Probability - Math 394/395/396 Notes

Brett Saiki

March 2022

This is a compilation of notes from the probability sequence, Math 394/395/396, at the University of Washington. Math 394 was taught by Alexander Giessing and borrows material from *Introduction to Probability*, 1982, by D. Anderson et. al. and *A First Course in Probability*, 10th edition, 2018 by S. Ross.

Contents

1 Introduction

1.1 Fundamental Concepts

Definition 1.1. An *experiment* is any activity or process whose outcome is subject to uncertainty.

Definition 1.2. An *sample space* of an experiment is the set of all possible outcomes of the experiment. We denote the sample space by Ω.

Definition 1.3. An *event*, A, is a subset of a sample space, Ω, that is, $A \subseteq \Omega$. Let $\omega \in \Omega$ be the outcome of an experiment. We say that the *event A occurs* if $\omega \in A$.

Definition 1.4. A *simple event* is a subset of the sample space that contains only one outcome.

1.2 Laplace Distribution

Definition 1.5. Let N give the number of simple events in an event. Suppose all outcomes of an experiment with finite sample space Ω are equally likely. Then, for all events $A \subseteq \Omega$,

$$\mathbb{P}(A) = \frac{N(A)}{N(\Omega)}.$$

We call $\mathbb{P}$ the *Laplace distribution (over Ω)*.

Lemma 1.6. The Laplace distributon $\mathbb{P}$ over Ω has the following properties:

(i) $\mathbb{P}(\Omega) = 1$.

(ii) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ for disjoint events A and B.

1.3 Probability and Set Theory

Theorem 1.7 (DeMorgan's Law). For any events A and B we have

(i) $(A \cup B)^c = A^c \cap B^c$,

(ii) $(A \cap B)^c = A^c \cup B^c$.

Definition 1.8. Given $A_1, \ldots, A_n \subseteq \Omega$ we define

$$\bigcup_{k=1}^{n} A_k = A_1 \cup \cdots \cup A_n = \{\omega \in \Omega \mid \exists k \in \{1, \ldots, n\} : \omega \in A_k\},$$
$$\bigcap_{k=1}^{n} A_k = A_1 \cap \cdots \cap A_n = \{\omega \in \Omega \mid \forall k \in \{1, \ldots, n\} : \omega \in A_k\}.$$

Theorem 1.9. Given $A_1, \ldots, A_n \subseteq \Omega$,

(i) $\left(\bigcup_{k=1}^{n} A_k\right)^c = \bigcap_{k=1}^{n} A_k^c$

(ii) $\left(\bigcap_{k=1}^{n} A_k\right)^c = \bigcup_{k=1}^{n} A_k^c$

Definition 1.10. Let $(A_k)_{k=1}^{\infty}$ be a sequence of subsets in Ω and define

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k = \{\omega \in \Omega \mid \exists n \geq 1 : \forall k \geq n : \omega \in A_k\},$$
$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k = \{\omega \in \Omega \mid \forall n \geq 1 : \exists k \geq n : \omega \in A_k\}.$$

1.4 Axioms of Probability Theory

Definition 1.11. Let Ω be a finite sample space and $\mathcal{A}$ be the collection of all subsets of Ω. A *probability measure on* $(\Omega, \mathcal{A})$ is a function $\mathbb{P}$ from $\mathcal{A}$ into the real numbers that satisfies

(i) $\mathbb{P}(A) \geq 0$ for all $A \in \mathcal{A}$;
(ii) $\mathbb{P}(\Omega) = 1$;
(iii) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ for all pairwise disjoint $A, B \in \mathcal{A}$.

The number of $\mathbb{P}(A)$ is called the probability that event A occurs. These properties are called *non-negativity*, *normalization*, and *additivity*.

Definition 1.12. A collection $\mathcal{A}$ of subsets of Ω is called a σ-algebra if it satisfies the following conditons:

(i) $\emptyset \in \mathcal{A}$;
(ii) if $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$;
(iii) if $A_1, A_2, \ldots \in \mathcal{A}$, then $\bigcup_{k=1}^{\infty} A_k \in \mathcal{A}$.

Properties (ii) and (iii) are called *closed under complement* and *countable additivity*.

Theorem 1.13. The smallest σ-algebra associated with Ω is $\mathcal{A} = \{\emptyset, \Omega\}$.

Theorem 1.14. If Ω is finite, then the power set 2^{Ω} is a σ-algebra.

Theorem 1.15. If A is any subset of Ω, then $\mathcal{A} = \{\emptyset, A, A^c, \Omega\}$ is a σ-algebra.

Definition 1.16. Let Ω be a sample space and $\mathcal{A}$ be a σ-algebra on Ω. A *probability measure on* $(\Omega, \mathcal{A})$ is a function $\mathbb{P}$ from $\mathcal{A}$ into the real numbers that satisfies

(i) $\mathbb{P}(A) \geq 0$ for all $A \in \mathcal{A}$;
(ii) $\mathbb{P}(\Omega) = 1$;
(iii) if $A_1, A_2, \ldots \in \mathcal{A}$ is a collection of pairwise disjoint events, in that $A_j \cap A_k = \emptyset$ for all pairs j, k satisfying $j \neq k$, then
$$\mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mathbb{P}(A_k).$$

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

Lemma 1.17. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $A, B \subseteq \Omega$.

(i) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$;
(ii) if $A \subseteq B$ then $\mathbb{P}(A) \leq \mathbb{P}(A) + \mathbb{P}(B \setminus A) = \mathbb{P}(B)$;
(iii) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Lemma 1.18 (Inclusion-Exclusion Formula). For any events $A_1, \dots, A_n$ we have

$$\mathbb{P}(A_1 \cup \cdots \cup A_n) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leq i_1 \leq \cdots \leq i_k \leq n} \mathbb{P}(A_1 \cap \cdots \cap A_k).$$

For $n = 2$, this equation simplifies to (iii) of Lemma 1.17.

Theorem 1.19. Let $A_1, A_2, \cdots$ be an increasing sequence of events, i.e. $A_1 \subset A_2 \subset A_3 \subset \cdots$, then

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \bigcup_{k=1}^{\infty} \mathbb{P}(A_k).$$

Let $B_1, B_2, \cdots$ be an decreasing sequence of events, i.e. $B_1 \supset B_2 \supset B_3 \supset \cdots$, then

$$\lim_{n\to\infty} \mathbb{P}(B_n) = \bigcap_{k=1}^{\infty} \mathbb{P}(B_k).$$

2 Combinatorics

2.1 Urn Models

Definition 2.1 (Falling Factorial). For $r \in \mathbb{R}$ and $k \in \mathbb{N}$ we define $(r)_k$, "r falling k", as

$$(r)_k = r \cdot (r-1) \cdots (r-k+1).$$

Definition 2.2 (Factorial). For $n \in \mathbb{N}$ we define $n!$, "n factorial", as

$$n! = \begin{cases} n \cdot (n-1) \cdots 2 \cdot 1 & \text{for } n > 1, \\ 1 & \text{for } n = 0. \end{cases}$$

Definition 2.3 (Binomial Coefficient).

For $r \in \mathbb{R}$ and $n \in \mathbb{N}$ we define binomial coefficient $\binom{r}{n}$, "r choose n" as

$$\binom{r}{n} = \frac{r \cdot (r-1) \cdots (r-n+1)}{n!} = \frac{r!}{n!(r-n)!}.$$

For $r \in \mathbb{R}$ and $n \in \mathbb{Z}$, $n \geq 0$ we define the binomial coefficient $\binom{r}{n}$ as

$$\binom{r}{n} = \begin{cases} 1 \text{ if } n = 0, \\ 0 \text{ if } n < 0. \end{cases}$$

Theorem 2.4 (Vandermonde's identity). For non-negative integers $m, n, r, k \in \mathbb{N}_0$,

$$\binom{m+n}{r} = \sum_{k=0}^{r} \binom{m}{k} \binom{n}{r-k}.$$

Definition 2.5 (Urn Model of Laplace experiments). Consider an urn with n balls which are labeled $1, \dots, n$. An urn model is an experiment in which k times a ball is drawn at random from the urn and its number is noted.

Definition 2.6 (Urn Model I, "Ordered Sampling with Replacement"). Draw k times from an urn with n balls. The number and the order of the ball are noted and the ball is put back into the urn. The outcome is $\omega = (a_1, \ldots, a_k)$ where a_i is the number of the ith draw (i.e. a k-tuple with values $\{1, \ldots, n\}$). The sample space is

$$\Omega_I = \{(a_1, \ldots, a_k) \mid a_1, \ldots, a_k \in \{1, \ldots, n\}\}.$$

(i.e. all possible k-tuples with values in $\{1, \ldots, n\}$).

Lemma 2.7. The cardinality of the set Ω_I is $|\Omega_I| = n^k$.

Definition 2.8 (Urn Model II, "Ordered Sampling without Replacement"). Draw k times from an urn with n balls. The number and the order of the ball are noted and the ball is not returned to the urn. The outcome is $\omega = (a_1, \ldots, a_k)$ where a_i is the number of the ith draw (i.e. an arrangement of k elements of $\{1, \ldots, n\}$). The sample space is

$$\Omega_{II} = \{(a_1, \ldots, a_k) \mid a_1, \ldots, a_k \in \{1, \ldots, n\}, a_i \neq a_j \text{ for } i \neq j\}.$$

Lemma 2.9. The cardinality of the set Ω_{II} is $|\Omega_{II}| = (n)_k = n \cdot (n-1) \cdots (n-k+1)$.

Definition 2.10 (Urn Model III, "Unordered Sampling without Replacement"). Draw k times from an urn with n balls. The number of the ball is noted but not the order, and the ball is not returned to the urn. The outcome is $\omega = (a_1, \ldots, a_k)$ (i.e. subsets of $\{1, \ldots, n\}$ of size k). The sample space is

$$\Omega_{III} = \{\omega \subseteq \{1, \ldots, n\} \mid |\omega| = k\}$$

(i.e. all possible subsets of $\{1, \ldots, n\}$ of size k).

Lemma 2.11. The cardinality of the set Ω_{III} is

$$|\Omega_{III}| = \binom{n}{k} = \frac{(n)_k}{k!} = \frac{n \cdots (n-1) \cdots (n-k+1)}{k!}.$$

Definition 2.12 (Urn Model IV, "Unordered Sampling with Replacement"). Draw k times from an urn with n balls. The number of the ball is noted but not the order, and the ball is returned to the urn. The outcome is $\omega = (k_1, \ldots, k_n)$ where k_i denotes how often the ith ball was drawn (i.e. a tuple whose values sum up to k). The sample space is

$$\Omega_{IV} = \{(k_1, \ldots, k_n) \mid k_i \in \mathbb{N}_0, k_1 + \ldots + k_n = k\}.$$

Lemma 2.13. The cardinality of the set Ω_{IV} is

$$|\Omega_{IV}| = \binom{k+n-1}{n-1} = \binom{k+n-1}{k}.$$

2.2 Discrete Probability Spaces

Definition 2.14. A probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is called *discrete* if there exists a finite or countable infinite subset $D \subseteq \Omega$ such that $\mathbb{P}(D) = 1$. The associated probability measure is also called *discrete*.

Lemma 2.15. Any discrete probability measure, $\mathbb{P}$ satisfies

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}),$$

that is, a discrete probability measure $\mathbb{P}$ is fully characterized by its values on simple events.

Lemma 2.16. Let $p : \Omega \to \mathbb{R}$ be a function that satisfies the following:

(i) $p(\omega) = 0$, except for countable many $\omega \in \Omega$,

(ii) $p(\omega) \geq 0$ for all $\omega \in \Omega$,

(iii) $\sum_{\omega\in\Omega} p(\omega) = 1$.

Then p is a probability measure on $(\Omega, \mathcal{A})$ and we call p the *probability (mass) function.*

Definition 2.17 (Urn Model with Colored Balls). Consider an urn with n balls which are labeled $1, \ldots, N$ with balls $\{1, \ldots, R\}$ being one color and $\{R + 1, \ldots, N\}$ being another color. We draw n times a ball at random from the urn and note its number and/or color.

2.3 Hypergeometric Distribution

Definition 2.18 (Hypergeometric Distribution). Under the urn model with colored balls, draw n balls at once from the urn. Consider the event E_r where exactly r balls are the first color, then

$$E_r = \{A \subseteq \{1, \ldots, N\} : |A| = n, |A \cap \{1, \ldots, R\}| = r, |A \cap \{R + 1, \ldots, N\}| = n - r\},$$

and

$$\Omega = \{\omega \subset \{1, \ldots, N\} : |\omega| = n\}.$$

Lemma 2.19. Define the probability mass function of the hypergeometric distribution as

$$p(r) = \frac{\binom{R}{r}\binom{N-R}{n-r}}{\binom{N}{n}} \quad \text{for} \quad r \in \{0, 1, \ldots, n\}.$$

Then $\mathbb{P}(E_r) = p(r)$.

2.4 Binomial Distribution

Definition 2.20 (Binomial Distribution). Under the urn model with colored balls, draw n times from the urn with replacement. Consider the event E_r where exactly r balls are the first color, then

$$E_r = \{(a_1, \ldots, a_n) : |\{i : a_i \in \{1, \ldots, R\}\}| = r\},$$

and

$$\Omega = \{(a_1, \ldots, a_n) : a_1, \ldots, a_n \in \{1, \ldots, N\}\}.$$

Lemma 2.21. Define the probability mass function of the binomial distribution as

$$p(r) = \binom{n}{r}\left(\frac{R}{N}\right)\left(1 - \frac{R}{N}\right)^{n-r} \quad \text{for} \quad r \in \{0, 1, \ldots, n\}.$$

Then $\mathbb{P}(E_r) = p(r)$.

2.5 Multinomial Distribution

Definition 2.22 (Urn Model With Many Colored Balls). Consider an urn with N balls which are labeled $1, \ldots, N$ with the first N_1 balls of color 1, the second N_2 balls of color 2, ..., the last N_r balls of color r. We draw n times a ball at random from the urn and its number and/or color is noted.

Lemma 2.23. The number of possible ways in which a set A with cardinality $|A| = k$ can be partitioned into n subsets $A_1, \dots, A_n$ with cardinalities $k_1, \dots, k_n$ such that $k_1 + \ldots + k_n = n$ is given by

$$\frac{k!}{k_1! \cdots k_n!}.$$

Definition 2.24. For $k, k_1, \dots, k_n \in \mathbb{Z}$ we define *multinomial coefficient* as

$$\binom{k}{k_1, \dots, k_n} = \begin{cases} \frac{k!}{k_1! \cdots k_n!} & \text{if } k_1 \geq 0, \text{ and } \text{ and } \sum_{i=1}^{n} k_i = k, \\ 0 & \text{otherwise.} \end{cases}$$

Definition 2.25 (Multinomial Distribution). Under the urn model with many colored balls, draw n balls with r colors with replacement. Consider the event E_r where exactly r balls are of one color can be written as

$$E_{n_1, \dots, n_r} = \{(a_1, \dots, a_n) : |\{i : a_i \in \{N_{k-1} + 1, \dots, N_k\}\}| = n_k, k \in \{1, \dots, r\}\},$$

where $N_0 = 0, N_1 + \cdots N_r = N$ and $n_1 + \cdots + n_r = n$, and

$$\Omega = \{(a_1, \dots, a_n) : a_1, \dots, a_n \in \{1, \dots, N\}\}.$$

Lemma 2.26. Define the probability mass function of the multinomial distribution as

$$p(n_1, \dots, n_r) = \binom{n}{n_1, \dots, n_r} \prod_{k=1}^{r} \left(\frac{N_k}{N}\right)^{n_k},$$

for $n_1, \dots, n_r \in \mathbb{N}_0$ and $n_1 + \cdots + n_r = n$. Then $\mathbb{P}(E_{n_1, \dots, n_r}) = p(n_1, \dots, n_r)$.

3 Independence and Conditional Events

3.1 Independence

Definition 3.1. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability triple. Two events A and B on $(\Omega, \mathcal{A}\mathbb{P})$ are called *indepedent* if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Definition 3.2. The events $A_1, \dots, A_n$ are called *indepedent* if for each $k \in \{1, \dots, n\}$ and each collection of indices $1 \leq i_1 < \ldots < i_k \leq n$

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k}).$$

Lemma 3.3. Let $A_1, \dots, A_n$ be indepedent events. Consider events $B_1, \dots, B_n$ such that

$$B_i = A_i \qquad \text{or} \qquad B_i = A_i^c.$$

Then the events $B_1, \dots, B_n$ are indepedent.

Definition 3.4. Let $(\Omega_i, \mathcal{A}_i, \mathbb{P}_i)$ be discrete probability spaces with $\mathbb{P}_i$ characterized by the probability mass function $p_i : \Omega_i \to [0, 1]$, $i = 1, \dots, n$. The *product space* $(\Omega, \mathbb{P})$ is the discrete probabilty space with sample space

$$\Omega = \Omega_1 \times \cdots \times \Omega_n = \{(\omega_1, \dots, \omega_n) : \omega_i \in \Omega_i, 1 \leq i \leq n\},$$

and *product measure* $\mathbb{P}$ defined by the probability mass function

$$p(\omega_1, \dots, \omega_n) = p_1(\omega_1) \cdots p_n(\omega_n).$$

Lemma 3.5. Let $A_i \in \Omega_i$ be any event concerning only the *ith* experiment and let A_i' be defined by

$$A_i' = \{\omega : \omega \in \Omega, \omega_i \in A_i\},$$

for $1 \leq i \leq n$. Then

$$\mathbb{P}(A_i') = \mathbb{P}_i(A_i) \quad \text{for all} \quad i = 1, \ldots, n,$$

and the events $A_1', \ldots, A_n'$ are stochastically independent.

3.2 Conditional Probability

Definition 3.6. Let $A, B \subseteq \Omega$ be events such that $\mathbb{P}(A) > 0$. The conditional probability of B given A is defined by as

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}.$$

Lemma 3.7. Events $A, B \subset \Omega$ are indepedent if and only if $\mathbb{P}(B \mid A) = \mathbb{P}(B)$.

Lemma 3.8 (Multiplication Rule). Let $A_1, \ldots, A_n \subseteq \Omega$ be events with $\mathbb{P}(A_1 \cap \ldots \cap A_{n-1} \neq 0)$. Then,

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2 \mid A_1) \cdots \mathbb{P}(A_n \mid A_1, \ldots, A_{n-1}).$$

Definition 3.9. Events $A_1, \ldots, A_n \subseteq \Omega$ are a *disjoint partition of* Ω when $B_1 \cup \cdots \cup B_n$ and $B_i \cap B_j = \emptyset$ for $i \neq j$.

Lemma 3.10 (Law of Total Probability). Let $B_1, \ldots, B_n$ be a disjoint partition of Ω. If $\mathbb{P}(B_i) > 0$ for all $1 \leq i \leq n$, then for any event $A \subseteq \Omega$,

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i).$$

Lemma 3.11 (Bayes' Rule). Let $B_1, \ldots, B_n$ be a disjoint partition of Ω. If $\mathbb{P}(B_i) > 0$ for all $1 \leq i \leq n$, then for any events $A \subseteq \Omega$ and $B_k \subseteq \Omega$,

$$\mathbb{P}(B_k \mid A) = \frac{\mathbb{P}(A \mid B_k)\mathbb{P}(B_k)}{\sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)}.$$

Definition 3.12. In the previous lemma, Lemma 3.11, $\mathbb{P}(B)$ is called the *prior* probability of B and $\mathbb{P}(B \mid A)$ is called the *posterior* probability of B given A.

Lemma 3.13 (Gambler's Ruin). Choose p to be some number such that $0 < p < 1$, choose an integer x such that $0 \leq x \leq K$ for some bound K, and let $q = 1 - p$. Consider a sequence $\{a_n\}$ generated by the following method:

$$a_n = \begin{cases} 0, & a_{n-1} = 0 \\ 1, & a_{n-1} = K \\ a_{n-1} + 1, & \text{with probability } p \\ a_{n-1} - 1, & \text{with probability } q. \end{cases}$$

That is, a_n moves by one in either direction but terminates once it reaches 0 or K. Let A_x be the event that a_n terminates at 0.

(i) If $p \neq q$, then the probability that A_x occurs is

$$\mathbb{P}(A_x) = \frac{(q/p)^x - (q/p)^K}{1 - (q/p)^K}.$$

(ii) If $p = q = 1/2$, then the probability that A_x occurs is

$$\mathbb{P}(A_x) = 1 - \frac{x}{K}.$$

Definition 3.14. A *linear first-order difference equation* is a recursive formula of the form

$$x_{t+1} = ax_t + b, \quad \text{for} \quad t = 0, 1, \ldots$$

where $a \neq 1$ and b are constants.

Lemma 3.15. The solution to the first-order linear difference equation is

$$x_t = a\left(x_0 - \frac{b}{1-a}\right) + \frac{b}{1-a}.$$

4 Discrete Random Variables

4.1 Random Variables

Definition 4.1 (Random Variable). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. A function $X : \Omega \to \mathbb{R}$ is called *measureable* if for all $\alpha \in \mathbb{R}$

$$\{\omega \in \Omega : X(\omega) \leq \alpha\} \in \mathcal{A}.$$

We call such a function a *random variable*.

Remark. In discrete probability spaces, the σ-algebra $\mathcal{A}$ is usually the power set 2^Ω, and therefore every function $X : \Omega \to \mathbb{R}$ is a random variable. For more generate probability spaces, this is not generally true.

Definition 4.2. If $X(\omega) = x$ for some $\omega \in \Omega$, we call x the *realization* or *observed value* of $X(\omega)$.

Remark. We often drop ω and write X instead of $X(\omega)$ and thus denote events of Ω by

$$\{X = a\} = \{\omega \in \Omega : X(\omega) = a\}.$$

Lemma 4.3. A random variable X defines a probability measure $\mathbb{P}_X$ on $\mathbb{R}$ by assigning each $A \subset \mathbb{R}$ the probability that X takes a value in A:

$$\mathbb{P}_X(A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}).$$

When $X^{-1}(A)$ is an event in $\mathcal{A}$,

$$\mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A)).$$

Lemma 4.4. Given a random variable X and a set $A \subset \mathbb{R}$, $X^{-1}(A) \in \mathcal{A}$ if X is measureable and A is Borel-measureable subset of $\mathbb{R}$.

Lemma 4.5. For our purposes it suffices to know that all intervals and all open and closed subsets of $\mathbb{R}$ are Borel-measureable.

Definition 4.6. Let X be a random variable on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The probability distribution $\mathbb{P}_X$ on $\mathbb{R}$ defined by

$$\mathbb{P}_X(A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}) \qquad A \subset \mathbb{R} \text{ is measureable}$$

is called the distribution of X. We generally denote $\mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\})$ by $\mathbb{P}(X \in A)$.

4.2 Discrete Random Variables

Definition 4.7. A random variable X is called *discrete*, if there exists a finite or countably infinite subset $D \subseteq \mathbb{R}$ such that $\mathbb{P}(X \in D) = 1$.

Definition 4.8. Let X be a discrete random variable with range $\{x_1, x_2, \ldots\}$. The function $p : X(\Omega) \to \mathbb{R}$ defined by

$$p(x_i) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = x_i\}) = \mathbb{P}(X = x_i).$$

is called the *probability mass function* of X. It is convenient to extend p to all of $\mathbb{R}$ by assigning $p(x) = 0$ for $x \in \mathbb{R} \setminus X(\Omega)$.

Lemma 4.9. Let X be a discrete random variable with range $X(\Omega) = \{x_1, x_2, \ldots\}$. Then x has a probability mass function that satisfies the following

(i) $p(x_i) \geq 0$,

(ii) $\sum_{i=1}^{\infty} p(x_i) = 1$.

Lemma 4.10. If a function $p : \mathbb{R} \to \mathbb{R}$ satisfies properties (i) and (ii) from Lemma 4.9, then it is a probability mass function for some random variable.

4.3 Distributions of Discrete Random Variables

Definition 4.11 (Laplace Distribution). A discrete random variable X has a *Laplace distribution* (or uniform distribution) on $\{1, 2, \ldots, N\}$ if its probability mass function is given by

$$p_X(k) = \mathbb{P}(X = k) = \frac{1}{N} \quad \text{for} \quad k \in \{1, 2, \ldots, N\}.$$

Definition 4.12. A *Bernoulli trial* (or binomial trial), X on $\Omega = \{S, F\}$ by

$$X(\omega) = \begin{cases} 1, & \omega = S, \\ 0, & \omega = F. \end{cases}$$

Usually, S is called a "success" and F is a "failure".

Definition 4.13 (Bernoulli Distribution). A Bernoulli trial X has a *Bernoulli distribution* with parameter p, where $0 \leq p \leq 1$, if its probability mass function is given by

$$p_X(1) = \mathbb{P}(X = 1) = p \quad \text{and} \quad p_X(0) = \mathbb{P}(X = 0) = 1 - p.$$

We denote this distribution by $\mathrm{Ber}(p)$.

Definition 4.14 (Binomial Distribution). A discrete random variable X has a *binomial distribution* with parameters n and p if its probability mass function is given by

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

for $k = 0, 1, \ldots, n$. We denote this distribution by $\mathrm{Binom}(n, p)$.

Lemma 4.15. Let $0 \leq p \leq 1$ be some probability and let $n \in \mathbb{N}$ be an integer. Suppose $X = Y_1 + Y_2 + \ldots Y_n$ is a discrete random variable where each Y_i is an independent and identically distributed random variable with a Bernoulli distribution of parameter p. Then X has a binomial distribution with parameters n and p.

Definition 4.16 (Geometric Distribution). A discrete random variable X has a *geometric distribution* with parameter p, where $0 \leq p \leq 1$, if its probability mass function is given by

$$p_X(k) = \mathbb{P}(X = k) = (1-p)^{k-1}p$$

for $k = 1, 2, \ldots$. We denote this distribution by $\text{Geo}(p)$.

Remark. The geometric distribution is obtained by running an infinite sequence of independent Bernoulli trials. X is the random variable defined by the number of trials conducted until the first "success" occurs.

Definition 4.17 (Negative Binomial Distribution). A discrete random variable X has a *negative binomial distribution* with parameters r and p, where $r \in \mathbb{N}$ and $0 \leq p \leq 1$ if its probability mass function is given by

$$p_X(k) = \mathbb{P}(X = k) = \binom{r+k-1}{k}(1-p)^k p^r$$

for $k = 0, 1, 2, \ldots$. We denote this distribution by $\text{NB}(r, p)$.

Remark. The negative binomial distribution is obtained by counting the number of "failures" before r "successes" occur.

Definition 4.18 (Hypergeometric Distribution). A discrete random variable X has a *hypergeometric distribution* with parameter N, M, and n if its probability mass function is given by

$$p_X(x) = \mathbb{P}(X = x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

where $\max\{0, n - N + M\} \leq x \leq \min\{n, M\}$. We denote this distribution by $\text{Hypergeo}(N, M, n)$.

Theorem 4.19 (Poisson Limit Theorem). Let $X_1, X_2, \ldots$ be a sequence of $\text{Binom}(n, p_n)$ distributed random variables. Suppose for some $\lambda \in (0, \infty)$, $np_n \to \lambda$ as $n \to \infty$. Then for all $k = 0, 1, 2, \ldots$,

$$\lim_{n\to\infty} \mathbb{P}(X_n = k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

Moreover, $p_\lambda(k) = e^{-\lambda}\lambda^k/k!$ is a probability mass function on $k = 0, 1, 2, \ldots$.

Definition 4.20 (Poisson Distribution). A discrete random variable X has a *Poisson distribution* with parameter $\lambda > 0$, if its probability mass function is given by

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

for $k = 0, 1, 2, \ldots$. We denote this distribution by $\text{Pois}(\lambda)$.

Remark. If $X \sim \text{Binom}(n, p)$ is a random variable where n is sufficiently large, then X can be approximated by $\text{Pois}(np)$.