2. Transformations and Expectations

Transformations

Recall that a random variable X is a function mapping a sample space S to $\mathcal{X} \subseteq \mathbb{R}$.

Consider a real-valued function g defined on $\mathbb{R}$. Then $Y = g(X)$ is a composition of functions mapping S to $\mathcal{Y} \subseteq \mathbb{R}$ and so is also a random variable.

What is the distribution of Y?

For any set $A \subseteq \mathcal{Y}$ we can define an inverse mapping

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

Then we define the event

$$\Big(Y \in A\Big) = \Big(g(X) \in A\Big) = \Big(X \in g^{-1}(A)\Big).$$

Thus we can define a probability measure

$$P\Big(Y \in A\Big) = P\Big(X \in g^{-1}(A)\Big) = P\Big(\{s \in S : X(s) \in g^{-1}(A)\}\Big)$$

This satisfies the Axioms of Probability and so is a valid probability measure.

The support $\mathcal{Y}$ of Y is given by

$$\mathcal{Y} = \{y : y = g(x) \text{ for some } x \in \mathcal{X}\}$$

For a discrete random variable we can find the probability mass function of Y from that for X.

$$f_Y(y) = \sum_{\{x \in \mathcal{X} : g(x) = y\}} f_X(x) = \sum_{x \in g^{-1}(y)} \qquad \text{for } y \in \mathcal{Y}$$

The cumulative distribution function for Y is found by summing its probability mass function

$$F_Y(y) = \sum_{t \leqslant y} f_Y(t) = \sum_{\{x \in \mathcal{X} : g(x) \leqslant y\}} f_X(x)$$

For a continuous random variable, it is easiest to get the cdf first.

$$F_Y(y) = \int_{\{x \in \mathcal{X} : g(x) \leqslant y\}} f_X(x)\, dx$$

We can then find the probability density function using the relation

$$f_Y(y) = \frac{d}{dy} F_Y(y)$$

Monotone Transformations

Suppose that g is monotone.

$$u > v \;\Rightarrow\; g(u) > g(v) \qquad \text{(increasing)}$$

$$u > v \;\Rightarrow\; g(u) < g(v) \qquad \text{(decreasing)}$$

Then g is one-to-one and so g^{-1} is single-valued and monotone.

Theorem 2.1

Suppose that X has cdf F_X on support $\mathcal{X}$ and let $Y = g(X)$ be defined on $\mathcal{Y} = g(\mathcal{X})$.

(i) *If g is an increasing function then $F_Y(y) = F_X\left(g^{-1}(y)\right)$ for any $y \in \mathcal{Y}$.*

(ii) *If g is a decreasing function and X is a continuous random variable then $F_Y(y) = 1 - F_X\left(g^{-1}(y)\right)$ for any $y \in \mathcal{Y}$.*

For continuous random variables we have

Theorem 2.2

Let X be a continuous random variable with continuous pdf f_X on a support $\mathcal{X}$ and let $Y = g(X)$ where g is a monotone function on $\mathcal{X}$. Let $\mathcal{Y} = g(\mathcal{X})$ and suppose that g^{-1} has a continuous derivative on $\mathcal{Y}$. Then the pdf of Y is

$$f_Y(y) = \begin{cases} f_X\left(g^{-1}(y)\right)\left|\frac{dg^{-1}(y)}{dy}\right| & \text{for } y \in \mathcal{Y} \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 2.3

Let X be a continuous random variable with pdf f_X on the support $\mathcal{X}$ and let $Y = g(X)$. Suppose that there exists a partition $A_0, A_1, \ldots, A_k$ of $\mathcal{X}$ such that $P(X \in A_0) = 0$ and f_X is continuous on each A_i. If there exist functions $g_1, \ldots, g_k$ defined on $A_1, \ldots, A_k$ such that

(i) $g(x) = g_i(x)$ *for every* $x \in A_i$*;*

(ii) g_i *is monotone on* A_i *for each* $i = 1, \ldots, k$*,*

(iii) *the set* $\mathcal{Y} = \{y : y = g_i(x) \text{ for some } x \in A_i\}$ *is the same for each* $i = 1, \ldots, k$*,*

(iv) g_i^{-1} *has continuous derivative on* $\mathcal{Y}$ *for each* $i = 1, \ldots, k$*,*

then the pdf of Y is

$$f_Y(y) = \begin{cases} \sum_{i=1}^{k} f_X\left(g_i^{-1}(y)\right) \left| \frac{dg_i^{-1}(y)}{dy} \right| & \text{for } y \in \mathcal{Y} \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 2.4 (Probability Integral Transform)

Let X have continuous cdf F_X and define the random variable $Y = F_X(x)$. Then Y is distributed as a uniform random variable on the interval $(0, 1)$.

That is the pdf of Y is

$$f_Y(y) = \begin{cases} 1 & 0 < y < 1 \\ 0 & \textit{otherwise.} \end{cases}$$

Expectations

Definition 2.1

If X is a discrete random variable with probability mass function f_{sssX} on support $\mathcal{X}$ then the expected value or mean of $g(X)$ for any real-valued function g is

$$\mathsf{E}[g(X)] = \sum_{x \in \mathcal{X}} g(x) f_X(x)$$

provided that $\sum |g(x)| f_X(x) < \infty$, otherwise we say that the mean does not exist.

If X is a continuous random variable with probability density function $f_X(x)$ the expected value of $g(X)$ is

$$\mathsf{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$$

provided that $\int |g(x)| f_X(x)\, dx < \infty$, otherwise we say that the mean does not exist.

Theorem 2.5

Let X be a random variable with support $\mathcal{X}$. Then for any real-valued functions g_1 and g_2 whose expectations exist and any real constants a, b and c

(i) $\mathsf{E}[ag_1(X) + bg_2(X) + c] = a\mathsf{E}[g_1(X)] + b\mathsf{E}[g_2(X)] + c$.

(ii) *If* $g_1(x) \geqslant 0$ *for all* $x \in \mathcal{X}$ *then* $\mathsf{E}[g_1(X)] \geqslant 0$

(iii) *If* $g_1(x) \geqslant g_2(x)$ *for all* $x \in \mathcal{X}$ *then* $\mathsf{E}[g_1(X)] \geqslant \mathsf{E}[g_2(X)]$

(iv) *If* $\leqslant g_1(x) \leqslant b$ *for all* $x \in \mathcal{X}$ *then* $a \leqslant \mathsf{E}[g_1(X)] \leqslant b$

Definition 2.2

Suppose that X is a random variable with cdf F_X.

For any positive integer r, the r^{th} moment of X (or more accurately of F_X) is

$$\mu'_r = \mathsf{E}[X^r]$$

The r^{th} central moment of X is

$$\mu_r = \mathsf{E}[(X-\mu)^r]$$

where $\mu = \mu'_1 = \mathsf{E}[X]$.

Definition 2.3

The second central moment is called the variance of X

$$\text{Var}(X) = \mathsf{E}[(X-\mu)^2] = \mathsf{E}[X^2] - \left(\mathsf{E}[X]\right)^2$$

The positive square root of the variance is called the standard deviation of X.

Definition 2.4

Let X be a random variable with cdf F_X. The moment generating function of X (or of F_X) is

$$M_X(t) \;=\; \mathsf{E}\left[\mathsf{e}^{tX}\right]$$

provided this expectation exists for t in some neighbourhood of 0.

Theorem 2.6

If X has moment generating function $M_X(t)$ then for any integer r

$$\mu_r' = \mathsf{E}[X^r] = \frac{d^r}{dt^r} M_X(t)\Big|_{t=0}.$$

Theorem 2.7

Let X be a random variable with moment generating function $M_X(t)$ which exists in a neighbourhood of 0 and let a and b be two real constants. The the moment generating function of the random variable $aX + b$ is

$$M_{aX+b}(t) = \mathrm{e}^{bt} M_X(at).$$

- If the moment generating function of a random variable exists then all moments of that random variable exist.

- The reverse is not true (e.g. the log-normal distribution).

Theorem 2.8

Let X and Y be two random variables with cdfs F_X and F_Y respectively, all of whose moments exist.

(i) *If X and Y have bounded support then $F_X(u) = F_Y(u)$ for all u if, and only if $\mathsf{E}[X^r] = \mathsf{E}[Y^r]$ for all positive integers r.*

(ii) *If the moment generating functions $M_X(t)$ and $M_Y(t)$ exist in a neighbourhood of 0 and $M_X(t) = M_Y(t)$ in that neighbourhood then $F_X(u) = F_Y(u)$ for all u.*

Theorem 2.9

Let $X_1, X_2, \ldots$ be a sequence of random variables each with moment generating function $M_{X_i}(t)$, and cumulative distribution function F_{X_i} for $i = 1, 2, \ldots$. Suppose that

$$\lim_{i \to \infty} M_{X_i}(t) \to M_X(t) \qquad \text{for all } t \text{ in a neighbourhood of 0}$$

where $M_X(t)$ is a moment generating function. Then there exists a unique cumulative distribution function F_X whose moments are determined by $M_X(t)$ and

$$\lim_{i \to \infty} F_{X_i}(x) \to F_X(x)$$

at every x where F_X is continuous.