Lecture 3

Working with Distributions

The distribution of a function of X

For example, if we know the distribution of earnings, what is the distribution of log-earnings.

If the random variable Y is defined to

$$Y = h(X)$$

then the CDF of Y is

$$F_Y(y) = P(Y \leq y) = P(h(X) \leq y)$$

The outcome space of Y is the outcome space of X transformed by h.

We usually can reexpress the last form in terms of the CDF of X.

We call $h(x)$ a *monotone function* of x, if either $h(x) < h(y)$ whenever $x < y$ or $h(x) > h(y)$ whenever $x < y$.

If $h(x)$ is a monotone function of x, we can always find $h^{-1}(x)$, the inverse of $h(x)$. If $u = h(x)$, the value of $h^{-1}(x)$ is just u.

In this case

$$F_Y(y) = P(X \leq h^{-1}(y)) = F(h^{-1}(y)).$$

The *uniform distribution* plays a role for distributions similar to the role played by unity for arithmetic.

Suppose we have a continuous for X and the corresponding CDF is strictly increasing when it is not zero or one
(i.e., the density does not have intervals where it is zero).

Consider transforming X by the function $F(x)$,

$$Z = F(X)$$

$F(x)$ gives the percentile in the distribution of x. $F(X)$ is the percentile of a value randomly selected from the distribution.

Intuitively, Z has a uniform distribution on the outcome space $[0, 1]$.

Ex. X represents the grade of a randomly chosen person in the class
$Z = F(X)$ represents the percentile in the class that the person appeared.

Let U be a random variable with a uniform distribution. Then

$$Q(U) \sim X$$

Ex. $Q(U)$ gives the exam grade corresponding to the percentile, and hence the randomly chosen class member.

Numerical Summary Measures

Numerical measures are important properties of population distributions

The overall level of a population is often summarized by the mean, or average value.

For a discrete random variable X, the corresponding concept is that of an *expectation* or *expected value.*

$$\mathbb{E}[X] = \sum_x xP(X = x)$$

where the sum is over the outcome set.

We can also think about expectations of functions of random variables. Let $h(x)$ be a real-valued function for x in the outcome space.

$$\mathbb{E}[h(X)] = \sum_x h(x)P(X = x)$$

For example, consider $h(x) = |x|$ so $\mathbb{E}[|X|]$ is the mean absolute value taken by X.

Other summary measures for probability distributions can be defined in correspondence with their population counterparts.

Ex. the spread of a distribution is often summarized by its *variance*,

$$\mathbb{V}[X] = \mathbb{E}\left[|X - \mathbb{E}[X]|^2\right] = \sum_x |x - \mathbb{E}[x]|^2 P(X = x)$$

where the sum is over the outcome set.

For continuous random variables the definitions of expectation and variance can be based on their probability density functions.

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) dx,$$

$$\mathbb{E}[h(X)] = \int_{-\infty}^{\infty} h(x) f(x) dx,$$

and

$$\mathbb{V}[X] = \mathbb{E}\left[|X - \mathbb{E}[X]|^2\right] = \int_{-\infty}^{\infty} |x - \mathbb{E}[x]|^2 f(x) dx.$$