Handbook for Mathematical Statistics and Probability Theory

Dany Entezari

Draft Version
Last Updated: Feb 2026

Contents

1 Preface

This handbook is an active draft. I am writing it with two goals in mind: first, to enable the interested reader to learn probability theory from mathematical first principles; second, to bridge what I consider a pedagogical gap between probability theory and its applications.

I originally began writing this as supplementary material for my students, but I am now generalizing it for anyone interested in learning probability theory and its applications—namely statistics and machine learning—from first principles.

2 Probability Theory

Probability theory is about mathematical modeling of the phenomena of randomness. In this section, we briefly define fundamental concepts in probability theory.

2.1 Important Terminology

Experiment A procedure that produces one outcome from a set of possible outcomes. *Example:* Rolling a die.

Trial An experiment with a binary outcome. *Example:* Flipping a coin.

Outcome The result of an experiment or trial. For example, one outcome of flipping a coin is "heads" and the other is "tails".

Sample Space A sample space is the set of possible outcomes in a (random) experiment.

Event An event is a subset of a sample space. For example, an event can be the set of outcomes of rolling a dice five times. If there were another event where the dice were rolled, say, five times, then both events would be subsets of the sample space.

2.2 Combinatorics

Combinatorics is the mathematics of counting things efficiently. In combinatorics, we have techniques for determining the number of possible outcomes of experiments without direct enumeration (i.e, without manually counting). In this section, we will look at some of these techniques which have applications in probability distributions and their functions.

2.2.1 Permutation with Repetition

The function for permutation with repetition (or replacement) is given by

$$P^r(n, r) = n^r$$

2.2.2 Permutation without Repetition

The function for permutation without repetition (or replacement) is given by

$$P(n, r) = nPr = \frac{n!}{(n-r)!}$$

2.2.3 Combination without Repetition and the Binomial Coefficient

The function for combination without repetition (or replacement) is given by

$$C(n, r) = nCr = \frac{n!}{r!(n-r)!}$$

$$\binom{n}{k} = \frac{n!}{r!(n-r)!}$$

The binomial coefficient notation is used to express combination, alternatively. Binomial coefficients are so called because they are coefficients of terms in the Binomial Theorem.

2.3 Important Concepts in Probability Theory

In this section, some properties of probability theory are highlighted because they are relevant throughout the statistics and probability theory.

2.3.1 Central Limit Theorem

The *central limit theorem* states that when samples are drawn from a population, where the size of the samples are 30 or greater, the samples will have a normal distribution. This theorem holds even if the population from which the samples are drawn does is not normally distributed.

2.3.2 Law of Large Numbers

The *law of large numbers* states that as the size of the sample increases, the mean of the sample will approach the mean of its population.

2.4 Bayes Theorem and Conditional Probability

The *bayesian theorem* is given by

$$P(A_i|B) = \frac{P(A_i) \times P(B|A_i)}{\sum_{i=k}^{n} P(B|A_i)}$$

where

- A_i is a given a-priori event
- B is the a-posteriori event

3 Random Variables

A *random variable* is a function, mathematically speaking. They map events in a sample space to a subset of the real numbers. But they can also be considered as sets (see §Convergence of Random Variables, Distributions, and Probabilities.)

A random variable is typically denoted by X and elements in the domain of the random variable denoted x; thus $x \in X$. Along with probability distributions, they are central components in Probability Theory.

3.1 Discrete Random Variables

A random variable is discrete if it has one of the following two characteristics:

1. a finite number of possible values
2. an infinite but countable sequence of possible values (see §Countable Sets.)

3.2 Continuous Random Variables

A random variable is continuous if its possible values are infinite and not countable (see §Countable Sets.)

4 Probability Distributions

A probability distribution is a description of data; in particular, the Observations in the data and their corresponding probability. Probability distributions are analogous to frequency distributions which are descriptions of data and their corresponding Frequencies.

4.1 Probability Distribution Functions

Probability distribution functions are models of various types of data. These functions allow us to determine probabilities as functions of Observations and Estimates. Put differently, probability distribution functions will return the probability for a given observation. There are three types of functions which are covered in the following section.

4.1.1 Probability Mass Function

The probability mass function of a discrete random variable will map, for every value in the random variable, the probability that the value will be observed. The probability mass function is denoted by,

$$P(X = k) = p$$

where X is a discrete random variable, k is an observable value, and p is a probability.

4.1.2 Probability Density Function

The probability density function of a continuous random variable will map, for every value in the random variable, the probability that the value will be observed. The probability density function is denoted by,

$$P(X = k) = p$$

where X is a continuous random variable, k is an observable value, and p is a probability.

4.1.3 Cumulative Probability Function

The cumulative probability function is denoted by

$$P(X \leq k) = p = \begin{cases} \sum_{i=1}^{k} f(x) \;\; ; & \text{if X is Discrete} \\ \int_{-\infty}^{k} f(x)\,dx \;\; ; & \text{if X is Continuous} \end{cases}$$

where

- X is a random variable
- $f(x)$ is either a PMF or CDF
- k is an observable value
- p is a probability

The cumulative probability function of a discrete or continuous random variable will map, for every value in the random variable, the probability that the value will be observed.

4.2 Important Probability Distributions

4.2.1 Binomial Distribution

The PMF of the binomial distribution is given by

$$f(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

where

- $\binom{n}{k}$ is the binomial coefficient
- n is a positive integer
- k a number between 0 and n; specifically, $0 \leq r \leq n$

The binomial distribution is used when a discrete random variable has the following characteristics

- fixed number of trials
- trials can have only two outcomes (e.g, heads or tails)

4.2.2 Binomial Distribution and the Binomial Coefficient

When expressing the PMF of the binomial distribution, sometimes the notation for the binomial coefficient is used.

4.2.3 Normal Distribution

The PDF of the normal distribution is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \text{Exp}\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]$$

where

- σ is the population standard deviation
- $n\sqrt{2\pi}$ is attributed to the central limit theorem
- μ is the population mean average

The normal distribution is a generalization of the binomial distribution. This also means that the normal distribution makes it possible to find the probability of values that are not observed in the random variable.

4.2.4 Poisson Distribution

The PDF of the chi-squared distribution is given by

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

where

- λ is the average number of occurrences
- k is the number of successes
- e^x is the exponential function

The Poisson distribution is used to approximate the binomial distribution when the number of trials (i.e, experiments) are large but the number of successes few.

4.2.5 Chi-Squared Distribution (χ^2)

The PDF of the chi-squared distribution is given by

$$f(x) = \begin{cases} \frac{x^{k/2-1} e^{-x/2}}{2^{k/2}\Gamma(k/2)} & ; \text{ for } x \geq 0 \\ 0 \ ; & \text{otherwise} \end{cases}$$

where

- k is parameter representing the degrees of freedom
- $\Gamma(x)$ is the gamma function (see Special Functions)
- e is the exponential function (see Special Functions)

The Chi-Squared distribution is used to determine significant differences between samples and their population with respect to two or more categorical variables.

4.2.6 (Student's) T-Distribution

The PDF of the t-distribution is given by

$$f(x) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{v\pi}\,\Gamma(\frac{v}{2})} \left(1 + \frac{t^2}{v}\right)^{-\frac{v+1}{2}}$$

where

- $\frac{1}{\sqrt{\pi\, n}} \frac{\Gamma((n+1)/2)}{\Gamma(n/2)}$ is the constant of proportionality

The t-distribution is used when the sample is almost normally distributed but has a size 30 or fewer.

4.2.7 (Fischer) F-Distribution

The PDF of the F-distribution is given by

$$f(x) = \begin{cases} \frac{\Gamma(\frac{v_1+v_2}{2})}{\Gamma(\frac{v_1}{2})\,\Gamma(\frac{v_2}{2})} v_1^{v_1/2}\ v_2^{v_2/2}\ x^{(v/2)-1}\ (v_2 + v_1 x)^{-(v_1+v_2)/2} \ ; \quad x > 0 \\ 0 \ ; \quad x \leq 0 \end{cases}$$

where

- v_1, v_2 are degrees of freedom
- $\Gamma(x)$ is the Gamma function (see Special Functions)

The F-distribution is the ratio of two random variables with chi-squared distribution. The F-distribution is used in ANOVA for two random variables and their mean square ratio.

5 Overview of Moments

Moments are specific descriptions of a probability distribution.

The kth moment is given by

$$E(X^k)$$

The kth central moment is given by:

$$E((X-\mu)^k)$$

5.0.1 Expected Value

The function of the expected value is given by

$$E(X) = \begin{cases} \sum_i^k x_i \cdot f(x_i) \ ; & \text{if } X \text{ is discrete} \\ \\ \int_{-\infty}^{\infty} x_i \cdot f(x_i)\, dx_i \ ; & \text{if } X \text{ is continuous} \end{cases}$$

where

- x_i is a value of a random variable
- $f(x_i)$ is a PMF or PDF

Expected Value is the average of the values of a random variable. The average is weighted, however, by the probabilities of the outcomes.

5.0.2 Variance

The function for variance is given by

$$Var(X) = E(X - \mu)^2 = \sigma^2$$

where

- X is a random variable
- μ is the mean average of the random variable

Variance is a measurement of the spread of values of the random variable about the mean. Note that the square root of variance is the standard deviation. In other words, to arrive at the standard deviation of a random variable, the variance must first be calculated. Also, variance can be expressed as

$$Var(X) = E[X^2] - (E[X])^2$$

where

- $E[X^2] = \begin{cases} \sum_i^k x_i \cdot f(x_i)^2 \ ; & \text{if } X \text{ is discrete} \\ \\ \int_{-\infty}^{\infty} x_i \cdot f(x_i)^2 \, dx_i \ ; & \text{if } X \text{ is continuous} \end{cases}$

Note, the expression $E[X^2]$ means that the probability function in the function of expected value is squared!

5.0.3 Skewness

The skewness coefficient is given by

$$\frac{E(X - \mu)^3}{\sigma^3}$$

where

- $E(X - \mu)^3$ is the third moment about the mean. The expression, $(X - \mu)$ is cubed to indicate asymmetry about the mean. If this term were not cubed, the terms in the expression would cancel out and the result would be zero.
- σ^3 is the cube of the standard deviation. This term is in the denominator to make skewness independent of any unit of measurement (e.g., cm, inch, etc.).

5.1 Methods of Moments

Methods of Moments are techniques for estimating Parameters of a population. This technique works by equating moments of samples to the theoretical moments of the population from which they are drawn, and then solving for the parameters.

5.1.1 Moment Generating Functions

Moment generating functions are used to find the characteristics (i.e, moments) of a probability distribution. Moment generating functions are a type of generating functions. Generating Functions are for representing sequences.

6 Maximum Likelihood

Maximum Likelihood is a technique for estimating, by maximizing likelihood, any parameter of a population given the sample.

$$L(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

where

- $L(x_1, \ldots, x_n; \theta)$ is the likelihood function (see next section)
- $f(x_i; \theta)$ is the probability distribution function (see Probability Distribution Functions).
- θ is some parameter
- n is the number of variables
- $x_1, \ldots, x_n$ are variables
- $\prod_{i=1}^{n}$ is the product operator

6.0.1 Likelihood Function

The likelihood function is denoted by

$$L(x_1, \ldots, x_n; \theta)$$

where

- $x_1, \ldots, x_n$ are variables
- θ is the parameter to be estimated

Note that the likelihood function can output very small probabilities, which is why it is necessary to apply the logarithm function to the likelihood function (see Log-Likelihood).

6.0.2 Log-Likelihood

The log-likelihood function is the natural logarithm of the likelihood function. log-likelihood function is given by

$$\ln(L(x_1, \dots, x_n; \theta)) = \ell(x_1, \dots, x_n; \theta) = \log_e(L(x_1, \dots, x_n; \theta))$$

Note that $\ln(L(x, \theta))$ is the natural logarithm; that is, the logarithm of base e. Thus, $\ln(L(x_1, \dots, x_n, \theta)) = \log_e(L(x_1, \dots, x_n, \theta))$.

Also note that capital L is used for likelihood and lowercase italic l for log-likelihood.

Log-likelihood is used, and preferably so, because it "scales" the products of the terms in the MLE — especially when the probabilities are small (e.g, p = 10^{-10} = 0.0000000001).

6.0.3 Differentiating the Log-Likelihood Function

Differentiating the log-likelihood function, $\ell(x_1, \dots, x_n, \theta)$, is done to find the critical values of the function. The critical value would then be the value at which the likelihood functions yield the maximum likelihood.

Since the (log) likelihood is a function of multiple variables, we apply the partial derivative to the likelihood function,

$$\frac{\partial}{\partial x_i} \ell(x_i; \theta) = \frac{\partial}{\partial x} \ln(L(x_i; \theta))$$

6.0.4 Maximum Likelihood Estimator (MLE)

An estimator is a function that is used to estimate a parameter.

$$\hat{\theta}(x_i) = \operatorname*{argmax}_{\theta} \ell(x_i; \theta)$$

where

- argmax is the function that chooses the argument for which the values of some function is maximized
- $\hat{\theta}(x_i)$ is the maximum likelihood estimator
- θ is some parameter
- $\ell(x_i; \theta)$ is the log-likelihood function
- x_i is some variable

7 Measure Theory

7.1 Measure

A *measure* is an abstraction of measurable quantities, such as length, weight, volume, and probability.

More formally, a measure is defined as a function that maps a σ-algebra to $[0, \infty)$. In other words, it maps sets to a real number that is non-negative or infinity. This non-negative property exists because measurable quantities are inherently non-negative.

Usually, in mathematics convention, μ (pronounced "mu") stands for measure, but since μ is reserved, we will denote a measure with the uppercase variant, $\mathcal{M}$.

$$\mu : S \to [0, \infty)$$

7.1.1 σ-algebra

A *σ-algebra* or *sigma-algebra* is a collection of subsets. More specifically, it is a type of collection that is closed under *countable unions* and *complements*. A σ-algebra is necessary, among other operations, for assigning probabilities to subsets of the sample space. The countable unions of the subsets of the sample space have the effect of allowing probabilities to be assigned to events including even infinitely many possible outcomes. A sigma-algebra is also known as a *sigma-field*.

7.1.2 Measure Space

A *space* is a set with certain conditions. In mathematics, a space is denoted as a pair of any elements; for example, $(A, B, ...)$ We define a measure space together with its conditions formally here.

Definition 1 (Measure Space)**.** *A* measure space *is a space* $(E, \mathcal{E}, \mathcal{M})$ *where:*

- *E is a set the subsets of which will be measured*
- *$\mathcal{E}$ is a σ-algebra of measurable subsets of E*
- *$\mathcal{M}$ is the measure on $(E, \mathcal{E})$*

In Definition 1, when we say $\mathcal{M}$ is the measure on $(E, \mathcal{E})$, we mean to say

$$\mathcal{M} : \mathcal{E} \to [0, \infty]$$

.

It is important to indicate that the measure is "on" the pair $(E, \mathcal{E})$ so that we know $\mathcal{E}$ are subsets of E.

7.1.3 Measurable Space

The following is the definition of the *measurable space*–not to be confused with the "measure space."

Definition 2 (Measurable Space)**.** *A* measurable space *is the space* $(X, \mathcal{E})$ *if*

- $\mathcal{E} \subset P(X)$

where:

- $\mathcal{E}$ *is* σ*-algebra*
- X *is a set*
- $P(X)$ *is the family of all subsets*

7.2 Lebesgue Measure

The Lebesgue measure, as a special case of a measure, is used for definite integration on areas, volumes, and other higher-dimensional analogues — n-dimensional measures, if you will.

Definite integration on Lebesgue measures is done using Lebesgue Integration. Here, we provide a formal definition.

Definition 3 (Lebesgue Meausure)**.** *...*

- μ_F *is the complete Lebesgue measure*
- $F(x) = x$ *function associated with* μ_F
- m *is the length of the measure*
- $\mathcal{L}$ *is the Lebesgue measurable set*

7.2.1 Lebesgue Integral

The *lebesgue integral* is a generalization of the Riemann integral.

Definition 4 (Lebesgue Integral)**.** *The* Lebesgue integral *is defined by*

$$\int_X f\, d\mathcal{M} = \sup \left\{ \int_X s\, d\mathcal{M} \;\middle|\; s : X \to \mathbb{R} \quad \textit{and}\ 0 \leq s \leq f \right\}.$$

where:

- $f : X \to [0, \infty]$ *is a non-negative measurable function*
- $\mathcal{M}$ *is the measure*
- X *is a set of the measure space*
- s *is a simple function*

7.3 Probability Measure

A probability measure is a special case of the general concept of a measure, where the measure of the entire sample space is 1.

More formally, a probability measure is expressed as:

$$P : \mathcal{F} \to [0, 1]$$

where

1. $\mathcal{F}$ is a σ-algebra of subsets of the sample space.

7.3.1 Probability Space

A *probability space* is a special case of a measure space which we define here formally.

Definition 5 (Probability Space)**.** *A* probability space *is a measure space* $(\Omega, \mathcal{F}, \mathbb{P})$ *where:*

- Ω *is the sample space*
- $\mathcal{F}$ *is the* σ*-algebra*
- $\mathbb{P}$ *is the measure with domain* $[0, 1]$

8 Markov Chain Monte–Carlo

Markov Chain Monte Carlo is a class of algorithms for obtaining information about distributions by sampling from them. An example is sampling from the posterior distribution in Bayesian probability.

8.1 Preliminaries

Let us first establish some definitions.

8.1.1 Stationary Distribution

A distribution is stationary if it is unchanged after a transition matrix is applied to it.

9 Fourier Analysis

9.1 Limiting Distribution

The distribution to which a sequence of random variables converges.

9.2 Characteristic Function

The *characteristic function* is a unique representation of a probability distribution. This means that, with the characteristic function, the densities of any continuous random variable or the mass of a discrete random variable are determined for the entirety of the probability function's domain.

In probability theory, the characteristic function is the Fourier transform. The characteristic function is also known as the *indicator function of a set*.

10 Convergence of Random Variables, Distributions, and Probabilities

Convergence of sequences is about how a sequence behaves as it approaches a specific limit. Convergence of functions is about how a function behaves through progression of its domain.

Understanding the convergence of random variables, distributions, and probabilities is necessary for understanding how distributions behave asymptotically; that is, as a random variable tends toward a limiting value, or a distribution approaches a limiting form, or a probability converges to a fixed value.

A random variable is a function when considered as a map from the sample space to real numbers, and a set when considered through its *level set* — the set of outcomes in its domain that map to values in the codomain.

A distribution is a function when considered as a map from the codomain of a random variable to probabilities, and a set when considered through the frequencies or densities associated with values in the codomain of a random variable.

11 Select Topics in Stochastic Calculus

11.1 Martingale

A *martingale* is, informally, a description about a sequence of random variables.

11.2 Stochastic Bridges

A *stochastic bridge* is a random sample path. Stochastic bridges are used to model and understand rare events.

A *random sample path*, in turn, is a sequence of random variables. Let us first start with a definition.

Definition 6 (Random Sample Path)**.** *A* random sample path *is a sequence of random variables,* $X_1, X_2, ..., X_n$ *such that*

1. $X_1, X_2, ..., X_n$ *are independent and*
2. $X_1, X_2, ..., X_n$ *have identical distributions*

When random variables are independent and have identical distributions, we use the abbreviation *i.i.d.*

11.3 Stochastic Process

Definition 7 (Stochastic Process)**.** *A* stochastic process *is a set of random variables,* $\{X(t)\, t \in I\}$ *defined on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*, where*

1. t *is the time variable in* I
2. I *is a real interval that is open, closed, or half-closed*

11.4 Stochastic Differential Equations

Stochastic differential equations have applications in modeling systems with non-deterministic behavior.

$$d\xi(t) = b(\xi(t), t)dt + \sigma(\xi(t), t)dw(t)$$

with initial condition

$$\xi(0) = \xi_0$$

Where:

- $\xi(t)$ is the n-dimensional stochastic process being modeled.
- $w(t)$ is an n-dimensional Brownian motion.
- $b(x, t) = (b_1(x, t), \ldots, b_n(x, t))$ is the drift coefficient (an n-vector).
- $\sigma(x, t) = (\sigma_{ij}(x, t))_{i,j=1}^n$ is the diffusion matrix (an n×n matrix).

12 Mathematical Appendix

In this section, we look at some special functions and other concepts widely used in mathematics. This section is for helping you fill in the gaps.

12.1 Infimum and Supremum

Definition 8 (Infimum)**.** *An* infimum *is the greatest lower bound of a set.*

- $m \leq x$ *for every* $x \in A$
- *there exists* $x \in A$ *such that* $x < m + \varepsilon$*, for every* $\varepsilon > 0$

where

- $m \in \mathbb{R}$ *is a real number*
- $x \in A$ *is a real number*
- $A \subset \mathbb{R}$ *is a set of real numbers*

Definition 9 (Supremum)**.** *An* supremum *is the least upper bound of a set.*

- $m \geq x$ *for every* $x \in A$
- *there exists* $x \in A$ *such that* $x > m - \varepsilon$*, for every* $\varepsilon > 0$

where

- $m \in \mathbb{R}$ *is a real number*
- $x \in A$ *is a real number*
- $A \subset \mathbb{R}$ *is a set of real numbers*

12.2 Exponential Function

The Exponential Function is defined by,

$$e^x = \text{Exp}(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

The letter e, which represents Euler's constant ~ 2.718, is the base unit for exponentiation. In other words, a function becomes exponential when the variable appears in the exponent:

$$f(x) = e^x.$$

Note that we are equating e^x to $\text{Exp}(x)$ because the latter is only an alternative (and more convenient) way to write the exponential function.

12.3 Gamma Function

The Gamma Function is defined by,

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x}\, dx$$

The Gamma Function is important because it generalizes the factorial function to all Real Numbers. The factorial operation ($n!$), as a function, is valid only for the Natural Numbers (1,2,3,..). The Gamma Function, however, extends the factorial operation to all Real Numbers greater than 0. For this reason, the Gamma Function is also known as the Generalized Factorial Function.

12.4 Manifolds

If a surface is a generalization of a curve, then a manifold is the generalization of both surfaces and curves. Here's another way to understand a manifold.

- A curve is a one-dimension manifold.
- A surface is a two-dimension manifold.

Beyond two dimensions, we just call these things n-manifold, where n is the number of dimensions.

Smooth Manifold A smooth manifold is a manifold that is locally homeomorphic to a linear space, which is a vector space.

Locally A description about how a manifold looks like around a point. When used, it means that it would take exactly n points to describe the manifold around a point.

Homeomorphic A description of a space that preserves the topological structure. An example is a sphere and a torus, which are homeomorphic to each other. By preserving the topological structure, we mean that the sphere and the torus have the same number of holes.

Topological space A topological space, $\mathcal{T}$, is a collection of subsets on $\mathcal{T}$ which contains

- the empty set
- the space $\mathcal{T}$ itself
- sets that are the result of unions and intersections

13 Table of Symbols

Symbol	**Concept**	**Read**
μ	Population Average	mu
σ	Population Standard Deviation	sigma
s	Sample Standard Deviation	s
$\bar{x}$	Mean Average	bar x
X	Random Variable	random variable X
σ^2	Population Variance	sigma squared
$X \sim N(\mu, \sigma)$	Normal Distribution	X "has" normal distribution
Ω	Sample Space	sample space
$\mathcal{F}$	Sigma Algebra	sigma algebra or sigma field
$\mathbb{P}$	Probability Measure	probability measure
$(\Omega, \mathcal{F}, \mathbb{P})$	Probability Space	probability space
$\mathcal{M}$	Measure	
$\mathcal{T}$	Topology	

Bibliography

1. J. L. Devore and K. N. Berk, Modern Mathematical Statistics with Applications, 3rd ed., Springer, 2021.

2. J. R. Munkres, Topology, 2nd ed., Prentice Hall, 2000.

3. S. Axler, Measure, Integration & Real Analysis, Springer, 2020.

4. D. van Ravenzwaaij, P. Cassey, and S. D. Brown, A Simple Introduction to Markov Chain Monte-Carlo Sampling, Psychonomic Bulletin & Review, vol. 25, no. 1, pp. 143–154, 2018.

5. Massachusetts Institute of Technology, Unified Engineering: Signals and Systems — Lecture 2, lecture notes, 2007–2008.
   Available at: https://web.mit.edu/16.unified/www/archives

6. Klenke, A. Probability Theory: A Comprehensive Course, 2nd ed., Springer, London, 2014.

7. Billingsley, P. Probability and Measure, 3rd ed., Wiley, New York, 1995.

8. Millson, John J., STAT 400: Applied Probability and Statistics I — Lecture 20: Random Samples, lecture notes, University of Maryland.
   Available at: https://math.umd.edu/ millson/teaching/STAT400/slides/article20.pdf

9. University of Cambridge, "Measure Theory," Part III lecture notes, accessed 8 August 2025, https://www.maths.cam.ac.uk/postgrad/part-iii/files/misc/measuretheory.pdf.

10. James Belk (Cornell University), "Defining the Integral," accessed 9 August 2025,
    https://e.math.cornell.edu/people/belk/measuretheory/DefiningTheIntegral.pdf

11. A. Friedman, Stochastic Differential Equations and Applications, Vol. 1. New York, NY, USA: Academic Press, 1975.