Basic Probability

Sample Space and Sigma-Field [1, §1.2.5]. The *sample space* Ω is the set of all *outcomes*, or *elementary events*, of a random experiment. The *power set* of Ω contains all subsets of Ω and is written as $\{0,1\}^{\Omega}$. A collection $\mathcal{F}$ of subsets of Ω is called a *σ-field* if it satisfies the following conditions

- $\emptyset \in \mathcal{F}$.
- If $A_1, \dots A_n \in \mathcal{F}$ then $\bigcup_{i=1}^{n} A_i \in \mathcal{F}$, where $n = 1, 2, \dots$ can be infinite.
- If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$.

A is called an *event*; σ-fields are closed under countable intersections.

Probability Space [1, §1.3.1]. A *probability measure* $\mathbb{P}$ on $(\Omega, \mathcal{F})$ is a function $\mathbb{P} : \mathcal{F} \to [0,1]$ satisfying

- $\mathbb{P}(\emptyset) = 0, \quad \mathbb{P}(\Omega) = 1$.
- If $A_1, A_2, \dots$ is a collection of disjoint members of $\mathcal{F}$, in that $A_i \cap A_j = \emptyset$ for all pairs i, j with $i \neq j$, then
$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *probability space*.

- An event A is called *null* if $\mathbb{P}(A) = 0$.
- An event B is said to occur *almost surely* if $\mathbb{P}(B) = 1$.
- A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called *complete* if all subsets of null sets, i.e., events of zero probability, are events themselves.

Properties of a Probability Space [1, §1.3].

- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
- If $B \supseteq A$ then $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A)$.
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(B \cap A)$.
- Let $A_1, A_2, \dots$ be an increasing sequence of events, so that $A_1 \subseteq A_2 \subseteq \cdots$, and write for their limit
$$A = \bigcup_{i=1}^{\infty} A_i = \lim_{i\to\infty} A_i.$$
The $\mathbb{P}(A) = \lim_{i\to\infty} \mathbb{P}(A_i)$. The same holds for a decreasing sequence of events and their intersection.

Conditional Probability [1, §1.4].

- If $\mathbb{P}(B) > 0$ then the *conditional probability* that A occurs given that B occurs is defined to be
$$\mathbb{P}(A \,|\, B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$
- For any events A and B such that $0 < \mathbb{P}(B) < 1$,
$$\mathbb{P}(A) = \mathbb{P}(A \,|\, B)\,\mathbb{P}(B) + \mathbb{P}(A \,|\, B^c)\,\mathbb{P}(B^c)$$
- More generally, let $B_1, B_2, \dots, B_n$ be a partitioning of Ω such that $\mathbb{P}(B_i) > 0$ for all i. Then
$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \,|\, B_i)\,\mathbb{P}(B_i).$$
- *Bayes Rule*: let B_i and A be as before, $\mathbb{P}(A) > 0$, then
$$\mathbb{P}(B_i \,|\, A) = \frac{\mathbb{P}(A \,|\, B_i)\,\mathbb{P}(B_i)}{\sum_{i=1}^{n} \mathbb{P}(A \,|\, B_i)\,\mathbb{P}(B_i)}.$$

Independence [1, §1.5]. A family $\{A_i : i \in \mathcal{I}\}$ is called *independent* if
$$\mathbb{P}\left(\bigcap_{i \in \mathcal{J}} A_i\right) = \prod_{i \in \mathcal{J}} \mathbb{P}(A_i)$$
for all finite subsets $\mathcal{J}$ of $\mathcal{I}$.

Random Variables

Basics

Random Variables and Distribution Functions [1, §2.1]. A *random variable* (RV) is a function $X : \Omega \to \mathbb{R}$ with the property that $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$ for each $x \in \mathbb{R}$. Such a function is said to be *$\mathcal{F}$-measurable*. The *(cumulative) distribution function* (CDF) of a RV X is the function $F_X : \mathbb{R} \to [0,1]$ given by $F_X(x) := \mathbb{P}(X \leq x)$. A distribution function has the following properties

- $\lim_{x\to-\infty} F_X(x) = 0, \quad \lim_{x\to\infty} F_X(x) = 1$.
- If $x < y$ then $F_X(x) \leq F_X(y)$.
- The CDF F_X is right-continuous, that is, $F_X(x+h) \to F_X(x)$ as $h \downarrow 0$.
- $\mathbb{P}(X > x) = 1 - F_X(x)$.
- $\mathbb{P}(x < X \leq y) = F_X(y) - F_X(x)$.
- $\mathbb{P}(X = x) = F_X(x) - \lim_{y\uparrow x} F_X(y)$.

The RV X is called *continuous* [2, §4.2] if its CDF F_X is continuous; in that case, $F_X(x^-) = F_X(x)\ \forall x$, and $\mathbb{P}(X = x) = 0$. It is *discrete* if it takes values in some countable subset $\{x_1, x_2, \dots\}$ of $\mathbb{R}$; in that case, $F_X(x)$ is constant except for a finite number of jump discontinuities, and $\mathbb{P}(X = x) = F_X(x) - F_X(x^-)$. It is of *mixed type* if $F_X(x)$ is piecewise continuous with a finite number of jump discontinuities.

The *indicator function* $I_A : \Omega \to \mathbb{R}$ is defined as the binary RV
$$I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \in A^c. \end{cases}$$

Independence [1, §4.2]. Random variables X and Y (discrete or continuous) are called *independent* if $\{X \leq x\}$ and $\{Y \leq y\}$ are independent events for all $x, y \in \mathbb{R}$, i.e., $F_{X,Y}(x,y) = F_X(x)F_Y(y)$ for all $x, y \in \mathbb{R}$.
Let $g, h : \mathbb{R} \to \mathbb{R}$. The functions $g(X)$ and $h(Y)$ map Ω into $\mathbb{R}$. Suppose that $g(X)$ and $h(Y)$ are random variables, i.e. they are $\mathcal{F}$-measurable. If X and Y are independent, then so are $g(X)$ and $h(Y)$.

Random Vectors [1, §2.5]. The *joint distribution function* of a random real-valued vector $\boldsymbol{X} := [X_1\, X_2\, \dots\, X_N]^T$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is the function $F_{\boldsymbol{X}} : \mathbb{R}^N \to [0,1]$ given by $F_{\boldsymbol{X}}(\mathbf{x}) := \mathbb{P}(\boldsymbol{X} \leq \mathbf{x})$ for $\mathbf{x} \in \mathbb{R}^N$, where the ordering $\mathbf{x} \leq \mathbf{y}$ means that $x_i \leq y_i$ for each $i = 1, 2, \dots$.
The joint distribution function $F_{X,Y}$ of the random vector $[X\, Y]$ has the following properties, which hold analogously for N-dimensional random vectors:

- $\lim_{x,y\to-\infty} F_{X,Y}(x,y) = 0, \ \lim_{x,y\to\infty} F_{X,Y}(x,y) = 1$.
- If $[x_1\ y_1] \leq [x_2\ y_2]$ then $F_{X,Y}(x_1, y_1) \leq F_{X,Y}(x_2, y_2)$.
- $F_{X,Y}$ is continuous from above in that
$$F_{X,Y}(x+u, y+v) \to F_{X,Y}(x,y)$$
as $u, v \downarrow 0$.
- The *marginal distribution functions* of X and Y are
$$\lim_{y\to\infty} F_{X,Y}(x,y) = F_X(x),$$
$$\lim_{x\to\infty} F_{X,Y}(x,y) = F_Y(y).$$
written as $F_X(x) = F_{X,Y}(x,\infty)$ and $F_Y(y) = F_{X,Y}(\infty, y)$, respectively.

The definitions of *discrete*, *continuous*, and *mixed* RVs extend to random vectors.

Relationship Between Real-Valued and Complex-Valued Operations [3, §5]. A *Complex RV* $U = X + jY$ can be treated as a random vector $[X\, Y]$. Consider arbitrary complex vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^N$ and a complex $M \times N$ matrix $\mathbf{A} \in \mathbb{C}^{M\times N}$. Define $\mathbf{u}_R := \Re\{\mathbf{u}\}$ and $\mathbf{u}_I := \Im\{\mathbf{u}\}$ and the real $2N$-dimensional vector
$$\underline{\mathbf{u}} = \begin{bmatrix} \mathbf{u}_R \\ \mathbf{u}_I \end{bmatrix} = \begin{bmatrix} \Re\{\mathbf{u}\} \\ \Im\{\mathbf{u}\} \end{bmatrix}.$$
Then the complex-valued linear operation $\mathbf{v} = \mathbf{A}\mathbf{u}$ can be expressed in terms of the real quantities as $\underline{\mathbf{v}} = \underline{\mathbf{A}}\,\underline{\mathbf{u}}$, where a matrix $\underline{\mathbf{A}}$ satisfying $\underline{\mathbf{v}} = \underline{\mathbf{A}}\,\underline{\mathbf{u}}$ exists and is given by
$$\underline{\mathbf{A}} = \begin{bmatrix} \mathbf{A}_R & -\mathbf{A}_I \\ \mathbf{A}_I & \mathbf{A}_R \end{bmatrix} = \begin{bmatrix} \Re\{\mathbf{A}\} & -\Im\{\mathbf{A}\} \\ \Im\{\mathbf{A}\} & \Re\{\mathbf{A}\} \end{bmatrix}.$$
Let $\mathbf{B}$ be another complex matrix, then

- $\underline{\mathbf{AB}} = \underline{\mathbf{A}}\,\underline{\mathbf{B}}$.
- $\underline{\mathbf{A}+\mathbf{B}} = \underline{\mathbf{A}} + \underline{\mathbf{B}}$.
- $\underline{\mathbf{A}^H} = \underline{\mathbf{A}}^T$.
- $\underline{\mathbf{A}^{-1}} = \underline{\mathbf{A}}^{-1}$.
- $\det(\underline{\mathbf{A}}) = |\det \mathbf{A}|^2 = \det(\mathbf{A}\mathbf{A}^H)$.
- $\underline{\mathbf{u}+\mathbf{v}} = \underline{\mathbf{u}} + \underline{\mathbf{v}}$.
- $\underline{\mathbf{Au}} = \underline{\mathbf{A}}\,\underline{\mathbf{u}}$.
- $\Re\left\{\mathbf{u}^H\mathbf{v}\right\} = \underline{\mathbf{u}}^T\underline{\mathbf{v}}$.
- If $\mathbf{A} \in \mathbb{C}^{N\times N}$ is unitary, then $\underline{\mathbf{A}} \in \mathbb{R}^{2N\times 2N}$ is orthogonal.
- If $\mathbf{A} \in \mathbb{C}^{N\times N}$ is positive semidefinite, then so is $\underline{\mathbf{A}} \in \mathbb{R}^{2N\times 2N}$; moreover, $\mathbf{u}^H\mathbf{A}\mathbf{u} = \underline{\mathbf{u}}^T\underline{\mathbf{A}}\,\underline{\mathbf{u}}$.

Discrete Random Variables

Discrete Random Variables [1, §3.1–§3.2]. The *probability mass function* (PMF) of a discrete RV X is the function $f : \mathbb{R} \to [0,1]$ given by $f_X(x) = \mathbb{P}(X = x)$. The joint PMF of a random vector $\boldsymbol{X}$ is defined analogously. The PMF of a discrete RV satisfies

- The set of x such that $f_X(x) \neq 0$ is countable.
- $\sum_i f_X(x_i) = 1$.

Discrete RVs $X_1, X_2, \dots X_N$ are *independent* if the events $\{X_1 = x_1\}, \{X_2 = x_2\}, \dots, \{X_N = x_N\}$ are independent for all $x_1, x_2, \dots, x_N$.

- If X and Y are independent and $g, h : \mathbb{R} \to \mathbb{R}$, then $g(X)$ and $h(X)$ are independent also.
- $X_1, X_2, \dots, X_N$ are independent iff $f_{X_1,X_2,\dots,X_N}(x_1, x_2, \dots, x_N) = f_{X_1}(x_1)f_{X_2}(x_2)\dots f_{X_N}(x_N)$ for all $x_1, x_2, \dots, x_N \in \mathbb{R}$.

Expectation [1, §3.3]. The *expectation* of the RV X with PMF f_X is defined as
$$\mathbb{E}[X] := \sum_{x: f_X(x)>0} x f_X(x)$$
whenever the sum is absolutely convergent. The expectation of random vectors is defined element wise.

- If $\boldsymbol{X}$ has PMF $f_{\boldsymbol{X}}$ and $g : \mathbb{R}^N \to \mathbb{R}$, then
$$\mathbb{E}[g(\boldsymbol{X})] = \sum_{\mathbf{x}} g(\mathbf{x}) f_{\boldsymbol{X}}(\mathbf{x})$$
whenever the sum is element-wise absolutely convergent.
- If $X \geq 0$ then $\mathbb{E}[X] \geq 0$.
- If $a, b \in \mathbb{R}$ then $\mathbb{E}[a\boldsymbol{X} + b\boldsymbol{Y}] = a\,\mathbb{E}[\boldsymbol{X}] + b\,\mathbb{E}[\boldsymbol{Y}]$ (*linearity*).
- The random variable 1, taking the value 1 always, has expectation $\mathbb{E}[1] = 1$.
- If X and Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$.

Sums of discrete RVs [1, §3.8]. The probability of the sum of two RVs X and Y having joint PMF $f_{X,Y}$ is given by
$$\mathbb{P}(X + Y = z) = \sum_{x} f_{X,Y}(x, z - x).$$
If X and Y are independent, then
$$f_{X+Y}(z) = \mathbb{P}(X + Y = z) = \sum_{x} f_X(x) f_Y(z - x).$$

Continuous Random Variables

Density Functions [1, §4.1,§4.5]. If X is a continuous RV, its CDF $F_X = \mathbb{P}(X \leq x)$ can be expressed as
$$F_X(x) = \int_{-\infty}^{x} f_X(\theta)\,d\theta.$$
The function f_X is called the *(probability) density function* of the continuous RV X.

- $\int_{-\infty}^{\infty} f_X(x)dx = 1$.
- $\mathbb{P}(X = x) = 0$ for all $x \in \mathbb{R}$.
- $\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)dx$.

The random variables X and Y are *jointly continuous* with *joint PDF* $f_{X,Y} : \mathbb{R}^2 \to [0, \infty)$ if
$$F_{X,Y}(x,y) = \int_{\theta=-\infty}^{y} \int_{\phi=-\infty}^{x} f_{X,Y}(\theta, \phi)\,d\theta\,d\phi$$
for each $x, y \in \mathbb{R}$. If $F_{X,Y}$ is sufficiently differentiable then
$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y).$$
The *marginal densities* are given as
$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$
$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx.$$
For continuous RVs, *independence* is equivalent to requiring that $f_{X,Y} = f_X(x)f_Y(y)$ whenever $F_{X,Y}$ is differentiable at (x,y).
The above properties hold analogously for higher-dimensional continuous random vectors.

Conditioning [1, §4.6]. The probability $\mathbb{P}(X \leq x \,|\, Y = y)$ is undefined because $\mathbb{P}(Y = y) = 0$ for continuous RVs. Hence, conditioning has to be understood as the limit of $\mathbb{P}(X \leq x \,|\, y \leq Y \leq y + dy)$ for $dy \downarrow 0$. The *conditional distribution function* of X given $Y = y$ is then defined as
$$F_{X\,|\,Y}(x \,|\, y) = \int_{-\infty}^{x} \frac{f_{X,Y}(\phi, y)}{f_Y(y)} d\phi$$
for any y such that $f_Y(y) > 0$. The *conditional density function* is given by
$$f_{X\,|\,Y}(x \,|\, y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$
for any y such that $f_Y(y) > 0$.

Expectation [1, §4.3]. The *expectation* of a continuous RV X is
$$\mathbb{E}[X] := \int_{-\infty}^{\infty} x f_X(x)dx$$
whenever the integral exists.

- If X and $g(X)$ are continuous random variables then
$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)dx.$$
- If X has PDF f_X with $f_X(x) = 0$ when $x < 0$, and distribution function F_X, then
$$\mathbb{E}[X] = \int_{0}^{\infty} (1 - F_X(x))dx.$$
- If $g : \mathbb{R}^N \to \mathbb{R}$ is an $\mathcal{F}$-measurable function then
$$\mathbb{E}[g(X_1, X_2, \dots, X_N)] =$$
$$\int \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \dots, x_N)$$
$$\times f_{X_1,X_2,\dots,X_N}(x_1, x_2, \dots, x_N)dx_1dx_2\dots dx_N$$
$$= \int_{-\infty}^{\infty} g(\mathbf{x}) f_{\boldsymbol{X}}(\mathbf{x})d\mathbf{x}.$$
- The expectation is linear: $\mathbb{E}[aX + bY] = a\,\mathbb{E}[X] + b\,\mathbb{E}[Y]$, for all $a, b \in \mathbb{C}$.

Functions of Random Variables [1, §4.7, §4.8]. Let X_1 and X_2 be RVs with joint density function f_{X_1,X_2}, and let $T : (x_1, x_2) \to (y_1, y_2)$ be a one-to-one mapping taking some domain $\mathcal{D} \subseteq \mathbb{R}^2$ onto some range $\mathcal{R} \subseteq \mathbb{R}^2$. If $g : \mathbb{R}^2 \to \mathbb{R}$ and T maps the set $\mathcal{A} \subseteq \mathcal{D}$ *onto* the set $\mathcal{B} \subseteq \mathcal{R}$, then
$$\iint_{\mathcal{A}} g(x_1, x_2)dx_1dx_2$$
$$= \iint_{\mathcal{B}} g\left(x_1(y_1, y_2), x_2(y_1, y_2)\right)$$
$$\times \left|J(y_1, y_2)\right| dy_1dy_2,$$
where J denotes the *Jacobian* of the transform
$$J(y_1, y_2) = \det \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_2}{\partial y_1} \\ \frac{\partial x_1}{\partial y_2} & \frac{\partial x_2}{\partial y_2} \end{bmatrix}.$$
Then the pair Y_1, Y_2, given by $(Y_1, Y_2) = T(X_1, X_2)$ has joint density function
$$f_{Y_1,Y_2}(y_1, y_2) =$$
$$f_{X_1,X_2}\left(x_1(y_1, y_2), x_2(y_1, y_2)\right)\left|J(y_1, y_2)\right|$$
if $(y_1, y_2) \in \mathcal{R}(T)$ and 0 otherwise.
If the transformation is not one-to-one but piecewise one-to-one and sufficiently smooth, the more general transformation rule is the following: Let $\mathcal{I}_1, \dots, \mathcal{I}_N$ be intervals which partition $\mathbb{R}$, and suppose $Y = g(X)$ is strictly monotone and continuously differentiable on every $\mathcal{I}_n$. For each n, the function $g : \mathcal{I}_n \to \mathbb{R}$ is invertible on $g(\mathcal{I}_n)$ and we write h_n for the inverse function. Then,
$$f_Y(y) = \sum_{n=1}^{N} f_X\left(h_n(y)\right)\left|h_n'(y)\right|,$$
with the convention that the nth summand is 0 if h_n is not defined at y.
If X and Y have joint density function $f_{X,Y}$, then the *sum of RVs* $X + Y$ has density function
$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z - x)dx.$$
If X and Y are independent, this simplifies to
$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x)dx.$$

Inverse Transform [1, §4.11.1]. Let F be a distribution function and let Y be uniformly distributed on the interval $[0,1]$.

- If F is a continuous function, the RV $X = F^{-1}(Y)$ has distribution function F.
- Let F be the distribution function of a RV taking on non-negative integer values. The RV X given by $X = k$ iff $F(k-1) < U < F(k)$ has distribution function F.

Order Statistics *[2, §7.1]*. Consider the N-dimensional random vector $\boldsymbol{X} = [X_1 X_2 \dots X_N]^T$. Ordering the elements of the vector for each outcome from smallest to largest yields a new random vector $\boldsymbol{Y}$. The kth element of $\boldsymbol{Y}$ is called the *kth-order statistic*. If the X_n are i.i.d. with CDF F_X and PDF f_X, then the PDF of the kth statistic Y_k is given as

$$f_k(y) = \frac{N!}{(k-1)!(N-k)!} F_X^{k-1}(y) \times \left(1 - F_X(y)\right)^{N-k} f_X(y).$$

Variance and Covariance

The following properties hold for both discrete and continuous random variables and vectors.

Moments *[1, §3.3]*. If $k \in \mathbb{N}$, the *kth moment* of the real RV X is defined as

$$m_k = \mathbb{E}\left[X^k\right].$$

The kth *central moment* is

$$\sigma_k := \mathbb{E}\left[(X - m_k)^k\right].$$

As a special case, the *variance* is defined as

$$\mathbb{V}\mathrm{ar}[X] := \sigma_2 = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right],$$

also denoted by $\mathbb{V}\mathrm{ar}[X] = \sigma^2$. The following properties hold:

- $\mathbb{V}\mathrm{ar}[X] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$.
- $\mathbb{V}\mathrm{ar}[aX] = a^2\, \mathbb{V}\mathrm{ar}[X]$, for $a \in \mathbb{R}$.

Conditional Expectation *[1, §3.7]*. Let $\Psi(Y) = \mathbb{E}[X \mid Y = y]$. Then $\Psi(Y)$ is called the *conditional expectation* of X given Y, written as $\mathbb{E}[X \mid Y]$. Conditioning for continuous random variables always has to be understood in the limit $dy \to 0$. The *conditional variance* is defined as $\mathbb{V}\mathrm{ar}[X \mid Y] := \mathbb{E}\left[(X - \mathbb{E}[X \mid Y])^2 \mid Y\right]$.

The conditional expectation satisfies

- $\mathbb{E}_X[\mathbb{E}_Y[Y \mid X]] = \mathbb{E}[Y]$.
- $\mathbb{E}_X[\mathbb{E}_Y[Y \mid X] g(X)] = \mathbb{E}[Y g(X)]$.
- $\mathbb{E}_{X_2}[\mathbb{E}_Y[Y \mid X_1, X_2] \mid X_1] = \mathbb{E}[Y \mid X_1]$.
- $\mathbb{E}_Y[Y g(X) \mid X]] = g(X)\, \mathbb{E}_Y[Y \mid X]$ for any suitable function $g(x)$.

The conditional variance satisfies

$$\mathbb{V}\mathrm{ar}[Y] = \mathbb{E}_X[\mathbb{V}\mathrm{ar}[Y \mid X]] + \mathbb{V}\mathrm{ar}[\mathbb{E}_Y[Y \mid X]].$$

Covariance *[1, §3.6]*. The *covariance* of the RVs X and Y is defined as

$$\mathbb{C}\mathrm{ov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

and the *correlation coefficient* as

$$\rho(X, Y) := \frac{\mathbb{C}\mathrm{ov}[X, Y]}{\sqrt{\mathbb{V}\mathrm{ar}[X]\, \mathbb{V}\mathrm{ar}[Y]}}$$

as long as the variances are non-zero.

- $\mathbb{C}\mathrm{ov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$.
- X and Y are called *uncorrelated* if $\mathbb{C}\mathrm{ov}[X, Y] = 0$.
- The correlation coefficient ρ satisfies $\left|\rho(X, Y)\right| \le 1$, with equality iff $\mathbb{P}(aX + bY = c) = 1$ fore some $a, b, c \in \mathbb{R}$.
- If X and Y are uncorrelated, then $\mathbb{V}\mathrm{ar}[X + Y] = \mathbb{V}\mathrm{ar}[X] + \mathbb{V}\mathrm{ar}[Y]$.

Covariance Matrices *[4, §2]*. The *covariance matrix* of a random vector $\boldsymbol{X}$ is defined as

$$\mathbf{K}_{\boldsymbol{X}} = \mathbb{E}\left[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^T\right].$$

Properties of covariance matrices:

- $\mathbf{K}_{\boldsymbol{X}}$ is symmetric.
- $\mathbf{K}_{\boldsymbol{X}}$ positive semidefinite, i.e., $\mathbf{a}^T \mathbf{K}_{\boldsymbol{X}} \mathbf{a} \ge 0$ for every vector $\mathbf{a}$, its eigenvalues are nonnegative, and it can be written as $\mathbf{K}_{\boldsymbol{X}} = \mathbf{A}^T\mathbf{A}$ for some matrix $\mathbf{A}$.
- The elements of $\boldsymbol{X}$ are *uncorrelated* if the covariance matrix is diagonal.
- Every positive semidefinite matrix is a covariance matrix.

The *cross-covariance* matrix between the random vectors X and Y is defined as

$$\mathbf{K}_{\boldsymbol{XY}} = \mathbb{E}\left[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{Y} - \mathbb{E}[\boldsymbol{Y}])^T\right],$$

the *correlation matrix* is defined as $\mathbf{R}_{\boldsymbol{X}} = \mathbb{E}\left[\boldsymbol{X}\boldsymbol{X}^T\right]$.

- $\mathbf{K}_{\boldsymbol{X}} = \mathbf{R}_{\boldsymbol{X}} - \mathbb{E}[\boldsymbol{X}]\,\mathbb{E}[\boldsymbol{X}]^T$.

Complex Extension

Complex-Valued Random Vectors *[3, §2]*. Let $\boldsymbol{U} = \boldsymbol{U}_R + i\boldsymbol{U}_I$ and $\boldsymbol{V} = \boldsymbol{V}_R + i\boldsymbol{V}_I$ be complex random vectors. The *expectation* is given as

$$\mathbb{E}[\boldsymbol{U}] = \mathbb{E}[\boldsymbol{U}_R] + i\,\mathbb{E}[\boldsymbol{U}_I].$$

The second order statistics are completely characterized either by the four real-valued covariance matrices

$$\mathbf{K}_{\boldsymbol{U}_R\boldsymbol{V}_R} := \mathbb{C}\mathrm{ov}[\boldsymbol{U}_R, \boldsymbol{V}_R],$$
$$\mathbf{K}_{\boldsymbol{U}_R\boldsymbol{V}_I} := \mathbb{C}\mathrm{ov}[\boldsymbol{U}_R, \boldsymbol{V}_I],$$
$$\mathbf{K}_{\boldsymbol{U}_I\boldsymbol{V}_R} := \mathbb{C}\mathrm{ov}[\boldsymbol{U}_I, \boldsymbol{V}_R],$$
$$\mathbf{K}_{\boldsymbol{U}_I\boldsymbol{V}_I} := \mathbb{C}\mathrm{ov}[\boldsymbol{U}_I, \boldsymbol{V}_I],$$

or the two complex-valued covariance matrices

$$\mathbf{K}_{\boldsymbol{UV}} := \mathbb{E}\left[\left(\boldsymbol{U} - \mathbb{E}[\boldsymbol{U}]\right)\left(\boldsymbol{V} - \mathbb{E}[\boldsymbol{V}]\right)^H\right],$$
$$\mathbf{J}_{\boldsymbol{UV}} := \mathbb{E}\left[\left(\boldsymbol{U} - \mathbb{E}[\boldsymbol{U}]\right)\left(\boldsymbol{V} - \mathbb{E}[\boldsymbol{V}]\right)^T\right].$$

$\mathbf{K}_{\boldsymbol{UV}}$ is called *covariance matrix*, while $\mathbf{J}_{\boldsymbol{UV}}$ is referred to as *pseudo-covariance matrix*. The following relations hold:

$$\mathbf{K}_{\boldsymbol{UV}} = \mathbf{K}_{\boldsymbol{U}_R\boldsymbol{V}_R} + \mathbf{K}_{\boldsymbol{U}_I\boldsymbol{V}_I} + i(\mathbf{K}_{\boldsymbol{U}_I\boldsymbol{V}_R} - \mathbf{K}_{\boldsymbol{U}_R\boldsymbol{V}_I}),$$
$$\mathbf{J}_{\boldsymbol{UV}} = \mathbf{K}_{\boldsymbol{U}_R\boldsymbol{V}_R} - \mathbf{K}_{\boldsymbol{U}_I\boldsymbol{V}_I} + i(\mathbf{K}_{\boldsymbol{U}_I\boldsymbol{V}_R} + \mathbf{K}_{\boldsymbol{U}_R\boldsymbol{V}_I}),$$

and

$$\mathbf{K}_{\boldsymbol{U}_R\boldsymbol{V}_R} = \tfrac{1}{2}\Re\{\mathbf{K}_{\boldsymbol{UV}} + \mathbf{J}_{\boldsymbol{UV}}\},$$
$$\mathbf{K}_{\boldsymbol{U}_I\boldsymbol{V}_I} = \tfrac{1}{2}\Re\{\mathbf{K}_{\boldsymbol{UV}} - \mathbf{J}_{\boldsymbol{UV}}\},$$
$$\mathbf{K}_{\boldsymbol{U}_I\boldsymbol{V}_R} = \tfrac{1}{2}\Im\{\mathbf{K}_{\boldsymbol{UV}} + \mathbf{J}_{\boldsymbol{UV}}\},$$
$$\mathbf{K}_{\boldsymbol{U}_R\boldsymbol{V}_I} = \tfrac{1}{2}\Im\{-\mathbf{K}_{\boldsymbol{UV}} + \mathbf{J}_{\boldsymbol{UV}}\}.$$

$\boldsymbol{U}$ and $\boldsymbol{V}$ are said to be *uncorrelated* if the four real covariance matrices above vanish. It follows that they are uncorrelated iff $\mathbf{K}_{\boldsymbol{UV}} = \mathbf{J}_{\boldsymbol{UV}} = \mathbf{0}$.

Complex Covariance Matrices.

- $\mathbf{K}_{\boldsymbol{U}} = \mathbf{K}_{\boldsymbol{UU}} = \mathbb{E}\left[\boldsymbol{U}\boldsymbol{U}^H\right] - \mathbb{E}[\boldsymbol{U}]\,\mathbb{E}[\boldsymbol{U}]^H$.
- $\mathbb{V}\mathrm{ar}[aU] = |a|^2\, \mathbb{V}\mathrm{ar}[U]$, for $a \in \mathbb{C}$.
- $\mathbb{E}\left[U_1 U_2^\star\right] = \mathbb{E}\left[U_2 U_1^\star\right]^\star \implies \mathbb{C}\mathrm{ov}[U_1, U_2] = \mathbb{C}\mathrm{ov}[U_2, U_1]^\star$.

Proper Complex Random Vectors *[3, §4]*. A complex random vector $\boldsymbol{U}$ is called *proper* if its pseudo-covariance $\mathbf{J}_{\boldsymbol{U}}$ vanishes. The complex vectors $\boldsymbol{U}$ and $\boldsymbol{V}$ are called *jointly proper* if the composite random vector $[\boldsymbol{U}^T\, \boldsymbol{V}^T]^T$ is proper.

- Any subvector of a proper random vector is also proper.
- Two jointly proper complex random vectors $\boldsymbol{U}$ and $\boldsymbol{V}$ are *uncorrelated* iff their covariance matrix $\mathbf{K}_{\boldsymbol{UV}}$ vanishes.
- A complex random vector $\boldsymbol{U}$ is proper iff $\mathbf{K}_{\boldsymbol{U}_R} = \mathbf{K}_{\boldsymbol{U}_I}$ and $\mathbf{K}_{\boldsymbol{U}_I\boldsymbol{U}_R} = -\mathbf{K}_{\boldsymbol{U}_I\boldsymbol{U}_R}^T$. Hence, $\mathbf{K}_{\boldsymbol{U}_I\boldsymbol{U}_R}$ is zero on the main diagonal; thus the real and imaginary parts of each component of $\boldsymbol{U}$ are uncorrelated.
- A real random vector is proper iff it is constant.
- Any random vector $\boldsymbol{V}$ obtained from $\boldsymbol{U}$ by an affine transformation is also proper, i.e., $\boldsymbol{V} = \mathbf{A}\boldsymbol{U} + \mathbf{b}$ is proper for all $\mathbf{A} \in \mathbb{C}^{M \times N}$ and $\mathbf{b} \in \mathbb{C}^M$. Then, $\boldsymbol{U}$ and $\boldsymbol{V}$ are jointly proper.
- Let $\boldsymbol{U}$ and $\boldsymbol{V}$ be two independent complex random vectors, and let $\boldsymbol{U}$ be proper. Then the linear combination $a\boldsymbol{U} + b\boldsymbol{V}$, with $a, b \in \mathbb{C}, b \neq 0$ is proper iff $\boldsymbol{V}$ is also proper.

Characteristic Functions

Characteristic Function *[2, §5.5]*. The *characteristic function* c_X of a RV X is defined as

$$c_X(\omega) = \int_{-\infty}^{\infty} f_X(x) e^{j\omega x} dx = \mathbb{E}\left[e^{j\omega X}\right].$$

- $|c_X(\omega)| \le c_X(0) = 1$
- $c_X(\omega)$ is uniformly continuous on $\mathbb{R}$.
- $c_X(\omega)$ is nonnegative definite, i.e., $\sum_{j,k} c_X(\omega_i - \omega_j) a_i a_j^\star \ge 0$ for all real $\omega_1, \omega_2, \dots, \omega_N$ and complex $a_1, a_2, \dots, a_N$.
- For $Y = aX + b$ and $a, b \in \mathbb{R}$: $c_Y(\omega) = e^{jb\omega} c_X(a\omega)$.
- If X and Y are independent, then $c_{X+Y}(\omega) = c_X(\omega) c_Y(\omega)$.
- RVs X and Y are independent iff $c_{X,Y}(\omega, \varphi) = c_X(\omega) c_Y(\varphi)$.

Moment Generating Function *[1, §5.7]*. The *moment generating function (MGF)* g_X of a RV X is defined as

$$g_X(s) = \int_{-\infty}^{\infty} f_X e^{sx} dx = \mathbb{E}\left[e^{sX}\right].$$

- For $Y = aX + b$ and $a, b \in \mathbb{R}$: $g_Y(\omega) = e^{jbs} g_X(as)$.
- Taylor expansion of the MGF within its circle of convergence yields

$$g_X(s) = \sum_{k=0}^{\infty} \frac{\mathbb{E}\left[X^k\right]}{k!} s^k.$$

The nth derivative of the MGF is $g_X^{(n)}(s) = \mathbb{E}\left[X^n e^{sX}\right]$. Therefore,

$$g_X^{(n)}(0) = \mathbb{E}[X^n] = m_n.$$

Jointly Gaussian Random Vectors

Gaussian Random Variables *[4, §2.2]*. A continuous RV $X \in \mathbb{R}$ is *Gaussian* or *normal* distributed if it has the PDF

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left(-\frac{(x - \mu_X)^2}{2\sigma_X^2}\right),$$
$$\mu_X \in \mathbb{R},\ \sigma_X \ge 0.$$

It is denoted $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$.

The *moment generating function* of $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ is given as

$$g_X(s) = \exp\left(s\mu_X + \frac{s^2\sigma_X^2}{2}\right).$$

The *moments* of the zero mean Gaussian RV $X \sim \mathcal{N}(0, \sigma_X)$ are

$$\mathbb{E}\left[X^{2k}\right] = \frac{(2k)!\sigma_X^{2k}}{k!2^k}.$$

The odd moments of X are zero.

The Q-Function *[5, §2.2]*. The integral over the tail of the Gaussian PDF is called the *Q-function*

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-u^2/2} du.$$

The Q-function is related to the complementary error function $\mathrm{erfc}(x) = (2/\sqrt{\pi}) \int_x^\infty e^{-u^2} du$ according to

$$Q(x) = \frac{1}{2} \mathrm{erfc}\left(\frac{x}{\sqrt{2}}\right).$$

$Q(x)$ can be bounded for $x > 0$ as

$$\left(1 - \frac{1}{x^2}\right) \frac{e^{-x^2/2}}{\sqrt{2\pi}x} < Q(x) < \frac{e^{-x^2/2}}{\sqrt{2\pi}x},$$

and

$$Q(x) \le \frac{1}{2} e^{-x^2/2}.$$

Gaussian Random Vectors *[4, §2.3–§2.5]*. $\boldsymbol{X} = [X_1 X_2 \dots X_N]^T$ is defined to be a *jointly Gaussian random vector* (JGRV) if, for all real vectors $\mathbf{s} = [s_1 s_2 \dots s_N]^T$, the linear combination $\mathbf{s}^T\boldsymbol{X} = s_1X_1 + s_2X_2 + \dots + s_NX_N$ is a Gaussian RV.

A JGRV is completely characterized by the mean $\boldsymbol{\mu}_{\boldsymbol{X}} := \mathbb{E}[\boldsymbol{X}]$ and the covariance matrix $\mathbf{K}_{\boldsymbol{X}} = \mathbb{E}\left[(\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{X}})(\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{X}})^T\right]$. $\boldsymbol{X}$ is a JGRV $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{X}}, \mathbf{K}_{\boldsymbol{X}})$ iff its *MGF* is

$$g_{\boldsymbol{X}}(\mathbf{s}) = \exp\left(\mathbf{s}^T\boldsymbol{\mu}_{\boldsymbol{X}} + \frac{\mathbf{s}^T\mathbf{K}_{\boldsymbol{X}}\mathbf{s}}{2}\right).$$

For JGRVs with a non-singular covariance matrix, the *PDF* is

$$f_{\boldsymbol{X}}(\mathbf{x}) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_{\boldsymbol{X}})^T\mathbf{K}_{\boldsymbol{X}}^{-1}(\mathbf{x} - \boldsymbol{\mu}_{\boldsymbol{X}})\right)}{(2\pi)^{N/2}\sqrt{\det \mathbf{K}_{\boldsymbol{X}}}}.$$

- For zero-mean JGRVs, *independence and uncorrelatedness are equivalent*.
- Pairwise independence of the elements of a JGRV implies overall independence of the elements of the vector.

Covariance Matrices *[4, 2.5]*. The matrix $\mathbf{K}_{\boldsymbol{X}}$ is a covariance matrix of a real JGRV iff it is *positive semidefinite*. In particular, $\mathbf{K}_{\boldsymbol{X}}$ is the covariance matrix of $\boldsymbol{X} = \mathbf{A}\boldsymbol{W}$, where $\mathbf{A}$ is the unique square root matrix $\mathbf{A} = \mathbf{Q}\boldsymbol{\Lambda}^{1/2}\mathbf{Q}^T$ arising from the spectral decomposition of $\mathbf{K}_{\boldsymbol{X}}$, and $\boldsymbol{W} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. For any covariance matrix $\mathbf{K}$, a zero mean JGRV $\boldsymbol{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{K})$ exists and can be expressed as $\boldsymbol{X} = \mathbf{A}\boldsymbol{W}$, where $\mathbf{A}\mathbf{A}^T = \mathbf{K}$.

Conditional Probabilities *[4, §2.7]*. Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be zero mean, jointly Gaussian and jointly non-singular. Then the dependence of $\boldsymbol{X}$ on $\boldsymbol{Y}$ and vice versa can be stated explicitly as

$$\boldsymbol{X} = \mathbf{G}\boldsymbol{Y} + \boldsymbol{V},$$

with

$$\mathbf{G} = \mathbf{K}_{\boldsymbol{XY}}\mathbf{K}_{\boldsymbol{Y}}^{-1}$$
$$\mathbf{K}_{\boldsymbol{V}} = \mathbf{K}_{\boldsymbol{X}} - \mathbf{G}\mathbf{K}_{\boldsymbol{Y}}\mathbf{G}^T = \mathbf{K}_{\boldsymbol{X}} - \mathbf{K}_{\boldsymbol{XY}}\mathbf{K}_{\boldsymbol{Y}}^{-1}\mathbf{K}_{\boldsymbol{XY}}^T,$$

and the conditional PDF of $\boldsymbol{X}$ given $\boldsymbol{Y}$ is

$$f_{\boldsymbol{X}|\boldsymbol{Y}}(\mathbf{x} \mid \mathbf{y}) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{G}\mathbf{y})^T\mathbf{K}_{\boldsymbol{V}}^{-1}(\mathbf{x} - \mathbf{G}\mathbf{y})\right)}{(2\pi)^{N/2}\sqrt{\det \mathbf{K}_{\boldsymbol{V}}}}.$$

Here, $\boldsymbol{V} \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_{\boldsymbol{V}})$, independent of $\boldsymbol{Y}$; it is sometimes called the *innovation*.

Jointly Complex Gaussian Random Vectors *[3, §6]*. Let $\boldsymbol{U} \in \mathbb{C}^N$ be a proper complex Gaussian random vector with mean $\boldsymbol{\mu}_{\boldsymbol{U}}$ and nonsingular covariance matrix $\mathbf{K}_{\boldsymbol{U}}$. Then the PDF is given by

$$f_{\boldsymbol{U}}(\mathbf{u}) = f_{\underline{\boldsymbol{U}}}(\underline{\mathbf{u}}) = \frac{\exp\left(-(\mathbf{u} - \boldsymbol{\mu}_{\boldsymbol{U}})^H\mathbf{K}_{\boldsymbol{U}}^{-1}(\mathbf{u} - \boldsymbol{\mu}_{\boldsymbol{U}})\right)}{\pi^N \det(\mathbf{K}_{\boldsymbol{U}})}.$$

Two jointly proper Gaussian random vectors $\boldsymbol{U}$ and $\boldsymbol{V}$ are *independent* iff $\mathbf{K}_{\boldsymbol{UV}} = 0$. Analogously to the real case, the covariance matrix $\mathbf{K}_{\boldsymbol{U}}$ is Hermitian and positive semidefinite.

Circularly Symmetric Random Vectors. A complex random vector $\boldsymbol{U}$ is called *circularly symmetric* if $f_{\boldsymbol{U}}(e^{i\theta}\mathbf{u})$ does not depend on $\theta \in \mathbb{R}$. For zero mean Gaussian random vectors, circular symmetry is equivalent to properness.

Inequalities and Limit Theorems

Modes of Convergence *[1, §7.2]*. Let $X_1, X_2, \dots$ be a sequence of RVs on some probability space Ω. We say:

- $X_n \to X$ *almmost surely*, *with probability 1*, or *almost everywhere*, written $X_n \xrightarrow{\text{a.s.}} X$, if $\mathbb{P}(\{\omega \in \Omega : X_n(\omega) \to X(\omega) \text{ as } n \to \infty\}) = 1$.
- $X_n \to X$ *in the rth mean*, $r \ge 1$, written $X_n \xrightarrow{r} X$, if $\mathbb{E}\left[\left|X_n^r\right|\right] < \infty$ for all n, and $\mathbb{E}[|X_n - X|^r] \to 0$ as $n \to \infty$.
- $X_n \to X$ *in probability*, written $X_n \xrightarrow{\text{P}} X$, if $\mathbb{P}(|X_n - X| > \epsilon) \to 0$ as $n \to \infty$ for all $\epsilon > 0$.
- $X_n \to X$ *in distribution*, also termed *weak convergence* or *convergence in law*, written $X_n \xrightarrow{\text{D}} X$, if $\mathbb{P}(X_n \le x) \to \mathbb{P}(X \le x)$ as $n \to \infty$ for all points x at which the function $F_X(x)$ is continuous.

The following implications hold

- $(X_n \xrightarrow{\text{a.s.}} X) \Rightarrow (X_n \xrightarrow{\text{P}} X)$.
- $(X_n \xrightarrow{r} X) \Rightarrow (X_n \xrightarrow{\text{P}} X)$ for any $r \ge 1$.
- $(X_n \xrightarrow{\text{P}} X) \Rightarrow (X_n \xrightarrow{\text{D}} X)$.
- If $r > s \ge 1$, then $(X_n \xrightarrow{r} X) \Rightarrow (X_n \xrightarrow{s} X)$.

Inequalities. Let $\epsilon > 0$ arbitrary.

- *Chebyshev inequality*: if X is a RV with mean μ and variance σ^2, then

$$\mathbb{P}\left(\left|X - \mu\right| \ge \epsilon\right) \le \frac{\sigma^2}{\epsilon^2}.$$

- *Markov inequality*: if X is a RV with finite mean μ, then

$$\mathbb{P}(|X| > \epsilon) \le \frac{\mu}{\epsilon}.$$

- *Chernoff bound*: if X is a RV with finite mean, then

$$\mathbb{P}(X \ge v) \le \min_{s \ge 0}\left(e^{-sv}\,\mathbb{E}\left[e^{sX}\right]\right).$$

- *Cauchy-Schwarz inequality*: if X and Y are RVs with finite second moments, then

$$\left(\mathbb{E}[XY]\right)^2 \le \mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right]$$

with equality iff $\mathbb{P}(aX + bY) = 1$ for some real a and b, at least one of which is non-zero.

Limit Theorems *[1, §5.10,§7.4]*. Let $X_1, X_2, \dots$ be i.i.d. RVs.

- *Weak law of large numbers*: if $\mathbb{E}[X_1] = \mu$, then

$$\frac{1}{N}\sum_{n=1}^{N} X_n \xrightarrow{\text{D}} \mu.$$

- *Strong law of large numbers*: if $\mathbb{E}[|X_1|] < \infty$, then

$$\frac{1}{N}\sum_{n=1}^{N} X_n \xrightarrow{\text{a.s.}} \mathbb{E}[X_1].$$

- *Central limit theorem*: Let $S_N = \sum_{n=1}^{N} X_n$. If $\mathbb{E}[X_1] = \mu < \infty$ and $\sigma^2 = \mathbb{V}\mathrm{ar}[X_1]$, $0 < \sigma^2 < \infty$, then

$$\frac{S_N - N\mu}{\sqrt{N\sigma^2}} \xrightarrow{\text{D}} \mathcal{N}(0, 1).$$

Random Processes

Random Process [1, §8.1],[2, §9.1]. A *random process* $X(t)$ is a family $\{X(t) : t \in \mathcal{T}\}$ of random variables that map the sample space into some set $\mathcal{S}$.

- A random process is called a *discrete-time* random process if $\mathcal{T}$ is a finite set.
- It is called a *continuous-time* process if $\mathcal{T}$ is uncountable.
- A *realization*, or *sample path*, is a collection $\{X(t, \omega) : t \in \mathcal{T}\}$ for a fixed $\omega \in \Omega$.
- The *first order distribution* of $X(t)$ is defined as
$$F_{X(t)}(x, t) = \mathbb{P}(X(t) \le x).$$
- The *n-th order distribution* is defined as
$$F_{X(t)}(x_1, x_2, \dots, x_n, t_1, t_2, \dots, t_n)$$
$$= \mathbb{P}(X(t_1) \le x_1, X(t_2) \le t_2, \dots, X(t_n) \le t_n).$$

A random process is completely specified if a joint distribution is given for *any* finite subset of $\mathcal{T}$.

Covariance and Correlation [1, §9]. The *autocorrelation function* of a complex-valued random process $U(t)$ is defined as
$$R_U(t, t') = \mathbb{E}[U(t)U^\star(t')],$$
and $R_U(t, t)$ is called the *average power* of the process. The autocorrelation function is *positive semidefinite*, i.e., for any a_i, a_j,
$$\sum_{i,j} R_U(t_i, t_j) a_i a_j^\star \ge 0.$$
The *autocovariance function* is defined as
$$K_U(t, t') = \mathbb{E}[(U(t) - \mu_U(t))(U(t') - \mu_U(t'))^\star],$$
where $\mu_U(t) = \mathbb{E}[U(t)]$. Covariance and correlation functions are related according to
$$K_U(t, t') = R_U(t, t') - \mu_U(t)\mu_U^\star(t').$$
The variance of the process is $\sigma^2(t) = K_U(t, t)$. The *pseudocovariance function* is defined as
$$J_U(t, t') = \mathbb{E}\Big[(U(t) - \mu_U(t))^2\Big].$$
The *cross-correlation* of two processes $U(t)$ and $V(t)$ is defined as
$$R_{UV}(t, t') = \mathbb{E}[U(t)V^\star(t')] = R_{VU}^\star(t', t),$$
and the *cross-covariance* is
$$K_{XY}(t, t') = R_{XY}(t, t') - \mu_X(t)\mu_Y^\star(t').$$
The two processes are called *uncorrelated* if $K_{UV}(t, t') = J_{UV}(t, t') = 0$ for every t and t'.

Stationarity [1, §8.1][2, §9.1]. The process $X(t)$ is called *(strongly) stationary*, or *strict sense stationary (SSS)* if the families
$$\{X(t_1), X(t_2), \dots, X(t_n)\}$$
and
$$\{X(t_1 + c), X(t_2 + c), \dots, X(t_n + c)\}$$
have the same joint distribution for all $t_1, t_2, \dots, t_n$ and $c \in \mathbb{R}$.
The process is called *wide sense stationary (WSS)*, or *weakly stationary*, if, for all t_1, t_2 and c,

- $\mu_X(t_1) = \mu_X(t_2) = \mu_X$,
- $K_X(t_1, t_2) = K_X(t_1 + c, t_2 + c)$
$= K_X(t_1 - t_2) = K_X(\tau)$.

For a complex-valued process $U(t)$, it is also required that $J_U(t_1, t_2) = J_U(t_1 + c, t_2 + c) = J_U(t_1 - t_2)$. Two processes $U(t)$ and $V(t)$ are called *jointly WSS* if each is WSS and their cross-correlation depends only on $\tau = t - t'$.

- $K_U(-\tau) = K_U^\star(\tau)$.
- $\sigma^2 = R_U(0)$.
- $|R_U(\tau)| \le R_U(0)\ \forall \tau$.
- If $R_U(\tau_1) = R_U(0)$ for some $\tau_1 \ne 0$, then $R_U(\tau)$ is periodic with period τ_1.
- $R_{UV}^2(\tau) \le R_U(0)R_V(0)$.

Power Spectral Density [2, §9.3]. The *power spectral density (PSD)* of a WSS process $U(t)$ is given by the Fourier transform
$$S_U(f) = \int_{-\infty}^{\infty} R_U(\tau)e^{-j2\pi f\tau}d\tau.$$
The *cross-PSD* $S_{UV}(f)$ of two jointly WSS processes $U(t)$ and $V(t)$ is the Fourier transform of $R_{UV}(\tau)$.

- $S_U(f)$ is real.
- For $X(t)$ real, $S_X(f)$ is real and even.
- $S_U(f) \ge 0$.
- $S_{UV}(f)$ is complex, and $S_{UV}(f) = S_{VU}^\star(f)$.

Random Processes in Systems [2, §9.2]. Let $\mathbb{L}$ denote a linear time invariant (LTI) system, i.e., $\mathbb{L}$ satisfies

- $\mathbb{L}[\alpha x(t) + \beta y(t)] = \alpha\mathbb{L}[x(t)] + \beta\mathbb{L}[y(t)]$, $\alpha, \beta \in \mathbb{C}$.
- If $y(t) = \mathbb{L}[x(t)]$, then $y(t + c) = \mathbb{L}[x(t + x)]$, $c \in \mathbb{R}$.

Consider an LTI system $\mathbb{L}$ with impulse response $h(\tau)$ and the random process $V(t) = \mathbb{L}[U(t)]$. Then,

- $\mathbb{E}[\mathbb{L}[U(t)]] = \mathbb{L}[\mathbb{E}[U(t)]]$.
- $R_{UV}(t_1, t_2) = \int_{-\infty}^{\infty} R_U(t_1, t_2 - \tau)h^\star(\tau)d\tau$.
- $R_V(t_1, t_2) = \int_{-\infty}^{\infty} R_{UV}(t_1 - \tau, t_2)h^\star(\tau)d\tau$.

Let $\mathbb{L}$ be a *differentiator*, i.e., $V(t) = U'(t)$. Then

- $R_{UU'}(t_1, t_2) = \partial R_U(t_1, t_2)/\partial t_2$.
- $R_{U'}(t_1, t_2) = \partial^2 R_U(t_1, t_2)/\partial t_1 t_2$.

If the input to a (not necessarily linear) memoryless system $g(x)$ is a SSS process $U(t)$, the resulting output $V(t)$ is also SSS.
For a *WSS process in an LTI system*, the second order properties of the output can be computed explicitly: consider $V(t) = \mathbb{L}[U(t)]$, where $U(t)$ is WSS and $\mathbb{L}$ LTI with impulse response $h(\tau)$ and transfer function H. Then

- $R_{UV}(\tau) = R_U(\tau) \star h^\star(-\tau)$.
- $S_{UV}(f) = S_U(f)H^\star(f)$
- $R_V(\tau) = R_{UV}(\tau) \star h(\tau)$.
- $S_V(f) = S_{UV}(f)H(f) = S_U(f)\left|H(f)\right|^2$.

Let $\mathbb{L}$ be a *differentiator* and U WSS. Then

- $R_{UU'}(\tau) = -R_U'(\tau)$.
- $S_{UU'}(f) = -j2\pi f S_U(f)$.
- $R_{U'}(\tau) = -R_U''(\tau)$.
- $S_{U'}(f) = 4\pi f S_U(f)$.

Gaussian Processes [1, §9.6]. A real-valued continuous-time process $X(t)$ is called a *Gaussian process* if each finite-dimensional vector $[X(t_1)\, X(t_2)\, \dots\, X(t_N)]^T$ is a JGRV. A complex-valued continuous-time process $U(t)$ is called a *complex Gaussian process* if each finite-dimensional vector $[U(t_1)\, U(t_2)\, \dots\, U(t_N)]^T$ is a proper complex JGVR.

- A Gaussian process (real or complex) is completely specified through its mean and autocovariance function.
- Real and complex Gaussian processes are (strict-sense) stationary iff they are WSS.

Linear Functionals of Random Processes. If $X(t)$ is a continuos-time random process with continuous covariance function, and $g(t)$ is a continuous function, nonzero only over a finite time interval, then the linear functional
$$Y = \langle g, X\rangle = \int_{-\infty}^{\infty} g(t)X(t)dt$$
is a random variable with mean
$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} g(t)\,\mathbb{E}[X(t)]\,dt$$
and variance
$$\mathbb{V}\mathrm{ar}[Y] = \iint_{-\infty}^{\infty} g(t)K_X(t, t')g(t')dtdt'.$$
If $X(t)$ is a Gaussian process, Y is also Gaussian.

White Gaussian Noise. A zero-mean stationary process $W(t)$ is called *white*, if the covariance of any linear functional $Y = \langle g_i, W\rangle$ satisfies
$$\mathbb{E}\Big[Y_iY_j\Big] = \iint_{-\infty}^{\infty} g(t)K_W(t - \tau)g(t')dtdt'$$
$$= \int_{-\infty}^{\infty} g_i(t)g_j(t)dt.$$
Such a process $W(t)$ is not a well-defined random process, but functionals of this process are; therefore, WGN is a *generalized random process*. Formally, the covariance function is written as $K_W = (N_0/2)\delta(\tau)$. If $W(t)$ is Gaussian, called *white Gaussian noise (WGN)*, then $W(t_1)$ and $W(t_2)$ are independent for every $t_1 \ne t_2$. The PSD of WGN is $S_W(f) = N_0/2$.
Let $W(t)$ be WGN, and let $\{\Phi_i(t)\}$ be a set of orthogonal functions. Then the random variables $Y_i = \langle W, \Phi_i\rangle$ are independent.

Common Distributions and Densities

Tabular Overview Including Moments and Characteristic Functions [1, 2].

Name	PMF/PDF	Domain	Mean	Variance	Skewness	Characteristic Function
Bernoulli	$f(1)=0,\ f(0)-q=1-p$	$\{0,1\}$	p	pq	$\frac{q-p}{\sqrt{pq}}$	$q+pe^{it}$
discrete Uniform	$\frac{1}{N}$	$\{1,\dots,N\}$	$\frac{1}{2}(N+1)$	$\frac{1}{12}(N^2-1)$	0	$\frac{e^{it}(1-e^{iNt})}{N(1-e^{it})}$
Binomial	$\binom{N}{k}p^k(1-p)^{(N-k)}$	$\{0,1,\dots,N\}$	Np	$Np(1-p)$	$\frac{1-2p}{\sqrt{Np(1-p)}}$	$(1-p+pe^{it})^N$
Poisson	$e^{-\lambda}\frac{\lambda^k}{k!}$	$k=0,1,2,\dots$	λ	λ	$\frac{1}{\sqrt{\lambda}}$	$\exp\left\{\lambda(e^{it}-1)\right\}$
continuous Unifrom	$\frac{1}{b-a}$	$[a,b]$	$\frac{1}{2}(a+b)$	$\frac{1}{12}(b-a)^2$	0	$\frac{e^{ibt}-e^{iat}}{it(b-a)}$
Exponential	$\lambda e^{-\lambda x}$	$[0,\infty),\ \lambda>0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	2	$\frac{\lambda}{\lambda-it}$
Normal $\mathcal{N}(\mu,\sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$	$\mathbb{R}$	μ	σ^2	0	$e^{i\mu t-\frac{1}{2}\sigma^2 t}$
Multivariat Normal $\mathcal{N}(\boldsymbol{\mu},\mathbf{K})$	$\frac{\exp\left\{-\frac{1}{2}\left(\mathbf{x}-\boldsymbol{\mu}\right)^T\mathbf{K}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)\}}{\sqrt{2\det(\pi\mathbf{K})}}$	$\mathbb{R}^n$	$\boldsymbol{\mu}$	$\mathbf{K}$		$\exp\left(i\mathbf{t}^T\boldsymbol{\mu}-\frac{\mathbf{t}^T\mathbf{K}\mathbf{t}}{2}\right)$
Cauchy $C(\alpha,\mu)$	$\frac{\frac{\alpha}{\pi}}{\alpha^2+(x-\mu)^2}$	$\mathbb{R}$	—	—	—	$e^{i\mu t}e^{-\alpha\lvert t\rvert}$
Rayleigh	$\frac{x}{\sigma^2}e^{-\frac{x^2}{2\sigma^2}}$	$[0,\infty)$	$\sqrt{\frac{\pi\sigma^2}{2}}$	$(2-\frac{\pi}{2})\sigma^2$	$\frac{2\sqrt{\pi}(\pi-3)}{(4-\pi)^{3/2}}$	$\left(1+i\sqrt{\frac{\pi\sigma^2}{2}}t\right)e^{\frac{\sigma^2t^2}{2}}$ (?)
Rice	$\frac{x}{\sigma^2}e^{-\frac{x^2+a^2}{2\sigma^2}}I_0\left(\frac{ax}{\sigma^2}\right)$ $r=\frac{a^2}{2\sigma^2}$	$\mathbb{R}$	$\sigma\frac{\sqrt{\pi}}{2}\Big[(1+r)\,I_0\left(\frac{r}{2}\right)$ $+rI_1\left(\frac{r}{2}\right)\Big]\exp\left\{-\frac{r}{2}\right\}$			
Log-normal	$\frac{1}{x\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(\ln x-\mu)^2}{2\sigma^2}\right\}$	$(0,\infty)$	$e^{\mu+\frac{\sigma^2}{2}}$	$e^{\sigma^2+2\mu}(e^{\sigma^2}-1)$	$\sqrt{e^{\sigma^2}-1}(2+e^{\sigma^2})$	
Central Chi-square χ_N^2	$\frac{x^{\frac{N}{2}-1}}{\Gamma(N/2)2^{N/2}}e^{-x/2}$	$[0,\infty)$	N	$2N$	$\sqrt{\frac{2}{N}}$	$\frac{1}{(1-i2t)^{N/2}}$
Non-Central Chi-Square	$\frac{e^{-\frac{x+\lambda}{2}}x^{\frac{N-1}{2}}\sqrt{\lambda}}{2(\lambda x)^{N/4}}I_{\frac{N}{2}-1}(\sqrt{\lambda x})$	$[0,\infty]$	$\lambda+N$	$2(2\lambda+N)$	$\frac{2\sqrt{2}(3\lambda+N)}{(2\lambda+N)^{3/2}}$	
Weibull	$\alpha\beta^{-\alpha}x^{\alpha-1}e^{-(x/\beta)^\alpha}$	$[0,\infty)$	$\beta\Gamma\left(1+\frac{1}{\alpha}\right)$	$\beta^2\left[\Gamma\left(1+\frac{2}{\alpha}\right)-\Gamma^2\left(1+\frac{1}{\alpha}\right)\right]$	$\frac{2\Gamma^3\left(1+\frac{1}{\alpha}\right)-3\Gamma\left(1+\frac{1}{\alpha}\right)\Gamma\left(1+\frac{2}{\alpha}\right)\Gamma\left(1+\frac{3}{\alpha}\right)}{\left[\Gamma\left(1+\frac{2}{\alpha}\right)-\Gamma^2\left(1+\frac{1}{\alpha}\right)\right]^{3/2}}$	
Nakagami-m	$\frac{2}{\Gamma(m)}\left(\frac{m}{\Omega}\right)^m\frac{x^{2m-1}}{e^{\frac{m}{\Omega}x^2}}$	$(0,\infty)$	$\frac{\Gamma(m+\frac{1}{2})}{\Gamma(m)}\sqrt{\frac{\Omega}{m}}$	$\Omega\left(1-\frac{1}{m}\left(\frac{\Gamma\left(m+\frac{1}{2}\right)}{\Gamma(m)}\right)^2\right)$		

References

[1] G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes*, 3rd ed. Oxford, U.K.: Oxford Univ. Press, 2001.

[2] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. Boston, MA, U.S.A.: McGraw-Hill, 2002.

[3] I. E. Telatar, "Mathematical preliminaries," Mar. 2002, lecture Notes for "Wireless Communication and Mobility", EPFL.

[4] R. G. Gallager, "Stochastic processes—a conceptual approach," Aug. 2001, University of California at Berkeley, EE226A Class Reader, Fall 2001.

[5] S. G. Wilson, *Digital Modulation and Coding*. Upper Saddle River, NJ, U.S.A.: Prentice Hall, 1996.