COOLIDGE ON PROBABILITY

An Introduction to Mathematical Probability. By Julian L. Coolidge. Oxford, Clarendon Press, 1925. xi+215 pp.

This book meets a real need for a textbook on probability written in English and giving an introduction to a wide range of topics. The work is based on lectures by Professor Coolidge given at Harvard University. The purpose expressed in the preface is to give the mathematical basis underlying each of the important applications of probability rather than to write a treatise on games of chance, or errors of observations, or statistics, or statistical mechanics, or insurance. In the opinion of the reviewer, the author has succeeded well in carrying out his purpose in brief space.

The first chapter deals with the meaning of probability. Various definitions of probability are described. Then by means of three broad empirical assumptions the statistical definition of probability is given and adhered to throughout the book. On the whole, a good case seems to be made for the underlying rationale adopted. Moreover, the definition adopted surely lends itself to important statistical applications. The elementary principles for the combination of probabilities and the meanings of expectation and risk are clearly developed in the second chapter. The third chapter deals with the Bernoulli theorem which is rather commonly regarded as the central theorem underlying the main applications of probability in practical statistics. A very commendable feature of this chapter is the clearness and emphasis with which approximate results are set down as approximations. For example, it is emphasized that the Gaussian probability integral is merely an approximation to the probability, in repeated trials, that the discrepancy is numerically equal to or less than an assigned value. Hence, it is properly called a "fortunate accident" that

$$\frac{2}{\pi}\int_0^\infty e^{-z^2}\,dz=1\,,$$

and it is another fortunate accident that

$$\frac{1}{\sqrt{2\pi npq}}\int_0^\infty x^2 e^{-\frac{x^2}{2npq}}\,dx=npq\,.$$

An unfortunate slip occurs on page 36. By simply taking the sense of the inequality $n-r\leqq r$ to be $n-r\geqq r$, certain incorrect conclusions are drawn. These errors are to be corrected at once by an erratum slip. The chapter on mean values and dispersion seems to the reviewer to be one of the most interesting chapters of the book, particularly because it presents the author's extension of the Lexis theory to mean values of observations instead of restricting the theory to mean values of relative frequencies. The chapter on probability of causes is a brief treatment along fairly conventional lines. Bayes' principle is regarded as open to grave question. Certain necessary conditions and limitations on the use of the principle are well discussed on page 100. No attention seems to be given to the claim of Karl Pearson for his recent improved basis for the probability of causes published in BIOMETRIKA, (vol. 13, 1920–21, p. 1). After considering objections to the Bayes' formula a question is asked and answered as follows: "Why not, then, reject the formula outright? Because, defective as it is, Bayes' formula is the only thing we have to answer certain important questions which do arise in the calculus of probability." In this connection the reader will surely enjoy the human touch in the following remark: "Therefore we use Bayes' formula with a sigh, as the only thing available under the circumstances: 'Steyning tuk him for the reason the thief tuk the hot stove—bekaze there was nothing else that season.' " The two chapters on the distribution and treatment of errors of observations are of more than ordinary interest. The feature of these chapters which appeals to the reviewer particularly is the explicit statement of the underlying assumptions. The derivation of the law of errors in many variables is taken from the author's article *The Gaussian law of errors for any number of variables*, TRANSACTIONS OF THIS SOCIETY, 1923. The idea of correlation comes up in connection with this generalized theory. Correlation is thus approached from the frequency surface standpoint. Later the regression equation standpoint is dealt with briefly. We are told (p. 149) that "The great trouble at present with the theory of correlations seems to be that there is no general agreement as to how large r^2 must be in order so that we may safely conclude there is a real connection between the two sets of phenomena." In the opinion of the reviewer this trouble may be largely overcome in statistical practice if we keep clearly before us the purpose for which we are to use the correlation coefficient. For example, if the purpose is merely to establish the existence of correlation, then a very small value of r may indicate a connection. In this case, the comparison of the correlation coefficient should be with multiples of its probable error. On the other hand, if the purpose is to predict the value of a variable y corresponding to an assigned x, it is not likely that a useful prediction can be made unless r is relatively large. Between these two extremes, we may find a moderate value of r useful in predicting the average values in the arrays of y's corresponding to assigned x's. Thus, the trouble mentioned is likely to be overcome in part at least when the purpose is clearly defined. The brief chapter on statistical theory of gases will serve the very useful purpose of introducing an important application of probability to a wider group of mathematical readers. In the chapter on principles of life insurance the difficulties of obtaining probabilities of life and death from statistical data are clearly discussed. The statement (p. 194) that "A life table calculated by Makeham's formula is better than any but the best table calculated by other means" is likely to be misleading. The Makeham formula gives a useful approximation to certain tables. For other well prepared tables such as the recent American-

Canadian Men Table the approximation is not satisfactory for many purposes.

The following minor typographical errors should be noted:

Page 46, formula (11), $\dfrac{1}{\sqrt{2npq}}$ should read $\dfrac{1}{\sqrt{2\pi npq}}$.

Page 48, formula (14), the $\sqrt{2npq}$ should read $\sqrt{2\pi npq}$.

Page 154, lines 19 and 24, the words "normal" and "residual" should be interchanged.

Page 170, line 6, $s_m = \sum_{k=0}^{k=m} x_k y_k$ should read $\dfrac{1}{m!} \sum_{k=0}^{k=n} x y_k^m$.

Then formula (36) following should begin with a factor $m!$

The author directed my attention to the last two errors noted.

In conclusion, the space available permits the reviewer to add only one or two instances of the very fine human touch that gives much life to the exposition. For example, on page 170 we note reference to "this unlovely formula." In addition to certain classical problems on probability involving a human element, we are pleased with such new ones as the following (p. 25) related to De Montmort's example and ascribed to one of the author's colleagues: "If all the inhabitants of Chicago should meet together in one place and get extremely drunk, and then try to go home by guesswork, the chances that at least one would get back to his own bed is almost two out of three." Then it is aptly remarked that "this is one of those cases where it is fortunate that the probability can be calculated beforehand, and we are not forced to seek it experimentally."

H. L. RIETZ