Course:	Mathematical Statistics
Term:	Fall 2017
Instructor:	Gordan Žitković

Lecture 6
Moment-generating functions

6.1 Definition and first properties

We use many different functions to describe probability distribution (pdfs, pmfs, cdfs, quantile functions, survival functions, hazard functions, etc.) Moment-generating functions are just another way of describing distributions, but they do require getting used as they lack the intuitive appeal of pdfs or pmfs.

Definition 6.1.1. The **moment-generating function (mgf)** of the (distribution of the) random variable Y is the function m_Y of a real parameter t defined by

$$m_Y(t) = \mathbb{E}[e^{tY}],$$

for all $t \in \mathbb{R}$ for which the expectation $\mathbb{E}[e^{tY}]$ is well defined.

It is hard to give a direct intuition behind this definition, or to explain at why it is useful, at this point. It is related to the notions of Fourier transform and generating functions. It will be only through examples in this and later lectures that a deeper understanding will emerge.

The first order of business is to compute the mgf for some of the more important (named) random variables. In the case of a continuous distribution, the main tool is the fundamental theorem which we use with the function $g(y) = \exp(ty)$ - we think of t as fixed, so that

$$m_Y(t) = \mathbb{E}[\exp(tY)] = \mathbb{E}[g(Y)] = \int_{-\infty}^{\infty} g(y) f_Y(y)\, dy = \int_{-\infty}^{\infty} e^{ty} f_Y(y)\, dy.$$

Example 6.1.2.

1. **Uniform distribution.** Let $Y \sim U(0,1)$, so that $f_Y(y) = \mathbf{1}_{\{0 \leq y \leq 1\}}$.

Then

$$m_Y(t) = \int_{-\infty}^{\infty} e^{ty} f_Y(y)\, dy = \int_0^1 e^{ty}\, dy = \tfrac{1}{t}(e^t - 1).$$

2. **Exponential distribution.** Let us compute the mgf of the exponential distribution $Y \sim E(\tau)$ with parameter $\tau > 0$:

$$m_Y(t) = \int_0^{\infty} e^{ty} \tfrac{1}{\tau} e^{-y/\tau}\, dy = \tfrac{1}{\tau} \int_0^{\infty} e^{-y(\frac{1}{\tau} - t)}\, dy = \tfrac{1}{\tau} \tfrac{1}{\frac{1}{\tau} - t} = \tfrac{1}{1 - \tau t}.$$

3. **Normal distribution.** Let $Y \sim N(0,1)$. As above,

$$m_Y(t) = \int_{-\infty}^{\infty} e^{ty} \tfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}\, dy.$$

This integral looks hard to evaluate, but there is a simple trick. We collect the exponential terms and complete the square:

$$e^{ty} e^{-\frac{1}{2}y^2} = e^{-\frac{1}{2}(y-t)^2} e^{\frac{1}{2}t^2}.$$

If we plug this into the expression above and pull out $e^{\frac{1}{2}t^2}$ which is constant, as far as the variable of integration is concerned, we get

$$m_Y(t) = e^{\frac{1}{2}t^2} \int_{-\infty}^{\infty} \tfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-t)^2}\, dy.$$

This does not look like a big improvement at first, but it is. The expression inside the integral is the pdf of a normal distribution with mean t and variance 1. Therefore, it must integrate to 1, as does any pdf. It follows that

$$m_Y(t) = e^{\frac{1}{2}t^2}.$$

As you can see from the first part of this example, the moment generating function does not have to be defined for all t. Indeed, the mfg of the exponential function is defined only for $t < \frac{1}{\tau}$. We will not worry too much for about this, and simply treat mgfs as expressions in t, but this fact is good to keep in mind when one goes deeper into the theory.

The fundamental formula for continuous distributions becomes a sum in the discrete case. When Y is discrete with support $\mathcal{S}_Y$ and pmf p_Y, the mgf can be computed as follows, where, as above, $g(y) = \exp(ty)$:

$$m_Y(t) = \mathbb{E}[e^{tY}] = \mathbb{E}[g(Y)] = \sum_{y \in \mathcal{S}_Y} \exp(ty) p_Y(y).$$

Here are some examples:

Example 6.1.3.

- **Bernoulli distribution.** For $Y \sim B(p)$, we have

$$m_Y(t) = e^{t\times 0} p_Y(0) + e^{t\times 1} p_Y(1) = q + pe^t,$$

  where $q = 1 - p$.

- **Geometric distribution.** If $Y \sim g(p)$, then $\mathbb{P}[Y = y] = q^y p$ and so

$$m_Y(t) = \sum_{y=0}^{\infty} e^{ty} pq^y = p \sum_{y=0}^{\infty} (qe^t)^y = \tfrac{p}{1-qe^t},$$

  where the last equality uses the familiar expression for the sum of a geometric series. We note that this only works for $qe^t < 1$, so that, like the exponential distribution, the geometric distribution comes with a mgf defined only for some values of t.

- **Poisson distribution.** Let Y have the Poisson distribution $P(\lambda)$ with parameter $\lambda > 0$, so that $\mathcal{S}_Y = \{0, 1, 2, 3, \dots\}$ and $p_Y(y) = e^{-\lambda} \frac{\lambda^y}{y!}$. Then

$$\begin{aligned} m_Y(t) &= \mathbb{E}[\exp(tY)] = \mathbb{E}[g(Y)] = \sum_{y \in \mathcal{S}_Y} g(y) p_Y(y) \\ &= \sum_{y=0}^{\infty} \exp(ty) \exp(-\lambda) \tfrac{\lambda^y}{y!} = e^{-\lambda} \sum_{y=0}^{\infty} \tfrac{(e^t\lambda)^y}{y!} \end{aligned}$$

  The last sum on the right is nothing else by the Taylor formula for the exponential function at $x = e^t\lambda$. Therefore,

$$m_Y(t) = e^{\lambda(e^t-1)}.$$

Here is how to compute the moment generating function of a linear transformation of a random variable. The formula follows from the simple fact that $\mathbb{E}[\exp(t(aY + b))] = e^{tb}\mathbb{E}[e^{(at)Y}]$:

Proposition 6.1.4. *Suppose that the random variable Y has the mgf $m_Y(t)$. Then mgf of the random variable $W = aY + b$, where a and b are constants, is given by*

$$m_W(t) = e^{tb} m_Y(at).$$

Example 6.1.5.

1. **Uniform distribution.** Let $W \sim U(l, r)$. We can represent W as $W = aY + b$ where $Y \sim U(0, 1)$, $a = (r - l)$ and $b = l$. We computed the mgf of Y in Example 6.1.2 above - $m_Y(t) = \frac{1}{t}(e^t - 1)$. Therefore, by Proposition 6.1.4

$$m_W(t) = e^{tl} \tfrac{1}{(r-l)t}(e^{(r-l)t} - 1) = \tfrac{e^{tr} - e^{ta}}{t(r-l)}.$$

2. **Normal distribution.** If $W \sim N(\mu, \sigma)$, then W has the same distribution as $\mu + \sigma Z$, where $Z \sim N(0, 1)$. Using the expression from Example 6.1.2 for the mgf of a *unit* normal distribution $Z \sim N(0, 1)$, we have

$$m_W(t) = e^{\mu t} e^{\frac{1}{2}\sigma^2 t^2} = e^{\mu t + \frac{1}{2}\sigma^2 t^2}.$$

6.2 Sums of independent random variables

One of the most important properties of the moment-generating functions is that they turn sums of independent random variables into products:

Proposition 6.2.1. *Let $Y_1, Y_2, \dots, Y_n$ be independent random variables with mgfs $m_{Y_1}(t), m_{Y_2}(t), \dots, m_{Y_n}(t)$. Then the mgf of their sum $Y = Y_1 + Y_2 + \dots + Y_n$ is given by*

$$m_Y(t) = m_{Y_1}(t) \times m_{Y_2}(t) \times \dots \times m_{Y_n}(t).$$

This proposition is true for all random variables, but here is a sketch of the argument in the continuous case. It is a consequence of the factorization theorem (Theorem **??**) and the fundamental formula (Theorem **??**). For simplicity, let us assume that $n = 2$:

$$m_{Y_1+Y_2}(t) = \mathbb{E}[e^{t(Y_1+Y_2)}] = \mathbb{E}[g(Y_1, Y_2)],$$

where $g(y_1, y_2) = e^{t(y_1+y_2)}$. The factorization criterion says that $f_{Y_1,Y_2}(y_1, y_2) =$

$f_{Y_1}(y_1) f_{Y_2}(y_2)$, and, so

$$\begin{aligned} m_{Y_1+Y_2}(y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{t(y_1+y_2)} f_{Y_1,Y_2}(y_1,y_2)\, dy_2\, dy_1 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{ty_1} e^{ty_2} f_{Y_1}(y_1) f_{Y_2}(y_2)\, dy_2\, dy_1 \\ &= \int_{-\infty}^{\infty} e^{ty_1} f_{Y_1}(y_1) \Big(\int_{-\infty}^{\infty} e^{ty_2} f_{Y_2}(y_2)\, dy_2 \Big)\, dy_1 \\ &= \int_{-\infty}^{\infty} e^{ty_1} f_{Y_1}(y_1) m_{Y_2}(t)\, dy_2 = m_{Y_2}(t) \int_{-\infty}^{\infty} e^{ty_2} f_{Y_2}(y_2)\, dy_2 \\ &= m_{Y_2}(t) m_{Y_1}(t). \end{aligned}$$

Example 6.2.2. Binomial distribution. Let $Y \sim b(n,p)$. We know that Y counts the number of successes in n independent Bernoulli trials, so we can represent (in distribution) as $Y = Y_1 + \cdots + Y_n$, where each Y_i is a $B(p)$-random variable. We know from Example 6.1.3 that the mgf $m_{Y_i}(t)$ of each Y_i is $q + pe^t$. Therefore

$$m_Y(t) = m_{Y_1}(t) \times m_{Y_2}(t) \times \cdots \times m_{Y_n}(t) = (q + pe^t)^n.$$

We could have obtained the same formula without the factorization criterion, but the calculation is trickier:

$$\begin{aligned} m_Y(t) &= \sum_{y=0}^{n} e^{ty} p_Y(y) = \sum_{y=0}^{n} e^{ty} \binom{n}{y} p^y q^{n-y} = \sum_{y=0}^{n} \binom{n}{y} (pe^t)^y q^{n-y} \\ &= (pe^t + q)^n, \end{aligned}$$

where the last inequality follows from the **binomial formula**

$$(a+b)^n = \sum_{y=0}^{n} \binom{n}{y} a^y b^{n-y}.$$

6.3 Why "moment-generating"?

The terminology "moment generating function" comes from the following nice fact:

Proposition 6.3.1. *Suppose that the moment-generating function $m_Y(t)$ of a random variable Y admits an expansion into a power series. Then the* *coefficients are related to the moments of Y in the following way:*

$$m_Y(t) = \sum_{k=0}^{\infty} \tfrac{\mu_k}{k!} t^k, \tag{6.3.1}$$

where $\mu_k = \mathbb{E}[Y^k]$ is the k-th moment of Y.

A fully rigorous argument of this proposition is beyond the scope of these notes, but we can see why it works is we do the following formal computation based on the Taylor expansion of the exponential function

$$e^x = \sum_{k=0}^{\infty} \tfrac{x^k}{k!}.$$

We plug in $x = tY$ and then take the expectation to get

$$m_Y(t) = \mathbb{E}[e^{tY}] = \mathbb{E}[\sum_{k=0}^{\infty} \tfrac{(tY)^k}{k!}] = \sum_{k=0}^{\infty} t^k \tfrac{1}{k!} \mathbb{E}[Y^k] = \sum_{k=0}^{\infty} \tfrac{\mu_k}{k!} t^k.$$

Formula (6.3.1) suggests the following approach to the computation of moments of a random variable:

1. Compute the mgf $m_Y(t)$.
2. Expand it in a power series in t, i.e., write

$$m_Y(t) = \sum_{k=0}^{\infty} a_k t^k.$$

3. Set $\mu_k = k! a_k$.

Example 6.3.2.

1. **Moments of the exponential distribution.** We know from Example 6.1.2 that the mgf $m_Y(t)$ of the exponential $E(\tau)$-distribution is $\frac{1}{1-\tau t}$. It is not hard to expand this into a power series because $\frac{1}{1-\tau t}$ is nothing by the sum of a geometric series

$$\tfrac{1}{1-\tau t} = \sum_{k=0}^{\infty} \tau^k t^k.$$

It follows immediately that

$$\mu_k = k! \tau^k.$$

2. **Moments of the uniform distribution.** The same example as above tells us that the mgf of the uniform distribution $U(l,r)$ is

$$m_Y(t) = \frac{e^{tr}-e^{tl}}{t(r-l)}.$$

We expand this into a series, buy expanding the numerator, first:

$$e^{tr} - e^{tl} = \sum_{k=0}^{\infty} \frac{1}{k!}(tr)^k - \sum_{k=0}^{\infty} \frac{1}{k!}(tl)^k = \sum_{k=0}^{\infty} \frac{r^k-l^k}{k!} t^k.$$

Then we divide by the denominator $t(r-l)$ to get

$$m_Y(t) = \sum_{k=0}^{\infty} \frac{r^k-l^k}{k!(r-l)} t^{k-1} = 1 + \frac{r^2-l^2}{2!(r-l)} t + \frac{r^3-l^3}{3!(r-l)} t^2 + \dots.$$

It follows that

$$\mu_k = \frac{r^{k+1}-l^{k+1}}{(k+1)(r-l)}.$$

3. **Normal distribution.** Let Y be a unit normal random variable, i.e., $Y \sim N(0,1)$. We have computed its mgf $e^{t^2/2}$ above and we expand it using the Taylor formula for the exponential function:

$$e^{t^2/2} = \sum_{k=0}^{\infty} \frac{1}{k!}(t^2/2)^k = \sum_{k=0}^{\infty} \frac{1}{2^k k!} t^{2k}.$$

The odd powers of t are all 0 so

$$\mu_k = 0 \text{ if } k \text{ is odd.}$$

For a moment of an even order $2k$, we get

$$\mu_{2k} = \frac{(2k)!}{2^k k!}.$$

In all examples above we managed to expand the mfg into a power series without using Taylor's theorem, i.e., without derivatives. Sometimes, the easiest approach is to differentiate (the notation in (6.3.2) means "take k derivatives in t, and then set $t=0$"):

Proposition 6.3.3. *The k-th moment μ_k of a random variable with the moment-generating function $m_Y(t)$ is given by*

$$\mu_k = \frac{d^k}{dt^k} m_Y(0)|_{t=0}, \qquad (6.3.2)$$

as long as m_Y is defined for t in some neighborhood of 0.

Example 6.3.4. Let Y be the Poisson random variable so that

$$m_Y(t) = e^{\lambda(e^t-1)}.$$

The first derivative $m'_Y(t)$ is given by $e^{\lambda(e^t-1)}\lambda e^t$ and, so

$$\mu_1 = \mathbb{E}[Y] = m'_Y(t)|_{t=0} = e^{\lambda(e^0-1)}\lambda e^0 = \lambda.$$

We can differentiate again to obtain

$$m''_Y(t) = \lambda(1+e^t\lambda)e^{t+\lambda(e^t-1)},$$

which yields $\mu_2 = \lambda(1+\lambda)$. One can continue and compute higher moments

$$\mu_3 = \lambda(1+3\lambda+\lambda^2),\ \mu_4 = \lambda(1+7\lambda+6\lambda^2+\lambda^3),\ \text{etc.}$$

There is no simple formula for the general term μ_k.

6.4 Recognizing the distribution

It is clear that different distributions come with different pdfs (pmf) and cdfs. It is also true for mgfs, but it is far from obvious and the proof is way outside the scope of these notes:

Theorem 6.4.1 (Uniqueness theorem)**.** *If two random variables Y_1 and Y_2 have the same moment generating functions, i.e., if*

$$m_{Y_1}(t) = m_{Y_2}(t) \text{ for all } t,$$

then the have the same distribution. In particular,

1. *if Y_1 is discrete, then so is Y_2, and Y_1 and Y_2 have the same support and the pmf, i.e.,*

$$\mathcal{S}_{Y_1} = \mathcal{S}_{Y_2} \text{ and } p_{Y_1}(y) = p_{Y_2}(y) \text{ for all } y.$$

2. *If Y_1 is continuous, then so is Y_2, and Y_1 and Y_2 have the same pdf, i.e.,*

$$f_{Y_1}(y) = f_{Y_2}(y), \text{ for all } y.$$

The way we use this result is straightforward:

1. Find an expression for the mgf of the random variable Y whose distribution you don't know (most often, in this class, using a combination of Propositions 6.1.4 and 6.2.1, but also see Problem 6.5.8).
2. Look it up in a table of mgfs to identify it.

Of course, for this to work you need to able to compute the mgf, and once you do, it has to be of the form you can recognize. When it works, it works very well, as the next example shows:

Example 6.4.2. Let Y_1 and Y_2 be independent normal variables with means μ_1 and μ_2, and standard deviations σ_1 and σ_2. We have computed mgf of normal random variables (with general parameters) in Example 6.1.5:

$$m_{Y_1}(t) = e^{\mu_1 t + \frac{1}{2}\sigma_1^2 t^2} \text{ and } m_{Y_2}(t) = e^{\mu_2 t + \frac{1}{2}\sigma_2^2 t^2}.$$

Since Y_1 and Y_2 are assumed to be independent, the mgf of their sum is the product of the individual mgfs, i.e.,

$$m_{Y_1+Y_2}(t) = e^{\mu_1 t + \frac{1}{2}\sigma_1^2 t^2} \times e^{\mu_2 t + \frac{1}{2}\sigma_2^2 t^2} = e^{(\mu_1+\mu_2)t + \frac{1}{2}(\sigma_1^2+\sigma_2^2)t^2}.$$

We recognize this as the mgf of a normal random variable, with mean $\mu = \mu_1 + \mu_2$ and the standard deviation $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$.

6.5 Problems

Problem 6.5.1. The distribution family of the random variable Y with moment generating function

$$m_Y(t) = 0.1\, e^t + 0.2\, e^{2t} + 0.3\, e^{3t} + 0.4\, e^{4t},$$

is

(a) binomial (b) geometric (c) Poisson (d) uniform (e) none of the above

Problem 6.5.2. Identify the distributions with the following mgfs:

1. $\frac{2}{2-t}$.
2. e^{2e^t-2},
3. $e^{t(t-2)}$,
4. $(3-2e^t)^{-1}$
5. $\frac{1}{9} + \frac{4}{9}e^t + \frac{4}{9}e^{2t}$.
6. $\frac{1}{t}(1-e^{-t})$.
7. $\frac{1}{4}(e^{4t} + 3e^{-t})$

If the distribution has a name, give the name and the parameters. If it does not, give the pdf or the pmf (table).

Problem 6.5.3. The standard deviation of the random variable Y whose mgf is given by $m_Y(t) = \frac{1}{\sqrt{1-2t}}$ is:

(a) 2 (b) $\sqrt{2}$ (c) $\frac{4}{3}$ (d) $\frac{\sqrt{3}}{2}$ (e) none of the above

Problem 6.5.4. Compute the standard deviation of the random variable Y whose mgf is given by $m_Y(t) = \frac{1}{(2-e^t)^3}$.

Problem 6.5.5. Let $Y_1, \dots, Y_n$ be independent random variables with distribution $N(\mu, \sigma)$. Then the mgf of $\frac{1}{n}(Y_1 + \dots + Y_n)$ is

(a) $e^{\frac{\mu}{n}t + \frac{1}{2}\sigma^2 t^2}$

(b) $e^{\mu t + \frac{1}{2}n\sigma^2 t^2}$

(c) $e^{\mu t + \frac{1}{2n}\sigma^2 t^2}$

(d) $e^{n\mu t + \frac{1}{2}\sigma^2 t^2}$

(e) none of the above

Problem 6.5.6. What is the distribution of the sum $S = Y_1 + Y_2 + \dots + Y_k$, if $Y_1, \dots, Y_k$ are independent and, for $i = 1, \dots, k$,

1. $Y_i \sim B(p)$, with $p \in (0,1)$,
2. $Y_i \sim b(n_i, p)$, with $n_1, \dots, n_k \in N$, $p \in (0,1)$,
3. $Y_i \sim P(\lambda_i)$, with $\lambda_1, \dots, \lambda_k > 0$,

Problem 6.5.7 (The mgf of the χ^2 distribution)**.** We learned in class that the distribution of the random variable $Y = Z^2$, where Z is a standard normal, is called the χ^2**-distribution**. We also computed its pdf using the cdf method. The goal of this exercise is to compute its mgf.

1. Compute $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\beta x^2}\, dx$, where $\beta > 0$ is a constant. (*Hint:* Use the fact that $\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(y-\mu)^2/\sigma^2}\, dy = 1$, as the expression inside the integral is the pdf of $N(\mu, \sigma)$. Plug in $\mu = 0$ and an appropriate value of σ.)
2. Compute the mgf of a χ^2-distributed random variable by computing $\mathbb{E}[e^{tZ^2}]$, where Z is a standard normal, i.e., $Z \sim N(0,1)$. (*Hint:* Compute this expectation by integrating the function $g(z) = e^{tz^2}$ against the standard normal density, and not the function $g(y) = e^{ty}$ against the density of Y, which we derived in class. It is easier this way.)

3. Let Z_1 and Z_2 be two independent standard normal random variables. Compute the mgf of $W = Z_1^2 + Z_2^2$, and then look through the notes for a (named) distribution which has the same mgf. What is it? What are its parameter(s)? (*Note:* If Z_1 and Z_2 are independent, then so are Z_1^2 and Z_2^2.)

Problem 6.5.8. (*) Let $Y_1, Y_2, \dots, Y_n$ be independent random variables each with the Bernoulli $B(p)$ distribution, for some $p \in (0,1)$.

1. Show that the mgf $m_{W_n}(t)$ of the random variable

$$W_n = \frac{Y_1+Y_2+\cdots+Y_n-np}{\sqrt{np(1-p)}},$$

can be written in the form

$$m_{W_n}(t) = (pe^{\frac{t}{\sqrt{n}}\alpha} + (1-p)e^{-\frac{t}{\sqrt{n}}\alpha^{-1}})^n, \tag{6.5.1}$$

for some α and find its value.

2. Write down the Taylor approximations in t around 0 for the functions $\exp(\frac{t}{\sqrt{n}}\alpha)$ and $\exp(-\frac{t}{\sqrt{n}}\alpha^{-1})$, up to and including the term involving t^2. Then, substitute those approximations in (6.5.1) above. What to you get? When n is large, $\frac{t}{\sqrt{n}}\alpha$ and $\frac{t}{\sqrt{n}}\alpha^{-1}$ are close to 0 and it can be shown that the expression you got is the limit of $m_{W_n}(t)$, as $n \to \infty$.

3. What distribution is that limit the mgf of?

(*Note:* Convergence of mgfs corresponds to a very important mode of convergence, called the *weak convergence*. We will not talk about it in this class, but it is exactly the kind of convergence that appears in the central limit theorem, which is, in turn, behind the effectiveness of the normal approximation to binomial random variables. In fact, what you just did is a fundamental part of the proof of the central limit theorem.)