Basic Probability
Sample Space and Sigma-Field [1, 2.2.5]. The sample space $$\Omega$$ is the set of all outcomes, of an elementary or a random experiment. The power set of $$\Omega$$ contains all subsets of $$\Omega$$ and is written as $$\{0, 1\}^\Omega$$. A collection $$\mathcal{F}$$ of subsets of $$\Omega$$ is called a field if it satisfies the following conditions

1. $$\emptyset \in \mathcal{F}$$.
2. If $$A_1, A_2, \ldots \in \mathcal{F}$$ then $$\bigcup_{i \geq 1} A_i \in \mathcal{F}$$.
3. If $$A \in \mathcal{F}$$ then $$\overline{A} \in \mathcal{F}$$.

A collection of events $$\mathcal{F}$$ are closed under countable intersections.

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

1. $$\mathbb{P}(\emptyset) = 0$$.
2. If $$(X_1, X_2, \ldots)$$ is a disjoint collection of members of $$\mathcal{F}$$, then $$\mathbb{P}(A_1 \cup A_2 \cup \cdots) = \sum_{i \geq 1} \mathbb{P}(A_i)$$.
3. 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, §1.3.1].

1. $$\mathbb{P}(\Omega) = 1$$.
2. If $$A \subseteq B$$ then $$\mathbb{P}(A) \leq \mathbb{P}(B)$$.
3. If $$\mathbb{P}(A) = 0$$, then $$\mathbb{P}(A') = 1$$.

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

1. $$F_X$$ is non-decreasing.
2. $$\lim_{x \to \infty} F_X(x) = 1$$.
3. $$\lim_{x \to -\infty} F_X(x) = 0$$.
4. $$F_X$$ is right-continuous, that is, $$F_X(x + 0) = F_X(x)$$. 
5. $$F_X$$ is continuous at $$x \in R$$ if and only if $$\mathbb{P}(X = x) = 0$$. 

The CDF $$F_X$$ is continuous in that case, i.e., $$F_X(x) = F_X(y)$$ for all $$x, y \in R$$ and $$F_X$$ 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 $$I_A : \Omega \rightarrow R$$ 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 $$F_{X,Y}(x,y) = F_X(x)F_Y(y)$$ for all $$x, y \in R$$. Let $$X : \Omega \rightarrow R$$ be a random variable and $$a \in R$$. The function $$g(x)$$ and $$h(Y)$$ map $$\Omega$$ to $$R$$. Suppose that $$g(x)$$ and $$h(Y)$$ are random variables, i.e., they are $$\mathcal{F}$$-measurable, and that if $$X, Y \in \mathcal{F}$$ is independent then so are $$g(x)$$ and $$h(Y)$$. Random Vectors [1, §2.3.5]. The joint distribution function of a random vector $$X = [X_1, X_2, \ldots, X_n]$$ on the probability space $$(\Omega, \mathcal{F}, \mathbb{P})$$ is the function $$F_{X_1, \ldots, X_n} : R^n \rightarrow [0,1]$$ given by $$F_{X_1, \ldots, X_n}(x) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n)$$ for all $$x = (x_1, x_2, \ldots, x_n)$$, where the ordering of $$x$$ and $$y$$ means that $$x_i \leq y_i$$ for $$i = 1, 2, \ldots, n$$.

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:

1. $$\lim_{x \to \infty} F_{X,Y}(x,y) = 1$$.
2. $$\lim_{x \to -\infty} F_{X,Y}(x,y) = 0$$.
3. $$F_{X,Y}$$ is continuous from below in that $$F_{X,Y}(x^-) = F_{X,Y}(x)$$ for all $$x \in R$$.

The RV $$X$$ is called continuous [2, §4.2] if the CDF $$F_X$$ is continuous; in that case, $$F_X(x) = F_X(y)$$ for all $$x, y \in R$$. It is discrete if it takes values in some countable subset $$\{x_1, x_2, \ldots\}$$ of $$R$$ and RVs are continuous and have density functions. The deﬁnitions of discrete, continuous, and mixed RVs extend to random vectors.

Relationship Between Real-Valued and Complex-Valued Observations [§2]. A Complex RV $$Z = X + jY$$ can be treated as a real vector $$[X, Y]$$.

Continuous Random Variables
Density Functions [1, §1.6.4.5]. If $$X$$ is a continuous RV, its CDF $$F_X$$ is $$F_X(x) = \mathbb{P}(X \leq x)$$ can be expressed as

$$F_X(x) = \int_{-\infty}^{x} f_X(\omega) d\omega.$$

The function $$f_X$$ is called the probability density function of the continuous RV $$X$$. Let $$f_X(x) = 0$$ for all $$x \in R$$. If $$f_X(x) \geq 0$$ for all $$x \in R$$ then $$X$$ is a nonnegative RV. If $$f_X(x) \geq 0$$ for all $$x \in R$$ and $$\int_{-\infty}^{\infty} f_X(x) dx = 1$$ then $$X$$ is a nonnegative RV. Let $$a < b$$ be a constant and $$f_X$$ is the probability density function of $$X$$.

$$\mathbb{P}(X \in [a, b]) = \int_{a}^{b} f_X(x) dx.$$

The function $$f_X$$ is bounded and has the property that $$\int_{-\infty}^{\infty} f_X(x) dx = 1$$.

Inverse Transform [1, §11.11]. Let $$F$$ be a distribution function and let $$Y$$ be uniformly distributed on the interval $$(0,1)$$. Then if $$f_X$$ 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 $$X = 0$$ if $$F(0) = 0$$ and $$X = 1$$ if $$F(1) = 1$$ has distribution function $$F$$.
Jointly Gaussian Random Vectors

**Gaussian Random Variables** [4, §2.2]. A Gaussian RV is characterized by the first and second moment and covariance between its components. If a RV is Gaussian, it is characterized by its mean vector $\mu = (\mu_1, \ldots, \mu_n)$ and covariance matrix $\Sigma = \begin{bmatrix} \Sigma_{11} & \cdots & \Sigma_{1n} \\ \vdots & \ddots & \vdots \\ \Sigma_{n1} & \cdots & \Sigma_{nn} \end{bmatrix}$

Jointly Gaussian Random Vectors

A Gaussian joint distribution can be specified by the mean vector and covariance matrix. If $(X_1, \ldots, X_n)$ are jointly Gaussian with mean $\mu = (\mu_1, \ldots, \mu_n)$ and covariance matrix $\Sigma = \begin{bmatrix} \Sigma_{11} & \cdots & \Sigma_{1n} \\ \vdots & \ddots & \vdots \\ \Sigma_{n1} & \cdots & \Sigma_{nn} \end{bmatrix}$, then for any $a \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$, $a^T (X - \mu) + b$ is normally distributed with mean $a^T \mu + b$ and variance $a^T \Sigma a$.

**Gaussian Random Vectors** [4, §2.3]. Let $U \in \mathbb{C}^n$ be a complex Gaussian random vector with mean $\mu_U = \mathbb{E}[U]$ and covariance matrix $\Sigma_U = \mathbb{E}[(U - \mu_U)(U - \mu_U)^*]^T$. Let $V \in \mathbb{C}^m$ be another complex Gaussian random vector with mean $\mu_V = \mathbb{E}[V]$ and covariance matrix $\Sigma_V = \mathbb{E}[(V - \mu_V)(V - \mu_V)^*]^T$. Then, the vector $W = \begin{bmatrix} U \\ V \end{bmatrix}$ is Gaussian with mean $\mu_W = \begin{bmatrix} \mu_U \\ \mu_V \end{bmatrix}$ and covariance matrix $\Sigma_W = \begin{bmatrix} \Sigma_U & \Sigma_U V^* \\ V \Sigma_U & V \Sigma_U V^* + \Sigma_V \end{bmatrix}$.

**Inequalities and Limit Theorems**

**Modes of Convergence** [1, §7.2]. Let $X_n$ be a sequence of random variables on some probability space $(\Omega, \mathcal{F}, P)$. The following modes of convergence are defined:

1. **Almost sure convergence**: $X_n \xrightarrow{a.s.} X$ if $\mathbb{P}(\lim_{n \to \infty} X_n = X) = 1$.
2. **Convergence in probability**: $X_n \xrightarrow{P} X$ if $\lim_{n \to \infty} \mathbb{P}(|X_n - X| > \epsilon) = 0$ for all $\epsilon > 0$.
3. **Convergence in distribution**: $X_n \xrightarrow{D} X$ if $\mathbb{P}(X_n \leq x) \xrightarrow{n \to \infty} \mathbb{P}(X \leq x)$.

**Markov inequality**. For any random variable $X$ and $c > 0$, $\mathbb{P}(X \geq c) \leq \frac{\mathbb{E}[X]}{c}$. This is a fundamental result in probability theory.

**Chebyshev’s inequality**. For any random variable $X$ and $c > 0$, $\mathbb{P}(X - \mathbb{E}[X] \geq c) \leq \frac{\mathbb{E}[X^2]}{c^2}$.

**Central limit theorem**. Let $S_n = \sum_{i=1}^n X_i$. If $X_i$ are i.i.d. RVs, then $S_n$ converges in distribution to a normal distribution as $n \to \infty$.

**Law of Large Numbers**. For any $X_i$, let $S_n = \sum_{i=1}^n X_i$. Then, $S_n/n$ converges in probability to $\mathbb{E}[X]$ as $n \to \infty$.
Random Processes

Random Process [1, §1.1][2, §3.1]. 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 \(\Omega\).

- A random process is called a discrete-time 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 \(F_X(x) = P(X(T) \leq x)\).
- The \(n\)-th order distribution is defined as \(F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = P(X(t_1) \leq x_1, \ldots, X(t_n) \leq x_n)\).

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 \(R_U(t,\tau) = E[U(t)\bar{U}(t-\tau)]\), and \(R_U(t) = \text{cavarl} \text{power of the process. The autocorrelation function is positive semi-definite, i.e., for any } a, a_i,\)

\[\sum a_i R_U(t_i,\tau) \geq 0.\]

The autocovariance function is defined as \(K_U(t,\tau) = E[(U(t) - \mu_U)(U(t-\tau) - \mu_U)]\), where \(\mu_U = E[U(t)]\). Covariance and correlation functions are related according to \(R_U(t,\tau) = K_U(t,\tau) - \mu_U^2\). The covariance function is defined as \(I_U(t,\tau) = E[U(t)(U(t-\tau)])\). The cross-covariance of two random processes \(U(t)\) and \(V(t)\) is defined as \(R_{UV}(t,\tau) = E[U(t)V(t-\tau)]\) and the cross-covariance is \(K_{UV}(t,\tau) = R_{UV}(t,\tau) - \mu_U\mu_V\). The two processes are called uncorrelated if \(K_{UV}(t,\tau) = 0\) for every \(t\) and \(\tau\).

Stationarity [1, §1.1][2, §1.1]. The random process \(X(t)\) is called (strongly) stationary, or strict sense stationary (SSS) if the families \(\{X(t_1), X(t_2), \ldots, X(t_n)\}\)

and \(\{X(t_1 + c), X(t_2 + c), \ldots, X(t_n + c)\}\) have the same joint distribution for all \(t_1, t_2, \ldots, t_n\) and \(c \in \mathbb{R}\).

If \(X(t) = \mu(t) + R(t)\), where \(\mu(t) = E[X(t)]\) and \(R(t)\) is a random variable with finite variance, \(\text{Var}[R(t)] = E[R(t)^2] - (E[R(t)])^2\).

Power Spectral Density [2, §3.1]. The power spectral density (PSD) of a WSS process \(U(t)\) is given by the Fourier transform \(S_U(f) = \int_{-\infty}^{\infty} R_U(t) e^{-2\pi i ft} dt\). The cross-PSD \(S_{UV}(f)\) of two jointly WSS processes \(U(t)\) and \(V(t)\) is the Fourier transform of \(R_{UV}(t)\).

\[S_{UV}(f) = S_U(f)S_V(f)\]

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

\[L\left[\alpha x(t) + \beta y(t)\right] = \alpha L[x(t)] + \beta L[y(t)], \quad \alpha, \beta \in \mathbb{C}\]

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

Consider an LTI system \(L\) with impulse response \(h(t)\) and the random process \(V(t) = L[U(t)]\).

- \(E[L[U(t)]] = \mathcal{L}[E[U(t)]])\)
- \(R_{UV}(t,\tau) = \int_{-\infty}^{\infty} R_{UU}(t,\tau) \ast h(t-\tau) dt\)
- \(R_V(t,\tau) = \int_{-\infty}^{\infty} R_U(t-\tau,\tau) d\tau\)

Let \(L\) be a differentiator, i.e., \(V(t) = U'(t)\).

Then \(R_{UV}(t,\tau) = R_U(t,\tau) \ast h(t-\tau)\)
- \(R_V(t,\tau) = -\mathcal{F}^{-1}[S_U(f)] \ast \mathcal{F}[H(f)]\)
- \(S_V(f) = \mathcal{F}^{-1}[S_U(f)] \ast \mathcal{F}[H(f)]^2\).

Let \(L\) be a differentiator and \(U\) WSS. Then \(R_{UV}(t,\tau) = \mathcal{F}^{-1}[S_U(f)] \ast \mathcal{F}[H(f)]\)
- \(R_V(t,\tau) = \mathcal{F}^{-1}[S_U(f)] \ast \mathcal{F}[H(f)]^2\)
- \(S_V(f) = \mathcal{F}^{-1} [S_U(f)] \ast \mathcal{F}[H(f)]^2\).

Gaussian Processes [1, §6]. A real-valued continuous-time process \(X(t)\) is called a Gaussian process if each finite-dimensional vector \(\{X(t_1), X(t_2), \ldots, X(t_n)\}\) is a GRV. A complex-valued continuous-time process \(U(t)\) is called a complex Gaussian process if each finite-dimensional vector \(\{U(t_1), U(t_2), \ldots, U(t_n)\}\) is a proper complex GFR.

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

Linear Functions of Random Processes. If \(X(t)\) is a continuous-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 \(E[Y] = \int_{-\infty}^{\infty} g(t)E[X(t)]dt\) and variance \(\text{Var}[Y] = \int_{-\infty}^{\infty} g(t)^2 \text{Var}[X(t)]dt\).

If \(X(t)\) is a Gaussian process, \(Y\) is also Gaussian.

While Gaussian Noise. A zero-mean stationary process \(W(t)\) is called white, if the covariance of any linear functional \(Y = \langle g, W \rangle\) satisfies

\[E[Y] = \int_{-\infty}^{\infty} g(t)K_W(t) dt = 0\]

\[\text{Var}[Y] = \int_{-\infty}^{\infty} g(t)^2 K_W(t) dt\]

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 \(K_W = \delta_W(t)/2\delta(t)\).

If \(W(t)\) is Gaussian, called white Gaussian noise (WGN), then \(W(t_1)\) and \(W(t_2)\) are independent for every \(t_1 \neq t_2\). The PSD of WGN is \(S_W(f) = \delta_W/2\).

Let \(W(t)\) be WGN, and let \(\{b(t)\}\) be a set of orthogonal functions. Then the random variables \(Y_i = \langle W, b_i \rangle\) are independent.
## Common Distributions and Densities

### Tabular Overview Including Moments and Characteristic Functions \([1, 2]\).

<table>
<thead>
<tr>
<th>Name</th>
<th>PMF/PDF</th>
<th>Domain</th>
<th>Mean</th>
<th>Variance</th>
<th>Skewness</th>
<th>Characteristic Function</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Discrete Uniform</strong></td>
<td>( f(k) = \frac{1}{N} \quad \text{for} \quad k = 0, 1, \ldots, N )</td>
<td>([0, N])</td>
<td>( \mu = \frac{N}{2} )</td>
<td>( \frac{N^2-1}{12N} )</td>
<td>0</td>
<td>( \frac{1}{\lambda} )</td>
</tr>
<tr>
<td><strong>Binomial</strong></td>
<td>( (\binom{N}{k}) p^k (1-p)^{N-k} )</td>
<td>([0, 1, \ldots, N])</td>
<td>( Np )</td>
<td>( Np(1-p) )</td>
<td>( \frac{Np(1-p)}{\sqrt{Np(1-p)}} )</td>
<td>( (1-p + pe^\lambda)^N )</td>
</tr>
<tr>
<td><strong>Poisson</strong></td>
<td>( \lambda^k e^{-\lambda} / k! )</td>
<td>( k = 0, 1, 2, \ldots )</td>
<td>( \lambda )</td>
<td>( \lambda )</td>
<td>( \frac{1}{\lambda} \exp[\lambda(e^{\mu} - 1)] )</td>
<td></td>
</tr>
<tr>
<td><strong>continuous Uniform</strong></td>
<td>( \frac{1}{b-a} )</td>
<td>([a, b])</td>
<td>( \frac{a+b}{2} )</td>
<td>( \frac{(b-a)^2}{12} )</td>
<td>0</td>
<td>( e^{\mu e^{-\lambda}} )</td>
</tr>
<tr>
<td><strong>Exponential</strong></td>
<td>( \lambda e^{-\lambda x} )</td>
<td>([0, \infty), \lambda &gt; 0 )</td>
<td>( \frac{1}{\mu} )</td>
<td>( \frac{1}{\lambda^2} )</td>
<td>2</td>
<td>( \lambda )</td>
</tr>
</tbody>
</table>

**Normal** \(N(\mu, \sigma^2)\)

- **Multivariate Normal** \(N(\mu, K)\): \( \mathbb{R}^n \)

**Cauchy** \(C(\alpha, \mu)\)

**Rayleigh**

**Rice**

**Log-normal**

**Central Chi-square** \(x^2_\nu\)

**Non-Central Chi-square**

**Weibull**

**Nakagami-m**

### References


