Chapter 3

Random Variables and Their Distributions

A **random variable** (r.v.) is a function that assigns one and only one numerical value to each simple event in an experiment.

We will denote r.vs by capital letters: X, Y or Z and their values by small letters: x, y or z respectively.

There are two types of r.vs: discrete and continuous. Random variables that can take a countable number of values are called **discrete**. Random variables that take values from an interval of real numbers are called **continuous**.

Example 3.1 Discrete r.v., Brockwell and Davis (2002)
Yearly results of the all-star baseball games over years 1933-1995, where

$$X_t = \begin{cases} 1 & \text{if the National League won in year } t, \\ -1 & \text{if the Americal League won in year } t. \end{cases}$$

In each of the realizations there is some probability p of $X_t = 1$ and probability $1 - p$ of $X_t = -1$.
□

Example 3.2 Continuous r.v.
The r.vs in all the examples given in Chapter 1 are continuous.
□

3.1 One Dimensional Random Variables

Definition 3.1. *If E is an experiment having sample space S, and X is a function that assigns a real number $X(e)$ to every outcome $e \in S$, then X is called a random variable.* □

Definition 3.2. *Let X denote a r.v. and x its particular value from the whole range of all values of X, say R_X. The probability of the event $(X \leq x)$ expressed as a function of x:*

$$F_X(x) = P_X(X \leq x) \tag{3.1}$$

is called the **cumulative distribution function** *(c.d.f.) of the r.v. X.* □

Properties of cumulative distribution functions

- $0 \leq F_X(x) \leq 1, \quad -\infty < x < \infty$
- $\lim_{x\to\infty} F_X(x) = 1$
- $\lim_{x\to-\infty} F_X(x) = 0$
- The function is nondecreasing. That is, if $x_1 \leq x_2$ then $F_X(x_1) \leq F_X(x_2)$.

3.1.1 Discrete Random Variables

Values of a discrete r.v. are elements of a countable set $\{x_1, x_2, \ldots\}$. We associate a number $p_X(x_i) = P_X(X = x_i)$ with each value $x_i,\ i = 1, 2, \ldots$, such that:

1. $p_X(x_i) \geq 0$ for all i
2. $\sum_{i=1}^{\infty} p_X(x_i) = 1$

Note that

$$F_X(x_i) = P_X(X \leq x_i) = \sum_{x \leq x_i} p_X(x) \tag{3.2}$$

$$p_X(x_i) = F_X(x_i) - F_X(x_{i-1}) \tag{3.3}$$

The function p_X is called the **probability mass function** of the random variable X, and the collection of pairs

$$\{(x_i, p_X(x_i)),\ i = 1, 2, \ldots\} \tag{3.4}$$

is called the **probability distribution** of X. The distribution is usually presented in either tabular, graphical or mathematical form.

Examples of known p.m.fs

Ex. 3.3 The Binomial Distribution (X denotes k successes in n independent trials). The p.m.f. of a binomially distributed r.v. X with parameters n and p is

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \ldots, n,$$

where n is a positive integer and $0 \le p \le 1$. □

Ex. 3.4 The Uniform Distribution The p.m.f. of a r.v. X uniformly distributed on $\{1, 2, \ldots, n\}$ is

$$P(X = k) = \frac{1}{n}, \quad k = 1, 2, \ldots, n,$$

where n is a positive integer. □

Ex. 3.5 The Poisson Distribution (X denotes a number of outcomes in a period of time). The p.m.f. of a r.v. X having a Poisson distribution with parameter $\lambda > 0$ is

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

□

Take as an example a r.v. X having the Binomial distribution: $X \sim \text{Bin}(8, 0.4)$. That is $n = 8$ and the probability of success $p = 0.4$. The distribution, shown in a mathematical, tabular and graphical way and a graph of the c.d.f. of the variable X follow.
Mathematical form:

$$\{(k, P(X = k) = {}^nC_k p^k (1-p)^{n-k}), \; k = 0, 1, 2, \ldots, 8\} \tag{3.5}$$

Tabular form:

k	0	1	2	3	4	5	6	7	8
$P(X = k)$	0.0168	0.0896	0.2090	0.2787	0.2322	0.1239	0.0413	0.0079	0.0007
$P(X \le k)$	0.0168	0.1064	0.3154	0.5941	0.8263	0.9502	0.9915	0.9993	1

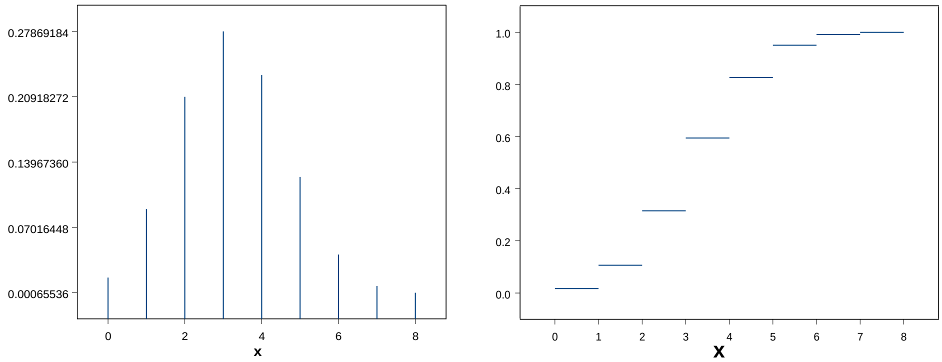

Figure 3.1: Graphical representation of the mass function and the cumulative distribution function for $X \sim \mathrm{Bin}(8, 0.4)$

Other important discrete distributions are:

- $Bernoulli(p)$
- $Geometric(p)$
- $Hypogeomeric(n, M, N)$

3.1.2 Continuous Random Variables

Values of a continuous r.v. are elements of an uncountable set, for example a real interval. A c.d.f. of a continuous r.v. is a continuous, nondecreasing, differentiable function. An interesting difference from a discrete r.v. is that for $\delta > 0$

$$P_X(X = x) = lim_{\delta \to 0}(F_X(x + \delta) - F_X(x)) = 0.$$

We define the **probability density function** (p.d.f.) of a continuous r.v. as:

$$f_X(x) = \frac{d}{dx} F_X(x) \tag{3.6}$$

Hence

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt \tag{3.7}$$

Similarly to the properties of the probability distribution of a discrete r.v. we have the following properties of the density function:

1. $f_X(x) \geq 0$ for all $x \in R_X$
2. $\int_{R_X} f_X(x)dx = 1$

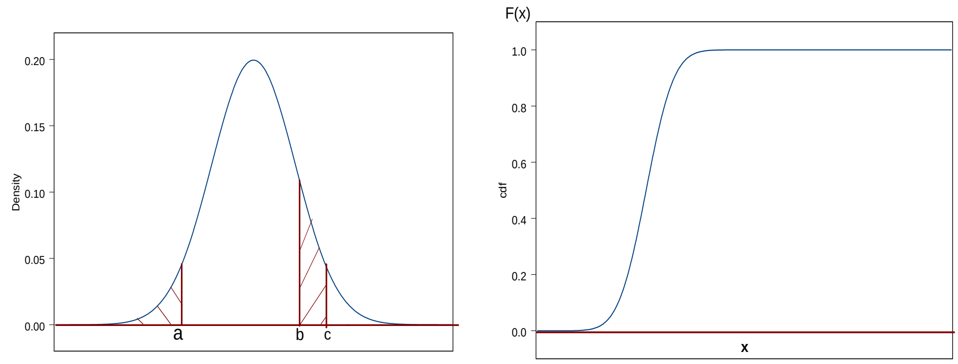

Figure 3.2: Distribution Function and Cumulative Distribution Function for $N(4.5, 2)$

Probability of an event $(X \in A)$, where A is an interval $(-\infty, a)$, is expressed as an integral

$$P_X(-\infty < X < a) = \int_{-\infty}^{a} f_X(x)dx = F_X(a) \tag{3.8}$$

or for a bounded interval (b, c)

$$P_X(b < X < c) = \int_{b}^{c} f_X(x)dx = F_X(c) - F_X(b). \tag{3.9}$$

Examples of continuous r.vs

Ex. 3.6 The normal distribution $N(\mu, \sigma^2)$
The density function is given by

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}. \tag{3.10}$$

There are two parameters which tell us about the position and the shape of the density curve: the expected value μ and the standard deviation σ. □

Ex. 3.7 The uniform distribution $U(a, b)$
The p.d.f. of a r.v. X uniformly distributed on $[a, b]$ is given by

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a \leq x \leq b, \\ 0, & \text{otherwise.} \end{cases}$$

□

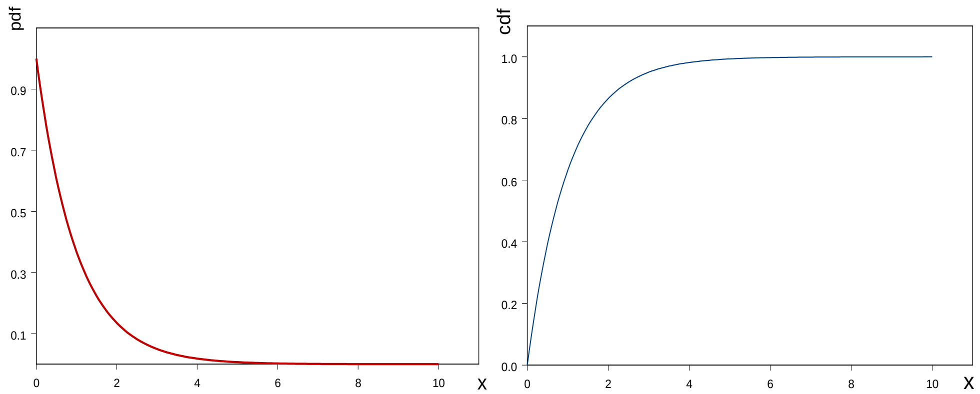

Figure 3.3: Distribution density function and cumulative distribution function for $Exp(1)$

Ex. 3.8 The exponential distribution $Exp(\lambda)$

The p.d.f. of an exponentially distributed r.v. X with parameter $\lambda > 0$ is

$$f_X(x) = \begin{cases} 0, & \text{if } x < 0, \\ \lambda e^{-\lambda x}, & \text{if } x \geq 0. \end{cases}$$

The corresponding c.d.f. is

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0, \\ 1 - e^{-\lambda x}, & \text{if } x \geq 0. \end{cases}$$

$\square$

Ex. 3.9 The gamma distribution $Gamma(\alpha, \lambda)$ with parameters representing shape (α) and scale (λ). The p.d.f. of a gamma distributed r.v. X is

$$f_X(x) = \begin{cases} 0, & \text{if } x < 0, \\ x^{\alpha-1}\lambda^{\alpha}e^{\lambda x}/\Gamma(\alpha), & \text{if } x \geq 0, \end{cases}$$

where the $\alpha > 0$, $\lambda > 0$ and Γ is the gamma function defined by

$$\Gamma(n) = \int_0^{\infty} x^{n-1}e^{-x}dx, \quad \text{for } n > 0.$$

A recursive relationship that may be easily shown integrating the above equation by parts is

$$\Gamma(n) = (n-1)\Gamma(n-1).$$

If n is a positive integer, then

$$\Gamma(n) = (n-1)!$$

since $\Gamma(1) = 1$.

$\square$

Other important continuous distributions are

- The chi-squared distribution with ν degrees of freedom, χ^2_ν
- The F distribution with ν_1, ν_2 degrees of freedom, F_{ν_1,ν_2}.
- Student t-distribution with ν degrees of freedom, t_ν.

These distributions are functions of normally distributed r.vs.

Note that all the distributions depend on some parameters, like p, λ, μ, σ or other. These values are usually unknown, and so their estimation is one of the important problems in statistical analyses.

3.1.3 Expectation

The expectation of a function g of a r.v. X is defined by

$$\mathrm{E}(g(X)) = \begin{cases} \displaystyle\int_{-\infty}^{\infty} g(x)f(x)dx, & \text{for a continuous r.v.,} \\ \displaystyle\sum_{j=0}^{\infty} g(x_j)p(x_j) & \text{for a discrete r.v.,} \end{cases} \tag{3.11}$$

and g is any function such that $\mathrm{E}\,|g(X)| < \infty$.

Two important special cases of g are:

Mean $\mathrm{E}(X)$, also denoted by $\mathrm{E}\,X$, when $g(X) = X$,

Variance $\mathrm{E}(X - \mathrm{E}\,X)^2$, when $g(X) = (X - \mathrm{E}\,X)^2$. The following relation is very useful while calculating the variance

$$\mathrm{E}(X - \mathrm{E}\,X)^2 = \mathrm{E}\,X^2 - (\mathrm{E}\,X)^2. \tag{3.12}$$

Example 3.10
Let X be a r.v. such that

$$f(x) = \begin{cases} \frac{1}{2}\sin x, & \text{for } x \in [0, \pi], \\ 0 & \text{otherwise.} \end{cases}$$

Then the expectation and variance are following

- Expectation
$$\operatorname{E} X = \frac{1}{2}\int_0^{\pi} x \sin x dx = \frac{\pi}{2}.$$
- Variance
$$\begin{aligned}\operatorname{var}(X) &= \operatorname{E} X^2 - (\operatorname{E} X)^2\\ &= \frac{1}{2}\int_0^{\pi} x^2 \sin x dx - \left(\frac{\pi}{2}\right)^2\\ &= \frac{\pi^2}{4}.\end{aligned}$$

□

Example 3.11 The Normal distribution, $X \sim N(\mu, \sigma^2)$

We will show that the parameter μ is the expectation of X and the parameter σ^2 is the variance of X.

First we show that $\operatorname{E}(X-\mu) = 0$.
$$\begin{aligned}\operatorname{E}(X-\mu) &= \int_{-\infty}^{\infty} (x-\mu) f(x) dx\\ &= \int_{-\infty}^{\infty} (x-\mu)\frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx\\ &= -\sigma^2 \int_{-\infty}^{\infty} f'(x) dx = 0.\end{aligned}$$

Hence $\operatorname{E} X = \mu$.

Similar arguments give
$$\begin{aligned}\operatorname{E}(X-\mu)^2 &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx\\ &= -\sigma^2 \int_{-\infty}^{\infty} (x-\mu) f'(x) dx = \sigma^2.\end{aligned}$$

□

A useful linearity property of expectation is
$$\operatorname{E}[ag(X) + bh(Y) + c] = a\operatorname{E}[g(X)] + b\operatorname{E}[h(Y)] + c, \tag{3.13}$$

for any real constants a, b and c and for any functions g and h of r.vs X and Y whose expectations exist.

3.2 Two Dimensional Random Variables

Definition 3.3. *Let S be a sample space associated with an experiment E, and X_1, X_2 be functions, each assigning a real number $X_1(e), X_2(e)$ to every outcome $e \in E$. Then the pair $\boldsymbol{X} = (X_1, X_2)$ is called a two-dimensional random variable. The range space of the two-dimensional random variable is*

$$R_{\boldsymbol{X}} = \{(x_1, x_2) : x_1 \in R_{X_1},\ x_2 \in R_{X_2}\} \subset R^2.$$

□

Definition 3.4. *The cumulative distribution function of a two-dimensional r.v. $\boldsymbol{X} = (X_1, X_2)$ is*

$$F_{\boldsymbol{X}}(x_1, x_2) = P(X_1 \le x_1, X_2 \le x_2) \tag{3.14}$$

□

3.2.1 Discrete Two-Dimensional Random Variables

If all values of $\boldsymbol{X} = (X_1, X_2)$ are countable, i.e., the values are in the range

$$R_{\boldsymbol{X}} = \{(x_{1i}, x_{2j}),\ i = 1, \ldots, n,\ j = 1, \ldots, m\}$$

then the variable is discrete. The c.d.f. of a discrete r.v. $\boldsymbol{X} = (X_1, X_2)$ is defined as

$$F_{\boldsymbol{X}}(x_1, x_2) = \sum_{x_{2j} \le x_2} \sum_{x_{1i} \le x_1} p(x_{1i}, x_{2j}) \tag{3.15}$$

where $p(x_{1i}, x_{2j})$ denotes the **joint probability mass function** and

$$p(x_{1i}, x_{2j}) = P(X_1 = x_{1i}, X_2 = x_{2j}).$$

As in the univariate case, the joint p.m.f. satisfies the following conditions.

1. $p(x_{1i}, x_{2j}) \ge 0$, for all i, j
2. $\sum_{all\ j} \sum_{all\ i} p(x_{1i}, x_{2j}) = 1$

Example 3.12 Tossing two fair dice.
Consider an experiment of tossing two fair dice and noting the outcome on each die. The whole sample space consists of 36 elements, i.e.,

$$\mathcal{S} = \{(i, j) : i, j = 1, \ldots, 6\}.$$

Now, with each of these 36 elements associate values of two variables, X_1 and X_2, such that

$$X_1 = sum\ of\ the\ outcomes\ on\ the\ two\ dice,$$
$$X_2 = |\,difference\ of\ the\ outcomes\ on\ the\ two\ dice\,|.$$

Then the bivariate r.v. $\boldsymbol{X} = (X_1, X_2)$ has the following joint probability mas function.

						x_1						
		2	3	4	5	6	7	8	9	10	11	12
	0	$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$
	1		$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$	
	2			$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$		
x_2	3				$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$			
	4					$\frac{1}{18}$		$\frac{1}{18}$				
	5						$\frac{1}{18}$					

□

Marginal p.m.fs

Theorem 3.1. *Let* $\boldsymbol{X} = (X_1, X_2)$ *be a discrete bivariate random variable with joint p.m.f.* $p(x_1, x_2)$. *Then the marginal p.m.fs of* X_1 *and* X_2, p_{X_1} *and* p_{X_2}, *are given respectively by*

$$p_{X_1}(x_1) = P(X_1 = x_1) = \sum_{R_{X_2}} p(x_1, x_2) \quad \textit{and}$$
$$p_{X_2}(x_2) = P(X_2 = x_2) = \sum_{R_{X_1}} p(x_1, x_2).$$

□

Example 3.12 cont. The marginal distributions of the variables X_1 and X_2 are following.

x_1	2	3	4	5	6	7	8	9	10	11	12
$P(X_1 = x_1)$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

x_2	0	1	2	3	4	5
$P(X_2 = x_2)$	$\frac{1}{6}$	$\frac{5}{18}$	$\frac{2}{9}$	$\frac{1}{6}$	$\frac{1}{9}$	$\frac{1}{18}$

□

Expectations of functions of bivariate random variables are calculated the same way as the univariate r.vs. Let $g(x_1, x_2)$ be a real valued function defined on $R_{\boldsymbol{X}}$. Then $g(\boldsymbol{X}) = g(X_1, X_2)$ is a r.v. and its expectation is

$$\mathrm{E}[g(\boldsymbol{X})] = \sum_{R_{\boldsymbol{X}}} g(x_1, x_2)p(x_1, x_2).$$

Example 3.12 cont. For $g(X_1, X_2) = X_1X_2$ we obtain

$$\mathrm{E}[g(\boldsymbol{X})] = 2 \times 0 \times \frac{1}{36} + \ldots + 7 \times 5 \times \frac{1}{18} = \frac{245}{18}.$$

□

3.2.2 Continuous Two-Dimensional Random Variables

If the values of $\boldsymbol{X} = (X_1, X_2)$ are elements of an uncountable set in the Euclidean plane, then the variable is continuous. For example the values might be in the range

$$R_{\boldsymbol{X}} = \{(x_1, x_2) : a \leq x_1 \leq b, c \leq x_2 \leq d\}$$

for some real a, b, c, d.

The c.d.f. of a continuous r.v. $\boldsymbol{X} = (X_1, X_2)$ is defined as

$$F_{\boldsymbol{X}}(x_1, x_2) = \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f(t_1, t_2) dt_1 dt_2, \tag{3.16}$$

where $f(x_1, x_2)$ is the probability density function such that

1. $f(x_1, x_2) \geq 0$ for all $(x_1, x_2) \in R^2$
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2) dx_1 dx_2 = 1$.

The equation (3.16) implies that

$$\frac{\partial^2 F(x_1, x_2)}{\partial x_1 \partial x_2} = f(x_1, x_2). \tag{3.17}$$

Also

$$P(a \leq X_1 \leq b, c \leq X_2 \leq d) = \int_c^d \int_a^b f(x_1, x_2) dx_1 dx_2. \tag{3.18}$$

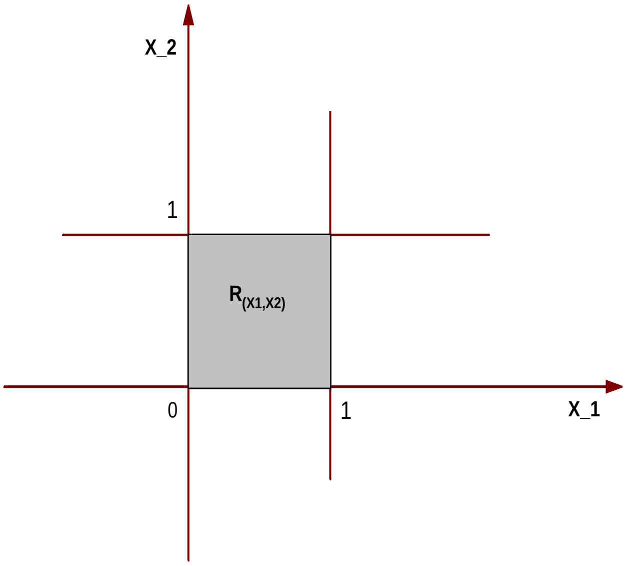

Figure 3.4: Domain of the p.d.f. (3.21)

The marginal p.d.fs of X_1 and X_2 are defined similarly as in the discrete case, here using integrals.

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2)dx_2, \quad \text{for} \quad -\infty < x_1 < \infty, \tag{3.19}$$

$$f_{X_2}(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2)dx_1, \quad \text{for} \quad -\infty < x_2 < \infty. \tag{3.20}$$

Example 3.13
Calculate $P[\boldsymbol{X} \subseteq A]$, where $A = \{(x_1, x_2) : x_1 + x_2 \geq 1\}$ and the joint p.d.f. of $\boldsymbol{X} = (X_1, X_2)$ is defined by

$$f_{\boldsymbol{X}}(x_1, x_2) = \begin{cases} 6x_1x_2^2 & \text{for } 0 < x_1 < 1, \ 0 < x_2 < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.21}$$

The probability is a double integral of the p.d.f. over the region A. The region is however limited by the domain in which the p.d.f. is positive, see Figure 3.4.

We can write

$$\begin{aligned} A &= \{(x_1,x_2) : x_1 + x_2 \geq 1,\ 0 < x_1 < 1,\ 0 < x_2 < 1\} \\ &= \{(x_1,x_2) : x_1 \geq 1 - x_2,\ 0 < x_1 < 1,\ 0 < x_2 < 1\} \\ &= \{(x_1,x_2) : 1 - x_2 < x_1 < 1,\ 0 < x_2 < 1\}. \end{aligned}$$

Hence, the probability is

$$P(\boldsymbol{X} \subseteq A) = \int\int_A f(x_1,x_2)dx_1dx_2 = \int_0^1 \int_{1-x_2}^1 6x_1x_2^2dx_1dx_2 = 0.9$$

Also, using formulae ((3.19)) and ((3.20)) we can calculate marginal p.d.fs.

$$f_{X_1}(x_1) = \int_0^1 6x_1x_2^2dx_2 = 2x_1x_2^3 \mid_0^1 = 2x_1,$$

$$f_{X_2}(x_2) = \int_0^1 6x_1x_2^2dx_1 = 3x_1^2x_2^2 \mid_0^1 = 3x_2^2.$$

These functions allow us to calculate probabilities involving only one variable. For example

$$P\left(\frac{1}{4} < X_1 < \frac{1}{2}\right) = \int_{\frac{1}{4}}^{\frac{1}{2}} 2x_1dx_1 = \frac{3}{16}.$$

□

Similarly to the case of a univariate r.v. the following linear property for the expectation holds.

$$\mathrm{E}[ag(\boldsymbol{X}) + bh(\boldsymbol{X}) + c] = a\,\mathrm{E}[g(\boldsymbol{X})] + b\,\mathrm{E}[h(\boldsymbol{X})] + c, \tag{3.22}$$

where a, b and c are constants and g and h are some functions of the bivariate r.v. $\boldsymbol{X} = (X_1, X_2)$.

3.2.3 Conditional Distributions and Independence

Definition 3.5. *Let $\boldsymbol{X} = (X_1, X_2)$ denote a discrete bivariate r.v. with joint p.m.f. $p_{\boldsymbol{X}}(x_1,x_2)$ and marginal p.m.fs $p_{X_1}(x_1)$ and $p_{X_2}(x_2)$. For any x_1 such that $p_{X_1}(x_1) > 0$, the conditional p.m.f. of X_2 given that $X_1 = x_1$ is the function of x_2 defined by*

$$p_{X_2}(x_2|x_1) = \frac{p_{\boldsymbol{X}}(x_1,x_2)}{p_{X_1}(x_1)}. \tag{3.23}$$

Analogously, we define the conditional p.m.f. of X_1 given $X_2 = x_2$

$$p_{X_1}(x_1|x_2) = \frac{p_{\boldsymbol{X}}(x_1,x_2)}{p_{X_2}(x_2)}. \tag{3.24}$$

□

It is easy to check that these functions are indeed p.d.fs. For example, for (3.23) we have

$$\sum_{R_{X_2}} p_{X_2}(x_2|x_1) = \sum_{R_{X_2}} \frac{p_{\boldsymbol{X}}(x_1, x_2)}{p_{X_1}(x_1)} = \frac{\sum_{R_{X_2}} p_{\boldsymbol{X}}(x_1, x_2)}{p_{X_1}(x_1)} = \frac{p_{X_1}(x_1)}{p_{X_1}(x_1)} = 1.$$

Example 3.12 cont.
The conditional p.m.f. of X_2 given, for example, $X_1 = 5$, is

x_2	0	1	2	3	4	5
$p_{X_2}(x_2\|5)$	0	$\frac{1}{2}$	0	$\frac{1}{2}$	0	0

□

Analogously to the conditional distribution for discrete r.vs, we define the conditional distribution for continuous r.vs.

Definition 3.6. *Let* $\boldsymbol{X} = (X_1, X_2)$ *denote a continuous bivariate r.v. with joint p.d.f.* $f_{\boldsymbol{X}}(x_1, x_2)$ *and marginal p.d.fs* $f_{X_1}(x_1)$ *and* $f_{X_2}(x_2)$*. For any* x_1 *such that* $f_{X_1}(x_1) > 0$*, the conditional p.d.f. of* X_2 *given that* $X_1 = x_1$ *is the function of* x_2 *defined by*

$$f_{X_2}(x_2|x_1) = \frac{f_{\mathbf{X}}(x_1, x_2)}{f_{X_1}(x_1)}. \tag{3.25}$$

Analogously, we define the conditional p.d.f. of X_1 *given* $X_2 = x_2$

$$f_{X_1}(x_1|x_2) = \frac{f_{\mathbf{X}}(x_1, x_2)}{f_{X_2}(x_2)}. \tag{3.26}$$

□

Here too, it is easy to verify that these functions are p.d.fs. For example, for the function (3.25) we have

$$\begin{aligned}\int_{R_{X_2}} f_{X_2}(x_2|x_1)dx_2 &= \int_{R_{X_2}} \frac{f_{\boldsymbol{X}}(x_1, x_2)}{f_{X_1}(x_1)} dx_2 \\ &= \frac{\int_{R_{X_2}} f_{\boldsymbol{X}}(x_1, x_2)dx_2}{f_{X_1}(x_1)} \\ &= \frac{f_{X_1}(x_1)}{f_{X_1}(x_1)} = 1.\end{aligned}$$

Example 3.13 cont.
The conditional p.d.fs are

$$f_{X_1}(x_1|x_2) = \frac{f_{\boldsymbol{X}}(x_1,x_2)}{f_{X_2}(x_2)} = \frac{6x_1x_2^2}{3x_2^2} = 2x_1$$

and

$$f_{X_2}(x_2|x_1) = \frac{f_{\boldsymbol{X}}(x_1,x_2)}{f_{X_1}(x_1)} = \frac{6x_1x_2^2}{2x_1} = 3x_2^2.$$

□

The conditional p.d.fs make it possible to calculate conditional expectations. The conditional expected value of a function $g(X_2)$ given that $X_1 = x_1$ is defined by

$$\mathrm{E}[g(X_2)|x_1] = \begin{cases} \sum_{R_{X_2}} g(x_2)p_{X_2}(x_2|x_1) & \text{for a discrete r.v.,} \\ \int_{R_{X_2}} g(x_2)f_{X_2}(x_2|x_1)dx_2 & \text{for a continuous r.v..} \end{cases} \tag{3.27}$$

Example 3.13 cont.
Equation ((3.27)) allows to calculate conditional mean and variance of a r.v. Here we have

$$\mu_{X_2|x_1} = \mathrm{E}(X_2|x_1) = \int_0^1 x_2 3x_2^2 dx_2 = \frac{3}{4},$$

and

$$\sigma^2_{X_2|x_1} = \mathrm{var}(X_2|x_1) = \mathrm{E}(X_2^2|x_1) - [\mathrm{E}(X_2|x_1)]^2 = \int_0^1 x_2^2 3x_2^2 dx_2 - \left(\frac{3}{4}\right)^2 = \frac{3}{80}.$$

□

Definition 3.7. *Let $\boldsymbol{X} = (X_1, X_2)$ denote a continuous bivariate r.v. with joint p.d.f. $f_{\boldsymbol{X}}(x_1,x_2)$ and marginal p.d.fs $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$. Then X_1 and X_2 are called independent random variables if, for every $x_1 \in R_{X_1}$ and $x_2 \in R_{X_2}$*

$$f_{\boldsymbol{X}}(x_1,x_2) = f_{X_1}(x_1)f_{X_2}(x_2). \tag{3.28}$$

□

If X_1 and X_2 are independent, then the conditional p.d.f. of X_2 given $X_1 = x_1$ is

$$f_{X_2}(x_2|x_1) = \frac{f_{\boldsymbol{X}}(x_1, x_2)}{f_{X_1}(x_1)} = \frac{f_{X_1}(x_1)f_{X_2}(x_2)}{f_{X_1}(x_1)} = f_{X_2}(x_2)$$

regardless of the value of x_1. Analogous property holds for the conditional p.d.f. of X_1 given $X_2 = x_2$.

Example 3.13 cont.
It is easy to notice that

$$f_{\boldsymbol{X}}(x_1, x_2) = 6x_1x_2^2 = 2x_1 3x_2^2 = f_{X_1}(x_1)f_{X_2}(x_2).$$

So, the variables X_1 and X_2 are independent. □

In fact, two r.vs are independent if and only if there exist functions $g(x_1)$ and $h(x_2)$ such that for every $x_1 \in R_{X_1}$ and $x_2 \in R_{X_2}$,

$$f_{\boldsymbol{X}}(x_1, x_2) = g(x_1)h(x_2).$$

Theorem 3.2. *Let X_1 and X_2 be independent random variables. Then*

1. *For any $A \subset \mathcal{R}$ and $B \subset \mathcal{R}$*

$$P(X_1 \in A, X_2 \in B) = P(X_1 \in A)P(X_2 \in B),$$

   that is, the events $\{X_1 \in A\}$ and $\{X_2 \in B\}$ are independent events.

2. *For $g(X_1)$, a function of X_1 only, and for $h(X_2)$, a function of X_2 only, we have*

$$\mathrm{E}[g(X_1)h(X_2)] = \mathrm{E}[g(X_1)]\,\mathrm{E}[h(X_2)].$$

□

3.2.4 Covariance and Correlation

Covariance and correlation are two measures of the strength of a relationship between two r.vs.

We will use the following notation.
$\mathrm{E}\,X_1 = \mu_{X_1}$
$\mathrm{E}\,X_2 = \mu_{X_2}$
$\mathrm{var}(X_1) = \sigma^2_{X_1}$
$\mathrm{var}(X_2) = \sigma^2_{X_2}$

Also, we assume that $\sigma^2_{X_1}$ and $\sigma^2_{X_2}$ are finite positive values.

A simplified notation $\mu_1,\ \mu_2,\ \sigma^2_1,\ \sigma^2_2$ will be used when it is clear which r.vs the notation refers to.

Definition 3.8. *The covariance of X_1 and X_2 is defined by*

$$\mathrm{cov}(X_1, X_2) = \mathrm{E}[(X_1 - \mu_{X_1})(X_2 - \mu_{X_2})]. \tag{3.29}$$

□

Definition 3.9. *The correlation of X_1 and X_2 is defined by*

$$\rho_{(X_1,X_2)} = \mathrm{corr}(X_1, X_2) = \frac{\mathrm{cov}(X_1, X_2)}{\sigma_{X_1}\sigma_{X_2}}. \tag{3.30}$$

□

Properties of covariance

- For any r.vs X_1 and X_2,

  $$\mathrm{cov}(X_1, X_2) = \mathrm{E}(X_1X_2) - \mu_{X_1}\mu_{X_2}.$$

- If r.vs X_1 and X_2 are independent then

  $$\mathrm{cov}(X_1, X_2) = 0.$$

  Then also $\rho_{(X_1,X_2)} = 0$.

- For any r.vs X_1 and X_2 and any constants a and b,

  $$\mathrm{var}(aX_1 + bX_2) = a^2\,\mathrm{var}(X_1) + b^2\,\mathrm{var}(X_2) + 2ab\,\mathrm{cov}(X_1, X_2).$$

Properties of correlation

For any r.vs X_1 and X_2

- $-1 \leq \rho_{(X_1,X_2)} \leq 1$,
- $|\rho_{(X_1,X_2)}| = 1$ iff there exist numbers $a \neq 0$ and b such that

$$P(X_2 = aX_1 + b) = 1.$$

If $\rho_{(X_1,X_2)} = 1$ then $a > 0$, and if $\rho_{(X_1,X_2)} = -1$ then $a < 0$.

3.2.5 Bivariate Normal Distribution

Here, we use matrix notation. A bivariate r.v. is treated as a random vector

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}.$$

Let $\boldsymbol{X} = (X_1, X_2)^{\mathrm{T}}$ be a bivariate random vector with expectation

$$\boldsymbol{\mu} = \mathrm{E}\,\boldsymbol{X} = \mathrm{E}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}.$$

and the variance-covariance matrix

$$\boldsymbol{V} = \begin{pmatrix} \operatorname{var}(X_1) & \operatorname{cov}(X_1, X_2) \\ \operatorname{cov}(X_2, X_1) & \operatorname{var}(X_2) \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Then the joint p.d.f. of the vector r.v. $\boldsymbol{X}$ is given by

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{2\pi\sqrt{\det(\boldsymbol{V})}} \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{V}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right\}, \tag{3.31}$$

where $\boldsymbol{x} = (x_1, x_2)^{\mathrm{T}}$.

The determinant of $\boldsymbol{V}$ is

$$\det \boldsymbol{V} = \det\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} = (1 - \rho^2)\sigma_1^2\sigma_2^2.$$

Hence, the inverse of $\boldsymbol{V}$ is

$$\boldsymbol{V}^{-1} = \frac{1}{\det \boldsymbol{V}} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} = \frac{1}{1-\rho^2} \begin{pmatrix} \sigma_1^{-2} & -\rho\sigma_1^{-1}\sigma_2^{-1} \\ -\rho\sigma_1^{-1}\sigma_2^{-1} & \sigma_2^{-2} \end{pmatrix}.$$

Then the exponent in formula (3.31) can be written as

$$\begin{aligned} &-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{V}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = \\ &= -\frac{1}{2(1-\rho^2)}(x_1-\mu_1, x_2-\mu_2) \begin{pmatrix} \sigma_1^{-2} & -\rho\sigma_1^{-1}\sigma_2^{-1} \\ -\rho\sigma_1^{-1}\sigma_2^{-1} & \sigma_2^{-2} \end{pmatrix} \begin{pmatrix} x_1-\mu \\ x_2-\mu \end{pmatrix} \\ &= -\frac{1}{2(1-\rho^2)} \left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} \right). \end{aligned}$$

So, the joint p.d.f. of the 2-dimensional vector r.v. $\boldsymbol{X}$ is

$$\begin{aligned} f_{\boldsymbol{X}}(\boldsymbol{x}) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1-\rho^2)}} \\ &\times \exp\left\{ \frac{-1}{2(1-\rho^2)} \left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} \right) \right\}. \end{aligned} \tag{3.32}$$

Remark 3.1. Note that when $\rho = 0$ the joint p.d.f. (3.32) simplifies to

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{ -\frac{1}{2} \left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} \right) \right\},$$

which can be written as a product of the marginal distributions of X_1 and X_2. Hence, if $\boldsymbol{X} = (X_1, X_2)^{\mathrm{T}}$ has a bivariate normal distribution and $\rho = 0$ then the variables X_1 and X_2 are independent. □

3.3 Multivariate Normal Distribution

Denote by $\boldsymbol{X} = (X_1, \ldots, X_n)^{\mathrm{T}}$ an n-dimensional random vector whose each component is a random variable. Then all the definitions given for bivariate r.vs extend to the multivariate r.vs.

$\boldsymbol{X}$ has a multivariate normal distribution if its p.d.f. can be written as in (3.31), where the mean is

$$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^{\mathrm{T}},$$

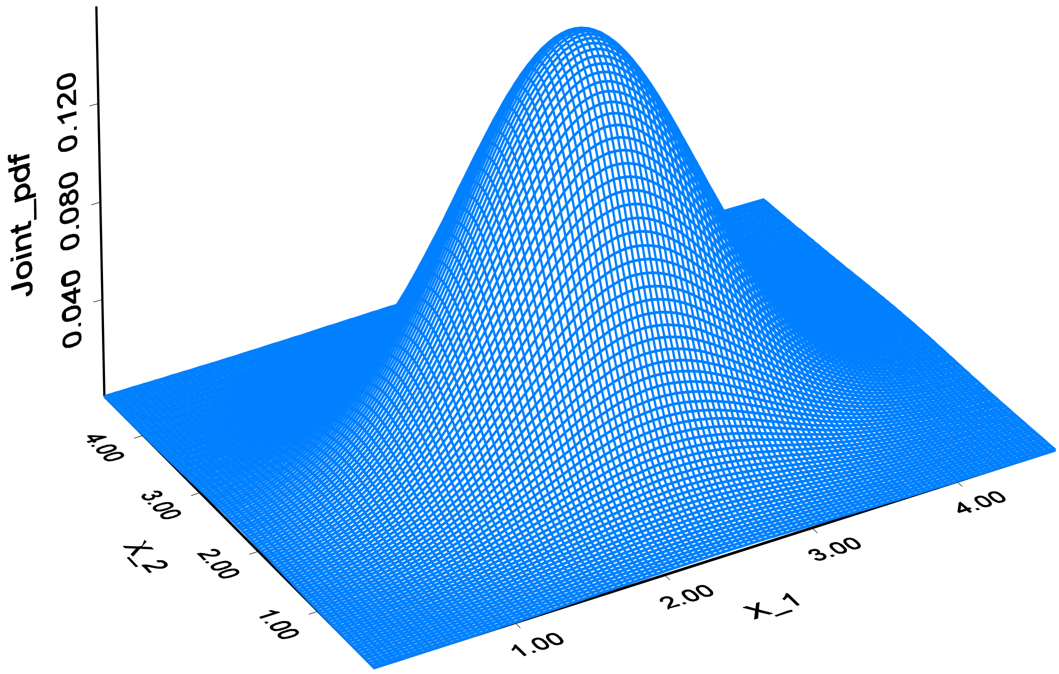

Figure 3.5: Bivariate Normal pdf

and the variance-covariance matrix has the form

$$\boldsymbol{V} = \begin{pmatrix} \mathrm{var}(X_1) & \mathrm{cov}(X_1, X_2) & \dots & \mathrm{cov}(X_1, X_n) \\ \mathrm{cov}(X_2, X_1) & \mathrm{var}(X_2) & \dots & \mathrm{cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(X_n, X_1) & \mathrm{cov}(X_n, X_2) & \dots & \mathrm{var}(X_n) \end{pmatrix}$$

Remark 3.2. If $\boldsymbol{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{V})$, $\boldsymbol{B}$ is an $m \times n$ matrix, and $\boldsymbol{a}$ is a real $m \times 1$ vector, then the random vector

$$\boldsymbol{Y} = \boldsymbol{a} + \boldsymbol{B}\boldsymbol{X}$$

is also multivariate normal with

$$\mathrm{E}(\boldsymbol{Y}) = \boldsymbol{a} + \boldsymbol{B}\,\mathrm{E}(\boldsymbol{X}) = \boldsymbol{a} + \boldsymbol{B}\boldsymbol{\mu},$$

and the variance-covariance matrix,

$$\boldsymbol{V}_{\boldsymbol{Y}} = \boldsymbol{B}\boldsymbol{V}\boldsymbol{B}^{\mathrm{T}}.$$

□

Remark 3.3. Taking $\boldsymbol{B} = \boldsymbol{b}^{\mathrm{T}}$, where $\boldsymbol{b}$ is an $n \times 1$ dimensional vector and $\boldsymbol{a} = \boldsymbol{0}$ we obtain

$$Y = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{X} = b_1 X_1 + \ldots + b_n X_n,$$

and

$$Y \sim \mathcal{N}(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{\mu}, \boldsymbol{b}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{b}).$$

□

Two important properties of multivariate normal random vectors are given in the following proposition.

Proposition 3.1. *Suppose that the normal r.v. is partitioned into two subvectors of dimensions* n_1 *and* n_2

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}^{(1)} \\ \boldsymbol{X}^{(2)} \end{pmatrix},$$

and, correspondingly, the mean vector and the variance-covariance matrix are partitioned as

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{pmatrix} \quad \boldsymbol{V} = \begin{pmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{21} & \boldsymbol{V}_{22} \end{pmatrix},$$

where

$$\boldsymbol{\mu}^{(i)} = \mathrm{E}(\boldsymbol{X}^{(i)}) \quad \textit{and} \quad \boldsymbol{V}_{ij} = \mathrm{cov}(\boldsymbol{X}^{(i)}, \boldsymbol{X}^{(j)}).$$

Then

1. $\boldsymbol{X}^{(1)}$ *and* $\boldsymbol{X}^{(2)}$ *are independent iff the* $(n_1 \times n_2)$ *dimensional covariance matrix* $\boldsymbol{V}_{12}$ *is a zero matrix.*

2. *The conditional distribution of* $\boldsymbol{X}^{(1)}$ *given* $\boldsymbol{X}^{(2)} = \boldsymbol{x}^{(2)}$ *is*

$$\mathcal{N}(\boldsymbol{\mu}^{(1)} + \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}(\boldsymbol{x}^{(2)} - \boldsymbol{\mu}^{(2)}),\ \boldsymbol{V}_{11} - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{V}_{21}).$$

□

For the bivariate normal r.v. we obtain

$$\mathrm{E}(X_1|X_2 = x_2) = \mu_1 + \rho\sigma_1\sigma_2^{-1}(x_2 - \mu_2) \tag{3.33}$$

and

$$\mathrm{var}(X_1|X_2 = x_2) = \sigma_1^2(1 - \rho^2). \tag{3.34}$$

Definition 3.10. X_t *is a* **Gaussian time series** *if all its joint distributions are multivariate normal, i.e., if for any collection of integers* $i_1, \ldots, i_n$*, the random vector* $(X_{i_1}, \ldots, X_{i_n})^{\mathrm{T}}$ *has a multivariate normal distribution.* □

For proofs of the results given in this chapter see Castella and Berger (1990) and Brockwell and Davis (2002) or other textbooks on probability and statistics.