4.1 Joint Distributions of Two Discrete Random Variables

Suppose the discrete random variables X and Y have supports S_X and S_Y, respectively.

The joint distribution of X and Y is given by the set $\{P(X = x, Y = y)\}_{x \in S_X, y \in S_Y}$.

This joint distribution satisfies

$$\begin{aligned} P(X = x, Y = y) &\geq 0 \\ \sum_{x \in S_X} \sum_{y \in S_Y} P(X = x, Y = y) &= 1 \end{aligned}$$

Joint Distributions of Two Discrete Random Variables

When S_X and S_Y are small sets, the joint distribution of X and Y is usually presented in a table, e.g.

	$Y = -1$	$Y = 0$	$Y = 1$
$X = 0$	0	$\frac{1}{3}$	0
$X = 1$	$\frac{1}{3}$	0	$\frac{1}{3}$

Note that by summing along a row (i.e. summing over all the possible values of Y for a particular realisation of X) we obtain the probability that X takes the appropriate value, i.e.

$$P(X = x) = \sum_{y \in S_Y} P(X = x, Y = y).$$

Similarly, by summing along a column (i.e. summing over all the possible values of X for a particular realisation of Y) we obtain the probability that Y takes the appropriate value, i.e.

$$P(Y = y) = \sum_{x \in S_X} P(X = x, Y = y).$$

Marginal Distributions of Discrete Random Variables

The distributions of X and Y obtained in this way are called the marginal distributions of X and Y, respectively.

These marginal distributions satisfy the standard conditions for a discrete distribution, i.e.

$$P(X = x) \geq 0; \qquad \sum_{x \in S_X} P(X = x) = 1.$$

4.1.2 Independence of Two Discrete Random Variables

Two discrete random variables X and Y are independent if and only if

$$P(X = x, Y = y) = P(X = x)P(Y = y), \quad \forall x \in S_X, y \in S_Y.$$

4.1.3 The Expected Value of a Function of Two Discrete Random Variables

The expected value of a function, $g(X, Y)$, of two discrete random variables is defined as

$$E[g(X, Y)] = \sum_{x \in S_X} \sum_{y \in S_Y} g(x, y) P(X = x, Y = y).$$

In particular, the expected value of X is given by

$$E[X] = \sum_{x \in S_X} \sum_{y \in S_Y} x P(X = x, Y = y).$$

It should be noted that if we have already calculated the marginal distribution of X, then it is simpler to calculate $E[X]$ using this.

4.1.4 Covariance and the Correlation Coefficient

The covariance between X and Y, $Cov(X, Y)$ is given by

$$Cov(X, Y) = E(XY) - E(X)E(Y)$$

Note that by definition $Cov(X, X) = E(X^2) - E(X)^2 = Var(X)$.
Also, for any constants a, b

$$Cov(X, aY + b) = aCov(X, Y).$$

The coefficient of correlation between X and Y is given by

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}.$$

Properties of the Correlation Coefficient

1. $-1 \leq \rho(X, Y) \leq 1$.
2. If $|\rho(X, Y)| = 1$, then there is a deterministic linear relationship between X and Y, $Y = aX + b$. When $\rho(X, Y) = 1$, then $a > 0$. When $\rho(X, Y) = -1$, $a < 0$.
3. If random variables X and Y are independent, then $\rho(X, Y) = 0$. Note that the condition $\rho(X, Y) = 0$ is not sufficient for X and Y to be independent (see Example 4.1).
4. If $a > 0$, $\rho(X, aY + b) = \rho(X, Y)$, i.e. the correlation coefficient is independent of the units a variable is measured in. If $a < 0$, $\rho(X, aY + b) = -\rho(X, Y)$.

Example 4.1

Discrete random variables X and Y have the following joint distribution:

	$Y = -1$	$Y = 0$	$Y = 1$
$X = 0$	0	$\frac{1}{3}$	0
$X = 1$	$\frac{1}{3}$	0	$\frac{1}{3}$

a) Calculate i) $P(X > Y)$, ii) the marginal distributions of X and Y, iii) the expected values and variances of X and Y, iv) the coefficient of correlation between X and Y.

b) Are X and Y independent?

Example 4.1

To calculate the probability of such an event, we sum over all the cells which correspond to that event. Hence,

$$P(X > Y) = P(X = 0, Y = -1) + P(X = 1, Y = -1) + \\ + P(X = 1, Y = 0) = \frac{1}{3}$$

Example 4.1

We have

$$P(X = x) = \sum_{y \in S_Y} P(X = x, Y = y).$$

Hence,

$$P(X = 0) = \sum_{y=-1}^{1} P(X = 0, Y = y) = 0 + \frac{1}{3} + 0 = \frac{1}{3}$$

$$P(X = 1) = \sum_{y=-1}^{1} P(X = 1, Y = y) = \frac{1}{3} + 0 + \frac{1}{3} = \frac{2}{3}.$$

Note that after calculating $P(X = 0)$, we could calculate $P(X = 1)$ using the fact that $P(X = 1) = 1 - P(X = 0)$, since X only takes the values 0 and 1.

Example 4.1

Similarly,

$$
\begin{aligned}
P(Y = -1) &= \sum_{x=0}^{1} P(X = x, Y = -1) = 0 + \frac{1}{3} = \frac{1}{3} \\
P(Y = 0) &= \sum_{x=0}^{1} P(X = x, Y = 0) = \frac{1}{3} + 0 = \frac{1}{3} \\
P(Y = 1) &= 1 - P(Y = -1) - P(Y = 0) = \frac{1}{3}.
\end{aligned}
$$

Example 4.1

We can calculate the expected values and variances of X and Y from these marginal distributions.

$$E[X] = \sum_{x=0}^{1} xP(X = x) = 0 \times \frac{1}{3} + 1 \times \frac{2}{3} = \frac{2}{3}$$

$$E[Y] = \sum_{y=-1}^{1} yP(Y = y) = -1 \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3} = 0.$$

Example 4.1

To calculate $Var(X)$ and $Var(Y)$, we use the formula $Var(X) = E(X^2) - E(X)^2$.

$$E[X^2]=\sum_{x=0}^{1} x^2 P(X = x) = 0^2 \times \frac{1}{3} + 1^2 \times \frac{2}{3} = \frac{2}{3}$$

$$E[Y^2]=\sum_{y=-1}^{1} yP(Y = y) = (-1)^2 \times \frac{1}{3} + 0^2 \times \frac{1}{3} + 1^2 \times \frac{1}{3} = \frac{2}{3}.$$

Thus,

$$Var(X)=\frac{2}{3} - \left(\frac{2}{3}\right)^2 = \frac{2}{9}$$

$$Var(Y)=\frac{2}{3} - 0^2 = \frac{2}{3}$$

Example 4.1

To calculate the coefficient of correlation, we first calculate the covariance between X and Y. We have

$$Cov(X, Y) = E[XY] - E[X]E[Y],$$

where

$$\begin{aligned} E[XY] =& \sum_{x=0}^{1} \sum_{y=-1}^{1} xyP(X = x, Y = y) \\ =& 0 \times 0 \times \frac{1}{3} + 1 \times (-1) \times \frac{1}{3} + 1 \times 1 \times \frac{1}{3} = 0. \end{aligned}$$

Example 4.1

Hence,

$$Cov(X, Y) = E[XY] - E[X]E[Y] = 0 - \frac{2}{3} \times 0 = 0.$$

From the definition of the correlation coefficient

$$Cov(X, Y) = 0 \Leftrightarrow \rho(X, Y) = 0.$$

Hence, $\rho(X, Y) = 0$.

Note that if X and Y are independent, then $E[XY] = E[X]E[Y]$. However, $E[XY] = E[X]E[Y]$ is not sufficient to ensure that X and Y are independent.

Example 4.1

Note that if $\rho(X, Y) \neq 0$, then we can immediately state that X and Y are dependent.

When $\rho(X, Y) = 0$, we must check whether X and Y are independent from the definition. X and Y are independent if and only if

$$P(X = x, Y = y) = P(X = x)P(Y = y), \quad \forall x \in S_X, y \in S_Y.$$

To show that X and Y are dependent it is sufficient to show that this equation does not hold for some pair of values x and y.

Example 4.1

We have

$$P(X = 0, Y = -1) = 0 \neq P(X = 0)P(X = -1) = \left(\frac{1}{3}\right)^2.$$

It follows that X and Y are dependent. It can be seen that $X = |Y|$.

4.1.5 Conditional Distributions and Conditional Expected Values

Define the conditional distribution of X given $Y = y$, where $y \in S_Y$ by $\{P(X = x|Y = y)\}_{x \in S_X}$.

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}.$$

Analogously, the conditional distribution of Y given $X = x$, where $x \in S_X$ is given by $\{P(Y = y|X = x)\}_{y \in S_Y}$.

$$P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(Y = y)}.$$

Note: This definition is analagous to the definition of conditional probability given in Chapter 1.

Conditional Expected Value

The expected value of the function $g(X)$ given $Y = y$ is denoted $E[g(X)|Y = y]$. We have

$$E[g(X)|Y = y] = \sum_{x \in S_X} g(x)P(X = x|Y = y).$$

In particular,

$$E[X|Y = y] = \sum_{x \in S_X} xP(X = x|Y = y).$$

The expected value of the function $g(Y)$ given $X = x$ is defined analogously.

Probabilistic Regression

The graph of the probabilistic regression of X on Y is given by a scatter plot of the points $\{(y, E[X|Y = y])\}_{y \in S_Y}$.

Note that the variable we are conditioning on appears on the x-axis.

Analogously, the graph of the probabilistic regression of Y on X is given by a scatter plot of the points $\{(x, E[Y|X = x])\}_{x \in S_X}$.

These graphs illustrate the nature of the (probabilistic) relation between the variables X and Y.

Probabilistic Regression

The graph of the probabilistic regression of Y on X, $E[Y|X = x]$, may be

1. Monotonically increasing in x. This indicates a positive dependency between X and Y. $\rho(X, Y)$ will be positive.
2. Monotonically decreasing in x. This indicates a negative dependency between X and Y. $\rho(X, Y)$ will be negative.
3. Flat. This indicates that $\rho(X, Y) = 0$. It does not, however, indicate that X and Y are independent.
4. None of the above (e.g. oscillatory). In this case, X and Y are dependent, but this dependency cannot be described in simple terms such as positive or negative. the correlation coefficient can be 0, positive or negative in such a case (but cannot be -1 or 1).

Example 4.2

Discrete random variables X and Y have the following joint distribution:

	$Y=-1$	$Y=0$	$Y=1$
$X=0$	0	$\frac{1}{3}$	0
$X=1$	$\frac{1}{3}$	0	$\frac{1}{3}$

a) Calculate the conditional distributions of X given i) $Y=-1$, ii) $Y=0$, iii) $Y=1$.

b) Calculate the conditional distributions of Y given i) $X=0$, ii) $X=1$.

c) Draw the graph of the probabilistic regression of i) Y on X, ii) X on Y.

Example 4.2

a) i) The distribution of X conditioned on $Y = -1$ is given by

$$P(X = x|Y = -1) = \frac{P(X = x, Y = -1)}{P(Y = -1)}, \quad x \in S_X = \{0, 1\}.$$

From Example 4.1, $P(Y = -1) = 1/3$. Thus

$$P(X = 0|Y = -1) = \frac{P(X = 0, Y = -1)}{1/3} = 0$$

$$P(X = 1|Y = -1) = \frac{P(X = 1, Y = -1)}{1/3} = 1$$

Note that since X only takes the values 0 and 1

$$P(X = 1|Y = -1) = 1 - P(X = 0|Y = -1).$$

Example 4.2

ii) Similarly, the distribution of X conditioned on $Y = 0$ is given by

$$P(X = 0|Y = 0) = \frac{P(X = 0, Y = 0)}{P(Y = 0)} = \frac{1/3}{1/3} = 1$$
$$P(X = 1|Y = 0) = \frac{P(X = 1, Y = 0)}{P(Y = 0)} = \frac{0}{1/3} = 0$$

iii) The distribution of X conditioned on $Y = 1$ is given by

$$P(X = 0|Y = 1) = \frac{P(X = 0, Y = 1)}{P(Y = 1)} = \frac{0}{1/3} = 0$$
$$P(X = 1|Y = 1) = \frac{P(X = 1, Y = 1)}{P(Y = 1)} = \frac{1/3}{1/3} = 1$$

Example 4.2

b) i) The distribution of Y conditioned on $X = 0$ is given by

$$P(Y = y|X = 0) = \frac{P(X = 0, Y = y)}{P(X = 0)}, \quad y \in S_Y = \{-1, 0, 1\}.$$

From Example 4.1, $P(X = 0) = 1/3$. Thus

$$P(Y = -1|X = 0) = \frac{P(X = 0, Y = -1)}{P(X = 0)} = \frac{0}{1/3} = 0$$

$$P(Y = 0|X = 0) = \frac{P(X = 0, Y = 0)}{P(X = 0)} = \frac{1/3}{1/3} = 1$$

Since Y takes the values $-1, 0$ and 1, it follows that

$$P(Y = 1|X = 0) = 1 - P(X = -1|X = 0) - P(X = 0|Y = 0) = 0$$

Example 4.2

Similarly, the distribution of Y conditioned on $X = 1$ is given by

$$
\begin{aligned}
P(Y = -1|X = 1) &= \frac{P(X = 1, Y = -1)}{P(X = 1)} = \frac{1/3}{2/3} = 1/2 \\
P(Y = 0|X = 1) &= \frac{P(X = 1, Y = 0)}{P(X = 1)} = \frac{0}{2/3} = 0 \\
P(Y = 1|X = 1) &= 1 - P(Y = -1|X = 1) - P(Y = 0|X = 1) = 1/2
\end{aligned}
$$

Example 4.2

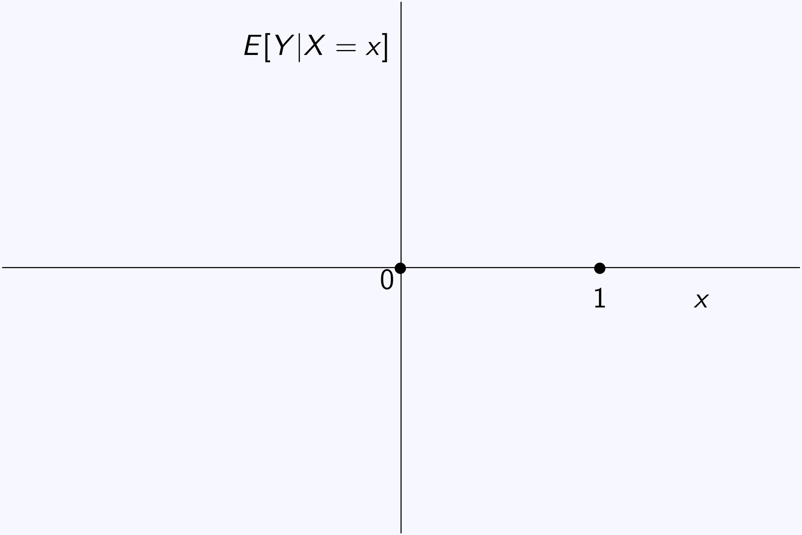
c) i) To plot the probabilistic regression of Y on X, we need first to calculate $E[Y|X = x]$, $x \in S_X = \{0, 1\}$. We have

$$E[Y|X = 0] = \sum_{y \in S_Y} yP(Y = y|X = 0) = -1 \times 0 + 0 \times 1 + 1 \times 0 = 0$$

$$E[Y|X = 1] = \sum_{y \in S_Y} yP(Y = y|X = 1) = -1 \times \frac{1}{2} + 0 \times 0 + 1 \times \frac{1}{2} = 0$$

Since $E[Y|X = x]$ is constant, it follows that $\rho(X, Y) = 0$.

Example 4.2



Example 4.2

c) ii) To plot the probabilistic regression of X on Y, we need first to calculate $E[X|Y = y]$, $y \in S_Y = \{-1, 0, 1\}$. We have

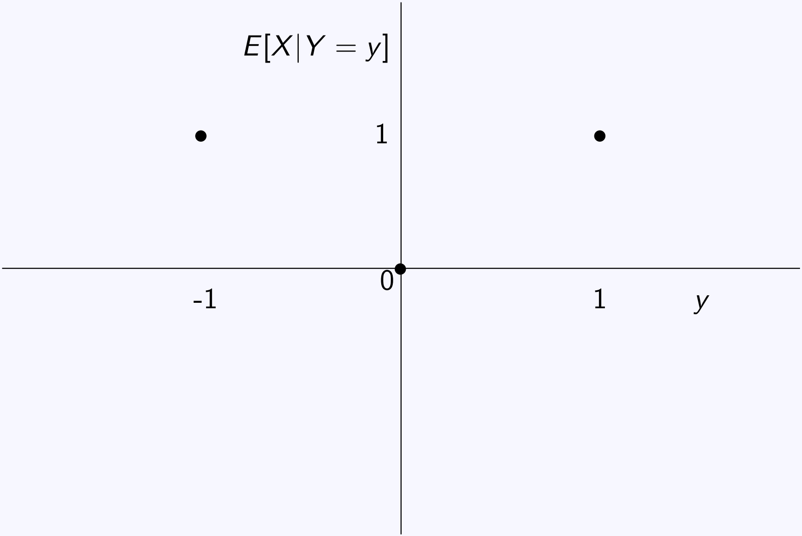
$$E[X|Y = -1] = \sum_{x \in S_X} xP(X = x|Y = -1) = 0 \times 0 + 1 \times 1 = 1$$

$$E[X|Y = 0] = \sum_{x \in S_X} yP(X = x|Y = 0) = 0 \times 1 + 1 \times 0 = 0$$

$$E[X|Y = 1] = \sum_{x \in S_X} xP(X = x|Y = 1) = 0 \times 0 + 1 \times 1 = 1$$

Since $E[X|Y = y]$ depends on y, it follows that X and Y are dependent.

Example 4.2



4.2 Joint Distributions of Two Continuous Random Variables

The joint distribution of two continuous random variables can be described by the joint density function $f_{X,Y}(x, y)$.

The joint density function $f_{X,Y}$ satisfies the following two conditions

$$f_{X,Y}(x, y) \geq 0$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)dxdy = 1$$

In practice, we integrate over the joint support of the random variables X and Y. This is the set of pairs (x, y) for which $f_{X,Y}(x, y) > 0$ (see Example 4.4).

Joint Distributions of Two Continuous Random Variables

The definitions used for continuous random variables are analogous to the ones used for discrete random variables. Summations are replaced by integrations.

Note that the **volume** under the density function is equal to 1.

The probability that $(X, Y) \in A$ is given by

$$P[(X, Y) \in A] = \int \int_A f_{X,Y}(x, y) dxdy$$

4.2.1 Marginal Density Functions for Continuous Random Variables

The marginal density function of X can be obtained by integrating the joint density function with respect to Y.

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

Similarly, the marginal density function of Y can be obtained by integrating the joint density function with respect to X.

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

If the joint support of X and Y is not a rectangle, some care should be made regarding the appropriate form of the joint density function according to the variable we are integrating with respect to (see Example 4.4).

These marginal density functions satisfy the standard conditions for the density function of a single random variable.

The continuous random variables X and Y are independent if and only if

$$f_{X,Y}(x, y) = f_X(x) f_Y(y), \quad \forall x, y.$$

4.2.3 The Expected Value of the Function of Two Continuous Random Variables

The expected value of the function $g(X, Y)$ is given by

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)dxdy.$$

In particular, the expected value of X is given by

$$E[X] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)dxdy.$$

As in the case of discrete variables, the expected values and variances of X and Y can be calculated from the respective marginal distributions.

4.2.4 Covariance and Coefficient of Correlation

The definitions of the covariance and coefficient of correlation, $Cov(X, Y)$ and $\rho(X, Y)$ are identical to the definitions used for discrete random variables.

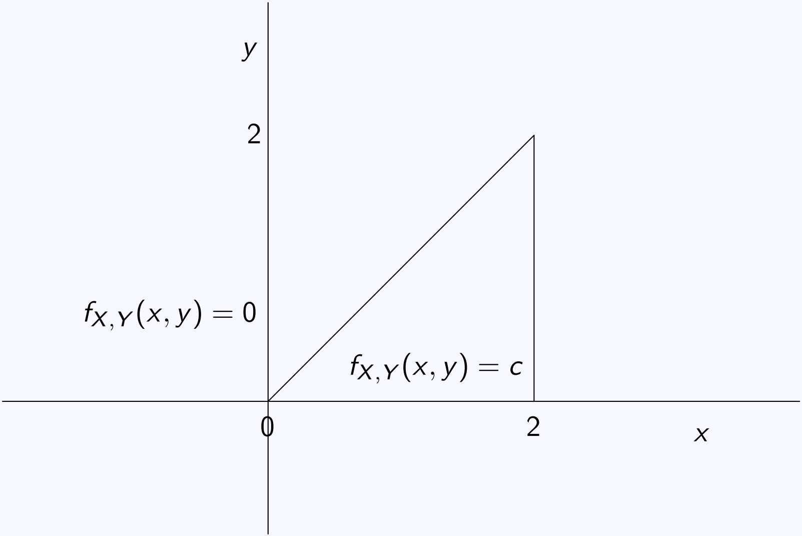
The properties of these measures and the interpretation of the coefficient of correlation are unchanged.

Example 4.3

The pair of random variables X and Y has a uniform joint distribution on the triangle A with apexes (0,0), (2,0) and (2,2), i.e. $f_{X,Y}(x,y) = c$ if $(x,y) \in A$, otherwise $f_{X,Y}(x,y) = 0$. The density function is illustrated on the following slide.

a) Find the constant c.
b) Calculate $P(X > 1, Y > 1)$.
c) Find the marginal densities of X and Y.
d) Find the expected values and variances of X and Y.
e) Find the coefficient of correlation, $\rho(X,Y)$.
f) Are X and Y independent?

Example 4.3



Example 4.3

a) **Method 1: The clever way**

Note that if the density function is constant on some set A, then the volume under the density curve (which must be equal to 1) is simply the height of the density curve, c, times the area of the set A.

The set A is a triangle with base and height both of length 2. The area of this triangle is thus $0.5 \times 2 \times 2 = 2$.

It follows that $2c = 1 \Rightarrow c = 0.5$.

Example 4.3

Method 2 The joint support of X and Y is the triangle A. Hence,

$$\int\int_A f_{X,Y}(x,y)dydx = 1.$$

We must be careful to get the boundaries of integration correct. Note that in A, x varies from 0 to 2. Fixing x, y can vary from 0 to x. Hence,

$$\begin{aligned}\int\int_A f_{X,Y}(x,y)dydx &= \int_{x=0}^{2}\left[\int_{y=0}^{x} cdy\right] dx \\ &= \int_{x=0}^{2}[cy]_{y=0}^{x}dx = \int_0^2 cxdx \\ &= c[x^2/2]_0^2 = 2c \Rightarrow c = 0.5.\end{aligned}$$

Example 4.3

It is possible to do the integration in the opposite order. Note that in A, y varies from 0 to 2. Fixing y, x can vary from y to 2. Hence,

$$\int\int_A f_{X,Y}(x,y)dydx = \int_{y=0}^{2}\left[\int_{x=y}^{2} cdx\right]dy.$$

Doing the integration this way obviously leads to the same answer, but the calculation is slightly longer.

Example 4.3

Method 1: The Clever Way If (X, Y) has a uniform distribution over the set A, then $P[(X, Y) \in B]$, where $B \subset A$ is given by

$$P[(X, Y) \in B] = \frac{|B|}{|A|},$$

where $|A|$ is the area of the set A.

Example 4.3

The area of A was calculated to be 2.

If both X and Y are greater than 1, then (X, Y) must belong to the interior of the triangle with apexes (1,1), (2,1) and (2,2).

The area of this triangle is 0.5. Hence,

$$P(X > 1, Y > 1) = \frac{0.5}{2} = \frac{1}{4}.$$

Example 4.3

Method 2: By integration When both X and Y are greater than 1, there is only a positive density on the triangle with apexes (1,1), (2,1) and (2,2).

In this triangle, $1 \leq x \leq 2$. Fixing x, $1 \leq y \leq x$. Hence,

$$P(X > 1, Y > 1) = \int_1^2 \int_1^x f_{X,Y}(x, y) dy dx.$$

Example 4.3

Thus

$$
\begin{aligned}
P(X > 1, Y > 1) &= \int_1^2 \int_1^x 0.5 dy dx \\
&= \int_1^2 [0.5y]_{y=1}^x \, dx \\
&= \int_1^2 (0.5x - 0.5) dx \\
&= \left[\frac{x^2}{4} - \frac{x}{2}\right]_1^2 \\
&= (1 - 1) - (1/4 - 1/2) = 1/4
\end{aligned}
$$

Example 4.3

c) Note that both X and Y take values in $[0, 2]$.

To get the marginal density function at x, $f_X(x)$, we integrate the joint density function with respect to y, i.e. we fix x and "sum" over y.

Given x, we should consider for what values of y the joint density is positive. From the graph of the density function, there is a positive density as long as $0 < y < x$. We have

$$f_X(x) = \int_{y=0}^{x} c dy = [cy]_0^x = cx = 0.5x.$$

Note that this holds on the support of X, i.e. $0 \leq x \leq 2$. Otherwise $f_X(x) = 0$.

Example 4.3

To get the marginal density function at y, $f_Y(y)$, we integrate the joint density function with respect to x, i.e. we fix y and "sum" over x.

For a given y, we should consider for what values of x the joint density is positive. From the graph of the density function, there is a positive density as long as $y < x < 2$. We have

$$f_Y(y) = \int_{x=y}^{2} c dx = [cx]_y^2 = c(2-y) = 1 - 0.5y.$$

Note that this holds on the support of Y, i.e. $0 \leq y \leq 2$. Otherwise $f_Y(y) = 0$.

Example 4.3

d) We can find $E(X)$ and $Var(X)$ from the marginal distribution

$$\begin{aligned} E(X) &= \int_0^2 x f_X(x) dx = \int_0^2 0.5x^2 dx \\ &= [x^3/6]_0^2 = \frac{4}{3} \\ E(X^2) &= \int_0^2 x^2 f_X(x) dx = \int_0^2 0.5x^3 dx \\ &= [x^4/8]_0^2 = 2 \end{aligned}$$

It follows that $Var(X) = E(X^2) - E(X)^2 = 2 - \left(\frac{4}{3}\right)^2 = \frac{2}{9}$.

Example 4.3

Similarly,

$$\begin{aligned} E(Y) &= \int_0^2 y f_Y(y) dy = \int_0^2 0.5y(2-y) dy \\ &= \int_0^2 (y - 0.5y^2) dy = [y^2/2 - y^3/6]_0^2 = \frac{2}{3} \\ E(Y^2) &= \int_0^2 y^2 f_Y(y) dy = \int_0^2 0.5y^2(2-y) dy \\ &= \int_0^2 (y^2 - 0.5y^3) dy = [y^3/3 - y^4/8]_0^2 = \frac{2}{3} \end{aligned}$$

It follows that $Var(Y) = E(Y^2) - E(Y)^2 = \frac{2}{3} - \left(\frac{2}{3}\right)^2 = \frac{2}{9}$.

Example 4.3

e) First, we find the covariance,
$Cov(X,Y) = E(XY) - E(X)E(Y)$. We have

$$E(XY) = \int\int_A xyf_{X,Y}(x,y)dxdy = \int_{x=0}^{2}\left[\int_{y=0}^{x} 0.5xydy\right]dx$$
$$= \int_{x=0}^{2}[xy^2/4]_{y=0}^{x}dx = \int_0^2 x^3/4dx$$
$$= [x^4/16]_0^2 = 1$$

It follows that $Cov(X,Y) = 1 - 4/3 \times 2/3 = 1/9$.

Example 4.3

We have

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$
$$= \frac{1/9}{\sqrt{2/9 \times 2/9}} = 1/2$$

Since $\rho(X,Y) \neq 0$, it follows that X and Y are dependent.

Example 4.3

Without calculating the correlation coefficient, we can show that X and Y are dependent from the definition.

Intuitively, X and Y are dependent, since the values Y can take depend on the realisation of X.

In such cases, it is easiest to show dependence by choosing x and y, such that $x \in S_X$ and $y \in S_Y$, but (x, y) is not in the joint support, i.e. $0 < x, y < 2$, but (x, y) does not lie in the triangle A.

$(0.5, 1.5)$ is such a point, thus

$$f_{X,Y}(0.5, 1.5) = 0 \neq f_X(0.5) f_Y(1.5) > 0.$$

Hence, X and Y are dependent.

4.2.5 Conditional Distributions and Conditional Expectations

The conditional density function of X given $Y = y$, where $y \in S_Y$ is given by $f_{X|Y=y}$, where

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

The conditional density function of Y given $X = x$, where $x \in S_X$ is given by $f_{Y|X=x}$, where

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

Conditional Expectations

The conditional expectation of $g(X)$ given $Y = y$, where $y \in S_Y$ is given by $E[g(X)|Y = y]$, where

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

In particular,

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

In practice, we integrate over the interval where $f_{X|Y=y}$ is positive. The conditional expected value of $g(Y)$ given $X = x$ is defined analogously.

Probabilistic Regression for Continuous Random Variables

In the case of discrete random variables, the probabilistic regression of Y on X is given by a scatter plot.

In the case of continuous random variables, the probabilistic regression of Y on X is given by a curve.

The x-coordinate in this plot is given by x, the y-coordinate is given by $E[Y|X = x]$.

Probabilistic Regression for Continuous Random Variables

The probabilistic regression of X on Y is defined analogously.

Since we are conditioning on y, the x-coordinate is given by y.

The y-coordinate is given by $E(X|Y = y)$.

A probabilistic regression curve can be interpreted in an analogous way to the interpretation of probabilistic regression for discrete random variables (see Example 4.4).

Example 4.4

For the joint distribution given in Example 4.3

i) Calculate the conditional density function of X given $Y = y$. Derive $E(X|Y = y)$.

ii) Calculate the conditional density function of Y given $X = x$. Derive $E(Y|X = x)$.

Example 4.4

i) We have $f_Y(y) = 1 - y/2$, $0 \leq y \leq 2$. Thus $S_Y = [0, 2)$. Hence, for $y \in [0, 2)$

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

Since the joint support of X and Y is not a rectangle, we must be careful with the form of $f_{X,Y}(x,y)$. Here, we are fixing y, there is a positive joint density if $y \leq x \leq 2$. Hence, for $y \leq x \leq 2$,

$$f_{X|Y=y}(x) = \frac{1/2}{1 - y/2} = \frac{1}{2 - y}.$$

Otherwise, $f_{X|Y=y}(x) = 0$.

Example 4.4

We have $f_{X|Y=y}(x) = 1/(2-y)$ for $y \leq x \leq 2$. Hence,

$$
\begin{aligned}
E[X|Y=y] &= \int_y^2 x f_{X|Y=y}(x)dx = \int_y^2 \frac{x}{2-y} dx \\
&= \left[\frac{x^2}{2(2-y)} \right]_{x=y}^2 = \frac{4-y^2}{2(2-y)} \\
&= \frac{(2-y)(2+y)}{2(2-y)} = 1 + y/2
\end{aligned}
$$

Since this regression curve is increasing, it follows that a) X and Y are dependent, b) $\rho(X, Y) > 0$.

Example 4.4

ii) We have $f_X(x) = x/2$, $0 \leq x \leq 2$. Thus $S_X = (0, 2]$. Hence, for $x \in (0, 2]$

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

Again, we must be careful with the form of $f_{X,Y}(x, y)$. Here, we are fixing x, there is a positive joint density if $0 \leq y \leq x$. Hence, for $0 \leq y \leq x$,

$$f_{Y|X=x}(y) = \frac{1/2}{x/2} = \frac{1}{x}.$$

Otherwise, $f_{Y|X=x}(y) = 0$.

Example 4.4

We have $f_{Y|X=x}(y) = 1/x$ for $0 \leq y \leq x$. Hence,

$$\begin{aligned} E[Y|X=x] &= \int_0^x y f_{Y|X=x}(y) dy = \int_0^x \frac{y}{x} dy \\ &= \left[\frac{y^2}{2x}\right]_{y=0}^x = \frac{x^2}{2x} \\ &= x/2 \end{aligned}$$

Again, since this regression curve is increasing, it follows that a) X and Y are dependent, b) $\rho(X, Y) > 0$.

4.3 The Multivariate Normal Distribution

Let $\mathbf{X} = (X_1, X_2, \ldots, X_k)$ be a vector of k random variables.

Let $E[X_i] = \mu_i$ and $\mu = (\mu_1, \mu_2, \ldots, \mu_k)$, i.e. μ is the vector of expected values

Let Σ be a $k \times k$ matrix where the entry in row i, column j, $\Sigma_{i,j}$ is given by $\mathrm{Cov}(X_i, X_j)$.

Σ is called the covariance matrix of the random variables $X_1, X_2, \ldots, X_k$.

Properties of the Covariance matrix, Σ

1. Σ is symmetric [since $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$].
2. $\Sigma_{i,i} = \text{Cov}(X_i, X_i) = \text{Var}(X_i) = \sigma_i^2$, i.e. the i-th entry on the leading diagonal is the variance of the i-th random variable.
3. Σ is positive definite, i.e. if $\mathbf{y}$ is any $1 \times k$ vector, $\mathbf{y}\Sigma\mathbf{y}^T \geq 0$.
4. Dividing each element $\Sigma_{i,j}$ by $\sigma_i \sigma_j$ (where σ_i is the standard deviation of the i-th variable), we obtain the correlation matrix $\mathbf{R}$. $\rho_{i,j}$ (the element in row i, column j of this matrix gives the coefficient of correlation between X_i and X_j. Note $\rho_{i,i} = 1$ by definition.

The Form of the Multivariate Distribution

$\mathbf{X} = (X_1, X_2, \ldots, X_k)$ has a multivariate normal distribution if and only if the joint density function is given by (in matrix form)

$$f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-k/2}|\Sigma|^{-1/2}\exp[-1/2(\mathbf{x}-\mu)\Sigma^{-1}(\mathbf{x}-\mu)^T]$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $|\Sigma|$ is the determinant of the covariance matrix.

If $\mathbf{X} = (X_1, X_2, \ldots, X_k)$ has a multivariate normal distribution, then any linear combination of the X_i has a (univariate) normal distribution. In particular, each X_i has a (univariate) normal distribution.

The Bivariate Normal distribution

In this case $k = 2$, $\mathbf{X} = [X_1, X_2]$, $\mu = [\mu_1, \mu_2]$. The covariance matrix is given by

$$\Sigma = \begin{pmatrix} \sigma_1^2 & Cov(X_1, X_2) \\ Cov(X_1, X_2) & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

where σ_i is the standard deviation of X_i and ρ is the coefficient of correlation between X_1 and X_2. The correlation matrix is given by

$$\mathbf{R} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

The Bivariate Normal distribution

In this case, the determinant of the covariance matrix is given by

$$|\Sigma| = \sigma_1^2\sigma_2^2 - \rho^2\sigma_1^2\sigma_2^2 = \sigma_1^2\sigma_2^2(1-\rho^2).$$

Hence, $|\Sigma|^{1/2} = \sigma_1\sigma_2\sqrt{1-\rho^2}$.

Also,

$$\Sigma^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}$$

It follows that

$$-1/2(\mathbf{x}-\mu)\Sigma^{-1}(\mathbf{x}-\mu) = -1/2(x_1-\mu_1, x_2-\mu_2)\Sigma^{-1} \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}$$

$$= \frac{-1}{2(1-\rho^2)} \left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} \right]$$

The Bivariate Normal distribution

Hence, the joint density function of $\mathbf{X} = [X_1, X_2]$ is given by

$$f_{\mathbf{X}}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{ \frac{-1}{2(1-\rho^2)} \left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \right.\right.$$
$$\left.\left. + \frac{(x_2-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} \right] \right\}.$$

The Bivariate Normal distribution

Note that if $\rho = 0$, then

$$
\begin{aligned}
f_{\mathbf{X}}(x_1, x_2) =& \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{ \frac{-1}{2} \left[\frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right] \right\} \\
=& f_{X_1}(x_1) f_{X_2}(x_2).
\end{aligned}
$$

Hence, it follows that if $[X_1, X_2]$ has a bivariate normal distribution, then $\rho = 0$ is a necessary and **sufficient** condition for X_1 and X_2 to be independent. Note that, In general, in order to prove independence it is not sufficient to show that $\rho = 0$ (see Example 4.2).

Linear Combinations of Random Variables

Note that for any variables $X_1, X_2, \ldots, X_k$.

$$E[\sum_{i=1}^{k} a_i X_i] = a_1 E[X_1] + a_2 E[X_2] + \ldots + a_n E[X_n].$$

$$Var[\sum_{i=1}^{k} a_i X_i] = \left[\sum_{i=1}^{k} a_i^2 Var(X_i)\right] + 2 \sum_{i=1}^{k} \sum_{j=i+1}^{k} a_i a_j Cov(X_i, X_j).$$

In particular, when $k = 2$,

$$Var[a_1 X_1 + a_2 X_2] = a_1^2 Var(X_1) + a_2^2 Var(X_2) + 2 a_1 a_2 Cov(X_1, X_2).$$

When the X_i are independent, then

$$Var[a_1 X_1 + a_2 X_2 + \ldots + a_k X_k] = \sum_{i=1}^{k} a_i^2 Var(X_i).$$

Example 4.5

Suppose the height of males has a normal distribution with mean 170cm and standard deviation 15cm and the correlation of the height of a father and son is 0.6.

1. Derive the joint distribution of the height of two males chosen at random.
2. Calculate the probability that the average height of two such individuals is greater than 179cm.
3. Assuming the joint distribution of the height of father and son is a bivariate normal distribution, derive the joint distribution of the height of a father and son.
4. Calculate the probability that the average height of a father and son is greater than 179cm.

Example 4.5

1. Since the individuals are picked at random, we may assume that the two heights are independent. The univariate density function of the i-th height ($i = 1, 2$) is

$$f_{X_i}(x_i) = \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left(\frac{-(x_i - \mu_i)^2}{2\sigma_i^2}\right) = \frac{1}{15\sqrt{2\pi}} \exp\left(\frac{-(x_i - 170)^2}{2 \times 15^2}\right),$$

The joint density function of the two heights is the product of these density functions, i.e.

$$\begin{aligned} f_{X_1,X_2}(x_1, x_2) &= f_{X_1}(x_1) f_{X_2}(x_2) \\ &= \frac{1}{15^2 \times 2\pi} \exp\left(\frac{-(x_1 - 170)^2 - (x_2 - 170)^2}{2 \times 15^2}\right). \end{aligned}$$

Example 4.5

2. Note that the sum of the two heights also has a normal distribution. Let $S = X_1 + X_2$. $E[S] = E[X_1] + E[X_2] = 340$.

Since X_1 and X_2 are independent,

$\text{Var}(S)=\text{Var}(X_1)+\text{Var}(X_2) = 2 \times 15^2 = 450$.

We must calculate

$$P([X_1 + X_2]/2 > 179) = P(X_1 + X_2 > 358) = P(S > 358).$$

Example 4.5

Since S has a normal distribution, by standardising we obtain,

$$\begin{aligned} P(S > 358) &= P\left(\frac{S - E(S)}{\sigma_S} > \frac{358 - 340}{\sqrt{450}}\right) \\ &= P(Z > 0.85) = 0.1977. \end{aligned}$$

Example 4.5

3. The joint distribution of the height of a father and his son is given by

$$
\begin{aligned}
f_{\mathbf{X}}(x_1, x_2) = & \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{ \frac{-1}{2(1-\rho^2)} \left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \right.\right. \\
& \left.\left. + \frac{(x_2-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} \right] \right\} \\
= & \frac{1}{360\pi} \exp\left\{ \frac{-1}{1.28 \times 225} \left[(x_1-170)^2 + \right.\right. \\
& \left.\left. + (x_2-170)^2 - 1.2(x_1-170)(x_2-170) \right] \right\}.
\end{aligned}
$$

Example 4.5

4. If the joint distribution of the height of father and son has a bivariate distribution, then the sum of their heights has a normal distribution. In this case, $E[S] = E[X_1] + E[X_2] = 340$.

$\mathrm{Var}(S)=\mathrm{Var}(X_1)+\mathrm{Var}(X_2) + 2\mathrm{Cov}(X_1, X_2)$. We have

$$\rho = \frac{Cov(X_1, X_2)}{\sigma_1\sigma_2}$$

$$Cov(X_1, X_2) = \rho\sigma_1\sigma_2 = 0.6 \times 15 \times 15 = 135.$$

Example 4.5

It follows that

$$\begin{aligned} Var(S) =& Var(X_1) + Var(X_2) + 2Cov(X_1, X_2) \\ =& 225 + 225 + 2 \times 135 = 720. \end{aligned}$$

By standardising, we obtain

$$\begin{aligned} P(S > 358) =& P\left(\frac{S - E(S)}{\sigma_S} > \frac{358 - 340}{\sqrt{720}}\right) \\ =& P(Z > 0.67) = 0.2514. \end{aligned}$$