4 Transformations and Generating Functions

4.1 Distributions of functions of random variables

4.1.1 Sums of discrete random variables

Much of the classical theory of probability concerns sums of random variables. The number of heads in n tosses of a coin is one of the simplest such examples, but we shall encounter many situations which are more complicated than this. One particular complication is when the summands are dependent. The first stage in developing a systematic technique is to find a formula for describing the mass function of the sum $Z = X + Y$ of two discrete random variables with joint mass function $p_{(X,Y)}(\cdot,\cdot)$, that is, to find $\mathbb{P}(Z = z)$ for any z. It is easy to see that

$$\{Z = z\} = \cup_x(\{X = x\} \cap \{Y = z - x\}).$$

This is a disjoint union, and at most countably many of its contributors have non-zero probability. Therefore

$$\mathbb{P}(Z = z) = \Sigma_x \mathbb{P}(X = x, Y = z - x) = \Sigma_x p_{(X,Y)}(x, z - x).$$

If X and Y are independent, then

$$\mathbb{P}(Z = z) = p_Z(z) = \Sigma_x p_X(x) p_Y(z - x) = \Sigma_y p_X(z - y) p_Y(y).$$

Definition 1 ***Convolution.***

In the case where X and Y are independent, the above expression for the mass function of $Z = X + Y$ is called the convolution of the mass functions of X and Y, and is written as

$$p_Z = p_X * p_Y.$$

Example 1 *Let X and Y be independent shifted geometric random variables with common mass function*

$$p(k) = p(1-p)^k, \qquad k = 0, 1, 2, \ldots$$

What is the mass function of Z?

$$\begin{aligned}
\mathbb{P}(Z = z) &= \Sigma_k \mathbb{P}(X = k)\mathbb{P}(Y = z - k) \\
&= \Sigma_{k=0}^{z} p(1-p)^k p(1-p)^{z-k} \\
&= (z+1)p^2(1-p)^z, \qquad z = 0, 1, 2, \ldots.
\end{aligned}$$

4.1.2 Distributions of functions of continuous random variables

Let X_1 and X_2 be jointly continuous random variables with joint probability density function $f_{(X_1,X_2)}(\cdot,\cdot)$. Now assume that we have other random variables, which are, at most, functions of X_1 and X_2. Then we might want to know the joint pdf of the new variables. One simple example is to look at the sum of random variables and ask what is the distribution of the sum.

Transformation I: $Y = g(X)$

Example 2 *On many occasions we will be interested in calculating the distribution function of a random variable $Y = g(X)$ where $g : \mathbb{R} \longrightarrow \mathbb{R}$ is a sufficiently nice function. Let X be a random variable with probability density function $f_X(\cdot)$. If $g(\cdot)$ is monotone increasing in $\mathbb{R}$, we proceed thus:*

$$\begin{aligned}
F_Y(y) &= \mathbb{P}(Y \le y) = \mathbb{P}(g(X) \le y) \\
&= \mathbb{P}(X \le g^{-1}(y)) = F_X(g^{-1}(y)) \\
&= \int_{-\infty}^{g^{-1}(y)} f_X(x)dx = \int_c^y f_X(g^{-1}(u)).\left(g^{-1}(u)\right)' du
\end{aligned}$$

by making the substitution $x = g^{-1}(u)$. Here c may be a constant or $-\infty$ determined by $u = g(x)$ when $x \to -\infty$. Then the density function of Y is

$$\begin{aligned}
f_Y(y) &= \frac{d}{dy}\left(\int_c^y f_X(g^{-1}(u))\left(g^{-1}(u)\right)' du\right) \\
&= f_X(g^{-1}(y))\left(g^{-1}(y)\right)'.
\end{aligned}$$

If g is monotone decreasing in $\mathbb{R}$, we have

$$\begin{aligned}
\mathbb{P}(Y \leq y) &= \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \geq g^{-1}(y)) \\
&= 1 - \mathbb{P}(X \leq g^{-1}(y)) = 1 - \int_{-\infty}^{g^{-1}(y)} f_X(x)dx \\
&= 1 - \int_c^y f_X(g^{-1}(u)) \left(g^{-1}(u)\right)' du
\end{aligned}$$

by making the substitution $x = g^{-1}(u)$. Then the density of Y is

$$\begin{aligned}
f_Y(y) &= \frac{d}{dy}\big(1 - \int_c^y f_X(g^{-1}(u))\big(g^{-1}(u)\big)' du\big) \\
&= f_X(g^{-1}(y))\big(- g^{-1}(y)\big)'.
\end{aligned}$$

So we have the following result: if g is a monotone function (increasing or decreasing), the density function of Y is

$$f_Y(y) = f_X(g^{-1}(y)) \left|\left(g^{-1}(y)\right)'\right|.$$

Here $|\cdot|$ means absolute value.

Example 3 *Let $X \sim N(0, 1)$ and let $g(x) = x^2$. What is the distribution of $Y = g(X) = X^2$?*

$$\begin{aligned}
\mathbb{P}(Y \leq y) &= \mathbb{P}(X^2 \leq y) \\
&= \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) \\
&= \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) \\
&= 2\Phi(\sqrt{y}) - 1 \quad for \quad y \geq 0.
\end{aligned}$$

Differentiation yields

$$\begin{aligned}
f_Y(y) &= 2\frac{d}{dy}\Phi(\sqrt{y}) \\
&= \frac{1}{\sqrt{y}}\phi(\sqrt{y}) \\
&= \frac{1}{\sqrt{2\pi y}}e^{-\frac{1}{2}y}
\end{aligned}$$

for $y \geq 0$. Note that X^2 is Gamma$(1/2, 1/2)$ or chi-squared with one degree of freedom.

Transformation II: $Y = g(X_1, X_2)$

When studying phenomena such as the evolution of stock prices we will mostly be interested in the distribution of the sum of many random variables; therefore we have to develop the tools to understand the 'sum' of many events. We start by looking at the distribution of the sum of two random variables.

Theorem 1 *If X and Y have joint density function $f_{(X,Y)}(x,y)(\cdot,\cdot)$, then $Z = X+Y$ has density function*

$$f_Z(z) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, z-x)dx.$$

Moreover, if they are independent we have

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)dx = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)dy$$

in which case, the function f_Z, also denoted by f_{X+Y}, is called the convolution of f_X and f_Y and is written as

$$f_{X+Y} = f_X * f_Y.$$

Proof.

$$\begin{aligned}
F_Z(z) = \mathbb{P}(Z \leq z) = \mathbb{P}(X+Y \leq z) &= \int\int_{x+y\leq z} f_{(X,Y)}(x,y)dxdy \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-x} f_{(X,Y)}(x,y)dy \right) dx \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z} f_{(X,Y)}(x,u-x)du \right) dx \\
&= \int_{-\infty}^{z} \left(\int_{-\infty}^{\infty} f_{(X,Y)}(x,u-x)dx \right) du
\end{aligned}$$

by making the substitution $y = u - x$ and changing the order of integration. Now

$$\begin{aligned}
f_Z(z) = \frac{dF_Z(z)}{dz} &= \frac{d}{dz}\left\{ \int_{-\infty}^{z} \left(\int_{-\infty}^{\infty} f_{(X,Y)}(x,u-x)dx \right) du \right\} \\
&= \int_{-\infty}^{\infty} f_{(X,Y)}(x, z-x)dx.
\end{aligned}$$

Example 4 *Let X and Y be independent $N(0,1)$ random variables. What is the distribution of $Z = X + Y$? Hint: you may consider 'completing the square' in simplifying a particular exponent that crops up in the calculation.*

Transformation III: $(X_1, X_2) \to (Y_1, Y_2)$

Let X_1 and X_2 be jointly continuous random variables with probability density function $f_{(X_1,X_2)}(\cdot,\cdot)$. Suppose that $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$ for some functions g_1 and g_2. The joint distribution function of Y_1 and Y_2 are

$$\begin{aligned} F_{(Y_1,Y_2)}(y_1, y_2) &= \mathbb{P}(Y_1 \le y_1, Y_2 \le y_2) \\ &= \int\int_D f_{(X_1,X_2)}(x_1, x_2) dx_1 dx_2 \end{aligned}$$

where $D = \{(x_1, x_2) : g_1(x_1, x_2) \le y_1; g_2(x_1, x_2) \le y_2\}$. Now assume that the functions g_1 and g_2 satisfy the following conditions:

1. the set of equations
$$\begin{aligned} y_1 &= g_1(x_1, x_2) \\ y_2 &= g_2(x_1, x_2) \end{aligned}$$
can be uniquely solved for x_1 and x_2 in terms of y_1 and y_2 with solutions given by, say, $x_1 = x_1(y_1, y_2)$ and $x_2 = x_2(y_1, y_2)$;

2. the functions g_1 and g_2 have continuous partial derivatives at all points (x_1, x_2) and are such that the following determinant
$$J(y_1, y_2) = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = \frac{\partial x_1}{\partial y_1}\frac{\partial x_2}{\partial y_2} - \frac{\partial x_2}{\partial y_1}\frac{\partial x_1}{\partial y_2} \neq 0$$
at all points (y_1, y_2).

Then under these two conditions it can be shown that the random variables Y_1 and Y_2 are jointly continuous with joint probability density function given by

$$f_{(Y_1,Y_2)}(y_1, y_2) = f_{(X_1,X_2)}(x_1, x_2)|J(y_1, y_2)|$$

where $x_1 = x_1(y_1, y_2)$ and $x_2 = x_2(y_1, y_2)$.

Remark: The determinant J is called the *Jacobian determinant* of (x_1, x_2) with respect to (y_1, y_2), where it is understood that $(x_1, x_2) = \big(x_1(y_1, y_2), x_2(y_1, y_2)\big)$.

Transformation IV: $(X_1, X_2, \cdots, X_n) \to (Y_1, Y_2, \cdots, Y_n)$

All this generalizes to any number of dimensions: if $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_n) = g(X_1, \ldots, X_n)$, where $g = (g_1, \ldots, g_n) : \mathbb{R}^n \to \mathbb{R}^n$, then

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{x}(\mathbf{y}))|J_{\mathbf{x}}(\mathbf{y})|,$$

where

$$J_{\mathbf{x}}(\mathbf{y}) = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & & \vdots \\ \frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}.$$

Example 5 *Let X_1 and X_2 be jointly continuous random variables with probability density function $f_{(X_1,X_2)}(\cdot,\cdot)$. Let $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$. Find the joint density function of Y_1 and Y_2 in terms of $f_{(X_1,X_2)}(\cdot,\cdot)$.*

First we note that

$$x_1 = \frac{y_1 + y_2}{2} \qquad \text{and} \qquad x_2 = \frac{y_1 - y_2}{2}.$$

The Jacobian determinant, J, is equal to $-1/2$, therefore we have that

$$f_{(Y_1,Y_2)}(y_1, y_2) = \frac{1}{2} f_{(X_1,X_2)}\left(\frac{y_1 + y_2}{2}, \frac{y_1 - y_2}{2}\right).$$

Example 6 *Let X and Y be random variables with joint probability density function given by*

$$f_{(X,Y)}(x, y) = 4xy, \quad 0 < x < 1, \quad 0 < y < 1.$$

Let $T = X^2$ and $W = XY$. Find the joint p.d.f. of T and W.

First we establish the range of (T, W), i.e. $R_{(T,W)}$. Note that $X = \sqrt{T}$ and so for each t if $T = t$ then $W = \sqrt{T}Y < \sqrt{t}$ (since $Y < 1$). T can take any value in $(0, 1)$ and so

$$R_{(T,W)} = \{(t, w) : \, t \in (0, 1), \, w \in (0, \sqrt{t})\}.$$

We now find $x = x(t,w)$ and $y = y(t,w)$. We have $x = \sqrt{t}$ and $w = xy$ so $y = w/x = w/\sqrt{t}$. Thus $\partial x/\partial t = \frac{1}{2}t^{-1/2}$, $\partial x/\partial w = 0$, $\partial y/\partial t = -\frac{1}{2}wt^{-3/2}$, $\partial y/\partial w = t^{-1/2}$. So the Jacobian is equal to $1/(2t)$ and we obtain

$$f_{(T,W)}(t,w) = f_{(X,Y)}(x,y)\frac{1}{2t} = 4x(t,w)y(t,w)\frac{1}{2t} = 2w/t,$$

for $(t,w) \in R_{(T,W)}$.

4.2 Generating Functions

4.2.1 Probability generating function

Definition 2 ***Probability generating function.***

Let X be a discrete random variable taking values in $0, 1, 2\ldots$. The probability generating function $G_X(\cdot)$ of X is defined as

$$G_X(s) = \mathbb{E}[s^X] = \sum_{n=0}^{\infty} s^n \mathbb{P}(X = n) = \sum_{n=0}^{\infty} s^n p_n \qquad \text{for} \quad -1 \le s \le 1.$$

The last term $\sum_{n=0}^{\infty} s^n p_n$ is a power series. Note that $G_X(1) = 1$, so the power series converges to $G_X(s)$ uniformly and absolutely when $|s| \le 1$. So the probability generating function always exists for the random variables only taking non-negative integers.

Remark: 1. By interchanging the operations of differentiation and expectation we have

$$G'_X(s) = \frac{d}{ds}\mathbb{E}[s^X] = \mathbb{E}\left[\frac{d}{ds}s^X\right] = \mathbb{E}\left[Xs^{X-1}\right] = \sum_{n=1}^{\infty} ns^{n-1}\mathbb{P}(X = n)$$

and similarly

$$G''_X(s) = \frac{d^2}{ds^2}\mathbb{E}[s^X] = \sum_{n=2}^{\infty} n(n-1)s^{n-2}\mathbb{P}(X = n)$$

for $-1 \leq s \leq 1$. Thus

$$\begin{aligned} G'_X(1) &= \sum_{n=1}^{\infty} n\mathbb{P}(X=n) = \mathbb{E}[X] \\ G''_X(1) &= \sum_{n=2}^{\infty} n(n-1)\mathbb{P}(X=n) = \mathbb{E}[X(X-1)] = \mathbb{E}[X^2] - \mathbb{E}[X], \end{aligned}$$

so we have that

$$\text{var}(X) = G''_X(1) + G'_X(1) - (G'_X(1))^2.$$

2. We can extend the definition of PGF to functions of X. The PGF of $Y = g(X)$ is

$$G_Y(s) = G_{g(X)}(s) = \mathbb{E}[s^{g(X)}] = \sum_{n=0}^{\infty} \mathbb{P}(X=n)s^{g(n)}.$$

Example 7

1. Bernoulli: $G_X(s) = q + ps$;
2. Binomial: $G_X(s) = (q+ps)^n$;
3. Poisson: $G_X(s) = e^{\lambda(s-1)}$;
4. Geometric: $G_X(s) = \frac{ps}{1-qs}$ if $|s| < q^{-1}$.

Theorem 2 *Uniqueness*

Let X and Y have PGFs $G_X(\cdot)$ and $G_Y(\cdot)$. Then $G_X(\cdot) = G_Y(\cdot)$ iff

$$\mathbb{P}(X=k) = \mathbb{P}(Y=k) \quad for\ k = 0, 1, 2, \cdots$$

From the definition above, it is obvious that probability functions determine generating functions. This theorem tells us that generating functions also determine probability functions. For example, if X has PGF $\frac{ps}{1-qs}$ with $q = 1-p$, then we conclude that $X \sim$ *Geometric*(p).

Theorem 3 Sum of a fixed number of independent random variables

Let X and Y be independent with PGFs $G_X(\cdot)$ and $G_Y(\cdot)$ respectively, and let $Z = X + Y$. Then, for each s for which the PGFs exist,

$$G_Z(s) = G_{X+Y}(s) = G_X(s)G_Y(s).$$

If $X_1, \cdots, X_n$ are independent with PGFs $G_{X_1}(s), \cdots, G_{X_n}(s)$ respectively (and n is a known integer), then, for each s for which the PGFs exist,

$$G_{X_1+\cdots+X_n}(s) = G_{X_1}(s)\cdots G_{X_n}(s).$$

Example 8 *Find the distribution of the sum of n independent random variables X_i, $i = 1, \cdots, n$, where $X_i \sim Poisson(\lambda_i)$.*

Solution:

$$G_{X_i}(s) = e^{\lambda_i(s-1)}.$$

So

$$G_{X_1+X_2+\cdots+X_n}(s) = \prod_{i=1}^{n} e^{\lambda_i(s-1)} = e^{(\lambda_1+\cdots+\lambda_n)(s-1)}.$$

This is the PGF of a poisson random variable, i.e,

$$\sum_{i=1}^{n} X_i \sim Poisson(\sum_{i=1}^{n} \lambda_i).$$

Theorem 4 Sum of a random number of independent random variables

Let $N, X_1, X_2, \cdots,$ be independent. If the $\{X_i\}$ are identically distributed, each with PGF $G_X(\cdot)$, then

$$S_N = X_1 + \cdots + X_N$$

has PGF

$$G_{S_N}(s) = G_N(G_X(s))$$

for each s for which the PGFs exist.

***Proof**: We have*

$$\begin{aligned} G_{S_N}(s) &= \mathbb{E}[s^{S_N}] \\ &= \mathbb{E}[\mathbb{E}[s^{S_N}|N]] \\ &= \sum_{n=0}^{\infty} \mathbb{E}[s^{S_N}|N=n]\mathbb{P}(N=n) \quad (conditioning\ on\ N) \\ &= \sum_{n=0}^{\infty} \mathbb{E}[s^{S_n}]\mathbb{P}(N=n) \\ &= \sum_{n=0}^{\infty} G_{X_1+\cdots+X_n}(s)\mathbb{P}(N=n) \\ &= \sum_{n=0}^{\infty} (G_X(s))^n\mathbb{P}(N=n) \\ &= G_N(G_X(s)) \quad by\ definition\ of\ G_N. \end{aligned}$$

Example 9 *A cat delivers a litter of N kittens, where $N \sim Poisson(\lambda)$. Each kitten is male with probability p, independently of the other kittens. Find the probability distribution of the number of male kittens, Z.*

***Solution**: We have*

$$Z = X_1 + \cdots + X_N,$$

where $X_i, i = 1, 2, \cdots$ are independent Bernoulli random variables with parameter p. Then

$$G_N(s) = e^{\lambda(s-1)}, \quad G_X(s) = q + ps.$$

So

$$G_Z(s) = G_N(G_X(s)) = e^{\lambda p(s-1)},$$

which is the PGF of a Poisson random variable, i.e, $Z \sim Poisson(\lambda p)$.

4.2.2 Moment Generating Function

Definition 3 ***Moment generating function.*** *For a random variable X, the moment generating function (mgf) is the function $M_X : \mathbb{R} \longrightarrow [0, \infty)$ given by*

$$M_X(\theta) = \mathbb{E}[e^{\theta X}].$$

More explicitly,

$$M_X(\theta) = \begin{cases} \sum_{n=1}^{\infty} p_n e^{\theta x_n} \text{ if } X \text{ discrete, } \mathbb{P}(X = x_n) = p_n, \\ \\ \int_{-\infty}^{\infty} f_X(x) e^{\theta x} \, dx \text{ if } X \text{ continuous with pdf } f_X(\cdot). \end{cases}$$

Note that since $e^{\theta X} \geq 0$, $M_X(\theta) \geq 0$. When $M_X(\theta) = \infty$, we say $M_X(\theta)$ doesn't exist.

Remark: If X has the generating function $M_X(\theta)$, then the generating function of $Y = g(X)$ can be found in the following way:

$$M_Y(\theta) = \mathbb{E}[e^{\theta Y}] = \mathbb{E}[e^{\theta g(X)}],$$

A simple and useful example is when $Y = aX + b$ where a and b are constants: the moment generating function is

$$M_Y(\theta) = \mathbb{E}[e^{\theta Y}] = \mathbb{E}[e^{\theta(aX+b)}] = e^{b\theta}\mathbb{E}[e^{a\theta X}] = e^{b\theta} M_X(a\theta).$$

Example 10 *1. If $X \sim Bin(n, p)$ and $q = 1 - p$, then $M_X(\theta) = (q + pe^{\theta})^n$.*

2. If $X \sim Poission(\lambda)$, then $M_X(\theta) = e^{\lambda(e^{\theta}-1)}$.

3. If $X \sim N(0, 1)$, then $M_X(\theta) = e^{\frac{1}{2}\theta^2}$; Moreover, if $Y \sim N(\mu, \sigma^2)$ then

$$M_Y(\theta) = e^{\mu\theta + \frac{1}{2}\sigma^2\theta^2}$$

Solution*: We only show (3), the rest being left as an exercise.*

If X is $N(0,1)$ then

$$\begin{aligned}
M_X(\theta) = \mathbb{E}[e^{\theta X}] &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(\theta x - \frac{1}{2}x^2) dx \\
&= \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}\theta^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-\theta)^2} dx \\
&= \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}\theta^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}u^2} du \quad (u = x - \theta) \\
&= e^{\frac{1}{2}\theta^2} \quad \left(\text{since } \int_{-\infty}^{\infty} e^{-\frac{1}{2}u^2} du = \sqrt{2\pi}\right)
\end{aligned}$$

If Y is $N(\mu, \sigma^2)$, i.e, $Y = \sigma X + \mu$, then

$$\begin{aligned}
M_Y(\theta) &= \mathbb{E}[e^{\theta Y}] = \mathbb{E}[e^{\theta(\sigma X + \mu)}] \\
&= e^{\theta\mu}\mathbb{E}[e^{\sigma\theta X}] = e^{\theta\mu} M_X(\sigma\theta) \\
&= e^{\mu\theta + \frac{1}{2}\sigma^2\theta^2}.
\end{aligned}$$

The moment generating function is used to generate moments (i.e. expectations of powers of X) and to identify the distribution of X. Its properties are given below without proof.

Uniqueness

If X and Y have the same moment generating function, then X and Y have the same distribution.

Moments

Let X be a random variable with moment generating function $M_X(\theta) = \mathbb{E}[e^{\theta X}] < \infty$ for θ in an interval containing the origin. Then, by expanding $e^{\theta X}$ in its power series, and interchanging integration and summation

$$\begin{aligned}
M_X(\theta) &= \mathbb{E}[e^{\theta X}] \\
&= \mathbb{E}[1 + \theta X + \frac{1}{2!}\theta^2 X^2 + \cdots] \\
&= 1 + \mu_1\theta + \frac{1}{2!}\mu_2\theta^2 + \cdots,
\end{aligned}$$

where μ_k is the kth moment of X, i.e, $\mu_r = \mathbb{E}[X^r]$ for $r = 1, 2, \ldots$. We recover the moments by the operation of differentiation, i.e,

$$\mu_k = \frac{\partial^k}{\partial \theta^k} M_X(\theta)|_{\theta=0}.$$

Remark: 1. From above, X has moments of all orders if $M_X(\theta)$ exists. It may be that the moment generating function does not exist, because some of the moments may be infinite. For example, the Cauchy distribution doesn't have any moments. Also, even if the moments are all finite and have definite values, the generating function may not converge for any value of θ other than 0. For example, the log-normal distribution has $M_X(\theta) = \infty$.

2. It is not always the case that if all moments of two distributions are the same, then the distributions must also be the same. This happens to log-normal distribution. See Remark 10, p.184 in 3rd edition of Grimmett and Stirzaker, for example.

Example 11 *Suppose X has Gamma(α, β) distribution, i.e,*

$$f_X(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}, \quad for\ x > 0,$$

where $\alpha > 0$ and $\beta > 0$. Find $M_X(\theta)$, $\mathbb{E}[X]$ and $\mathrm{var}(X)$.

Solution:

$$\begin{aligned} M_X(\theta) = \mathbb{E}[e^{\theta X}] &= \int_0^\infty e^{\theta x} \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} dx \\ &= \int_0^\infty \frac{\beta^\alpha x^{\alpha-1} e^{-(\beta-\theta)x}}{\Gamma(\alpha)} dx \\ &= \int_0^\infty \frac{\beta^\alpha u^{\alpha-1} e^{-u}}{(\beta-\theta)^\alpha \Gamma(\alpha)} du \quad \Big((\beta-\theta)x = u; \theta < \beta\Big) \\ &= \frac{\beta^\alpha}{(\beta-\theta)^\alpha \Gamma(\alpha)} \int_0^\infty u^{\alpha-1} e^{-u} du \\ &= (\frac{\beta}{\beta-\theta})^\alpha \quad for\ \theta < \beta. \end{aligned}$$

It can be verified that

$$\frac{d}{d\theta} M_X(\theta) = \alpha\beta^\alpha (\beta-\theta)^{-(\alpha+1)}; \quad \frac{d^2}{d\theta^2} M_X(\theta) = \alpha\beta^\alpha (\alpha+1)(\beta-\theta)^{-(\alpha+2)}.$$

Then we get

$$\mathbb{E}[X] = \frac{d}{d\theta} M_X(\theta)|_{\theta=0} = \frac{\alpha}{\beta}; \quad \mathbb{E}[X^2] = \frac{d^2}{d\theta^2} M_X(\theta)|_{\theta=0} = \frac{\alpha(\alpha+1)}{\beta^2}.$$

Therefore

$$\mathrm{var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{\alpha}{\beta^2}.$$

Independence

If X and Y are independent random variables then

$$M_{X+Y}(\theta) = M_X(\theta) M_Y(\theta)$$

for all θ for which both $M_X(\theta)$ and $M_Y(\theta)$ are defined.

This follows from the fact that e^X and e^Y will also be independent, and hence

$$\mathbb{E}[e^{\theta(X+Y)}] = \mathbb{E}[e^{\theta X} e^{\theta Y}] = \mathbb{E}[e^{\theta X}]\, \mathbb{E}[e^{\theta Y}].$$

Similarly, if $X_1, X_2, \cdots, X_n$ are independent random variables, then

$$M_{X_1+\cdots+X_n}(\theta) = M_{X_1}(\theta) \cdots M_{X_n}(\theta).$$

Example 12 *Assume that $X_1, \cdots, X_n$ are independent random variables with $X_i \sim N(\mu_i, \sigma_i^2)$ and $a_1, \cdots, a_n$ are constants, then*

$$Y = \sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right).$$

***Solution**:*

$$\begin{aligned}
M_Y(\theta) = \mathbb{E}[e^{\theta Y}] &= \mathbb{E}[e^{\theta(\sum_{i=1}^{n} a_i X_i)}] \\
&= \mathbb{E}[e^{\theta a_1 X_1}] \cdots \mathbb{E}[e^{\theta a_n X_n}] \quad (a_1 X_1, \cdots, a_n X_n \textit{ are independent}) \\
&= M_{X_1}(a_1\theta) \cdots M_{X_n}(a_n\theta) \\
&= e^{\mu_1(a_1\theta) + \frac{1}{2}\sigma_1^2(a_1\theta)^2} \cdots e^{\mu_n(a_n\theta) + \frac{1}{2}\sigma_n^2(a_n\theta)^2} \\
&= e^{(\sum a_i\mu_i)\theta + \frac{1}{2}(\sum a_i^2\sigma_i^2)\theta^2}.
\end{aligned}$$

4.2.3 Characteristic Function

To get around the problem that $M_X(\theta)$ need not always be defined, we replace θ by the complex number $i\theta$ ($i = \sqrt{-1}$):

Definition 4 ***Characteristic Function.***

The characteristic function $\Psi_X : \mathbb{R} \mapsto \mathbb{C}$ of a random variable X is defined by

$$\Psi_X(\theta) = \mathbb{E}[e^{i\theta X}].$$

The point here is that since the absolute value $|e^{i\theta}| = 1$, the random variable of which we take the expectation is bounded, and therefore the expectation will always exist. In general $\Psi_X(\theta)$ is a complex-valued function of θ. Obviously $\Psi_X(0) = 1$.

Remark: 1. Suppose X has the characteristic function $\Psi_X(\theta)$ and a and b are constants, then the characteristic function of $Y = aX + b$ is

$$\begin{aligned}\Psi_Y(\theta) &= \mathbb{E}[e^{i\theta Y}] = \mathbb{E}[e^{i\theta(aX+b)}] \\ &= e^{ib\theta}\mathbb{E}[e^{ia\theta X}] = e^{ib\theta}\Psi_X(a\theta)\end{aligned}$$

2. Suppose X has moment generating function $M_X(\theta)$ that satisfies $M_X(\theta) < \infty$ for θ in some interval I about 0. Then the characteristic function of X satisfies $\Psi_X(\theta) = M_X(i\theta)$ for $\theta \in I$.
3. It is often hard to calculate the characteristic function of a continuous random variable as the standard method involves using contour integrals and complex analysis. In some instances there are alternative simpler (but perhaps non-intuitive) approaches.

Example 13 *Normal distribution*
If X is $N(0,1)$ then

$$\begin{aligned}\Psi_X(\theta) = \mathbb{E}[e^{i\theta X}] &= \int_{-\infty}^{\infty} e^{i\theta x}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\cos(\theta x)e^{-x^2/2}dx + i\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\sin(\theta x)e^{-x^2/2}dx.\end{aligned}$$

For the term involving sine we have

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \sin(\theta x) e^{-x^2/2} dx = \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} \sin(\theta x) e^{-x^2/2} dx + \int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} \sin(\theta x) e^{-x^2/2} dx.$$

Making a change of variable in the first integral ($y = -x$) we have

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \sin(\theta x) e^{-x^2/2} dx &= \int_{\infty}^{0} -\frac{1}{\sqrt{2\pi}} \sin(-\theta y) e^{-y^2/2} dy + \int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} \sin(\theta x) e^{-x^2/2} dx \\ &= \int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} (\sin(-\theta x) + \sin(\theta x)) e^{-x^2/2} dx = 0. \end{aligned}$$

Thus we obtain

$$\Psi_X(\theta) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cos(\theta x) e^{-x^2/2} dx.$$

We now differentiate both sides with respect to θ to get

$$\Psi_X'(\theta) = \int_{-\infty}^{\infty} -\frac{x}{\sqrt{2\pi}} \sin(\theta x) e^{-x^2/2} dx$$

and integrating by parts gives

$$\Psi_X'(\theta) = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \theta \cos(\theta x) e^{-x^2/2} dx = -\theta \Psi_X(\theta),$$

i.e. we have $\Psi_X'(\theta)/\Psi_X(\theta) = -\theta$, i.e.

$$\frac{d}{d\theta} \log \Psi_X(\theta) = -\theta \implies \log \Psi_X(\theta) = -\frac{1}{2}\theta^2 + c,$$

for some constant c to be determined. Hence $\Psi_X(\theta) = e^C e^{-\frac{1}{2}\theta^2}$, and since $\Psi_X(0) = 1$ we deduce that $e^C = 1$ which gives $\Psi_X(\theta) = e^{-\frac{1}{2}\theta^2}$.

If Y is $N(\mu, \sigma^2)$, i.e, $Y = \sigma X + \mu$, then

$$\begin{aligned} \Psi_Y(\theta) &= \mathbb{E}[e^{i\theta Y}] = \mathbb{E}[e^{i\theta(\sigma X + \mu)}] \\ &= e^{i\theta\mu} \mathbb{E}[e^{i\sigma\theta X}] = e^{i\theta\mu} \Psi_X(\sigma\theta) \\ &= e^{i\mu\theta - \frac{1}{2}\sigma^2\theta^2} \end{aligned}$$

The following lists several important properties of characteristic functions. We skip the proofs, some of which are rather deep and subtle, and usually done in a course on Fourier Analysis.

Uniqueness

X and Y have the same characteristic function if and only if they have the same distribution.

Inversion: If X is a continuous random variable with pdf $f_X(\cdot)$, then

$$\Psi_X(\theta) = \int_{-\infty}^{\infty} f_X(x) e^{i\theta x}\, dx,$$

which some of the readers may recognize as being the so-called *Fourier transform* of $f_X(\cdot)$. By a famous result in Analysis, there is a formula called the *Fourier inversion formula* which expresses $f_X(\cdot)$ in terms of the function $\Psi_X(\cdot)$, by a similar integral:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-i\theta x} \Psi_X(\theta) d\theta.$$

This is a non-trivial result, whose proof we will skip. We note that, as a consequence, the characteristic function completely determines the pdf.

Moments

As with the other generating functions, the characteristic function can be used to find the moments of X. Moreover, this can be done even if only some of the moments exist.

1. If $\frac{d^k \Psi(\theta)}{d\theta^k}|_{\theta=0} \equiv \Psi^{(k)}(0)$ exists then

   $$\mathbb{E}[|X^k|] < \infty \qquad \text{if } k \text{ is even,}$$

   and

   $$\mathbb{E}[|X^{k-1}|] < \infty \qquad \text{if } k \text{ is odd.}$$

2. If $\mathbb{E}[|X^n|] < \infty$ then

   $$\Psi_X(\theta) = \sum_{k=0}^{n} \frac{\mathbb{E}[X^k]}{k!} (i\theta)^k + o(\theta^n)$$

   and therefore $\Psi_X^{(n)}(0) = i^n \mathbb{E}[X^n]$.

   Note: $o(\theta^n)$ is just a term such that $o(\theta^n)/\theta^n$ goes to 0 as $\theta \to 0$.

Independence

Suppose that X and Y are independent, real-valued random variables with characteristic functions $\Psi_X(\theta)$ and $\Psi_Y(\theta)$ respectively. Then, for any θ, the characteristic function of $X + Y$ is

$$\Psi_{X+Y}(\theta) = \Psi_X(\theta)\Psi_Y(\theta).$$

Characteristic functions can also be introduced for pairs of random variables (and more generally, vectors of random variables); in that case we will obtain a function of two auxiliary variables, θ and ξ (say):

Definition 5 ***Joint Characteristic Function***

The joint characteristic function of X and Y, for any θ and ξ, is the function

$$\Psi_{(X,Y)}(\theta, \xi) = \mathbb{E}[e^{i\theta X + i\xi Y}].$$