} \cdot \sup_{m\geqslant 1}\left(M_m(t+\eta) + M_m(t-\eta)\right) + \mathbb{E}[e^{tX}\mathbf{1}_{\{|X|>A\}}].
\end{aligned}$$

Since A is arbitrary, by letting $A \to \infty$ the right hand side of the above inequality vanishes and we conclude that $\mathbb{E}[e^{tX_n}] \to \mathbb{E}[e^{tX}]$.

Necessity. It suffices to prove Part (i) as Part (ii) is trivial. The idea is to apply a change of distribution so that we are led to the setting of Theorem 3.1 where the interval contains the origin. To achieve this, let $c \triangleq (a+b)/2$, and for each n we define a new distribution function

$$\tilde{F}_n(x) \triangleq \int_{-\infty}^{x} \frac{e^{cu}}{M_n(c)} dF_n(u), \quad x \in \mathbb{R}.$$

It is routine to see that the m.g.f. of $\tilde{F}_n$ is given by

$$\tilde{M}_n(t) = \int_{-\infty}^{\infty} e^{tx} d\tilde{F}_n(x) = \int_{-\infty}^{\infty} \frac{e^{(t+c)x}}{M_n(c)} dF_n(x) = \frac{M_n(t+c)}{M_n(c)}.$$

In particular, $\tilde{M}_n(t)$ is well defined on $(-\frac{b-a}{2}, \frac{b-a}{2})$. Moreover, by the assumption, $\tilde{M}_n(t)$ converges pointwisely on this interval to $\frac{M(t+c)}{M(c)}$ which is the m.g.f. of $\tilde{F}$ defined in a similar way as $\tilde{F}_n$ but using $M(c)$. According to Theorem 3.1 and the uniqueness theorem, we know that $\tilde{F}_n$ converges weakly to $\tilde{F}$.

Let us now show that F_n converges weakly to F. For this purpose, let φ be a continuous function on $\mathbb{R}$ which satisfies $0 \leqslant \varphi \leqslant 1$ (to prove weak convergence it is enough to test against this class of functions). For any $\varepsilon > 0$, we have

$$\begin{aligned}\int_{-\infty}^{\infty} \varphi(x) dF_n(x) &= \int_{-\infty}^{\infty} \varphi(x) \cdot \frac{M_n(c)}{e^{cx}} d\tilde{F}_n(x) \\ &\geqslant M_n(c) \cdot \int_{-\infty}^{\infty} \frac{\varphi(x)}{\max\{e^{cx}, \varepsilon\}} d\tilde{F}_n(x).\end{aligned}$$

Note that $0 \leqslant \frac{\varphi(x)}{\max\{e^{cx}, \varepsilon\}} \leqslant \frac{1}{\varepsilon}$. Since $\tilde{F}_n \to \tilde{F}$ weakly and $M_n(c) \to M(c)$, by taking $n \to \infty$ we see that

$$\begin{aligned}\liminf_{n\to\infty} \int_{-\infty}^{\infty} \varphi(x) dF_n(x) &\geqslant M(c) \cdot \int_{-\infty}^{\infty} \frac{\varphi(x)}{\max\{e^{cx}, \varepsilon\}} d\tilde{F}(x) \\ &= \int_{-\infty}^{\infty} \frac{\varphi(x) e^{cx}}{\max\{e^{cx}, \varepsilon\}} dF(x).\end{aligned}$$

This is true for arbitrary $\varepsilon > 0$. If we send $\varepsilon \to 0^+$, by the dominated convergence theorem we obtain that

$$\liminf_{n\to\infty} \int_{-\infty}^{\infty} \varphi(x) dF_n(x) \geqslant \int_{-\infty}^{\infty} \varphi(x) dF(x). \tag{3.2}$$

Replacing φ with $1 - \varphi$ yields

$$\limsup_{n\to\infty} \int_{-\infty}^{\infty} \varphi(x) dF_n(x) \leqslant \int_{-\infty}^{\infty} \varphi(x) dF(x). \tag{3.3}$$

The relations (3.2) and (3.3) together imply

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} \varphi(x) dF_n(x) = \int_{-\infty}^{\infty} \varphi(x) dF(x)$$

which concludes the desired convergence property. □

The main benefit of Theorem 3.2 is that the interval of convergence for the m.g.f.'s needs not contain the origin. This is particularly useful when one deals with non-negative random variables and their Laplace transforms (cf. Section 4). However, it should be pointed out that part of the assumptions is stronger than Theorem 3.1 as the a priori existence of the random variable X with m.g.f. $M(t)$ is presumed. Without this assumption the theorem may fail to hold, as seen from the simple example below.

Example 3.2. Let X_n be a random variable whose p.m.f. is given by

$$\mathbb{P}(X_n = -n) = \frac{1}{2},\ \mathbb{P}(X_n = 0) = \frac{1}{2}.$$

Then the m.g.f. of X_n is given by

$$M_n(t) = \frac{1}{2}e^{-nt} + \frac{1}{2}.$$

Apparently $M_n(t)$ converges to $1/2$ for all $t > 0$. However, the family $\{X_n\}$ of random variables is not tight and does not contain weakly convergent subsequences.

Using the generalised convergence theorem, we obtain a generalised uniqueness theorem easily.

Corollary 3.1. *Let X, Y be random variables whose m.g.f.'s are well defined and equal on some interval $I = (a, b)$. Then $X \stackrel{d}{=} Y$.*

Proof. Applying Theorem 3.2 to the situation when $X_n = Y$ (for all n), we know that X_n converges to X weakly. This plainly implies $X \stackrel{d}{=} Y$. □

4 The Laplace transform and its inversion formula

There are at least two shortcomings for the m.g.f. The first one is that it always comes with its intrinsic domain of definition. Another one is that, although we have the nice uniqueness theorem, an explicit formula recovering the distribution from the m.g.f. is in general not available. There is a special situation where these two disadvantages are overcome: non-negative random variables. For a non-negative random variable, it is obvious that the m.g.f. is well defined for $t \leqslant 0$. In addition, as we will see, there is an explicit inversion formula recovering the distribution using values of the m.g.f. and its derivatives "at $-\infty$". For convenience, in this context the m.g.f. is usually flipped to the positive axis and we are led to the notion of Laplace transform.

Definition 4.1. Let X be a non-negative random variable. The *Laplace transform* of X, is the function defined by

$$L_X(t) \triangleq \mathbb{E}[e^{-tX}],\ t \geqslant 0.$$

The Laplace transform is related with the m.g.f. by $L_X(t) = M_X(-t)$. It is clear that $L_X(t)$ is non-increasing in t, and $0 < L_X \leqslant 1$. By the dominated convergence theorem, $L_X(0+) = 1$ and $L_X(\infty) = \mathbb{P}(X = 0)$. It could be possible that the definition of L_X is extended beyond the origin to the negative axis but we do not need this information.

From our previous study on the m.g.f., the analyticity property (cf. Lemma 1.1), the generalised convergence theorem (cf. Theorem 3.2) and uniqueness theorem (cf. Corollary 3.1) carry through directly. To summarise:

Theorem 4.1. *Let X_n, X, Y be non-negative random variables with Laplace transforms L_{X_n}, L_X, L_Y respectively.*

(i) L_X is analytic on $(0, \infty)$. In particular, it is infinitely differentiable on $(0, \infty)$ and

$$L_X^{(n)}(t) = (-1)^n \mathbb{E}[X^n e^{-tX}], \quad t \in (0, \infty). \tag{4.1}$$

(ii) If L_{X_n} converges pointwisely to L_X on some given interval $I \subseteq [0, \infty)$, then X_n converges to X weakly.

(iii) If $L_X = L_Y$ on some given interval $I \subseteq [0, \infty)$, then $X \stackrel{d}{=} Y$.

An advantage of the Laplace transform is the availability of inversion formulae. We conclude our discussion by establishing one inversion formula in full generality (which may be less useful than other inversion formulae in the context of continuous random variables when the p.d.f. exists).

Theorem 4.2. *Let X be a non-negative random variable with Laplace transform L_X. Then at every continuity point $x \geqslant 0$ of its c.d.f. F_X, we have*

$$F_X(x) = \lim_{n \to \infty} \sum_{j=0}^{[nx]} \frac{n^j}{j!} (-1)^j L_X^{(j)}(n), \tag{4.2}$$

where $[nx]$ denotes the largest integer not exceeding nx.

Remark 4.1. The formula (4.2) holds at $x = 0$ regardless of whether F_X is continuous at 0 or not, since $L_X(n) \to \mathbb{P}(X = 0) = F_X(0)$ when $n \to \infty$. For $x > 0$, the assumption that x is a continuity point of F_X is necessary. Indeed, consider $X \equiv 1$ and thus $L_X(t) = e^{-t}$. The summation on the right hand side of (4.2) at $x = 1$ becomes $\sum_{j=0}^{n} \frac{n^j}{j!} e^{-n}$, which converges to $1/2 \neq F_X(1)$! There is a nice probabilistic way to see that

$$\lim_{n \to \infty} \sum_{j=0}^{n} \frac{n^j}{j!} e^{-n} = \frac{1}{2}.$$

Let $\{X_n : n \geqslant 1\}$ be an i.i.d. sequence of Poisson random variables with parameter $\lambda = 1$. Then

$$\sum_{j=0}^{n} \frac{n^j}{j!} e^{-n} = \mathbb{P}(X_1 + \cdots + X_n \leqslant n) = \mathbb{P}\big(\frac{X_1 + \cdots + X_n - n}{\sqrt{n}} \leqslant 0\big), \tag{4.3}$$

which converges to $1/2$ by the central limit theorem.

Proof of Theorem 4.2. The theorem can be proved by using an analytic approach. Namely, one can first show that the Laplace transform of F_n converges pointwisely to L_X, and then a density argument (e.g. using the Stone-Weierstrass theorem) shows that

$$\int_0^\infty \varphi(x) dF_n(x) \to \int_0^\infty \varphi(x) dF_X(x) \tag{4.4}$$

for any continuous function φ with compact support. By approximating the indicator function $\mathbf{1}_{[0,x]}$ with compactly supported functions on both sides, it is standard to conclude that $F_n(x) \to F(x)$ at every continuity point x of F_X.

Instead of following the usual analytic approach, we use a probabilistic argument inspired by (4.3) in Remark 4.1, which appears to be more enlightening. With aid of (4.1), we start by observing that

$$F_n(x) = \sum_{j=0}^{[nx]} \frac{n^j}{j!} (-1)^j L_X^{(j)}(n) = \mathbb{E}\left[\sum_{j=0}^{[nx]} \frac{(nX)^j}{j!} e^{-nX}\right].$$

Now fix $x > 0$ to be a continuity point of F_X. Let $\omega \in \Omega$ (the underlying sample space) be such that $X(\omega) \neq 0$. Consider an i.i.d. sequence $\{X_n\}$ of Poisson random variables with parameter $\lambda = X(\omega)$. Then

$$\begin{aligned}
&\sum_{j=0}^{[nx]} \frac{(nX(\omega))^j}{j!} e^{-nX(\omega)} \\
&= \mathbb{Q}\left(X_1 + \cdots + X_n \leqslant [nx]\right) \\
&= \mathbb{Q}\big(\frac{X_1 + \cdots + X_n - nX(\omega)}{\sqrt{nX(\omega)}} \leqslant \frac{[nx] - nX(\omega)}{\sqrt{nX(\omega)}}\big).
\end{aligned}$$

Note that here $\mathbb{P}$ is an auxiliary probability measure on some other probability space on which the sequence $\{X_n\}$ are defined. According to the central limit theorem, we know that

$$\frac{X_1 + \cdots + X_n - nX(\omega)}{\sqrt{nX(\omega)}} \to N(0, 1)$$

in distribution. This has two implications:

(i) If $X(\omega) > x$, then

$$c_n \triangleq \frac{[nx] - nX(\omega)}{\sqrt{nX(\omega)}} \to -\infty.$$

This implies that

$$\sum_{j=0}^{[nx]} \frac{(nX(\omega))^j}{j!} e^{-nX(\omega)} \to 0. \tag{4.5}$$

Indeed, for any $c < 0$, we have

$$\begin{aligned}
&\limsup_{n\to\infty} \mathbb{Q}\big(\frac{X_1 + \cdots + X_n - nX(\omega)}{\sqrt{nX(\omega)}} \leqslant c_n\big) \\
&\leqslant \lim_{n\to\infty} \mathbb{Q}\big(\frac{X_1 + \cdots + X_n - nX(\omega)}{\sqrt{nX(\omega)}} \leqslant c\big) \\
&= \Phi(c),
\end{aligned}$$

where Φ is the c.d.f. of the standard normal distribution. By taking $c \to -\infty$, we obtain (4.5).

(ii) If $X(\omega) < x$, we have

$$\frac{[nx] - nX(\omega)}{\sqrt{nX(\omega)}} \to \infty,$$

and a similar argument to Part (i) shows that

$$\sum_{j=0}^{[nx]} \frac{(nX(\omega))^j}{j!} e^{-nX(\omega)} \to 1 \tag{4.6}$$

in this case. Note that (4.6) holds trivially when $X(\omega) = 0$.

To summarise, we have

$$\lim_{n\to\infty} \sum_{j=0}^{[nx]} \frac{(nX)^j}{j!} e^{-nX} = \begin{cases} 1, & X < x, \\ 0, & X > x. \end{cases}$$

Since x is a continuity point of F_X (i.e. $\mathbb{P}(X = x) = 0$), we conclude that

$$\sum_{j=0}^{[nx]} \frac{(nX)^j}{j!} e^{-nX} \to \mathbf{1}_{[0,x]}(X) \quad \text{a.s.}$$

It is also clear that

$$0 \leqslant \sum_{j=0}^{[nx]} \frac{(nX)^j}{j!} e^{-nX} \leqslant 1.$$

By the dominated convergence theorem, we have

$$F_n(x) = \mathbb{E}\Big[\sum_{j=0}^{[nx]} \frac{(nX)^j}{j!} e^{-nX}\Big] \to \mathbb{E}[\mathbf{1}_{[0,x]}(X)] = \mathbb{P}(X \leqslant x) = F_X(x).$$

This completes the proof of Theorem 4.2.

□

Remark 4.2. Although it reaches full generality, the inversion formula (4.2) is not as useful as one may expect due to the complication of summation in the asymptotics as $n \to \infty$. If X has a continuous p.d.f. $f_X(x)$, there is a so-called *Bromwich's inversion formula*:

$$f_X(x) = \lim_{R \to +\infty} \frac{1}{2\pi i} \int_{t-iR}^{t+iR} e^{zx} L_X(z) dz$$

which is more commonly used in practice. Here $L_X(z) \triangleq \mathbb{E}[e^{zX}]$ is the uniquely holomorphic extension of $L_X(t)$ to the complex half-plane $\{z : \mathrm{Re}(z) > 0\}$, and the complex integral is performed along the line segment from $t - iR$ to $t + iR$ with any given $t > 0$.

5 An application: large deviations

We discuss an important application of the m.g.f. to large deviations. This is a rich research area in modern probability theory.

One way to motivate the study of large deviations is the following. The law of large numbers tells us that, the sample average S_n/n of an i.i.d. sequence $\{X_n\}$ of random variables converges to the mean $\mu \triangleq \mathbb{E}[X_1]$ almost surely. The central limit theorem tells us that, S_n deviates from its mean $n\mu$ roughly in the order of $\sqrt{n}$. Large deviations is concerned with events where S_n deviates from its mean $n\mu$ with a larger order, say n. For instance, we know for sure that the probability $\mathbb{P}(|S_n - n\mu| > n\theta)$ $(\theta > 0)$ is small when n is large. Large deviations studies its precise decay rate as $n \to \infty$. It is typical that such probabilities will decay exponentially:

$$\mathbb{P}(|S_n - n\mu| > n\theta) \sim e^{-nI(\theta)}$$

with certain rate $I(\theta)$ depending on θ. One important aspect in large deviations is to identify the underlying rate function along with the study of convergence.

In this section, we do not pursue full generality and restrict ourselves in the simplest one dimensional setting, in which several essential ideas in the general theory are already involved. To be precise, throughout the rest of this section, we assume that: $\{X_n : n \geqslant 1\}$ is an i.i.d. sequence of random variables whose *m.g.f. $M(t)$ is well defined in a neighbourhood of the origin.* In particular, the m.g.f. is analytic in this neighbourhood and X_n has finite moments of all orders.

Let $S_n \triangleq X_1 + \cdots + X_n$ be the corresponding partial sum sequence. Given $a > \mu \triangleq \mathbb{E}[X_1]$, we are interested in the behaviour of the probability $\mathbb{P}(S_n > na)$ as $n \to \infty$. The event $\{S_n > na\}$ captures the positive deviation of S_n with respect to its mean $n\mu$ in the order of n. Symmetrically, we can also consider the probability $\mathbb{P}(S_n < nb)$ $(b < \mu)$ of having a negative deviation correspondingly. A simple reflection $Y_n \triangleq -X_n$ allows us to focus on the first case.

We start by deriving an upper estimate which also motivates our later study. Given $t > 0$, according to Chebyshev's inequality, we have

$$\mathbb{P}(S_n > na) = \mathbb{P}(e^{tS_n} > e^{tna}) \leqslant e^{-tna}\mathbb{E}[e^{tS_n}] = e^{-tna}M(t)^n. \tag{5.1}$$

Note that the above estimate holds trivially if $M(t) = \infty$. If we introduce the so-called *cumulant generating function*

$$\Lambda(t) \triangleq \log M(t) = \log \mathbb{E}[e^{tX_1}],$$

then (5.1) can be written as

$$\mathbb{P}(S_n > na) \leqslant e^{-n(at-\Lambda(t))}, \quad \text{for all } t > 0.$$

In particular,

$$\frac{1}{n}\log \mathbb{P}(S_n \geqslant na) \leqslant -(at - \Lambda(t)), \quad \text{for all } t > 0. \tag{5.2}$$

The right hand side of (5.2) motivates the following definition which plays a central role in the theory of large deviations.

Definition 5.1. The *Fenchel-Legendre transform* of Λ is the function Λ^* defined by

$$\Lambda^*(a) \triangleq \sup_{t\in\mathbb{R}}(at - \Lambda(t)), \quad a \in \mathbb{R}. \tag{5.3}$$

Remark 5.1. It is obvious that $\Lambda^*(a) \geqslant 0$, but it is also typical that $\Lambda^*(a)$ can be ∞ (cf. Example 5.1 below). Note that the generating functions $M(t)$ and $\Lambda(t)$ can be defined on the whole real line as long as one allows $+\infty$ to be their values. Thus there is no need to restrict the supremum in (5.3) to the domain of definition for $M(t)$.

As we will see later on (cf. Lemma 5.1 below), when $a > \mu$, under the current assumptions the supremum in (5.3) can equivalently be taken over $t > 0$. Therefore, the upper estimate (5.2) becomes

$$\frac{1}{n} \log \mathbb{P}(S_n > na) \leqslant -\Lambda^*(a). \tag{5.4}$$

What is non-trivial and surprising is that, under mild conditions the above estimate becomes asymptotically sharp as $n \to \infty$. That is to say, there is also a matching lower estimate in the limit as $n \to \infty$, leading us to the following large deviation theorem in the current context.

Theorem 5.1. *Let $a > \mu$ and assume that $\mathbb{P}(X_1 > a) > 0$. Then we have $\Lambda^*(a) \in (0, \infty)$ and*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n > na) = -\Lambda^*(a). \tag{5.5}$$

Remark 5.2. Theorem 5.1 tells us that the large deviation probability $\mathbb{P}(S_n > na)$ decays exponentially like $e^{-n\Lambda^*(a)}$. Note that the assumption $\mathbb{P}(X_1 > a) > 0$ is necessary for the assertion to hold. Indeed, if $X_1 \leqslant a$ almost surely, the left hand side of (5.5) is $-\infty$ while $\Lambda^*(a)$ can still be finite (cf. Example 5.1 below with $a = 1$).

Before proving Theorem 5.1, we give a simple example to support our intuition about the rate function $\Lambda^*(a)$.

Example 5.1. Let X be the symmetric two-point random variable with distribution

$$\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = \frac{1}{2}.$$

We have $\mathbb{E}[X] = 0$ and the m.g.f. of X is

$$M_X(t) = \frac{e^t + e^{-t}}{2}.$$

For each a, define

$$\varphi_a(t) \triangleq at - \Lambda(t) = at - \log\left(\frac{e^t + e^{-t}}{2}\right).$$

Simple calculus shows that when $a \in (-1,1)$, $\varphi_a(t)$ attains its maximum at $\tau = \frac{1}{2}\log\left(\frac{1+a}{1-a}\right)$ with value

$$\Lambda^*(a) = \frac{1+a}{2}\log(1+a) + \frac{1-a}{2}\log(1-a).$$

If $a = \pm 1$, the supremum $\Lambda^*(a) = \log 2$ is attained at $\tau = \pm\infty$. If $a \notin (-1,1)$, we have $\Lambda^*(a) = \infty$. Therefore,

$$\Lambda^*(a) = \begin{cases} \frac{1+a}{2}\log(1+a) + \frac{1-a}{2}\log(1-a), & -1 \leqslant a \leqslant 1, \\ \infty, & \text{otherwise}. \end{cases}$$

The rest of this section is devoted to the proof of Theorem 5.1. Since we have already obtained the upper estimate (cf. (5.4)), it remains to establish the lower asymptotic estimate

$$\liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}(S_n > na) \geqslant -\Lambda^*(a). \tag{5.6}$$

We start by deriving the following property of $\Lambda^*(a)$.

Lemma 5.1. *Suppose that $a > \mu$ and $\mathbb{P}(X_1 > a) > 0$. Then $\Lambda^*(a) \in (0,\infty)$ and we have*

$$\Lambda^*(a) = \sup_{t>0}(at - \Lambda(t)). \tag{5.7}$$

Proof. Let $\varphi_a(t) \triangleq at - \Lambda(t)$. By assumption, we have

$$\varphi_a'(0) = a - \Lambda'(0) = a - \frac{M'(0)}{M(0)} = a - \mu > 0 \tag{5.8}$$

Since $\varphi_a(0) = 0$, it follows that $\varphi_a(t)$ achieves positive values when $t > 0$ is small. By the definition of $\Lambda^*(a)$, it must be positive in this case.

Next, we prove (5.7). For this purpose, we first show that $\Lambda(t)$ is a convex function. Indeed, given $s \leqslant t$ and $\alpha \in [0,1]$, by Hölder's inequality we have

$$\begin{aligned} M(\alpha s + (1-\alpha)t) &= \mathbb{E}[e^{\alpha(sX_1)+(1-\alpha)(tX_1)}] \\ &\leqslant \mathbb{E}[e^{sX_1}]^\alpha \cdot \mathbb{E}[e^{tX_1}]^{1-\alpha} \\ &= M(s)^\alpha \cdot M(t)^{1-\alpha}. \end{aligned}$$

Therefore,

$$\Lambda(\alpha s + (1-\alpha)t) \leqslant \alpha\Lambda(s) + (1-\alpha)\Lambda(t).$$

It follows that $\varphi_a(t)$ is a concave function. As a consequence, we have $\varphi_a(t) \leqslant 0$ for all $t \leqslant 0$. Indeed, from (5.8) we know that $\varphi_a(t) < 0$ at least for $t \in [-\eta, 0]$ with some small $\eta > 0$. If $t < -\eta$, with $\alpha \triangleq \eta/|t|$ and by the concavity of $\varphi_a(t)$ we have

$$\alpha\varphi_a(t) = \alpha\varphi_a(t) + (1-\alpha)\varphi_a(0) \leqslant \varphi_a(-\eta) < 0,$$

showing that $\varphi_a(t) < 0$. Therefore, $\varphi_a \leqslant 0$ on $(-\infty, 0]$, which then trivially implies (5.7).

Finally, for the finiteness of $\Lambda^*(a)$, one simply observes that:

$$\begin{aligned}\Lambda^*(a) &= \sup_{t>0}\big(at - \log \mathbb{E}[e^{tX_1}]\big)\\ &\leqslant \sup_{t>0}\big(at - \log \mathbb{E}\big[e^{at}\mathbf{1}_{\{X_1>a\}}\big]\big)\\ &= -\log \mathbb{P}(X_1 > a),\end{aligned}$$

which is finite by assumption. □

The key idea for proving (5.6) is to apply an exponential change of measure, so that under the new measure the mean of X_n is shifted from μ to a. Such an idea of change of measure is far reaching and it is robust enough to establish large deviation (lower) estimates in much more general situations. One who has some exposure to stochastic calculus will recognise that, what we are doing here is essentially the analogue of the Cameron-Martin-Girsanov transformation.

Before developing the analysis, we make one extra simplifying assumption. Let $t_1 \triangleq \sup\{t > 0 : M(t) < \infty\}$. We assume for the moment that the supremum in (5.7) is attained at some interior point $\tau \in (0, t_1)$. In other words,

$$\Lambda^*(a) = a\tau - \Lambda(\tau) \tag{5.9}$$

for some $\tau \in (0, t_1)$. In the end, we will get rid of this assumption through a truncation argument.

Suppose that the sequence $\{X_n : n \geqslant 0\}$ is realised on the canonical product space $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty), \mathbb{P})$. This is always possible by a standard measure-theoretic construction. More precisely, X_n is assumed to be the n-th coordinate function on $\mathbb{R}^\infty$ defined by

$$X_n(\omega) \triangleq \omega_n, \ \omega = (\omega_1, \omega_2, \cdots) \in \mathbb{R}^\infty,$$

and $\mathbb{P}$ is the unique probability measure on $\mathcal{B}(\mathbb{R}^\infty)$ induced by the (identical) laws of those X_n's. Let $\mathcal{F}_n = \sigma(X_1, \cdots, X_n)$ $(n \geqslant 0)$ be the filtration generated by the sequence $\{X_n\}$ ($\mathcal{F}_0 \triangleq \{\emptyset, \mathbb{R}^\infty\}$). Over $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$, $\mathcal{F}_n$ is merely the cylindrical σ-algebra given by

$$\mathcal{F}_n = \big\{\{\omega \in \mathbb{R}^\infty : (\omega_1, \cdots, \omega_n) \in \Gamma\big\} : \Gamma \in \mathcal{B}(\mathbb{R}^n)\}.$$

We define the exponential martingale

$$Z_0 \triangleq 1,\ Z_n \triangleq \frac{e^{\tau S_n}}{M(\tau)^n}, \quad n \geqslant 1,$$

where τ satisfies (5.9) under the previous extra assumption. It is easily checked that Z_n is a positive $\mathcal{F}_n$-martingale with mean one. Using this exponential martingale, we can construct a probability measure (exponential change of measure) on $\mathcal{B}(\mathbb{R}^\infty)$ through the following two steps:

(i) For $A \in \mathcal{F}_n$, we define

$$\tilde{\mathbb{P}}(A) \triangleq \mathbb{E}[Z_n \mathbf{1}_A].$$

The martingale property shows that $\tilde{\mathbb{P}}$ is well defined on $\cup_{n\geqslant 0}\mathcal{F}_n$ (the class of cylindrical measurable sets).

(ii) A standard measure-theoretic argument allows one to extend $\tilde{\mathbb{P}}$ to a unique probability measure on $\mathcal{B}(\mathbb{R}^\infty)$.

We leave the details to the reader as a good exercise.

In the lemma below, we summarise the basic properties of S_n under the new measure $\tilde{\mathbb{P}}$ that are needed for our purpose.

Lemma 5.2. *Under the new probability measure $\tilde{\mathbb{P}}$, $\{X_n\}$ is again an i.i.d. sequence whose m.g.f. is well defined in a neighbourhood of the origin. Moreover, we have $\tilde{\mathbb{E}}[X_n] = a$.*

Proof. We first show that $\{X_n\}$ are identically distributed under $\tilde{\mathbb{P}}$. For $A \in \mathcal{B}(\mathbb{R})$, we have

$$\begin{aligned}\tilde{\mathbb{P}}(X_n \in A) &= \mathbb{E}\big[\frac{e^{\tau S_n}}{M(\tau)^n}\mathbf{1}_{\{X_n\in A\}}\big] \\ &= \mathbb{E}\big[\frac{e^{\tau S_{n-1}}}{M(\tau)^{n-1}}\big] \cdot \mathbb{E}\big[\frac{e^{\tau X_n}}{M(\tau)}\mathbf{1}_{\{X_n\in A\}}\big] \\ &= \mathbb{E}\big[\frac{e^{\tau X_n}}{M(\tau)}\mathbf{1}_{\{X_n\in A\}}\big].\end{aligned}$$

This implies that the distribution of X_n under $\tilde{\mathbb{P}}$ is independent of n. A similar argument shows that the random variables $\{X_n\}$ are independent under $\tilde{\mathbb{P}}$. The m.g.f. of X_n under $\tilde{\mathbb{P}}$ is given by

$$\tilde{M}(t) = \tilde{\mathbb{E}}[e^{tX_n}] = \mathbb{E}\big[\frac{e^{(\tau+t)X_n}}{M(\tau)}\big] = \frac{M(t+\tau)}{M(\tau)}.$$

Since τ is an interior point of the domain of $M(t)$, we see that $\tilde{M}(t)$ is well defined in a neighbourhood of the origin. In addition,

$$\tilde{\mathbb{E}}[X_n] = \tilde{M}'(0) = \frac{M'(\tau)}{M(\tau)} = \Lambda'(\tau) = a,$$

where the last equality holds since τ is an interior extremal point of φ_a by assumption. $\square$

With the above preparations, we can now establish the lower estimate (5.6) in the current setting.

First of all, for any $b > a$ we have

$$\begin{aligned}
\mathbb{P}(S_n > na) &= \tilde{\mathbb{E}}\big[\frac{M(\tau)^n}{e^{\tau S_n}}\mathbf{1}_{\{S_n > na\}}\big] \\
&\geqslant \tilde{\mathbb{E}}\big[\frac{M(\tau)^n}{e^{\tau S_n}}\mathbf{1}_{\{nb \geqslant S_n > na\}}\big] \\
&\geqslant e^{-n(b\tau - \Lambda(\tau))} \cdot \tilde{\mathbb{P}}(na < S_n \leqslant nb).
\end{aligned}$$

Consequently,

$$\frac{1}{n}\log \mathbb{P}(S_n > na) \geqslant -(b\tau - \Lambda\tau) + \frac{1}{n}\log \tilde{\mathbb{P}}(na < S_n \leqslant nb). \tag{5.10}$$

Next, we claim that

$$\lim_{n\to\infty} \tilde{\mathbb{P}}(na < S_n \leqslant nb) = \frac{1}{2}. \tag{5.11}$$

Indeed, we have

$$\tilde{\mathbb{P}}(na < S_n \leqslant nb) = \tilde{\mathbb{P}}(S_n > na) - \tilde{\mathbb{P}}(S_n > nb).$$

Since X_n has mean a under $\tilde{\mathbb{P}}$, the central limit theorem implies

$$\tilde{\mathbb{P}}(S_n > na) = \tilde{\mathbb{P}}\big(\frac{S_n - na}{\sqrt{\mathrm{Var}[S_n]}} > 0\big) \to \frac{1}{2},$$

and the law of large numbers shows $\frac{S_n}{n} \to a$ a.s., which then implies that

$$\tilde{\mathbb{P}}(S_n > nb) \to 0.$$

Therefore, (5.11) holds.

It follows that the second term on the right hand side of (5.10) vanishes as $n \to \infty$, leading us to the estimate

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n > na) \geqslant -(b\tau - \Lambda(\tau)).$$

Since b is arbitrary, by letting $b \downarrow a$ we conclude that

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n > na) \geqslant -(a\tau - \Lambda(\tau)) = -\Lambda^*(a).$$

This completes the proof of the desired lower estimate in the current setting.

As the last piece of the puzzle, we treat the case when the extra assumption (5.9) is not met. The argument is quite technical but important.

Completing the proof of Theorem 5.1. Let us fix $c > a$ and consider the truncated sequence $X_n^c \triangleq X_n \wedge c$. The functions $M^c(t), \Lambda^c(t), \varphi_a^c(t)$ and $\Lambda^{c*}(a)$ are defined accordingly for X_1^c. We claim that X_1^c satisfies the extra assumption (5.9) we made previously. This is a consequence of the following three observations.

(i) Since $e^{tX_1^c} \leqslant e^{ct}$ for $t > 0$, the m.g.f. of X_1^c is well defined for all $t > 0$.
(ii) We pick an arbitrary $b \in (a, c)$ so that $\mathbb{P}(X_1 > b) > 0$. It follows that

$$\begin{aligned}\varphi_a^c(t) &\leqslant at - \log \mathbb{E}[e^{tX_1\wedge c}\mathbf{1}_{\{X_1>b\}}] \\ &\leqslant at - \log\big(e^{tb}\mathbb{P}(X_1 > b)\big) \\ &= t(a-b) - \log \mathbb{P}(X_1 > b).\end{aligned}$$

By taking $t \to \infty$, we obtain

$$\lim_{t\to\infty} \varphi_a^c(t) = -\infty. \tag{5.12}$$

(iii) We also have

$$(\varphi_a^c)'(0) = a - \mathbb{E}[X_1^c] \geqslant a - \mu > 0,$$

showing that $at - \Lambda^c(t) > 0$ when $t > 0$ is small.

The above three properties together imply that the function φ_a^c ($t \in [0,\infty)$) attains its maximum at some $\tau_c \in (0,\infty)$. As a consequence, we have

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n > na) \geqslant \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n^c > na) \geqslant -\Lambda^{c*}(a).$$

Since this is true for all $c > a$, and the function

$$c \mapsto \Lambda^{c*}(a) = \sup_{t>0} \big(at - \log \mathbb{E}[e^{tX_1 \wedge c}]\big)$$

is decreasing, we se that

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n > na) \geqslant -\lambda \triangleq -\lim_{c\to\infty} \Lambda^{c*}(a).$$

To finish the proof, it remains to show that $\Lambda^*(a) \geqslant \lambda$. For this purpose, we set

$$I_c \triangleq \{t \geqslant 0 : \varphi_a^c(t) \geqslant \lambda\}.$$

Note that $I_c \neq \emptyset$ since $\varphi_a^c(\tau_c) = \Lambda^{c*}(a) \geqslant \lambda$. In addition, the property (5.12) clearly implies that I_c is bounded. Therefore, as c increases, I_c forms a decreasing family of non-empty compact sets. As a result, they must have at least one common element, say τ, which satisfies

$$\varphi_a^c(\tau) = a\tau - \log \mathbb{E}[e^{\tau X_1 \wedge c}] \geqslant \lambda \quad \text{for all } c > a.$$

By the monotone convergence theorem, after sending $c \to \infty$ we conclude that

$$\Lambda^*(a) \geqslant a\tau - \log \mathbb{E}[e^{\tau X_1}] \geqslant \lambda.$$

Now the proof of Theorem 5.1 is complete.

□

There is a modern perspective to motivate the study of large deviations. Let $\mathbb{P}_n$ denote the law of $\frac{S_n}{n}$. Since $S_n/n \to \mu$ almost surely, we know that $\mathbb{P}_n(F) \to 0$ if $\mu \notin F$, at least when F is closed. The so-called large deviation principle quantifies such kind of convergence with certain rate function. Mathematically, this is the following abstract definition.

Definition 5.2. Let $\{\mathbb{P}_n : n \geqslant 1\}$ be a sequence of probability measures defined over a topological space E equipped with its Borel σ-algebra. We say that $\{\mathbb{P}_n\}$ satisfies the *large deviation principle* with a rate function $I : E \to [0,\infty]$, if:

(i) for any closed subset $F \subseteq E$, we have

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_n(F) \leqslant -\inf_{x\in F} I(x);$$

(ii) for any open subset $G \subseteq E$, we have

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_n(G) \geqslant -\inf_{x\in G} I(x).$$

Without many extra technical efforts, Theorem 5.1 leads to the following classical *Cramér's theorem*.

Theorem 5.2. *Let $\{X_n : n \geqslant 1\}$ be an i.i.d. sequence of random variables on $\mathbb{R}$. Define $\mathbb{P}_n$ to be the law of $\frac{S_n}{n}$. Then $\{\mathbb{P}_n\}$ satisfies the large deviation principle with rate function given by the Fenchel-Legendre transform of the cumulant generating function of X_1.*

Cramér's theorem has a multivariate version. The study of large deviation principle in infinite dimensions (e.g. for families of stochastic processes) is a significant area of active research. We refer the reader to the beautiful presentation of Varadhan [6] for an introduction to this exciting area.

Appendix A Some important theorems on holomorphic functions

In this appendix, we recall some properties of holomorphic functions over $\mathbb{C}$ that are used in the present notes. If one has not yet taken a course on complex analysis, with probability one he/she will find these properties quite surprising (and probably also quite elegant) when compared with the situation over $\mathbb{R}$.

Basic properties of holomorphic functions

Let $f : \Omega \to \mathbb{C}$ be a complex-valued function defined on a given domain Ω (i.e. an open subset) in $\mathbb{C}$. Let $z_0 \in \Omega$. The function f is said to be *differentiable at* z_0 if there is a complex number w_0 such that

$$\lim_{z\to z_0} \frac{f(z) - f(z_0)}{z - z_0} = w_0.$$

In this case, w_0 is called the *derivative* of f at z_0 and it is denoted as $f'(z_0)$.

Definition A.1. A function $f : \Omega \to \mathbb{C}$ is said to be *holomorphic* on Ω if it is differentiable at every point in Ω.

At first glance, holomorphicity does not seem to be a new concept compared to usual differentiability except for switching from the real field to the complex one. However, the magic of complex structure will make properties for holomorphic functions drastically different from the ones for real differentiable functions. If it has to be one, the analogue over $\mathbb{R}$ of holomorphic functions is harmonic functions.

The first theorem of this kind asserts that holomorphicity is equivalent to analyticity, and in particular implies infinite differentiability. Recall that f is *analytic at* $z_0 \in \Omega$ if f admits a convergent power series expansion in a neighbourhood of z_0. We say that f is *analytic on* Ω if it is analytic at every point in Ω. The analysis of power series over $\mathbb{C}$ is identical to the real case. For instance, if f is analytic at z_0, it is infinitely differentiable in a neighbourhood of z_0, and one can differentiate the power series term by term within its intrinsic radius of convergence.

Theorem A.1. *A function $f : \Omega \to \mathbb{C}$ is holomorphic on Ω if and only if it is analytic on Ω. In particular, it is infinitely differentiable on Ω.*

Another important theorem is known as the *Identity Theorem* which is stated as follows.

Theorem A.2. *Let f and g be holomorphic functions on a connected open subset Ω of $\mathbb{C}$. Suppose that there exists an infinite subset $D \subseteq \Omega$ such that:*

(i) $f = g$ on D;
(ii) D has a limit point in Ω.

Then $f = g$ on Ω.

The Identity Theorem tells us that holomorphic functions are global and rigid objects. It is not quite possible to cook up a holomorphic function by locally tuning its values at one's wish. For instance, in contrast to a common technique in real analysis, constructing a holomorphic "bump function" is usually of no hope in the complex world. The rigidity of holomorphic functions is also reflected by the following theorem. Note that a complex-valued function $f(z)$ can be equivalently viewed as a pair of real-valued functions $u(x, y)$ and $v(x, y)$ in two real variables (x, y):

$$f(z) = u(x, y) + iv(x, y), \quad z = x + iy.$$

The real functions u, v are called the *real* and *imaginary* parts of f respectively.

Theorem A.3. *Let f be a complex-valued function defined on a given domain Ω. Then f is holomorphic on Ω if and only if its real and imaginary parts $u(x,y), v(x,y)$ are continuously differentiable on Ω and satisfy the following so-called Cauchy-Riemann equations:*

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \ \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \quad \text{on } \Omega.$$

Before proceeding further, we recall the notion of complex integration. Let $f : \Omega \to \mathbb{C}$ be a continuous function on a given domain Ω. Let γ be a piecewise smooth path in Ω, which is given by a piecewise smooth function $z : [a,b] \to \Omega$ parametrised on some finite interval $[a,b]$. The *integral of f along γ*, denoted as $\int_\gamma f(z)dz$, is defined by the definite integral

$$\int_\gamma f(z)dz \triangleq \int_a^b f(z(t)) \cdot z'(t)dt,$$

where the product is the complex multiplication.

It is not hard to show that the integral is independent of the parametrisation of γ. However, it dependents on the orientation of γ, for if γ is run backward, the integral is changed by a sign. If γ is a loop (i.e. $z(a) = z(b)$), the integral $\int_\gamma f(z)dz$ depends only on the orientation of γ but not on its starting point. When we consider a loop integral for a non-self-intersecting loop γ, we always assume that the orientation of γ is taken in the way that the region enclosed by γ always lies on the left of the underlying orientation. Complex integration is a natural generalisation of real integrals $\int_{x_1}^{x_2} g(x)dx$ if we think of the latter as integrating along the line segment joining x_1 to x_2.

Note that the integral of f depends on the path γ in general. The following so-called *Morera's theorem* shows that holomorphicity is equivalent to the path-independence for complex integrals. This is a useful tool for proving holomorphicity.

Theorem A.4. *Let f be a continuous function defined on a given domain Ω. Then f is holomorphic on Ω if and only if*

$$\int_\gamma f(z)dz = 0 \tag{A.1}$$

for any piecewise smooth loop γ in Ω (i.e. a piecewise smooth function $z : [a,b] \to \Omega$ such that $z(a) = z(b)$).

One of the most elegant and powerful results in complex analysis is the *Cauchy's integral formula*. It can be used to establish most of the aforementioned properties for holomorphic functions.

Theorem A.5. *Let f be a holomorphic function on a given domain Ω. Let*

$$U \triangleq \{z \in \mathbb{C} : |z - z_0| < r\}$$

be a disk whose closure is contained in Ω. Then

$$f(z) = \frac{1}{2\pi i} \int_\gamma \frac{f(\zeta)}{\zeta - z} d\zeta$$

for all $z \in U$, where γ is the boundary of U and the orientation of γ is taken to be counter-clockwise.

Vitali's convergence theorem

We now discuss compactness and convergence properties for families of holomorphic functions. The core result is Vitali's convergence theorem. We provide complete proofs as this part is not always contained in a standard complex analysis course.

Let $\mathcal{F}$ be a family of complex-valued continuous functions defined on a given domain Ω.

Definition A.2. We say that the family $\mathcal{F}$ is *uniformly bounded on compact subsets* if

$$\sup_{f \in \mathcal{F}} \sup_{z \in K} |f(z)| < \infty$$

for any compact subset $K \subseteq \Omega$.

In real analysis, a uniformly bounded family of continuous functions defined on a compact set need not contain any convergent subsequences. According to the classical Arzelà-Ascoli theorem, the missing piece is precisely the property of equicontinuity. However, for holomorphic functions equicontinuity will be a consequence of uniform boundedness!

Lemma A.1. *Let $\mathcal{F}$ be a family of holomorphic functions defined on a given domain Ω. Suppose that $\mathcal{F}$ is uniformly bounded on compact subsets. Then it is equicontinuous on compact subsets. That is to say, for any compact subset $K \subseteq \Omega$ and any $\varepsilon > 0$, there exists $\delta > 0$ such that*

$$|f(z_1) - f(z_2)| < \varepsilon$$

for all $f \in \mathcal{F}$ and all $z_1, z_2 \in K$ satisfying $|z_1 - z_2| < \delta$.

Proof. Let K be a fixed compact subset of Ω. Choose $r > 0$ so that

$$K_r \triangleq \{z \in \mathbb{C} : \operatorname{dist}(z, K) \leqslant r\} \subseteq \Omega.$$

By assumption,

$$M_r \triangleq \sup_{f\in\mathcal{F}} \sup_{z\in K_r} |f(z)| < \infty.$$

Given $z_1, z_2 \in K$ with $|z_2 - z_1| < r/2$, we apply Cauchy's integral formula on the disk $\{z : |z - z_1| \leqslant r\}$ to each $f \in \mathcal{F}$:

$$f(z_i) = \frac{1}{2\pi i}\int_{\{z:|z-z_1|=r\}} \frac{f(\zeta)d\zeta}{\zeta - z_i}, \quad i = 1, 2.$$

By taking difference, we have

$$\begin{aligned}
&f(z_1) - f(z_2) \\
&= \frac{1}{2\pi i}\int_{\{z:|z-z_1|=r\}} f(\zeta)\cdot\left(\frac{1}{\zeta - z_1} - \frac{1}{\zeta - z_2}\right) d\zeta \\
&= \frac{z_1 - z_2}{2\pi i}\int_{\{z:|z-z_1|=r\}} \frac{f(\zeta)d\zeta}{(\zeta - z_1)(\zeta - z_2)}.
\end{aligned}$$

Since

$$|\zeta - z_1| = r,\ |\zeta - z_2| \geqslant r - |z_1 - z_2| > \frac{r}{2}$$

for any $\zeta \in \{z : |z - z_1| = r\}$, it follows that

$$\begin{aligned}
|f(z_1) - f(z_2)| &\leqslant \frac{|z_1 - z_2|}{2\pi}\cdot 2\pi r\cdot M_r\cdot\frac{1}{r\cdot r/2} \\
&= \frac{2M_r|z_1 - z_2|}{r}.
\end{aligned}$$

This inequality is true for all $f \in \mathcal{F}$ and all $z_1, z_2 \in K$ with $|z_1 - z_2| < r/2$. The desired equicontinuity property for $\mathcal{F}$ on K thus follows. □

Proposition A.1. *Let $\mathcal{F}$ be a family of holomorphic functions defined on a given domain Ω, which is uniformly bounded on compact subsets. Then there exists a subsequence $\{f_n\} \subseteq \mathcal{F}$ which converges to some holomorphic function f on Ω uniformly on compact subsets.*

Proof. With the aid of Lemma A.1, this is a simple consequence of the Arzelà-Ascoli theorem together with a standard diagonal selection argument. For each $m \geqslant 1$, define

$$K_m \triangleq \left\{ z \in \Omega : |z| \leqslant m \text{ and } \mathrm{dist}(z, \Omega^c) \geqslant \frac{1}{m} \right\}.$$

It is easy to see that K_m is a compact subset of Ω for each m, and $K_m \uparrow \Omega$. For $m = 1$, since the family $\mathcal{F}$ is uniformly bounded and is also equicontinuous (by Lemma A.1) on K_1, there exists a subsequence $\{f_{n_1(k)}\}$ converging uniformly to some function $f^{(1)}$ defined on K_1. By applying the same argument to the family $\{f_{n_1(k)}\}$ on K_2, we obtain a subsequence $\{f_{n_2(k)}\}$ of $\{f_{n_1(k)}\}$ converging uniformly to some function $f^{(2)}$ defined on K_2, and we can continue this argument to extract further subsequences $\{f_{n_m(k)}\} \subseteq \{f_{n_{m-1}(k)}\}$ so that $f_{n_m(k)}$ converges uniformly to some function $f^{(m)}$ defined on K_m. As the uniform limit, each $f^{(m)}$ must be continuous, and they all patch to a single continuous function f on Ω such that $f = f^{(m)}$ on K_m. Now if we consider the diagonal subsequence $\{f_{n_k(k)}\}$, then $f_{n_k(k)}$ converges to f uniformly on compact subsets. Indeed, given any compact subset K, there exists m such that $K \subseteq K_m \subseteq \Omega$, and we have $f_{n_k(k)} \to f^{(m)} = f|_{K_m}$ uniformly on K_m. The holomorphicity of f is a consequence of Morera's theorem (cf. Theorem A.4), since the integral property (A.1) is true for all $f_{n_k(k)}$'s and is thus preserved after taking limit. □

We conclude this section by establishing *Vitali's convergence theorem.*

Theorem A.6. *Let $\{f_n\}$ be a sequence of holomorphic functions on a given domain Ω, such that:*

(i) $\{f_n\}$ is uniformly bounded on compact subsets;
(ii) $\{f_n\}$ converges pointwisely on an infinite subset S of Ω, and S has a limit point in Ω.

Then there exists a holomorphic function f on Ω such that f_n converges to f uniformly on compact subsets.

Proof. It is enough to show that $\{f_n\}$ is uniformly convergent on every compact subset of Ω. Assume on the contrary that there exists a compact subset K, such that $\{f_n|_K\}$ is not a Cauchy sequence (under the uniform metric). Then there exists $\varepsilon > 0$, and two sequences m_k, n_k satisfying

$$m_1 < n_1 < m_2 < n_2 < \cdots < m_k < n_k < \cdots \uparrow \infty,$$

such that

$$\sup_{z\in K} |f_{m_k}(z) - f_{n_k}(z)| \geqslant \varepsilon \quad \text{for all } k. \tag{A.2}$$

On the other hand, according to Proposition A.1, both f_{m_k} and f_{n_k} will have convergent subsequences. For simplicity, we may assume that for some holomorphic functions g, h on Ω, f_{m_k} converges to g and f_{n_k} converges to h uniformly on compact subsets. The relation (A.2) implies that

$$\sup_{z\in K} |g(z) - h(z)| \geqslant \varepsilon. \tag{A.3}$$

However, Assumption (ii) ensures that $g = h$ on S, which by the Identity Theorem implies that $g = h$ on Ω. This is a contradiction to (A.3). $\square$

References

[1] P. Chareka, The converse to Curtiss' theorem for one-sided moment generating functions, *arXiv preprint*, arXiv:0807.3392, 2008.

[2] J.H. Curtiss, A note on the theory of moment generating functions, *Ann. Math. Stat.* 13 (4) (1942) 430–433.

[3] R. Durrett, *Probability: theory and examples*, Cambridge university press, 2019.

[4] G. Grimmett and D. Stirzaker, Probability and random processes, Third Edition, Oxford University Press, 2001.

[5] A. Mukherjea, M. Rao and S. Suen, A note on moment generating functions, *Statist. Probab. Lett.* 76 (2006) 1185–1189.

[6] S.R.S. Varadhan, *Large deviations and applications*, S.I.A.M., 1984.