2.5 TRANSFORMATIONS OF RANDOM VARIABLES

2.5.1 GENERAL TRANSFORMATIONS

Consider a discrete/continuous random variable X with range X and probability distribution described by mass/pdf f_X, or cdf F_X. Suppose g is a real-valued function whose domain includes X, and suppose that

$$\begin{aligned} g: \quad \mathbb{X} &\longrightarrow \mathbb{Y} \\ x &\longmapsto y \end{aligned}$$

Then $Y = g(X)$ is also a random variable as Y is a function from Ω to $\mathbb{R}$. For $A \subseteq \mathbb{R}$, the event $[Y \in A]$ is an event in terms of the transformed variable Y. If f_Y is the mass/density function for Y, then

$$P[Y \in A] = \begin{cases} \displaystyle\sum_{y \in A} f_Y(y) & Y \text{ discrete} \\ \displaystyle\int_A f_Y(y)dy & Y \text{ continuous} \end{cases}$$

We wish to derive the probability distribution of random variable Y; in order to do this, we first consider the inverse transformation g^{-1} from Y to X defined for set $A \subseteq Y$ (and for $y \in Y$) by

$$g^{-1}(A) = \{x \in \mathbb{X} : g(x) \in A\} \qquad\qquad g^{-1}(y) = \{x \in \mathbb{X} : g(x) = y\}$$

that is, $g^{-1}(A)$ is the set of points in X that map into A, and $g^{-1}(y)$ is the set of points in X that map to y, under transformation g. By construction, we have

$$P[Y \in A] = P[X \in g^{-1}(A)]$$

and hence $[Y \in A]$ and $[X \in g^{-1}(A)]$ are equivalent events.

Consider first the cdf of Y, F_Y, evaluated at a point $y \in \mathbb{R}$. We have

$$F_Y(y) = P[Y \le y] = P[g(X) \le y] = \begin{cases} \displaystyle\sum_{x \in A_y} f_X(x) & X \text{ discrete} \\ \displaystyle\int_{A_y} f_X(x)dx & X \text{ continuous} \end{cases}$$

where $A_y = \{x \in \mathbb{X} : g(x) \le y\}$. This result gives the "*first principles*" approach to computing the distribution of the new variable: the approach can be summarized as follows:

- consider the range $\mathbb{Y}$ of the new variable
- consider the cdf $F_Y(y)$; step through the arguments as follows

$$F_Y(y) \quad = P[Y \le y] = P[g(X) \le y] = P[X \in A_y]$$

Note that it is usually a good idea to start with the cdf, not the pmf or pdf.

Often, the set is A_y is easy to identify for a given y, that is,

$$F_Y(y) = P[g(X) \le y] = P[x_1 \le X \le x_2]$$

where x_1 and x_2 depend on y and g or g^{-1}. The main objective is therefore to identify the set A_y.

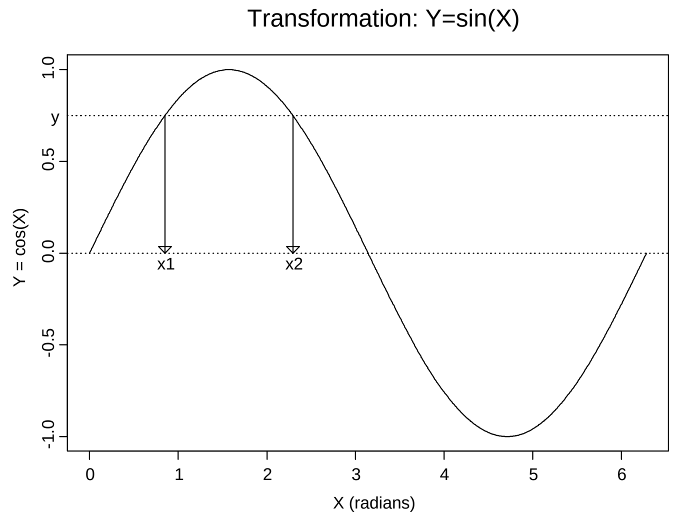

Figure 2.3: Computation of A_y for $Y = \sin X$

Example 2.5.1 Suppose that X is a continuous random variable with range $\mathbb{X} \equiv (0, 2\pi)$ whose pdf f_X is constant

$$f_X(x) = \frac{1}{2\pi} \qquad 0 < x < 2\pi$$

and zero otherwise. This pdf has corresponding continuous cdf

$$F_X(x) = \frac{x}{2\pi} \qquad 0 < x < 2\pi$$

Consider transformed random variable

$$Y = \sin X$$

Then the range of Y, $\mathbb{Y}$ is $[-1, 1]$, but the transformation is not 1-1. However, from first principles, we have

$$F_Y(y) = P[Y \le y] = P[\sin X \le y]$$

Now, by inspection of Figure 2.3,we can easily identify the required set A_y : it is the union of two disjoint intervals

$$A_y = [0, x_1] \cup [x_2, 2\pi] = \left[0, \sin^{-1} y\right] \cup \left[\pi - \sin^{-1} y, 2\pi\right]$$

so that

$$\begin{aligned} F_Y(y) = P[\sin X \le y] &= P[X \le x_1] + P[X \ge x_2] \\ &= \{P[X \le x_1]\} + \{1 - P[X < x_2]\} \\ = \left\{\frac{1}{2\pi}\sin^{-1} y\right\} + \left\{1 - \frac{1}{2\pi}\left(\pi - \sin^{-1} y\right)\right\} &= \frac{1}{2} + \frac{1}{\pi}\sin^{-1} y \end{aligned}$$

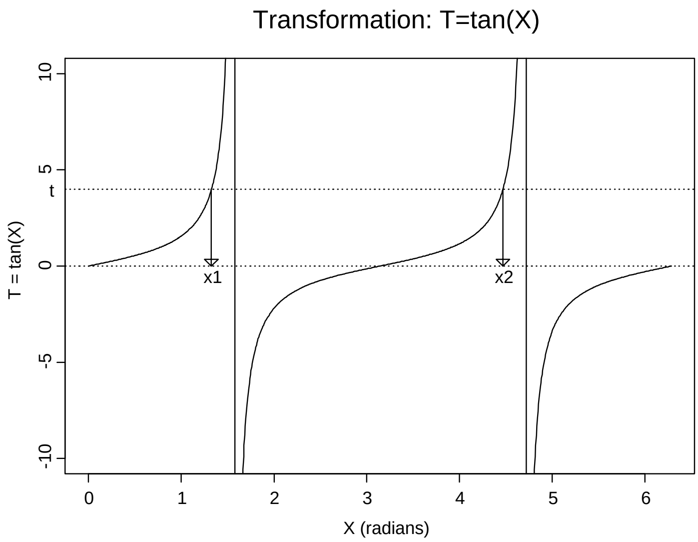

Figure 2.4: Computation of A_t for $T = \tan X$

and hence, by differentiation

$$f_Y(y) = \frac{1}{\pi}\frac{1}{\sqrt{1-y^2}}$$

Example 2.5.2 Consider transformed random variable

$$T = \tan X$$

Then the range of Y, $\mathbb{T}$ is $\mathbb{R}$, but the transformation is not 1-1. However, from first principles, we have, for $t > 0$

$$F_T(t) = P\left[T \leq t\right] = P\left[\tan X \leq t\right]$$

Figure 2.4 helps identify the required set A_t: in this case, it is the union of three disjoint intervals

$$A_y = [0, x_1] \cup [x_1, x_2] \cup \left[\frac{3\pi}{2}, 2\pi\right] = \left[0, \tan^{-1} t\right] \cup \left[\pi, \pi + \tan^{-1} t\right] \cup \left[\frac{3\pi}{2}, 2\pi\right]$$

(note, for values of $t < 0$, the union will be of only two intervals, but the calculation proceeds identically) so that

$$\begin{aligned} F_Y(y) &= P\left[\tan X \leq y\right] = P\left[X \leq x_1\right] + P\left[x_1 \leq X \leq x_2\right] + P\left[\frac{3\pi}{2} \leq X \leq 2\pi\right] \\ &= \left\{\frac{1}{2\pi}\tan^{-1} t\right\} + \frac{1}{2\pi}\left\{\pi + \tan^{-1} t - \pi\right\} + \frac{1}{2\pi}\left\{2\pi - \frac{3\pi}{2}\right\} = \frac{1}{2} + \frac{1}{\pi}\tan^{-1} t \end{aligned}$$

and hence, by differentiation

$$f_T(t) = \frac{1}{\pi}\frac{1}{1+t^2}$$

2.5.2 1-1 TRANSFORMATIONS

The mapping $g(X)$ is a function of X from X which is 1-1 and onto Y if,
(i) for each $x \in X$, there exists one and only one y such that $y = g(x)$, and
(ii) for each $y \in Y$, there exists an $x \in X$ such that $g(x) = y$.

The following theorem gives the distribution for random variable $Y = g(X)$ when g is 1-1.

THEOREM
Let X be a random variable with mass/density function f_X and support $\mathbb{X}$. Let g be a 1-1 function from $\mathbb{X}$ onto $\mathbb{Y}$ with inverse g^{-1}. Then $Y = g(X)$ is a random variable with support $\mathbb{Y}$ and

Discrete Case : The mass function of random variable Y is given by

$$f_Y(y) = f_X(g^{-1}(y))y \in Y = \{y|f_Y(y) > 0\}$$

.where x is the unique solution of $y = g(x)$ (so that $x = g^{-1}(y)$).

Continuous Case : The pdf of random variable Y is given by

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dt}\left\{g^{-1}(t)\right\}_{t=y} \right| \qquad y \in Y = \{y|f_Y(y) > 0\}$$

where $y = g(x)$, provided that the derivative

$$\frac{d}{dt}\left\{g^{-1}(t)\right\}$$

is continuous and non-zero on $\mathbb{Y}$.

PROOF
Discrete case: by direct calculation,

$$f_Y(y) = P[Y = y] = P[g(X) = y] = P[X = g^{-1}(y)] = f_X(x)$$

where $x = g^{-1}(y)$, and hence $f_Y(y) > 0 \iff f_X(x) > 0$.

Continuous case: function g is either (I) a monotonic increasing, or (II) a monotonic decreasing function.

Case (I): If g is increasing, then for $x \in \mathbb{X}$ and $y \in \mathbb{Y}$, we have that

$$g(x) \leq y \iff x \leq g^{-1}(y).$$

Therefore, for $y \in \mathbb{Y}$,

$$F_Y(y) = P[Y \leq y] = P[g(X) \leq y] = P[X \leq g^{-1}(y)] = F_X(g^{-1}(y))$$

and, by differentiation, because g is monotonic increasing,

$$f_Y(y) = f_X(g^{-1}(y)) \frac{d}{dt}\left\{g^{-1}(t)\right\}_{t=y} = f_X(g^{-1}(y)) \left| \frac{d}{dt}\left\{g^{-1}(y)\right\}_{t=y} \right| \qquad \text{as } \frac{d}{dt}\left\{g^{-1}(t)\right\} > 0.$$

Case (II): If g is decreasing, then for $x \in \mathbb{X}$ and $y \in \mathbb{Y}$ we have

$$g(x) \leq y \Longleftrightarrow x \geq g^{-1}(y)$$

Therefore, for $y \in Y$,

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

so

$$f_Y(y) = -f_X(g^{-1}(y))\frac{d}{dt}\left\{g^{-1}(y)\right\} = f_X(g^{-1}(y))\left|\frac{d}{dt}\left\{g^{-1}(t)\right\}_{t=y}\right| \qquad \text{as } \frac{d}{dt}\left\{g^{-1}(t)\right\} < 0.$$

Definition 2.5.1 Suppose transformation $g : X \longrightarrow Y$ is 1-1, and is defined by $g(x) = y$ for $x \in X$. Then the **Jacobian** of the transformation, denoted $J(y)$, is given by

$$J(y) = \frac{d}{dt}\left\{g^{-1}(t)\right\}_{t=y}$$

that is, the first derivative of g^{-1} evaluated at $y = g(x)$. Note that the inverse transformation $g^{-1} : Y \longrightarrow X$ has Jacobian $1/J(x)$.

Note : This is precisely the same term that appears as a change of variable term in an integration.

Note : To compute the expectation of $Y = g(X)$, we now have two alternative methods of computation; we either compute the expectation of $g(x)$ with respect to the distribution of X, or compute the distribution of Y, and then its expectation. It is straightforward to demonstrate that the two methods are equivalent, that is

$$E_{f_X}\left[g(X)\right] = E_{f_Y}\left[Y\right]$$

This result is sometimes known as the *Law of the Unconscious Statistician.*

IMPORTANT NOTE: Note that the apparently appealing "plug-in" approach that sets

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

will almost always fail as the Jacobian term must be included. For example, if $Y = e^X$ so that $X = \log Y$, then merely setting

$$f_Y(y) = f_X(\log y)$$

is **insufficienct**, you **must** have

$$f_Y(y) = f_X(\log y) \times \frac{1}{y}$$

2.6 GENERATING FUNCTIONS

2.6.1 MOMENT GENERATING FUNCTIONS

Definition 2.6.1 For random variable X with mass/density function f_X, the **moment generating function**, or mgf, of X, M_X, is defined by

$$M_X(t) = E_{f_X}[e^{tX}]$$

if this expectation exists for all values of $t \in (-h, h)$ for some $h > 0$, that is,

$$\text{DISCRETE CASE} \qquad M_X(t) = \sum_{x \in \mathbb{X}} e^{tx} f_X(x)$$

$$\text{CONTINUOUS CASE} \quad M_X(t) = \int_{x \in \mathbb{X}} e^{tx} f_X(x)dx$$

Note : It can be shown that if X_1 and X_2 are random variables taking values on X with mass/density functions f_{X_1} and f_{X_2}, and mgfs M_{X_1} and M_{X_2} respectively, then

$$f_{X_1}(x) \equiv f_{X_2}(x), x \in X \Longleftrightarrow M_{X_1}(t) \equiv M_{X_2}(t), t \in (-h, h)$$

Hence there is a 1-1 correspondence between generating functions and distributions: this provides a key technique for identification of probability distributions

2.6.2 KEY PROPERTIES OF MGFS

(i) If X is a discrete random variable, the rth derivative of M_X evaluated at t, $M_X^{(r)}(t)$, is given by

$$M_X^{(r)}(t) = \frac{d^r}{ds^r}\{M_X(s)\}_{s=t} = \frac{d^r}{ds^r}\left\{\sum_{x \in \mathbb{X}} e^{sx} f_X(x)\right\}_{s=t} = \sum_{x \in \mathbb{X}} x^r e^{tx} f_X(x)$$

and hence

$$M_X^{(r)}(0) = \sum_{x \in \mathbb{X}} x^r f_X(x) = E_{f_X}[X^r]$$

If X is a continuous random variable, the rth derivative of M_X is given by

$$M_X^{(r)}(t) = \frac{d^r}{ds^r}\left\{\int_{x \in \mathbb{X}} e^{sx} f_X(x) d_x\right\}_{s=t} = \int_{x \in \mathbb{X}} x^r e^{tx} f_X(x)dx$$

and hence

$$M_X^{(r)}(0) = \int_{x \in \mathbb{X}} x^r f_X(x)dx = E_{f_X}[X^r]$$

(ii) If X is a discrete random variable, then

$$
\begin{aligned}
M_X(t) &= \sum_{x \in \mathbb{X}} e^{tx} f_X(x) \\
&= \sum_{x \in \mathbb{X}} \left\{ \sum_{r=0}^{\infty} \frac{(tx)^r}{r!} \right\} f_X(x) \\
&= 1 + \sum_{r=1}^{\infty} \frac{t^r}{r!} \left\{ \sum_{x \in \mathbb{X}} x^r f_X(x) \right\} = 1 + \sum_{r=1}^{\infty} \frac{t^r}{r!} E_{f_X}[X^r]
\end{aligned}
$$

(identical result holds for the continuous case).

(iii) From the general result for expectations of functions of random variables

$$E_{f_Y}[e^{tY}] \equiv E_{f_X}[e^{t(aX+b)}] \Longrightarrow M_Y(t) = E_{f_X}[e^{t(aX+b)}] = e^{bt} E_{f_X}[e^{atX}] = e^{bt} M_X(at).$$

Therefore, if

$$Y = aX + b, M_Y(t) = e^{bt} M_X(at)$$

THEOREM

Let $X_1, ..., X_k$ be independent random variables with mgfs $M_{X_1}, ..., M_{X_k}$ respectively. Then if random variable Y is defined by $Y = X_1 + ... + X_k$,

$$M_Y(t) = \prod_{i=1}^{k} M_{X_i}(t)$$

PROOF

Using the general result for expectations of functions of independent random variables,

$$M_Y(t) = E_{f_Y}[e^{tY}] = E_{f_{X_1,...,X_k}}[e^{t(X_1+...+X_k)}] = \prod_{i=1}^{k} E_{f_{X_i}}[e^{tX_i}] = \prod_{i=1}^{k} M_{X_i}(t).$$

Special Case : If $X_1, ..., X_k$ are identically distributed, then $M_{X_i}(t) \equiv M_X(t)$, say, for all i, so

$$M_Y(t) = \prod_{i=1}^{k} M_X(t) = \{M_X(t)\}^k$$

2.6.3 OTHER GENERATING FUNCTIONS

Definition 2.6.2 For random variable X, with mass/density function f_X, the **factorial moment** or **probability generating function**, fmgf or pgf , of X, G_X, is defined by

$$G_X(t) = E_{f_X}[t^X] = E_{f_X}[e^{X\log t}] = M_X(\log t)$$

if this expectation exists for all values of $t \in (1-h, 1+h)$ for some $h > 0$.

Properties :
(i) Using similar techniques to those used for the mgf, it can be shown that

$$\begin{aligned} G_X^{(r)}(t) &= \frac{d^r}{ds^r}\{G_X(s)\}_{s=t} = E_{f_X}\left[X(X-1)...(X-r+1)t^{X-r}\right] \\ &\implies G_X^{(r)}(1) = E_{f_X}[X(X-1)...(X-r+1)] \end{aligned}$$

where $\mathrm{E}_{f_X}[X(X-1)...(X-r+1)]$ is the rth factorial moment.

(ii) For discrete random variables, it can be shown by using a Taylor series expansion of G_X that, for $r = 1, 2, ...$,

$$\frac{G_X^{(r)}(0)}{r!} = P[X = r]$$

Definition 2.6.3 For random variable X with mass/density function f_X, the **cumulant generating function** of X, K_X, is defined by

$$K_X(t) = \log[M_X(t)]$$

for $t \in (-h, h)$ for some $h > 0$.

Definition 2.6.4 The **characteristic function**, or cf, of X, C_X, is defined by

$$C_X(t) = E_{f_X}\left[e^{itX}\right]$$

if this expectation exists for $t \in \mathbb{R}$. By definition

$$\begin{aligned} C_X(t) &= \int_{x\in\mathbb{X}} e^{itx} f_X(x)dx = \int_{x\in\mathbb{X}} [\cos tx + i\sin tx]\, f_X(x)dx \\ &= \int_{x\in\mathbb{X}} \cos tx f_X(x)dx + i\int_{x\in\mathbb{X}} \sin tx f_X(x)dx \\ &= E_{f_X}[\cos tX] + iE_{f_X}[\sin tX] \end{aligned}$$