EC203: Applied Econometrics
A review of random variables

Dr. Tom Martin

University of Warwick

2019

Illustrative reading:

- Thomas: chapters 1 and 2.
- First year materials: $EC122/124$

Random variables

A random variable (RV) is any variable whose outcome has uncertainty associated with it. RVs do not have a single value associated to them, instead they can take a range of values. We will use RVs in econometrics for three main purposes:

1. Descriptive statistics: describing variables of interest, such as, income, treatments, health outcomes, unemployment rates, ... etc.
2. Bias and efficiency: evaluating the properties of estimators using sampling distributions.
3. Population inference: using estimators, samples and associated sampling distributions to make statistical statements about the population of interest.

Random Variables

There are two types of random variables:

1. discrete:
   - the outcomes are countable.
   - each outcome has an associated probability of occurring.
2. continuous:
   - the outcomes are continuous or, at least, are not feasibly countable.
   - all particular outcomes have a zero probability of occurring.
   - probabilities are associated with ranges of outcomes.

Notation: we denote random variables in upper case X, Y, Z and specific realisations in lower case x, y, z.

Useful definitions

Definition 1: The **sample space**, S, of an experiment is the set of all possible outcomes of that experiment. This is typically written $S = \{s_1, s_2, ..., s_k\}$, where s_j, $i = 1, ..., k$ represent the k possible outcomes.

Definition 2: An **event**, A, is any set of possible outcomes to an experiment. For example, $A = \{s_1\}$, $A = \{s_1, s_k\}$, or $A = \{s_i, s_j\}$ where $i \neq j$, ... etc.

Useful probability results

Given a sample space S then the following are **useful probability axioms**:

- The probaility of any event (A) occuring lies between zero and one, inclusive: $P(A) \in [0, 1]$.
- If events A_i, $i = 1, 2, ..., n$ are mututally exclusive, such that $A_i \cap A_j = \varnothing$ for all $i \neq j$, then $P(A_1 \cup A_2 \cup ... \cup A_n) = P(A_1) + P(A_2) + ... + P(A_n)$.
- If events A_i, $i = 1, 2, ..., n$ are statistically independent then $P(A_1 \cap A_2 ... \cap A_n) = P(A_1)P(A_2)P(A_3)...P(A_n)$.

Discrete random variables: a single coin toss

As an example of a discrete random variable take a single coin toss:

- The **sample space** for a single coin toss is $S = \{s_1, s_2\} = \{T, H\}$
- Then a **random variable X is a rule**, which assigns each outcome in the sample space to a real number, $X(S) = x$.
- Here let X(S) represent the number of heads on a coin toss, that is, $X(T) = 0$ and $X(H) = 1$.
- Here $x = 0$ and $x = 1$ are the **specific outcomes**.

Discrete random variables: a single coin toss

The **probability distribution function (pdf)** associated with X, written $f_X(x)$, is:

- $P(X = 0) = 0.5$, the probability that $X = 0$ is 50%.
- $P(X = 1) = 0.5$, the probability that $X = 1$ is 50%.

The **cumulative distribution function (cdf)** associated with X, written $F_X(x)$, is:

- $P(X \leq 0) = P(X = 0) = 0.5$, the probability X is less than or equal to 0.
- $P(X \leq 1) = P(X = 0) + P(X = 1) = 1$, the probability X is less than or equal to 1.

Discrete random variables: three coin tosses

Now take the example of three coin tosses:

- The **sample space** for three coin tosses is
- $S = \{s_1, s_2, ..., s_8\}$
- $S = \{TTT, TTH, THT, THH, HTT, HTH, HHT, HHH\}$

Then let the random variable $X(S) =$ **the number of heads**, such that,

- $X(TTT) = 0$
- $X(HTT) = X(THT) = X(TTH) = 1$
- $X(HHT) = X(HTH) = X(THH) = 2$
- $X(HHH) = 3$

Discrete random variables: three coin toss

- The pdf associated with X is:
  - $P(X=0) = P(TTT) = P(T)P(T)P(T) = (0.5)(0.5)(0.5) = 0.125$
  - $P(X=1) = P(HTT) + P(THT) + P(TTH) = 0.125 + 0.125 + 0.125 = 0.375$
  - $P(X=2) = P(HHT) + P(HTH) + P(THH) = 0.125 + 0.125 + 0.125 = 0.375$
  - $P(X=3) = P(HHH) = 0.125$
- The pdf $f_X(x)$ representing X is,

x_j	0	1	2	3
$f_X(x_j)$	0.125	0.375	0.375	0.125

- That is, $f_X(x)$ tracks the probability that exactly x_j heads are attained in 3 coin tosses.

Discrete random variables: three coin toss

- the cdf associated with X is:
  - $P(X \leq 0) = P(X = 0) = 0.125$
  - $P(X \leq 1) = P(X = 0) + P(X = 1) = 0.5$
  - $P(X \leq 2) = P(X = 0) + P(X = 0) + P(X = 0) = 0.875$
  - $P(X \leq 3) = P(X = 0) + P(X = 0) + P(X = 0) + P(X = 0) = 1$
- The cdf $F_X(x)$ is summarised in the table below,

x_j	0	1	2	3
$F_X(x_j)$	0.125	0.5	0.875	1

- That is, $F_X(x_j)$ tracks the probability that up to or less than x_j heads are attained in 3 coin tosses.

Discrete random variables: the general case

Let X represent a random variable that can take k possible values $\{x_1, x_2, x_3, ..., x_k\}$ each with an associated probability $p_1, p_2, p_3, ..., p_k$, where,

1. The probability that X takes the value x_j is written $p_j = P(X = x_j)$.
2. It is the case that $p_j \in [0, 1]$.
3. Further, for $f_X(x)$ to represent a true pdf it must be the case that $\sum_{j=1}^{k} p_j = 1$.

Discrete random variables: the general case

The pdf $f_X(x)$ summarises the possible outcomes of X and the associated probabilities, such that,

1. $f_X(x_j) = p_j = P(X = x_j)$
2. $f_X(x_j) \geq 0$ and $f(x_j) \leq 1$
3. $\sum_{j=1}^{k} f(x_j) = 1$

The cdf $F_X(x)$ is such that,

1. $F_X(x_l) = P(X \leq x_l),$ where $l \in (1, 2, 3, ..., k)$
2. $F_X(x_l) = \sum_{j=1}^{l} p_j = \sum_{j=1}^{l} f_X(x_j)$
3. $F_X(x_k) = P(X \leq x_k) = \sum_{j=1}^{k} f_X(x_j) = 1$

Random variables: expected values and dispersion

Given a random variable, X, in this course we will we will often work with:

1. the mean, or expected value, of X: $E[X]$.
2. the variance of X: $V(X)$.
3. the standard deviation of X: $sd(X)$.

Random variables: the expected value of X

Suppose X takes on a finite number of values, $x = x_1, ...x_k$ and each has an associated probability $p_1, ..., p_k$. The expected value of X is,

$$E[X] = \sum_{j=1}^{k} p_j(x)x_j = p_1x_1 + p_2x_2 + ... + p_kx_k$$

For example, in a population of size N, where each individual appears once and has an associated probability of occuring equal to the frequency of occurring,

$$E[X] = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Random variables: the expected value of X

Some important properties include:

1. For any constant c: $E[c] = c$.
2. For any constants any two constants a and b and random variable X: $E[aX + b] = aE[X] + b$.
3. If $\{a_1, a_2, ..., a_k\}$ are constants and $\{X_1, X_2, ..., X_k\}$ are random variables then $\sum_{i=1}^{k} E[a_i X_i] = \sum_{i=1}^{k} a_i E[X_i]$.
4. In general $E[g(X)] = \sum_X p_X(x)g(x)$. For instance higher order moments can be calculated. The n^{th} moment is $E[X^n] = \sum_{j=1}^{k} p_j(x)x_j^n$
5. Note, it is normally the case that $E[g(X)] \neq g(E[X])$. For instance, typically $E[X^2] \neq E[X]^2$.

Discrete random variables: the variance of X

Suppose X takes on a finite number of values, $x = x_1, ...x_k$ and each has an associated probability $p_1, ..., p_k$. The variance of X is,

$$\begin{aligned} V[X] =& E[(X - E[X])^2] \\ =& \sum_{j=1}^{k} p_j(x)(x_j - E[X])^2 \\ =& p_1(x_1 - E[X])^2 + p_2(x_2 - E[X])^2 + ... \\ & ... + p_k(x_k - E[X])^2 \end{aligned}$$

Or, alternatively,

$$V[X] = E[X^2] - E[X]^2$$

The standard deviation is the positive square root of the variance: $sd(X) = +\sqrt{V(X)}$

Discrete random variables: the variance of X

Some important properties of the variance include:

1. For any constant c: $V(c) = 0$
2. For any constants a and b and random variable X, $V(aX + b) = a^2V(X)$
3. For any constants a and b and random variables X and Y, $V(aX + bY) = a^2V(X) + b^2V(Y) + 2abCov(X, Y)$
4. Let $\{a_1, a_2, ..., a_k\}$ be constants and $\{X_1, X_2, ..., X_k\}$ be random variables, then, $V(\sum_{i=1}^{k} a_iX_i) = \sum_{i=1}^{k} a_i^2V(X_i) + \sum_{i>j} 2a_ia_jCov(X_i, X_j)$
5. Let $\{a_1, a_2, ..., a_k\}$ be constants and $\{X_1, X_2, ..., X_k\}$ be pairwise uncorrelated (independent), then, $V(\sum_{i=1}^{k} a_iX_i) = \sum_{i=1}^{k} a_i^2V(X_i)$

Discrete random variables: the variance of X

Some important properties of the standard deviation:

1. For any constant c: $sd(c) = 0$
2. For any constants a and b and random variable X, $sd(aX + b) = |a|sd(X)$
3. Let $\{a_1, a_2, ..., a_k\}$ be constants and $\{X_1, X_2, ..., X_k\}$ be random variables , then, $sd(\sum_{i=1}^{k} a_i X_i) = \sqrt{\sum_{i=1}^{k} a_i^2 V(X_i) + \sum_{i>j} 2a_i a_j Cov(X_i, X_j)}$
4. Let $\{a_1, a_2, ..., a_k\}$ be constants and $\{X_1, X_2, ..., X_k\}$ be pairwise uncorrelated (independent), then, $sd(\sum_{i=1}^{k} a_i X_i) = \sqrt{\sum_{i=1} a_i^2 V(X_i)}$

Bivariate distributions

Now we consider bivariate probability density functions. That is, probability distributions that represent the likelihood of two random variables taking specific values.

Definition 1: We now have two **sample spaces** $S_X = \{x_1, x_2, ..., x_k\}$ and $S_Y = \{y_1, y_2, ..., y_l\}$, one for each random variable X and Y respectively.

Definition 2: An **event**, is now defined in terms of X and Y's sample space. For example, the event $x_j \cap y_l$ is the event that both x_j and y_l occur.

Discrete bivariate distributions

A joint discrete *pdf* is represented in the table below,

		Y				
		y_1	y_2	$\dots$	y_l	
X	x_1	$P(x_1 \cap y_1)$	$P(x_1 \cap y_2)$	$\dots$	$P(x_1 \cap y_l)$	
	x_2	$P(x_2 \cap y_1)$	$P(x_2 \cap y_2)$	$\dots$	$P(x_2 \cap y_l)$	
	$\vdots$	$\vdots$			$\vdots$	
	x_k	$P(x_k \cap y_1)$	$P(x_k \cap y_2)$	$\dots$	$P(x_k \cap y_l)$	

Discrete bivariate distributions

Where:

- The joint pdf is determined by $P(X = x_i, Y = y_i) = p_{ij}(x, y)$ for $i = 1, .., k$ and $j = 1, ..., l$.
- The joint pdf defines a probability for each of the possible combinations (x, y).
- The joint pdf is also commonly written as $f_{X,Y}(X = x, Y = y)$ or just $f(x, y)$.

As with univariate pdfs, to be a valid bivariate pdf we require:

- $P(X = x_i, Y = y_j) \in [0, 1]$
- $\sum_{i=1}^{k} \sum_{j=1}^{l} P(X = x_i, Y = y_j) = 1$

Discrete bivariate distributions: marginal distribution

Moving from the joint distribution to the marginal distribution. Suppose we are only interested in the random variable X. Then it is the case that,

$$
\begin{aligned}
P(X = x_i, y_1 \leq Y \leq y_l) =& \sum_{j=1}^{l} p_{X,Y}(x_i, y_j) \\
=& P(x_i \cap y_1) + P(x_i \cap y_2) + ... + P(x_i \cap y_l)
\end{aligned}
$$

This gives us the marginal distribution for X, i.e. $f_X(x)$, which summarises the possible outcomes of X and the associated probabilities.

Discrete bivariate distributions: marginal distribution

Moving from the joint distribution to the marginal distribution. Suppose we are only interested in the random variable Y. Then it is the case that,

$$
\begin{aligned}
P(Y = y_j, x_1 \leq X \leq x_k) =& \sum_{i=1}^{k} p_{X,Y}(x_i, y_j) \\
=& P(y_j \cap x_1) + P(y_j \cap x_2) + ... + P(y_j \cap x_k)
\end{aligned}
$$

This gives us the marginal distribution for X, i.e. $f_Y(y)$, which summarises the possible outcomes of Y and the associated probabilities.

Discrete bivariate distributions: marginal distributions

The marginal distributions can be usefully tabulated along the sides of the joint pdf,

		Y				
		y_1	y_2	$\ldots$	y_l	$f_X(x)$
X	x_1	$P(x_1 \cap y_1)$	$P(x_1 \cap y_2)$	$\ldots$	$P(x_1 \cap y_l)$	$P(x_1)$
	x_2	$P(x_2 \cap y_1)$	$P(x_2 \cap y_2)$	$\ldots$	$P(x_2 \cap y_l)$	$P(x_2)$
	$\vdots$	$\vdots$			$\vdots$	$\vdots$
	x_k	$P(x_k \cap y_1)$	$P(x_k \cap y_2)$	$\ldots$	$P(x_k \cap y_l)$	$P(x_k)$
	$f_Y(y)$	$P(y_1)$	$P(y_2)$	$\ldots$	$P(y_l)$	1

Discrete bivariate distributions: conditional distributions

We can also calculate conditional probabilities and associated distributions. Given two random variables X and Y, a joint pdf, $P(x,y)$, and the marginal distributions, $P(x)$ and $P(y)$, the conditional pdf for the random variable X conditional on Y is:

$$P(x|y) = \frac{P(x,y)}{P(y)} \text{ also written } f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Or, rearranging the expression,

$$P(x,y) = P(x|y)P(y) \text{ also written } f_{X,Y}(x,y) = f_{X|Y}(x|y)f_Y(y)$$

Example: wages conditional on height

Consider the following table which represents the joint pdf for height (H) against wages (W).

			Height		
			1	2	3
			below	average	above
	1	low	0.253	0.092	0.032
Wage	2	medium	0.123	0.132	0.106
	3	high	0.017	0.104	0.141

Example: wages conditional on height

The marginal probability distibutions $p(h)$ and $p(w)$ are,

			Height			
			1	2	3	p(w)
			below	average	above	
	1	low	0.253	0.092	0.032	0.377
Wage	2	medium	0.123	0.132	0.106	0.361
	3	high	0.017	0.104	0.141	0.263
		p(h)	0.393	0.328	0.279	

Example: wages conditional on height

The conditional probability distributions $p(w|h)$ are,

			Height		
			1	2	3
			below	average	above
	1	low	0.643	0.280	0.115
Wage	2	medium	0.313	0.402	0.379
	3	high	0.044	0.317	0.506
			1.000	1.000	1.000
			$p(w\|h=1)$	$p(w\|h=2)$	$p(w\|h=3)$

Example: wages conditional on height

The conditional probability probability distributions $p(h|w)$,

			Height				
			1	2	3		
			below	average	above		
	1	low	0.671	0.244	0.085	1	$p(h\|w=1)$
Wage	2	medium	0.341	0.366	0.293	1	$p(h\|w=2)$
	3	high	0.066	0.397	0.538	1	$p(h\|w=3)$

Discrete bivariate distributions: expected values

From the discrete bivariate and marginal distributions we can calculate moments of interest. For example:

1. Expected values: $E[X] = \sum_{i=1}^{k} p(X = x_i)x_i$ and $E[Y] = \sum_{j=1}^{l} p(Y = y_j)y_j$
2. Variances: $V(X) = E[X^2] - E[X]^2$ and $V(Y) = E[Y^2] - E[Y]^2$
3. Higher order moments: $E[X^n] = \sum_{i=1}^{k} p(X = x_i)x_i^n$ and $E[Y^n] = \sum_{j=1}^{l} p(Y = y_j)y_j^n$

Discrete bivariate distributions: expected values

We can also calculate measure of dependence. For example, the covariance between X and Y:

$$
\begin{aligned}
cov(X,Y) &= E\{[X - E(X)][Y - E(Y)]\} \\
&= \sum_{i=1}^{k} \sum_{j=1}^{l} p_{i,j}(x,y)(x_i - E[X])(y_j - E[Y]) \\
&= E[XY] - E[X]E[Y] \\
&= \sum_{i=1}^{k} \sum_{j=1}^{l} x_i y_j p_{i,j}(x,y) - \sum_{i=1}^{k} x_i p_i(x) \sum_{j=1}^{l} y_j p_j(y)
\end{aligned}
$$

Covariance

Covariance is an extremely useful measure. However, for our purposes there are two main problems with covariance as a measure of the association between two random variables:

1. It is not unit free. Since for any constants a_1, a_2, b_1 and b_2, $Cov(a_1X + b_1, a_2Y + b_2) = a_1a_2Cov(X, Y)$.
2. It is only a measure of linear association.

Correlation

Correlation addresses the first problem:

$$Corr(X,Y) = \rho_{XY} = \frac{Cov(X,Y)}{sd(X)sd(Y)} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

1. Given $\sigma_X, \sigma_Y > 0$ means the $Corr(X,Y)$ and $Cov(X,Y)$ have the same sign.
2. Correlation is easier to interpret since, $-1 \leq Corr(X,Y) \leq 1$
3. For any constants a_1, a_2, b_1 and b_2, with $a_1 a_2 > 0$: $Corr(a_1 X + b_1, a_2 X + b_2) = Corr(X,Y)$
4. For any constants a_1, a_2, b_1 and b_2, with $a_1 a_2 < 0$: $Corr(a_1 X + b_1, a_2 X + b_2) = -Corr(X,Y)$
5. Properties 3 and 4 show explicitly that correlation is unit free.

Conditional expectations

The conditional expectations operator addresses the second problem. It can be represented as,

$$E[Y|X = x] = \sum_{j=1}^{l} p(y_j|x)y_j$$

- This conditional expectation operator calculates the expected value of Y at certain levels of X.
- It is possible to model non-linear associations.

Conditional expectations

Some important properties of conditional expectations are:

1. For any function $c(X)$, $E[c(X)|X] = c(X)$. For example $E[X^2|X] = X^2$, intuitively if we know X we know $c(X)$.
2. For two functions $a(X), b(X)$, and two random variables X and Y, $E[a(X)Y + b(X)|X] = a(X)E[Y|X] + b(X)$
3. If two random variables are mean independent then $E[Y|X] = E[Y]$. That is, the expected (mean) value of Y does not depend on the value the X takes.

Height and wage example continued

Reconsider the height and wages associations

			Height			
			1	2	3	p(w)
			below	average	above	
	1	low	0.253	0.092	0.032	0.377
Wage	2	medium	0.123	0.132	0.106	0.361
	3	high	0.017	0.104	0.141	0.263
		p(h)	0.393	0.328	0.279	

Height and wage example continued

Expected values:

- $E[H] = 0.393 * 1 + 0.328 * 2 + 0.279 * 3 = 1.886$
- $E[W] = 0.377 * 1 + 0.361 * 2 + 0.262 * 3 = 1.886$

Variances:

- $E[H^2] = 0.393 * (1^2) + 0.328 * (2^2) + 0.279 * (3^2) = 4.2159$
- $E[W^2] = 0.377 * (1^2) + 0.361 * (2^2) + 0.262 * (3^2) = 4.1831$
- $V(H) = E[H^2] - E[H]^2 = 0.6585$
- $V(W) = E[W^2] - E[W]^2 = 0.6257$

Height and wage example continued

Covariance:

- $cov(H,W) = E[HW] - E[H]E[W]$
- $E[HW] = 0.253(1)(1) + 0.092(1)(2) + 0.032(1)(2) + ... + 0.104(3)(2) + 0.141(3)(3) = 3.886$
- $cov(H,W) = 3.886 - (1.886)(1.886) = 0.331$

Conditional expectations:

- $E[W|H=1] = \sum w_i p(w|h=1) = 1(0.643) + 2(0.313) + 3(0.044) = 1.400$
- $E[W|H=2] = \sum w_i p(w|h=2) = 1(0.280) + 2(0.402) + 3(0.317) = 2.037$
- $E[W|H=3] = \sum w_i p(w|h=3) = 1(0.115) + 2(0.379) + 3(0.506) = 2.391$
- From this we can see that expected wages increases with height.

Random variables

There are two types of random variables:

1. discrete:
   - the outcomes are countable.
   - each outcome has an associated probability of occurring.
2. continuous:
   - the outcomes are continuous or at least are not feasibly countable.
   - all particular outcomes have a zero probability of occurring.
   - probabilities are associated with ranges of outcomes.

Notation: we denote random variables in upper case X, Y, Z and specific realisations in lower case x, y, z.

Continuous random variables

Let the random variable X take any value between $[-\infty, \infty]$ but specific values have a zero probability of occurring.[1] Then the pdf for X, denoted $f_X(x)$, can be used to track the probability that $X \in [a, b]$,

$$P[a \leq X \leq b] = \int_a^b f_X(x)dx \;\forall\; a \leq b$$

In words, the probability that X lies between a and b is equal to the area under the *pdf* from a to b. Further, a true *pdf* is such that,

$$P[-\infty \leq X \leq \infty] = \int_{-\infty}^{\infty} f_X(x)dx = 1$$

That is, the entire area beneath the *pdf* must be equal to 1.

---

[1] That is, $P[X = a] = \int_a^a f_X(x)dx = 0$.

Continuous random variables

The cumulative density function (*cdf*), $F_X(x)$, summarises the probability that $X \in [-\infty, x]$, which is equivalent to $P[X \leq x]$, where x is some level of X. Therefore, the *cdf* is the area under the *pdf* up to x,

$$P[-\infty \leq X \leq x] = F(x) = \int_{-\infty}^{x} f(t)dt \ \forall \ -\infty \leq x$$

Further, given the total area under the *pdf* is equal to 1, it is also the case that,

$$P[-\infty \leq X \leq \infty] = F(x) = \int_{-\infty}^{\infty} f(x)dx = 1$$

Useful properties

The following properties are useful. When X is continuous:

1. For any constant c $P(X = c) = 0$. This implies that
2. $P(X \geq c) = P(X > c)$
3. $P(a < X < b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b)$

Further, when working with *cdf s*:

1. $F(x) \in [0, 1]$
2. If $x_1 \leq x_2$ then $P(X \leq x_1) \leq P(X \leq x_2) \implies F(x_1) \leq F(x_2)$
3. For any number c: $P(X > c) = 1 - F(c)$
4. For any number $a < b$: $P(a < X \leq b) = F(b) - F(a)$

Random variables: expected values and dispersion

There are also continuous representations for each of the following,

1. the mean, or expected value, of X.
2. the variance of X.
3. the standard deviation of X.

Random variables: the expected value of X

For a continuous random variable, X, the expected value is:

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

Further, the standard rules, covered for the discrete case, also hold for the continuous case, such as, $E[aX] = aE[X]$. Further, we can calculate higher moments as,

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

Continuous random variables: the variance of X

For continuous random variables the variance of X is,

$$\begin{aligned} V[X] &= E[(X - E[X])^2] \\ &= \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x)dx \\ &= E[X^2] - E[X]^2 \\ &= \int_{-\infty}^{\infty} x^2 f_X(x)dx - E[X]^2 \end{aligned}$$

The standard deviation is the positive square root of the variance. Further, the standard rules, covered for the discrete case, also hold for the continuous case.

Continuous bivariate distributions: joint distributions

There are also the equivalent continuous bivariate distributions. Suppose we have two random variables, X and Y, then their joint pdf $f_{X,Y}(x, y)$ has the following properties:

1. All probabilities lie between zero and one, inclusive: $f_{X,Y}(x, y) \in [0, 1]$
2. The total area under the joint pdf is equal to one: $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) \partial x \partial y = 1$

Continuous bivariate distribution: marginal distributions

Suppose we have two continuous random variables, say X and Y, and their joint pdf $f_{X,Y}(x, y)$. To calculate the marginal distribution of X, $f_X(x)$, integrate over the joint distribution with respect to Y. (Think of this as being equivalent to, for each value of X, summing over the discrete values of Y in the discrete case.)

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy$$

To calculate the marginal distribution of Y, $f_Y(y)$, integrate over the joint distribution with respect to X.

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dx$$

Continuous bivariate distributions: expected values

From the continuous bivariate and marginal distributions we can calculate the usual moments of interest:

1. Expected values: $E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$ and $E[y] = \int_{-\infty}^{\infty} y f_Y(y) dy$.
2. Variances:
   $V(X) = E[X^2] - E[X]^2 = \int_{-\infty}^{\infty} x^2 f_X(x) dx - E[X]^2$ and
   $V(Y) = E[Y^2] - E[Y]^2 = \int_{-\infty}^{\infty} y^2 f_Y(y) dy - E[Y]^2$

Continuous bivariate distributions: expected values

Further, we can calculate measures of covariance

$$
\begin{aligned}
cov(X,Y) =& E\{[X - E(X)](Y - E(Y)]\} \\
=& \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - E[X)(y - E[Y])f_{X,Y}(x,y)\partial x \partial y \\
=& E[XY] - E[X]E[Y] \\
=& \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y)\partial x \partial y \\
&- \int_{-\infty}^{\infty} x f_X(x)dx \int_{-\infty}^{\infty} y f_Y(y)dy
\end{aligned}
$$

Conditional distributions

Given a joint pdf, $f_{X,Y}(x, y)$, and the marginal distributions, $f_X(x)$ and $f_Y(y)$, the conditional pdf for the random variables X and Y is:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

Or, rearranging the expression,

$$f_{X,Y}(x, y) = f_{X|Y}(x|y)f_Y(y)$$

Conditional expectations

As usual, we will be more concerned with expected values. The conditional expectation operator is represented as,

$$E[Y|X = x] = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy$$

This tells us the expected value of Y, given certain levels of X.

Note: we will not explicilty use integration on this course. However, it is important you understand the basic idea to, for example, understand the calculation of the area under the normal distribution, which we consider next.