Chapter 2.

Random Variable:

A random variable is a variable that associates a real number with each element in the sample. It is an arbitrary choice, though it is generally made to make interpretation simpler. A random variable can take on any of the set of real numbered outcomes associated with the sample space.

Another way of putting this is that a random variable is a function mapping a probability space into the real line. More formally, associated with each point s in the domain S the function X assigns one and only one value X(s) in the range R. (The set of possible values of X(s) is usually a proper subset of the real line; i.e., not all real numbers need occur. If S is a finite set with m elements, then X(s) can assume at most m different values as s varies in S.) A function $X : S \rightarrow \mathbb{R}$, that associates a real number X(s) to each outcome s is called a random variable.

Less formally, we are mapping a construct that represents the outcomes in our experiment into a number, that we will use to represent that outcome.

Example:

Let an experiment consist of rating a product into three categories: easy to use, average to use and difficult to use. The three outcomes, or events are constructs that are not numbers. We will first map these to the real line. We can pick any combination of real numbers, but generally we will pick random numbers that help with interpretation. So, in this example we can see that we have an increasing level of difficulty associated with the three outcomes.

1) Easy
2) Average
3) Difficult

This is a straightforward mapping, where we associate 1, 2, 3 with each of our possible outcomes. But we should realize, the mapping

0) Easy
1) Average
2) Difficult

Is also perfectly acceptable and easy to interpret, such that the score is increasing with difficulty.
However, we could also map this so that Easy is rated higher than difficult and this too would be a perfectly acceptable mapping.

3) Easy
2) Average
1) Difficult

The choice of mapping is generally done in such a way as to make the interpretation of the random variables outcomes easier.

Discrete Distribution:
(Def 2.2 and 2.4 text); (Def 2.5 and 2.7 text)
If a sample space contains a finite or countable (an unending sequence with as many elements as there are whole numbers) number of possibilities it is called a discrete sample space.
Consider X is a discrete random variable. We will specify probabilities associated with the random variable with either

1) a **probability density function** f, pdf, where

   $f(x) = P(X = x)$; or

2) a **cumulative distribution function** F, cdf, where

$$F(x) = P(X \leq x) = \sum_{y \leq x} f(y)$$

We note the following:

3) If x is an impossible value for X, then $f(x) = 0$.
4) $f(x) \geq 0$.
5) $0 \leq f(x) \leq 1$
6) $\sum_x f(x) = 1$.
7) These probabilities associated with each x are called the distribution of X. That is $P(X=x) = f(x)$.

Example Random Variable:
In each batch of 100 objects we select 10 at random. Let there be more than 10 defective units in any batch of 100. Then define the random variable X, as the number of items in the sample of 10 that are defective.
We can see that X can take on the following values:
x = 0, 1, 2, … 10

Here we sample units until we find a fixed number of defective. Here we define X, the random variable as the number of units observed before we find a defective.
We can see that X can take on the following values:
x=0,1, 2,…..

An electronic device has 2 parts that work independently of each other. The probability that the first component is defective is 0.1 and the probability that the second component is defective is 0.2. Let X be the number of defective components in the device.
We can see that x can take on the following values:
X = 0, 1, 2

Probability Density Function (pdf):
Example, discrete distribution:
An engineer has produced a new car with an onboard computer navigation which is both inexpensive and easy to use. To understand the present market for different combinations of price and ease of use the marketing department tests 100 cars to find the distribution of the 6 combinations of car navigation products in the present market. These outcomes were expensive or inexpensive cars, with either easy, average, or difficult to use navigation systems.

Let the marketing departments results of the testing resulted in the following percentages (these could be shown in a tree):

Expensive + Easy	0.6
Expensive + Average	0.1
Expensive + Difficult	0.05
Inexpensive + Easy	0.05
Inexpensive + Average	0.1
Inexpensive + Difficult	0.1

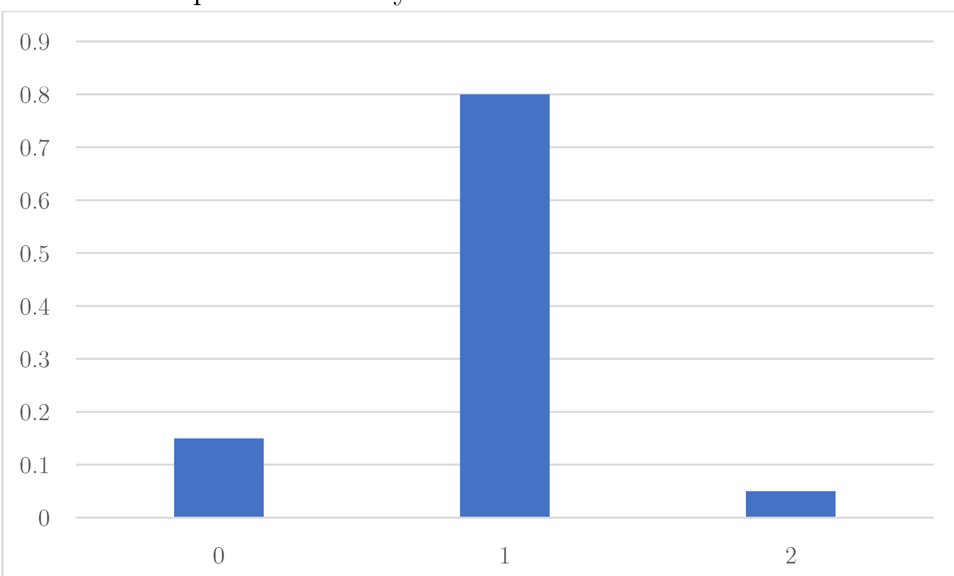
Now let X be the number of preferred options chosen by the consumers. Where preferred consisted of either inexpensive or easy to use and not preferred was all other outcomes. What is the probability density function of X?

We can see that X can take on the following values:
X = 0, 1, or 2 preferences.

Therefore, using the previous table, we can see that:

X	0	1	2
P(x) = f(x)	0.15	0.80	0.05

If we were to plot the density it would look like this:



We can now ask questions like "What is the percentage of the market that has 2 preferences, like the engineers' new product?"
Using the pdf above, we can see that P(X=2) = 0.05, so only 5% of the market has the two preferable traits.

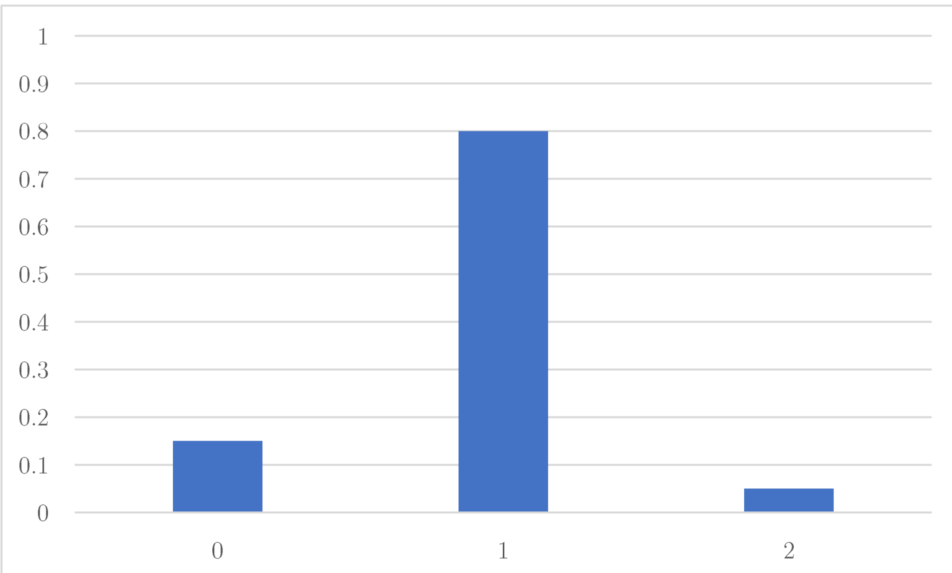
Cumulative Distribution Function (cdf):

Example:

Continuing with the above example, what is the cumulative distribution function of X?

X	0	1	2
$F(x) = P(X \leq x)$	0.15	0.95	1

If we were to plot the cdf it would look like this:



Mean

(Def 2.14)

Let X be a discrete random variable with probability mass function *f*. The expected value of X is defined as:

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

where the sum is taken over all possible values x for the variable X

E[X] is the weighted average of the possible values taken by X, where f(x) are the weights. If we were to repeat the experiment many times, then we expect the values of X approximately equal to E[X] on average. Thus, we say that the expected value of X is E[X].

This is exactly like how you calculate a mark in a course, you take the weighted average of your assignments and your tests.

Example:

Using the previous table:

X	0	1	2
P(x) = f(x)	0.15	0.80	0.05

The expected value of this random variable is:

E(X) = 0*0.15 + 1*.80 + 2*0.05 = 0.9

Variance:

(Def 2.16)

Variance is the expected value of the squared difference between X and E[X]. The variance is denoted with σ^2. It is defined as:

$\sigma^2 = Var(X) = E[(X - \mu)^2] = \sum_x (x-\mu)^2 f(x)$

where the sum is taken over all possible values x for X.

It can be shown that Var(X) an alternative (and computationally more efficient) formula is:

$\sigma^2 = E[X^2] - (E[X])^2 = \sum_x x^2 f(x) - \mu^2$

We can show this by the following:

$$\sigma^2 = E[(X-\mu)^2] = E[(X-\mu)(X-\mu)] = E[(X^2 - 2X\mu + \mu^2) = E(X) - 2\mu^2 + \mu^2$$
$$= E(X^2) - \mu^2$$

The square root of the variance is called the standard deviation:

$\sigma = \sqrt{Var(x)}$

Example:

Using the previous table and the mean 0.9:

X	0	1	2
P(x) = f(x)	0.15	0.80	0.05

$Var(X) = (0-0.9)^2 * 0.15 + (1-0.9)^2 * 0.8 + (2-0.9)^2 * 0.05 = 0.19$

Linear combinations: Y = a + bX

E(aX) = a E(X); E(aX + c) = aE(X) + c

$Var(aX) = a^2 Var(X)$; $Var(aX + c) = a^2 Var(X)$

Continuous Distributions:

(Def 2.3, 2.6 and 2.7 text)

A continuous random variable takes on an uncountably infinite number of possible values. The **probability density function** of a continuous random variable X with sample space (or support) S is an integrable function $f(x)$ satisfying the following:

(1) $f(x)$ is positive everywhere in S, that is, $f(x) > 0$, for all x in S

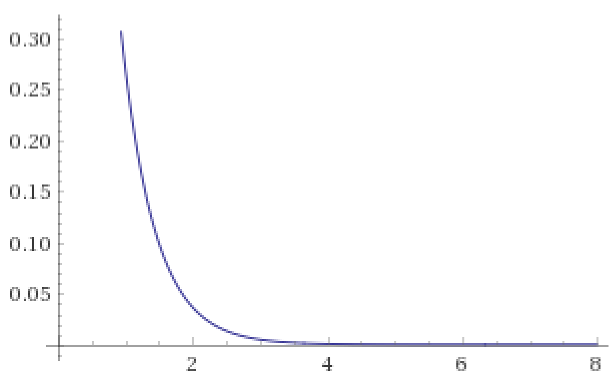
(2) The area under the curve $f(x)$ for the entire sample space S is 1, that is:
$\int_x f(x)dx=1$

(3) If $f(x)$ is the p.d.f. of x, then the probability that x belongs to A, where A is some interval ($a \leq x \leq b$), is given by the integral of $f(x)$ over that interval, that is:
$P(X\in A)=\int_A f(x)dx$ or $P(a \leq x \leq b) = \int_a^b f(x)dx$

In the continuous case, it does not matter if you use > or ≥ as the integral is the same. The probability in a continuous random variable can be thought of as the area under the curve.

Example:

$$\text{Let } f(x) = \begin{cases} 2e^{-2x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$



Find the probability that $1 \leq x \leq 3$.

$$\int_1^3 2e^{-2x}dx = e^{-2} - e^{-6} = 0.133$$

Find the probability that $x \geq 0.5$

$$\int_{0.5}^{\infty} 2e^{-2x} dx = e^{-1} = 0.368$$

Cumulative distribution function.

The cumulative distribution function F(x) of all continuous random variables X with density function *f*(x) is:

$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx$ for $-\infty \leq x \leq \infty$

Example:

Using the previous pdf *f*(x), we have

$$F(x) = \begin{cases} \int_0^x 2e^{-2t}dt = 1 - e^{-2x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Use the cdf to find $P(X \leq 1)$.

$F(1) = 1 - e^{-2} = 0.865$

Mean

(Def 2.14)

Like in the discrete case, the mean is a weighted average. Let X be a discrete random variable with probability mass function *f*(x). Then the mean is given by

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

Variance.

(Def 2.16)

$Var(X) = E([X - E(X)]^2) = E(X^2) - [E(X)]^2$

Linear combinations: $Y = a + bX$

$E(aX) = a\,E(X)$;

$E(aX + c) = aE(X) + c$

$Var(aX) = a^2\,Var(X)$;

$Var(aX + c) = a^2\,Var(X)$

Joint probability distribution function, discrete case:

(Def 2.8)

We can have cases with more than one random variable. In these cases, we need to use the joint probability distribution. The function *f*(x.y) is the joint probability distribution (or probability mass function) of the discrete random variables X and Y if:

1) $f(x,y) \geq 0$ for all (x,y)
2) $\sum_x \sum_y f(x,y) = 1$
3) $P(X = x, Y = y) = f(x,y)$
4) For any region a in the xy plane $P[(x,y) \in A] = \sum\sum_A f(x,y)$

Example:

2 scanners are needed for an experiment of the five available, 2 have electronic defects, 1 has a memory defect and 2 are in good working order. If 2 units are selected at random, find the joint probability distribution of X_1 = the number of electrical defects and X_2 = the number of memory defects in the sample.

We can see that our random variables can take on the following values:

(0,0), (0,1), (1,0), (1,1), (2,0), (2,1)

If we use a table (or a grid) we can see the sample space

		X_2	
	f(x,y)	0	1
X_1	0	0.1	0.2
	1	0.4	0.2
	2	0.1	0

We cam see that the joint distribution is calculated as

$$f(x_1, x_2) = \frac{\binom{2}{x_1}\binom{1}{x_2}\binom{2}{2-x_1-x_2}}{\binom{5}{2}}$$

Marginal distribution function

(Def 2.10)

The marginal distribution of X alone and Y alone are

$g(x) = \sum_y f(x,y)$ and $h(y) = \sum_x f(x,y)$

Discrete Example:

Using the joint distribution from the previous example we can see the row and column totals are the marginal distributions of our variables X_1 and X_2.

		X_2		Total
	f(x,y)	0	1	Marginal X_1
X_1	0	0.1	0.2	0.3
	1	0.4	0.2	0.6
	2	0.1	0	0.1
Total	Marginal X_2	0.6	0.4	1

Joint probability distribution function, continuous case:

(Def 2.9)

The function $f(x,y)$ is a joint probability density function of the continuous random variables X and Y if

1) $f(x,y) \geq 0$ for all (x,y)
2) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$
3) For any region a in the xy plane $\mathrm{P}[(\mathrm{x},\mathrm{y}) \in \mathrm{A}] = \int\int f(x,y)\,dx\,dy$

Example:

Marginal distribution function

(Def 2.10)

The marginal distribution of X alone and Y alone are

$g(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$ and $h(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$

for the continuous case.

Example:

Let X and Y have joint probability density function:

$f(x,y)=4xy$ for $0 < x < 1$ and $0 < y < 1$.

What is $f_X(x)$, the marginal pdf of X, and $f_Y(y)$, the marginal p.d.f. of Y?

In order to calculate the marginal pdf of X, we need to integrate the joint pdf $f(x,y)$ over $0 < y < 1$. We get:

$$g(x) = \int_0^1 4xy\,dy = 4x\left[\frac{y^2}{2}\right]_{y=0}^{y=1} = 2x, \qquad 0<x<1$$

In order to calculate the marginal pdf. of Y, we need to integrate the joint pdf $f(x,y)$ over $0 < x < 1$. We get:

$$h(y) = \int_0^1 4xy\,dx = 4y\left[\frac{x^2}{2}\right]_{x=0}^{x=1} = 2y\,, \qquad 0<y<1$$

Conditional distribution function:

(Def 2.11)

Let X and Y be 2 random variables discrete or continuous. The conditional distribution of the random variable Y given X is

$$f(y|x) = \frac{f(x,y)}{g(x)} \text{ provided } g(x) > 0.$$

And the conditional distribution of X given Y is

$$f(x|y) = \frac{f(x,y)}{h(y)} \text{ provided } h(y) > 0.$$

Example:

Using the previous joint pdf and margins, we can see that the conditional distribution of X given Y is

$$f(y|x) = \frac{f(x,y)}{g(x)} = \frac{4xy}{2x} = 2y$$ and the conditional distribution of Y given X is

$$f(x|y) = \frac{f(x,y)}{h(y)} = \frac{4xy}{2y} = 2x.$$

Independence

(Def 2.12)

If X and Y are independent, then

$E(XY) = E(X)\,E(Y)$

Covariance

(Def 2.17)

The covariance between the random variables X and Y is

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

It is a measure of how much two variables vary together.

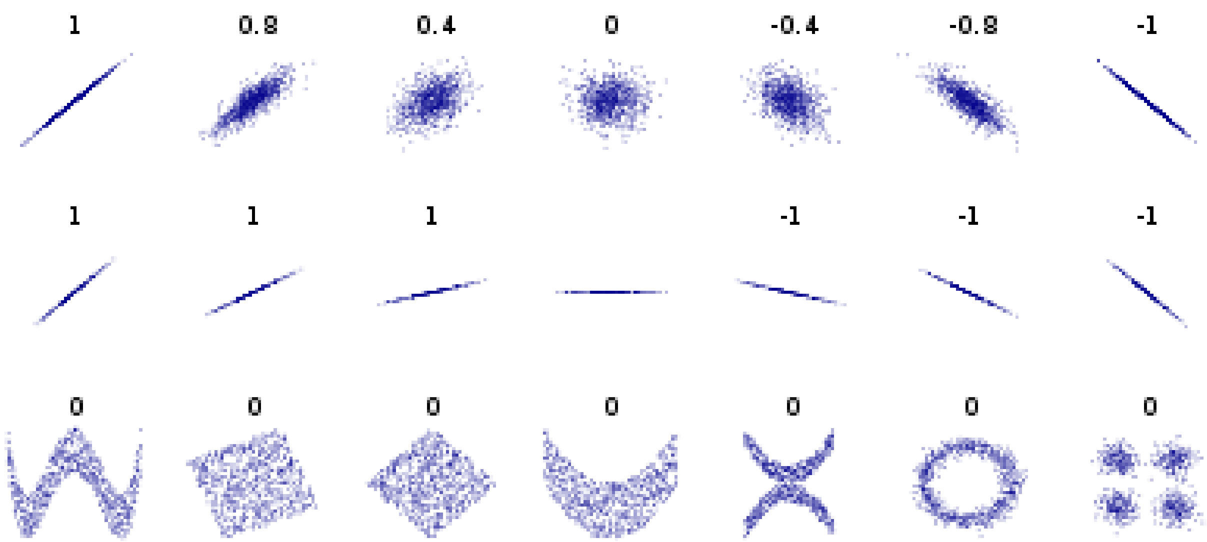
Correlation

(Def 2.18)

The correlation between X and Y is a measure of **linear** association. The correlation coefficient between X and Y is $\rho_{XY} = \sigma_{XY} / \sigma_X \sigma_Y$. Correlation has the following property:

1) $-1 \leq \rho_{xy} \leq 1$

Some visualizations of different correlations. You can see that 0 correlation, does not just mean no relationship.



Source: https://en.wikipedia.org/wiki/Correlation_and_dependence

Correlation will be discussed in more detail with regression.