Chapter 2: A Review of Probability

Instructor: Sergi Quintana Garcia
Notes obtained from Manuel V. Montesinos (UAB and IDEA)

Statistics Brush-Up Course
Competition and EPP Master Programs

Fall 2022

Barcelona School of Economics

Introduction

Probability theory allows us to explain how data are generated from a population by means of statistical models.

The process that generated the data that we observe is a **random experiment** or **trial**. As opposed to a **deterministic experiment**, the result is not known with certainty (more than one possible outcome).

The mutually exclusive potential results of a random experiment are called **outcomes**.

The set of all possible outcomes is the **sample space**, and an **event** is a subset of it.

Introduction

Examples:

1. Tossing a coin: the result of this experiment is either “tails” (T) or “heads” (H) and thus the sample space is $\{H, T\}$.

2. Rolling a die: the result can be one of the numbers between 1 and 6.
   - A natural sample space is $\{1, 2, 3, 4, 5, 6\}$.
   - Another sample space could be $\{odd, even\}$.

3. Tossing a coin twice: there are two possible results in each round and thus four possible results overall. The sample space is $\{HH, HT, TH, TT\}$.

A sample space is **discrete (countable)** if there exists a one-to-one function from the sample space to the natural numbers. Otherwise, the sample space is **continuous**.

Events and Probabilities

Events and Probabilities

The **probability** of an event A is given by $\Pr(A)$, such that $0 \leq \Pr(A) \leq 1$.

The **union** of events A and B is denoted by $A \cup B$ (A **or** B happens).

The **intersection** of events A and B is denoted by $A \cap B$ (A **and** B happen).

Conditional probability:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} \text{ if } \Pr(B) > 0.$$

Rearranging, we get the multiplication rule:

$$\Pr(A \cap B) = \Pr(A|B)\Pr(B).$$

Events and Probabilities

The **law of total probability** tells us that if we partition the sample space into B_1 and B_2, then

$$\begin{aligned}\Pr(A) &= \Pr(A \cap B_1) + \Pr(A \cap B_2)\\ &= \Pr(A|B_1)\Pr(B_1) + \Pr(A|B_2)\Pr(B_2).\end{aligned}$$

Bayes' theorem tells us that if we partition the sample space into B_1 and B_2, then

$$\Pr(B_1|A) = \frac{\Pr(A \cap B_1)}{\Pr(A \cap B_1) + \Pr(A \cap B_2)}$$

In general, suppose that $A_1, A_2, \ldots, A_n$ are mutually exclusive events whose union if the sample space S. Then, if A is any event,

$$\Pr(A_k|A) = \frac{\Pr(A_k)\Pr(A|A_k)}{\sum_{j=1}^{n}\Pr(A_j)\Pr(A|A_j)}$$

Example

Suppose that, when rolling a die, there are two events, A and B, defined as:

- $A =$"get an odd number" ($A = \{1, 3, 5\}$)
- $B =$"get a number greater than 3" ($B = \{4, 5, 6\}$)

Then, the union of A and B, $A \cup B$ is the set $\{1, 3, 4, 5, 6\}$, as this is the set of either A or B. The intersection of A and B, given by $A \cap B$, is the set $\{5\}$.

The conditional probability of A given B is given by

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{1/6}{1/2} = \frac{1}{3}.$$

Exercise 1

There are 5 red and 2 green balls in an urn. A random ball is selected and replaced by a ball of the other color; then a second ball is drawn.

1. What is the probability that the second ball is red?
2. What is the probability that the first ball was red given the second ball was red?

Solution

1. We use the law of total probability to obtain $\Pr(R_2)$:

$$\begin{aligned}\Pr(R_2) &= \Pr(R_2|R_1)\Pr(R_1) + \Pr(R_2|G_1)\Pr(G_1)\\ &= \frac{4}{7}\frac{5}{7} + \frac{6}{7}\frac{2}{7} = \frac{32}{49}\end{aligned}$$

2. We need to obtain the following conditional probability:

$$\begin{aligned}\Pr(R_1|R_2) &= \frac{\Pr(R_1 \cap R_2)}{\Pr(R_2)} = \frac{\Pr(R_1)\Pr(R_2|R_1)}{\Pr(R_2)}\\ &= \frac{5/7 \times 4/7}{32/49} = \frac{20/49}{32/49} = \frac{20}{32}\end{aligned}$$

Random Variables and Probability Distributions

Random Variables and Probability Distributions

A **random variable** is a function which assigns to each sample point $s \in S$ a real number, the value of the random variable at s.

Random variables are **discrete** if they can take on a finite or countably infinite number of values (e.g. 0 or 1). Otherwise, they are **continuous**, and can take a continuous range of values (e.g. $[0, 1]$, $[0, \infty)$, $(-\infty, \infty)$).

Example

Suppose that we are tossing a coin twice and we have a sample space $S = \{HH, HT, TH, TT\}$. Now let's define the random variable X as the number of heads that can come up. Thus, with each sample point we can associate a number as shown in the table below.

Sample point	HH	HT	TH	TT
X	2	1	1	0

Let X be a discrete random variable and let the possible values that X takes be given by $x_1, x_2, ...$. Then, assume that the probabilities that X takes a particular value x is given by

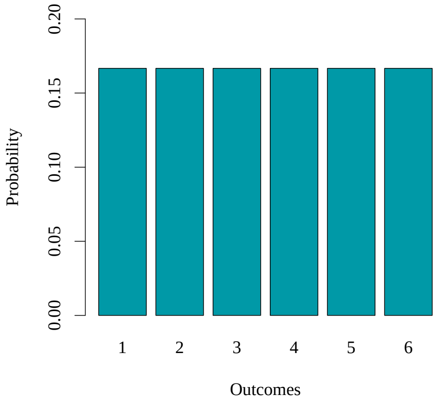
$$\Pr(X = x) = f(x).$$

This function is called the **probability (density) function (pdf)** of x if it satisfies the following properties:

1. $f(x) \geq 0$.
2. $\sum_x f(x) = 1$ (the sum is taken over all possible values of x).

Example

Figure 1: Probability Distribution Function of a Dice Roll



Exercise 2

Going back to the example of tossing a coin twice, find the probability function which corresponds to the random variable X, which represents the number of heads.

Solution

We assume that the coin is fair, meaning that the probability of obtaining heads or tails is 1/2. Thus, we can determine the probability of each of the sample points HH, HT, TH, TT.

First,

$$\begin{aligned}
\Pr(HH) &= 1/2 \times 1/2 = 1/4, \\
\Pr(HT) &= 1/2 \times 1/2 = 1/4, \\
\Pr(TH) &= 1/2 \times 1/2 = 1/4, \\
\Pr(TT) &= 1/2 \times 1/2 = 1/4.
\end{aligned}$$

Solution

Once we have the probabilities of each event, we can find the probability associated to each value of the random variable X:

$$\Pr(X = 0) = \Pr(TT) = 1/4,$$
$$\Pr(X = 1) = \Pr(HT) + \Pr(TH) = 1/4 + 1/4 = 1/2,$$
$$\Pr(X = 2) = \Pr(HH) = 1/4.$$

Then the probability function is given by

x	0	1	2
$f(x)$	$1/4$	$1/2$	$1/4$

The **cumulative distribution function (cdf)** for a random variable X is defined by

$$F(x) = \Pr(X \leq x),$$

where x is any real number i.e. $-\infty < x < \infty$.

Random Variables and Probability Distributions

Discrete random variables

The distribution function has the following properties:

1. $0 \leq F(x) \leq 1$.
2. $F(x)$ is non-decreasing: $F(x) \leq F(y)$ if $x \leq y$.
3. $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$.
4. $F(x)$ is continuous from the right: $\lim_{h \to 0^+} F(x+h) = F(x)\ \forall x$.
5. The pdf can be obtained from the cdf as $f(x) = F(x) - \lim_{u \to x^-} F(u)$.

Random Variables and Probability Distributions

Discrete random variables

The **distribution function for a discrete variable** X can be obtained from the probability function noting that for all x in $(-\infty, \infty)$

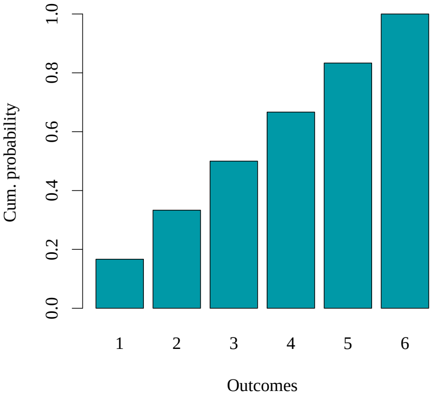
$$F(x) = \Pr(X \leq x) = \sum_{u \leq x} f(u).$$

Thus, if X takes a finite number of values x_1, …, x_n, the distribution function is given by

$$F(X) = \Pr(X \leq x) = \begin{cases} 0 & -\infty < x < x_1 \\ f(x_1) & x_1 \leq x < x_2 \\ f(x_1) + f(x_2) & x_2 \leq x < x_3 \\ \vdots & \vdots \\ f(x_1) + \ldots + f(x_n) & x_n \leq x < \infty \end{cases}$$

Example

Figure 2: Cumulative Distribution Function of a Dice Roll



Exercise 3

Going back to the example of tossing a coin twice, find the distribution function which corresponds to the random variable X, which represents the number of heads.

Solution

The random variable X takes on values 0, 1 and 2, so these will be the points at which we will see a jump in the distribution function:

$$F(X) = \begin{cases} 0 & -\infty < x < 0 \\ \frac{1}{4} & 0 \leq x < 1 \\ \frac{1}{4} + \frac{1}{2} = \frac{3}{4} & 1 \leq x < 2 \\ \frac{1}{4} + \frac{1}{2} + \frac{1}{4} = 1 & 2 \leq x < \infty. \end{cases}$$

Solution

- The magnitude of the jumps at 0, 1 and 2 are $\frac{1}{4}$, $\frac{1}{2}$ and $\frac{1}{4}$, which correspond to the probability function.

- The value of the function at an integer is obtained from the right step: the function is continuous from the right at 0, 1 and 2.

- From left to right, the function either stays the same or increases: it is a monotonically increasing function (even though not strictly).

The **probability density function (pdf)** of a **continuous random variable** X, $f(x)$ has the following properties:

1. $f(x) \geq 0$.
2. $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$.

The integral can equivalently be taken over the support of x, which is the range within which x takes values. For example, if $x \in [0, \infty)$ then $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = \int_{0}^{\infty} f(x)\mathrm{d}x$.

Random Variables and Probability Distributions

Continuous random variables

Note that **for continuous variables, the density is not the probability.** If X is a continuous random variable, the probability that X takes on a particular value is zero, but the interval probability that X lies between a and b is given by

$$\Pr(a < x < b) = \int_a^b f(x)\mathrm{d}x$$

The fact that the probability that X takes on a particular value is zero implies:

- $\Pr(a \leq x \leq a) = \Pr(a) = 0.$
- Does $\Pr(X = a) = 0$ mean X never equals a? NO! For a continuous variable, any single value has probability zero, and only a range of values has non-zero probability.

Exercise 4

Find the constant c such that the function

$$f(x) = \begin{cases} cx^2 & \text{if } 0 < x < 3 \\ 0 & \text{otherwise} \end{cases}$$

is a density function and compute $\Pr(1 < X < 2)$.

Solution

Condition 1 tells us that $f(x) \geq 0$. Then, for this function to be a density function, $c \geq 0$.

Condition 2 tells us that

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1 \Leftrightarrow \int_0^3 cx^2\mathrm{d}x = 1.$$

Therefore,

$$\int_0^3 cx^2\mathrm{d}x = \left[\frac{c}{3}x^3\right]_0^3 = \frac{c}{3}27 = 9c = 1 \Rightarrow c = \frac{1}{9}.$$

Random Variables and Probability Distributions

Continuous random variables

A **cumulative distribution function (cdf)** of a continuous variable is given by

$$F(x) = \Pr(X \leq x) = \int_{-\infty}^{x} f(u)\mathrm{d}u \text{ for } -\infty < x < \infty.$$

The properties of the cdf (the first three are the same as in the discrete case) are:

1. $0 \leq F(x) \leq 1$.
2. $F(x)$ is non-decreasing: $F(x) \leq F(y)$ if $x \leq y$.
3. Zero to the left: $\lim_{x\to-\infty} F(x) = 0$; one to the right: $\lim_{x\to\infty} F(x) = 1$.
4. The derivative of the cdf is the pdf: $\frac{\partial F(x)}{\partial x} = f(x)$.
5. $\Pr(c \leq x \leq d) = F(d) - F(c)$.

Exercise 5

Suppose Y has range $[0, b]$ and cumulative distribution function (cdf) $F(y) = \frac{y^2}{9}$.

1. What is b?
2. Find the pdf of y.

Solution

Since we know that the total probability must be 1, we know that $F(b) = \frac{b^2}{9} = 1$. Thus,

$$\begin{aligned} \frac{b^2}{9} &= 1 \\ b^2 &= 9 \\ b &= 3. \end{aligned}$$

We know that $f(y) = F'(y)$, so the pdf of y is

$$f(y) = F'(y) = \frac{2y}{9}.$$

Expected Value and Variance

Expected Value and Variance

The **expected value** of a random variable Y, denoted by $\mathbb{E}[Y]$ or μ_Y, is the long-run average value of the random variable over many repeated trials or occurrences.

For a **discrete random variable**, the expected value is computed as a weighted average of the possible outcomes of that random variable, where the weights are the probabilities of each outcome:

$$\mathbb{E}(X) = \sum_{i=1}^{n} p(x_i)x_i = p(x_1)x_1 + p(x_2)x_2 + \ldots + p(x_n)x_n.$$

Example

Let's compute the expected value of X, a discrete random variable:

x	-2	-1	0	1	2
$f(x)$	1/5	1/5	1/5	1/5	1/5

The expected value is the sum of the values of the random variable weighted by the probability of the corresponding outcome of the random variable:

$$\mathbb{E}(X) = -2 \times \frac{1}{5} - 1 \times \frac{1}{5} + 0 \times \frac{1}{5} + 1 \times \frac{1}{5} + 2 \times \frac{1}{5} = 0,$$

or noticing that each outcome occurs with equal probability,

$$\mathbb{E}(X) = \frac{-2 - 1 + 0 + 1 + 2}{5} = 0.$$

Expected Value and Variance

Properties of $\mathbb{E}(X)$:

1. $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$ where a and b are constants.
2. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$.
3. $\mathbb{E}(h(X)) = \sum_i h(x_i)p(x_i)$.

Example

Following the previous example, let's compute $\mathbb{E}(X^2)$. We can compute X^2 for each sample point of the random variable X:

x	-2	-1	0	1	2
$f(x)$	$1/5$	$1/5$	$1/5$	$1/5$	$1/5$
x^2	4	1	0	1	4

We can compute $\mathbb{E}(X^2)$ as the sum of X^2 weighted by the probabilities:

$$\mathbb{E}\left(X^2\right) = 4 \times \frac{1}{5} + 1 \times \frac{1}{5} + 0 \times \frac{1}{5} + 1 \times \frac{1}{5} + 4 \times \frac{1}{5} = 2,$$

which coincides with

$$\mathbb{E}\left(x^2\right) = \sum_i x_i^2 p(x_i).$$

Expected Value and Variance

The **variance** of a random variable X, denoted by $\mathrm{Var}[X]$ or σ_X^2, is the expected value of the square of the deviation of X from its mean:

$$\mathrm{Var}(X) = \mathbb{E}\left[(X-\mu)^2\right] = \sum_{i=1}^{n}(x_i-\mu)^2 p(x_i).$$

The **standard deviation** of a random variable X is

$$\sigma = \sqrt{\mathrm{Var}(X)},$$

which is measured in the same units as X.

Example

Similarly to the expected value, we can compute the variance from a table:

x	1	2	3	4	5
$f(x)$	$1/10$	$2/10$	$4/10$	$2/10$	$1/10$

First, we need to compute the mean

$$\begin{aligned}\mathbb{E}(X) = \mu &= 1 \times \frac{1}{10} + 2 \times \frac{2}{10} + 3 \times \frac{4}{10} + 4 \times \frac{2}{10} + 5 \times \frac{1}{10} \\ &= \frac{1+4+12+8+5}{10} = 3.\end{aligned}$$

Example

Now let's compute $(X-\mu)^2$ for each value of X add it to the table:

x	1	2	3	4	5
$f(x)$	$1/10$	$2/10$	$4/10$	$2/10$	$1/10$
$(X-\mu)^2$	4	1	0	1	4

Thus we can compute $\mathbb{E}\left[(X-\mu)^2\right]$ from the table as follows:

$$\begin{aligned}\mathbb{E}\left[(X-\mu)^2\right] &= 4\times\frac{1}{10}+1\times\frac{2}{10}+0\times\frac{4}{10}+1\times\frac{2}{10}+4\times\frac{1}{10}\\ &= \frac{4+2+2+4}{10} = 1.2,\end{aligned}$$

and the standard deviation is

$$\sigma = \sqrt{1.2}.$$

Expected Value and Variance

Properties of the variance (for discrete and continuous random variables):

1. $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$.
2. If X is constant then $\mathrm{Var}(X) = 0$. Note that if X takes two different values with positive probability, then the variance will be a sum of two positive terms, so it will not be zero.
3. $\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.

Expected Value and Variance

If X is **continuous** with range $[a, b]$ and pdf $f(x)$, then

$$\mathbb{E}(X) = \int_a^b x f(x) \mathrm{d}x,$$
$$\mathrm{Var}(X) = \int_a^b (x - \mu)^2 f(x) \mathrm{d}x.$$

Exercise 6

Suppose that X is distributed in the range $[0, 1]$ and its pdf is given by $3x^2$. Find the mean and the variance.

Solution

Mean:

$$\mathbb{E}(X) = \int_0^1 x3x^2 \mathrm{d}x = \int_0^1 3x^3 = \left[\frac{3}{4}x^4\right]_0^1 = \frac{3}{4}.$$

Variance (long way):

$$\begin{aligned}
\mathrm{Var}(X) &= \int_0^1 (x - 3/4)^2 3x^2 \mathrm{d}x = \int_0^1 \left(x^2 - \frac{6}{4}x + \frac{9}{16}\right) 3x^2 \mathrm{d}x \\
&= \int_0^1 \left(3x^4 - \frac{18}{4}x^3 + \frac{27}{16}x^2\right) \mathrm{d}x \\
&= \left[\frac{3}{5}x^5 - \frac{18}{16}x^4 + \frac{27}{16 \times 3}x^3\right]_0^1 \\
&= \frac{3}{5} - \frac{18}{16} + \frac{27}{16 \times 3} = \frac{3}{5} - \frac{9}{8} + \frac{9}{16} = \frac{3}{80}.
\end{aligned}$$

Solution

Variance ("lazy way"):

Use $\mathbb{E}(X^2) - (\mathbb{E}(X))^2$. We already know $\mathbb{E}(X)$ so we just need to calculate $\mathbb{E}(X^2)$:

$$\mathbb{E}(X^2) = \int_0^1 x^2 3x^2 \mathrm{d}x = \int_0^1 3x^4 \mathrm{d}x = \left[\frac{3}{5}x^5\right]_0^1 = \frac{3}{5}.$$

Thus,

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \frac{3}{5} - \left(\frac{3}{4}\right)^2 = \frac{3}{5} - \frac{9}{16} = \frac{3}{80}.$$

Two Random Variables

Two Random Variables

Joint and marginal distributions of discrete random variables

Suppose that X takes any of m values $x_1, x_2, ..., x_m$, and Y takes any of n values $y_1, y_2, ..., y_n$. Then, the probability of the event that $X = x_j$ and $Y = y_k$ is given by

$$\Pr(X = x_j, Y = y_k) = f(x_j, y_k),$$

where $f(x, y)$ is called the **joint probability function** for X and Y and can be represented by a joint probability table.

Properties of the joint pdf:

1. $f(x, y) \geq 0$.
2. $\sum_{j=1}^{m} \sum_{k=1}^{n} f(x_j, y_k) = 1$ (the sum is over all the values of x and y).

Exercise 7

The joint pdf of random variables X and Y is given by

$$f(x,y) = c(2x+y)$$

and X and Y can take integers as follows: $0 \le X \le 2$ and $0 \le Y \le 3$. Find c.

Solution

We can tabulate the values that the random variables X and Y take, and for each point (x, y) substitute the values into the density function:

$X \setminus Y$	0	1	2	3
0	0	c	$2c$	$3c$
1	$2c$	$3c$	$4c$	$5c$
2	$4c$	$5c$	$6c$	$7c$

From the second property of the joint pdf, we know that the sum of the probabilities over all values of x and y is equal to 1. Thus, we can sum up all the entries of the table to obtain c, as follows:

$$c(1 + 2 + 3 + 2 + 3 + 4 + 5 + 4 + 5 + 6 + 7) = 1$$

$$42c = 1 \Rightarrow c = \frac{1}{42}.$$

Two Random Variables

Joint and marginal distributions of discrete random variables

The probability that $X = x_j$ is obtained by adding all entries in the row corresponding to x_j and is given by (sum over all values of y):

$$\Pr(X = x_j) = f_1(x_j) = \sum_{k=1}^{n} f(x_j, y_k)$$

This is a **marginal probability function** for X. The same can be obtained for Y by summing the density function $f(x, y)$ over all the values of x for each value of y.

The **joint distribution function** is given by

$$F(x, y) = \Pr(X \le x, Y \le y) = \sum_{u \le x} \sum_{v \le y} f(u, v).$$

Exercise 8

Find the marginal probability functions of X and Y using the joint probability function given in Exercise 7.

Solution

We can sum the probabilities for each row and column, corresponding to the probability of $X = x$ and $Y = y$, respectively:

$X \setminus Y$	0	1	2	3	Total
0	0	c	$2c$	$3c$	$6c$
1	$2c$	$3c$	$4c$	$5c$	$14c$
2	$4c$	$5c$	$6c$	$7c$	$22c$
Total	$6c$	$9c$	$12c$	$15c$	$42c$

Then, the marginal densities of x and y are given by

$$f_1(x) = \begin{cases} 6c & \text{if } x = 0 \\ 14c & \text{if } x = 1 \\ 22c & \text{if } x = 2 \end{cases} \qquad f_2(y) = \begin{cases} 6c & \text{if } y = 0 \\ 9c & \text{if } y = 1 \\ 12c & \text{if } y = 2 \\ 15c & \text{if } y = 3 \end{cases}$$

Solution

We can also check that each of the marginal densities sums up to one:

$$\sum_{j=1}^{m} f_1(x_j) = 1, \sum_{k=1}^{n} f_2(y_k) = 1,$$

Two Random Variables

Joint and marginal distributions of continuous random variables

Here we simply replace the sums by integrals and obtain the **joint density function** $f(x, y)$, which satisfies:

1. $f(x, y) \geq 0$.
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \mathrm{d}y \mathrm{d}x = 1$.

Then, the probability that X lies between a and b, and Y lies between c and d is given by

$$\Pr(a < X < b, c < Y < d) = \int_{x=a}^{b} \int_{y=c}^{d} f(x, y) \mathrm{d}y \mathrm{d}x,$$

and the **marginal density functions** are given by

$$f_1(x) = \int_{v=-\infty}^{\infty} f(x, v) \mathrm{d}v \quad \text{and} \quad f_2(y) = \int_{u=-\infty}^{\infty} f(u, y) \mathrm{d}u.$$

Exercise 9

The joint density of two continuous variables X and Y is

$$f(x, y) = \begin{cases} cxy & \text{if } 0 < x < 4, 1 < y < 5 \\ 0 & \text{otherwise} \end{cases}$$

1. Find the value of c.
2. Compute $\Pr(1 < X < 2, 2 < Y < 3)$.

Solution

We know that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\mathrm{d}x\mathrm{d}y = 1$. Then we can write

$$\int_{y=1}^{5}\int_{x=0}^{4} cxy\mathrm{d}x\mathrm{d}y = 1.$$

Let's compute the integral

$$\begin{aligned}
\int_{y=1}^{5}\int_{x=0}^{4} cxy\mathrm{d}x\mathrm{d}y &= \int_{y=1}^{5}\left[\frac{c}{2}x^2y\right]_0^4\mathrm{d}y = \int_{y=1}^{5}\frac{c}{2}16y\mathrm{d}y \\
&= \int_{y=1}^{5} 8cy\mathrm{d}y = \left[\frac{8c}{2}y^2\right]_1^5 = \left[4cy^2\right]_1^5 \\
&= 4c25 - 4c = c(100-4) = 96c \\
1 = 96c &\Rightarrow c = \frac{1}{96}.
\end{aligned}$$

Solution

Now, we have to calculate $\Pr(1 < X < 2, 2 < Y < 3)$:

$$\begin{aligned}
\Pr(1 < X < 2, 2 < Y < 3) &= \int_{y=2}^{3}\int_{x=1}^{2} \frac{xy}{96}\mathrm{d}x\mathrm{d}y \\
&= \int_{y=2}^{3} \left[\frac{y}{96\times 2}x^2\right]_1^2 \mathrm{d}y \\
&= \int_{y=2}^{3} \left(\frac{4y}{96\times 2} - \frac{y}{96\times 2}\right) \mathrm{d}y \\
&= \int_{y=2}^{3} \left(\frac{3y}{96\times 2}\right) \mathrm{d}y = \int_{y=2}^{3} \left(\frac{y}{64}\right) \mathrm{d}y \\
&= \left[\frac{y^2}{64\times 2}\right]_2^3 = \frac{9-4}{128} = \frac{5}{128}.
\end{aligned}$$

Two Random Variables

Joint and marginal distributions of continuous random variables

The **joint distribution function** is given by

$$F(x,y) = \Pr(X \leq x, Y \leq y) = \int_{u=-\infty}^{x} \int_{v=-\infty}^{y} f(u,v)\mathrm{d}v\mathrm{d}u.$$

Thus, we have again the case that

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

The **marginal distribution functions** for X and Y are given by:

$$\begin{aligned}\Pr(X \leq x) =& F_1(x) = \int_{u=-\infty}^{x} \int_{v=-\infty}^{\infty} f(u,v)\mathrm{d}v\mathrm{d}u \\ \Pr(Y \leq y) =& F_2(y) = \int_{u=-\infty}^{\infty} \int_{v=-\infty}^{y} f(u,v)\mathrm{d}v\mathrm{d}u.\end{aligned}$$

Two Random Variables

Conditional distributions

We already know that

$$\Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)}.$$

If X and Y are discrete random variables, an equivalent definition of the **conditional probability** of the event that $Y = y$ given that $X = x$ is given by the **conditional density**:

$$\Pr(Y = y|X = x) = \frac{f(x, y)}{f_1(x)},$$

where $f(x, y)$ is the joint probability function and $f_1(x)$ is the marginal probability function.

Exercise 10

Find $f(y|x)$ if X and Y have the joint density function

$$f(x,y) = \begin{cases} \frac{3}{4} + xy & \text{if } 0 < x < 1, 0 < y < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Solution

First, we need to obtain the marginal density of X, which can be obtained by integrating out the y in the following way:

$$f_1(x) = \int_0^1 \left(\frac{3}{4} + xy\right) \mathrm{d}y = \frac{3}{4} + \int_0^1 xy\mathrm{d}y = \frac{3}{4} + \left[\frac{x}{2}y^2\right]_0^1 = \frac{3}{4} + \frac{x}{2}.$$

Once we have the marginal density of X, we can get the conditional density $f(y|x) = \frac{f(x,y)}{f_1(x)}$:

$$f(y|x) = \frac{\frac{3}{4} + xy}{\frac{3}{4} + \frac{x}{2}} = \frac{3 + 4xy}{3 + 2x} = \begin{cases} \frac{3+4xy}{3+2x} & \text{if } 0 < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Two Random Variables

Independence

If random variables X and Y are discrete and the two events that $X = x$ an $Y = y$ are independent for all x and y, then these are called **independent random variables**. This means that

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

or

$$f(x, y) = f_1(x) f_2(y).$$

If X and Y are independent, then

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$$
$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

Two Random Variables

Independence

If random variables X and Y are continuous, then these are independent random variables if the events $X \leq x$ and $Y \leq y$ are independent events for all x and y. This is equivalent to

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

or equivalently

$$F(x, y) = F_1(x) F_2(y),$$

and also implies

$$f(x, y) = f_1(x) f_2(y).$$

This means that the knowledge of the probability of one event does not provide any information on the probability of the other event occurring.

Exercise 11

Roll two dice and consider the following three events:

- $A =$"first die is 3"
- $B =$ "the sum is 6"
- $C =$ "the sum is 7"

Is A independent of B, C, both or neither?

Solution

The question we should ask ourselves is: by knowing that B or C occurred, does the probability of A change?

The probability of A is the probability that the first die is equal to 3, which is is $1/6$. This corresponds to the unconditional probability $\Pr(A)$.

If B occurs, then we know that the possible outcomes on the first die are not $\{1, 2, 3, 4, 5, 6\}$ but only $\{1, 2, 3, 4, 5\}$. Thus, if we know B occurred, it is now more likely that A occurred (i.e., the conditional probability of A given B, $\Pr(A|B) \neq \Pr(A)$). Thus, A and B are not independent.

On the other hand, if C occurred, the first die can still take all the possible values with the same probability. Thus, A and C are independent, and $\Pr(A|C) = \Pr(A)$.

Two Random Variables

Covariance and correlation

The **covariance** measures the degree to which two random variables vary together, e.g. height and weight of people.

If X and Y are two random variables with mean μ_x and μ_y, then

$$\mathrm{Cov}(X, Y) = \mathbb{E}\left[(X - \mu_X)(Y - \mu_Y)\right].$$

The properties of the covariance are:

1. $\mathrm{Cov}(aX + b, cY + d) = ac\mathrm{Cov}(X, Y)$
2. $\mathrm{Cov}(X_1 + X_2, Y) = \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y)$
3. $\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mu_X\mu_Y$
4. $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$

If X and Y are independent, then $\mathrm{Cov}(X, Y) = 0$. This is because if X and Y are independent, then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, which means that

$$\begin{aligned}\mathrm{Cov}(X, Y) &= \mathbb{E}(XY) - \mu_X \mu_Y \\ &= \mathbb{E}(X)\mathbb{E}(Y) - \mu_X \mu_Y \\ &= \mu_X \mu_Y - \mu_X \mu_Y = 0.\end{aligned}$$

The converse is not true: if the covariance is 0, the variables might not be independent.

Exercise 12

Consider the following table, which shows the probability function $f(x, y)$ of the random variables X and Y. Show that $\mathrm{Cov}(X, Y) = 0$ but X and Y are not independent.

$Y \setminus X$	-1	0	1	$f_2(y)$
-1	$1/6$	$1/3$	$1/6$	$2/3$
1	$1/6$	0	$1/6$	$1/3$
$f_1(x)$	$1/3$	$1/3$	$1/3$	1

Solution

To get the covariance between X and Y, let's use $\text{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$:

$$\begin{aligned}
\mathbb{E}(X) &= -1 \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3} = 0 \\
\mathbb{E}(Y) &= -1 \times \frac{2}{3} + 0 \times 0 + 1 \times \frac{1}{3} = -\frac{1}{3} \\
\mathbb{E}(XY) &= 1 \times \frac{1}{6} + 0 \times \frac{1}{3} - 1 \times \frac{1}{6} - 1 \times \frac{1}{6} + 0 \times 0 + 1 \times \frac{1}{6} = 0
\end{aligned}$$

Thus,

$$\text{Cov}(X,Y) = 0 - 0 \times \frac{1}{3} = 0.$$

For independence, we need that $f(x,y) = f(x)f(y)$ for all (x,y), which is not the case here. For instance,

$$f(-1,-1) = \frac{1}{6} \neq f_x(-1)f_y(-1) = \frac{2}{3}\frac{1}{3} = \frac{2}{9}.$$

Two Random Variables

Covariance and correlation

The **correlation coefficient** between X and Y is defined by

$$\mathrm{Corr}(X,Y) = \rho = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}, \tag{1}$$

and it measures the linear relationship between X and Y.

Properties:

1. $-1 \leq \rho \leq 1$. $\rho = 1$ if and only if $Y = aX + b$ and $a > 0$, and $\rho = -1$ if and only if $Y = aX + b$ and $a < 0$.
2. When $\rho = 1$ ($\rho = -1$), X and Y are perfectly, positively (negatively) correlated.
3. When $\rho = 0$, X and Y are uncorrelated.

Important: correlation is not causation.

Normal Distribution

Normal Distribution

The **normal distribution**, also called Gaussian distribution, is characterized by the pdf

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},$$

for $-\infty < x < \infty$ ($x \in \mathbb{R}$), where μ and σ are the mean and the standard deviation, and $\sigma > 0$. The notation we use to describe a normally distributed variable with mean μ and variance σ^2 is

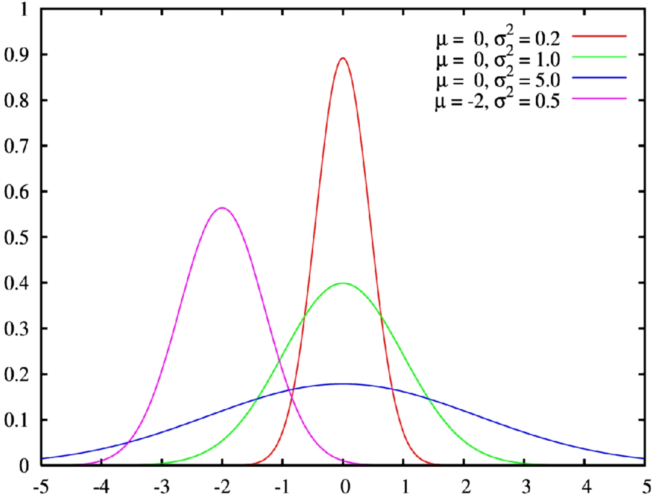
$$X \sim \mathcal{N}(\mu, \sigma^2).$$

The distribution function is given by

$$F(x) = \Pr(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{v-\mu}{\sigma}\right)^2} \mathrm{d}v.$$

Normal Distribution

Figure 3: Probability Density Function of the Normal Distribution



Normal Distribution

A random variable Z has the **standard normal distribution** (or is standard normal) if $Z \sim \mathcal{N}(0,1)$. Its density is then

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then the standardized variable $Z = \frac{X-\mu}{\sigma}$ is normal with mean 0 and variance 1, so $Z \sim \mathcal{N}(0,1)$:

$$\mathbb{E}(Z) = \mathbb{E}\left(\frac{X-\mu}{\sigma}\right) = \frac{\mathbb{E}(X)-\mu}{\sigma} = \frac{\mu-\mu}{\sigma} = 0$$

$$\mathrm{Var}(Z) = \mathrm{Var}\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}\mathrm{Var}(X) = \frac{\sigma^2}{\sigma^2} = 1.$$

A **property** of the normal distribution: a sum of normally distributed variables is also normally distributed.

Normal Distribution

Normal densities are **symmetric** around the mean μ. This has the following implications:

- If $a < 0$, then $\Phi(a) = 1 - \Phi(-a)$ where Φ denotes the cdf of the standard normal distribution.
- If $a < b < 0$, then $\Phi(b) - \Phi(a) = \Phi(-a) - \Phi(-b)$.
- If $a < 0$ and $b > 0$, then $\Phi(b) - \Phi(a) = \Phi(b) + \Phi(-a) - 1$.

Then, to find the probability that $a \leq X \leq b$:

$$\begin{aligned}
\Pr(a \leq X \leq b) &= \Pr\left(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) \\
&= \Pr\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right) \\
&= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).
\end{aligned}$$

Example

If $X \sim \mathcal{N}(\mu, \sigma^2)$ with $\mu = 4$ and $\sigma^2 = 49$, then

$$\begin{aligned}
\Pr(-2 \le X \le 5) &= \Pr\left(\frac{-2-4}{7} \le \frac{X-\mu}{\sigma} \le \frac{5-4}{7}\right) \\
&= \Pr(-0.8571 \le Z \le 0.1429) \\
&= \Phi(0.1429) - \Phi(-0.8571) \\
&= \Phi(0.1429) + \Phi(0.8571) - 1 \\
&= 0.5557 + 0.8051 - 1 = 0.3608.
\end{aligned}$$

Chi-Squared Distribution

Chi-Squared Distribution

The ***chi-squared distribution*** is the distribution of the sum of m squared, independent, standard normal random variables:

$$W = Z_1^2 + ... + Z_M^2 = \sum_{m=1}^{M} Z_m^2 \sim \chi_M^2 \text{ with } Z_m \overset{i.i.d.}{\sim} \mathcal{N}(0,1).$$

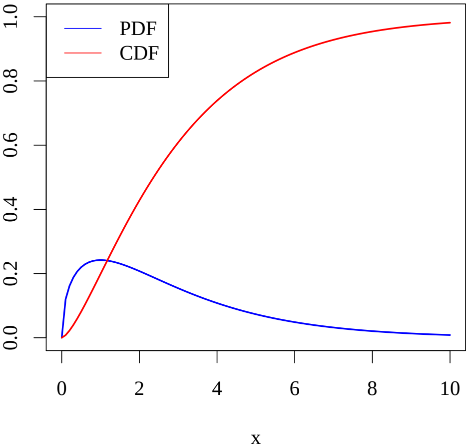
The mean and the variance of the chi-squared distribution are given by:

$$\mathbb{E}(W) = m$$
$$\mathrm{Var}(W) = 2m.$$

A **property** of the chi-squared distribution: if $W_1 \sim \chi^2_{m_1}$,
$W_2 \sim \chi^2_{m_2}$, ..., $W_n \sim \chi^2_{m_n}$, then
$W_1 + W_2 + \ldots + W_n \sim \chi^2_{m_1+m_2+\ldots+m_n}$.

Chi-Squared Distribution

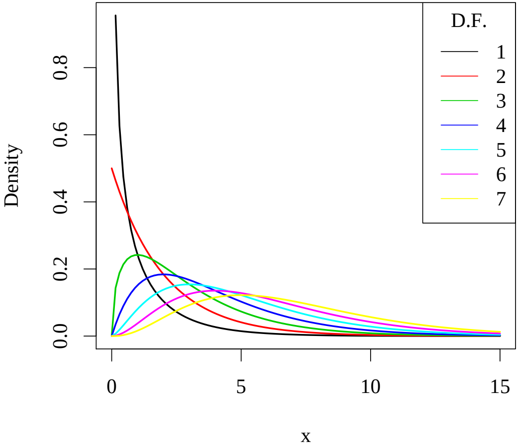
Figure 4: PDF and CDF of the Chi-Squared Distribution with $m = 3$



Chi-Squared Distribution

The shape of the distribution changes with the number of degrees of freedom m.

Figure 5: Chi-Squared Distributed Random Variables



Student's-t Distribution

Student's-t Distribution

Let Z and W be two independently distributed random variables, such that $Z \sim \mathcal{N}(0,1)$ and $W \sim \chi^2_m$. Then,

$$X = \frac{Z}{\sqrt{W/M}} \sim t_m, \tag{2}$$

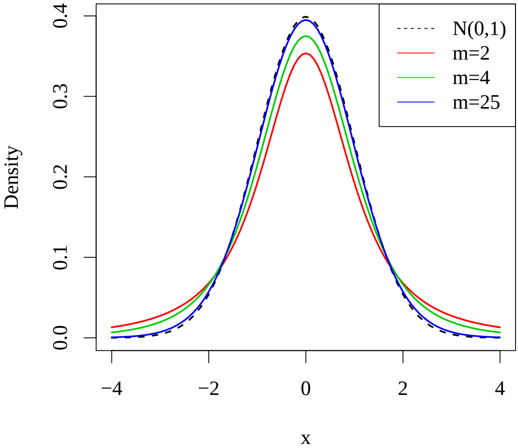
that is, X has a **Student's-*t* distribution** with m degrees of freedom.

A t_m distributed random variable X has expectation $\mathbb{E}[X] = 0$ if $m > 1$, and variance $\mathrm{Var}[X] = m/(m-2)$ if $m > 2$.

For a sufficiently large m, the t_m distribution can be approximated by the standard normal distribution. Actually, $t_\infty = \mathcal{N}(0,1)$.

Student's-t Distribution

Figure 6: Densities of t Distributions compared to $\mathcal{N}(0,1)$



F Distribution

F Distribution

Let W be a chi-squared random variable with m degrees of freedom and let V be a chi-squared random variable with n degrees of freedom. Assume that W and V are independently distributed. Then,

$$\frac{W/m}{V/n} \sim F_{m,n} \text{ with } W \sim \chi^2_m, V \sim \chi^2_n,$$

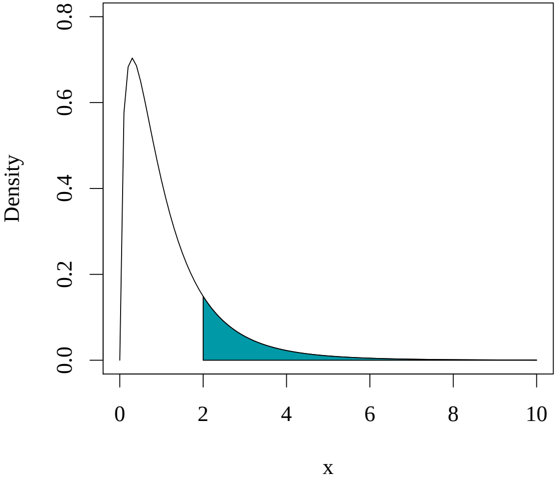
which is an ***F* distribution** with m and n degrees of freedom.

Note that:

- $F_{m,\infty} = \chi^2_m/m$.
- $F_{1,n} = \frac{W}{V/n} = t_n^2$ ($W \sim \chi^2_1$, and Z^2 for $Z \sim \mathcal{N}(0,1)$ is also χ^2_1).

F Distribution

Figure 7: Density of $x \sim F_{3,14}$



Appendix

Appendix: Integration

- The function $\int g(x)\mathrm{d}x$ is an integral of the function $g(x)$ with respect to x (indicated by $\mathrm{d}x$).
- **Constant**: $\int a\mathrm{d}x = ax$.
- **Multiplication by a constant**: $\int ag(x)\mathrm{d}x = a\int g(x)\mathrm{d}x$.
- **Power rule**: $\int x^k = \frac{1}{k+1}x^{k+1}$.
- **Sum rule**: $\int \left(g(x) + f(x)\right)\mathrm{d}x = \int g(x)\mathrm{d}x + \int f(x)\mathrm{d}x$.

Appendix: Integration

- **Definite integrals**: $\int_a^b g(x) = [G(X)]_a^b = G(b) - G(a)$ where $G(x)$ is the integral of $g(x)$ (indeed, $\frac{\partial G(x)}{\partial x} = g(x)$).

- **Functions of more than one variable**: $f(x, y)$ can be integrated over either or both variables.
  - Integrating $f(x, y)$ over y : $\int f(x, y)\mathrm{d}y$.
  - Integrating $f(x, y)$ over both x and y : $\int_x \int_y f(x, y)\mathrm{d}y\mathrm{d}x$.
  - With definite integrals: $\int_a^b \int_c^d f(x, y)\mathrm{d}y\mathrm{d}x$.