trim

TEXTS AND READINGS IN MATHEMATICS 33

Introduction to Probability and Measure

K. R. Parthasarathy

HINDUSTAN BOOK AGENCY

TEXTS AND READINGS IN MATHEMATICS 33

Introduction to Probability and Measure

Texts and Readings in Mathematics

Introduction to Probability and Measure

K. R. Parthasarathy
Indian Statistical Institute
New Delhi

HINDUSTAN
BOOK AGENCY

Originally published by Macmillan India Ltd. in 1977

Corrected edition, published by

Hindustan Book Agency (India)
P 19 Green Park Extension
New Delhi 110 016
India

email: hba@vsnl.com
http://www.hindbook.com

Produced from camera ready copy supplied by the Author.

ISBN 978-81-85931-55-5 ISBN 978-93-86279-27-9 (eBook)
DOI 10.1007/978-93-86279-27-9

Digitally Printed at Replika Press Pvt. Ltd.

TO MY GRANDFATHER
S. Raghunathachari

Contents

Preface

In 1902 the French mathematician Henri Lebesgue wrote his famous dissertation *Intégrale, Longueur, Aire* (Integral, Length and Area). Since 1914 the theory of the Lebesgue measure has become a part of the undergraduate curriculum in analysis in all the technologically advanced countries of the world. In 1933 the Russian mathematician A. N. Kolmogorov wrote the famous book *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Foundations of Probability) in which he formulated the basic axioms of probability theory. The appearance of *Measure Theory* by P. R. Halmos and *An Introduction to Probability Theory and Its Applications* by W. Feller in 1950 made both subjects accessible to all undergraduate and graduate students in mathematics all over the world.

The present book has been written in the hope that it will provide the impetus to introduce in undergraduate and graduate programmes both measure theory and probability theory as a one-year course. Since the study of probability theory in its advanced stage depends on a knowledge of measure theory, special effort has been made to integrate the two subjects into a single volume.

The material of the book grew out of the lectures delivered by the author to M. Sc. students at the Centre of Advanced Study in mathematics in the University of Bombay, M. Stat. students of the Indian Statistical Institute, Delhi, and M. Sc. students at the Indian Institute of Technology, Delhi.

The book is divided into eight chapters. Chapter 1 deals with combinatorial probability and the classical limit theorems of Poisson and Laplace-De Moivre. It provides a motivation for extending measures from boolean algebras to σ-algebras. Chapter 2 is devoted to extension of measures from boolean semialgebras and classes of topologically important subsets to σ-algebras. Chapter III deals with properties of borel maps from a measure space into a separable metric space. In particular, Lusin's theorem and the isomorphism theorem are proved. Extension of measures to projective limits of borel spaces is also studied. Chapter 4 deals with integration, Reisz's representation theorem that integration is the only linear operation on good function spaces, and properties of function spaces arising out of a measure space.. Chapter 5 contains a discussion of measures and transition measures on product spaces. The Lebesgue measure in R^k, the change of variable formula for Lebesgue integrals and construction of infinitely differentiable functions are also considered. Chapter 6, which is the longest in the book, introduces the notion of conditional expectation through orthogonal projection and avoids the customary use of the Radon-Nikodym theorem. The Radon-Nikodym theorem and the Lebesgue decomposition are deduced as a corollary of a more general decomposition theorem due to von Neumann. Convergence of conditional expectations in various senses, the idea of regular conditional probability, ergodic theorems and ergodic decomposition are also treated in this chapter. Chapter 7 gives a brief introduction to weak convergence of probability measures and characteristic functions. The last chapter introduces the construction of Haar measure on a locally compact group and invariant and quasi invariant measures on homogenous spaces. The Mackey-Weil theorem on groups with a quasi invariant measure is also proved. For the benefit of the student a number of exercises are included. Connections between measure theory and probability theory on the one hand and various topics like functional analysis, statistics, ergodic theory, etc., on the other are indicated through Remarks, Examples and Exercises.

New Delhi, 1977 K.R.P.

Preface to the Revised Edition

Probability theory had a humble origin during the middle of the seventeenth century in the study of games of chance whereas measure theory was born in the quest to understand Fourier series in the beginning of the twentieth century. To borrow a phrase attributed to Mark Kac and its adaptation by my friend Luigi Accardi, probability theory and quantum probability are respectively measure and operator theory with a soul. This sentiment is well brought out in A. N. Kolmogorov's "Foundations of Probability" (1933) and von Neumann's "Mathematical Foundations of Quantum Mechanics" (1932). A recognition of this aesthetic aspect depends very much on a good understanding of measure theory. Owing to recent explosive developments in probability theory from the view point of applications there is a similar need among a wide spectrum of scholars ranging from economists to engineers and physicists to psychologists. Furthermore, measure theory has its ramifications in topics like function spaces, operator theory, generalized functions, ergodic theory, group representation, quantum probability etc.

Taking all the aspects mentioned above into our view the manuscript for the 1977 edition of the book was prepared by using the notes based on the M.Sc. and M.Stat courses which I taught at the University of Bombay, I.I.T., Delhi and the Delhi Centre of the Indian Statistical Institute. It was typed on an ancient mechanical typewriter and printed by a publisher, not very much accustomed to the delicate needs of a subject like mathematics. However, the main aim of bringing out an inexpensive local edition and making the study of probability and measure occupy a visible position in our graduate and undergraduate programmes was achieved. Unfortunately, the book disappeared from the market around 1982. Students who took courses based on this book have communicated their appreciation on several occasions and also expressed their disappointment at its nonavailability in the market. The present, albeit infinitesimally revised, edition with several corrections has been skilfully TEXed by Anil Shukla and printed by the most competent professional publisher in India in the field of mathematics and, I believe, it would go a long way in helping students get a firm foothold in the twin themes of probability and measure and understand the sentiments expressed above. To Anil Shukla and the Hindustan Publisher I offer my sincere thanks.

New Delhi, 2005 K.R.P.

Acknowledgements

Thanks to Professor S. S. Shrikhande whose enthusiasm brought me back from Manchester into a teaching career in India. It was in his department that I enjoyed my considerable freedom to violate the 'regular' syllabus and teach anything I wanted. My special thanks to Professor C. R. Rao who invited me to teach this subject to the students of the Indian Statistical Institute and provided me with the opportunity to go much above the accepted levels of other degree-awarding institutions. Thanks to the authorities of the Indian Institute of Technology for providing me with a comfortable house in their pleasant campus. Thanks to Sri S. Ramasubramanian who read the manuscript and made many corrections. Thanks to Sri Dev Raj Joshi for his efficient typing of the manuscript on his heavy and ancient mathematical typewriter. Finally, thanks to my wife Shyama who cheerfully exerted herself in no small measure to shield me from the children and provide the required solitude for writing this volume.

Chapter 1

Probability on Boolean Algbras

1.1 Sets and Events

In probability theory we look at all possible basic outcomes of a statistical experiment and assume that they constitute a set X, called the *sample space.* The points or elements of X are called *elementary outcomes.* We shall illustrate by a few examples.

Example 1.1.1. The simplest statistical experiment is one with two elementary outcomes, for example, tossing a coin where the outcome is a head or a tail; observing the sex of a new born baby where the outcome is male or female; examining whether a manufactured item is defective or not, etc.

In these cases we denote the basic outcomes by 0 and 1. It is customary to call them *failure* and *success* respectively. The sample space X contains exactly two points, namely, 0 and 1.

Example 1.1.2. Throw a die and observe the score. The die has six faces and the possible scores form the set $X = \{1, 2, 3, 4, 5, 6\}$.

Example 1.1.3. Go on tossing a coin till you get the first head and observe the outcome at every stage. If we denote head by H and tail by T, any elementary outcome of this experiment is a finite sequence of the form $TTT \dots TH$. The sample space consists of all such finite sequences.

Example 1.1.4. Shuffle a pack of cards and observe the order from top to bottom. The space X consists of 52! permutations.

Example 1.1.5. Observe the atmospheric temperature at a specific place. The elementary outcomes are just real numbers. Thus the sample space is the real line.

Example 1.1.6. Observe the pressure and the temperature of a gas in a box. Here X may be assumed to be the plane R^2.

Example 1.1.7. Observe the temperature graph of the atmosphere during a fixed hour. The sample space X may be identified with the set of all continuous curves in the interval $[0, 1]$.

Let $A \subset X$ be any subset of the sample space X of a statistical experiment. The performance of the experiment leads to the observation of an elementary outcome x which is an element of X. If $x \in A$, we say that the *event A has occurred.* If $x \notin A$, we say that the event A has not occurred or, equivalently, $X - A$ (the complement of A) has occurred. From a practical point of view not every event may be of interest. For example, in Example 1.1.5 above, consider the event ' the temperature measured is a transcendental number'. Such an event is not of any practical significance. However, an event of the kind 'the temperature measured lies in the interval $[a, b]$' is of value. We can sum up this discussion as follows: there is a collection $\mathcal{F}$ of subsets of the sample space, the events corresponding to the elements of which are of 'practical value'. We assume that such a collection $\mathcal{F}$ of events or subsets of the sample space X is clearly specified. We simply say that $\mathcal{F}$ *is the collection of all events* concerning the statistical experiment whose sample space is X. By an *event* we mean an element of $\mathcal{F}$.

We shall now examine what are the natural conditions which the collection or family $\mathcal{F}$ of all events should satisfy. Let $A \subset B \subset X$ be such that $A, B \in \mathcal{F}$. If $x \in A$, then $x \in B$. In other words, whenever A occurs B also occurs. Thus set theoretic inclusion is equivalent to the logical notion of implication.

If $A, B \in \mathcal{F}$, consider the sets $A \cup B$, $A \cap B$ and $X - A$. Note that the occurrence of one of the events A, B is equivalent to saying that the experiment yields an observation x belonging to $A \cup B$. It is natural to expect that $A \cup B$ is also an event. The occurrence of both A and B means that the experimental observation x belongs to $A \cap B$. The nonoccurrence of A means that x lies in $X - A$. So it is natural to demand that $\mathcal{F}$ is closed under finite union, finite intersection and complementation. Nothing is lost by assuming that the whole space X and hence its complement, the empty set $\emptyset$ also belong to $\mathcal{F}$. This leads us to the following.

Definition 1.1.8 A collection $\mathcal{F}$ of subsets of a set X is called a *boolean algebra* if the following conditions are satisfied :

(1) If $A, B \in \mathcal{F}$, then $A \cup B \in \mathcal{F}$ and $A \cap B \in \mathcal{F}$;

(2) If $A \in \mathcal{F}$, the complement $X - A \in \mathcal{F}$;

(3) the empty set $\emptyset$ and whole space X belong to $\mathcal{F}$.

Remark 1.1.9 Hereafter throughout the text we shall write A' (read A prime) for the complement $X - A$ of the set A. For any two subsets A, B of X we shall write AB for the intersection $A \cap B$, $A - B$ for the set AB' and $A \Delta B$ for the symmetric difference $(A - B) \cup (B - A)$.

Example 1.1.10. Let X be any nonempty set and let $\mathcal{F}$ be the class of all subsets of X. Then $\mathcal{F}$ is a boolean algebra.

Example 1.1.11. Let $X = R$ be the real line and let the family $\mathcal{I}$ be defined by

$$\begin{aligned} \mathcal{I} &= \{\text{all intervals of the form}(-\infty, +\infty), \\ &\quad (-\infty, a], (a, \infty), (a, b], \text{ where } a, b \in R\}. \end{aligned}$$

Then the collection

$$\begin{aligned} \mathcal{F} &= \{A : A \subset R, A = \cup_{i=1}^{n} A_i, A_i \in \mathcal{I}, A_i \cap A_j = \emptyset \\ &\quad \text{for } i \neq j, \text{ for some positive integer } n\} \end{aligned}$$

is a boolean algebra. (Here we consider the empty set as the interval $(a, b]$ when $b \geq a$.)

Example 1.1.12. Let Y be any set and let X be the space of all sequences of elements from Y, i.e., any $x \in X$ can be written as $x = (y_1, y_2, \ldots)$ where $y_i \in Y$ for every $i = 1, 2, \ldots$. Let A be any subset of the cartesian product $Y \times Y \times \ldots \times Y$, taken k times. A subset $C \subset X$ of the form

$$C = \{x = (y_1, y_2, \ldots) : (y_{i_1}, y_{i_2}, \ldots, y_{i_k}) \in A\},$$

(where $i_1 < i_2 < \ldots < i_k$ are fixed positive integers) is called a *k-dimensional cylinder set.* Then the collection $\mathcal{F}$ of all finite dimensional cylinder sets is a boolean algebra.

Going back to the relation between the language of set theory and the language of events we summarise our conclusions in the form of a table. Let $\mathcal{F}$ be a boolean algebra of subsets of the sample space of a statistical experiment so that $\mathcal{F}$ is the collection of all events. Then we have the following dictionary :

Language of events	*Language of set theory*
A is an event	$A \in \mathcal{F}$
event A implies event B	$A \subset B$
event A does not occur	A'
one of the events A, B occurs	$A \cup B$
both the events A and B occur	AB
event which always occurs	X
event which never occurs	$\emptyset$
events A and B cannot occur at the same time	$A \cap B = \emptyset$

1.2 Probability on a Boolean algebra

Consider a statistical experiment whose elementary outcomes are described by a sample space X together with a boolean algebra $\mathcal{F}$ of subsets of X. Let the experiment be performed n times resulting in the elementary outcomes $x_1, x_2, \ldots, x_n \in X$. Let $A \subset X$ be an element of $\mathcal{F}$. Let

$$\rho_n(A) = m(A)/n$$

where $m(A)$ is the number of elementary outcomes x_i that lie in the set A. The number $\rho_n(A)$ may be called the *frequency* of occurrence of the event A in the given n trials. First of all we note that $A \to \rho_n(A)$ is a map from $\mathcal{F}$ into the unit interval $[0, 1]$. It is clear that

(i) $\rho_n(A \cup B) = \rho_n(A) + \rho_n(B)$ if $A \cap B = \emptyset, A, B \in \mathcal{F}$;

(ii) $\rho_n(X) = 1$.

It follows from property (i) that

$$\rho_n(A_1 \cup A_2 \cup \ldots \cup A_k) = \sum_{i=1}^{k} \rho_n(A_i) \text{ if } A_i \cap A_j = \emptyset$$

for all $i \neq j$ and $A_1, A_2, \ldots A_k \in \mathcal{F}$. We say that ρ_n is a nonnegative finitely additive function on $\mathcal{F}$ such that $\rho_n(X) = 1$. If there is a 'statistical regularity' in the occurrence of the observations $x_1, x_2, \ldots,$ we expect that, for $A \in \mathcal{F}$, $\rho_n(A)$ will stabilise to a number $\rho(A)$. If it is indeed so, then the map $A \to \rho(A)$ will share the properties (i) and (ii). Motivated by these considerations we introduce the following definitions.

Definition 1.2.1. Let $\{A_\alpha, \alpha \in I\}$ be a family subsets of a set X, where I is some index set. Such a family is said to be *pairwise disjoint* if $A_\alpha \cap A_\beta = \emptyset$ whenever $\alpha \neq \beta$ and $\alpha, \beta \in I$.

Definition 1.2.2. Let $\mathcal{F}$ be a boolean algebra of subsets of a set X. A map $m : \mathcal{F} \to [0, \infty]$ is said to be *finitely additive* if

$$m(A \cup B) = m(A) + m(B) \text{ whenever } A, B \in \mathcal{F} \text{ and } A \cap B = \emptyset.$$

It is said to be *countably additive* if for any sequence $\{A_n\}$ of pairwise disjoint sets belonging to $\mathcal{F}$

$$m\left(\cup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} m(A_n), \text{ if } \cup_{n=1}^{\infty} A_n \in \mathcal{F}.$$

A map $p : \mathcal{F} \to [0, 1]$ is called a *probability distribution* on $\mathcal{F}$ if it is finitely additive and $p(X) = 1$.

We shall now introduce a few examples.

Example 1.2.3. Let X be a finite or countable set and let $\mathcal{F}$ be the boolean algebra of all subsets of X. Let $\{x_1, x_2, \ldots\}$ be an enumeration of all the points of X. Let $\{p_1, p_2, \ldots\}$ be a sequence of nonnegative numbers. For any $A \subset X$, let

$$m(A) = \sum_{j : x_j \in A} p_j.$$

Then it is clear that m is a countably additive function on $\mathcal{F}$. If $\sum_{j : x_j \in X} p_j = 1, m$ is a probability distribution on $\mathcal{F}$.

Example 1.2.4. Let F be a monotonically increasing real valued function defined on the real line R. Let

$$m((a, b]) = F(b) - F(a) \text{ if } a < b \text{ and } a, b \in R.$$

Write $F(+\infty) = \lim_{a \to +\infty} F(a)$ and $F(-\infty) = \lim_{a \to -\infty} F(a)$. Put

$$\begin{aligned} m((-\infty, a]) &= F(a) - F(-\infty), \\ m((b, +\infty)) &= F(+\infty) - F(b), \\ m((-\infty, +\infty)) &= F(+\infty) - F(-\infty). \end{aligned}$$

Then m is a finitely additive set function defined on the class $\mathcal{I}$ of intervals (Example 1.1.11), i.e.,

$$m\left(\cup_{j=1}^{k} I_j\right) = \sum_{j=1}^{k} m(I_j)$$

whenever $I_1, I_2, \ldots, I_k$ and $\cup_{j=1}^{k} I_j$ belong to $\mathcal{I}$ and the family $\{I_j, 1 \leq j \leq k\}$ is pairwise disjoint. Let now A be any set of the form

$$A = \cup_{r=1}^{k} I_r \tag{1.2.1}$$

where $I_1, I_2, \ldots, I_k$ belong to $\mathcal{I}$ and are pairwise disjoint. Define

$$\tilde{m}(A) = \sum_{r=1}^{k} m(I_r).$$

Now the question arises whether $\tilde{m}$ is well defined. For, it is quite possible that A is also of the form

$$A = \cup_{s=1}^{l} F_s \tag{1.2.2}$$

where $F_1, F_2, \ldots, F_l$ belong to $\mathcal{I}$ and are pairwise disjoint. Thus A has two representations (1.2.1) and (1.2.2). However,

$$\sum_{r=1}^{k} m(I_r) = \sum_{s=1}^{l} m(F_s) \tag{1.2.3}$$

Indeed, we have

$$\begin{aligned} I_r = I_r \cap A &= \cup_{s=1}^{l} (I_r \cap F_s), \\ F_s = F_s \cap A &= \cup_{r=1}^{k} (F_s \cap I_r). \end{aligned}$$

We note that the family $\mathcal{I}$ is closed under finite intersection. Since m is additive on $\mathcal{I}$, it follows that

$$\begin{aligned} m(I_r) &= \sum_{s=1}^{l} m(I_r \cap F_s), \\ m(F_s) &= \sum_{r=1}^{k} m(F_s \cap I_r). \end{aligned}$$

Now (1.2.3) is an immediate consequence of the above two equations. This argument implies that $\tilde{m}$ is a well defined finitely additive map on the boolean algebra $\mathcal{F}$ of all subsets which are finite disjoint unions of intervals from $\mathcal{I}$. In other words, corresponding to every monotonic increasing function F on R, one can construct a unique nonnegative finitely additive function on the boolean algebra $\mathcal{F}$ of Example 1.1.11. This becomes a probability distribution if

$$\lim_{\substack{b \to +\infty \\ a \to -\infty}} F(b) - F(a) = 1.$$

Proposition 1.2.5. Let m be a nonnegative finitely additive function on a boolean algebra $\mathcal{F}$ of subsets of X. If $A \subset B$ and $A, B \in \mathcal{F}$, then $m(A) \leq m(B)$. If $A_1, A_2, \ldots A_k \in \mathcal{F}$, then

$$m\left(\cup_{i=1}^k A_i\right) \leq \sum_{i=1}^k m(A_i). \tag{1.2.4}$$

Proof. To prove the first part we observe that

$$B = A \cup (BA') \text{ if } A \subset B.$$

Since $\mathcal{F}$ is a boolean algebra A and BA' are disjoint subsets belonging to $\mathcal{F}$. Hence

$$m(B) = m(A) + m(BA') \geq m(A).$$

To prove the second part we note that

$$\cup_{i=1}^k A_i = \cup_{i=1}^k B_i,$$

where $B_1 = A_1$, $B_2 = A_2 A_1', \ldots,$ $B_i = A_i A_{i-1}' A_{i-2}' \cdots A_1', \ldots,$ $B_k = A_k A_{k-1}' \, A_{k-2}' \cdots, A_1'$. Then $B_1, B_2, \ldots$ are disjoint sets belonging to $\mathcal{F}$ and $B_i \subset A_i$, for all $i = 1, 2, \ldots, k$. Hence

$$\begin{aligned} m\left(\cup_{i=1}^k A_i\right) &= m\left(\cup_{i=1}^k B_i\right) \\ &= \sum_{i=1}^k m(B_i) \\ &\leq \sum_{i=1}^k m(A_i). \end{aligned}$$

This completes the proof. □

Remark 1.2.6. Property (1.2.4) is called *finite subadditivity* . If m is countably additive (1.2.4) holds with $k = \infty$, provided $\cup_{i=1}^{\infty} A_i$ belongs to $\mathcal{F}$. The same proof goes through. When $k = \infty$, (1.2.4) is called the property of *countable subadditivity* .

1.3 Probability Distributions and Elementary Random Variables

Consider a statistical experiment, the performance of which leads to an observation x in the sample space X. Very often one is not interested in the exact observation but a function of the observation. We shall illustrate by means of a few examples.

Example 1.3.1. Consider an individual performing an experiment with two elementary outcomes called 'success' and 'failure' (see Example 1.1.1). Suppose he gets a rupee if success occurs and loses a rupee if failure occurs. Then his gain can be expressed through the function f defined by

$$f(0) = -1, \;\; f(1) = +1,$$

where 0 and 1 denote the outcomes failure and success respectively.

Example 1.3.2. Suppose r objects are distributed in n cells. Assume the objects to be distinguishable from one another. Observe the configuration. It should be noted that more than one object can occupy a cell. Then the sample space X of all possible configurations contains n^r points. For each configuration $x \in X$, let $f(x)$ be the number of empty cells.

Example 1.3.3. Let a bullet be shot from a gun and let the experiment consist of observing the trajectory of the bullet. For every trajectory x, let $f(x)$ be the coordinates of the point at which the bullet hits the ground.

From the above examples we understand that the value of the function depends on the outcome which is subject to chance. Thus the value of the function varies in a 'random' manner. Till we make further progress in the subject we shall consider functions on X, which take only a finite number of values.

Let $f : X \to Y$ be a map from the sample X into a set Y. Let $\mathcal{F}$ be a boolean algebra of subsets of X, on which we shall consider probability distributions. Suppose we wish to raise the following question: what is the probability that the experiment yields as elementary outcome $x \in X$ such that the function $f(x)$ takes a given value $y \in Y$? Consider the set

$$\{x : f(x) = y\} = f^{-1}(\{y\}).$$

If we wish to find the probability of the above event, it is necessary that $f^{-1}(\{y\}) \in \mathcal{F}$. For this reason we introduce the following definition.

Definition 1.3.4. Let X be a sample space with a boolean algebra $\mathcal{F}$ of subsets of X. A map $f : X \to Y$ is called a *Y-valued simple random variable* if f takes only a finite number of values and, for every $y \in Y$,

$$f^{-1}(\{y\}) \in \mathcal{F}.$$

If Y is the real line we shall call f a *simple random variable.* We denote by $S(X, \mathcal{F})$ the set of all simple random variables.

For any set $A \subset X$, let

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$$

Then χ_A is called the *characteristic or indicator function of the set* A. If $A \in \mathcal{F}$, χ_A is a simple random variable assuming two values, namely 0 and 1. If $a_1, a_2, \ldots, a_k$ are real numbers and $A_1, A_2, \ldots, A_k \in \mathcal{F}$, then $\sum_{j=1}^{k} a_j \chi_{A_j}$ is a simple random variable. Conversely every simple random variable can be expressed in this form with A_j's pairwise disjoint. It is clear that

$$\begin{aligned} \chi_{A \cup B} &= \chi_A + \chi_B - \chi_{AB}, \\ \chi_A \chi_B &= \chi_{AB}, \\ |\chi_A - \chi_B| &= \chi_{A \Delta B}, \text{ for all } A, B \subset X. \end{aligned}$$

In particular, it follows that the set $S(X, \mathcal{F})$ of all simple random variables is an algebra under the usual operations of addition, multiplication and scalar multiplication.

Definition 1.3.5. By a *boolean space* we mean a pair $(X, \mathcal{F})$ where X is a set and $\mathcal{F}$ is a boolean algebra of subsets of X. By *a boolean probability space* we mean a triple $(X, \mathcal{F}, P)$ where $(X, \mathcal{F})$ is a boolean space and P is a probability distribution on $\mathcal{F}$. If s is a simple random variable on $(X, \mathcal{F})$ and P is a probability distribution on $\mathcal{F}$, we define the *integral* of s with respect to P as the number

$$\sum_i a_i P(s^{-1}(\{a_i\})),$$

where the summation is over all the values a_i which s can take. We denote this number by the symbol $\int s dP$ or simply $\mathbf{E}s$ when P is fixed. $\mathbf{E}s$ is also called the *expectation* of s with respect to P.

Proposition 1.3.6. If $s = \sum_{i=1}^k a_i \chi_{A_i}$, where $A_1, A_2, \ldots, A_k$ are disjoint sets in $\mathcal{F}$ and $a_1, a_2, \ldots, a_k$ are real numbers then

$$\int s dP = \sum_{i=1}^k a_i P(A_i). \tag{1.3.1}$$

Further

(i) $\int (as_1 + bs_2) dP = a \int s_1 dP + b \int s_2 dP$ for any two simple random variables s_1 and s_2 and any two real constants a and b;

(ii) the function Q on $\mathcal{F}$ defined by

$$Q(F) = \int s\chi_F dP, F \in \mathcal{F}$$

is finitely additive, i.e.,

$$Q\left(\cup_{i=1}^j F_i\right) = \sum_{i=1}^j Q(F_i)$$

whenever $F_1, F_2, \ldots, F_j$ are pairwise disjoint elements in $\mathcal{F}$;

(iii) $\int s dP \geq 0$ if $P(\{x : s(x) < 0\}) = 0$;

(iv) $\inf_{x \in X} s(x) \leq \int s dP \leq \sup_{x \in X} s(x)$.

Proof. Without loss of generality we may assume that the a_i's are distinct and $\cup_{i=1}^k A_i = X$. Then $s^{-1}(\{a_i\}) = A_i$ and the range of s is the set $\{a_1, a_2, \ldots, a_k\}$. Hence (1.3.1) follows immediately from the definition of integral. To prove property (i) we may assume that

$$s_1 = \sum_{i=1}^k \alpha_i \chi_{A_i}, s_2 = \sum_{j=1}^l \beta_j \chi_{B_j}$$

where $A_1, A_2, \ldots, A_k$ and $B_1, B_2, \ldots, B_l$ are two partitions of X into disjoint sets belonging to $\mathcal{F}$. Then

$$as_1 + bs_2 = \sum_{i=1}^{k} \sum_{j=1}^{l} (a\alpha_i + b\beta_j)\chi_{A_iB_j},$$

where the sets A_iB_j constitute another partition of X. Then

$$\begin{aligned}
\int (as_1 + bs_2)dP &= \sum_i \sum_j (a\alpha_i + b\beta_j)P(A_iB_j) \\
&= a\sum_i \alpha_i \left\{\sum_j P(A_iB_j)\right\} \\
&\quad + b\sum_j \beta_j \left\{\sum_i P(A_iB_j)\right\} \\
&= a\sum_i \alpha_i P(A_i) + b\sum_j \beta_j P(B_j) \\
&= a\int s_1 dP + b\int s_2 dP.
\end{aligned}$$

Here we have used the fact that P is finitely additive and

$$\begin{aligned}
\cup_j A_iB_j &= A_i(\cup_j B_j) = A_iX = A_i, \\
\cup_i A_iB_j &= (\cup_i A_i)B_j = XB_j = B_j.
\end{aligned}$$

Property (ii) follows from property (i). Properties (iii) and (iv) follow from (1.3.1) immediately. □

Remark 1.3.7. It should be noted that property (ii) indicates a method of manufacturing new finitely additive functions on $\mathcal{F}$ from a given one by the process of integration.

We shall now prove some elementary results by using the notion of expectation and its properties described in Proposition 1.3.6.

Proposition 1.3.8. Let $A_1, A_2, \ldots, A_n$ be subsets of X, and let $B = \cup_{i=1}^{n} A_i$. Then

$$\begin{aligned}
\chi_B &= \sum_{i=1}^{n} \chi_{A_i} - \sum_{1 \le i < j \le n} \chi_{A_iA_j} + \cdots \\
&+ (-1)^{r-1} \sum_{1 \le i_1 < i_2 \ldots < i_r \le n} \chi_{A_{i_1}A_{i_2}\ldots A_{i_r}} + \cdots \\
&+ (-1)^{n-1}\chi_{A_1A_2\ldots A_n}, \qquad (1.3.2)
\end{aligned}$$

where $A_{i_1}A_{i_2}\dots A_{i_r}$ stands for the intersection $\cap_{j=1}^{r} A_{i_j}$.

Proof. Let x belong to none of the A_i's. Then $\chi_B(x) = 0$ and every term on the right hand side of (1.3.2) is zero. If x belongs to exactly k of the sets $A_1, A_2, \dots, A_n$ then the rth term on the right hand side is $(-1)^{r-1}\binom{k}{r}$ if $r \leq k$ and 0 otherwise. Thus the right hand side of (1.3.2) is

$$\sum_{r=1}^{k}(-1)^{r-1}\binom{k}{r} = 1 - (1-1)^k = 1.$$

Then $\chi_B(x) = 1$ and (1.3.2) always holds. This completes the proof. □

Corollary 1.3.9. If P is a probability distribution on $(X, \mathcal{F})$, then for any $A_1, A_2, \dots, A_n \in \mathcal{F}$.

$$P\left(\cup_{i=1}^{n} A_i\right) = \sum_{r=1}^{n}(-1)^{r-1}S_r, \tag{1.3.3}$$

where

$$S_r = \sum_{1\leq i_1<i_2<\cdots i_r\leq n} P(A_{i_1}A_{i_2}\dots A_{i_r}). \tag{1.3.4}$$

Proof. Equation (1.3.3) follows from (1.3.2) by integration of both sides and Proposition 1.3.6. □

Proposition 1.3.10. Let $A_1, A_2, \dots, A_n$ be any subsets of X. Let B be the subset of all those points which belong to exactly k of the sets $A_1, A_2, \dots A_n$. Then

$$\chi_B = \sum_{r=k}^{n}(-1)^{r-k}\binom{r}{k}\left[\sum_{1\leq i_1<i_2<\dots<i_r\leq n}\chi_{A_{i_1}A_{i_2}\dots A_{i_r}}\right]. \tag{1.3.5}$$

Proof. Let x be a point which belongs to exactly m of the sets $A_1, A_2, \dots A_n$. If $m < k$, then $\chi_B(x) = 0$ and every term on the right hand side of (1.3.5) vanishes. If $m = k$, $\chi_B(x) = 1$. On the right hand side of (1.3.5) the term within square brackets is one if $r = k$ and zero otherwise. Thus the right hand side is also unity. Now let $m > k$. Then $\chi_B(x) = 0$. The right hand side is

$$\sum_{r=k}^{m}(-1)^{r-k}\binom{r}{k}\binom{m}{r} = \binom{m}{k}(1-1)^{m-k} = 0.$$

Thus (1.3.5) always holds and the proof is complete. □

Corollary 1.3.11. Let P be a probability distribution on $(X, \mathcal{F})$. Let $A_1, A_2, \ldots, A_n$ be n events belonging to $\mathcal{F}$, and let p_k be the probability that exactly k of the events $A_1, A_2, \ldots, A_n$ occur. Then

$$p_k = \sum_{r=k}^{n} (-1)^{r-k} \binom{r}{k} S_r \tag{1.3.6}$$

where S_r is defined by (1.3.4).

Proof. This is obtained by integrating both sides of (1.3.5) and using Proposition 1.3.6. □

Example 1.3.12. Let X be the set of all permutations of the integers $1, 2, 3, \ldots, N$ and let $\mathcal{F}$ be the boolean algebra of all subsets of X. Let P be the probability distribution which assigns the same probability $\frac{1}{N!}$ to every permutation (see Example 1.2.3). Let A_i be the set of all permutations which leave i fixed. In this case

$$P(A_{i_1} A_{i_2} \ldots A_{i_r}) = \frac{(N-r)!}{N!} \text{ if } i_1 < i_2 < \ldots < i_r.$$

Thus

$$S_r = \binom{N}{r} \frac{(N-r)!}{N!} = \frac{1}{r!}.$$

Corollary 1.3.9 implies that, the probability that a 'random' permutation leaves at least one of the i's fixed is given by the expression $1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + \frac{(-1)^{N-1}}{N!}$. An application of Corollary 1.3.11 shows that the probability of exactly m of the elements $1, 2, \ldots, N$ being fixed by a random permutation is given by the expression

$$\begin{aligned} p_m &= \sum_{r=m}^{N} (-1)^{r-m} \binom{r}{m} \frac{1}{r!} \\ &= \frac{1}{m!} \left[1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^{N-m}}{(N-m)!} \right]. \end{aligned} \tag{1.3.7}$$

We also observe that p_m converges to $\frac{e^{-1}}{m!}$ as $N \to \infty$. For every fixed N, consider the experiment of selecting a random permutation x from the set X of all permutations of $1, 2, \ldots, N$. Let $f(x)$ be the number of elements in the set $\{1, 2, \ldots, N\}$ left fixed by the permutation x. Then f is a simple random variable taking values $0, 1, 2, \ldots, N$. Then the probability that $f(x) = m$ is given by (1.3.7) for $m \leq N$. As $N \to \infty$, we get a distribution on the set of all nonnegative integers such that the probability for m is $\frac{e^{-1}}{m!}$. This is a special example of many limit theorems to come in our subject. This particular limiting distribution is a special case of the Poisson distribution.

Example 1.3.13. Suppose r objects (which are distinguishable from each other) are distributed randomly in n cells. The experiment consists in observing the configuration. Since any object can occupy any of the n cells there are n^r possible configurations. Thus the sample space X consists of n^r points. Let $\mathcal{F}$ be the class of all subsets of X. Let A_i be the set of all configurations in which cell number i is empty. By random distribution of objects we mean that all configurations have equal probability n^{-r}. Now we can ask the question: what is the probability that exactly k cells are empty? Let it be p_k. Let $i_1 < i_2 < \ldots < i_j$. Then

$$P\left(A_{i_1} A_{i_2} \ldots A_{i_j}\right) = \frac{(n-j)^r}{n^r}.$$

In the notation of (1.3.4)

$$S_j = \binom{n}{j}\left(1 - \frac{j}{n}\right)^r.$$

By Corollary 1.3.11,

$$p_k = \sum_{j=k}^{n} (-1)^{j-k} \binom{j}{k} \binom{n}{j} \left(1 - \frac{j}{n}\right)^r. \tag{1.3.8}$$

As in the preceding example the number of empty cells in any observed configuration x can be thought of as a simple random variable assuming the values $0, 1, 2, \ldots (n-1)$. The probability that this random variable assumes the value k is equal to p_k given by (1.3.8). Its expectation is $\sum_{k=0}^{n-1} k p_k$.

We see that p_k is a function of r and n, namely, the number of objects and cells. We can ask the question: what happens when r and n tend to ∞ in some suitable manner? Does p_k converge to a limit for every fixed k in such a case?

Putting $j - k = s$ in (1.3.8), an elementary computation shows that

$$\begin{aligned} p_k &= \frac{1}{k!} \sum_{s=0}^{n-k} \frac{(-1)^s}{s!} \left[\left(1 - \frac{s+k}{n}\right)^r n^{s+k} \left(1 - \frac{s+k-1}{n}\right) \right. \\ & \quad \left. \times \left(1 - \frac{s+k-2}{n}\right) \ldots 1 \right]. \end{aligned} \tag{1.3.9}$$

Since $1 - x \leq e^{-x}$ for $0 \leq x < 1$, we have

$$\left(1 - \frac{s+k}{n}\right)^r n^{s+k} \leq \left(ne^{-r/n}\right)^{s+k}.$$

Thus the term within square brackets in (1.3.9) is dominated by $\left(ne^{-r/n}\right)^{s+k}$. Let now r and n tend to ∞ in such a manner that $ne^{-r/n} \to \lambda > 0$. Thus every term in the summation (1.3.9) has absolute value dominated by $\frac{(\lambda+1)^{s+k}}{s!}$ for all r and n sufficiently large. Since the series $\sum\limits_{s=0}^{\infty} \frac{(\lambda+1)^s}{s!}$ is convergent we may take the limit in (1.3.9) under the summation sign. Since

$$\lim_{n\to\infty} \left[\left(1 - \frac{x}{n}\right)^n e^x\right]^{\log n} = 1 \text{ for all } x \geq 0,$$

and

$$\lim_{r,n\to\infty} \left(\log n - \frac{r}{n}\right) = \log \lambda,$$

we have

$$\lim_{r,n\to\infty} \left(1 - \frac{s+k}{n}\right)^r n^{s+k} =$$
$$\lim_{r,n\to\infty} \left[\left(1 - \frac{s+k}{n}\right)^n e^{s+k}\right]^{r/n} e^{(s+k)\left(\log n - \frac{r}{n}\right)} = \lambda^{s+k}.$$

Thus

$$\lim_{r,n\to\infty} p_k = e^{-\lambda}\frac{\lambda^k}{k!}.$$

The distribution on the set of all nonnegative integers with probability for the single point set $\{k\}$ equal to $e^{-\lambda}\frac{\lambda^k}{k!}$ is called the *Poisson distribution* with parameter λ.

Example 1.3.14. Consider now r indistinguishable objects being placed randomly in n cells. In this case a configuration corresponds to a vector $(x_1, x_2, \ldots, x_n)$ where x_i is the number of objects in the ith cell. If X is the sample space of all possible configurations we shall find out the number of points in X. We can look at a configuration as

$$\underbrace{00\ldots0}_{x_1} \mid \underbrace{00\ldots0}_{x_2} \mid \cdots\cdots \mid \underbrace{00\ldots0}_{x_n},$$

where 0 denotes the object and $\mid$ denotes a wall of the cell. Since there are n cells the number of vertical bars in the above picture is $n-1$.

The total number of positions occupied either by an object or a bar is $n-1+r$. Out of these, r positions are taken over by the objects. Thus the total number of configurations is $\binom{n-1+r}{r}$. If all configurations are equally likely, then each point of X has probability $\frac{1}{\binom{n-1+r}{r}}$. This distribution is known as *Bose-Einstein Statistics.* This distribution, discovered by the physicist S. N. Bose in the context of quantum statistical mechanics, is basic in the theory of Bose-Einstein statistics.

In the above problem, let A_i be the event that the ith cell is empty. Then for $i_1 < i_2 < \ldots < i_j$, $A_{i_1}A_{i_2}\ldots A_{i_j}$ is the event that the j cells $i_1, i_2, \ldots, i_j$ are empty. The number of such configurations is the number of ways in which r identical objects can be distributed in $n-j$ cells. Thus

$$P(A_{i_1}A_{i_2}\ldots A_{i_j}) = \binom{n-j+r-1}{r} \Big/ \binom{n+r-1}{r}.$$

If p_k denotes the probability of exactly k cells being empty, then

$$p_k = \sum_{j=k}^{n}(-1)^{j-k}\binom{j}{k}\binom{n}{j}\binom{n-j+r-1}{r} \Big/ \binom{n+r-1}{r} \qquad (1.3.10)$$

Exercise 1.3.15. In (1.3.10), $p_k \to e^{-\lambda}\frac{\lambda^k}{k!}$ for $k = 0, 1, 2, \ldots$, if $n, r \to \infty$ in such a manner that $\frac{n^2}{r} \to \lambda$.

Example 1.3.16. Consider an urn with m white balls and n black balls. Choose a sample of k balls at *random without replacement.* This means that the probability for any sample of k balls out of the total of $m+n$ balls is the same and hence is equal to $\frac{1}{\binom{m+n}{k}}$. We now ask the question: what is the probability of r white balls occurring in the sample? If there are r white balls in the sample then there are $k-r$ black balls in the same sample. Such a choice can be made in $\binom{m}{r}\binom{n}{k-r}$ ways. If we denote the required probability by p_r then

$$p_r = \frac{\binom{m}{r}\binom{n}{k-r}}{\binom{m+n}{k}}, 0 \le r \le \min(k, m).$$

The above distribution on the set of integers $0, 1, 2, \ldots \min(k, m)$ is known as the *hypergeometric distribution.*

Exercise 1.3.17. Consider an urn with balls of r different colours. Let there be n_1 balls of the first colour, n_2 of the second and so on.