On Moment-Generating Functions

Horst Hohberger and Runze Cai

For a discrete random variable, the moment-generating function is defined, formally, as

$$m_X(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \operatorname{E}[X^k]$$

We would like to show that

$$\operatorname{E}[e^{tX}] = m_X(t)$$

if $\operatorname{E}[e^{tX}]$ is real-analytic (i.e., has a power series expansion) for $-\varepsilon < t < \varepsilon$ and some $\varepsilon > 0$.
Formally, this looks simple, since we can just write

$$m_X(t) = \sum_{j=0}^{\infty} \frac{t^j}{j!} \operatorname{E}[X^j] = \operatorname{E}\Big[\sum_{j=0}^{\infty} \frac{t^j}{j!} X^j\Big] = \operatorname{E}[e^{tX}]$$

but if we want to properly justify taking the infinite series into the expectation, then we have to do some work. (A finite sum would be no problem, since we assume that $\operatorname{E}[X+Y] = \operatorname{E}[X] + \operatorname{E}[Y]$ and this additivity can easily be extended to finite sums. But the infinite series is a little tricky.)
We only discuss the case of a discrete random variable. Suppose that $X : S \to \Omega$ for some sample space S and a countable $\Omega \subset \mathbb{R}$. Since Ω is countable, we can write $\Omega = \{x_k : k \in \mathbb{N}\}$ for some sequence of real numbers (x_k).
Then, the calculation we would like to do runs as follows:

$$\begin{aligned} m_X(t) &= \lim_{n\to\infty} \sum_{j=0}^{n} \frac{t^j}{j!} \operatorname{E}[X^j] = \lim_{n\to\infty} \operatorname{E}\Big[\sum_{j=0}^{n} \frac{t^j}{j!} X^j\Big] = \lim_{n\to\infty} \sum_{k=0}^{\infty} \sum_{j=0}^{n} \frac{t^j}{j!} x_k^j f_X(x_k) \\ &= \lim_{n\to\infty} \sum_{k=0}^{\infty} \varphi_n(k) \qquad (1) \\ &= \sum_{k=0}^{\infty} \lim_{n\to\infty} \varphi_n(k) \\ &= \sum_{k=0}^{\infty} e^{tx_k} f_X(x_k) \\ &= \operatorname{E}[e^{tX}]. \end{aligned}$$

Here, we have set

$$\varphi_n(k) := f_X(x_k) \sum_{j=0}^{n} \frac{t^j}{j!} x_k^j$$

and of course

$$\lim_{n\to\infty} \varphi_n(k) = f_X(x_k) \sum_{j=0}^{\infty} \frac{t^j}{j!} x_k^j = e^{tx_k} f_X(x_k)$$

exists for every $k \in \mathbb{N}$. The problematic step is clearly the exchange of limit and summation in (1). To justify this, we need to bring in one of the "big guns" of analysis, the *dominated convergence theorem.* For the discrete sums here, a simplified form is sufficient:

Lemma 1. *Let (φ_n) be a sequence of functions $\varphi_n : \mathbb{N} \to \mathbb{R}$ such that $|\varphi_n| \le g$ for some function $g : \mathbb{N} \to \mathbb{R}$. Suppose that the pointwise limit*

$$\lim_{n\to\infty} \varphi_n(k) =: \varphi(k) \qquad \text{exists for any } k \in \mathbb{N}.$$

and that

$$\sum_{k=0}^{\infty} g(k) < \infty. \qquad (2)$$

Then

$$\lim_{n\to\infty} \sum_{k=0}^{\infty} \varphi_n(k) = \sum_{k=0}^{\infty} \varphi(k). \qquad (3)$$

In our case, we know that the limits of the φ_n exist, so we only need to find a dominating function g that satisfies (2). By our assumptions,

$$\mathrm{E}[e^{tX}] = \sum_{k=0}^{\infty} e^{x_k t} f_X(x_k) \qquad \text{converges for any } -\varepsilon < t < \varepsilon.$$

Then, for $0 \leq t < \varepsilon$,

$$\sum_{k=0}^{\infty} e^{x_k t} f_X(x_k) + \sum_{k=0}^{\infty} e^{-x_k t} f_X(x_n)$$

converges (absolutely) and so does

$$\sum_{k=0}^{\infty} (e^{x_k t} + e^{-x_k t}) f_X(x_k)$$

Since $e^{|x_n| t} \leq e^{x_k t} + e^{-x_k t}$, this implies that

$$\sum_{k=0}^{\infty} e^{|x_k| t} f_X(x_k)$$

converges for $t > 0$ and

$$\sum_{k=0}^{\infty} e^{|x_k t|} f_X(x_k)$$

converges for any $-\varepsilon < t < \varepsilon$. So we can set

$$g(k) := e^{|x_k t|} f_X(x_k).$$

Since

$$|\varphi_n(x_k)| \leq f_X(x_k) \sum_{j=0}^{n} \frac{|t x_k|^j}{j!} \leq e^{|x_k t|} f_X(x_k) = g(k),$$

we can apply Lemma 1 to justify the exchaange of summation and limit in (1).

Proof of Lemma 1. Let (φ_n), g and φ be given as in the lemma. We want to prove that for every $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ so that

$$\sum_{k=0}^{\infty} \varphi_n(k) > \sum_{k=0}^{\infty} \varphi(k) - \varepsilon \qquad \text{for all } n > N. \tag{4}$$

If we can show (4) for our (φ_n) with $\varphi_n(k) \to \varphi(k)$ for each k, then by applying the result to $-\varphi_n$, we obtain

$$\sum_{k=0}^{\infty} \varphi_n(k) < \sum_{k=0}^{\infty} \varphi(k) + \varepsilon \qquad \text{for all } n > N. \tag{5}$$

Putting (4) and (5) together, we will have shown that for every $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that for all $n > N$

$$\left| \sum_{k=0}^{\infty} \varphi_n(k) - \sum_{k=0}^{\infty} \varphi(k) \right| < \varepsilon.$$

This is just the statement (3) and we are done.
To prove (4), we want to be able to assume that $\varphi_n(k) \geq 0$. This is where the dominating function g comes in: $\varphi_n(k) \to \varphi(k)$ if and only if $\varphi_n(k) + g(k) \to \varphi(k) + g(k)$. Since $|\varphi_n(k)| \leq g(k)$, $\varphi_n(k) + g(k) \geq 0$ and, adding $\sum_{k=0}^{\infty} g(k)$ to both sides of (4), we have

$$\sum_{k=0}^{\infty} \underbrace{(\varphi_n(k) + g(k))}_{\geq 0} > \sum_{k=0}^{\infty} \underbrace{(\varphi(k) + g(k))}_{\geq 0} - \varepsilon \qquad \text{for all } n > N.$$

So, due to the existence of the dominating function, we can henceforth assume that $\varphi_n(k) \geq 0$ and $\varphi(k) \geq 0$ and return to proving (4).

Given $\varepsilon > 0$, we can choose $N_1 \in \mathbb{N}$ such that

$$\sum_{k=0}^{N_1} \varphi(k) > \sum_{k=0}^{\infty} \varphi(k) - \frac{\varepsilon}{2}$$

(this uses that $\varphi_n(k) \geq 0$). Next, choose $N_2 \in \mathbb{N}$ such that for all $n > N_2$

$$\varphi_n(k) > \varphi(k) - \varepsilon/[2(N_1 + 1)] \qquad \text{for all } k = 0, \ldots, N_1.$$

(It's not possible to do this for an infinite number of ks, but we can find such an N_2 for each individual k and then simply take the largest one, which will work for all $k = 0, \ldots, N$.) Then we put everything together:

$$\begin{aligned}\sum_{k=0}^{\infty} \varphi_n(k) &\geq \sum_{k=0}^{N_1} \varphi_n(k) \\ &> \sum_{k=0}^{N_1} \varphi(k) - \frac{\varepsilon}{2} \\ &> \sum_{k=0}^{\infty} \varphi(k) - \varepsilon\end{aligned}$$

This completes the proof. □

Remark 1. The proof for continuous random variable is basically the same, but one needs the dominated convergence theorem for integrals, which is a bit more technically difficult to prove.