# Introduction to Probability Theory

**Quick reference book created from 531 Lecture Notes of Timo**

**Seppalainen and Benedek Valkó**

**Author:** Chris Cai/ Linrong Cai

**Institute:** University of Wisconisn Madison

**Date:** May 15, 2023

**Version:** 1.0

**Instructor**: Benedek Valkó

*The theory of probabilities is at bottom nothing but common sense reduced to calculus; it enables us to appreciate with exactness that which accurate minds feel with a sort of instinct for which of times they are unable to account.*

# Contents

# Chapter 1   Events and Probability

## 1.1  Kolmogorov's axioms and probability space

These are Kolmogorov's axioms for probability theory.

---
**Definition 1.1**

*A **probability space** is a triple $(\Omega, \mathcal{F}, P)$, with the following components.*

*(a) $\Omega$ is a set, called the **sample space**.*

*(b) $\mathcal{F}$ is a collection of subsets of $\Omega$. Members of $\mathcal{F}$ are called events. $\mathcal{F}$ is assumed to be a $\sigma$-algebra, which means that it satisfies the following three properties.*

*(b.1) $\Omega \in \mathcal{F}$. That is, the whole sample space itself is an event.*

*(b.2) If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$.*

*(b.3) If $\{A_k\}_{1 \leq k < \infty}$ is a sequence of members of $\mathcal{F}$, then their union $\bigcup_{k=1}^{\infty} A_k$ is also a member of $\mathcal{F}$.*

*(c) $P$ is a function from $\mathcal{F}$ into real numbers, called the **probability measure**. $P$ satisfies the following axioms.*

*(c.1) $0 \leq P(A) \leq 1$ for each event $A \in \mathcal{F}$.*

*(c.2) $P(\varnothing) = 0$ and $P(\Omega) = 1$.*

*(c.3) If $\{A_k\}_{1 \leq k < \infty}$ is a sequence of pairwise disjoint events then*

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k).$$

♠

---

**Note** *In mathematical analysis and in probability theory, a $\boldsymbol{\sigma}$-algebra (also $\boldsymbol{\sigma}$-field) on a set $X$ is a nonempty collection $\Sigma$ of subsets of $X$ closed under complement, countable unions, and countable intersections. The ordered pair $(X, \Sigma)$ is called a measurable space.*

## 1.2  Inclusion Exclusion Principle

---
**Theorem 1.1 (inclusion-exclusion principle)**

*Let $A_1, A_2, A_3, \ldots$ be events in some probability space $(\Omega, \mathcal{F}, P)$. Then for each integer $n \geq 2$,*

$$
\begin{aligned}
P(A_1 \cup \cdots \cup A_n) = &\sum_{i=1}^{n} P(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} \cap A_{i_2}) \\
&+ \sum_{1 \leq i_1 < i_2 < i_3 \leq n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) \\
&- \sum_{1 \leq i_1 < i_2 < i_3 < i_4 \leq n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3} \cap A_{i_4}) \\
&+ \cdots + (-1)^{n+1} P(A_1 \cap \cdots \cap A_n). \\
= &\sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leq i_1 < \cdots < i_k \leq n} P(A_{i_1} \cap \cdots \cap A_{i_k}).
\end{aligned}
$$

*This is called the inclusion-exclusion identity.*

♡

---

## 1.3 Monotonicity and Countable Subadditivity

> **Proposition 1.1**
>
> *Let $A, B, A_1, A_2, A_3, \ldots$ be events in some probability space $(\Omega, \mathcal{F}, P)$*
>
> *(i) Monotonicity: if $A \subset B$ then $P(A) \leq P(B)$.*
>
> *(ii) Countable subadditivity: for any sequence of events $\{A_k\}$,*
>
> $$P\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} P(A_k).$$
>
> *Countable subadditivity generalizes the countable additivity axiom in a natural way. Its truth should be fairly obvious because the union $\bigcup_{k=1}^{\infty} A_k$ can have overlaps whose probabilities are then counted several times over in the sum $\sum_{i=1}^{\infty} P(A_k)$. By taking $A_k = \varnothing$ for all $k > n$ we get a finite version of subadditivity:*
>
> $$P(A_1 \cup \cdots \cup A_n) \leq P(A_1) + \cdots + P(A_n)$$
>
> *valid for all events $A_1, \ldots, A_n$.* ♣

> **Corollary 1.1**
>
> *Let $\{A_k\}$ be a sequence of events on $(\Omega, \mathcal{F}, P)$.*
>
> *(i) If $P(A_k) = 0$ for all $k$, then $P(\bigcup_k A_k) = 0$.*
>
> *(ii) If $P(A_k) = 1$ for all $k$, then $P(\bigcap_k A_k) = 1$.* ♡

## 1.4 Continuity of Probability

> **Definition 1.2**
>
> *Suppose $\{A_k\}_{k \in \mathbb{Z}_{>0}}, \{B_k\}_{k \in \mathbb{Z}_{>0}}, A$, and $B$ are events in a probability space $(\Omega, \mathcal{F}, P)$. We say that $A_k$ increases up to $A$ and use the notation*
>
> $$A_k \nearrow A$$
>
> *if the events $A_k$ are nested nondecreasing, which means that $A_1 \subset A_2 \subset A_3 \subset \cdots \subset A_k \subset \cdots$, and $A = \bigcup_k A_k$. Figure 1 illustrates. Analogously, we say that $B_k$ decreases down to $B$ and use the notation*
>
> $$B_k \searrow B$$
>
> *if the events $B_k$ are nested nonincreasing, which means that $B_1 \supset B_2 \supset B_3 \supset \cdots \supset B_k \supset \cdots$, and $B = \bigcap_k B_k$* ♠

> **Theorem 1.2**
>
> *If $A_k \nearrow A$ or $A_k \searrow A$, then the probabilities converge: $\lim_{k \to \infty} P(A_k) = P(A)$* ♡

## 1.5 Conditional Probability

> **Definition 1.3 (conditional probability)**
>
> *Let $B$ be an event on the probability space $(\Omega, \mathcal{F}, P)$ such that $P(B) > 0$. Then for all events $A \in \mathcal{F}$ the conditional probability of $A$ given $B$ is defined as*
> $$P(A \mid B) = \frac{P(AB)}{P(B)}$$
> ♠

> **Proposition 1.2**
>
> *Let $B$ be an event on the probability space $(\Omega, \mathcal{F}, P)$ such that $P(B) > 0$. Then as a function of the event $A$, $P(A \mid B)$ is a probability measure on $(\Omega, \mathcal{F})$*
> ♣

> **Theorem 1.3**
>
> *In each statement below all events are on the same probability* space$(\Omega, \mathcal{F}, P)$
>
> *(a) Let $A$ and $B$ be two events and assume $P(B) > 0$. Then*
> $$P(AB) = P(B)P(A \mid B)$$
> *Let $A_1, \ldots, A_n$ be events and assume $P(A_1 \cdots A_{n-1}) > 0$. Then*
> $$P(A_1 A_2 \cdots A_n) = P(A_1) P(A_2 \mid A_1) P(A_3 \mid A_1 A_2) \cdots P(A_n \mid A_1 \cdots A_{n-1})$$
> *(b) Let $\{B_i\}$ be a countable partition of $\Omega$ and $A$ any event. Then*
> $$P(A) = \sum_{i:P(B_i)>0} P(A \mid B_i) P(B_i).$$
> *The sum above ranges over those indices $i$ such that $P(B_i) > 0$.*
> ♡

## 1.6 Bayes' Formula

> **Theorem 1.4 (Bayes' formula)**
>
> *Let $\{B_k\}$ be a countable partition of the sample space $\Omega$. Then for any event $A$ with $P(A) > 0$ and each $k$ such that $P(B_k) > 0$,*
> $$P(B_k \mid A) = \frac{P(AB_k)}{P(A)} = \frac{P(A \mid B_k) P(B_k)}{\sum_{i:P(B_i)>0} P(A \mid B_i) P(B_i)}$$
> ♡

## 1.7 Independent Events

> **Definition 1.4**
>
> *Two events $A$ and $B$ are independent if*
> $$P(AB) = P(A)P(B)$$
> ♠

**Theorem 1.5**

*Suppose that $A$ and $B$ are independent events. Then the same is true for each of these pairs: $A^c$ and $B$, $A$ and $B^c$, and $A^c$ and $B^c$.*

♡

The definition of independence of more than two events requires that the product property hold for any subcollection of events.

**Definition 1.5**

*(a) Events $A_1, \ldots, A_n$ are independent (or mutually independent) if for every collection $A_{i_1}, \ldots, A_{i_k}$, where $2 \le k \le n$ and $1 \le i_1 < i_2 < \cdots < i_k \le n$*

$$P\left(A_{i_1} A_{i_2} \cdots A_{i_k}\right) = P\left(A_{i_1}\right) P\left(A_{i_2}\right) \cdots P\left(A_{i_k}\right)$$

*(b) Let $\{A_k\}_{k \in \mathbb{Z}_{>0}}$ be an infinite sequence of events in a probability space $(\Omega, \mathcal{F}, P)$. Then events $\{A_k\}_{k \in \mathbb{Z}_{>0}}$ are independent if for each $n \in \mathbb{Z}_{>0}$, events $A_1, \ldots, A_n$ are independent.*

♠

**Theorem 1.6**

*(a) Suppose events $A_1, \ldots, A_n$ are mutually independent. Then for every collection $A_{i_1}, \ldots, A_{i_k}$, where $2 \le k \le n$ and $1 \le i_1 < i_2 < \cdots < i_k \le n$, we have*

$$P\left(A_{i_1}^* A_{i_2}^* \cdots A_{i_k}^*\right) = P\left(A_{i_1}^*\right) P\left(A_{i_2}^*\right) \cdots P\left(A_{i_k}^*\right)$$

*where each $A_i^*$ can represent either $A_i$ or $A_i^c$.*

*(b) Let $\{A_k\}_{k \ge 1}$ be a finite or infinite sequence of independent events. Let $0 = k_0 < k_1 < \cdots < k_n$ be integers. Let $B_1, \ldots, B_n$ be events constructed from the $A_k$s so that, for each $j = 1, \ldots, n$, $B_j$ is formed by applying set operations to $A_{k_{j-1}+1}, \ldots, A_{k_j}$. Then the events $B_1, \ldots, B_n$ are independent.*

♡

**Definition 1.6**

*Let $A_1, \ldots, A_n$ and $B$ be events on $(\Omega, \mathcal{F}, P)$ and assume $P(B) > 0$. Then events $A_1, \ldots, A_n$ are conditionally independent, given $B$, if for every collection $A_{i_1}, \ldots, A_{i_k}$, where $2 \le k \le n$ and $1 \le i_1 < i_2 < \cdots < i_k \le n$,*

$$P\left(A_{i_1} A_{i_2} \cdots A_{i_k} \mid B\right) = \prod_{j=1}^{k} P\left(A_{i_j} \mid B\right)$$

♠

# Chapter 2　Random Variables and Probability Distributions

## 2.1  Random Variables

Often we are interested in some numerical value associated to the outcome of a random experiment. This just means that we are interested in the value of a function that maps the elements of the sample space into the real numbers. These functions are called random variables.

---
**Definition 2.1**

Let $(\Omega, \mathcal{F}, P)$ be a probability space. A random variable on $\Omega$ is a real valued function $X : \Omega \to \mathbb{R}$, for which $\{\omega \in \Omega : X(\omega) \le c\} \in \mathcal{F}$ for any $c \in \mathbb{R}$.

♠

---

There is a simple way to encode an events as a random variable.

---
**Definition 2.2**

Let $B$ be an event in a probability space. Then the indicator function (or indicator random variable) of $B$ is defined as the function

$$I_B(\omega) = \begin{cases} 1, & \text{if } \omega \in B \\ 0, & \text{if } \omega \notin B \end{cases}$$

♠

---

**Note**  *Note that $I_B$ is a random variable.*

## 2.2  Probability Distributions

Through the probabilities of events of type $\{X \in B\}$, a random variable induces a probability measure on the real line.

---
**Definition 2.3**

Let $X$ be a random variable defined on the probability space $(\Omega, \mathcal{F}, P)$. The probability distribution of $X$ is the probability measure $\mu$ on $\mathbb{R}$ defined by

$$\mu(B) = P(X \in B)$$

for Borel subsets $B$ of $\mathbb{R}$.

♠

---

**Note**  *In mathematics, a Borel set is any set in a topological space that can be formed from open sets (or, equivalently, from closed sets) through the operations of countable union, countable intersection, and relative complement. Borel sets are named after Émile Borel.*

## 2.3  Discrete Random Variables

---
**Definition 2.4**

A random variable $X$ is discrete if there exists a finite or countably infinite set $B \subset \mathbb{R}$ such that $P(X \in B) = 1$.

---

We say that those values $k$ for which $P(X = k) > 0$ are the possible values of the discrete random variable $X$.

As for functions in general, the range of a random variable $X$ is the set of all its values: the range of $X$ is the set $\{X(\omega) : \omega \in \Omega\}$. In particular, if the range of the random variable $X$ is finite or countably infinite, then $X$ is a discrete random variable.

## 2.4 Probability Mass Functions

**Definition 2.5**

The probability mass function (p.m.f.) of a discrete random variable $X$ is the function $p$ ( or $p_X$ ) defined by

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

for the possible values $k$ of $X$. (In some cases it is convenient to extend the function $p$ for other values as well, we can use the same definition even if $P(X = k) = 0$.)

Probabilities of events involving $X$ come by summing values of the probability mass function: for any subset $B \subseteq \mathbb{R}$

$$P(X \in B) = \sum_{k \in B} P(X = k) = \sum_{k \in B} p_X(k)$$

where the sum is over the possible values $k$ of $X$ that lie in $B$.

## 2.5 Bernoulli Distribution

**Definition 2.6**

Let $0 \le p \le 1$. We say that a random variable $X$ has Bernoulli distribution with parameter $p$ if $X$ is a discrete random variable with probability mass function

$$p_X(1) = p, \quad p_X(0) = 1 - p$$

We abbreviate this as $X \sim \text{Ber}(p)$.

**Note** The distribution of an indicator random variable $I_B$ is always Bernoulli, its parameter is $P(B)$.

## 2.6 Cumulative distribution functions

**Definition 2.7**

Let $X$ be a random variable defined on the probability space $(\Omega, \mathcal{F}, P)$. The cumulative distribution function (c.d.f.) of $X$ is defined by $F(s) = P(X \le s)$ for all $s \in \mathbb{R}$.

**Proposition 2.1**

(a) Let $F : \mathbb{R} \to [0, 1]$ be the cumulative distribution function of a random variable $X$. Then $F$ has the following properties.

*(i) Monotonicity: if $s < t$ then $F(s) \leq F(t)$.*
*(ii) Right continuity: $F(t) = F(t+)$ for each $t \in \mathbb{R}$.*
*(iii) $\lim_{t \to -\infty} F(t) = 0$ and $\lim_{t \to \infty} F(t) = 1$.*
*(b) Conversely, if a function $F : \mathbb{R} \to [0,1]$ has properties (i)-(iii) above, then $F$ is the cumulative distribution function of some random variable.*
*(c) Let $X$ be a random variable with cumulative distribution function $F$. Then for any $s \in \mathbb{R}$ we have these identities:*

$$P(X < s) = F(s-)$$

*and*

$$P(X = s) = F(s) - F(s-).$$

## 2.7 Probability Density Function

**Definition 2.8**

*Let $X$ be a random variable on $(\Omega, \mathcal{F}, P)$. If a function $f$ on $\mathbb{R}$ satisfies $f(x) \geq 0$ for all $x$ and*

$$P(X \leq b) = \int_{-\infty}^{b} f(x)dx$$

*for all real values $b$, then $f$ is the probability density function (p.d.f.) of $X$. When $X$ has a density function, we call $X$ an absolutely continuous random variable.*

**Theorem 2.1**

*Suppose the cumulative distribution function $F$ of the random variable $X$ is continuous and the derivative $F'(x)$ exists everywhere on the real line, except possibly at countably many points. Then $X$ is an absolutely continuous random variable and $f(x) = F'(x)$ is the density function of $X$. If $F$ is not differentiable at a point $x$, then the value $f(x)$ can be set arbitrarily.*

**Definition 2.9**

*Let $f$ be a piecewise continuous function on $\mathbb{R}$. Then $f$ is the density function of a random variable if and only if $f(x) \geq 0$ for all $x \in \mathbb{R}$ and $\int_{-\infty}^{\infty} f(x)dx = 1$*

**Corollary 2.1**

*Random variables can be neither discrete nor absolutely continuous.*

**Example 2.1** Fix $a < b$ and define $F : \mathbb{R} \to [0,1]$ by

$$F(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{1}{3} \cdot \frac{x-a}{b-a} & \text{if } a \leq x < b \\ 1 & \text{if } x \geq b. \end{cases}$$

**Note** *It is natural to generalize the above to random vectors.*

## 2.8 Uniform Distribution

> **Definition 2.10 (Uniform distribution on an interval.)**
>
> *Let $[a, b]$ be a bounded interval on the real line. Random variable $X$ has the uniform distribution on the interval $[a, b]$ if $X$ has density function*
>
> $$f(x) = \begin{cases} \frac{1}{b-a}, & \text{if } x \in [a, b] \\ 0, & \text{if } x \notin [a, b] \end{cases}$$
>
> *Abbreviate this by $X \sim \text{Unif}[a, b]$ If $X \sim \text{Unif}[a, b]$ and $[c, d] \subset [a, b]$, then*
>
> $$P(c \le X \le d) = \int_c^d \frac{1}{b-a} dx = \frac{d-c}{b-a}$$
> ♠

> **Definition 2.11**
>
> *Let $\Omega$ be a subset of $d$-dimensional Euclidean space $\mathbb{R}^d$ with finite volume. Then the random point $\mathbf{X}$ is uniformly distributed on $\Omega$ if its joint density function is*
>
> $$f(\mathbf{x}) = \begin{cases} \frac{1}{\text{vol}(\Omega)} & \text{if } \mathbf{x} \in \Omega \\ 0 & \text{if } \mathbf{x} \notin \Omega \end{cases}$$
> ♠

## 2.9 Notion of equality

> **Definition 2.12 ( Almost sure equality)**
>
> *Let $X$ and $Y$ be random variables defined on $(\Omega, \mathcal{F}, P)$. Then $X$ and $Y$ are equal almost surely if $P(X = Y) = 1$. This is abbreviated by $X = Y$ a.s.*
> ♠

**Note** *Almost sure equality is also expressed by saying $X = Y$ with probability one, abbreviated $X = Y$ w.p.1. Below is a discrete and an absolutely continuous example of almost sure equality $X = Y$ where pointwise equality fails.*

**Example 2.2** Let $\Omega = \{1, 2, 3\}$ with probability measure $P\{1\} = P\{2\} = \frac{1}{2}$ and $P\{3\} = 0$. Define random variables $X$ and $Y$ on $\Omega$ by

$$X(1) = Y(1) = 1, X(2) = Y(2) = 2, X(3) = 3 \text{ and } Y(3) = 0.$$

Then $P(X = Y) = P\{1, 2\} = 1$.

> **Definition 2.13 (Equality in distribution)**
>
> *Random variables $X$ and $Y$ are equal in distribution if $P(X \in B) = P(Y \in B)$ for all Borel subsets $B$ of $\mathbb{R}$. This is abbreviated by $X \overset{d}{=} Y$.*
> ♠

> **Theorem 2.2**
>
> *Suppose $X$ and $Y$ are random variables on the same probability space $(\Omega, \mathcal{F}, P)$. Then $P(X = Y) = 1$ implies $X \overset{d}{=} Y$.*
> ♡

# Chapter 3 Independent and dependent random variables

## 3.1 Independence

---
**Definition 3.1**

*(a) Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be random vectors defined on the same probability space. Let $\mathbf{X}_i$ be $\mathbb{R}^{d_i}$-valued. Then $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are independent if*

$$P\left(\mathbf{X}_1 \in B_1, \mathbf{X}_2 \in B_2, \ldots, \mathbf{X}_n \in B_n\right) = \prod_{k=1}^{n} P\left(\mathbf{X}_k \in B_k\right)$$

*for all Borel subsets $B_i \subset \mathbb{R}^{d_i}, i = 1, \ldots, n$.*

*(b) Let $\{\mathbf{X}_k\}_{k \in \mathbb{Z}_{>0}}$ be an infinite sequence of random vectors defined on some probability space $(\Omega, \mathcal{F}, P)$. Then the random vectors $\{\mathbf{X}_k\}_{k \in \mathbb{Z}_{>0}}$ are independent if, for each $n \in \mathbb{Z}_{>0}$, the random vectors $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are independent.* ♠

---

---
**Theorem 3.1**

*Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be random vectors defined on the same probability space. For $1 \leq i \leq n$ let $d_i$ be the dimension of $\mathbf{X}_i$ and set $d = d_1 + \cdots + d_n$. Define the d-dimensional random vector $\mathbf{Y} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ by putting all the coordinates of the $\mathbf{X}_i$ s together. Denote the joint cumulative distribution functions of these random vectors by $F_{\mathbf{Y}}, F_{\mathbf{X}_1}, \ldots, F_{\mathbf{X}_n}$. Then $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are independent if and only if*

$$F_{\mathbf{Y}}\left(\mathbf{x}_1, \ldots, \mathbf{x}_n\right) = F_{\mathbf{X}_1}\left(\mathbf{x}_1\right) \cdot F_{\mathbf{X}_2}\left(\mathbf{x}_2\right) \cdots F_{\mathbf{x}_n}\left(\mathbf{x}_n\right)$$

*for all vectors $\mathbf{x}_1 \in \mathbb{R}^{d_1}, \mathbf{x}_2 \in \mathbb{R}^{d_2}, \ldots, \mathbf{x}_n \in \mathbb{R}^{d_n}$.* ♡

---

## 3.2 Independent and identically distributed random variables

---
**Definition 3.2**

*Random variables $X_1, X_2, X_3, \ldots$ are independent and identically distributed (abbreviated i.i.d.) if they are independent and each $X_k$ has the same probability distribution. That is, $X_k \overset{d}{=} X_\ell$ for any two indices $k, \ell$.* ♠

---

## 3.3 Independence of jointly absolutely continuous random variables.

---
**Theorem 3.2**

*Theorem 3.11. Let $X_1, \ldots, X_d$ be random variables on the same sample space. Assume that for each $j = 1, 2, \ldots, d, X_j$ has density function $f_{X_j}$. (a) If $X_1, \ldots, X_d$ have joint density function $f$ given by*

$$f\left(x_1, x_2, \ldots, x_d\right) = f_{X_1}\left(x_1\right) f_{X_2}\left(x_2\right) \cdots f_{X_d}\left(x_d\right)$$

*then $X_1, \ldots, X_d$ are independent.*

---

*(b) Suppose $X_1, \ldots, X_d$ are independent. Then they are jointly absolutely continuous with joint density function*

$$f(x_1, x_2, \ldots, x_d) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_d}(x_d)$$

## 3.4 Finding Independence

**Theorem 3.3**

*(a) Suppose $X_1, \ldots, X_n$ are independent random variables and for each index $i$, $g_i$ is a function on the range of $X_i$. Then the random variables $g_1(X_1), g_2(X_2), \ldots, g_n(X_n)$ are independent.*
*One needs to assume that $g_1, \ldots, g_n$ are measurable, but this does not come up usually in applications.*

*(b) Let $\{X_k\}_{k \geq 1}$ be a finite or infinite sequence of independent random variables. Let $0 = k_0 < k_1 < \cdots < k_n$ be integers. Let $g_1, \ldots, g_n$ be functions such that $g_j$ is defined on the range of the random vector $(X_{k_{j-1}+1}, \ldots, X_{k_j})$. Define new random variables $Y_j = g_j(X_{k_{j-1}+1}, \ldots, X_{k_j})$ for $j = 1, \ldots, n$. Then the random variables $Y_1, \ldots, Y_n$ are independent.*

## 3.5 Binomial Distribution

**Definition 3.3 (Binomial Distribution)**

*Let $n$ be a positive integer and $0 \leq p \leq 1$. A random variable $X$ has the binomial distribution with parameters $n$ and $p$ if the possible values of $X$ are $\{0, 1, \ldots, n\}$ and the probabilities are*

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for } k = 0, 1, \ldots, n.$$

*Abbreviate this by $X \sim \text{Bin}(n, p)$.*

## 3.6 Geometric Distribution

**Definition 3.4 (Geometric Distribution)**

*Let $0 < p \leq 1$. A random variable $X$ has the **geometric distribution** with success parameter $p$ if the set of possible values of $X$ is $\mathbb{Z}_{>0}$ and $X$ satisfies $P(X = k) = (1-p)^{k-1}p$ for positive integers $k$. Abbreviate this by $X \sim \text{Geom}(p)$.*

## 3.7 Negative Binomial Distribution

**Definition 3.5 (Negative binomial distribution)**

*Let $k$ be a positive integer and $0 < p \leq 1$. A random variable $X$ has the negative binomial distribution with parameters $(k, p)$ if the set of possible values of $X$ is the set of integers $\mathbb{Z}_{\geq k} = \{k, k+1, k+2, \ldots\}$ and the probability mass function is*

$$P(X = n) = \binom{n-1}{k-1} p^k (1-p)^{n-k} \quad \text{for } n \in \mathbb{Z}_{\geq k}$$

*Abbreviate this by $X \sim \text{Negbin}(k, p)$.*

**Note** *The Negbin $(1, p)$ distribution is the same as the $\text{Geom}(p)$ distribution.*

**Corollary 3.1**

*Consider a sequence of independent trials with success probability $0 < p \leq 1$. Let $N_k$ be the number of trials needed for the $k$ th success. Set $Y_1 = N_1$ and $Y_k = N_k - N_{k-1}$ for $k \geq 2$. Then the random variables $Y_1, Y_2, Y_3, \ldots$ are i.i.d. In particular, $N_k - N_{k-1} \sim \text{Geom}(p)$ for each $k \geq 2$.*

**Note** *Can think of negative binomial as sum of geometric random variables.*

## 3.8 Possion Distribution

**Definition 3.6 (Possion Distribution)**

*Let $\lambda > 0$. A random variable $X$ has the Poisson distribution with parameter $\lambda$ if the possible values of $X$ are the nonnegative integers and the probability mass function is*

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{for } k \in \{0, 1, 2, \ldots\}$$

*Abbreviate this by $X \sim \text{Poisson}(\lambda)$.*

**Theorem 3.4**

*Fix $\lambda > 0$. For positive integers $n$ for which $\lambda/n < 1$, let $S_n \sim \text{Bin}(n, \lambda/n)$. Then*

$$\lim_{n \to \infty} P(S_n = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{for all} \quad k \in \mathbb{Z}_{\geq 0}$$

## 3.9 Exponential Distributions

**Definition 3.7**

*Let $0 < \lambda < \infty$. A random variable $X$ has the exponential distribution with parameter $\lambda$ if $X$ has density function*

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

*on the real line. Abbreviate this by $X \sim \text{Exp}(\lambda)$.*

Exponential Distribution

**Theorem 3.5**

*Suppose that $X \sim \mathrm{Exp}(\lambda)$. Then for any $s, t > 0$, (3.27)*

$$P(X > t + s \mid X > t) = P(X > s).$$

**Note** *This is also called the memoryless property.*

**Theorem 3.6**

*Fix $\lambda > 0$. Consider $n$ large enough so that $\lambda/n < 1$. Suppose that for each $n$, the random variable $T_n$ satisfies $nT_n \sim \mathrm{Geom}(\lambda/n)$. Then*

$$\lim_{n \to \infty} P\left(T_n > t\right) = e^{-\lambda t} \quad \text{for all real } t \geq 0$$