### 0.0.1 Moment Generating Functions

There are many uses of generating functions in mathematics. We often study the properties of a sequence $a_n$ of numbers by creating the function

$$\sum_{n=0}^{\infty} a_n s^n$$

In statistics the most commonly used generating functions are the probability generating function (for discrete variables), the moment generating function, the characteristic function and the cumulant generating function. I begin with moment generating functions:

**Definition**: The moment generating function of a real valued random variable $X$ is

$$M_X(t) = \mathrm{E}(e^{tX})$$

defined for those real $t$ for which the expected value is finite.

**Definition**: The moment generating function of a random vector $X \in R^p$ is

$$M_X(u) = \mathrm{E}[e^{u^t X}]$$

defined for those vectors $u$ for which the expected value is finite.

This function has a formal connection to moments obtained by taking expected values term by term; in fact if $M_X(t)$ is finite for all $|t| < \epsilon$ then it is legitimate to take expected values term by term for $|t| < \epsilon$. We get

$$\begin{aligned} M_X(t) &= \sum_{k=0}^{\infty} \mathrm{E}[(tX)^k]/k! \\ &= \sum_{k=0}^{\infty} \mu_k' t^k/k!\,. \end{aligned}$$

Sometimes we can find the power series expansion of $M_X$ and read off the moments of $X$ from the coefficients of $t^k/k!$.

**Theorem 1** *If $M$ is finite for all $t \in [-\epsilon, \epsilon]$ for some $\epsilon > 0$ then*

1. *Every moment of $X$ is finite.*

2. *$M$ is $C^\infty$ (in fact $M$ is analytic).*

*3.* $\mu'_k = \frac{d^k}{dt^k} M_X(0)$.

**Note**: A function is $C^\infty$ if it has continuous derivatives of all orders.

**Note**: Analytic means the function has a convergent power series expansion in neighbourhood of each $t \in (-\epsilon, \epsilon)$.

The proof, and many other facts about moment generating functions, rely on advanced techniques in the field of complex variables. I won't be proving any of these assertions.

### 0.0.2 Moment Generating Functions and Sums

One of the most useful facts about moment generating functions is that the moment generating function of a sum of independent variables is the product of the individual moment generating functions.

**Theorem 2** *If $X_1, \ldots, X_p$ are independent random vectors in $\mathbb{R}^p$ and $Y = \sum X_i$ then the moment generating function of $Y$ is the product of those of the individual $X_i$:*

$$M_Y(u) = \mathrm{E}(e^{u^t Y}) = \prod_i \mathrm{E}(e^{u^t X_i}) = \prod_i M_{X_i}(u).$$

If we could find the power series expansion of $M_Y$ then we could find the moments of $M_Y$. The problem, however, is that the power series expansion of $M_Y$ not nice function of the expansions of individual $M_{X_i}$. There is a related fact, namely, that the first 3 moments (meaning $\mu$, $\sigma^2$ and $\mu_3$) of $Y$ are sums of those of the $X_i$:

$$\begin{aligned}
\mathrm{E}(Y) &= \sum \mathrm{E}(X_i) \\
\mathrm{Var}(Y) &= \sum \mathrm{Var}(X_i) \\
\mathrm{E}[(Y - \mathrm{E}(Y))^3] &= \sum \mathrm{E}[(X_i - \mathrm{E}(X_i))^3]
\end{aligned}$$

(I have given the univariate versions of these formulas but the multivariate versions are correct as well. The first line is a vector, the second a matrix and the third an object with 3 subscripts.) However:

$$\begin{aligned}
\mathrm{E}[(Y - \mathrm{E}(Y))^4] = &\sum \{\mathrm{E}[(X_i - \mathrm{E}(X_i))^4] - 3\mathrm{E}^2[(X_i - \mathrm{E}(X_i))^2]\} \\
&+ 3\left\{\sum \mathrm{E}[(X_i - \mathrm{E}(X_i))^2]\right\}^2
\end{aligned}$$

These observations lead us to consider cumulants and the cumulant generating function. Since the logarithm of a product is a sum of logarithms we are led to consider taking logs of the moment generating function. The result will give us *cumulants* which add up properly.

**Definition**: the cumulant generating function of a a random vector $X$ by

$$K_X(u) = \log(M_X(u))\,.$$

Then if $X_1, \ldots, X_n$ are independent and $Y = \sum_1^n X_i$ we have

$$K_Y(t) = \sum K_{X_i}(t)\,.$$

Note that moment generating functions are all positive so that the cumulant generating functions are defined wherever the moment generating functions are.

Now $K_Y$ has a power series expansion. I consider here only the univariate case.

$$K_Y(t) = \sum_{r=1}^{\infty} \kappa_r t^r / r!\,.$$

**Definition**: the $\kappa_r$ are the cumulants of $Y$.

Observe that

$$\kappa_r(Y) = \sum \kappa_r(X_i)\,.$$

In other words cumulants of independent quantities add up. Now we examine the relation between cumulants and moments by relating the power series expansion of $M$ with that of its logarithm. The cumulant generating function is

$$\begin{aligned} K(t) &= \log(M(t)) \\ &= \log(1 + [\mu_1 t + \mu_2' t^2/2 + \mu_3' t^3/3! + \cdots]) \end{aligned}$$

Call the quantity in $[\ldots]$ $x$ and expand

$$\log(1+x) = x - x^2/2 + x^3/3 - x^4/4 \cdots\,.$$

Stick in the power series

$$x = \mu t + \mu_2' t^2/2 + \mu_3' t^3/3! + \cdots\,;$$

Expand out powers of $x$ and collect together like terms. For instance,

$$\begin{aligned}
x^2 &= \mu^2 t^2 + \mu\mu_2' t^3 + [2\mu_3'\mu/3! + (\mu_2')^2/4]t^4 + \cdots \\
x^3 &= \mu^3 t^3 + 3\mu_2'\mu^2 t^4/2 + \cdots \\
x^4 &= \mu^4 t^4 + \cdots .
\end{aligned}$$

Now gather up the terms. The power $t^1$ occurs only in $x$ with coefficient $\mu$. The power $t^2$ occurs in $x$ and in $x^2$ and so on. Putting these together gives

$$\begin{aligned}
K(t) =& \mu t + [\mu_2' - \mu^2]t^2/2 + [\mu_3' - 3\mu\mu_2' + 2\mu^3]t^3/3! \\
& + [\mu_4' - 4\mu_3'\mu - 3(\mu_2')^2 + 12\mu_2'\mu^2 - 6\mu^4]t^4/4! \cdots
\end{aligned}$$

Comparing coefficients of $t^r/r!$ we see that

$$\begin{aligned}
\kappa_1 &= \mu \\
\kappa_2 &= \mu_2' - \mu^2 = \sigma^2 \\
\kappa_3 &= \mu_3' - 3\mu\mu_2' + 2\mu^3 = \mathrm{E}[(X-\mu)^3] \\
\kappa_4 &= \mu_4' - 4\mu_3'\mu - 3(\mu_2')^2 + 12\mu_2'\mu^2 - 6\mu^4 \\
&= \mathrm{E}[(X-\mu)^4] - 3\sigma^4 .
\end{aligned}$$

**Reference**: Kendall and Stuart (or a new version called *Kendall's Theory of Advanced Statistics* by Stuart and Ord) for formulas for larger orders $r$.

**Example**: The normal distribution: Suppose $X_1, \ldots, X_p$ independent, $X_i \sim N(\mu_i, \sigma_i^2)$ so that

$$\begin{aligned}
M_{X_i}(t) &= \int_{-\infty}^{\infty} e^{tx} e^{-\frac{1}{2}(x-\mu_i)^2/\sigma_i^2} dx/(\sqrt{2\pi}\sigma_i) \\
&= \int_{-\infty}^{\infty} e^{t(\sigma_i z + \mu_i)} e^{-z^2/2} dz/\sqrt{2\pi} \\
&= e^{t\mu_i} \int_{-\infty}^{\infty} e^{-(z-t\sigma_i)^2/2 + t^2\sigma_i^2/2} dz/\sqrt{2\pi} \\
&= e^{\sigma_i^2 t^2/2 + t\mu_i} .
\end{aligned}$$

The cumulant generating function is then

$$K_{X_i}(t) = \log(M_{X_i}(t)) = \sigma_i^2 t^2/2 + \mu_i t .$$

The cumulants are $\kappa_1 = \mu_i$, $\kappa_2 = \sigma_i^2$ and every other cumulant is 0. Cumulant generating function for $Y = \sum X_i$ is

$$K_Y(t) = \sum \sigma_i^2 t^2/2 + t \sum \mu_i$$

which is the cumulant generating function of $N(\sum \mu_i, \sum \sigma_i^2)$.

**Example**: The $\chi^2$ distribution: In you homework I am asking you to derive the moment and cumulant generating functions and moments of a Gamma random variable. Now suppose $Z_1, \ldots, Z_\nu$ independent $N(0,1)$ rvs. By definition the random variable $S_\nu = \sum_1^\nu Z_i^2$ has $\chi_\nu^2$ distribution. It is easy to check $S_1 = Z_1^2$ has density

$$(u/2)^{-1/2} e^{-u/2}/(2\sqrt{\pi})$$

and then the moment generating function of $S_1$ is

$$(1-2t)^{-1/2}.$$

It follows that

$$M_{S_\nu}(t) = (1-2t)^{-\nu/2}$$

which is (from the homework) the moment generating function of a Gamma$(\nu/2, 2)$ random variable. So the $\chi_\nu^2$ distribution has a Gamma$(\nu/2, 2)$ density given by

$$(u/2)^{(\nu-2)/2} e^{-u/2}/(2\Gamma(\nu/2)).$$

**Example**: The Cauchy distribution: The Cauchy density is

$$\frac{1}{\pi(1+x^2)};$$

the corresponding moment generating function is

$$M(t) = \int_{-\infty}^{\infty} \frac{e^{tx}}{\pi(1+x^2)} dx$$

which is $+\infty$ except for $t = 0$ where we get 1. *Every* $t$ distribution has exactly same moment generating function. So we cannot use moment generating functions to distinguish such distributions. The problem is that these distributions do not have infinitely many finite moments. So we now develop a substitute substitute for the moment generating function which is defined for every distribution, namely, the characteristic function.

### 0.0.3 Aside on complex arithmetic

Complex numbers are a fantastically clever idea. The idea is to imagine that $-1$ has a square root and see what happens. We add $i \equiv \sqrt{-1}$ to the real numbers. Then, we insist that all the usual rules of algebra are unchanged. So, if $i$ and any real numbers $a$ and $b$ are to be complex numbers then so must be $a + bi$. Now let us look at each of the arithmetic operations to see how they have to work:

- Multiplication: If we multiply a complex number $a+bi$ with $a$ and $b$ real by another such number, say $c + di$ then the usual rules of arithmetic (associative, commutative and distributive laws) require

$$\begin{aligned}(a+bi)(c+di) =& ac + adi + bci + bdi^2 \\ =& ac + bd(-1) + (ad+bc)i \\ =& (ac-bd) + (ad+bc)i\end{aligned}$$

  so this is precisely how we define multiplication.

- Addition: we follow the usual rules (commutative, associative and distributive laws) to get

$$(a+bi) + (c+di) = (a+c) + (b+d)i\,.$$

- Additive inverses:
$$-(a+bi) = -a + (-b)i.$$

  Notice that $0 + 0i$ functions as 0 – it is an additive identity. In fact we normally just write 0.

- Multiplicative inverses:

$$\begin{aligned}\frac{1}{a+bi} &= \frac{1}{a+bi}\frac{a-bi}{a-bi} \\ &= \frac{a-bi}{a^2-abi+abi-b^2i^2} = \frac{a-bi}{a^2+b^2}\,.\end{aligned}$$

- Division:

$$\frac{a+bi}{c+di} = \frac{(a+bi)}{(c+di)}\frac{(c-di)}{(c-di)} = \frac{ac-bd+(bc+ad)i}{c^2+d^2}\,.$$

This rule for clearing the complex number from the denominator is a perfect match for the technique taught in high school and used in calculus, for dealing with fractions involving $a + b\sqrt{c}$ in the denominator.

- You should now notice that the usual rules of arithmetic don't require any more numbers than
$$x + yi$$
where $x$ and $y$ are real. So the complex numbers $\mathbb{C}$ are just all these numbers.

- **Transcendental functions**: For real $x$ have $e^x = \sum x^k/k!$ and $e^{a+b} = e^a e^b$ so we want to insist that
$$e^{x+iy} = e^x e^{iy}\,.$$
The problem is how to compute $e^{iy}$?

- Remember $i^2 = -1$ so $i^3 = -i$, $i^4 = 1$ $i^5 = i^1 = i$ and so on. Then
$$\begin{aligned} e^{iy} =& \sum_0^\infty \frac{(iy)^k}{k!} \\ =& 1 + iy + (iy)^2/2 + (iy)^3/6 + \cdots \\ =& 1 - y^2/2 + y^4/4! - y^6/6! + \cdots \\ & + iy - iy^3/3! + iy^5/5! + \cdots \\ =& \cos(y) + i\sin(y) \end{aligned}$$

- We can thus write
$$e^{x+iy} = e^x(\cos(y) + i\sin(y))$$

- Identify $x + yi$ with the corresponding point $(x, y)$ in the plane.

- Picture the complex numbers as forming a plane.

- Now every point in the plane can be written in polar co-ordinates as $(r\cos\theta, r\sin\theta)$ and comparing this with our formula for the exponential we see we can write
$$x + iy = \sqrt{x^2 + y^2}\, e^{i\theta} = re^{i\theta}$$
for an angle $\theta \in [0, 2\pi)$.

- Multiplication revisited: if $x+iy = re^{i\theta}$ and $x'+iy' = r'e^{i\theta'}$ then when we multiply we get

$$(x+iy)(x'+iy') = re^{i\theta}r'e^{i\theta'} = rr'e^{i(\theta+\theta')}\,.$$

- We will need from time to time a couple of other definitions:
- **Definition**: The **modulus** of $x+iy$ is

$$|x+iy| = \sqrt{x^2+y^2}\,.$$

- **Definition**: The **complex conjugate** of $x+iy$ is $\overline{x+iy} = x-iy$.
- Some identities: $z = x+iy = re^{i\theta}$ and $z' = x'+iy' = r'e^{i\theta'}$.
- Then

$$z\overline{z} = x^2+y^2 = r^2 = |z|^2$$
$$\frac{z'}{z} = \frac{z'\overline{z}}{|z|^2} = rr'e^{i(\theta'-\theta)}$$
$$\overline{re^{i\theta}} = re^{-i\theta}.$$

### 0.0.4 Notes on calculus with complex variables

The rules for calculus with complex numbers are really very much like the usual rules. For example,

$$\frac{d}{dt}e^{it} = ie^{it}\,.$$

We will (mostly) be doing only integrals over the real line; the theory of integrals along paths in the complex plane is a very important part of mathematics, however.

**Fact**: (This fact is not used explicitly in course). If $f : \mathbb{C} \mapsto \mathbb{C}$ is differentiable then $f$ is analytic (has power series expansion).

### 0.0.5 Characteristic Functions

**Definition**: The characteristic function of a real random variable $X$ is

$$\phi_X(t) = \mathrm{E}(e^{itX})$$

where $i = \sqrt{-1}$ is the imaginary unit.

Since

$$e^{itX} = \cos(tX) + i\sin(tX)$$

we find that

$$\phi_X(t) = \mathrm{E}(\cos(tX)) + i\mathrm{E}(\sin(tX)) .$$

Since the trigonometric functions are bounded by 1 the expected values must be finite for all $t$. This is precisely the reason for using characteristic rather than moment generating functions in probability theory courses.

The characteristic function is called "characteristic" because if you know it you know the distribution of the random variable involved. That is what is meant in mathematics when we say something characterizes something else.

**Theorem 3** *For any two real random vectors $X$ and $Y$ (say $p$-dimensional) the following are equivalent:*

1. *$X$ and $Y$ have the same distribution, that is, for any (Borel) set $A \subset \mathbb{R}^p$ we have*
$$P(X \in A) = P(Y \in A) .$$
2. *$F_X(t) = F_Y(t)$ for all $t \in \mathbb{R}^p$.*
3. *$\phi_X(u) = \mathrm{E}(e^{iu^tX}) = \mathrm{E}(e^{iu^tY}) = \phi_Y(u)$ for all $u \in \mathbb{R}^p$.*

*Moreover, all these are implied if there is $\epsilon > 0$ such that for all $|t| \leq \epsilon$*

$$M_X(t) = M_Y(t) < \infty .$$