Synthesis Lectures on
Engineering, Science, and Technology

SYNTHESIS
COLLECTION OF TECHNOLOGY

Rajan Chattamvelli · Ramalingam Shanmugam

# Random Variables for Scientists and Engineers

Springer

# Synthesis Lectures on Engineering, Science, and Technology

The focus of this series is general topics, and applications about, and for, engineers and scientists on a wide array of applications, methods and advances. Most titles cover subjects such as professional development, education, and study skills, as well as basic introductory undergraduate material and other topics appropriate for a broader and less technical audience.

Rajan Chattamvelli · Ramalingam Shanmugam

# Random Variables for Scientists and Engineers

Springer

Rajan Chattamvelli
School of Computer Science and Engineering
Amrita University
Amaravati, Andhra Pradesh, India

Ramalingam Shanmugam
School of Health Administration
Texas State University
San Marcos, TX, USA

ISSN 2690-0300 ISSN 2690-0327 (electronic)
Synthesis Lectures on Engineering, Science, and Technology
ISBN 978-3-031-58930-0 ISBN 978-3-031-58931-7 (eBook)
https://doi.org/10.1007/978-3-031-58931-7

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

# Preface

This book gives an introduction to random variables (RVs) and their transformations. The aim is to give a clear exposition of mathematical expectations, univariate RVs, and joint distributions.

Chapter 1 introduces RVs and mathematical expectations. Discrete and continuous random variables and their basic properties are discussed. Cumulative distribution functions (CDF), survival functions (SF), quantile functions (QF), and their properties are detailed. Arithmetic operations on RVs give rise to new RVs. A simple technique to find the expectation of functions of RVs is given. This is followed by a discussion on moments (ordinary, central, factorial), variance, and co-variance as expected values. The conditional expectation is introduced and its applications to finding the moments of infinite mixture distributions such as noncentral chi-square, noncentral beta, and noncentral F distributions are demonstrated. Chapter 1 ends with a discussion on the applications of random variables.

Chapter 2 discusses the distributions of functions of single RVs. Topics discussed include distribution of absolute value, method of distribution functions (MoDF), change of variable technique (CoVT), distribution of sums, squares, square-roots, reciprocals, trigonometric, and transcendental functions, minimum and maximum, integer and fractional parts, arbitrary functions, and ratio of sums. A summary table of common single-variable transformations is provided in Sect. 2.11.1. These results are used to express the mean deviation of continuous distributions as a simple integral from lower limit to F(mean) where F() is the CDF. The chapter ends with a discussion of transformations of normal variates and some applications of functions of random variables in various fields.

Distribution of functions of several random variables is introduced in Chap. 3. Marginal and conditional distributions are briefly discussed. The Jacobian of matrix transformation is described and its applications in various fields are cited. This is illustrated in finding the distribution of a variety of transformations including products, ratios, and nonlinear functions of two or more RVs. A “do-little” technique to quickly find the Jacobian of transformation of random variables useful in statistics is described. Plane-polar, spherical-polar, cylindrical-polar, toroidal-polar, Helmert, and Rosenblatt transformations are also discussed. Integral calculus is heavily used in this chapter. A summary table of common transformation in two variables is provided in Sect. 3.4. The chapter ends with a discussion on copula-based methods.

Suggestions for changes are always welcome. For any suggestions on improvement please contact rajancv@am.amrita.edu.

Amaravati, India
February 2024

Rajan Chattamvelli
Ramalingam Shanmugam

# Contents

# About the Authors

**Rajan Chattamvelli** is a professor in the school of computer science and engineering at Amrita University, Amaravati. He has published more than 20 research articles in international journals of repute and at various conferences. His research interests are in computational statistics, design of algorithms, parallel computing, cryptography, data mining, machine learning, and big data analytics. His prior assignments include Denver Public Health, Colorado; Metromail Corporation, Lincoln, Nebraska; Frederick University, Cyprus; Indian Institute of Management; Periyar Maniammai University, Thanjavur; Presidency University, Bengaluru; and VIT University, Vellore.

**Ramalingam Shanmugam** is an honorary professor in the school of Health Administration at Texas State University, San Marcos. He is the editor of the journals: *Advances in Life Sciences*, *Global Journal of Research and Review*, *International Journal of Research in Medical Sciences*, *Kenkyu Journal of Epidemiology and Community Medicine*, and *Journal of Obesity and Metabolism* and book-review editor of the *Journal of Statistical Computation and Simulation*. He has published more than 200 research articles and 120 conference papers. His areas of research include theoretical and computational statistics, number theory, operations research, biostatistics, decision making, infectious disease modeling, patient risk management, cost-effective analysis, and epidemiology. His prior assignments include University of South Alabama, University of Colorado at Denver, Argonne National Labs, Indian Statistical Institute, and Mississippi State University. He is the president of the San Antonio chapter of the *American Statistical Association* and a fellow of the International Statistical Institute.

# List of Tables

# Mathematical Expectation

# 1

This chapter introduces random variables and mathematical expectation. Discrete and continuous random variables and their basic properties are discussed. A simple technique to find the expectation of functions of random variables is given. This is followed by a discussion on moments (ordinary, central, factorial) and variance as expected values. The conditional expectation is introduced and its applications to finding the low-order moments of mixture distributions are demonstrated. A brief discussion of inverse and incomplete moments, and distance as expected value follows it. The chapter ends with some common applications of random variables.

## 1.1 Meaning of Expectation

The concept of "expected value" appeared for the first time in the works of Christian Huygens (1629–1695) around 1657. It was used to predict the possible gains in gambling and games of chance. It can be associated either with a single random variable (RV) or a well-defined function of the RV. Location measures (such as the mean, median, mode) condense the information in a sample as a single number (in univariate case). Analogous measures are needed to succinctly present the characteristics of statistical populations or random experiments. This is where the concept of expectation comes in. The functional form of the population is known precisely in most of the discussions below. But theoretically, the concept is valid even when the exact form is either unknown or is partially known (as in random experiments involving circuits, transmission medium, moving objects, etc.) The expected value of observed phenomena are applicable in the long-run during which an event of interest is going to occur repeatedly under identical experimental conditions. This may sometimes be observed from past data. For instance, consider the price of a stock that fluctuates randomly

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024 
R. Chattamvelli and R. Shanmugam, *Random Variables for Scientists and Engineers*,
Synthesis Lectures on Engineering, Science, and Technology,
https://doi.org/10.1007/978-3-031-58931-7_1

over time. We could find the expected value of stock price by averaging the observed values over a suitable time period using an uncertainly measure (probability for it to rise, fall or remain steady). Insurance companies use expected values to predict various quantities. For example, flooding and power outages are quite common during monsoon season in tropical cities with poor drainage facilities. If a weatherman predicts that a heavy thunderstorm is likely in the next few days with a probability of 0.90, the insurance companies can use this information using past data to estimate the expected amount on insurance claims that will be received after the event. This is discussed further in Sect. 1.3. The notion of mathematical expectation (or simply called expectation) relies on one or more *random variables* defined below.

## 1.2 Random Variables

The concept of RV is of prime importance in mathematical expectation. It is defined on the sample space of a random experiment, which is an experiment that can be repeated any number of times under (more or less) identical conditions.

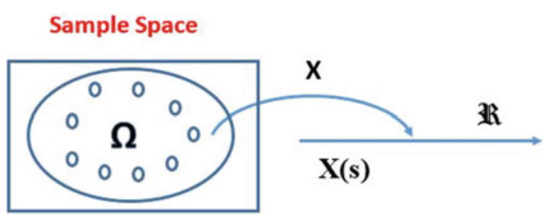
**Definition 1.1** The set of all possible outcomes of a random experiment is called the sample space. It is usually denoted by the Greek letter Ω or the letter S.

The outcome of a random experiment can be given names, labels or an enumeration. Thus when a coin is tossed, the possible outcomes are represented as {H, T}, {Head, Tail} or simply as {0, 1} where 0 denotes the nonoccurence and 1 denotes the occurrence of an event.

**Definition 1.2** A random variable is a function defined on the sample space of a random experiment that maps each possible outcome of the sample space to real numbers such that the associated probabilities sum to one.

Mathematically, an RV is a rule that assigns a unique numerical value to each event (outcome) of a random experiment (Fig. 1.1). An indicator function is a special type of RV in which each element in the sample space is mapped to either 0 or 1. If E is an arbitrary event

**Fig. 1.1** Random variable maps sample space to real line



$$I_E(s) = \begin{cases} 1 \text{ if } s \in E \\ 0 \text{ if } s \notin E. \end{cases}$$

### 1.2.1 Realisation of Random Variables

An RV ensues when (i) an experimenter performs a random experiment, (ii) defines a random experiment on a hypothetical experiment (like tossing a die or coin), (iii) observes the results of an experimental outcome, or (iv) a physical or natural process generates data that can be approximated by a statistical law. Such occurrences are denoted by lowercase letters (in which case it is an implicit assumption that the corresponding uppercase letter denotes the RV). RandVar is an R-package to implement random variables. (r-project.org)

Random processes are RVs where the values of a variable vary systematically over time. Note that the outcomes can be anything including numbers, labels, symbols or even text strings. Thus in an industrial experiment that checks whether a machine or part is defective, the outcomes could be {DD, DN, ND, NN} where 'D' denotes defective or non-working and 'N' denotes non-defective or in good working condition. There are multiple ways in which these outcomes can be mapped numerically. If the aim of a study is to identify defectives, we could map 'D' to a '1' and 'N' to a '0' so that the possible values the random variable can take are {2, 1, 0}. However, some people prefer to always map a defective to a zero and non-defective to a one. In this case the probabilities simply get reversed. An RV can be denoted as X:$\Omega \rightarrow \mathbb{R}$ where the values that it takes in $\mathbb{R}$ are known from the mapping used (see Fig. 1.1). All RVs are denoted by uppercase letters and particular values by lowercase letters in the following discussion.

We will denote "distributed as" by the symbol "$\sim$" (which is the *tilde* symbol), and 'independently and identically distributed' as IID. Abbreviations will be used for distributions in an unambiguous way (POIS for Poisson, CUNI for continuous uniform, DUNI for discrete uniform, EXP for exponential, BINO(n, p) for binomial, etc.). Thus X$\sim$CUNI(0, 1) = U(0, 1) is read as "X is distributed as continuous uniform in [0, 1]".

### 1.2.2 Discrete and Continuous Random Variables

An RV can be discrete, continuous or mixed type. Among these discrete RVs are often used with count data, and continuous RVs are used when measurements are done by machines or computed using mathematical formula (like BMI).

**Definition 1.3** An RV is discrete if the set of possible values (outcomes) that it could take is finite or countably infinite. Mathematically, X is discrete if x$\in \{x_1, x_2, \ldots, x_n\}$ or x$\in \{x_1, x_2, \ldots, \}$. This concept is easy to understand for discrete RVs as the number of events in the sample space (domain) is countably finite. The domain can also be a part of an integer (such as half-integer). As an example, suppose X represents the number of medical leaves taken by an employee where the employer allows either full-day leave or half-day leave types only. In this case the domain can be integers or half-integers per month. The domain of several discrete RVs are an ordered sequence. Consider a vehicle insurance company who screens new customers for the number of past accidents. A great majority of customers might have no severe accidents at all (x = 0). There could be several customers with one accident (x = 1), two accidents (x = 2) and so on. If the number of applicants is large, we expect x to take values in (0, 1, . . ., m) where m is a fixed number. However, there could be breaks in such a sequence in which case we simply insert missing values and assume the corresponding frequency to be zero. This makes it into an ordered sequence.

In some practical applications, the discrete RV represents *count* data that may occur over time or space. As examples, the number of defective items in a shipment, number of patients with a particular symptom, number of alpha-particles emitted by a radioactive source in a small time-interval, number of crossovers that have occurred in a genome, number of earthquakes in a particular locality are all examples of count data. Here the last two examples are occurrences over space while others occur over time. Any number of RVs can be defined on a given sample space. These may be related or independent. There are some numerical values associated with each outcome of a random experiment. For instance, consider the toss of a die with six faces. The possible outcomes are $\{1, 2, 3, 4, 5, 6\}$, each with probability 1/6. If X denotes the face that turns up, we express it mathematically as f(x) = 1/6 for x = 1, 2, . . . 6. We have simply assigned a mathematical function to each outcome of a random experiment. This experiment is called equally likely (and the RV as equiprobable) because the associated probabilities are all equal. The corresponding mapping from the sample space to the real line is called probability mass function (PMF) in the discrete case and probability density function (PDF) in the continuous case. This is the most common way to define a discrete RV. Similarly, f(x) = $q^x$p is a mathematically defined RV associated with a countably infinite sample space consisting of the sequence 0, 1, 2, . . . , $\infty$ associated with a random experiment (with two mutually exclusive outcomes) that is repeated until a special event occurs. The above RV occurs when a coin is tossed repeatedly until a Head (with $\Pr(Head)$ = p) occurs. The values of a well-defined RV are subject to chance (i.e. it is stochastic). Thus even if we toss a coin thousands of times, we can't predict in advance what the next outcome is going to be. Some RVs are time-dependent. Consider the number of phone calls passing through an automated telephone exchange in a fixed time interval (say 10 s). This varies from time to time, but could be modeled using a statistical law if the time interval is properly chosen. Similarly, the number of phone calls received at an emergency number or fire-station can be modeled using a statistical law such as the Poisson law. As these are rare events, the time interval is large.

There is one more way to define discrete RVs. It is called complete enumeration method. Consider the RV p(1) = 0.2, p(2) = 0.6, p(3) = 0.2. Here x takes 3 values $\{1, 2, 3\}$. It is a well-defined RV as the probabilities add up to one. This can also be written as p(x = 1) = 0.2, p(x = 2) = 0.6, p(x = 3) = 0.2 for a univariate RV X. This notation can be extended to