Statistical Models in Engineering

Statistical Models in Engineering

Gerald J. Hahn & Samuel S. Shapiro
Research and Development Center, General Electric Company

JOHN WILEY & SONS
New York • Chichester • Brisbane • Toronto

ISBN 0 471 33915 6

Library of Congress Catalog Card Number: 67-12562

Printed in the United States of America

10 9

Statistical Models in Engineering

Wiley Series on Systems Engineering and Analysis

HAROLD CHESTNUT, Editor

Chestnut

Systems Engineering Tools

Wilson and Wilson

Information, Computers, and System Design

Hahn and Shapiro

Statistical Models in Engineering

To Bea and Yevette for their help, patience, and understanding
and
To Adrienne, Susan, and Judith and Bonnie and Michael

SYSTEMS ENGINEERING AND ANALYSIS SERIES

In a society which is producing more people, more materials, more things, and more information than ever before, systems engineering is indispensable in meeting the challenge of complexity. This series of books is an attempt to bring together in a complementary as well as unified fashion the many specialties of the subject, such as modeling and simulation, computing, control, probability and statistics, optimization, reliability, and economics, and to emphasize the interrelationship between them.

The aim is to make the series as comprehensive as possible without dwelling on the myriad details of each specialty and at the same time to provide a broad basic framework on which to build these details. The design of these books will be fundamental in nature to meet the needs of students and engineers and to ensure they remain of lasting interest and importance.

Preface

This book is the outgrowth of many years of consulting with electrical, mechanical, chemical, industrial, and reliability engineers in widely divergent product areas in the General Electric Company and other organizations. In this work we have frequently felt the need for a detailed treatment, *directed at engineers*, on the use of statistical models to represent physical phenomena, and in this book we attempt to satisfy this need.

Our main purpose is to show the engineer responsible for a product whose performance is subject to chance fluctuations how to choose a reasonable statistical model and how to use this model in the evaluation of practical problems. Our choice of subjects is in line with this objective.

After some introductory comments in Chapter 1, the basic concepts of probability and distributional models are developed in Chapter 2. Chapters 3 and 4, which deal with continuous and discrete distributions, respectively, provide a detailed development of many models and the type of situations to which each is applicable. The relevance of the widely (and sometimes incorrectly) used normal distribution is considered, as is the applicability of the equally popular (and equally misused) exponential distribution in reliability problems. Alternate models, such as the gamma distribution, the beta distribution, the Weibull distribution and others are introduced and their use discussed. We have attempted to summarize the information in these chapters in a summary figure and table, which we hope will provide a fast reference guide.

The choice of a distribution among those presented in Chapters 3 and 4 is based principally on theoretical considerations. In some situations the underlying physical phenomena are insufficiently known for this to be possible, and a distributional fit must be evolved from the given data. Some versatile distributions for this purpose, known as Johnson and Pearson distributions, are discussed in Chapter 6.

Often the engineer is concerned with a system, ranging from a simple circuit to a complicated space vehicle, in which each component is subject to statistical variability. From knowledge of the system structure and the nature of the component behavior we draw conclusions about system performance. Because of its importance to engineering problems, we cover this problem in much greater detail than has been done to date in the literature accessible (and understandable) to the practicing engineer. In particular, an exact method, known as the transformation of variables, is described in Chapter 5. The approximate methods of generation of system moments (sometimes referred to as "statistical error propagation") and Monte Carlo simulation are considered in Chapter 7. The three methods are compared and the practical applicability of each is discussed. In the consideration of generation of system moments, expressions for evaluating the average, the spread, the symmetry, and the peakedness of the distribution of system performance are given and used to obtain distribution fits with the empirical models developed in Chapter 6. In the discussion of Monte Carlo simulation we include a description of methods for generating random values from the various models introduced in Chapters 3, 4, and 6 and also consider the number of simulations required to obtain a prespecified degree of precision.

The last chapter is devoted to methods of evaluating the adequacy of a chosen model and deals with both graphical and analytical procedures. The use of probability plots for estimating distribution parameters, as well as for evaluating the applicability of the model, is described for many of the distributions previously introduced (and not just for the normal distribution). Normal probability plots of data from various known distributions are shown to indicate the sensitivity of the method. In the second part of the chapter we present some new and powerful statistical tests for evaluating the assumption of normal and exponential distribution models, developed recently by one of us (S. S. Shapiro in conjunction with M. B. Wilk) and published previously only in statistical journals.

Except for indicating procedures for estimating parameters for many of the models and determining confidence intervals for some of these parameters, we include little of the material generally associated with the term "statistical inference." These subjects are excellently covered in many available books (see bibliography). Their omission here has hopefully allowed us to cover the subject of statistical models in some depth within the confines of a reasonably small volume.

We assume that the reader has the minimum undergraduate engineering mathematics and professional sophistication; however, no previous formal training in statistics is required. Those who have had an introductory course in probability and statistics will be acquainted with much of Chapter 2 and parts of Chapters 3 and 4 but should find most of the remainder of the book new. We attempt to give the theoretical justification necessary for the correct application of the methods, but in most cases derivation of results is either omitted or relegated to the appendices. A large number of illustrative examples are given, most of which have been adapted from actual problems; however, the name of the product or the actual numerical values have been changed "to protect the innocent."

Although the book is directed principally at practicing engineers, it has been used successfully as a supplementary text in an introductory graduate course in statistics for engineers at Union College, Schenectady, New York, with whose adjunct faculty we are both associated.

We are indebted to the Literary Executor of the late Sir Ronald A. Fisher, F.R.S., Cambridge, to Dr. Frank Yates, F.R.S., Rothamsted and to Messrs. Oliver & Boyd Ltd., Edinburgh, for permission to reprint Table III from their book *Statistical Tables for Biological, Agricultural, and Medical Research.* We are also indebted to Professor E. S. Pearson and the Biometrika Trustees for permission to reprint various tables from *Biometrika Tables for Statisticians*, Volume 1, by E. S. Pearson and H. O. Hartley; to Dr. Harold Chestnut and John Wiley and Sons for permission to reprint various figures from *Systems Engineering Tools*; to Dr. A. Hald and John Wiley and Sons for permission to reprint material from *Statistical Theory with Engineering Applications* and from *Statistical Tables and Formulas*; to Professors H. F. Dodge and H. G. Romig and John Wiley and Sons for permission to reprint material from *Sampling Inspection Tables*; to the Computational Laboratory, Harvard University, and the Harvard University Press for permission to reprint material from *Tables of the Cumulative Binomial Probability Distribution*; to the RAND Corporation and The Free Press, Macmillan, for permission to reprint material from *A Million Random Digits with 100,000 Normal Deviates*; and to Professors N. L. Johnson and E. S. Pearson for permission to reprint additional published and unpublished material.

It is a pleasure to acknowledge the substantial help we have had from many friends and associates in the preparation of this book. In particular, we are indebted to Messrs. Joseph J. Buckley, Charles B. Chandler, Ralph R. Duersch, Joel J. Fleck, Albert Fox, Paul Gunther, George E. Henry, John L. Jaech, William J. MacFarland, James E. Mulligan, James L. Myracle, Wayne B. Nelson, John Oldenkamp, Glenn M. Roe, Henry J. Scudder, III, and Hermann von Schelling for their constructive comments on preliminary versions of the manuscript; to Messrs. Harold Chestnut, Gilbert C. Dodson and Richard L. Shuey for their encouragement and guidance; to Professor Alfred L. Thimm and Dean William L. Weifenbach for arranging reproduction of the manuscript at Union College; to Mrs. Clara Doring for her excellent typing; and last (but not least) to Mrs. Gerald J. Hahn for her able preparation of the illustrations.

Schenectady, New York GERALD J. HAHN
January 1967 SAMUEL S. SHAPIRO

Contents

Chapter 1

Introduction

An important development in modern science and engineering is the study of systems in a probabilistic rather than a deterministic framework. The modern engineer, like his counterpart in many other fields, is becoming increasingly aware that deterministic models are inadequate for designing or evaluating the complex equipments of the twentieth century. Performance of supposedly identical systems differs because of many factors, such as component differences and fluctuations in the operating environment. Consequently, the engineer must be concerned with statistical models that describe these variations.

Once an understanding of such basic concepts is gained, it will seem natural to speak of the *statistical distribution* for output voltage of some system, rather than to specify only its design value or to talk about the *probability* that a component will not fail during a given time, instead of merely saying that it is not expected to fail. It will also seem reasonable to calculate circuit tolerances by statistical methods, rather than by the generally overconservative method of combining worst cases.

The application of statistical models in engineering is recent. The role of statistics in industry was limited prior to World War II. However, the widespread use of statistics in the war effort led to the rapid growth of the field in the immediate postwar period in such areas as quality control and communication theory. The present-day need for complex systems with high reliability has provided an even greater impetus to the more general use of statistics in industry.

As in other rapidly developing fields, education has lagged behind need, and only in very recent years have statistics courses found their way into the typical engineering curriculum. It is fair to say that even today probability and statistics in general, and the use of statistical models in particular, carry with them an aura of mystery for many practicing engineers. This should not be so, because statistics is no more difficult to comprehend than many other engineering disciplines.