Transformations and Expectations

1 Distributions of Functions of a Random Variable

If X is a random variable with cdf $F_X(x)$, then any function of X, say $g(X)$, is also a random variable. Sine $Y = g(X)$ is a function of X, we can describe the probabilistic behavior of Y in terms of that of X. That is, for any set A,

$$P(Y \in A) = P(g(X) \in A),$$

Showing that the distribution of Y depends on the function F_X and g.

Formally, if we write $y = g(x)$, the function $g(x)$ defines a mapping from the original sample space of X, $\mathcal{X}$, to a new sample space, $\mathcal{Y}$, the sample space of the random variable Y. That is,

$$g(x) : \mathcal{X} \longrightarrow \mathcal{Y}.$$

Conveniently, we can write

$$\mathcal{X} = \{x : f_X(x) > 0\} \quad \text{and} \quad \mathcal{Y} = \{y : y = g(x) \text{ for some } x \in CX\}. \tag{1}$$

The pdf of X is positive only on the set $\mathcal{X}$ and is 0 elsewhere. Such a set is called the support set or support of a distribution. We associate with g an inverse mapping, denoted by g^{-1}, which is a mapping from subsets of $\mathcal{Y}$ to subsets of $\mathcal{X}$, and is defined by

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

It is possible for A to be a point set, say $A = \{y\}$. Then

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

In this case, we often write $g^{-1}(y)$ instead of $g^{-1}(\{y\})$.

The probability distribution of Y can be defined as follows. For any set $A \subset \mathcal{Y}$,

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

It is straightforward to show that this probability function satisfies the Kolmogorov Axioms.

If X is a discrete random variable, then $\mathcal{X}$ is countable. The sample space for $Y = g(X)$ is $\mathcal{Y} = \{y : y = g(x), x \in \mathcal{X}\}$, which is also a countable set. Thus, Y is also a discrete random variable. The pmf for Y is

$$\begin{aligned} f_Y(y) = P(Y = y) &= \sum_{x \in g^{-1}(y)} P(X = x) \\ &= \sum_{x \in g^{-1}(y)} f_X(x), \quad \text{for } y \in \mathcal{Y} \end{aligned}$$

and $f_Y(y) = 0$ for $y \notin \mathcal{Y}$. In this case, finding the pmf of Y involves simply identifying $g^{-1}(y)$, for each $y \in \mathcal{Y}$, and summing the appropriate probabilities.

Example 1.1 *(Binomial transformation) A discrete random variable X has a binomial distribution if its pmf is of the form*

$$f_X(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \ldots, n,$$

where n is a positive integer and $0 \leq p \leq 1$. Consider the random variable $Y = g(X)$, where $g(x) = n - x$. Thus, $g^{-1}(y)$ is the single point $x = n - y$, and

$$\begin{aligned} f_Y(y) &= \sum_{x \in g^{-1}(y)} f_X(x) = f_X(n-y) \\ &= \binom{n}{n-y} p^{n-y} (1-p)^{n-(n-y)} \\ &= \binom{n}{y} (1-p)^y p^{n-y}. \end{aligned}$$

Thus, we see that Y also has a binomial distribution, but with parameters n and $1-p$.

If X and Y are continuous random variables, the cdf of $Y = g(X)$ is

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(g(X) \leq y) \\ &= P(\{x \in \mathcal{X} : g(x) \leq y\}) = \int_{\{x \in \mathcal{X} : g(x) \leq y\}} f_X(x)dx. \end{aligned}$$

Sometimes there may be difficulty in identifying $\{x \in \mathcal{X} : g(x) \leq y\}$ and carrying out the integration of $F_X(x)$ over this region.

Example 1.2 *(Uniform transformation) Suppose X has a uniform distribution on the interval $(0, 2\pi)$, that is,*

$$f_X(x) = \begin{cases} 1/(2\pi) & 0 < x < 2\pi \\ 0 & \textit{otherwise.} \end{cases}$$

Consider $Y = \sin^2(X)$. *Then*

$$P(Y \le y) = P(X \le x_1) + P(x_2 \le X \le x_3) + P(X \ge x_4)$$
$$= 2P(X \le x_1) + 2P(x_2 \le X \le \pi),$$

where x_1 *and* x_2 *are the two solutions to*

$$\sin^2(x) = y, \quad 0 < x < \pi.$$

Thus, even though this example dealt with a seemingly simple situation, the cdf of Y *was not simple.*

It is easiest to deal with functions $g(x)$ that are monotone, that is, those that satisfy either

$$u > v \Rightarrow g(u) > g(v) \text{ (increasing) or } u < v \Rightarrow g(u) > g(v) \text{ (decreasing).}$$

If g is monotone, then g^{-1} is single-valued; that is, $g^{-1}(y) = x$ if and only if $y = g(x)$. If g is increasing, this implies that

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

If g is decreasing, this implies that

$$\{x \in \mathcal{X} : g(x) \le y\} = \{x \in \mathcal{X} : x \ge g^{-1}(y)\}.$$

If $g(x)$ is increasing, we can write

$$F_Y(y) = \int_{\{x \in \mathcal{X} : x \le g^{-1}(y)\}} f_X(x)dx = \int_{-\infty}^{g^{-1}(y)} f_X(x)dx = F_X(g^{-1}(y)).$$

If $g(x)$ is decreasing, we have

$$F_Y(y) = \int_{g^{-1}(y)}^{\infty} f_X(x)dx = 1 - F_X(g^{-1}(y)).$$

The continuity of X is used to obtain the second equality. We summarize these results in the following theorem.

Theorem 1.1 *Let* X *have cdf* $F_X(x)$, *let* $Y = g(X)$, *and let* $\mathcal{X}$ *and* $\mathcal{Y}$ *be defined as in (1).*

a. *If* g *is an increasing function on* $\mathcal{X}$, $F_Y(y) = F_X(g^{-1}(y))$ *for* $y \in \mathcal{Y}$.

b. *If* g *is a decreasing function on* $\mathcal{X}$ *and* X *is a continuous random variable,* $F_Y(y) = 1 - F_X(g^{-1}(y))$ *for* $y \in \mathcal{Y}$.

Example 1.3 *(Uniform-exponential relationship-I) Suppose* $X \sim f_X(x) = 1$ *if* $0 < x < 1$ *and 0 otherwise, the uniform(0,1) distribution. It is straightforward to check that* $F_X(x) = x, 0 < x < 1$. *We now make the transformation* $Y = g(X) = -\log(X)$. *Since*

$$\frac{d}{dx}g(x) = -\frac{1}{x} < 0, \quad \textit{for } 0 < x < 1,$$

$g(x)$ *is a decreasing function. Therefore, for* $y > 0$,

$$F_Y(y) = 1 - F_X(g^{-1}(y)) = 1 - F_X(e^{-y}) = 1 - e^{-y}.$$

Of course, $F_Y(y) = 0$ *for* $y \leq 0$.

If the pdf of Y is continuous, it can be obtained by differentiating the cdf.

Theorem 1.2 *Let* X *have pdf* $f_X(x)$ *and* $Y = g(X)$, *where* g *is a monotone function. Let* $\mathcal{X}$ *and* $\mathcal{Y}$ *be define by (1). Suppose that* $f_X(x)$ *is continuous on* $\mathcal{X}$ *and that* $g^{-1}(y)$ *has a continuous derivative on* $\mathcal{Y}$. *Then the pdf of* Y *is given by*

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))|\frac{d}{dy}g^{-1}(y)| & y \in \mathcal{Y} \\ 0 & \textit{otherwise.} \end{cases}$$

Proof: From Theorem 1.1 we have, by the chain rule,

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \begin{cases} f_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y) & \text{if } g \text{ is increasing} \\ -f_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y) & \text{if } g \text{ is decreasing.} \end{cases}$$

□

Example 1.4 *(Inverted gamma pdf) Let* $f_X(x)$ *be the gamma pdf*

$$f(x) = \frac{1}{(n-1)!\beta^n}x^{n-1}e^{-x/\beta}, \quad 0 < x < \infty,$$

where β *is a positive constant and* n *is a positive integer. If we let* $y = g(x) = 1/x$, *then* $g^{-1}(y) = 1/y$ *and* $\frac{d}{dy}g^{-1}(y) = -1/y^2$. *Applying the above theorem, for* $0 < y < \infty$, *we get*

$$\begin{aligned} f_Y(y) &= f_X(g^{-1}(y))|\frac{d}{dy}g^{-1}(y)| \\ &= \frac{1}{(n-1)!\beta^n}\Big(\frac{1}{y}\Big)^{n-1}e^{-1/(\beta y)}\frac{1}{y^2} \\ &= \frac{1}{(n-1)!\beta^n}\Big(\frac{1}{y}\Big)^{n+1}e^{-1/(\beta y)}, \end{aligned}$$

a special case of a pdf known as the inverted gamma pdf.

Theorem 1.3 *Let X have pdf $F_X(x)$, let $Y = g(X)$, and define the sample space $\mathcal{X}$ as in (1). Suppose there exists a partition, $A_0, A_1, \ldots, A_k$, of $\mathcal{X}$ such that $P(X \in A_0) = 0$ and $f_X(x)$ is continuous on each A_i. Further, suppose there exist functions $g_1(x), \ldots, g_k(x)$, defined on $A_1, \ldots, A_k$, respectively, satisfying*

i. $g(x) = g_i(x)$, *for* $x \in A_i$,

ii. $g_i(x)$ *is monotone on* A_i,

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

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

Then

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

Example 1.5 *(Normal-Chi squared relationship) Let X have the standard normal distribution*

$$f_X(x) = \frac{1}{2\pi}e^{-x^2/2}, \quad -\infty < x < \infty.$$

Consider $Y = X^2$. The function $g(x) = x^2$ is monotone on $(-\infty, 0)$ and $(0, \infty)$. The set $\mathcal{Y} = (0, \infty)$. Applying Theorem 1.3, we take

$$A_0 = \{0\};$$

$$A_1 = (-\infty, 0), \quad g_1(x) = x^2, \quad g_1^{-1}(y) = -\sqrt{y};$$

$$A_2 = (0, \infty), \quad g_2(x) = x^2, \quad g_2^{-1}(y) = \sqrt{y}.$$

The pdf of Y is

$$\begin{aligned} f_Y(y) &= \frac{1}{\sqrt{2\pi}}e^{-(-\sqrt{y})^2/2}|-\frac{1}{2\sqrt{y}}| + \frac{1}{\sqrt{2\pi}}e^{-(\sqrt{y})^2/2}|\frac{1}{2\sqrt{y}}| \\ &= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{y}}e^{-y/2}, \quad 0 < y < \infty. \end{aligned}$$

So Y is a chi-squared random variable with 1 degree of freedom.

Let F_X^{-1} denote the inverse of the cdf F_X. If F_X is strictly increasing, then F_X^{-1} is well defined by

$$F_X^{-1}(y) = x \Leftrightarrow F_X(x) = y. \tag{2}$$

However, if F_X is constant on some interval, then F_X^{-1} is not well defined by (2). The problem is avoided by defining $F_X^{-1}(y)$ for $0 < y < 1$ by

$$F_X^{-1}(y) = \inf\{x : F_X(x) \geq y\}. \tag{3}$$

At the end point of the range of y, $F_X^{-1}(1) = \infty$ if $F_X(x) < 1$ for all x and, for any F_X, $F_X^{-1}(0) = -\infty$.

Theorem 1.4 *(Probability integral transformation) Let X have continuous cdf $F_X(x)$ and define the random variable Y as $Y = F_X(X)$. Then Y is uniformly distributed on $(0, 1)$, that is, $P(Y \leq y) = y$, $0 < y < 1$.*

Proof: For $Y = F_X(X)$ we have, for $0 < y < 1$,

$$\begin{aligned} P(Y \leq y) &= P(F_X(X) \leq y) \\ &= P(F_X^{-1}[F_X(X)] \leq F_X^{-1}(y)) \\ &= P(X \leq F_X^{-1}(y)) \\ &= F_X(F_X^{-1}(y)) = y. \end{aligned}$$

At the endpoints we have $P(Y \leq y) = 1$ for $y \geq 1$ and $P(Y \leq y) = 0$ for $y \leq 0$, showing that Y has a uniform distribution.

The reasoning behind the equality

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

is somewhat subtle and deserves additional attention. If F_X is strictly increasing, then it is true that $F_X^{-1}(F_X(x)) = x$. However, if F_X is flat, it may be that $F_X^{-1}(F_X(x)) \neq x$. Then $F_X^{-1}(F_X(x)) = x_1$, since $P(X \leq x) = P(X \leq x_1)$ for any $x \in [x_1, x_2]$. The flat cdf denotes a region of 0 probability $P(x_1 < X \leq x) = F_X(x) - F_X(x_1) = 0$. □

2 Expected values

Definition 2.1 *The expected value or mean of a random variable $g(X)$, denoted by $Eg(X)$, is*

$$Eg(X) = \begin{cases} \int_{-\infty}^{\infty} g(x) f_X(x) dx & \text{if } X \text{ is continuous} \\ \sum_{x \in \mathcal{X}} g(x) f_X(x) = \sum_{x \in \mathcal{X}} g(x) P(X = x) & \text{if } X \text{ is discrete,} \end{cases}$$

provided that the integral or sum exists. If $E|g(X)| = \infty$, we say that $Eg(X)$ does not exist.