4 Random Vectors

Everything that holds for random *variables* (one-dimensional case) can be easily generalized to any dimension, i.e. to random *vectors*. We restrict our discussion to two-dimensional random vectors $(X, Y) : S \to \mathbb{R}^2$.

Let $(S, \mathcal{K}, P)$ be a probability space. A **random vector** is a function $(X, Y) : S \to \mathbb{R}^2$ satisfying the condition

$$(X \le x, Y \le y) = \{e \in S \mid X(e) \le x, Y(e) \le y\} \in \mathcal{K},$$

for all $(x, y) \in \mathbb{R}^2$.

- if the set of values that it takes, $(X, Y)(S)$, is at most countable in $\mathbb{R}^2$, then (X, Y) is a **discrete random vector**,
- if $(X, Y)(S)$ is a continuous subset of $\mathbb{R}^2$, then (X, Y) is a **continuous random vector**.
- the function $F : \mathbb{R}^2 \to \mathbb{R}$ defined by

  $$F(x, y) \;=\; P(X \le x, Y \le y)$$

  is called the **joint cumulative distribution function (joint cdf)** of the vector (X, Y).

The properties of the cdf of a random variable translate very naturally for a random vector, as well: Let (X, Y) be a random vector with joint cdf $F : \mathbb{R}^2 \to \mathbb{R}$ and let $F_X, F_Y : \mathbb{R} \to \mathbb{R}$ be the cdf's of X and Y, respectively. Then following properties hold:

- If $a_k < b_k,\ k = \overline{1, 2}$, then

  $$\begin{aligned} P(a_1 < X \le b_1, a_2 < Y \le b_2) &= F(b_1, b_2) - F(b_1, a_2) \\ &- F(b_2, a_1) + F(a_1, a_2). \end{aligned}$$

- $\lim\limits_{x,y \to \infty} F(x, y) = 1$,
  $\lim\limits_{y \to -\infty} F(x, y) = \lim\limits_{x \to -\infty} F(x, y) = 0,\ \forall x, y \in \mathbb{R}$,
  $\lim\limits_{y \to \infty} F(x, y) = F_X(x),\ \forall x \in \mathbb{R}$,
  $\lim\limits_{x \to \infty} F(x, y) = F_Y(y),\ \forall y \in \mathbb{R}$.

4.1 Discrete Random Vectors

Let $(X, Y) : S \to \mathbb{R}^2$ be a two-dimensional discrete random vector. The **joint probability distribution (function)** of (X, Y) is a two-dimensional array of the form

$$
\begin{array}{c|ccccc|c}
X \setminus Y & y_1 & \dots & y_j & \dots & & \\
\hline
x_1 & & & & & & \\
\vdots & & & \vdots & & & \\
x_i & & \dots & p_{ij} & \dots & & p_i \\
\vdots & & & \vdots & & & \\
 & & & & & & \\
\hline
 & & & q_j & & &
\end{array}
\tag{4.1}
$$

where $(x_i, y_j) \in \mathbb{R}^2,\ (i, j) \in I \times J$ are the values that (X, Y) takes and $p_{ij} = P(X = x_i, Y = y_j)$.

An important property is that

$$\sum_{j \in J} p_{ij} = p_i,\ \sum_{i \in I} p_{ij} = q_j \ \text{ and } \ \sum_{i \in I} \sum_{j \in J} p_{ij} = \sum_{j \in J} \sum_{i \in I} p_{ij} = 1,$$

where $p_i = P(X = x_i),\ i \in I$ and $q_j = P(Y = y_j),\ j \in J$. The probabilities p_i and q_j are called **marginal** pdf's.

For discrete random vectors, the computational formula for the cdf is

$$F(x, y) \ = \ \sum_{x_i \le x} \sum_{y_j \le y} p_{ij},\ x, y \in \mathbb{R}.$$

Operations with discrete random variables

Let X and Y be two discrete random variables with pdf's

$$X \begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i \in I} \ \text{ and } \ Y \begin{pmatrix} y_j \\ q_j \end{pmatrix}_{j \in J}.$$

Sum. The sum of X and Y is the random variable with pdf given by

$$X + Y \begin{pmatrix} x_i + y_j \\ p_{ij} \end{pmatrix}_{(i,j) \in I \times J}. \tag{4.2}$$

Product. The product of X and Y is the random variable with pdf given by

$$X \cdot Y \begin{pmatrix} x_i y_j \\ p_{ij} \end{pmatrix}_{(i,j) \in I \times J}. \tag{4.3}$$

Scalar Multiple. The random variable αX, $\alpha \in \mathbb{R}$, with pdf given by

$$\alpha X \begin{pmatrix} \alpha x_i \\ p_i \end{pmatrix}_{i \in I}. \tag{4.4}$$

Quotient. The quotient of X and Y is the random variable with pdf given by

$$X/Y \begin{pmatrix} x_i/y_j \\ p_{ij} \end{pmatrix}_{(i,j) \in I \times J}, \tag{4.5}$$

provided that $y_j \neq 0$, for all $j \in J$.
In general, if $h : \mathbb{R} \to \mathbb{R}$ is a function, then we can define the random variable $h(X)$, with pdf given by

$$h(X) \begin{pmatrix} h(x_i) \\ p_i \end{pmatrix}_{i \in I}. \tag{4.6}$$

Variables X and Y are said to be **independent** if

$$p_{ij} = P\left(X = x_i, Y = y_j\right) = P\left(X = x_i\right) P\left(Y = y_j\right) = p_i q_j, \tag{4.7}$$

for all $(i, j) \in I \times J$.
If X and Y are independent, then in (4.2), (4.3) and (4.5), $p_{ij} = p_i q_j$, for all $(i, j) \in I \times J$.

4.2 Continuous Random Vectors

Let (X, Y) be a continuous random vector with joint cdf $F : \mathbb{R}^2 \to \mathbb{R}$. Then F is *absolutely continuous*, i.e. there exists a real function $f : \mathbb{R}^2 \to \mathbb{R}$, such that

$$F(x, y) \;=\; \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\, du\, dv, \tag{4.8}$$

for all $x, y \in \mathbb{R}$. The function f is called the **joint probability density function (joint pdf)** of (X, Y).

The usual properties of continuous pdf's (and their relationship with cdf's) hold for the two-dimensional case, as well: Let (X, Y) be a continuous random vector with joint cdf F and joint density function f. Let $F_X, F_Y : \mathbb{R} \to \mathbb{R}$ be the cdf's of X and Y and $f_X, f_Y : \mathbb{R} \to \mathbb{R}$ be the pdf's of X and Y, respectively. Then the following properties hold:

- $\dfrac{\partial^2 F(x,y)}{\partial x \partial y} = f(x,y)$, for all $(x,y) \in \mathbb{R}^2$.
- $\displaystyle\iint\limits_{\mathbb{R}^2} f(x,y)\, dxdy = 1.$
- for any domain $D \subseteq \mathbb{R}^2$, $P\big((X,Y) \in D\big) = \displaystyle\iint\limits_{D} f(x,y)\, dxdy.$
- $f_X(x) = \displaystyle\int\limits_{\mathbb{R}} f(x,y)\, dy,\ \forall x \in \mathbb{R}$ and $f_Y(y) = \displaystyle\int\limits_{\mathbb{R}} f(x,y)\, dx,\ \forall y \in \mathbb{R}.$

When obtained from the vector (X, Y), the pdf's f_X and f_Y are called *marginal* densities.
The continuous random variables X and Y are said to be **independent** if

$$f_{(X,Y)}(x,y) = f_X(x) f_Y(y), \tag{4.9}$$

for all $(x,y) \in \mathbb{R}^2$.

5 Common Distributions

5.1 Common Discrete Distributions

Bernoulli Distribution $Bern(p)$

A random variable X has a Bernoulli distribution with parameter $p \in (0,1)$ $(q = 1-p)$, if its pdf is

$$X \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}. \tag{5.1}$$

Then

$$\begin{aligned} E(X) &= p, \\ V(X) &= pq. \end{aligned}$$

A Bernoulli r.v. models the occurrence or nonoccurrence of an event.

Discrete Uniform Distribution $U(m)$

A random variable X has a Discrete Uniform distribution ($\boxed{\text{unid}}$) with parameter $m \in \mathbb{N}$, if its pdf is

$$X \begin{pmatrix} k \\ \frac{1}{m} \end{pmatrix}_{k=\overline{1,m}}, \tag{5.2}$$

with mean and variance

$$\begin{aligned} E(X) &= \frac{m+1}{2}, \\ V(X) &= \frac{m^2-1}{12}. \end{aligned}$$

The random variable that denotes the face number shown on a die when it is rolled, has a discrete uniform distribution $U(6)$.

Binomial Distribution $B(n, p)$

A random variable X has a Binomial distribution ($\boxed{\text{bino}}$) with parameters $n \in \mathbb{N}$ and $p \in (0, 1)$ ($q = 1 - p$), if its pdf is

$$X \begin{pmatrix} k \\ C_n^k p^k q^{n-k} \end{pmatrix}_{k=\overline{0,n}}, \tag{5.3}$$

with

$$\begin{aligned} E(X) &= np, \\ V(X) &= npq. \end{aligned}$$

This distribution corresponds to the Binomial model. Given n Bernoulli trials with probability of success p, let X denote the number of successes. Then $X \in B(n, p)$. Also, notice that the Bernoulli distribution is a particular case of the Binomial one, for $n = 1$, $Bern(p) = B(1, p)$.

Geometric Distribution $Geo(p)$

A random variable X has a Geometric distribution ($\boxed{\text{geo}}$) with parameter $p \in (0,1)$ $(q = 1-p)$, if its pdf is given by

$$X \begin{pmatrix} k \\ pq^k \end{pmatrix}_{k=0,1,\ldots}. \tag{5.4}$$

Its cdf, expectation and variance are given by

$$\begin{aligned} F(x) &= 1 - q^{x+1}, \\ E(X) &= \frac{q}{p}, \\ V(X) &= \frac{q}{p^2}. \end{aligned}$$

If X denotes the number of failures that occurred before the occurrence of the 1^{st} success in a Geometric model, then $X \in Geo(p)$.

Remark 5.1. In a Geometric model setup, one might count the number of *trials* needed to get the 1^{st} success. Of course, if X is the number of failures and Y the number of trials, then we simply have $Y = X + 1$ (the number of failures plus the one success). The variable Y is said to have a Shifted Geometric distribution with parameter $p \in (0,1)$ ($Y \in SGeo(p)$). Its pdf is

$$X \begin{pmatrix} k \\ pq^{k-1} \end{pmatrix}_{k=1,2,\ldots} \tag{5.5}$$

and the rest of its characteristics are given by

$$\begin{aligned} F(x) &= 1 - q^{x}, \\ E(X) &= \frac{1}{p}, \\ V(X) &= \frac{q}{p^2}. \end{aligned}$$

In some books, *this* is considered to be a Geometric variable (not in Matlab, though).

Negative Binomial (Pascal) Distribution $NB(n,p)$

A random variable X has a Negative Binomial (Pascal) ($\boxed{\text{nbin}}$) distribution with parameters $n \in \mathbb{N}$ and $p \in (0,1)$ $(q = 1-p)$, if its pdf is

$$X \begin{pmatrix} k \\ C_{n+k-1}^{k} p^n q^k \end{pmatrix}_{k=0,1,\dots}. \tag{5.6}$$

Then

$$\begin{aligned} E(X) &= \frac{nq}{p}, \\ V(X) &= \frac{nq}{p^2}. \end{aligned}$$

This distribution corresponds to the Negative Binomial model. If X denotes the number of failures that occurred before the occurrence of the n^{th} success in a Negative Binomial model, then $X \in NB(n,p)$. It is a generalization of the Geometric distribution, $Geo(p) = NB(1,p)$.

Poisson Distribution $\mathcal{P}(\lambda)$

A random variable X has a Poisson distribution ($\boxed{\text{poiss}}$) with parameter $\lambda > 0$, if its pdf is

$$X \begin{pmatrix} k \\ \dfrac{\lambda^k}{k!} e^{-\lambda} \end{pmatrix}_{k=0,1,\dots} \tag{5.7}$$

with

$$E(X) = V(X) = \lambda.$$

Poisson's distribution is related to the concept of "rare events", or Poissonian events. Essentially, it means that two such events are extremely unlikely to occur simultaneously or within a very short period of time. Arrivals of jobs, telephone calls, e-mail messages, traffic accidents, network blackouts, virus attacks, errors in software, floods, earthquakes are examples of rare events.

A Poisson variable X counts the number of rare events occurring during a fixed time interval. The parameter λ represents the average number of occurrences of the event in that time interval.

Remark 5.2.

1. The sum of n independent $Bern(p)$ random variables is a $B(n,p)$ variable.
2. The sum of n independent $Geo(p)$ random variables is a $NB(n,p)$ variable.

5.2 Common Continuous Distributions

Uniform Distribution $U(a,b)$

A random variable X has a Uniform distribution ($\boxed{\text{unif}}$)with parameters $a, b \in \mathbb{R},\ a < b$, if its pdf is

$$f(x) \;=\; \begin{cases} \dfrac{1}{b-a}, & \text{if } x \in [a,b] \\ 0, & \text{if } x \notin [a,b]. \end{cases} \tag{5.8}$$

Then its cdf is

$$F(x) = \int_{-\infty}^{x} f(t)dt \;=\; \begin{cases} 0, & \text{if } x \le a \\ \dfrac{x-a}{b-a}, & \text{if } a < x \le b \\ 1, & \text{if } x \ge b \end{cases} \tag{5.9}$$

and its numerical characteristics are

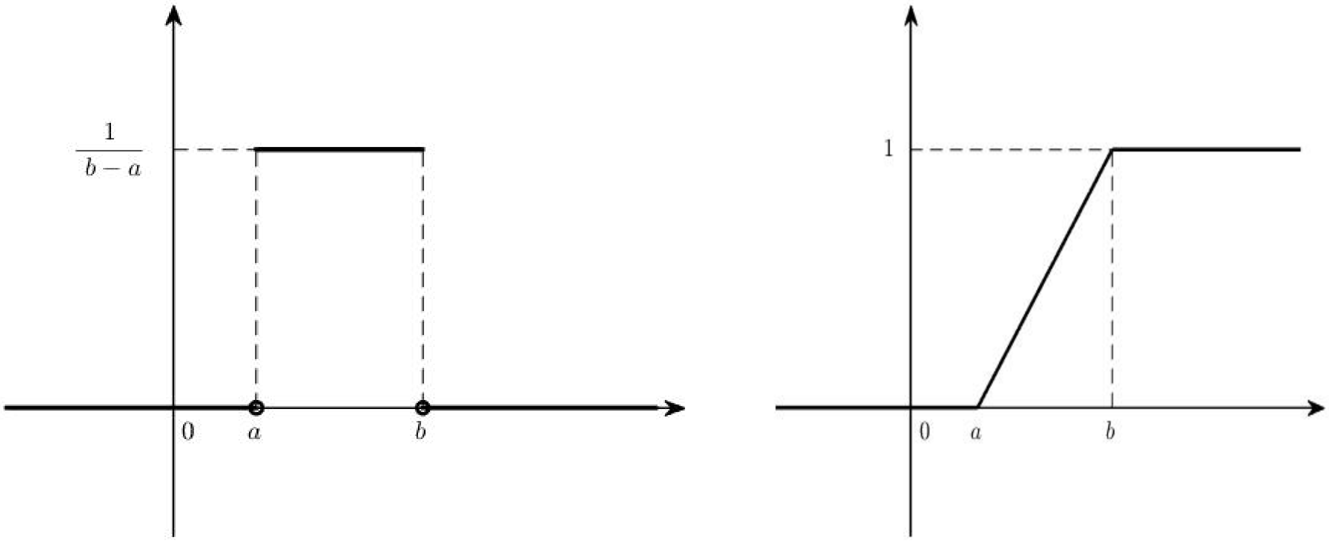
$$\begin{aligned} E(X) &= \frac{a+b}{2}, \\ V(X) &= \frac{(b-a)^2}{12}. \end{aligned}$$

The Uniform distribution is used when a variable can take *any* value in a given interval, equally probable. For example, locations of syntax errors in a program, birthdays throughout a year, arrival times of customers, etc.

A special case is that of a **Standard Uniform Distribution**, where $a = 0$ and $b = 1$. The pdf and cdf are given by

$$f_U(x) = \begin{cases} 1, & x \in [0,1] \\ 0, & x \notin [0,1] \end{cases}, \quad F_U(x) = \begin{cases} 0, & x \le 0 \\ x, & 0 < x \le 1 \\ 1, & x \ge 1\,. \end{cases} \tag{5.10}$$

Standard Uniform variables play an important role in stochastic modeling; in fact, *any* random



(a) Density Function (pdf) (b) Cumulative Distribution Function (cdf)

Fig. 1: Uniform Distribution

variable, with any thinkable distribution (discrete or continuous) can be generated from Standard Uniform variables.

Normal Distribution $N(\mu, \sigma)$

A random variable X has a Normal distribution ($\boxed{\text{norm}}$) with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$, if its pdf is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}. \tag{5.11}$$

The cdf of a Normal variable is then given by

$$F(x) \quad = \quad \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\dfrac{(t-\mu)^2}{2\sigma^2}} dt = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\dfrac{t^2}{2}} dt \tag{5.12}$$

and its mean and variance are

$$\begin{aligned} E(X) &= \mu, \\ V(X) &= \sigma^2. \end{aligned}$$

There is an important particular case of a Normal distribution, namely $N(0,1)$, called the **Standard (or Reduced) Normal Distribution**. A variable having a Standard Normal distribution is usually denoted by Z. The density and cdf of Z are given by

$$f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R} \quad \text{and} \quad F_Z(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt. \tag{5.13}$$

The function F_Z given in (5.13) is known as *Laplace's function* and its values can be found in tables or can be computed by any mathematical software. One can notice that there is a relationship between the cdf of any Normal $N(\mu, \sigma)$ variable X and that of a Standard Normal variable Z, namely,

$$F_X(x) = F_Z\left(\frac{x-\mu}{\sigma}\right).$$

Exponential Distribution $Exp(\lambda)$

A random variable X has an Exponential distribution ($\boxed{\text{exp}}$) with parameter $\lambda > 0$, if its pdf and cdf are given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases} \quad \text{and} \quad F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}, \tag{5.14}$$

respectively. Its mean and variance are given by

$$\begin{aligned} E(X) &= \frac{1}{\lambda}, \\ V(X) &= \frac{1}{\lambda^2}. \end{aligned}$$

Remark 5.3.

1. The Exponential distribution is often used to model *time*: lifetime, waiting time, halftime, interarrival time, failure time, time between rare events, etc. The parameter λ represents the frequency of rare events, measured in time^{-1}.

2. A word of **caution** here: The parameter μ in Matlab (where the Exponential pdf is defined as $\frac{1}{\mu}e^{-\frac{1}{\mu}x}, x \geq 0$) is actually $\mu = 1/\lambda$. It all comes from the different interpretation of the "frequency". For instance, if the frequency is "2 per hour", then $\lambda = 2/\text{hr}$, but this is equivalent to " one every half an hour"', so $\mu = 1/2$ hours. The parameter μ is measured in time units.

3. The Exponential distribution is a special case of a more general distribution, namely the $Gamma(a, b)$, $a, b > 0$, distribution ($\boxed{\text{gam}}$). The Gamma distribution models the *total* time of a multistage scheme, e.g. total compilation time, total downloading time, etc.

4. If $\alpha \in \mathbb{N}$, then the sum of α independent $Exp(\lambda)$ variables has a $Gamma(\alpha, 1/\lambda)$ distribution.

5. In a Poisson process, where X is the number of rare events occurring in time t, $X \in \mathcal{P}(\lambda t)$, the time between rare events and the time of the occurrence of the first rare event have $Exp(\lambda)$ distribution, while T, the time of the occurrence of the α^{th} rare event has $Gamma(\alpha, 1/\lambda)$ distribution.

Gamma-Poisson formula

Let $T \in Gamma(\alpha, 1/\lambda)$ with $\alpha \in \mathbb{N}$ and $\lambda > 0$. Then T represents the time of the occurrence of the α^{th} rare event. Then, the event $(T > t)$ means that the α^{th} event occurs <u>after</u> the moment t. That means that <u>before</u> the time t, <u>fewer</u> than α rare events occur. So, if X is the number of rare events that occur before time t, then the two events

$$(T > t) = (X < \alpha)$$

are equivalent (equal). Now, X has a $\mathcal{P}(\lambda t)$ distribution. So, we have:

$$\begin{aligned} P(T > t) &= P(X < \alpha) \quad \text{and} \\ P(T \leq t) &= P(X \geq \alpha). \end{aligned} \tag{5.15}$$

Remark 5.4. This formula is useful in applications where this setup can be used (seeing a Gamma variable as a sum of times between rare events, if $\alpha \in \mathbb{N}$), as it avoids lengthy computations of Gamma probabilities. However, one should be **careful**, T is a *continuous* random variable, for which $P(T > t) = P(T \geq t)$, whereas X is a discrete one, so on the right-hand sides of (5.15) the inequality signs cannot be changed.

Remark 5.5. The Exponential distributions has the so-called "memoryless property". Suppose that an Exponential variable T represents waiting time. Memoryless property means that the fact of having waited for t minutes gets "forgotten" and it does not affect the future waiting time. Regardless of the event $(T > t)$, when the total waiting time exceeds t, the remaining waiting time still has

Exponential distribution with the same parameter. Mathematically,

$$P(T > t + x | T > t) \;\; = \;\; P(T > x), \; t, x > 0. \tag{5.16}$$

The Exponential distribution is the only continuous variable with this property. Among discrete ones, the Shifted Geometric distribution also has this property. In fact, there is a close relationship between the two families of variables. In a sense, the Exponential distribution is a continuous analogue of the Shifted Geometric one, one measures time (continuously) until the next rare event, the other measures time (discretely) as the number of trials until the next success.