Lecture 2. Random Variables[1]

So far we have been discussing general experiments. Random variables are experiments with real-valued outcomes. On the other hand, any experiment can be numerically coded into a random variable. Thus, in measure theory probability, random variables are defined as a function mapping from the sample space S to the real line R. Most of the time, we forget (without much loss) about the sample space, and just think of the random variable as an experiment with real-valued outcomes. As such, the sample space that we talk about will be R, and events will be subsets of R. Probability functions will be functions mapping from the Borel σ-field to $[0, 1]$ that satisfy the three axioms of probabilities.

Obviously, fully describing a probability functon for a random variable is difficult since the Borel σ-field contains so many elements. Fortunately, there are much easier functions that convey the same information as the probability function for a random variable. They are the distribution functions and the probability density functions.

1 Distribution Function and Density (or Mass) Function

Distribution Function: $F(x) = P(X \leq x) = P((-\infty, x])$. Note that we use the capital letter X to denote the random variable, and lower case letter x to denote a specific value that the random variable can take. We will follow this convention throughout the course.

The distribution function is also called the cumulative distribution function (CDF). We often write this as $X \sim F$.

A CDF has some unique properties:

1. $\lim_{x\to\infty} F(x) = 1$, $\lim_{x\to-\infty} F(x) = 0$.
2. $F(x)$ is non-decreasing.
3. $F(x)$ is right-continuous, that is $\lim_{x\downarrow x_0} F(x) = F(x_0)$.

Continuous Random Variable. A random variable X is continuous if F is continuous for all x.

Discrete Random Variable. A random variable X is discrete if F is a step function.

Density Function. If F is continuous and differentiable, then we define its density function as

$$f(x) = \frac{d}{dx}F(x).$$

[1] This lecture note is largely adapted from Professor Bruce Hansen's handwritten notes, however, all errors are mine.

In general, when F is continuous, $f(x)$ is the function which satisfies

$$F(x) = \int_{\infty}^{x} f(t)dt.$$

A density function is also called the probability density function (PDF). The definition of the PDF immediately implies that

$$P(a \leq X \leq b) = \int_{a}^{b} f(x)dx.$$

If F is discrete, we then define the probability mass function (PMF):

$$f(x) = P(X = x) = \lim_{z \to x^+} F(x) - \lim_{z \to x^-} F(x).$$

Notation-wise, we often write $X \sim F$ or $X \sim f$ if X is a random variable with distribution function F or density function (or mass function for discrete r.v.s) f.

The density function also has several unique properties: f is a density function if

1. $f(x) \geq 0$ for all x.
2. $\int_{-\infty}^{\infty} f(x)dx = 1$.

Support Set of a distribution: Suppose that $f_X(\cdot)$ is the PDF or PMF of X. The support set of the distribution of X, often denoted $\mathcal{X}$ is defined as

$$\mathcal{X} = \{x : f_X(x) > 0\}. \tag{1}$$

2 Transformation of Random Variables

In econometrics, some variables of interest are transformations of other variables. Suppose that $X \sim F$. Let $Y = g(X)$ be a transformation of X, where $g : \mathcal{X} \to \mathcal{Y}$ is a known function and $\mathcal{Y} = \{y : y = g(x) \text{ for some } x \in \mathcal{X}\}$. What is the distribution of Y? Finding the CDF and PDF (or PMF) of Y based on those of X is not always straigtforward.

If X is discrete, the task is easy: Suppose that

$$P(X = \tau_j) = \pi_j \text{ for } j = 1, 2, 3, \ldots$$

Then we have

$$P(Y = g(\tau_j)) = \sum_{j' : g(\tau'_j) = g(\tau_j)} \pi_j \text{ for } j = 1, 2, 3, \ldots.$$

If X is continuous, the CDF and PDF have analytical solutions only in some special cases.
Case 1. Assume that g is strictly monotone. Then there is an inverse function

$$h(y) = g^{-1}(y),$$

such that $X = h(Y)$. Suppose that g is increasing, then

$$\begin{aligned} F_Y(y) &= P(g(X) \leq y) \\ &= P(X \leq h(y)) \\ &= F_X(h(y)). \end{aligned} \tag{2}$$

This gives us the CDF of Y. To find the PDF, take the first derivative w.r.t. y:

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_X(h(y)) \\ &= f_X(h(y))h'(y) \\ &= f_X(g^{-1}(y))\frac{d}{dy} g^{-1}(y). \end{aligned} \tag{3}$$

On the other hand, suppose that g is decreasing, then

$$\begin{aligned} F_Y(y) &= P(g(X) \leq y) \\ &= P(X \geq h(y)) \\ &= 1 - F_X(h(y)). \end{aligned} \tag{4}$$

This gives us the CDF of Y. To find the PDF, take the first derivative w.r.t. y:

$$\begin{aligned} f_Y(y) &= -\frac{d}{dy} F_X(h(y)) \\ &= -f_X(h(y))h'(y) \\ &= f_x(g^{-1}(y))\frac{-d}{dy} g^{-1}(y). \end{aligned} \tag{5}$$

The above arguments prove the following theorms:

Theorem (Theorems 2.1.3 of CB)**.** *Let X have CDF $F_X(x)$, and let $\mathcal{X}$ and $\mathcal{Y}$ be defined as above. Then*

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

Theorem (Theorem 2.1.5 of CB). *Let X have PDF $f_X(x)$, and let $Y = g(X)$, where g is a monotone function. Let $\mathcal{X}$ and $\mathcal{Y}$ be defined as above. 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*

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| & y \in \mathcal{Y} \\ 0 & \textit{otherwise.} \end{cases} \tag{6}$$

Here $J(y) := \left| \frac{d}{dy} g^{-1}(y) \right|$ is called the Jacobian of the transformation.

Example (Uniform-exponential relationship). *Suppose that $X \sim U[0,1]$, that is,*

$$f_X(x) = \begin{cases} 1 & \textit{if } x \in \mathcal{X} := [0,1] \\ 0 & \textit{else} \end{cases}$$

Suppose that $Y = g(X) = -\log(X)$. Then $\mathcal{Y} = (0, \infty)$. This is an increasing transformation, and

$$h(y) = g^{-1}(y) = \exp(-y).$$

Take derivative to obtain the Jacobian of the transformation:

$$J(y) = \left| \frac{d}{dy} h(y) \right| = |-\exp(-y)| = \exp(-y).$$

Therefore, we have $f_Y(y) = \exp(-y), y \geq 0$. This density is called the exponential density.

Example (Probability Integral Transformation). *Let X have continuous CDF $F_X(x)$ and define the random variable $Y = F_X(X)$. Then Y is uniformly distributed on $(0,1)$.*

Proof. Here we prove only the case where $F_X(\cdot)$ is strictly increasing, so that its inverse F_X^{-1} exists. Consider the CDF of Y, for $y \in (0,1)$:

$$\begin{aligned} P(Y \leq y) &= P(F_X(X) \leq y) \\ &= P(X \leq F_X^{-1}(y)) \\ &= F_X(F_X^{-1}(y)) \\ &= y. \end{aligned} \tag{7}$$

Thus, $f_Y(y) = 1\{y \in (0,1)\}$. This is a uniform density on $(0,1)$.

For the case where $F_X(\cdot)$ is not strictly increasing, we define

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

This defines the quantile function of X. The statement above still holds in this case. Chapter 2 of CB has detailed arguments. □

Case 2. Assume that g is the quadratic function: $g(x) = x^2$. This function is not monotonic. We can find the CDF of Y as follows: For $y \geq 0$,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X^2 \leq y) \\ &= P(|X| \leq \sqrt{y}) = P(-\sqrt{y} \leq x \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}). \end{aligned} \tag{8}$$

The PDF is

$$f_Y(y) = f_X(\sqrt{y})/(2(\sqrt{y}) + f_X(-\sqrt{y})/(2\sqrt{y}) = \frac{f_X(\sqrt{y}) + f_X(-\sqrt{y})}{2\sqrt{y}}. \tag{9}$$

Example. *Suppose that* $X \sim f_X(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$, *that is,* X *has a standard normal density. Then*

$$f_X(\sqrt{y}) = \frac{1}{\sqrt{2\pi}} \exp(-y/2).$$

And thus

$$f_Y(y) = \frac{1}{\sqrt{2\pi y}} \exp(-y/2).$$

This is known as the χ_1^2 *distribution (Ki-squared one).*

3 Expectation of Random Variables

While probability function, distribution function, and density functions are all complete descriptions of the distribution of a random variable, the expectation, or expected value, is a crude summary of the distribution. Roughly speaking, it reflects the average value, the typical value, or the central tendency. It competes with mode, median, etc. for such roles, and may not be the most accurate or appropriate candidate for any of them. However, the expectation wins the statistician's favor because of its simplicity and ease of use.

The expectation is defined as

$$E(X) = \sum_{x \in \mathcal{X}} f_X(x)x,$$

if X is discrete, and as

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

if X is continuous, if the sum or the integral exists.

Example (Uniform)**.** *$X \sim U[0,1]$. Then $E(X) = \int_0^1 x dx = 1/2$.*

Example (Exponential)**.** *$X \sim f_X(x) = \frac{1}{\lambda} \exp\left(-\frac{x}{\lambda}\right)$. Then $E(X) = \int_0^\infty \frac{x}{\lambda} \exp\left(-\frac{x}{\lambda}\right) dx = \lambda$.*

Expectation of Transformation. If $Y = g(X)$ and $X \sim f_X$, then

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

No need to work out the density function of Y.

Example. *Suppose that $X \sim U[0,1]$. Then $E[X^2] = \int_0^1 x^2 dx = 1/3$.*

Theorem (Linearity of Expectation)**.** *For any constants a and b, we have $E[a + bX] = a + bE[X]$.*

Proof. omitted. □

Moments and Central Moments. We say that the mth moment of X is $\mu_m := E(X^m)$. We also define the central moments:

$$E(X - E(X))^m.$$

Example. *The variance is the second central moment: $\sigma^2(X) = E(X - E(X))^2 = E(X^2) - (E(X))^2$.*

The variance is not *a linear operator. Instead:*

$$\sigma^2(a + bX) = b^2 \sigma^2(X).$$

We typically use the expectation (mean) and variance to summarize the center and the spread of the distribution.

Example (Bernoulli)**.** *$P(X = 1) = p$ and $P(X = 0) = 1 - p$. Then $E(X) = 1 * p + 0 * (1 - p) = p$, and $Var(X) = (1 - p)^2 * p + (0 - p)^2 * (1 - p) = p(1 - p)$.*

Moment Generating Function. This is a technical tool used to facilitate some proofs. We define the following function of t for $t \geq 0$ as the moment generating function of $f_X(x)$:

$$M_X(t) = E[\exp(tX)] = \int_{-\infty}^{\infty} \exp(tx) f_X(x) dx.$$

Note that the moment generating function may or may not exist. When it does exist, it ahs the following properties:

1. $\frac{d}{dt} M(t)\big|_{t=0} = \int_{-\infty}^{\infty} x \exp(tx) f_X(x) dx\Big|_{t=0} = \int_{-\infty}^{\infty} x f_X(x) dx = E(X)$.
2. $\frac{d^m}{dt^m} M(t)\big|_{t=0} = \int_{-\infty}^{\infty} x^m \exp(tx) f_X(x) dx\Big|_{t=0} = \int_{-\infty}^{\infty} x^m f_X(x) dx = E(X^m)$.

That is why it is called the moment generating function.

Because the moment generating function does not always exist, we also define the characteristic function (CF):

$$C_X(t) = E[\exp(itX)], \text{ where } i = \sqrt{-1}.$$

It has similar properties as $M_X(t)$, and always exists.

So far we have defined $E(X)$ seperately for discrete and continuous X's. There is actually a unified notation for both, which we may use in the rest of the course:

$$E(X) = \int_{-\infty}^{\infty} x dF_X(x),$$

This notation is the notation of the Riemann-Stieltijes integral, defined as:[2]

$$\int_{-\infty}^{\infty} g(x)dF(x) = \lim_{M\to\infty} \lim_{N\to\infty} \sum_{j=1}^{N} g\left(\frac{2(j-1)M}{N} - M\right)\left(F_X\left(\frac{2jM}{N} - M\right) - F_X\left(\frac{2(j-1)M}{N} - M\right)\right).$$

4 Problems

1. A CDF F_X is stochastically greater than a CDF F_Y if $F_X(t) \leq F_Y(t)$ for all t and $F_X(t) < F_Y(t)$ for some t. Prove that if $X \sim F_X$ and $Y \sim F_Y$, then

   $$P(X > t) \geq P(Y > t) \text{ for every } t$$

   and

   $$P(X > t) > P(Y > t) \text{ for some } t,$$

   that is, X tends to be bigger than Y.

2. Show that the function $F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - \exp(-x) & \text{if } x \geq 0 \end{cases}$ is a CDF, and find $f_X(x)$ and $F_X^{-1}(y)$.

3. Suppose that $Y = X^3$ and $f_X(x) = 42x^5(1-x), x \in (0,1)$. Find the PDF of Y, and show that the PDF integrates to 1.

[2]This definition is shown here for completeness (or to impress you:-)), and we will not need to use it anywhere else in this course.

4. Consider the CDF $F_X(x) = \begin{cases} 1.2x & \text{if } x \in [0, 0.5) \\ 0.2 + 0.8x & \text{if } x \in [0.5, 1] \end{cases}$, and the function

$$f_X(x) = \begin{cases} 1.2 & \text{if } x \in [0, 0.5) \\ a & \text{if } x = 0.5 \\ 0.8 & \text{if } x \in (0.5, 1] \end{cases}$$

Show that f_X is the density function of F_X as long as $a \geq 0$. That is, show that for all $x \in [0, 1]$, $F_X(x) = \int_0^x f_X(t)dt$.

5. Let X have PDF $f_X(x) = \frac{2}{9}(x+1),\ x \in [-1, 2]$. Find the PDF of $Y = X^2$. Note that this is a bit different from the exercise in the lecture note.

6. A median of a distribution is a value m such that $P(X \leq m) \geq 1/2$ and $P(X \geq m) \geq 1/2$. Find the median of the distribution $f(x) = \frac{1}{\pi(1+x^2)}\ \ x \in R$.

7. Show that if X is a continuous random variable, then

$$\min_a E|X - a| = E|X - m|,$$

where m is the median of X. (Hint: workout the integral expression of $E|X - a|$ and notice that it is differentiable.)

8. Let μ_n denote the nth central moment of a random variable X. Two quantities of interest, in addition to the mean and variance are

$$\alpha_3 = \frac{\mu_3}{\mu_2^{3/2}} \text{ and } \alpha_4 = \frac{\mu_4}{\mu_2^2}.$$

The values α_3 is called the skewness and α_4 is called the kurtosis. The skewness measures the lack of symmetry in the density function. The kurtosis, althugh harder to interpret, measures the peakedness or flatness of the density function.

   (a) Show that if a density function is symmetric about a point a, then $\alpha_3 = 0$.

   (b) Calculate α_3 for $f(x) = \exp(-x),\ x \geq 0$, a density function that is skewed to the right.

   (c) Calculate α_4 for the following density functions and comment on the peakedness of each:

$$f(x) = \frac{1}{\sqrt{2\pi}}\exp(-x^2/2),\ x \in R$$
$$f(x) = 1/2,\ x \in (-1, 1)$$
$$f(x) = \frac{1}{2}\exp(-|x|),\ x \in R.$$