10 Moment generating functions

If X is a random variable, then its *moment generating function* is

$$\phi(t) = \phi_X(t) = E(e^{tX}) = \begin{cases} \sum_x e^{tx} P(X=x) & \text{in discrete case,} \\ \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx & \text{in continuous case.} \end{cases}$$

Example 10.1. Assume that X is Exponential(1) random variable, that is,

$$f_X(x) = \begin{cases} e^{-x} & x > 0, \\ 0 & x \le 0. \end{cases}$$

Then,

$$\phi(t) = \int_0^\infty e^{tx} e^{-x}\,dx = \frac{1}{1-t},$$

only when $t < 1$. Otherwise the integral diverges and the moment generating function does not exist. Have in mind that moment generating function is only meaningful when the integral (or the sum) converges.

Here is where the name comes from. Writing its Taylor expansion in place of e^{tX} and exchanging the sum and the integral (which can be done in many cases)

$$\begin{aligned} E(e^{tX}) &= E[1 + tX + \frac{1}{2}t^2X^2 + \frac{1}{3!}t^3X^3 + \ldots] \\ &= 1 + tE(X) + \frac{1}{2}t^2E(X^2) + \frac{1}{3!}t^3E(X^3) + \ldots \end{aligned}$$

The expectation of the k-th power of X, $m_k = E(X^k)$, is called the k-th *moment* of x. In combinatorial language, then, $\phi(t)$ is the exponential generating function of the sequence m_k. Note also that

$$\frac{d}{dt}E(e^{tX})|_{t=0} = EX,$$
$$\frac{d^2}{dt^2}E(e^{tX})|_{t=0} = EX^2,$$

which lets you compute the expectation and variance of a random variable once you know its moment generating function.

Example 10.2. Compute the moment generating function for a Poisson(λ) random variable.

By definition,

$$\begin{aligned}\phi(t) &= \sum_{n=0}^{\infty} e^{tn} \cdot \frac{\lambda^n}{n!} e^{-\lambda} \\ &= e^{-\lambda} \sum_{n=0}^{\infty} \frac{(e^t\lambda)^n}{n!} \\ &= e^{-\lambda+\lambda e^t} \\ &= e^{\lambda(e^t-1)}.\end{aligned}$$

Example 10.3. Compute the moment generating function for a standard Normal random variable.

By definition,

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

where from the first to the second line we have used, in the exponent,

$$tx - \frac{1}{2}x^2 = -\frac{1}{2}(-2tx + x^2) = \frac{1}{2}((x-t)^2 - t^2).$$

Lemma 10.1. *If $X_1, X_2, \ldots, X_n$ are independent and $S_n = X_1 + \ldots + X_n$, then*

$$\phi_{S_n}(t) = \phi_{X_1}(t) \ldots \phi_{X_n}(t).$$

If X_i is identically distributed as X, then

$$\phi_{S_n}(t) = (\phi_X(t))^n .$$

Proof. This follows from multiplicativity of expectation for independent random variables:

$$E[e^{tS_n}] = E[e^{tX_1} \cdot e^{tX_2} \cdot \ldots \cdot e^{tX_n}] = E[e^{tX_1}] \cdot E[e^{tX_2}] \cdot \ldots \cdot E[e^{tX_n}].$$

□

Example 10.4. Compute the moment generating functions of a Binomial(n, p) random variable.

Here we have $S_n = \sum_{k=1}^n I_k$ where I_k are independent and $I_k = I_{\{\text{success on } k\text{th trial}\}}$, so that

$$\phi_{S_n}(t) = (e^t p + 1 - p)^n.$$

Why are moment generating functions useful? One reason is the computation of *large deviations*. Let $S_n = X_1 + \cdots + X_n$, where X_i are independent and identically distributed as X, with expectation $EX = \mu$ and moment generating function ϕ. At issue is the probability that S_n is far away from its expectation $n\mu$, more precisely $P(S_n > an)$, where $a > \mu$. We can of course use Chebyshev's inequality to get a bound of order $\frac{1}{n}$. But it turns out that this probability tends to be much smaller.

Theorem 10.2. *Large deviation bound.*

> *Assume that $\phi(t)$ is finite for some $t > 0$. For any $a > \mu$,*
>
> $$P(S_n \geq an) \leq \exp(-n\, I(a)),$$
>
> *where*
>
> $$I(a) = \sup\{at - \log\phi(t) : t > 0\} > 0.$$

Proof. For any $t > 0$, using the Markov's inequality,

$$P(S_n \geq an) = P(e^{tS_n - tan} \geq 1) \leq E[e^{tS_n - tan}] = e^{-tan}\phi(t)^n = \exp\left(-n(at - \log\phi(t))\right).$$

Note that $t > 0$ is arbitrary, so we can optimize over t to get what the theorem claims. We need to show that $\psi(a) > 0$ when $a > \mu$. For this, note that $\Phi(t) = at - \log\phi(t)$ satisfies $\Phi(0) = 0$ and, assuming that one can differentiate under the integral sign (which one can in this case but proving this requires a bit of abstract analysis beyond our scope),

$$\Phi'(t) = a - \frac{\phi'(t)}{\phi(t)} = a - \frac{E(Xe^{tX})}{\phi(t)},$$

and then

$$\Phi'(0) = a - \mu > 0,$$

so that $\Phi(t) > 0$ for some small enough positive t. □

Example 10.5. Roll a fair die n times and let S_n be the sum of the numbers you roll. Estimate the probability that S_n exceeds its expectation by at least n, for $n = 100$ and $n = 1000$.

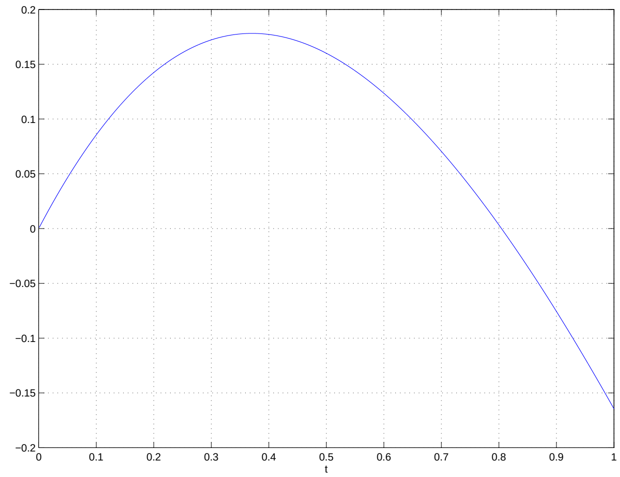
We fit this into the above theorem: observe that $\mu = 3.5$ and so $ES_n = 3.5\,n$, and that we need to find an upper bound on $P(S_n \geq 4.5\,n)$, i.e., $a = 4.5$. Moreover

$$\phi(t) = \frac{1}{6}\sum_{i=1}^{6} e^{it} = \frac{e^t(e^{6t} - 1)}{6(e^t - 1)}.$$

and we need to compute $I(4.5)$, which by definition is the maximum, over $t > 0$, of the function

$$4.5\,t - \log\phi(t),$$

whose graph is in the figure below.



It would be nice if we could solve this problem by calculus, but unfortunately we cannot (which is very common in such problems), so we resort to numerical calculations. The maximum is at $t \approx 0.37105$ and as a result $I(4.5)$ is a little larger than 0.178. This gives the upper bound

$$P(S_n \geq 4.5\, n) \leq e^{-0.178 \cdot n},$$

which is about 0.17 for $n = 10$, $1.83 \cdot 10^{-8}$ for $n = 100$, and $4.16 \cdot 10^{-78}$ for $n = 1000$. The bound $\frac{35}{12n}$ for the same probability, obtained by Chebyshev's inequality, is much much too large for large n.

Another reason why moment generating functions are useful is that they characterize the distribution, and convergence of distributions. We will state the following theorem without proof.

Theorem 10.3. *Assume that the moment generating functions for random variables X, Y, and X_n are finite for all t.*

1. *If $\phi_X(t) = \phi_Y(t)$ for all t, then $P(X \leq x) = P(Y \leq x)$ for all x.*
2. *If $\phi_{X_n}(t) \to \phi_X(t)$ for all t, and $P(X \leq x)$ is continuous in x, then $P(X_n \leq x) \to P(X \leq x)$ for all x.*

Example 10.6. Show that the sum of independent Poisson random variables is Poisson.

Here is the situation, then. We have n independent random variables $X_1, \ldots, X_n$, such that:

$$\begin{array}{lll} X_1 & \text{is Poisson}(\lambda_1), & \phi_{X_1}(t) = e^{\lambda_1(e^t-1)}, \\ X_2 & \text{is Poisson}(\lambda_2), & \phi_{X_2}(t) = e^{\lambda_2(e^t-1)}, \\ & \vdots & \\ X_n & \text{is Poisson}(\lambda_n), & \phi_{X_n}(t) = e^{\lambda_n(e^t-1)}. \end{array}$$

Then

$$\phi_{X_1+\ldots+X_n}(t) = e^{(\lambda_1+\ldots+\lambda_n)(e^t-1)}$$

and so $X_1 + \ldots + X_n$ is Poisson($\lambda_1 + \ldots + \lambda_n$). Very similarly, one could also prove that the sum of independent Normal random variables is Normal.

We will now reformulate and prove the Central Limit Theorem in a special case when moment generating function is finite. This assumption is not needed, and you should apply it as we did in the previous chapter.

Theorem 10.4. *Assume that X is a random variable with $EX = \mu$ and $\mathrm{Var}(X) = \sigma^2$, and assume that $\phi_X(t)$ is finite for all t. Let $S_n = X_1 + \ldots + X_n$, where $X_1, \ldots, X_n$ are i. i. d., and distrubuted as X. Let*

$$T_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

Then, for every x,

$$P(T_n \le x) \to P(Z \le x),$$

as $n \to \infty$, where Z is a standard Normal random variable.

Proof. Let $Y = \frac{X-\mu}{\sigma}$ and $Y_i = \frac{X_i-\mu}{\sigma}$. Then Y_i are independent, distributed as Y, $E(Y_i) = 0$, $\mathrm{Var}(Y_i) = 1$, and

$$T_n = \frac{Y_1 + \ldots + Y_n}{\sqrt{n}}.$$

To finish the proof, we show that $\phi_{T_n}(t) \to \phi_Z(t) = \exp(t^2/2)$ as $n \to \infty$:

$$\begin{aligned}
\phi_{T_n}(t) &= E\left[e^{t\,T_n}\right] \\
&= E\left[e^{\frac{t}{\sqrt{n}}Y_1+\ldots+\frac{t}{\sqrt{n}}Y_n}\right] \\
&= E\left[e^{\frac{t}{\sqrt{n}}Y_1}\right]\cdots E\left[e^{\frac{t}{\sqrt{n}}Y_n}\right] \\
&= E\left[e^{\frac{t}{\sqrt{n}}Y}\right]^n \\
&= \left(1+\frac{t}{\sqrt{n}}EY+\frac{1}{2}\frac{t^2}{n}E(Y^2)+\frac{1}{6}\frac{t^3}{n^{3/2}}E(Y^3)+\ldots\right)^n \\
&= \left(1+0+\frac{1}{2}\frac{t^2}{n}+\frac{1}{6}\frac{t^3}{n^{3/2}}E(Y^3)+\ldots\right)^n \\
&\approx \left(1+\frac{t^2}{2}\frac{1}{n}\right)^n \\
&\to e^{\frac{t^2}{2}}.
\end{aligned}$$

□

Problems

1. The player pulls three cards at random from a full deck, and collects as many dollars as the number of red cards among the three. Assume 10 people each play this game once, and let X be the number of their combined winnings. Compute the moment generating function of X.

2. Compute the moment generating function of a uniform random variable on $[0, 1]$.

3. This exercise was in fact the original motivation for the study of large deviations, by the Swedish probabilist Harald Cramèr, who was working as an insurance company consultant in 1930's. Assume that the insurance company receives a steady stream of payments, amounting to (a deterministic number) λ per day. Also every day, they receive a certain amount in claims; assume this amount is Normal with expectation μ and variance σ^2. Assume also day-to-day independence of the claims. The regulators require that within a period of n days, the company must be able to cover its claims by the payments received in the same period, *or else*. Intimidated by the fierce regulators, the company wants to fail to satisfy the regulators with probability less than some small number ϵ. The parameters n, μ, σ and ϵ are fixed, but λ is something the company controls. Determine λ.

4. Assume that S_n is Binomial(n, p). For every $a > p$, determine by calculus the large deviation bound for $P(S_n \geq an)$.

5. Using the central limit theorem for a sum of Poisson random variables, compute

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

Solutions to problems

1. Compute the moment generating function for a single game, then raise it to the 10th power:

$$\phi(t) = \left(\frac{1}{\binom{52}{3}} \left(\binom{26}{3} + \binom{26}{1}\binom{26}{2} \cdot e^t + \binom{26}{2}\binom{26}{1} \cdot e^{2t} + \binom{26}{3} \cdot e^{3t} \right) \right)^{10}.$$

2. Answer: $\phi(t) = \int_0^1 e^{tx}\, dx = \frac{1}{t}(e^t - 1)$.

3. By the assumption, a claim Y is Normal $N(\mu, \sigma^2)$ and so $X = (Y - \mu)/\sigma$ is standard normal. Note that then $Y = \sigma X + \mu$. The combined amount of claims thus is $\sigma(X_1 + \cdots + X_n) + n\mu$,

where X_i are i. i. d. standard Normal, so we need to bound

$$P(X_1 + \cdots + X_n \geq \frac{\lambda - \mu}{\sigma} n) \leq e^{-In}.$$

As $\log \phi(t) = \frac{1}{2}t^2$, we need to maximize, over $t > 0$,

$$\frac{\lambda - \mu}{\sigma} t - \frac{1}{2} t^2,$$

and the maximum equals

$$I = \frac{1}{2} \left(\frac{\lambda - \mu}{\sigma} \right)^2.$$

Finally we solve the equation

$$e^{-In} = \epsilon,$$

to get

$$\lambda = \mu + \sigma \cdot \sqrt{\frac{-2 \log \epsilon}{n}}.$$

4. After a computation, the answer you should get is

$$I(a) = a \log \frac{a}{p} + (1 - a) \log \frac{1 - a}{1 - p}.$$

5. Let S_n be the sum of i. i. d. Poisson(1) random variables. Thus S_n is Poisson(n), and $ES_n = n$. By the central limit theorem $P(S_n \leq n) \to \frac{1}{2}$, but $P(S_n \leq n)$ is exactly the expression in question. So the answer is $\frac{1}{2}$.