

Random Variables

ECON 340: Economic Research Methods

Instructor: Div Bhagia

1 Single Random Variable

A random variable is a variable that takes different values under different scenarios. The likelihood of these different scenarios is summarized by the distribution of the random variable. We will denote a random variable by X and realizations of it by x .

Random variables can either be discrete or continuous. A *discrete* random variable has a countable number of possible values. While *continuous* random variables can take any value in a given interval.

1.1 Discrete Random Variables

The *probability distribution function* (PDF) for a discrete random variable X is given by:

$$f(x) = Pr(X = x)$$

where $0 \leq f(x) \leq 1$ for all x and $\sum_x f(x) = 1$.

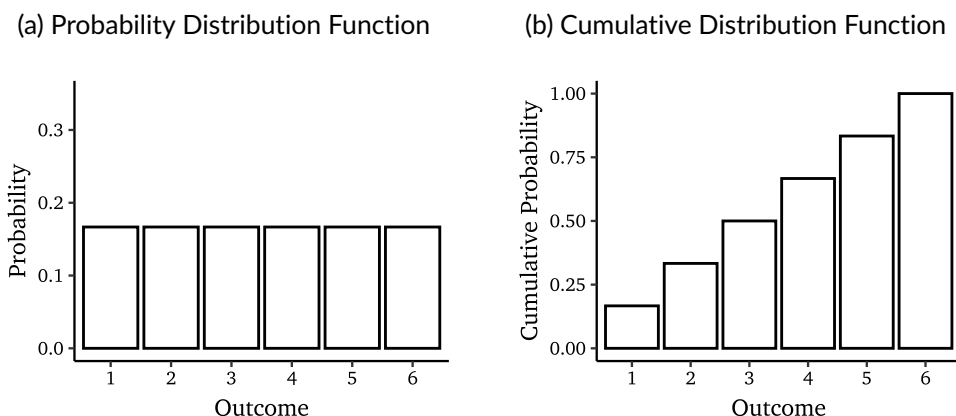
The *cumulative distribution function* (CDF) for a discrete random variable X is given by:

$$F(x_0) = Pr(X \leq x_0) = \sum_{x \leq x_0} f(x)$$

Example. X is the outcome of rolling a die.

x	$f(x)$	$F(x)$
1	1/6	1/6
2	1/6	2/6
3	1/6	3/6
4	1/6	4/6
5	1/6	5/6
6	1/6	1

Figure 1: Outcome from a Die Roll



Bernoulli Random Variable is a special type of discrete random variable that only takes two values 1 and 0. It is also called a *binary* variable.

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

1.2 Continuous Random Variables

Because of continuum of possible values, it is not feasible to list the probability of each possible value of a continuous random variable. So we instead have the *probability density function* (PDF), denoted by $f(x)$. The area under the curve $f(x)$ gives us the probability of X falling in certain intervals. The *probability density function* is defined as:

$$Pr(a \leq x \leq b) = \int_a^b f(x) dx$$

where $f(x) > 0$ for all x and $\int_{-\infty}^{\infty} f(x) dx = 1$.

Note that for continuous random variables, $Pr(X = x) = 0$. This is just to say that it is very very unlikely that any particular value will be realized because there are infinite possibilities.

The integral \int_a^b is the continuous analog of the sum. You don't need to know how to solve an integral, but remember it is like taking a sum over continuous values. The limits

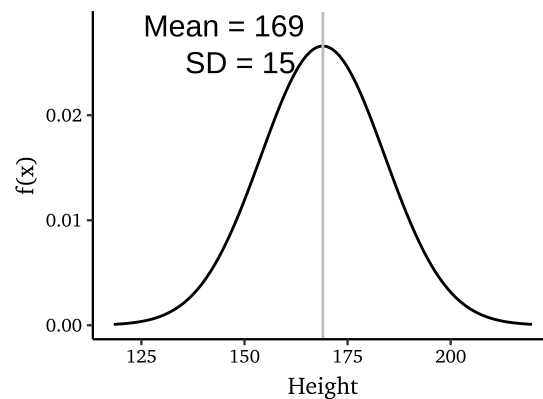
of the integral, a and b , define the interval over which we are taking this sum.

The *cumulative density function* (CDF) for a continuous random variable X is given by:

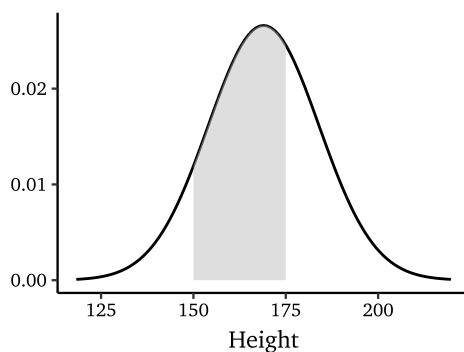
$$F(x_0) = \Pr(X \leq x_0) = \int_{-\infty}^{x_0} f(x) \partial x$$

Note that, $\Pr(a \leq x \leq b) = F(b) - F(a)$.

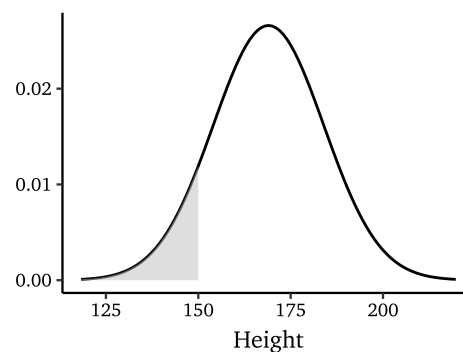
Example. Let's say the distribution for the height of individuals in the world is given by the following probability density function:



We need to find the corresponding area under the curve to find the probability of height being in particular intervals.

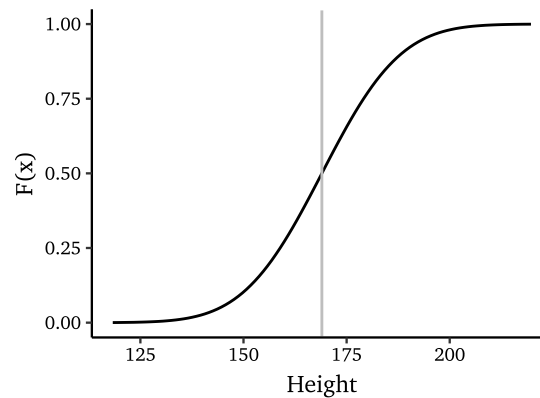


(a) $\Pr(150 < X < 175)$



(b) $\Pr(X < 150)$

The CDF corresponding to the above PDF looks like:



1.3 Expectation and Variance of Random Variables

The *expectation* or *expected value* or the *mean* of a random variable gives us the average value of this variable over many repeated trials or occurrences. We can compute the expectation as a weighted average of possible outcomes, where the weights are probabilities.

The expectation of a discrete random variable is given by:

$$\mu_X = E(X) = \sum_x x f(x)$$

The expected value for a continuous random variable is given by:

$$\mu_X = E(X) = \int_x x f(x) dx$$

The *variance* measures the dispersion or the “spread” of a probability distribution. The variance for a random variable is given by:

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

In other words, variance is the expected value of the square of deviations of X from its mean.

Using the definition of expectation for a discrete random variable, we can write:

$$\sigma_X^2 = \text{Var}(X) = E[(X - \mu_X)^2] = \sum_x (x - \mu_X)^2 f(x)$$

Alternative formula for the variance: $\text{Var}(X) = E[(X - \mu_X)^2] = E(X^2) - \mu_X^2$.

Variance is in units of the square of X . Therefore we use *standard deviation*, which is the square-root of variance:

$$\sigma_X = \sqrt{\sigma_X^2}$$

Example. We can calculate the expected value and variance for the outcome of rolling a die as follows:

x	$f(x)$	$xf(x)$	$(x - \mu_X)^2 f(x)$
1	1/6	1/6	$(-2.5)^2/6$
2	1/6	2/6	$(-1.5)^2/6$
3	1/6	3/6	$(-0.5)^2/6$
4	1/6	4/6	$(0.5)^2/6$
5	1/6	5/6	$(1.5)^2/6$
6	1/6	6/6	$(2.5)^2/6$
Total	-	21/6	17.5/6

$$\mu_X = E(X) = \sum_x xf(x) = \frac{21}{6} = 3.5$$

$$\text{Var}(X) = \sum_x (x - \mu_X)^2 f(x) = \frac{17.5}{6}$$

Example. The expected value and variance for a binary random variable that takes value 1 with probability p and 0 with probability $1 - p$ is given by:

$$\mu_X = E(X) = \sum_x xf(x) = 1 \cdot p + 0 \cdot (1 - p) = p$$

$$\text{Var}(X) = \sum_x (x - \mu_X)^2 f(x) = (1 - p)^2 \cdot p + (0 - p)^2 \cdot (1 - p) = p(1 - p)$$

Using the alternative formula for variance:

$$\text{Var}(X) = E(X^2) - \mu_X^2 = 1^2 \cdot p + 0^2 \cdot (1 - p) - \mu_X^2 = p - p^2 = p(1 - p)$$

2 Normal and Standard Normal Distribution

2.1 Normal Distribution

There are lots and lots of probability distributions that are used for modeling different variables. For example, *uniform* distribution is a distribution with constant probability. There are many more, Bernoulli, binomial, gamma, beta, Poisson, and so on.

However, one distribution that appears over and over again is the *normal distribution*. The distribution of a lot of things like height, birthweight, IQ, etc. is normal. The normal distribution is symmetric (i.e., the left and right tails are the same sizes and there's no skew). For this reason, sometimes it is informally referred to as a bell curve. We express normal distribution with mean μ and variance σ^2 as follows:

$$N(\mu, \sigma^2)$$

Figure 2 presents normal distributions with different means and variances.

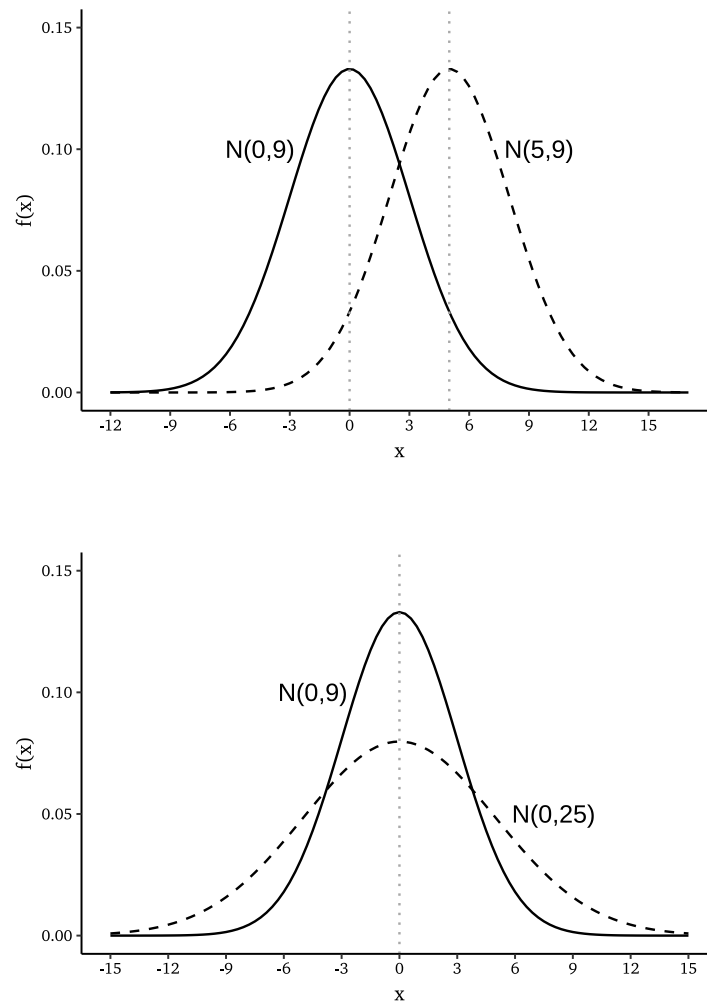
The distribution of height that we saw in the above example was a normal distribution with a mean of 169 and a variance of 225. In which case, we can write $\text{height} \sim N(169, 225)$, which is short-form for saying height is normally distributed with mean 169 and variance 225.

2.2 Standard Normal Distribution

The normal distribution with mean 0 and variance 1 is called the *standard normal distribution* and is denoted by $N(0, 1)$. Random variables that have a $N(0, 1)$ distribution are often denoted by Z .

In fact, we can standardize any normally distributed variable into a standard normal variable by subtracting the mean of the normal variable from each value in the dis-

Figure 2: Normal distributions with different means and variances



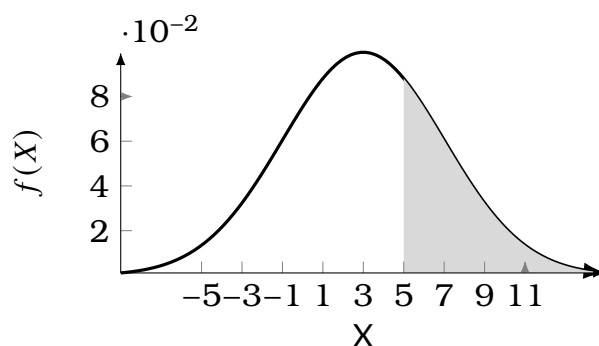
tribution and then dividing by the standard deviation of the distribution. Given $X \sim N(\mu, \sigma^2)$, the standardized random variable is given by:

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

Here, $Z \sim N(0, 1)$. To see why this is the case, note that when we standardize X , we subtract the mean μ from each value of X . This has the effect of shifting the distribution so that its center is at 0. Next, we divide each value by the standard deviation σ . This has the effect of scaling the distribution so that its spread equals 1.

2.3 Finding the area under the curve

We are often interested in finding the probability that a random variable lies in a particular interval. For example, say $X \sim N(3, 16)$, and we want to calculate $Pr(X \geq 5)$. As we said before, to find this probability, we need to calculate the following area under the curve:



However, it can be difficult and time-consuming to calculate the area by hand. Instead, we can standardize this variable and use the *standard normal table* to find this area. The standard normal table provides the area under the standard normal distribution curve for different values of Z .

Given $X \sim N(3, 16)$,

$$Z = \frac{X - 3}{4} \sim N(0, 1)$$

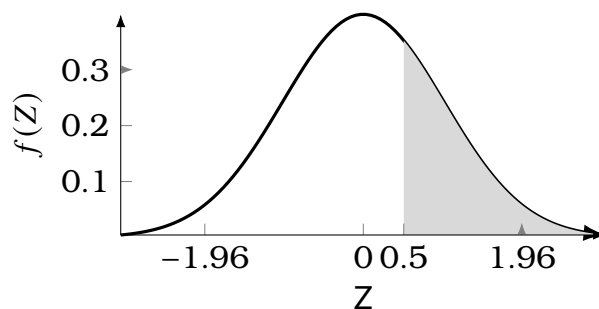
Note that,

$$Pr(X \geq 5) = Pr\left(\frac{X - 3}{4} \geq \frac{5 - 3}{4}\right) = Pr(Z \geq 0.5)$$

We can now refer to the standard normal table and find that $Pr(Z \geq 0.5)$ equals 0.3085. The figure below presents this probability as the area under the curve standard normal curve.

Given $X \sim N(\mu, \sigma^2)$, general recipe to find $Pr(x_0 < X < x_1)$:

- Find $z_0 = (x_0 - \mu)/\sigma$ and $z_1 = (x_1 - \mu)/\sigma$
- Use standard normal table to find $Pr(z_0 < Z < z_1)$



Example. Given $X \sim N(3, 16)$, we want to find $Pr(2 < X < 5)$. Here $x_0 = 2, x_1 = 5$, we can find that $z_0 = (2 - 3)/4 = -0.25$ and $z_1 = (5 - 3)/4 = 0.5$. Now we just need to look at the standard normal table and find $Pr(-0.25 < Z < 0.5)$.

Alternatively, sometimes we wish to identify the value of x for which the probability $Pr(X < x)$ or $Pr(X > x)$ equals a specific value p . This can be achieved in a similar manner by using the standard normal table. In particular, we need to find the corresponding value z that satisfies $Pr(Z < z)$ or $Pr(Z > z)$, and then transform it back to obtain x .

Given $X \sim N(\mu, \sigma^2)$ and $Pr(X < x) = p$, to find x :

- Use standard normal table to find z where $Pr(Z < z) = p$
- Find $x = \mu + z \cdot \sigma$

This follows analogously for when we are given $Pr(X > x) = p$.

3 Multiple Random Variables

3.1 Joint and Marginal Distribution

The *joint probability distribution* of two discrete random variables X and Y , is the probability that the random variables simultaneously take on certain values.

$$f(x, y) = Pr(X = x, Y = y)$$

Example. The table below gives us the probability of possible commute times and rain on a given day.

	Rain ($X = 1$)	No Rain ($X = 0$)	Total
60-min commute ($Y = 60$)	0.3	0.2	0.5
30-min commute ($Y = 30$)	0.1	0.4	0.5
Total	0.4	0.6	1

The *marginal probability distribution* of a random variable Y is just another name for its probability distribution. In particular,

$$f(y) = Pr(Y = y) = \sum_x Pr(X = x, Y = y)$$

For example, the probability of having a 60-minute commute is given by the sum of probability of having a 60-minute commute and no rain and the probability of having a 60-minute commute and rain.

$$Pr(Y = 60) = Pr(X = 1, Y = 60) + Pr(X = 0, Y = 60) = 0.3 + 0.2 = 0.5$$

3.2 Conditional Probability and Bayes Rule

The distribution of a random variable Y conditional on another random variable X taking on a specific value is called the *conditional distribution* of Y given X .

$$f(y|x) = Pr(Y = y|X = x) = \frac{Pr(X = x, Y = y)}{Pr(X = x)} = \frac{f(x, y)}{f(x)}$$

For example, the probability of having a 60-minute commute conditional on rain is given by:

$$Pr(Y = 60|X = 1) = \frac{Pr(X = 1, Y = 60)}{Pr(X = 1)} = \frac{0.3}{0.4} = \frac{3}{4}$$

While the probability of having a 60-minute commute conditional on no rain is given

by:

$$Pr(Y = 60|X = 0) = \frac{Pr(X = 0, Y = 60)}{Pr(X = 0)} = \frac{0.2}{0.6} = \frac{1}{3}$$

Remember, the unconditional probability of having a 60-minute commute was given by 0.5.

The *conditional expectation* of Y given X is the mean of the conditional distribution of Y given X .

$$E(Y|X = x) = \sum_y y Pr(Y = y|X = x)$$

For the example above,

$$E(Y|X = 1) = 60.Pr(Y = 60|X = 1) + 30.Pr(Y = 30|X = 1) = 60 \cdot \frac{3}{4} + 30 \cdot \frac{1}{4} = 52.5$$

Similarly,

$$E(Y|X = 0) = 60.Pr(Y = 60|X = 0) + 30.Pr(Y = 30|X = 0) = 60 \cdot \frac{1}{3} + 30 \cdot \frac{2}{3} = 40$$

So by comparing $E(Y|X = 1)$ and $E(Y|X = 0)$, we can find out how rain affects the average commute time.

Bayes' rule says that the conditional probability of Y given X is the conditional probability of X given Y times the relative marginal probabilities of Y and X :

$$Pr(Y = y|X = x) = \frac{Pr(X = x|Y = y)Pr(Y = y)}{Pr(X = x)}$$

This is a very useful law because it says that we can deduce conditional probabilities from the reverse conditional probability with the help of marginal probabilities.

3.3 Law of Iterated Expectations

The mean of Y is the weighted average of the conditional expectation of Y given X , weighted by the probability distribution of X .

$$E(Y) = \sum_x E(Y|X = x)Pr(X = x)$$

More compactly,

$$E(Y) = E(E(Y|X))$$

E.g. the mean height of adults is the weighted average of the mean height of men and women, weighted by their proportions.

3.4 Independence and Uncorrelatedness

Two random variables X and Y are independently distributed, or *independent*, if knowing the value of one of the variables provides no information about the other. That is,

$$Pr(Y = y|X = x) = Pr(Y = y)$$

Then by Bayes' rule:

$$Pr(X = x, Y = y) = Pr(X = x)Pr(Y = y)$$

Example: Two consecutive coin tosses.

Note: We can equivalently say that X and Y are independent if $E(Y|X) = E(Y)$.

Covariance is a measure of the extent to which two random variables move together. Let X and Y be a pair of random variables, then the *covariance* of X and Y is given by:

$$\sigma_{XY} = Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - \mu_X\mu_Y$$

The *correlation* between X and Y is given by:

$$\rho_{XY} = corr(X, Y) = \frac{Cov(X, Y)}{\sigma_X\sigma_Y} \quad \text{where } -1 \leq \rho \leq 1$$

X and Y are uncorrelated if $\rho_{XY} = \sigma_{XY} = 0$ i.e. $E(XY) = E(X)E(Y)$.

If X and Y are independent, then they are also uncorrelated.

$$E(Y|X) = E(Y) \rightarrow \rho_{XY} = 0$$

However, it is not necessarily true that if X and Y are uncorrelated, then they are also independent.

4 Linear Functions of Random Variables

4.1 Linear Functions of a Single Random Variable

If X is a random variable and $Y = a + bX$, then Y is also a random variable with

$$E(Y) = a + bE(X) \quad \text{Var}(Y) = b^2\text{Var}(X)$$

In addition, a linear transformation of a random variable does not change the shape of the distribution. So if X is normal, Y will also be normal.

Example. Let us calculate the expectation and variance for the Z score:

$$Z = \frac{X - \mu_X}{\sigma_X}$$

Note that we can write,

$$Z = -\frac{\mu_X}{\sigma_X} + \frac{1}{\sigma_X} \cdot X$$

So $a = -\frac{\mu_X}{\sigma_X}$ and $b = \frac{1}{\sigma_X}$, then by the above formulas:

$$E(Z) = a + bE(X) = -\frac{\mu_X}{\sigma_X} + \frac{1}{\sigma_X}\mu_X = 0$$

$$\text{Var}(Z) = b^2\text{Var}(X) = \frac{1}{\sigma_X^2} \cdot \sigma_X^2 = 1$$

So if $X \sim N(\mu_X, \sigma_X^2)$, then $Z \sim N(0, 1)$.

4.2 Linear Combination of Two Random Variables

X and Y is a pair of random variables, define

$$W = aX + bY$$

Then the expectation of W is given by:

$$E(W) = aE(X) + bE(Y)$$

And the variance of W is given by:

$$Var(W) = a^2Var(X) + b^2Var(Y) + 2abCov(X, Y)$$

If X and Y are independent then $Cov(X, Y) = 0$, so in that case $Var(W) = a^2Var(X) + b^2Var(Y)$.

4.3 Linear Combination of Several Random Variables

For random variables, X_1, X_2, \dots, X_n :

$$E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n)$$

$$Var(X_1 + X_2 + \dots + X_n) = \sum_{i=1}^n Var(X_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n Cov(X_i, X_j)$$

If X_1, X_2, \dots, X_n are independent random variables, then $Var(\sum_i X_i) = \sum_i Var(X_i)$.

Note that a linear combination of several normally distributed random variables will also be normal.