Transformations and Expectations of random variables

$X \sim F_X(x)$: a random variable X distributed with CDF F_X.

Any function $Y = g(X)$ is also a random variable.

If both X, and Y are continuous random variables, can we find a simple way to characterize F_Y and f_Y (the CDF and PDF of Y), based on the CDF and PDF of X?

■■■

For the CDF:

$$\begin{aligned}
F_Y(y) &= P_Y(Y \leq y) \\
&= P_Y(g(X) \leq y) \\
&= P_X(x \in \mathcal{X} : g(X) \leq y) \text{ ($\mathcal{X}$ is sample space for X)} \\
&= \int_{\{x \in \mathcal{X}: g(X) \leq y\}} f_X(s) ds.
\end{aligned}$$

PDF: $f_Y(y) = F'_y(y)$

Caution: need to consider support of y.

Consider several examples:

1. $X \sim U[-1, 1]$ and $y = \exp(x)$

   That is:

$$\begin{aligned}
f_X(x) &= \begin{cases} \frac{1}{2} & \text{if } x \in [-1, 1] \\ 0 & \text{otherwise} \end{cases} \\
F_X(x) &= \frac{1}{2} + \frac{1}{2}x, \text{ for } x \in [-1, 1].
\end{aligned}$$

$$\begin{aligned}
F_Y(y) &= Prob(\exp(X) \leq y) \\
&= Prob(X \leq \log y) \\
&= F_X(\log y) = \frac{1}{2} + \frac{1}{2}\log y, \quad \text{for } y \in [\frac{1}{e}, e].
\end{aligned}$$

   Be careful about the bounds of the support!

$$\begin{aligned}
f_Y(y) &= \frac{\partial}{\partial y} F_Y(y) \\
&= f_X(\log y)\frac{1}{y} = \frac{1}{2y}, \text{ for } y \in [\frac{1}{e}, e].
\end{aligned}$$

2. $X \sim U[-1, 1]$ and $Y = X^2$

$$\begin{aligned} F_Y(y) &= Prob(X^2 \le y) \\ &= Prob(-\sqrt{y} \le X \le \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \\ &= 2F_X(\sqrt{y}) - 1, \text{ by symmetry: } F_X(-\sqrt{y}) = 1 - F_X(\sqrt{y}). \end{aligned}$$

$$\begin{aligned} f_Y(y) &= \frac{\partial}{\partial y} F_Y(y) \\ &= 2f_X(\sqrt{y})\frac{1}{2\sqrt{y}} = \frac{1}{2\sqrt{y}}, \text{ for } y \in [0, 1]. \end{aligned}$$

■■■

As the first example above showed, it's easy to derive the CDF and PDF of Y when $g(\cdot)$ is a strictly monotonic function:

Theorems 2.1.3, 2.1.5: When $g(\cdot)$ is a strictly increasing function, then

$$\begin{aligned} F_Y(y) &= \int_{-\infty}^{g^{-1}(y)} f_X(x)dx = F_X(g^{-1}(y)) \\ f_Y(y) &= f_X(g^{-1}(y))\frac{\partial}{\partial y} g^{-1}(y) \text{ using chain rule.} \end{aligned}$$

Note: by the inverse function theorem,

$$\frac{\partial}{\partial y} g^{-1}(y) = 1/\left[g'(x)\right]|_{x=g^{-1}(y)}.$$

When $g(\cdot)$ is a strictly decreasing function, then

$$\begin{aligned} F_Y(y) &= \int_{g^{-1}(y)}^{\infty} f_X(x)dx = 1 - F_X(g^{-1}(y)) \\ f_Y(y) &= -f_X(g^{-1}(y))\frac{\partial}{\partial y} g^{-1}(y) \text{ using chain rule.} \end{aligned}$$

These are the *change of variables* formulas for transformations of univariate random variables.

Thm 2.1.8 generalizes this to piecewise monotonic transformations.

■■■

Here is a special case of a transformation:

Thm 2.1.10: Let X have a continuous CDF $F_X(\cdot)$ and define the random variable $Y = F_X(X)$. Then $Y \sim U[0,1]$, i.e., $F_Y(y) = y$, for $y \in [0,1]$.

Note: all that is required is that the CDF F_X is continuous, not that it must be strictly increasing. The result also goes through when F_X is continuous but has flat parts (cf. discussion in CB, pg. 34).

■■■

Expected value (Definition 2.2.1): The expected value, or mean, of a random variable $g(X)$ is

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

provided that the integral or the sum exists

The expectation is a *linear operator* (just like integration): so that

$$E\left[\alpha * \sum_{i=1}^{n} g_i(X) + b\right] = \alpha * \sum_{i=1}^{n} Eg_i(X) + b.$$

Note: Expectation is a *population average*, i.e., you average values of the random variable $g(X)$ weighting by the population density $f_X(x)$.

A statistical experiment yields sample observations $X_1, X_2, \ldots, X_n \sim F_X$. From these sample observations, we can calculate sample avg. $\bar{X}_n \equiv \frac{1}{n} \sum_i X_i$. In general: $\bar{X}_n \neq EX$. But under some conditions, as $n \to \infty$, then $\bar{X}_n \to EX$ in some sense (which we discuss later).

■■■

Expected value is commonly used measure of "central tendency" of a random variable X.

Example: But mean may not exist: Cauchy random variable with density $f(x) = \frac{1}{\pi(1+x^2)}$ for $x \in (-\infty, \infty)$. Note that

$$\begin{aligned}
\int_{-\infty}^{\infty} \frac{x}{\pi(1+x^2)} dx &= \int_{-\infty}^{0} \frac{x}{\pi(1+x^2)} dx + \int_{0}^{\infty} \frac{x}{\pi(1+x^2)} dx \\
&= \overset{0}{\lim_{a \to -\infty}} \int_{a}^{0} \frac{x}{\pi(1+x^2)} dx + \lim_{b \to \infty} \int_{0}^{b} \frac{x}{\pi(1+x^2)} dx \\
&= \overset{0}{\lim_{a \to -\infty}} \frac{1}{2\pi} [\log(1 - x^2)]_a^0 + \lim_{b \to \infty} \frac{1}{2\pi} [\log(1 - x^2)]_0^b \\
&= -\infty + \infty \quad \text{undefined}
\end{aligned}$$

■

Other measures:

1. Median: $\text{med}(X) = m$ such that $F_X(x) = 0.5$. Robust to outliers, and has nice invariance property: for $Y = g(X)$ and $g(\cdot)$ monotonic increasing, then $\text{med}(Y) = g(\text{med}(X))$.
2. Mode: $\text{Mode}(X) = \max_x f_X(x)$.

■■■

Moments: important class of expectations

For each integer n, the n-th (uncentred) moment of $X \sim F_X(\cdot)$ is $\mu'_n \equiv EX^n$.

The n-th centred moment is $\mu_n \equiv E(X - \mu)^n = E(X - EX)^n$. (It is centred around the mean EX.)

For $n = 2$: $\mu_2 = E(X - EX)^2$ is the *Variance* of X. $\sqrt{\mu_2}$ is the *standard deviation*.

Important formulas:

- $Var(aX + b) = a^2 Var X$ (variance is not a linear operation)
- $Var X = E(X^2) - (EX)^2$: alternative formula for the variance

■■■

The moments of a random variable are summarized in the *moment generating function*.

Definition: the moment-generating function of X is $M_X(t) \equiv E\exp(tX)$, provided that the expectation exists in some neighborhood $t \in [-h, h]$ of zero.

Specifically:

$$M_x(t) = \begin{cases} \int_{-\infty}^{\infty} e^{tx} f_X(x) dx & \text{for } X \text{ continuous} \\ \sum_{x \in \mathcal{X}} e^{tx} P(X = x) & \text{for } X \text{ discrete.} \end{cases}$$

The uncentred moments of X are generated from this function by:

$$EX^n = M_X^{(n)}(0) \equiv \frac{d^n}{dt^n} M_X(t) \bigg|_{t=0},$$

which is the n-th derivative of the MGF, evaluated at $t = 0$.

When it exists (see below), then MGF provides alternative description of a probability distribution. Mathematically, it is a *Laplace transform*, which can be convenient for certain mathematical calculations.

Example: standard normal distribution:

$$\begin{aligned}
M_X(t) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(tx - \frac{x^2}{2}\right) dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}((x-t)^2 - t^2)\right) dx \\
&= \exp(\frac{1}{2}t^2) \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x-t)^2\right) dx \\
&= \exp(\frac{1}{2}t^2) \cdot 1
\end{aligned}$$

where last term on RHS is integral over density function of $N(t,1)$, which integrates to one.

First moment: $EX = M_X^1(0) = t \cdot \exp(\frac{1}{2}t^2)\big|_{t=0} = 0$.
Second moment: $EX^2 = M_X^2(0) = \exp(\frac{1}{2}t^2) + t^2 \exp(\frac{1}{2}t^2) = 1$.

$\blacksquare\blacksquare\blacksquare$

In many cases, the MGF can characterize a distribution. But problem is that it may not exist (eg. Cauchy distribution)

For a RV X, is its distribution uniquely determined by its moment generating function?

Thm 2.3.11: For $X \sim F_X$ and $Y \sim F_Y$, if M_X and M_Y exist, and $M_X(t) = M_Y(t)$ for all t in some neighborhood of zero, then $F_X(u) = F_Y(u)$ for all u.

Note that if the MGF exists, then it characterizes a random variable with an *infinite* number of moments (because the MGF is infinitely differentiable). Converse not necessarily true. (ex. log-normal random variable: $X \sim N(0,1)$, $Y = \exp(X)$)

$\blacksquare\blacksquare\blacksquare$

Characteristic function:

The characteristic function of a random variable $g(x)$, defined as

$$\phi_{g(x)}(t) = E_x \exp(itg(x)) = \int_{-\infty}^{+\infty} \exp(itg(x)) f(x) dx$$

where $f(x)$ is the density for x.

This is also called the “Fourier transform”.

Features of characteristic function:

- The CF always exists. This follows from the equality $e^{itx} = \cos(tx) + i \cdot \sin(tx)$, and both the real and complex parts of the integrand are bounded functions.
- Consider a symmetric density function, with $f(-x) = f(x)$ (symmetric around zero). Then resulting $\phi(t)$ is *real-valued*, and symmetric around zero.
- The CF completely determines the distribution of X (every cdf has a unique characteristic function).
- Let X have characteristic function $\phi_X(t)$. Then $Y = aX + b$ has characteristic function $\phi_Y(t) = e^{ibt}\phi_X(at)$.
- X and Y, independent, with characteristic functions $\phi_X(t)$ and $\phi_Y(t)$. Then $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$
- $\phi(0) = 1$.
- For a given characteristic function $\phi_X(t)$ such that $\int_{-\infty}^{+\infty} |\phi_X(t)| dt < \infty$,[1] the corresponding density $f_X(x)$ is given by the inverse Fourier transform, which is

  $$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \phi_X(t) \exp(-itx) dt.$$

  Example: $N(0,1)$ distribution, with density $f(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$.

  Take as given that the characteristic function of $N(0,1)$ is

  $$\phi_{N(0,1)}(t) = \frac{1}{\sqrt{2\pi}} \int \exp\left(itx - x^2/2)\right) dx = \exp(-t^2/2). \tag{1}$$

  Hence the inversion formula yields

  $$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-t^2/2) \exp(-itx) dt.$$

  Now making substitution $z = -t$, we get

  $$\begin{aligned} &\frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp\left(izx - z^2/2\right) dz \\ =&\frac{1}{\sqrt{2\pi}} \phi_{N(0,1)}(x) = \frac{1}{\sqrt{2\pi}} \exp(x^2/2) = f_{N(0,1)}(x). \quad \text{(Use Eq. (1))} \end{aligned}$$

[1] Here $|\cdot|$ denotes the modulus of a complex number. For $x + iy$, we have $|x + iy| = \sqrt{x^2 + y^2}$.

- Characteristic function also summarizes the moments of a random variable. Specifically, note that the h-th derivative of $\phi(t)$ is

$$\phi^h(t) = \int_{-\infty}^{+\infty} i^h g(x)^h \exp(itg(x)) f(x)dx. \tag{2}$$

Hence, assuming the h-th moment, denoted $\mu^h_{g(x)} \equiv E[g(x)]^h$ exists, it is equal to

$$\mu^h_{g(x)} = \phi^h(0)/i^h.$$

Hence, assuming that the required moments exist, we can use Taylor's theorem to expand the characteristic function around $t = 0$ to get:

$$\phi(t) = 1 + \frac{it}{1}\mu^1_{g(x)} + \frac{(it)^2}{2}\mu^2_{g(x)} + ... + \frac{(it)^k}{k!}\mu^k_{g(x)} + o(t^k).$$

- **Cauchy distribution, cont'd:** The characteristic function for the Cauchy distribution is

$$\phi(t) = \exp(-|t|).$$

This is not differentiable at $t = 0$, which by Eq. (2) is saying that its mean does not exist. Hence, the expansion of the characteristic function in this case is invalid.