Chapter 3*

September 2, 2018

1 Random Variables

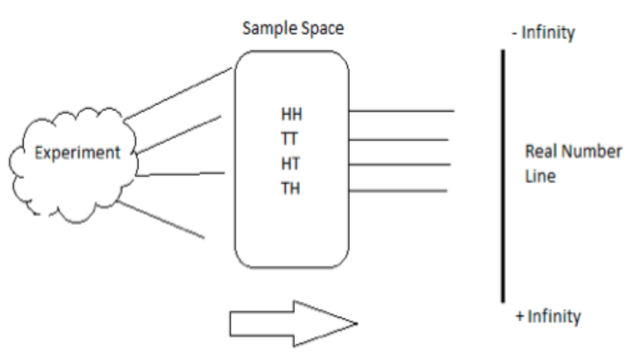
In many situations we are not concerned directly with the outcome of an experiment, but instead with some *function* of the outcome.

For example, when rolling two dice we are generally not interested in the separate values of the dice, but instead we are concerned with the sum of the two dice.

Definition: A *random variable* X is a function from the sample space into the real numbers, i.e., $X : S \to \mathbb{R}$.

A random variable that can take on at most a *countable* number of possible values is said to be **discrete**.

Example: Flip two coins. $S = \{TT, HT, TH, HH\}$. Let X be the "number of heads".



Thus, X is a random variable that takes values on $\{0, 1, 2\}$.

---

*Notes for Chapter 3 of DeGroot and Schervish adapted from Giovanni Motta's, Bodhisattva Sen's and Martin Lindquists notes for STAT W4109/W4105.

Table 1: Values of the r.v X

outcomes	value of X
TT	0
HT	1
TH	1
HH	2

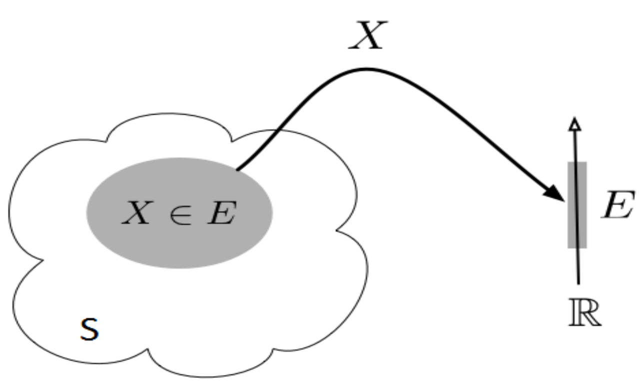
Table 2: Values of the r.v X

outcomes	value of X
TT	0
HT	0
TH	0
HH	4

We assign probabilities to random variables, e.g.,

$$\mathbb{P}(X=0)=\frac{1}{4}, \qquad \mathbb{P}(X=1)=\frac{1}{2}, \qquad \mathbb{P}(X=2)=\frac{1}{4}.$$

In general, for any $E \subset \mathbb{R}$, by $\mathbb{P}(X \in E)$ we mean $\mathbb{P}(\{s \in S : X(s) \in E\})$.



Note that you could define any number of random variables on an experiment.

Example: In a game, which consists of flipping two coins, you start with 1 USD. On each flip, if you get a H you double your current fortune, while you lose everything if you get a T.

Let X = your total fortune after two flips. Then, X is a random variable that takes values on $\{0, 4\}$. Thus,

$$\mathbb{P}(X = 0) = \frac{3}{4}, \qquad \mathbb{P}(X = 4) = \frac{1}{4}.$$

When a probability has been specified on the sample space of an experiment, we can determine probabilities associated with the possible values of each random variable X. The probability in the sample space induces a probability on the set of real numbers.

Definition: The *distribution* of X is the collection of all probabilities of the form $\mathbb{P}(X \in C)$ for all sets C of real numbers such that $\{X \in C\}$ is an event.

For a discrete random variable we define the *probability mass function* $p : \mathbb{R} \to [0, 1]$ of X, as

$$p(x) = \mathbb{P}(X = x) = \mathbb{P}(\{s \in S : X(s) = x\}).$$

Note that random variables are often capitalized, while their values are often written by lower-case letters.

If a discrete random variable X assumes the values $x_1, x_2, x_3, \ldots$ then

(i) $p(x_i) \geq 0$ for $i = 1, 2, 3, \ldots$

(ii) $p(x) = 0$ for all other values of x.

(iii) $\sum_{i=1}^{\infty} p(x_i) = 1$.

Example: Independent trials that consist of flipping a coin that has probability p of turning up heads, is performed until a head appears.

Let X = number of times the coin is flipped.

X is a discrete random variable that can take values in $\{1, 2, 3, 4, \ldots\}$. The outcomes of the experiment can be represented as

$$\{H, TH, TTH, TTTH, TTTTH, TTTTTH, TTTTTTH, \ldots\}.$$

Then,

$$\begin{aligned}
\mathbb{P}(X = 1) &= p \\
\mathbb{P}(X = 2) &= (1 - p)p \\
\mathbb{P}(X = 3) &= (1 - p)^2 p \\
&\vdots \\
\mathbb{P}(X = n) &= (1 - p)^{n-1} p \\
&\ldots
\end{aligned}$$

Example: The probability mass function of a random variable X is given by

$$p(i) = c\frac{L^i}{i!}, \text{ for } i = 0, 1, 2, \ldots$$

where $L > 0$ if given and c is a constant. Find

(a) The value of c.

(b) $\mathbb{P}(X = 0)$.

(c) $\mathbb{P}(X > 2)$.

Solution: (a) Since $\sum_{i=0}^{\infty} p(i) = 1$, we have

$$c\sum_{i=0}^{\infty} \frac{L^i}{i!} = 1.$$

Since $e^x = \sum_{i=0}^{\infty} x^i/i!$, we have that $ce^L = 1$. So $c = e^{-L}$ and hence $p(i) = e^{-L}\frac{L^i}{i!}$.

(b) $\mathbb{P}(X = 0) = p(0) = e^{-L}L^0/0! = e^{-L}$.

(c)

$$\begin{aligned} \mathbb{P}(X > 2) = 1 - \mathbb{P}(X \le 2) &= 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) - \mathbb{P}(X = 2) \\ &= 1 - p(0) - p(1) - p(2) \\ &= 1 - e^{-L} - Le^{-L} - L^2e^{-L}/2. \end{aligned}$$

1.1 The cumulative distribution function

The *cumulative distribution function* (c.d.f) $F : \mathbb{R} \to [0, 1]$ of the random variable X, is a defined for all real numbers b, by

$$F(b) = \mathbb{P}(X \le b).$$

Relationship between p.m.f and c.d.f for discrete random variables:

$$F(b) = \sum_{x \le b} p(x).$$

Example: Flip two coins. Let X = number of heads. Then $p(0) = 1/4$, $p(1) = 1/2$, $p(2) = 1/4$. Let F be the c.d.f of X.

[Draw the c.d.f!]

Thus, F is a **step function**. Properties of the c.d.f:

Table 3: Cumulative distribution function

b	$F(b)$
$(-\infty, 0)$	0
$[0, 1)$	$1/4$
$[1, 2)$	$3/4$
$[2, \infty)$	1

(i) $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$ and $F(\infty) = \lim_{x \to \infty} F(x) = 1$.

(ii) $F(x)$ is a non-decreasing function of x.

(iii) $F(x)$ is right-continuous, i.e., $\lim_{x \to a+} F(x) = F(a)$.

How do we show (i)?

Theorem 1.1. *Let $A_1, A_2, \ldots$ be an infinite sequence of events such that $A_1 \subset A_2 \subset \ldots$. Prove that*

$$\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \lim_{n \to \infty} \mathbb{P}(A_n).$$

Proof. Let us define the sequence of events $\{B_i\}_{i=1}^n$ such that $B_1 = A_1$, $B_2 = A_1^c \cap A_2$, $B_3 = A_1^c \cap A_2^2 \cap A_3$, $\ldots$, $B_n = A_1^c \cap A_2^c \cap \ldots \cap A_{n-1}^c \cap A_n$. Then B_i's are disjoint and for every $n \geq 1$,

$$\cup_{i=1}^{n} A_i = \cup_{i=1}^{n} B_i \qquad \text{and} \qquad \cup_{i=1}^{\infty} A_i = \cup_{i=1}^{\infty} B_i.$$

Thus,

$$\mathbb{P}\left(\cup_{i=1}^{\infty} A_i\right) = \mathbb{P}\left(\cup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(B_i),$$

where the last equality follows from Axiom 3. Now, when $A_1 \subset A_2 \subset \ldots$, then $A_n = \cup_{i=1}^{n} A_i = \cup_{i=1}^{n} B_i$, and thus

$$\sum_{i=1}^{\infty} \mathbb{P}(B_i) = \lim_{n \to \infty} \sum_{i=1}^{n} \mathbb{P}(B_i) = \lim_{n \to \infty} \mathbb{P}(A_n).$$

□

Theorem 1.2. *Let $A_1, A_2, \ldots$ be an infinite sequence of events such that $A_1 \supset A_2 \supset \ldots$. Prove that*

$$\mathbb{P}(\cap_{i=1}^{\infty} A_i) = \lim_{n \to \infty} \mathbb{P}(A_n).$$

Proof. Consider the sequence $A_1^c, A_2^c, \ldots$, and apply Theorem 1.1. □

To prove (i) take a non-increasing sequence $x_n \to -\infty$. Note that $\{X \leq x_1\} \supset \{X \leq x_2\} \supset \ldots$. Then note that $\cap_{i=1}^{\infty}\{X \leq x_i\} = \emptyset$. Thus, by Theorem 1.2,

$$\lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} \mathbb{P}(X \leq x_n) = \mathbb{P}(\emptyset) = 0.$$

Example: Independent trials that consist of flipping a coin that has probability p of turning up heads, is performed until a head appears. Let X = number of times the coin is flipped. What is the c.d.f of X?

X is a discrete random variable that can take on values in $\{1, 2, 3, 4, \ldots\}$. We also know its p.m.f.

For $b = 1, 2, 3, \ldots$, we have

$$\begin{aligned} F(b) = \mathbb{P}(X \leq b) &= \sum_{i=1}^{b} \mathbb{P}(X = i) = \sum_{i=1}^{b} p(1-p)^{i-1} \\ &= \sum_{j=0}^{b-1} p(1-p)^j = p\frac{1-(1-p)^b}{1-(1-p)} = 1-(1-p)^b. \end{aligned}$$

Then, $F(y) = \begin{cases} 0, & \text{for } y < 1 \\ 1-(1-p)^b & \text{for } b \leq y < (b+1). \end{cases}$

It is important to note that all probability questions about a random variable X can be answered in terms of the p.m.f or c.d.f: these are equivalent, and both contain all the information we need.

For example, we may like to calculate $\mathbb{P}(a < X \leq b)$:

$$\begin{aligned} \{X \leq b\} &= \{X \leq a\} \cup \{a < X \leq b\} \\ \mathbb{P}(X \leq b) &= \mathbb{P}(X \leq a) + \mathbb{P}(a < X \leq b) \\ \mathbb{P}(a < X \leq b) &= \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a) = F(b) - F(a). \end{aligned}$$

In order to calculate $\mathbb{P}(X < b)$, we use

$$\mathbb{P}(X < b) = \mathbb{P}\left(\cup_{n=1}^{\infty}\left\{X \leq b - \frac{1}{n}\right\}\right) = \lim_{n\to\infty} \mathbb{P}\left(X \leq b - \frac{1}{n}\right) = \lim_{n\to\infty} F\left(b - \frac{1}{n}\right) = F(b-).$$

where we have used the preceding result.

If we know the c.d.f we can calculate the p.d.f by

$$p(x) = \mathbb{P}(X = x) = \mathbb{P}(X \leq x) - \mathbb{P}(X < x) = F(x) - F(x-).$$

Example: The distribution function of the random variable X is given by

$$F(x) = \begin{cases} 0, & x < 0 \\ 1/2 & 0 \leq x < 1 \\ 2/3 & 1 \leq x < 2 \\ 11/12 & 2 \leq x < 3 \\ 1 & 3 \leq x. \end{cases}$$

Draw the c.d.f F. Compute

(a) $\mathbb{P}(X < 3)$

(b) $\mathbb{P}(X = 1)$

(c) $\mathbb{P}(X > 1/2)$

(d) $\mathbb{P}(2 < X \leq 4)$.

Solution: (a) $\mathbb{P}(X < 3) = F(3-) = 11/12$.

(b) $\mathbb{P}(X = 1) = \mathbb{P}(X \leq 1) - \mathbb{P}(X < 1) = F(1) - F(1-) = 2/3 - 1/2 = 1/6$.

(c) $\mathbb{P}(X > 1/2) = 1 - \mathbb{P}(X \leq 1/2) = 1 - F(1/2) = 1 - 1/2 = 1/2$.

(d) $\mathbb{P}(2 < X \leq 4) = F(4) - F(2) = 1 - 11/12 = 1/12$.

1.2 Continuous random variables

Often there are random variables of interest whose possible values are **not** *countable*.

Example: Let X be the lifetime of a battery. Then X takes values in the interval $[0, \infty)$ (all non-negative real numbers).

Example: Pick a number randomly in the interval $(0, 1)$ so that it is just as likely to take any value between 0 and 1. Call it X.

Intuition: Cannot assign positive probability to any real number, but can assign probabilities to *intervals*.

The probability that a continuous random variable will assume any fixed value is zero. There are an uncountable number of values that X can take, thus the probability of X taking any particular value cannot exceed zero.

Definition: A random variable X is called *continuous* if there exists a non-negative function $f : \mathbb{R} \to \mathbb{R}$, such that for any set $B \subset \mathbb{R}$ of real numbers

$$\mathbb{P}(X \in B) = \int_B f(x)dx.$$

The function f is called the probability density function (p.d.f) of X; sometimes written f_X.

Compare this with the probability mass function. All probability statements about X can be answered in terms of the p.d.f f, i.e.,

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)dx.$$

Note that

$$\mathbb{P}(X \in (-\infty, \infty)) = \int_{-\infty}^{\infty} f(x)dx.$$

The density can be larger than one (in fact, it can be unbounded) – it just has to integrate to one (compare with p.m.f's, which has to be less than one).

Example: Suppose that X is a continuous random variable whose p.d.f is given by

$$f(x) = \begin{cases} C(4x - 2x^2), & \text{for } 0 < x < 2 \\ 0 & \text{otherwise.} \end{cases}$$

(a) What is the value of C?

(b) Find $\mathbb{P}(X > 1)$.

Solution: (a) Since f is a p.d.f $\int_{-\infty}^{\infty} f(x)dx = 1$. Show that $C = 3/8$.

(b) $\mathbb{P}(X > 1) = \int_1^{\infty} f(x)dx = ... = 1/2$.

Thus, for a continuous random variable X, we have $\mathbb{P}(X < x) = \mathbb{P}(X \leq x)$.

The cumulative distribution function of X is defined by

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(y)dy.$$

As before we can express probabilities in terms of the c.d.f:

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)dx = \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx = F(b) - F(a).$$

The relationship between the p.d.f and the c.d.f can also be expressed as

$$\frac{d}{dx}F(x) = f(x).$$

Non-uniqueness of the p.d.f: The p.d.f can be changed at a finite number of points, or even a certain infinite sequences of points, without changing the value of the integral of the p.d.f over any subset A.

1.2.1 Uniform random variables

A random variable is said to be *uniformly distributed* over the interval $(0,1)$ if its p.d.f is given by

$$f(x) = \begin{cases} 1 & \text{for } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

This is a valid density function, since $f(x) \geq 0$ and $\int_{-\infty}^{\infty} f(x)dx = \int_0^1 dx = 1$.

The c.d.f is given by

$$F(x) = \int_{-\infty}^{x} f(y)dy = \int_0^x dy = x$$

for $x \in (0,1)$. Thus,

$$F(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ x & \text{for } 0 < x < 1 \\ 1 & \text{for } x \geq 1. \end{cases}$$

[Draw p.d.f and c.d.f.]

Intuition: X is just as likely to take any value between 0 and 1.

For any $0 < a < b < 1$, $\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)dx = b - a$.

The probability that X is in any particular subinterval of $(0,1)$ equals the length of the interval.

Example: X is uniformly distributed over $(0,1)$. Calculate $\mathbb{P}(X < 0.3)$.

In general, a random variable is said to be uniformly distributed over the interval (α, β) if its p.d.f is given by

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{for } \alpha < x < \beta \\ 0 & \text{otherwise.} \end{cases}$$

The c.d.f is given by

$$F(x) = \begin{cases} 0 & \text{for } x \leq \alpha \\ \frac{x-\alpha}{\beta-\alpha} & \text{for } \alpha < x < \beta \\ 1 & \text{for } x \geq \beta. \end{cases}$$

1.2.2 Quantile function

Let X be a random variable with c.d.f F. For each $p \in (0,1)$, define

$$F^{-1}(p) = \inf\{x : F(x) \geq p\},$$

where 'inf' denotes the infimum of the set. Then $F^{-1}(p)$ is called the *p-quantile* of X or the $100p$ *percentile* of X.

The function $F^{-1} : (0, 1) \to \mathbb{R}$ is called the *quantile function*.

Example: Standardized test scores: Many universities in the US rely on standardized test scores as part of their admission process.

When the c.d.f of a random variable X is continuous and one-to-one over the whole set of possible values of X, the inverse F^{-1} of F exists and equals the quantile function of X.

Median: The 0.5-quantile or the 50th percentile of a distribution is called its *median*.

1.3 Jointly distributed random variables

So far we have dealt with one random variable at a time. Often it is interesting to work with several random variables at once.

Definition: An n-dimensional random vector is a function from a sample space S into $\mathbb{R}^n$.

For a two-dimensional (or bivariate) random vector, each point in the sample space is associated with a point on the plane.

Example: Roll two dice. Let X = sum of the dice, and Y = the absolute value of the difference of the two dice.

Solution: $S = \{(i, j) : i, j = 1, \ldots, 6\}$. S consists of 36 points. X takes values from 2 to 12. Y takes values between 0 and 5.

For the sample point: $(1, 1)$, $X = 2$ and $Y = 0$; for $(1, 2)$, $X = 3$ and $Y = 1$, etc.

The random vector above is an example of a discrete random vector.

Definition: If (X, Y) is a random vector then the *joint cumulative distribution* function F_{XY} of (X, Y) is defined by

$$F_{XY}(a, b) = \mathbb{P}(X \le a, Y \le b), \qquad \text{for } a, b \in \mathbb{R}.$$

The function F_{XY} has the following properties:

$$\begin{aligned}
F_{XY}(\infty, \infty) &= \lim_{a\to\infty, b\to\infty} \mathbb{P}(X \le a, Y \le b) = 1 \\
F_{XY}(-\infty, b) &= \lim_{a\to-\infty} \mathbb{P}(X \le a, Y \le b) = 0 \\
F_{XY}(a, -\infty) &= \lim_{b\to-\infty} \mathbb{P}(X \le a, Y \le b) = 0 \\
F_{XY}(a, \infty) &= \lim_{b\to\infty} \mathbb{P}(X \le a, Y \le b) = \mathbb{P}(X \le a) = F_X(a) \\
F_{XY}(\infty, b) &= \lim_{a\to\infty} \mathbb{P}(X \le a, Y \le b) = \mathbb{P}(Y \le b) = F_Y(b).
\end{aligned}$$

F_X and F_Y are called the *marginal distributions* of X and Y, respectively.

All joint probability statements about X and Y can be answered in terms of their joint distribution function.

For $a_1 < a_2$ and $b_1 < b_2$,

$$\mathbb{P}(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F(a_2, b_2) - F(a_2, b_1) - F(a_1, b_2) + F(a_1, b_1).$$

Definition: If X and Y are both discrete random variables, we define the *joint probability mass function* of X and Y by

$$p_{XY}(a, b) = \mathbb{P}(X = a, Y = b), \qquad \text{for } a, b \in \mathbb{R}.$$

The probability mass function of X can be obtained from p_{XY} by

$$p_X(x) = \mathbb{P}(X = x) = \sum_{y: p(x,y)>0} p(x, y).$$

Similarly, we can obtain the p.m.f of Y by

$$p_Y(y) = \mathbb{P}(Y = y) = \sum_{x: p(x,y)>0} p(x, y).$$

Example: Suppose two balls are chosen from a box containing 3 white, 2 red and 5 blue balls. Let X= the number of white balls chosen, and Y= the number of blue balls chosen. Find the joint p.m.f of X and Y. Also, find the p.m.f of X.

Solution: Both X and Y takes values in $\{0, 1, 2\}$. Note that,

$$\begin{aligned}
p(0,0) &= \binom{2}{2} / \binom{10}{2} = \frac{1}{45} \\
p(0,1) &= \binom{2}{1}\binom{5}{1} / \binom{10}{2} = \frac{10}{45} \\
p(0,2) &= \binom{5}{2} / \binom{10}{2} = \frac{10}{45} \\
p(1,0) &= \binom{2}{1}\binom{3}{1} / \binom{10}{2} = \frac{6}{45} \\
p(1,1) &= \binom{5}{1}\binom{3}{1} / \binom{10}{2} = \frac{15}{45} \\
p(1,2) &= 0 \\
p(2,0) &= \binom{3}{2} / \binom{10}{2} = \frac{3}{45} \\
p(2,1) &= 0 \\
p(2,2) &= 0.
\end{aligned}$$

Show that X takes the values 0,1,2 with probabilities 21/45, 21/45 and 3/45, respectively.

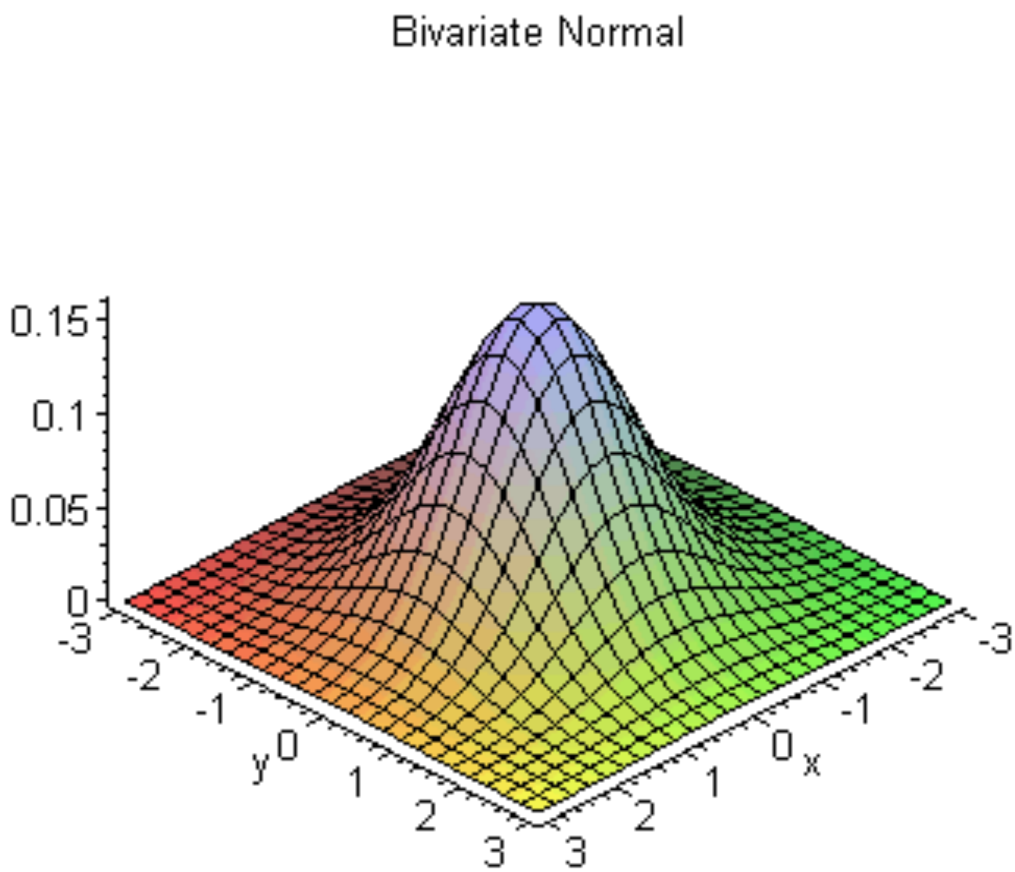
Note that if $(X_1, X_2, \ldots, X_n)$ is an n-dimensional discrete random vector, we can define probability distributions in exactly the same manner as in the 2-D case.

Definition: A random vector (X, Y) is called a *continuous* random vector if, for any set $C \subset \mathbb{R}^2$ the following holds:

$$\mathbb{P}((X,Y) \in C) = \int_C f(x,y)dxdy$$

for some non-negative function $f : \mathbb{R}^2 \to \mathbb{R}$.

The function $f(x, y)$ is called the *joint probability density* function of (X, Y).



The following relationship holds:

$$f(x,y) = \frac{\partial^2}{\partial x \partial y} F(x,y).$$

The p.d.f f_X of X can be obtained from $f(x,y)$ as

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

Similarly, we can obtain the p.d.f of Y by

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

Example. The joint density function of X and Y is given by

$$f(x,y) = \begin{cases} 2e^{-x}e^{-2y} & \text{for } 0 < x < \infty, 0 < y < \infty \\ 0 & \text{otherwise.} \end{cases}$$

Compute:

(a) $\mathbb{P}(X > 1, Y < 1)$.

(b) $\mathbb{P}(X < Y)$.

(c) $\mathbb{P}(X < a)$.

Solution: Show that $\mathbb{P}(X > 1, Y < 1) = e^{-1}(1 - e^{-2})$.

$$\begin{aligned} \mathbb{P}(X < Y) &= \int\int_{x<y} 2e^{-x}e^{-2y}dxdy = \int_0^{\infty}\int_0^{y} 2e^{-x}e^{-2y}dxdy = \int_0^{\infty} 2e^{-2y}\left[-e^{-x}\right]_0^y dy \\ &= \int_0^{\infty} 2e^{-2y}(1-e^{-y})dy = \int_0^{\infty} 2e^{-2y}dy - \int_0^{\infty} 2e^{-3y} = \frac{1}{3}. \end{aligned}$$

Show that for $a > 0$, $\mathbb{P}(X < a) = 1 - e^{-a}$.

Example: The joint density function of X and Y is given by

$$f(x,y) = \begin{cases} e^{-(x+y)} & \text{for } 0 < x < \infty, 0 < y < \infty \\ 0 & \text{otherwise.} \end{cases}$$

Find the density function of the random variable X/Y.

Solution:

$$\begin{aligned} F_{X/Y}(a) &= \mathbb{P}(X/Y \le a) = \mathbb{P}(X \le aY) = \int\int_{x<ay} e^{-(x+y)}dxdy = \int_0^{\infty}\int_0^{ay} e^{-(x+y)}dxdy \\ &= \int_0^{\infty} e^{-y}\left[e^{-x}\right]_0^{ay} dy = \int_0^{\infty} e^{-y}(1-e^{-ay})dy = 1 - \frac{1}{a+1}. \end{aligned}$$

Hence, $f_{X/Y}(a) = \frac{d}{da}F_{X/Y}(a) = \frac{1}{(1+a)^2}$, for $0 < a < \infty$.

1.4 Independent random variables

Definition: Two random variables X and Y are said to be *independent* if for any two sets of real numbers A and B,

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

If X and Y are discrete, then they are independent if and only if

$$p_{XY}(x,y) = p_X(x)p_Y(y) \quad \text{for all } x, y \in \mathbb{R},$$

where p_X and p_Y are the marginal p.m.f's of X and Y respectively.

Interpretation: Note the close resemblance between independence of random variables and events. Suppose that X and Y have a discrete joint distribution. If for each y learning that $Y = y$ does not change any of the probabilities of the events $\{X = x\}$, we would like to say that X and Y are independent. Indeed, if X and Y are independent then for y such that $\mathbb{P}(Y = y) > 0$, we have

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

for every x, i.e., learning about the value of Y does not change any of the probabilities associated with X.

Similarly, if X and Y are continuous, they are independent if and only if

$$f_{XY}(x,y) = f_X(x)f_Y(y) \quad \text{for all } x, y \in \mathbb{R}.$$

If you are given the joint p.d.f of the random variables X and Y, you can determine whether or not they are independent by calculating the marginal p.d.f's of X and Y and determining whether or not the above relationship holds.

Proposition: Suppose that (X, Y) have a joint p.d.f f_{XY}. Then the continuous random variables X and Y are independent if and only if their joint p.d.f can be expressed as

$$f_{XY}(x,y) = h(x)g(y), \qquad \text{for } x, y \in \mathbb{R}, \tag{1}$$

for some non-negative functions h and g.

The same result holds for discrete random variables.

Proof. X and Y independent implies that $f_{XY}(x,y) = h(x)g(y)$, for $x, y \in \mathbb{R}$ where we can take $h = f_X$ and $g = f_Y$.

Now, suppose that (1) holds for some h, g. Define $C = \int_{-\infty}^{\infty} h(x)dx$ and $D = \int_{-\infty}^{\infty} g(y)dy$. Then

$$CD = \int_{-\infty}^{\infty} h(x)dx \int_{-\infty}^{\infty} g(y)dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x)g(y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)dxdy = 1.$$

Furthermore,

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{XY}(x,y)dy = \int_{-\infty}^{\infty} h(x)g(y)dy = h(x)D \\ f_Y(y) &= \int_{-\infty}^{\infty} f_{XY}(x,y)dx = \int_{-\infty}^{\infty} h(x)g(y)dx = g(y)C \end{aligned}$$

Hence $f_X(x)f_Y(y) = h(x)Dg(y)C = h(x)g(y) = f_{XY}(x,y)$ and thus X and Y are independent. $\square$

Indicator function:

$$I_{(a,b)}(x) = \begin{cases} 1 & \text{if } a < x < b \\ 0 & \text{otherwise.} \end{cases}$$

Example: The joint p.d.f of (X, Y) is

$$f(x,y) = \begin{cases} 6e^{-2x}e^{-3y} & \text{for } 0 < x < \infty, 0 < y < \infty \\ 0 & \text{otherwise.} \end{cases}$$

Are X and Y independent?

Solution: Note that

$$f(x,y) = 6e^{-2x}I_{(0,\infty)}(x)e^{-3y}I_{(0,\infty)}(y) = h(x)g(y),$$

where we take $h(x) = 6e^{-2x}I_{(0,\infty)}(x)$ and $g(y) = e^{-3y}I_{(0,\infty)}(y)$. Thus, X and Y are independent.

Example: The joint p.d.f of (X, Y) is

$$f(x,y) = \begin{cases} 24xy & \text{for } 0 < x < 1, 0 < y < 1, 0 < x + y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Are X and Y independent?

Solution: Note that $f(x,y) = 24xy\ I_{(0,1)}(x)I_{(0,1)}(y)I_{(0,1)}(x+y)$. Thus, X and Y are **not** independent.

Example: A man and woman decide to meet at a certain location. If each person independently arrives at a time uniformly distributed between 2 and 3 PM, find the probability that the man arrives at least 10 minutes before the woman.

Solution: Let $X = \#$ minutes past 2 the man arrives, and let $Y = \#$ minutes past 2 the woman arrives. X and Y are independent random variables each uniformly distributed over $(0, 60)$. We want to find the probability $\mathbb{P}(X + 10 < Y)$.

$$\begin{aligned}\mathbb{P}(X+10<Y) = \int\int_{x+10<y} f(x,y)dxdy &= \int_{10}^{60}\int_{0}^{y-10}\left(\frac{1}{60}\right)^2 dxdy \\ = \left(\frac{1}{60}\right)^2\int_{10}^{60}(y-10)dy &= \left(\frac{1}{60}\right)^2\left[\frac{y^2}{2}-10y\right]_{10}^{60} = \frac{1250}{3600}.\end{aligned}$$

1.4.1 Sums of independent random variables

It is often important to calculate the distribution of $X + Y$ when X and Y are independent.

Theorem 1.3. *If X and Y are independent continuous random variables with p.d.f's f_X and f_Y then the p.d.f of $Z = X + Y$ is*

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y)f_Y(y)dy.$$

Proof. Observe that

$$\begin{aligned}F_Z(z) &= \mathbb{P}(Z \le z) = \mathbb{P}(X+Y \le z) = \int_{x+y\le z} f_X(x)f_Y(y)dxdy \\ &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{z-y} f_X(x)dx\right)f_Y(y)dy = \int_{-\infty}^{\infty} F_X(z-y)f_Y(y)dy.\end{aligned}$$

Now,

$$f_Z(z) = \frac{d}{dz}\int_{-\infty}^{\infty} F_X(z-y)f_Y(y)dy = \int_{-\infty}^{\infty}\frac{d}{dz}F_X(z-y)f_Y(y)dy = \int_{-\infty}^{\infty} f_X(z-y)f_Y(y)dy.$$

$\square$

Convolution: f_Z is usually denoted as the convolution of f_X and f_Y and represented as $f_Z = f_X \star f_Y$.

Example: X and Y are independent uniform r.v.'s on $[0,1]$. What is the distribution of $Z = X + Y$?

Solution: $f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y)f_Y(y)dy = \int_{-\infty}^{\infty} I_{(0,1)}(z-y)I_{(0,1)}(y)dy = \int_0^1 I_{(0,1)}(z-y)dy = (1-|z-1|)$.

Note that we started with two discontinuous functions f_X and f_Y and ended up with a continuous function f_Z.

In general, convolution is a smoothing operation: if $h = f \star g$, then h is at least as differentiable as f or g.

1.5 Conditional probability

Often when two random variables, (X, Y), are observed, their values are related.

Example: Let X= a person's height, and Y= the person's weight.

If we are told that $X = 73$ inches the likelihood that $Y > 200$ pounds is greater then if we are told that $X = 41$ inches.

1.5.1 Discrete case

Suppose that (X, Y) is a jointly distributed discrete random vector. Given that $Y = y$, X still has a distribution, which is called the conditional distribution of X given $Y = y$.

Definition: Let X and Y be discrete random variables, with joint a p.m.f. $p_{XY}(\cdot,\cdot)$ and marginal p.m.f's $p_X(\cdot)$ and $p_Y(\cdot)$. For any y such that $\mathbb{P}(Y = y) > 0$, the *conditional probability mass function* of X given $Y = y$, $p_{X|Y}(\cdot|y)$, is defined by

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

Similarly, the conditional mass function of Y given $X = x$, written $p_{Y|X}(\cdot|x)$, is defined by

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

Note that for each y, $p_{X|Y}(\cdot|y)$ is a valid p.m.f over x, i.e.,

- $p_{X|Y}(x|y) \geq 0$

- $\sum_x p_{X|Y}(x|y) = \sum_x \frac{p_{XY}(x,y)}{p_Y(y)} = \frac{1}{p_Y(y)} \sum_x p_{XY}(x,y) = \frac{1}{p_Y(y)} p_Y(y) = 1.$

If X and Y are independent then the following relationship holds:

$$p_{X|Y}(x|y) = \frac{p_{XY}(x,y)}{p_Y(y)} = \frac{p_X(x)p_Y(y)}{p_Y(y)} = p_X(x).$$

Definition: The *conditional distribution function* of X given $Y = y$, written $F_{X|Y}(\cdot|y)$, is defined by

$$F_{X|Y}(x|y) = \mathbb{P}(X \le x | Y = y)$$

for any y such that $\mathbb{P}(Y = y) > 0$.

Example: Suppose the joint p.m.f of (X, Y) is given by

$$p(0,0) = 0.4, \quad p(0,1) = 0.2, \quad p(1,0) = 0.1, \quad p(1,1) = 0.3.$$

Calculate the conditional p.m.f of X, given that $Y = 1$, i.e., find $p_{X|Y}(x|1)$, for $x = 0$ and 1.

1.5.2 Continuous case

Definition: Let X and Y be continuous random variables with joint density $f_{XY}(\cdot,\cdot)$ and marginal p.d.f's $f_X(\cdot)$ and $f_Y(\cdot)$. For any y such that $f_Y(y) > 0$, the *conditional p.d.f* of X given $Y = y$, written $f_{X|Y}(\cdot|y)$, is defined by

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

Similarly, the conditional p.d.f of Y given $X = x$, written $f_{Y|X}(\cdot|x)$, is defined by

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

Note that $f_{X|Y}(\cdot|y)$ is a valid p.d.f. (Verify at home)

If X and Y are independent then the following relationship holds:

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

Exercise: Suppose (X, Y) has joint density

$$f(x,y) = \begin{cases} \frac{21}{4}x^2 y & x^2 \le y < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Compute $f_{X|Y}(x|y)$ and $\mathbb{P}(X \geq Y|Y=y)$.

Solution: We compute first $f_Y(y) = \frac{7}{2}y^{5/2}$, for $y \in [0,1]$. Then,

$$f_{X|Y}(x|y) = \frac{3}{2}x^2 y^{-3/2},$$

where $-\sqrt{y} \leq x \leq \sqrt{y}$. Note that

$$\mathbb{P}(X \geq Y|Y=y) = \int_y^\infty f_{X|Y}(x|y)dx = \frac{1}{2}(1-y^{3/2}).$$

1.6 The distribution of a function of a random variable

Often we know the probability distribution of a random variable and are interested in determining the distribution of some function of it.

In the discrete case, this is pretty straightforward. Let us say g sends X to one of m possible discrete values. (Note that $g(X)$ must be discrete whenever X is). To get $p_g(y)$, the p.m.f of g at y, we just look at the p.m.f at any values of X that g mapped to y. More mathematically,

$$p_g(y) = \sum_{x \in g^{-1}(y)} p_X(x).$$

Here the "inverse image" $g^{-1}(A) = \{x : g(x) \in A\}$.

Things are slightly more subtle in the continuous case.

Example: If X is a continuous random variable with probability density f_X, then the distribution of $Y = cX$, $c > 0$, is obtained as follows: For $y \in \mathbb{R}$,

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(cX \leq y) = \mathbb{P}(X \leq y/c) = F_X(y/c).$$

Differentiation yields:

$$f_Y(y) = \frac{dF_X(y/c)}{dy} = \frac{1}{c}f_X(y/c).$$

Where did this extra factor of $1/c$ come from? We have stretched out the x-axis by a factor of c, which means that we have to divide everything by c to make sure that f_Y integrates to one. (This is similar to the usual change-of-variables trick for integrals from calculus.)

Example: If X is a continuous random variable with probability density f_X, then the distribution of $Y = X^2$ is obtained as follows: For $y \geq 0$,

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X^2 \leq y) = \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}).$$

Differentiation yields:

$$f_Y(y) = \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(-\sqrt{y}) = \frac{1}{2\sqrt{y}} \{f_X(\sqrt{y}) + f_X(-\sqrt{y})\}.$$

Theorem 1.4. *Let X be a continuous random variable having probability density function f_X. Suppose that $g(X)$ is a strictly monotone (increasing or decreasing), differentiable function of x. Then the random variable $Y = g(X)$ has a probability density function given by*

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| & \text{if } y = g(x) \text{ for some } x \\ 0 & \text{if } y \neq g(x) \text{ for all } x, \end{cases}$$

where $g^{-1}(y)$ is defined to be equal to that value of x such that $g(x) = y$.

Proof. If g is increasing,

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \leq g^{-1}(y)) = F_X(g^{-1}(y)).$$

Thus,

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{dF_X(g^{-1}(y))}{dy} = f_X(g^{-1}(y)) \frac{dg^{-1}(y)}{dy}.$$

If g is decreasing,

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y)).$$

Thus,

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{d[1 - F_X(g^{-1}(y))]}{dy} = -f_X(g^{-1}(y)) \frac{dg^{-1}(y)}{dy}.$$

$\square$

Example: Let X be a continuous non-negative random variable with density function f, and let $Y = X^n$. Find f_Y, the probability density function of Y.

Solution: In this case $g(x) = x^n$, so $g^{-1}(y) = y^{1/n}$. Then, $\frac{d}{dy} g^{-1}(y) = \frac{1}{n} y^{1/n-1}$.

Hence, $f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| = \frac{1}{n} y^{1/n-1} f_X(y^{1/n})$.

More generally, if g is non-monotonic, we have to form a sum over all of the points x such that $g(x) = y$, that is,

$$f_Y(y) = \begin{cases} \sum_{x:g(x)=y} f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(g(x)) \right| & \text{if } y = g(x) \text{ for some } x \\ 0 & \text{if } y \neq g(x) \text{ for all } x, \end{cases}$$

Compare this with the example $Y = X^2$ above.

1.6.1 The probability integral transform

Let X be a continuous random variable with c.d.f F and let $Y = F(X)$. This transformation from X to Y is called the *probability integral transformation.*

Theorem 1.5. *The distribution of Y is uniform on the interval $[0,1]$.*

Proof. As $F(x) \in [0,1]$ for all $x \in \mathbb{R}$, we must have $\mathbb{P}(Y < 0) = 0 = \mathbb{P}(Y > 1)$. Suppose that F is strictly increasing. Then,

$$\mathbb{P}(Y \leq u) = \mathbb{P}(F(X) \leq u) = \mathbb{P}(X \leq F^{-1}(u)) = F(F^{-1}(u)) = u.$$

The proof for the general case is given in the text book. $\square$

Corollary 1.6. *Let Y have a uniform distribution on the interval $[0,1]$, and let F be a continuous c.d.f with quantile function F^{-1}. Then $X = F^{-1}(Y)$ has c.d.f F.*

Example: Suppose that we want to generate a random variable Y with p.d.f $f(x) = e^{-x} I(0 < x < \infty)$.

Solution: Show that the c.d.f F of Y is given by

$$F(x) = \begin{cases} 1 - e^{-x} & \text{for } x > 0, \\ 0 & \text{for } x \leq 0. \end{cases}$$

Then, $F^{-1} : (0,1) \to \mathbb{R}$ is defined as

$$F^{-1}(u) = -\log(1-u) \qquad \text{for } u \in (0,1).$$

Thus, if we have access to a random variable Y drawn uniformly in the interval $(0,1)$, then $F^{-1}(Y)$ is a random variable that has F as its c.d.f.

1.7 Joint probability distribution of functions of random variables

Let X_1 and X_2 be continuous random variables with joint density $f_{X_1 X_2}(\cdot,\cdot)$. Assume that there is a subset $S \subset \mathbb{R}^2$ such that $\mathbb{P}((X_1, X_2) \in S) = 1$. Suppose that $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$ for some functions g_1 and g_2 which satisfy the following conditions:

(1) The equations $y_1 = g_1(x_1, x_2)$ and $y_2 = g_2(x_1, x_2)$ can be uniquely solved for x_1 and x_2 (in S) in terms of y_1 and y_2 with solutions given by $x_1 = h_1(y_1, y_2)$ and $x_2 = h_2(y_1, y_2)$.

(2) The functions g_1 and g_2 have continuous partial derivatives at all points $(x_1, x_2) \in S$ and are such that:

$$J = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{vmatrix} = \frac{\partial g_1}{\partial x_1}\frac{\partial g_2}{\partial x_2} - \frac{\partial g_2}{\partial x_1}\frac{\partial g_1}{\partial x_2} \neq 0,$$

holds at all points $(x_1, x_2) \in S$. Let $T = (g_1, g_2)(S) \subset \mathbb{R}^2$ be the image of (g_1, g_2).

Under these conditions the random variables Y_1 and Y_2 have the following joint p.d.f:

$$f_{Y_1Y_2}(y_1, y_2) = f_{X_1X_2}(h_1(y_1, y_2), h_2(y_1, y_2))|J(h_1(y_1, y_2), h_2(y_1, y_2))|^{-1}, \quad (y_1, y_2) \in T.$$

Example: Let X_1 and X_2 be continuous distributed random variables with joint density $f_{X_1X_2}(\cdot, \cdot)$. Let $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$. Find the joint p.d.f of (Y_1, Y_2).

Solution: Let $y_1 = g_1(x_1, x_2) = x_1 + x_2$ and $y_2 = g_2(x_1, x_2) = x_1 - x_2$. Then, $J(x_1, x_2) = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2$. Now, $x_1 = h_1(y_1, y_2) = (y_1 + y_2)/2$ and $x_2 = h_2(y_1, y_2) = (y_1 - y_2)/2$. Thus,

$$f_{Y_1Y_2}(y_1, y_2) = \frac{1}{2} f_{X_1X_2}\left(\frac{y_1 + y_2}{2}, \frac{y_1 - y_2}{2}\right).$$

Assume that X_1 and X_2 are independent Uniform$(0, 1)$ random variables. Then,

$$f_{Y_1Y_2}(y_1, y_2) = \frac{1}{2} f_{X_1}\left(\frac{y_1 + y_2}{2}\right) f_{X_2}\left(\frac{y_1 - y_2}{2}\right) = \frac{1}{2} I(0 < y_1 + y_2 < 2) I(0 < y_1 - y_2 < 2).$$

Example: Suppose that (X_1, X_2) has the joint density

$$f_{X_1X_2}(x_1, x_2) = \lambda e^{-\lambda x_1} \lambda e^{-\lambda x_2} I(x_1 > 0, x_2 > 0).$$

Let $(Y_1, Y_2) = (g_1(X_1, X_2), g_2(X_1, X_2)) = (X_1 + X_2, \frac{X_1}{X_1 + X_2})$. Find the joint p.d.f of (Y_1, Y_2).

Solution: Let $S = \{(x_1, x_2) : x_1 > 0, x_2 > 0\}$. Observe that (Y_1, Y_2) takes values in $T = \{(y_1, y_2) : y_1 > 0, y_2 \in (0, 1)\}$. Computing the partial derivatives of g we have,

$$\frac{\partial g_1}{\partial x_1} = 1, \qquad \frac{\partial g_2}{\partial x_1} = \frac{x_2}{(x_1 + x_2)^2},$$

and

$$\frac{\partial g_1}{\partial x_2} = 1, \qquad \frac{\partial g_2}{\partial x_2} = \frac{x_1}{(x_1 + x_2)^2}.$$

Thus, the functions g_1 and g_2 have continuous partial derivatives at all points (x_1, x_2); also g is clearly a one-to-one function on S. Furthermore,

$$J(x_1, x_2) = \left| -\frac{x_1 + x_2}{(x_1 + x_2)^2} \right| = \frac{1}{x_1 + x_2} > 0$$

for every $(x_1, x_2) \in S$.

The inverse transformation is

$$h_1(y_1, y_2) = y_1 y_2, \qquad h_2(y_1, y_2) = y_1 - y_1 y_2.$$

Thus the density of (Y_1, Y_2) at the point (y_1, y_2) is

$$\begin{aligned} f_Y(y_1, y_2) &= f_{X_1 X_2}(h_1(y_1, y_2), h_2(y_1, y_2))|J(h_1(y_1, y_2), h_2(y_1, y_2))|^{-1} \\ &= \lambda^2 e^{-\lambda(h_1(y_1,y_2)+h_2(y_1,y_2))}|h_1(y_1, y_2) + h_2(y_1, y_2)| \\ &= \lambda^2 e^{-\lambda y_1} y_1. \end{aligned}$$

Thus, we know that Y_1 and Y_2 are independent and that Y_2 has a uniform distribution on $(0, 1)$.

Theorem 1.7. *(Multivariate transformation) Let $(X_1, X_2, \ldots, X_n)$ have a continuous joint distribution for which the joint pd.f if f. Assume that there is a subset $S \subset \mathbb{R}^n$ such that $\mathbb{P}((X_1, \ldots, X_n) \in S) = 1$. Define n new random variables $Y_1, Y_2, \ldots, Y_n$ as follows:*

$$\begin{aligned} Y_1 &= r_1(X_1, \ldots, X_n), \\ Y_2 &= r_2(X_1, \ldots, X_n), \\ &\vdots \\ Y_n &= r_n(X_1, \ldots, X_n), \end{aligned}$$

where we assume that the n functions $r_1, r_2, \ldots, r_n$ define one-to-one differentiable transformations of S onto a subset $T \subset \mathbb{R}^n$. Let the inverse of this transformation be given as follows:

$$\begin{aligned} x_1 &= s_1(y_1, \ldots, y_n), \\ x_2 &= s_2(y_1, \ldots, y_n), \\ &\vdots \\ x_n &= s_n(y_1, \ldots, y_n). \end{aligned}$$

Then the joint p.d.f g of $(Y_1, \ldots, Y_n)$ is

$$g(y_1, y_2, \ldots, y_n) = \begin{cases} f(s_1, \ldots, s_n)|J| & \text{for } (y_1, \ldots, y_n) \in T, \\ 0 & \text{otherwise,} \end{cases}$$

where $J = \begin{vmatrix} \frac{\partial s_1}{\partial y_1} & \cdots & \frac{\partial s_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial s_n}{\partial y_1} & \cdots & \frac{\partial s_n}{\partial y_n} \end{vmatrix}$ *is the determinant.*

Theorem 1.8. *(Linear transformations) Let* $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ *have a continuous joint distribution for which the joint p.d.f is* f*. Define* $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$ *by*

$$\mathbf{Y} = A\mathbf{X},$$

where A *is a nonsingular* $n \times n$ *matrix. Then* $\mathbf{Y}$ *has a continuous joint distribution with p.d.f*

$$g(\mathbf{y}) = \frac{1}{|\det A|} f(A^{-1}\mathbf{y}), \qquad \textit{for } \mathbf{y} \in \mathbb{R}^n,$$

where A^{-1} *is the inverse of* A*.*

1.8 Maximum and minimum of a random sample

Suppose that $X_1, X_2, \ldots, X_n$ form a random sample of size n from a distribution for which the p.d.f is f and the c.d.f is F. The largest value Y_n and the smallest value Y_1 in the random sample are defined as:

$$\begin{aligned} Y_n &= \max\{X_1, \ldots, X_n\} \\ Y_1 &= \min\{X_1, \ldots, X_n\}. \end{aligned}$$

Find the distributions of Y_n and Y_1.

Let $Y_n \sim G_n$ (i.e., G_n is the c.d.f of Y_n) and let g_n be its p.d.f. For any $y \in \mathbb{R}$,

$$\begin{aligned} G_n(y) &= \mathbb{P}(Y_n \le y) = \mathbb{P}(X_1 \le y, \ldots, X_n \le y) \\ &= \mathbb{P}(X_1 \le y) \cdot \mathbb{P}(X_2 \le y) \cdots \mathbb{P}(X_n \le y) \\ &= F(y) \cdot F(y) \cdots F(y) = [F(y)]^n, \end{aligned}$$

where the third equality follows from the fact that the X_i's are independent and the fourth follows from the fact that X_i's all have the same c.d.f F. Now, g_n can be determined by differentiating the c.d.f G_n. Thus, for $y \in \mathbb{R}$,

$$g_n(y) = n[F(y)]^{n-1} f(y).$$

Let $Y_1 \sim G_1$ and let g_1 be its p.d.f. For any $y \in \mathbb{R}$,

$$\begin{aligned} G_1(y) &= \mathbb{P}(Y_1 \le y) = 1 - \mathbb{P}(Y_1 > y) \\ &= 1 - \mathbb{P}(X_1 > y, \ldots, X_n > y) \\ &= 1 - \mathbb{P}(X_1 > y) \cdot \mathbb{P}(X_2 > y) \cdots \mathbb{P}(X_n > y) \\ &= 1 - [1 - F(y)]^n, \end{aligned}$$

Now, g_1 can be determined by differentiating the c.d.f G_1. Thus, for $y \in \mathbb{R}$,

$$g_1(y) = n[1 - F(y)]^{n-1} f(y).$$

Exercise: Let X and Y be independent random variables, both uniformly distributed on $[0, 1]$. Let $Z = \min(X, Y)$ be the smaller value of the two.

(a) Compute the density function of Z.

(b) Compute $\mathbb{P}(X \le 0.5 | Z \le 0.5)$.

(c) Are X and Z independent?

Solution: (a) For $z \in [0,1]$,

$$\mathbb{P}(Z \le z) = 1 - \mathbb{P}(\text{both } X \text{ and } Y \text{ are above } z) = 1 - (1-z)^2 = 2z - z^2,$$

so that $f_Z(z) = 2(1-z)$, for $z \in [0,1]$ and 0 otherwise.

(b) From (a), we conclude that $\mathbb{P}(Z \le 0.5) = \frac{3}{4}$ and $\mathbb{P}(X \le 0.5, Z \le 0.5) = \mathbb{P}(X \le 0.5) = \frac{1}{2}$, so the answer is $\frac{2}{3}$.

(c) No: $Z \le X$.