

Chapter 2

Elements of Probability Theory

In this chapter, we formally review some basic concepts of probability theory. Most of this material is standard and available in classical references, such as [108, 189, 319]; more advanced material on multivariate statistical analysis can be found in [22]. The definitions introduced here are instrumental to the study of randomized algorithms presented in subsequent chapters.

2.1 Probability, Random Variables and Random Matrices

2.1.1 Probability Space

Given a sample space Ω and a σ -algebra \mathcal{S} of subsets S of Ω (the events), a probability $\text{PR}\{S\}$ is a real-valued function on \mathcal{S} satisfying:

1. $\text{PR}\{S\} \in [0, 1]$;
2. $\text{PR}\{\Omega\} = 1$;
3. If the events S_i are mutually exclusive (i.e., $S_i \cap S_k = \emptyset$ for $i \neq k$), then

$$\text{PR}\left\{\bigcup_{i \in \mathcal{I}} S_i\right\} = \sum_{i \in \mathcal{I}} \text{PR}\{S_i\}$$

where \mathcal{I} is a countable¹ set of positive integers.

The triple $(\Omega, \mathcal{S}, \text{PR}\{S\})$ is called a *probability space*.

A *discrete probability space* is a probability space where Ω is countable. In this case, \mathcal{S} is given by subsets of Ω and the probability $\text{PR} : \Omega \rightarrow [0, 1]$ is such that

$$\sum_{\omega \in \Omega} \text{PR}\{\omega\} = 1.$$

¹By countable we mean finite (possibly empty) or countably infinite.

2.1.2 Real and Complex Random Variables

We denote with \mathbb{R} and \mathbb{C} the real and complex field respectively. The symbol \mathbb{F} is also used to indicate either \mathbb{R} or \mathbb{C} . A function $f : \Omega \rightarrow \mathbb{R}$ is said to be measurable with respect to a σ -algebra \mathcal{S} of subsets of Ω if $f^{-1}(A) \in \mathcal{S}$ for every Borel set $A \subseteq \mathbb{R}$.

A *real random variable* \mathbf{x} defined on a probability space $(\Omega, \mathcal{S}, \text{Pr}\{S\})$ is a measurable function mapping Ω into $\mathcal{Y} \subseteq \mathbb{R}$, and this is indicated with the shorthand notation $\mathbf{x} \in \mathcal{Y}$. The set \mathcal{Y} is called the range or *support* of the random variable \mathbf{x} . A *complex random variable* $\mathbf{x} \in \mathbb{C}$ is a sum $\mathbf{x} = \mathbf{x}_{\mathbb{R}} + j\mathbf{x}_{\mathbb{I}}$, where $\mathbf{x}_{\mathbb{R}} \in \mathbb{R}$ and $\mathbf{x}_{\mathbb{I}} \in \mathbb{R}$ are real random variables, and $j \doteq \sqrt{-1}$. If the random variable \mathbf{x} maps the sample space Ω into a subset $[a, b] \subset \mathbb{R}$, we write $\mathbf{x} \in [a, b]$. If Ω is a discrete probability space, then \mathbf{x} is a *discrete* random variable mapping Ω into a countable set.

Distribution and Density Functions The (cumulative) *distribution function* (cdf) of a random variable \mathbf{x} is defined as

$$F_{\mathbf{x}}(x) \doteq \text{Pr}\{\mathbf{x} \leq x\}.$$

The function $F_{\mathbf{x}}(x)$ is nondecreasing, right continuous (i.e., $F_{\mathbf{x}}(x) = \lim_{z \rightarrow x+} F_{\mathbf{x}}(z)$), and $F_{\mathbf{x}}(x) \rightarrow 0$ for $x \rightarrow -\infty$, $F_{\mathbf{x}}(x) \rightarrow 1$ for $x \rightarrow \infty$. Associated with the concept of distribution function, we define the α *percentile* of a random variable

$$x_{\alpha} = \inf\{x : F_{\mathbf{x}}(x) \geq \alpha\}.$$

For random variables of continuous type, if there exists a Lebesgue measurable function $f_{\mathbf{x}}(x) \geq 0$ such that

$$F_{\mathbf{x}}(x) = \int_{-\infty}^x f_{\mathbf{x}}(x) \, dx$$

then the cdf $F_{\mathbf{x}}(x)$ is said to be absolutely continuous, and

$$f_{\mathbf{x}}(x) = \frac{dF_{\mathbf{x}}(x)}{dx}$$

holds except possibly for a set of measure zero. The function $f_{\mathbf{x}}(x)$ is called the probability *density function* (pdf) of the random variable \mathbf{x} .

For discrete random variables, the cdf is a staircase function, i.e. $F_{\mathbf{x}}(x)$ is constant except at a countable number of points x_1, x_2, \dots having no finite limit point. The total probability is hence distributed among the “mass” points x_1, x_2, \dots at which the “jumps” of size

$$f_{\mathbf{x}}(x_i) \doteq \lim_{\epsilon \rightarrow 0} F_{\mathbf{x}}(x_i + \epsilon) - F_{\mathbf{x}}(x_i - \epsilon) = \text{Pr}\{\mathbf{x} = x_i\}$$

occur. The function $f_{\mathbf{x}}(x_i)$ is called the *mass density* of the discrete random variable \mathbf{x} . The definition of random variables is extended to real and complex random matrices in the next section.

2.1.3 Real and Complex Random Matrices

Given n random variables $\mathbf{x}_1, \dots, \mathbf{x}_n$, their *joint distribution* is defined as

$$F_{\mathbf{x}_1, \dots, \mathbf{x}_n}(x_1, \dots, x_n) \doteq \text{PR} \{ \mathbf{x}_1 \leq x_1, \dots, \mathbf{x}_n \leq x_n \}.$$

When the above distribution is absolutely continuous, we can define the joint density function $f_{\mathbf{x}_1, \dots, \mathbf{x}_n}(x_1, \dots, x_n)$

$$f_{\mathbf{x}_1, \dots, \mathbf{x}_n}(x_1, \dots, x_n) \doteq \frac{\partial^n F_{\mathbf{x}_1, \dots, \mathbf{x}_n}(x_1, \dots, x_n)}{\partial x_1 \cdots \partial x_n}.$$

The random variables $\mathbf{x}_1, \dots, \mathbf{x}_n$ are said to be *independent* if

$$F_{\mathbf{x}_1, \dots, \mathbf{x}_n}(x_1, \dots, x_n) = \prod_{i=1}^n F_{\mathbf{x}_i}(x_i)$$

where $F_{\mathbf{x}_i}(x_i) = \text{PR} \{ \mathbf{x}_i \leq x_i \}$.

A real random matrix $\mathbf{X} \in \mathbb{R}^{n,m}$ is a measurable function $\mathbf{X} : \Omega \rightarrow \mathcal{Y} \subseteq \mathbb{R}^{n,m}$. That is, the entries of \mathbf{X} are real random variables $[\mathbf{X}]_{i,k}$ for $i = 1, \dots, n$ and $k = 1, \dots, m$. A complex random matrix $\mathbf{X} \in \mathbb{C}^{n,m}$ is defined as the sum $\mathbf{X} = \mathbf{X}_{\mathbb{R}} + j\mathbf{X}_{\mathbb{I}}$, where $\mathbf{X}_{\mathbb{R}}$ and $\mathbf{X}_{\mathbb{I}}$ are real random matrices. A random matrix is *discrete* if its entries are discrete random variables.

The distribution function $F_{\mathbf{X}}(X)$ of a real random matrix \mathbf{X} is the joint cdf of the entries of \mathbf{X} . If \mathbf{X} is a complex random matrix, then its cdf is the joint cdf of $\mathbf{X}_{\mathbb{R}}$ and $\mathbf{X}_{\mathbb{I}}$. The pdf $f_{\mathbf{X}}(X)$ of a real or complex random matrix is analogously defined as the joint pdf of the real and imaginary parts of its entries. The notation $\mathbf{X} \sim f_{\mathbf{X}}(X)$ means that \mathbf{X} is a random matrix with probability density function $f_{\mathbf{X}}(X)$.

Let $\mathbf{X} \in \mathbb{F}^{n,m}$ be a real or complex random matrix (of continuous type) with pdf $f_{\mathbf{X}}(X)$ and support $\mathcal{Y} \subseteq \mathbb{F}^{n,m}$. Then, if $Y \subseteq \mathcal{Y}$, we have

$$\text{PR} \{ \mathbf{X} \in Y \} = \int_Y f_{\mathbf{X}}(X) dX.$$

Clearly, $\text{PR} \{ \mathbf{X} \in \mathcal{Y} \} = \int_{\mathcal{Y}} f_{\mathbf{X}}(X) dX = 1$. When needed, to further emphasize that the probability is relative to the random matrix \mathbf{X} , we explicitly write $\text{PR}_{\mathbf{X}} \{ \mathbf{X} \in Y \}$.

2.1.4 Expected Value and Covariance

Let $\mathbf{X} \in \mathcal{Y} \subseteq \mathbb{F}^{n,m}$ be a random matrix and let $J : \mathbb{F}^{n,m} \rightarrow \mathbb{R}^{p,q}$ be a Lebesgue measurable function. The *expected value* of the random matrix $J(\mathbf{X})$ is defined as

$$\text{E}_{\mathbf{X}}(J(\mathbf{X})) \doteq \int_{\mathcal{Y}} J(X) f_{\mathbf{X}}(X) dX$$

where \mathcal{Y} is the support of \mathbf{X} . We make use of the symbol $\text{E}_{\mathbf{X}}(J(\mathbf{X}))$ to emphasize the fact that the expected value is taken with respect to \mathbf{X} . The suffix is omitted when clear from the context.

If $\mathbf{X} \in \mathbb{R}^{n,m}$ is a discrete random matrix with countable support $\mathcal{Y} = \{X_1, X_2, \dots\}$, $X_i \in \mathbb{R}^{n,m}$ and $Y \subseteq \mathcal{Y}$, then

$$\text{PR}\{\mathbf{X} \in Y\} = \sum_{X_i \in Y} f_{\mathbf{X}}(X_i) = \sum_{X_i \in Y} \text{PR}\{\mathbf{X} = X_i\}.$$

The expected value of $J(\mathbf{X})$ is defined as

$$\mathbb{E}(J(\mathbf{X})) \doteq \sum_{X_i \in \mathcal{Y}} J(X_i) f_{\mathbf{X}}(X_i).$$

The expected value of $\mathbf{X} \in \mathbb{R}^{n,m}$ is usually called the *mean*. The *covariance matrix* of $\mathbf{x} \in \mathbb{R}^n$ is defined as

$$\text{Cov}(\mathbf{x}) \doteq \mathbb{E}_{\mathbf{x}}((\mathbf{x} - \mathbb{E}_{\mathbf{x}}(\mathbf{x}))^T (\mathbf{x} - \mathbb{E}_{\mathbf{x}}(\mathbf{x})))$$

where X^T denotes the transpose of X . The covariance of $\mathbf{x} \in \mathbb{R}$ is called the *variance* and is given by

$$\text{Var}(\mathbf{x}) \doteq \mathbb{E}_{\mathbf{x}}((\mathbf{x} - \mathbb{E}_{\mathbf{x}}(\mathbf{x}))^2).$$

The square root of the variance $(\text{Var}(\mathbf{x}))^{1/2}$ is called the *standard deviation*.

2.2 Marginal and Conditional Densities

Consider a random vector $\mathbf{x} = [\mathbf{x}_1 \cdots \mathbf{x}_n]^T \in \mathbb{R}^n$ with joint density function

$$f_{\mathbf{x}}(x) = f_{\mathbf{x}_1, \dots, \mathbf{x}_n}(x_1, \dots, x_n).$$

The *marginal density* of the first i components of the random vector $\mathbf{x} = [\mathbf{x}_1 \cdots \mathbf{x}_n]^T$ is defined as

$$f_{\mathbf{x}_1, \dots, \mathbf{x}_i}(x_1, \dots, x_i) \doteq \int \cdots \int f_{\mathbf{x}}(x_1, \dots, x_n) dx_{i+1} \cdots dx_n. \quad (2.1)$$

The *conditional density* $f_{\mathbf{x}_i | x_1, \dots, x_{i-1}}(x_i | x_1, \dots, x_{i-1})$ of the random variable \mathbf{x}_i conditioned to the event $\mathbf{x}_1 = x_1, \dots, \mathbf{x}_{i-1} = x_{i-1}$ is given by the ratio of marginal densities

$$f_{\mathbf{x}_i | x_1, \dots, x_{i-1}}(x_i | x_1, \dots, x_{i-1}) \doteq \frac{f_{\mathbf{x}_1, \dots, \mathbf{x}_i}(x_1, \dots, x_i)}{f_{\mathbf{x}_1, \dots, \mathbf{x}_{i-1}}(x_1, \dots, x_{i-1})}. \quad (2.2)$$

2.3 Univariate and Multivariate Density Functions

We next present a list of classical univariate and multivariate density functions. The reader is referred to Chap. 14 for numerical methods for generating random variables with the mentioned densities.

Binomial Density The binomial density with parameters n, p is defined as

$$b_{n,p}(x) \doteq \binom{n}{x} p^x (1-p)^{n-x}, \quad x \in \{0, 1, \dots, n\} \quad (2.3)$$

where $\binom{n}{x}$ indicates the binomial coefficient $\binom{n}{x} = \frac{n!}{x!(n-x)!}$. The binomial distribution is denoted as

$$B_{n,p}(x) \doteq \sum_{k=0}^x \binom{n}{k} p^k (1-p)^{n-k}, \quad x \in \{0, 1, \dots, n\}. \quad (2.4)$$

Normal Density The normal (Gaussian) density with mean $\bar{x} \in \mathbb{R}$ and variance $\sigma^2 \in \mathbb{R}$ is defined as

$$\mathcal{N}_{\bar{x}, \sigma^2}(x) \doteq \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2}(x-\bar{x})^2/\sigma^2}, \quad x \in \mathbb{R}. \quad (2.5)$$

Multivariate Normal Density The multivariate normal density with mean $\bar{x} \in \mathbb{R}^n$ and symmetric positive definite covariance matrix $W \in \mathbb{S}^n$, $W \succ 0$, is defined as

$$\mathcal{N}_{\bar{x}, W}(x) \doteq (2\pi)^{-n/2} |W|^{-1/2} e^{-\frac{1}{2}(x-\bar{x})^T W^{-1} (x-\bar{x})}, \quad x \in \mathbb{R}^n. \quad (2.6)$$

Uniform Density The uniform density on the interval $[a, b]$ is defined as

$$\mathcal{U}_{[a,b]}(x) \doteq \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b]; \\ 0 & \text{otherwise.} \end{cases} \quad (2.7)$$

Uniform Density over a Set Let S be a Lebesgue measurable set of nonzero volume (see Sect. 3.1.3 for a precise definition of volume). The uniform density over S is defined as

$$\mathcal{U}_S(X) \doteq \begin{cases} \frac{1}{\text{Vol}(S)} & \text{if } X \in S; \\ 0 & \text{otherwise.} \end{cases} \quad (2.8)$$

If instead S is a finite discrete set, i.e. it consists of a finite number of elements $S = \{X_1, X_2, \dots, X_N\}$, then the uniform density over S is defined as

$$\mathcal{U}_S(X) \doteq \begin{cases} \frac{1}{\text{Card}(S)} & \text{if } X \in S; \\ 0 & \text{otherwise} \end{cases}$$

where $\text{Card}(S)$ is the cardinality of S .

Chi-Square Density The unilateral chi-square density with $n > 0$ degrees of freedom is defined as

$$\chi_n^2(x) \doteq \frac{1}{\Gamma(n/2) 2^{n/2}} x^{n/2-1} e^{-x/2}, \quad x \in \mathbb{R}_+ \quad (2.9)$$

where $\Gamma(\cdot)$ is the Gamma function

$$\Gamma(x) \doteq \int_0^\infty \xi^{x-1} e^{-\xi} d\xi, \quad x > 0.$$

Weibull Density The Weibull density with parameter $a > 0$ is defined as

$$W_a(x) \doteq ax^{a-1}e^{-x^a}, \quad x \in \mathbb{R}. \quad (2.10)$$

Laplace Density The unilateral Laplace (or exponential) density with parameter $\lambda > 0$ is defined as

$$L_\lambda(x) \doteq \lambda e^{-\lambda x}, \quad x \in \mathbb{R}_+. \quad (2.11)$$

Gamma Density The unilateral Gamma density with parameters $a > 0, b > 0$ is defined as

$$G_{a,b}(x) \doteq \frac{1}{\Gamma(a)b^a} x^{a-1} e^{-x/b}, \quad x \in \mathbb{R}_+. \quad (2.12)$$

Generalized Gamma Density The unilateral generalized Gamma density with parameters $a > 0, c > 0$ is defined as

$$\overline{G}_{a,c}(x) \doteq \frac{c}{\Gamma(a)} x^{ca-1} e^{-x^c}, \quad x \in \mathbb{R}_+. \quad (2.13)$$

2.4 Convergence of Random Variables

We now recall the formal definitions of convergence almost everywhere (or almost sure convergence), convergence in the mean square sense and convergence in probability. Other convergence concepts not discussed here include vague convergence, convergence of moments and convergence in distribution, see e.g. [108].

Definition 2.1 (Convergence almost everywhere) A sequence of random variables $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots$ converges *almost everywhere* (a.e.) (or with probability one) to the random variable \mathbf{x} if

$$\text{PR} \left\{ \lim_{N \rightarrow \infty} \mathbf{x}^{(N)} = \mathbf{x} \right\} = 1.$$

Definition 2.2 (Convergence in the mean square sense) A sequence of random variables $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots$ converges *in the mean square sense* to the random variable \mathbf{x} if

$$\lim_{N \rightarrow \infty} \text{E}(|\mathbf{x} - \mathbf{x}^{(N)}|^2) = 0.$$

Definition 2.3 (Convergence in probability) A sequence of random variables $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots$ converges *in probability* to the random variable \mathbf{x} if, for any $\epsilon > 0$, we have

$$\lim_{N \rightarrow \infty} \text{PR} \{ |\mathbf{x} - \mathbf{x}^{(N)}| > \epsilon \} = 0.$$

Convergence a.e. and convergence in the mean square sense both imply convergence in probability, while there is no implicative relationship between convergence a.e. and convergence in the mean square sense.

Randomized Algorithms for Analysis and Control of
Uncertain Systems

With Applications

Tempo, R.; Calafiore, G.; Dabbene, F.

2013, XXIV, 360 p., Hardcover

ISBN: 978-1-4471-4609-4