In this lecture, we return to the situation in which a random variable Y is defined in terms of another random variable X by $Y = g(X)$, where $g : \mathbb{R} \to \mathbb{R}$ is a function. We treated this situation in a complete, and therefore somewhat abstract, way in the "Change of variables" section of Lecture 5. We now focus on a few important particular cases.

1 Change of units transformations

A change of unit can be viewed as the application of the function $g(x) = ax + b$ to the old unit to obtain the new unit, with $a > 0$. As an example, the conversion from Celsius degrees ($°C$) to Fahrenheit degrees ($°F$) is given by:

$$T(°F) = g(T(°C)) = \frac{9}{5}T(°C) + 32$$

Let us now see how random variables are modified under these changes.

1.1 Continuous random variables

Let X and Y be two continuous random variables such that $Y = g(X)$, with $g(x) = ax + b$. Since $a > 0$, g is an increasing function. Following Lecture 5, we then have

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

Now,

$$g(x) = ax + b \quad \Leftrightarrow \quad x = g^{-1}(y) = \frac{y-b}{a}$$

Hence,

$$P(X \leq g^{-1}(y)) = P\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right)$$

We just proved that

$$\boxed{F_Y(y) = F_X\left(\frac{y-b}{a}\right)}$$

We can then compute the probability density function f_Y of Y:

$$f_Y(y) = \frac{d}{dy}\left[F_Y(y)\right] = \frac{d}{dy}\left[F_X\left(\frac{y-b}{a}\right)\right] = \frac{1}{a}f_X\left(\frac{y-b}{a}\right)$$

where we have used the chain rule for the last equality. We therefore have

$$\boxed{f_Y(y) = \frac{1}{a}f_X\left(\frac{y-b}{a}\right)}$$

Example: Consider a normally distributed random variable X with mean μ and variance σ^2. Then we expect the random variable

$$Y = \frac{X-\mu}{\sigma}$$

to be normally distributed, with mean 0 and variance 1. Let us verify this.

In our notation, we have $Y = g(X)$, with $g = ax + b$, $a = 1/\sigma$, $b = -\mu/\sigma$. Applying the formula we just derived

$$f_Y(y) = \sigma f_X(\sigma y + \mu) = \frac{\sigma}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(\frac{\sigma y+\mu-\mu}{\sigma}\right)^2} = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2} = \phi(y)$$

as expected. This shows, in a slightly different way, why only ϕ (or the c.d.f. Φ) needs to be tabulated for us to be able to construct any normally distributed probability density function: for any normally distributed random variable X, with mean μ and variance σ^2,

$$\boxed{f_X(x) = \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)} \qquad \text{for } f_X \text{ p.d.f. of } N(\mu, \sigma^2)$$

1.2 Discrete random variables

As we have seen in Lecture 5, changes of variables are even more straightforward for discrete random variables. It can essentially be seen as careful bookkeeping.

Let X be a discrete random variable with probability mass function

$$\begin{cases} p_X(a_1) = p_1 \\ p_X(a_2) = p_2 \\ \vdots \end{cases}$$

The discrete random variable $Y = g(X)$ with $g(x) = ax+b$, $a > 0$ takes on the values $\alpha_1 = g(a_1)$, $\alpha_2 = g(a_2)$, etc. Now, for any i,

$$p_Y(\alpha_i) = P(Y = \alpha_i) = P(g(X) = g(a_i)) = P(X = a_i) = p_X(a_i) = p_i$$

In other words,

$$\boxed{p_Y(\alpha_i) = p_X(g^{-1}(\alpha_i)) = p_X\left(\frac{\alpha_i - b}{a}\right)}$$

1.3 Examples

- Let X be the continuous random variable corresponding to the ice thickness in inches of the ice over Turtle Pond in Central Park on February 27$^{\text{th}}$ of any year. X takes on values in $0, \infty)$, and is well represented by the following probability density function:

$$\begin{cases} f_X(x) = 0 \text{ if } x < 0 \\ f_X(x) = 2e^{-2x} \text{ if } x \geq 0 \end{cases}$$

  Hence, the cumulative distribution function for X is

$$\begin{cases} F_X(x) = 0 \text{ if } x < 0 \\ F_X(x) = 1 - e^{-2x} \text{ if } x \geq 0 \end{cases}$$

  What is the probability density function and cumulative distribution function of the random variable Y corresponding to the ice thickness in centimeters?

  We have $Y = 2.54X$, so

$$\begin{cases} f_Y(x) = 0 \text{ if } x < 0 \\ f_Y(x) = \frac{1}{2.54} f_X\left(\frac{x}{2.54}\right) = \frac{1}{1.27} e^{-\frac{x}{1.27}} \text{ if } x \geq 0 \end{cases}$$

  and

$$\begin{cases} F_Y(x) = 0 \text{ if } x < 0 \\ F_Y(x) = 1 - e^{-\frac{x}{1.27}} \text{ if } x \geq 0 \end{cases}$$

- To attract customers, a casino has a special offer for the game we discussed in the first page of Lecture 5: it will multiply the gains or losses by 1.5, and give you 3$ per round played. Is it profitable to play this new game?

  Let Y be the amount of money one can win or lose in a round of this new game, and X the amount of money one can win or lose in a round of the old game. We can write

$$Y = 1.5X + 3$$

Thus, Y takes on values in $\{-27, -12, 3, 18, 33\}$, and its probability mass function is given by

$$\begin{cases} p_Y(-27) = p_X\left(\frac{-27-3}{1.5}\right) = p_X(-20) = 0.2 \\ p_Y(-12) = p_X\left(\frac{-12-3}{1.5}\right) = p_X(-10) = 0.3 \\ p_Y(3) = p_X\left(\frac{3-3}{1.5}\right) = p_X(0) = 0.1 \\ p_Y(18) = p_X\left(\frac{18-3}{1.5}\right) = p_X(10) = 0.2 \\ p_Y(33) = p_X\left(\frac{33-3}{1.5}\right) = p_X(20) = 0.2 \end{cases}$$

The expected value of Y is

$$E[Y] = (-27) \cdot 0.2 + (-12) \cdot 0.3 + 3 \cdot 0.1 + 18 \cdot 0.2 + 33 \cdot 0.2 = 1.5 \ \$$$

The casino seems to be keen to have you play this game, so you should do so!

Observe that based on what we learned in Lecture 5, we could have written the formula for $E[Y]$ directly, according to

$$E[Y] = \sum_i g(a_i) p_X(a_i)$$

For illustrative purposes, we instead derived the result from scratch here, by first constructing the change of random variable explicit. Both methods are obviously equivalent.

2 Jensen's formula

In Lecture 5, we derived the formula for the expected value $E[Y]$ of a random variable Y given by $Y = g(X)$:

- For a discrete random variable, we have

$$E[Y] = E[g(X)] = \sum_i g(a_i) p_X(a_i)$$

  as we just saw in the previous section.

- For a continuous random variable, we have

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

We will now see that when g is *convex*, we can provide a lower bound on $E[Y] = E[g(X)]$ without having to actually compute $E[Y]$. This useful result is known as *Jensen's inequality*, or *Jensen's formula*. To understand it, let us first remind ourselves of what convex functions are.

2.1 Convex functions

Definition: A function $f : \mathbb{R} \to \mathbb{R}$ is called **convex** if for all $x_1 \in \mathbb{R}$, $x_2 \mathbb{R}$, and all $t \in [0, 1]$,

$$\underline{f(tx_1 + (1-t)x_2)} \leq tf(x_1) + (1-t)f(x_2)$$

Graphically, a convex function is such that the line segment between any two points on the graph of the function lies above (or on) the graph of the function. This is shown in Figure 1. For illustrative purposes, we chose $t = \frac{2}{3}$ in that figure.

The Danish mathematician Jensen showed that this definition implies the following more general result:
If p_1, p_2,...,p_n are nonnegative real numbers such that $\sum_{i=1}^{n} p_i = 1$, then a function f is convex if for all real numbers $x_1, x_2, \ldots, x_n$,

$$\underline{f(p_1 x_1 + p_2 x_2 + \ldots + p_n x_n) \leq p_1 f(x_1) + p_2 f(x_2) + \ldots + p_n f(x_n)} \tag{1}$$

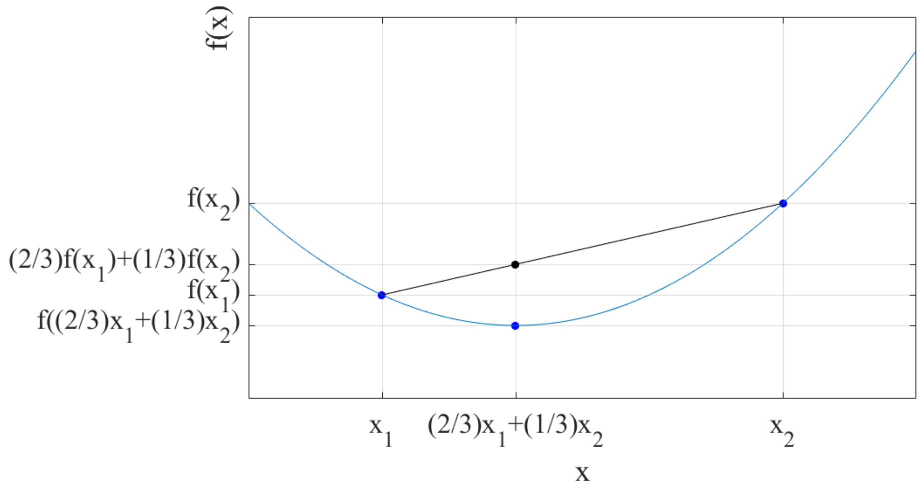

Figure 1: For any convex function, the line segment (shown in black) between any two points on the graph of the function lies above (or on) the graph of the function (shown in blue).

Note that the first definition of a convex function, and the related result given in (1) are not always the most practical methods for verifying whether a function is convex or not. Instead, it is often more convenient to rely on the following simple criterion for determining whether a function is convex:

A twice differentiable function f is convex if and only if

$$\underline{f''(x) \geq 0}$$

for all x in its domain of definition.

2.2 Convexity and expected value

Imagine we have a discrete random variable X with probability mass function $p_X(a_i) = p_i$ for $i = 1, 2, \ldots$, and a discrete random variable Y given by $Y = g(X)$ with g a *convex* function. Then

$$g(E[X]) = g(\sum_i a_i p_i)$$

Now, by definition of a p.m.f., the p_i are all positive, and are also such that

$$\sum_i p_i = 1$$

Hence, since g is convex, we can use (1) to write:

$$g(E[X]) = g\left(\sum_i a_i p_i\right) \leq \sum_i p_i g(a_i) = E[Y]$$

We have just proved that **if g is convex**,

$$\boxed{E[g(X)] \geq g(E[X])}$$

This is known as **Jensen's inequality** in the field of probability and statistics. It applies to discrete random variables as well as to continuous random variables, although we will not prove the result for continuous random variables in this course.

Illustration

Let $g(x) = x^2$. For all $x \in \mathbb{R}$, $g''(x) = 2$, so g is a convex function. Using Jensen's inequality, we can therefore write, for any random variable X:

$$E[X^2] \geq (E[X])^2$$

In other words,

$$\text{Var}(X) = E[X^2] - (E[X])^2 \geq 0$$

so we recover the result we had observed based on a different formula in Lecture 5: the variance of a random variable is always a positive number.

3 Extremes

Let us imagine we roll n fair dice, and are interested in the maximum result among the n dice.

We can introduce n discrete random variables X_1, X_2,...,X_n, each corresponding to the result of one of the dice. To construct a distribution function for the maximum result, we consider the discrete random variable

$$Y = \max(X_1, \dots, X_n)$$

Here is how we can compute the distribution of Y, and in particular its cumulative distribution function.

For any $x \in \mathbb{R}$,

$$F_Y(x) = P(Y \leq x) = P(\max(X_1, \dots, X_n) \leq x) = P(\{X_1 \leq x\} \cap \{X_2 \leq x\} \cap \dots \cap \{X_n \leq x\})$$

In our examples, the events $\{X_i \leq x\}$ are all independent of one another, so

$$P(\{X_1 \leq x\} \cap \{X_2 \leq x\} \cap \dots \cap \{X_n \leq x\}) = P(X_1 \leq x)P(X_2 \leq x) \dots P(X_n \leq x) = F_{X_1}(x)F_{X_2}(x) \dots F_{X_n}(x)$$

In our example, all the F_{X_i} are equal; we may call them all F_X, where X is the random variable corresponding to the result of rolling one dice. This leads to our final expression

$$\underline{F_Y(x) = (F_X(x))^n}$$

It is a good exercise for you to compute the cumulative distribution function F_Z for the random variable $Z = \min(X_1, \dots, X_n)$, and verify the following formula:

$$\underline{F_Z(x) = 1 - (1 - F_X(x))^n}$$