ISEG - Lisbon School of Economics and Management
Statistics I
(Date of this version: 15/10/2017)
Handout 2 (Part 1) – Random Variables

Issues covered:

2.1 Random variable

Random variable: is a variable that takes on numerical values and has an outcome that is determined by an experiment.

Formal Definition: Let $\mathbf{S}$ be a sample space with a probability measure. A random variable (or stochastic variable) X is a real-valued function defined over the elements of $\mathbf{S}$:

$$X : s \in \mathbf{S} \to X(s) \in \mathbb{R}.$$

Remark: Although a random variable is a function of s, usually we drop the argument, that is we write X, rather than $X(s)$.

Example: We flip a coin 2 times and count the number of times the coin turns up heads. $X =$"number of Heads".

Remarks:

1. The function X makes each $s \in \mathbf{S}$ correspond to one and just one $X(s) \in \mathbb{R}$
2. Typically random variables are expressed with capital letters and the values that the random variable can assume are expressed as lower case letters
3. Once the random variable is defined, $\mathbb{R}$ is the space in which we work with.
4. The fact that definition of random variable is limited to real-valued functions does not impose any restrictions. If, for example, the outcomes of an experiment are of the categorical type, like the colour of a person's hair, we can arbitrarily make the descriptions real-valued by coding the various colours, perhaps by representing them with the numbers 1, 2, 3, and so on.
5. Although the definition of random variable does not rely explicitly on the concept of probability, it is introduced to make easier the computation of probabilities. Suppose that we want to compute the probability that a random variable is in a set $B \subset \mathbb{R}$. Note $P(X \in B) = P(A)$ for $A = \{s \in \mathbf{S} : X(s) \in B\} \subset \mathbf{S}$, by the definition of random variable. Hence we are computing the probability of an event in $\mathbf{S}$.

Example: We flip a balanced coin 2 times and define a random variable $X =$"number of Heads". We obtain:

X	Possible Outcomes	Probability
0	(Tails, Tails)	1/4
1	(Tails, Heads)(Heads, Tails)	2/4
2	(Heads, Heads)	1/4

There are 4 possible outcomes and all are equally likely.

2.2 Cumulative Distribution Function

Definition: If X is a random variable, the real function F_X with domain $\mathbb{R}$ is defined as

$$F_X(x) = P(X \leq x) = P(X \in (-\infty, x])$$

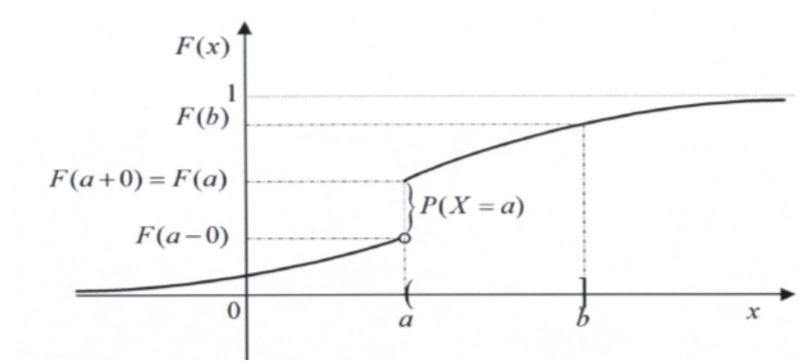
is the cumulative distribution function of the random variable X.

The properties of the measure probabilities of events imply the following properties of cumulative distribution functions:

1. $0 \leq F_X(x) \leq 1$;
2. $F_X(x)$ is non-decreasing: $\forall \Delta_x > 0 : F_X(x) \leq F_X(x + \Delta_x)$.
3. $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to +\infty} F_X(x) = 1$.
4. $P(a < X \leq b) = F_X(b) - F_X(a)$, for $b > a$
5. $\lim_{x \to a^+} F_X(x) = F_X(a)$; therefore X is **right continuous**
6. $P(X = a) = F_X(a) - \lim_{x \to a^-} F_X(a)$ for any real finite number.

Remark: The set of discontinuities of the cumulative distribution function D_X is given by $D_X = \{x \in \mathbb{R} : P(X = x) > 0\}$. Note that by property 6 this the same as $D_X = \{a \in \mathbb{R} : F_X(a) - \lim_{x \to a^-} F_X(x) > 0\}$.

Example:



Further properties:

1. $P(X < b) = \lim_{x \to b^-} F_X(x)$

2. $P(X > a) = 1 - F_X(a)$

3. $P(X \geq a) = 1 - \lim_{x\to a^-} F_X(x)$.

4. $P(a < X < b) = \lim_{x\to b^-} F_X(x) - F_X(a)$.

5. $P(a \leq X < b) = \lim_{x\to b^-} F_X(x) - \lim_{x\to a^-} F_X(x)$

6. $P(a \leq X \leq b) = F_X(b) - \lim_{x\to a^-} F_X(x)$

The random variables that we typically work with are of the following type:

1. Discrete

2. Continuous

3. Mixed

Remark: Random variables may neither be discrete, nor continuous, or mixed.

2.3 Discrete random variable

Discrete random variable: is a random variable that takes only a discrete set of values.

Example: We flip a balanced coin 2 times and define a random variable $X =$"number of Heads. In the example X takes 3 possible values $(0, 1, 2)$ and the associated probabilities are $(1/4, 1/2, 1/4)$ respectively.

In general a discrete random variable takes a finite number of possible values or a countably infinite. For the sake of generality we consider the latter case.

Formal Definition: X is a discrete random variable if $D_X \neq \varnothing$ and if it takes the values x_1, x_2, x_3... with associated probabilities $P(X = x_1)$, $P(X = x_2)$, $P(X = x_3)$... , then these probabilities satisfy $P(X \in D_X) = \sum_{j=1}^{\infty} P(X = x_j) = 1$.

Remark: In this case $D_X = \{x_1, x_2, x_3...\}$.

Formal Definition: The *probability function (also known probability distribution or as probability mass function)* $f_X(x)$ is the function $f_X(x) = \mathcal{P}(X = x)$ for each x within the range of X.

Theorem: A function can serve as the probability function of a discrete random variable X if and only if its values, $f_X(x)$, satisfy the conditions

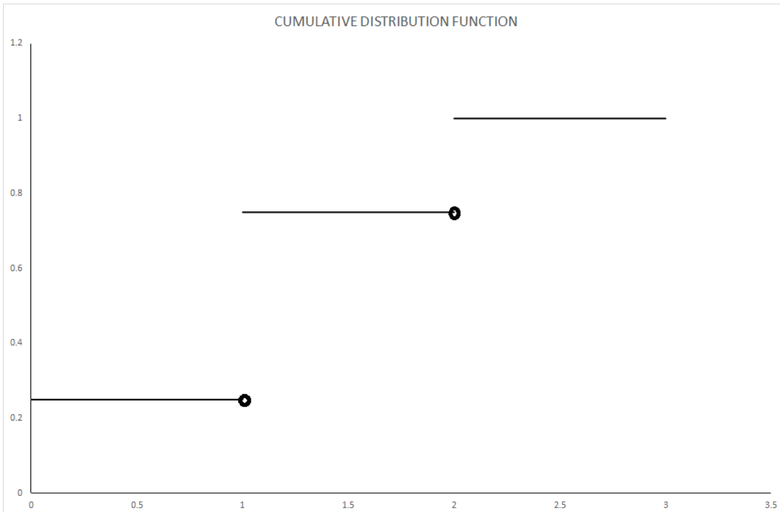
1. $0 \leq f_X(x_j) \leq 1$, $j = 1, 2, 3, ...$

2. $\sum_{j=1}^{\infty} f_X(x_j) = 1$.

Therefore the *cumulative probability distribution* is given by : $F_X(x) = P(X \leq x) = \sum_{x_j \leq x} f_X(x_j)$.

Example (cont):

X	Probability distribution	Cumulative probability distribution
0	1/4	1/4
1	2/4	3/4
2	1/4	1



Example: Bernoulli Random variable

We flip a coin and define a random variable

$$X = \begin{cases} 1 & \text{if Heads} \\ 0 & \text{if Tails} \end{cases}$$

Let us denote

$$P(Heads) = P(X = 1) = p$$

Then

$$P(Tails) = P(X = 0) = 1 - p.$$

This can be written as

$$f_X(x) = p^x(1-p)^{1-x}, x = 0, 1.$$

If the coin is balanced $p = 0.5$.

A random variable that is defined as

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases}$$

is known as a *Bernoulli Random variable*, named after the Swiss mathematician Jacob Bernoulli (1654-1705). One outcome is arbitrarily labeled a "success" (denoted $X = 1$) and the other a "failure" (denoted $X = 0$).

Exercise: Find the distribution function of the total number of heads obtained in four tosses of a balanced coin.

Exercise: Check whether the function given by $f(x) = \frac{x+2}{25}$, for $x = 1, 2, 3, 4, 5$ can serve as the probability distribution of a discrete random variable

2.4 Continuous random variable

Continuous random variable: are random variables that take a continuum of possible values.

Example: Suppose that we are concerned with the possibility that an accident will occur on a highway that is 200 kilometers long and that we are interested in the probability that it will occur on a given stretch of the road. The sample space of this "experiment" consists of a continuum of points, those on the interval from 0 to 200, and we shall assume, for the sake of argument, that the probability that an accident will occur on any interval of length d is $d/200$, with d measured in kilometers. Note that this assignment of probabilities is consistent with Postulates 1 and 2. (Postulate 1 states that probability of an event is a nonnegative real number; that is, $P(A) \geq 0$ for any subset A of $\mathbf{S}$ but in Postulate 2 $P(\mathbf{S}) = 1$.) The probabilities $d/200$ are all nonnegative and $P(\mathbf{S}) = 200/200 = 1$. So far this assignment of probabilities applies only to intervals on the line segment from 0 to 200, but if we use Postulate 3 or 3* (Postulate 3 or 3*: If A_1, A_2 , A_3, ... , is a finite or infinite sequence of mutually exclusive events of $\mathbf{S}$, then $P(A_1 \cup A_2 \cup A_3 \cup ...) = P(A_1) + P(A_2) + P(A_3) + ...$, we can also obtain probabilities for the union of any finite or countably infinite sequence of nonoverlapping intervals. For instance, the probability that an accident will occur on either of two nonoverlapping intervals of length d_1 and d_2 is $(d_1 + d_2)/200$ and the probability that it will occur on any one of a countably infinite sequence of nonoverlapping intervals of length $d_1, d_2, d_3, ...$ is $(d_1 + d_2 + d_3 + ...)/200$.

Formal Definition: X is a continuous random variable if and only if $D_X = \varnothing$ and there is a non-negative function $f_X(x)$, such that

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt, \ \forall x \in \mathbb{R}. \tag{1}$$

Remarks:

1. Condition (1) implies that $F_X(x)$ is a continuous function. However, continuity of $F_X(x)$ does not imply condition (1).
2. Continuous cumulative probability functions are smooth functions. Unlike discrete distributions, the probability of any single point = 0.
3. Note that $P(X \in D_X) = P(X \in \varnothing) = 0$.
4. The function $f_X(x)$ is known as *probability density function.* It that provides information on how likely the outcomes of the random variable are.

Theorem. A function can serve as a probability density function of a continuous random variable X if its values, $f_X(x)$, satisfy the conditions:

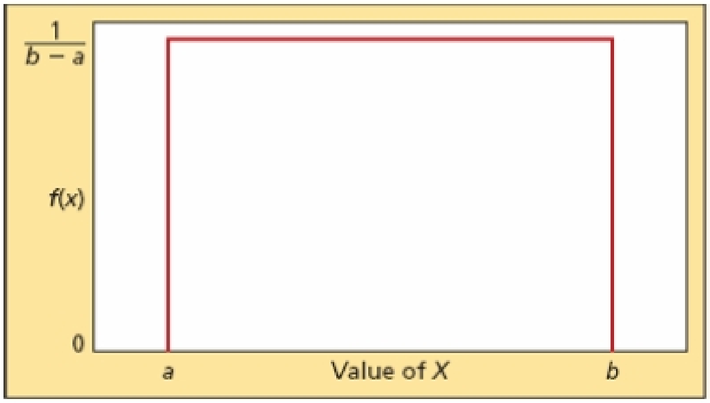
1. $f_X(x) \geq 0$ for$-\infty < x < +\infty$;
2. $\int_{-\infty}^{+\infty} f_X(x)dx = 1$

Example: *Uniform Continuous Distribution* $X \sim U(a, b)$

If X is a random variable that is uniformly distributed between a and b, its probability density function has constant height:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{otherwise} \end{cases},$$

Probability density function of $U(a, b)$



The cumulative distribution function is

$$F_X(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \le x \le b \\ 1 & x > b \end{cases}$$

Theorem. If $f_X(x)$ and $F_X(x)$ are the values of the probability density and the distribution function of X at x, then

$$\begin{aligned} P(a \le X \le b) &= F_X(b) - F_X(a) \\ &= \int_a^b f_X(t)dt \end{aligned}$$

for any real constants a and with $a \le b$, and

$$f_X(x) = \frac{dF_X(x)}{dx},$$

except perhaps at a finite number of points.

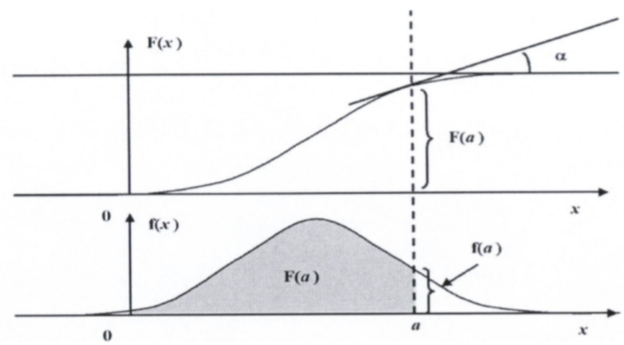
Remarks:

1. In each of the points x at which $F_X(x)$ has no derivative at a finite it does not matter the value that we give to $f_X(x)$ as it does not affect the computation of $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$.
2. The probability density function is not a probability and therefore it can assume values bigger than one.
3. If X is a continuous random variable $P(X = a) = \int_a^a f_X(t)dt = 0$.

Theorem. If X is a continuous random variable and a and b are real constants with $a \le b$, then

$$
\begin{aligned}
P(a &\leq X \leq b) = P(a \leq X < b) \\
&= P(a < X \leq b) \\
&= P(a < X < b)
\end{aligned}
$$

Probability density function and cumulative distribution function of a continuous random variable



2.5 Mixed random variables

Definition: X is a mixed random variable if $D_X \neq \varnothing$ but $0 < P(X \in D_X) < 1$ and

$$F_X(x) = \lambda F_{X_1}(x) + (1-\lambda) F_{X_2}(x), \lambda \in (0,1)$$

and X_1 is a discrete random variable and X_2 is a continuous random variable.

Exercise: Let

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{12} + \frac{3}{4}(1 - e^{-x}) & 0 \leq x < 1 \\ \frac{1}{4} + \frac{3}{4}(1 - e^{-x}) & x \geq 1 \end{cases},$$

Compute $P(X = 0)$, $P(X = 1)$, $P(0.5 < X < 1)$ and $P(0.5 < X < 2)$

2.6 Distribution of functions of random variables

Let X be a random variable with known cumulative distribution function $F_X(x)$. Let $Y = g(X)$ where $g : \mathbb{R} \to \mathbb{R}$. Let $F_Y(y)$ be the cumulative distribution function of Y. How can we derive $F_Y(y)$ from $F_X(x)$?

Discrete random variables

If X is a discrete random variable then $Y = g(X)$ is also a discrete random variable. We determine the probability function of Y, $f_Y(y)$ from $f_X(x)$.

Let $D_X = \{x_1, x_2, x_3...\}$ be the set of discontinuities of $F_X(x)$, then $D_Y = \{g(x_1), g(x_2), g(x_3)...\}$ is the set of discontinuities of $F_Y(y)$. Given $y \in D_Y$, $A_y = \{x : g(x) = y, x \in D_X\}$, The probability function of Y is given by $f_Y(y) = P(Y = y) = P(X \in A_y) = \sum_{x_i \in A_y} f(x_i)$

Example: Consider the discrete random variable X with probability function

x	-2	-1	0	1	2
$f_X(x)$	$12/60$	$15/60$	$10/60$	$6/60$	$17/60$

Let $Y = X^2$, what is $f_Y(y)$? Note that $D_Y = \{0, 1, 4\}$
Consequently

- $f_Y(0) = P(Y = 0) = P(X \in A_0)$, where $A_0 = \left\{x : x^2 = 0\right\}$, hence $P(X \in A_0) = P(X^2 = 0) = P(X = 0) = 10/60$
- $f_Y(1) = P(Y = 1) = P(X \in A_1)$, where $A_1 = \left\{x : x^2 = 1\right\}$, hence $P(X \in A_1) = P(X^2 = 1) = P(X = 1 \text{ or } X = -1) = P(X = 1) + P(X = -1) = 6/60 + 15/60 = 21/60$.
- $f_Y(4) = P(Y = 4) = P(X \in A_4)$, where $A_4 = \left\{x : x^2 = 4\right\}$, hence $P(X \in A_4) = P(X^2 = 4) = P(X = 2 \text{ or } X = -2) = P(X = 2) + P(X = -2) = 17/60 + 12/60 = 29/60$.

The General Case

Let X be any random variable and $Y = g(X)$, the derivation of $F_Y(y)$ is based on the equality $F_Y(y) = P(Y \leq y) = P(X \in A_y^*)$, where $A_y^* = \{x : g(x) \leq y\}$

Example: Let $Y = |X|$, note that

$$\begin{aligned} P(Y \leq y) &= P(|X| \leq y) = P(-y \leq X \leq y) \\ &= F_X(y) - \lim_{x \to (-y)^-} F_X(x) \end{aligned}$$

Exercise: Derive the distribution function of $Y = aX + b$, where $a < 0$ and $Y = X^2$.