3 Mappings and Generating Functions of Random Variables

3.1 Functions of a r.v.

If $X$ is a r.v. on some underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $g : \mathbb{R} \to \mathbb{R}$, then we can also define a r.v. $Y$ on $(\Omega, \mathcal{F}, \mathbb{P})$ by setting $Y(\omega) = g(X(\omega))$ for $\omega \in \Omega$.

Example 3.1.1

$X$ is the lifetime of a randomly chosen fluorescent tube.
Running cost per unit time is £$a$.
Purchase cost is £$b$.
Total cost (in £) of operating the tube is

$$aX + b = g(X) = Y.$$ 

In order to ask questions about the r.v. $Y$, we need to find its distribution; as we saw in the previous chapter, this task can sometimes be bypassed by utilizing the Law of the Unconscious Statistician. Here, we will be concerned with finding the distribution of $Y$ for certain special cases.

Example 3.1.2

Suppose $X$ is a continuous r.v. with p.d.f. $f_X$, and let $Y = aX + b$, for constants $a, b \in \mathbb{R}$, $a \neq 0$. Then

$$f_Y(y) = \frac{1}{|a|}f_X\left(\frac{y - b}{a}\right) = y \in R_Y = \mathbb{R}.$$ 

Procedure

The general procedure used in this proof is quite important to understand. It runs as follows:
(i) Find the distribution function of $Y$, in terms of the distribution function of $X$,
(ii) Find the p.d.f. of $Y$ via differentiation.

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(aX + b \leq y) = \begin{cases} \mathbb{P}(X \leq \frac{y-b}{a}) & a > 0 \\ \mathbb{P}(X \geq \frac{y-b}{a}) & a < 0 \end{cases}$$
\[
\begin{aligned}
= \begin{cases} 
F_X \left( \frac{y-b}{a} \right) & a > 0 \\
1 - F_X \left( \frac{y-b}{a} \right) & a < 0 
\end{cases}
\end{aligned}
\]

Hence, by using the chain rule, we find that
\[
f_Y(y) = \frac{d}{dy} F_Y(y) = \begin{cases} 
\frac{1}{a} f_X \left( \frac{y-b}{a} \right) & a > 0 \\
-\frac{1}{a} f_X \left( \frac{y-b}{a} \right) & a < 0 
\end{cases}
= \frac{1}{|a|} f_X \left( \frac{y-b}{a} \right).
\]

**Example 3.1.3**

Suppose that \( X \) has a p.d.f. \( f_X \), and let \( Y = X^2 \). Then
\[
f_Y(y) = \frac{1}{2\sqrt{y}} (f_X(\sqrt{y}) + f_X(-\sqrt{y})) \quad y \geq 0.
\]

**Procedure**

\( Y \geq 0 \Rightarrow f_Y(y) = 0 \) for \( y < 0 \), so assume \( y \geq 0 \).

\[
F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y})
= P(X \leq \sqrt{y}) - P(X \leq -\sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}).
\]

Therefore,
\[
f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X(\sqrt{y}) - \frac{d}{dy} F_X(-\sqrt{y})
= \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) - \left( -\frac{1}{2\sqrt{y}} f_X(-\sqrt{y}) \right)
\]

We can make use of the following, more general, result in certain situations.

**Proposition 3.1.4**

Suppose \( g : \mathbb{R} \rightarrow \mathbb{R} \) is strictly monotone increasing (or decreasing) and differentiable on \( R_X \), with \( Y = g(X) \), where \( X \) is a continuous random variable. Then
\[
f_Y(y) = |(g^{-1})'(y)| f_X(g^{-1}(y))
= |h'(y)| f_X(h(y))
\]

where \( h(\cdot) = g^{-1}(\cdot) \).

**Proof**

\( g \) could either be strictly increasing on all of \( R_X \), or strictly decreasing on all of \( R_X \); this ensures that the inverse function \( h = g^{-1} \) exists.

First, consider the case of \( g \) increasing.
\[ F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \leq g^{-1}(y)) = F_X(g^{-1}(y)) = F_X(h(y)). \]

Therefore,
\[
f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X(h(y))
= f_X(h(y)) \times h'(y)
\] (1)

where the final equality follows from the chain rule for differentiation.

Now suppose that \( g \) is decreasing.

\[ F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \geq g^{-1}(y)) = \mathbb{P}(X \geq h(y)) = 1 - F_X(h(y)). \]

Therefore,
\[
f_Y(y) = -h'(y) f_X(h(y)).
\] (2)

In the first case, \( h(y) = g^{-1}(y) \) is increasing in \( y \), and so \( h'(y) \geq 0 \) in (1); in the latter case, \( h(y) \) is decreasing in \( y \), and so \( h'(y) \leq 0 \) in (2).

Hence,
\[
f_Y(y) = |h'(y)| f_X(h(y)).
\]

**Remarks 3.1.5**

Regrettably, we cannot apply this result to our previous example if \( R_X = \mathbb{R} \), since \( y = g(x) = x^2 \) is NOT strictly monotone there. However, if \( R_X = \{ x : x \geq 0 \} \), then we can apply the result, since \( y = g(x) = x^2 \) is strictly increasing there.

**Example 3.1.6 (Example 3.1.2 revisited)**

\( y = g(x) = ax + b \) is \( \nearrow \) for \( a > 0 \), and \( \searrow \) for \( a < 0 \).

\[ x = g^{-1}(y) = h(y) = \frac{y - b}{a}, \quad \text{and} \quad (g^{-1})'(y) = h'(y) = \frac{1}{a}. \]

Hence,
\[
f_Y(y) = \frac{1}{|a|} f_X \left( \frac{y - b}{a} \right) = \frac{1}{|a|} f_X \left( \frac{y - b}{a} \right).
\]
3.2 Generating Functions

The calculation of the mean, variance, and various other quantities can be facilitated via the use of generating functions. We will focus attention on two such types of function.

3.2.1 Probability Generating Functions

Definition 3.2.1 (p.g.f.)

Suppose \( X \) is a discrete r.v. with p.m.f \( p_X(\cdot) \) and range \( R_X \). Then the PROBABILITY GENERATING FUNCTION (p.g.f) of \( X \) is

\[
G_X(z) = E[z^X] = \sum_k z^k p_X(k)
\]

provided that the R.H.S exists.

Remarks 3.2.2

(i) In fact, it can be shown that \( G_X(z) \) exists and that its \( n \)-th derivative exists for \( |z| \leq 1 \) provided that \( X \) takes values only in the set \( \{0, 1, 2, \ldots\} \).

(ii) The coefficients in the power series expansion of \( G_X(z) \) in \( z \) are unique.

(iii) The p.g.f. quite rightly deserves its name because

\[
G_X(z) = z^0 p_X(0) + z^1 p_X(1) + z^2 p_X(2) + \ldots
\]

We see that the probabilities of the distribution of \( X \) are indeed generated! \( G_X(z) \) compresses the information \( \{ p_X(0), p_X(1), \ldots \} \) into a single function.

(iv) Notice that the p.g.f. takes a mapping, \( p_X(\cdot) \), as its input, and produces another mapping, \( G_X(\cdot) \), as its output.

We differentiate \( G_X(z) \) term by term to give:

\[
G_X'(z) = \frac{d}{dz} \left\{ \sum_{k=0}^{\infty} z^k p_X(k) \right\} = \sum_{k=0}^{\infty} \frac{d}{dz} \{ z^k p_X(k) \} = \sum_{k=1}^{\infty} k z^{k-1} p_X(k).
\]

Also,

\[
G_X''(z) = \sum_{k=2}^{\infty} k(k-1) z^{k-2} p_X(k).
\]

Restricting attention to those p.m.f.’s for which \( R_X \subseteq \{0, 1, 2, \ldots\} \) then

\[
G_X'(1) = \sum_{k=1}^{\infty} k p_X(k) = \sum_{k=0}^{\infty} k p_X(k) = E[X]
\]
and

\[ G''_X(1) = \sum_{k=2}^{\infty} k(k-1)p_X(k) = \sum_{k=0}^{\infty} k(k-1)p_X(k) \]

\[ = \sum_{k=0}^{\infty} k^2p_X(k) - \sum_{k=0}^{\infty} kp_X(k) = E[X^2] - E[X]. \]

Therefore

\[ \mu = E[X] = G'_X(1), \]

\[ \sigma^2 = \text{var}(X) = E[X^2] - E[X]^2 = G''_X(1) + \mu - \mu^2. \]

**Example 3.2.3**

\( X \sim \text{Po}(\lambda) \) i.e. \( p_X(k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad k = 0, 1, \ldots \)

So

\[ G_X(z) = \sum_{k=0}^{\infty} z^k \left( e^{-\lambda} \frac{\lambda^k}{k!} \right) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda z)^k}{k!} = e^{-\lambda} e^{\lambda z} = e^{\lambda(z-1)}. \]

Now

\[ G'_X(z) = \lambda e^{\lambda(z-1)} \quad G''_X(z) = \lambda^2 e^{\lambda(z-1)}. \]

Therefore

\[ \mu = G'_X(1) = \lambda, \quad \sigma^2 = G''_X(1) + \mu - \mu^2 = \lambda^2 + \lambda - \lambda^2 = \lambda. \]

**Example 3.2.4**

\( X \sim \text{Geometric}(p) \) i.e. \( p_X(k) = (1-p)^{k-1}p \quad k = 1, 2, \ldots \)

So

\[ G_X(z) = \sum_{k=1}^{\infty} z^k(1-p)^{k-1}p \]

\[ = pz \sum_{k=1}^{\infty} [z(1-p)]^{k-1} = pz \sum_{k=1}^{\infty} (zq)^{k-1} = \frac{pz}{1-qz} \quad \text{for} \quad |z| < 1/q \]

where \( q = 1 - p. \)

\[ G'_X(z) = \frac{p}{1-qz} + \frac{pqz}{(1-qz)^2} = \frac{p(1-qz) + pqz}{(1-qz)^2} = \frac{p}{(1-qz)^2}. \]

\[ G''_X(z) = \frac{2pq}{(1-qz)^3}. \]

So

\[ G'_X(1) = \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}. \]

Also

\[ G''_X(1) = \frac{2pq}{(1-q)^3} = \frac{2p(1-p)}{p^3} = \frac{2(1-p)}{p^2}. \]
Hence,

\[ \mu = G'_X(1) = \frac{1}{p} \]

\[ \sigma^2 = G''_X(1) + \mu - \mu^2 = \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}. \]

3.2.2 Moment Generating Functions

The **moment generating function** is often used for calculating the mean, variance and various other moments for both continuous and discrete r.v’s. However, it is often more convenient to use the p.g.f. for discrete r.v’s.

**Definition 3.2.5 (m.g.f.)**

The moment generating function (m.g.f.) of a r.v. \( X \) is the function of a real variable defined by

\[
M_X(t) = E[e^{tX}] = \left\{ \begin{array}{ll}
\sum_k e^{tk}p_X(k) & \text{discrete} \\
\int e^{tx}f_X(x)dx & \text{cts}
\end{array} \right.
\]

for all \( t \) for which the R.H.S. exists.

Many of our remarks for existence relating to the p.g.f. also hold for the m.g.f. since \( M_X(t) = G_X(e^t) \) (check for yourself). We also note that \( M_X(0) \) exists, and is equal to 1.

The m.g.f. is so-called because it *generates* the moments of \( X \), as can be seen by expanding the exponential function inside the expectation operator:

\[
M_X(t) = E[e^{tX}] = E \left[ 1 + tX + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \ldots \right]
\]

\[ = 1 + tE[X] + \frac{t^2}{2!}E[X^2] + \frac{t^3}{3!}E[X^3] + \ldots \]

It is clear that information about the moments of \( X \) about the origin, i.e. \( \{E[X^k], k = 0, 1, 2, \ldots \} \) has been *compressed* into the function \( M_X(t) \).

So if \( a_k \) is the coefficient of \( t^k \) in the series expansion of \( M_X(t) \), then \( a_k = \frac{E[X^k]}{k!} \), and so

\[ E[X^k] = a_k \times k! \]

We can also use calculus to extract the \( k \)-th moment. To see this, observe that

\[
M'_X(t) = E[X] + tE[X^2] + \frac{t^2}{2!}E[X^3] + \ldots
\]

\[
M''_X(t) = E[X^2] + tE[X^3] + \frac{t^2}{2!}E[X^4] + \ldots
\]

In general,

\[
M^{(k)}_X(t) = E[X^k] + tE[X^{k+1}] + \frac{t^2}{2!}E[X^{k+2}] + \ldots
\]
\( k = 1, 2, \ldots \)

Therefore

\[ M'(0) = E[X], \quad M''(0) = E[X^2], \quad \text{and} \quad M^{(k)}(0) = E[X^k]. \]

**Example 3.2.6**

\( X \sim \text{Exp}(\theta) \) i.e. \( f_X(x) = \theta e^{-\theta x}, \ x \geq 0. \)

\[
M_X(t) = E[e^{tX}] = \int_0^\infty e^{tx} f_X(x) \, dx
\]

\[
= \int_0^\infty \theta e^{-\theta x} t^x \, dx = \frac{\theta}{\theta - t} \quad t < \theta.
\]

Let us find the 1st and 2nd moments of \( X \) (about the origin) using the two methods outlined above.

**By Power Series Expansion:**

\[
M_X(t) = \frac{\theta}{\theta - t} = \frac{1}{1 - (t/\theta)} = 1 + \frac{t}{\theta} + \frac{t^2}{\theta^2} + \frac{t^3}{\theta^3} + \ldots
\]

for \(|t| < \theta\).

\[
a_1 = \frac{1}{\theta}, \quad a_2 = \frac{1}{\theta^2}.
\]

Therefore

\[
E[X] = \frac{1}{\theta} \times 1! = \frac{1}{\theta} \quad E[X^2] = \frac{1}{\theta^2} \times 2! = \frac{2}{\theta^2}.
\]

**By Calculus:**

\[
M'(t) = \frac{\theta}{(\theta - t)^2} \quad M''(t) = \frac{2\theta}{(\theta - t)^3}.
\]

Therefore

\[
E[X] = M'(0) = \frac{1}{\theta} \quad E[X^2] = M''(0) = \frac{2}{\theta^2}.
\]

We can now also compute the variance of the distribution:

\[
\sigma^2 = E[X^2] - \mu^2 = \frac{2}{\theta^2} - \left( \frac{1}{\theta} \right)^2 = \frac{1}{\theta^2}.
\]

**Remarks 3.2.7**

(i) We can identify the distribution of a r.v. through its m.g.f. due to the following uniqueness result:

\( X \) and \( Y \) have the same p.d.f.

\( \iff M_X(t) = M_Y(t) \) for suitable \( t \).

This follows from a uniqueness result relating to LAPLACE TRANSFORMS.

(ii) The m.g.f. of a r.v. is the LAPLACE TRANSFORM of its p.d.f.
Proposition 3.2.8 (linear transformations)
Suppose that $Y = aX + b$ for constants $a, b \in \mathbb{R}$. Then

$$M_Y(t) = e^{bt} M_X(at).$$

Proof

$$M_Y(t) = E[e^{tY}] = E[e^{t(aX+b)}] = E[e^{atX}e^{bt}] = e^{bt} E[e^{atX}] = e^{bt} M_X(at).$$