

BOOK REVIEWS

Foundations of information theory. By Amiel Feinstein. McGraw-Hill Electrical and Electronic Engineering Series, New York, McGraw-Hill Book Company, Inc., 1958. 10+137 pp. \$6.50.

In the event that the mathematical readers of this review have been approached by electrical engineers interested in understanding the Radon-Nikodym theorem, they should not be surprised. The question may have been inspired by attempts to read the volume under review. Interest of electrical engineers in mathematically sophisticated probability theory is reaching the dimensions of a sociological phenomenon in several universities. Registration of electrical engineers in introductory evening graduate courses in probability theory on the order of five times the day enrollment from other university sources in corresponding day courses is not at all uncommon.

For those mathematicians who do not care to abdicate from this rapidly growing area of mathematical training for engineers, it would be well to present a little of the background. The term, information theory, may be best understood by considering the content of the I.R.E. *Transactions on information theory*. This content is characterized by the use of probability and statistics in the problems of communication and control, and for engineering publication is sophisticated in mathematical methods, with material close to parts of the *Annals of Mathematical Statistics*. Some of the broad categories covered in this publication follow: (1) *Stochastic* (most frequently gaussian) *processes*. The material here is influenced by the pioneering papers of S. O. Rice, *Mathematical analysis of random noise*, Bell System Tech. J. vol. 23 (1944) pp. 282-332; vol. 24 (1945) pp. 46-156 and is available in textbook form (reasonably packaged for the engineering reader) in Davenport and Root, *An introduction to the theory of random signals and noise*. A good deal of this material is treated at a mathematical level not readily accessible to the average engineer in J. L. Doob, *Stochastic processes*.

A typical result of the kind included here is the Kac-Siebert result on the theory of noise in radio receivers with square law detectors. (2) *Prediction and filtering*. This includes developments from Wiener's generalization of regression theory applied to problems of automatic control. This is conveniently packaged for the engineer in the textbook by Laning and Battin. A typical result of interest to engineers is the one by Zadeh and Ragazzini on polynomial signal with finite observation time. Some recent mathematical develop-

ments in the area of prediction not covered by the Transactions is the work of Helson and Lowdenslager, *Prediction theory and Fourier series in several variables*, Acta Math. vol. 99 (1958) pp. 165–202. (3) *Applications of modern theories of statistical inference to detection of signal in noise, radar tracking, etc.* This includes the application of the Neyman-Pearson theory of testing hypotheses and the Cramér-Rao inequality in estimation. A good deal of this is formulated in the context of statistical decision theory. From the engineering point of view the papers of Middleton and his collaborators have been most voluminous in this area, e.g. Middleton and Van Meter, *Detection and extraction of signals in noise from the point of view of statistical decision theory*, J. Soc. Indust. Appl. Math. vol. 13 (1955) pp. 192–253; vol. 114 (1956) pp. 86–119. A reasonably compact summary available to electrical engineers can be found in the last chapter of the previously cited book of Davenport and Root. (4) *Theory of coding so as to combat noise*. It is with information theory in this sense that Feinstein's book is concerned.

C. E. Shannon in his fundamental paper *A mathematical theory of communication*, Bell System Tech. J. vol. 27 (1948) pp. 379–423, pp. 623–656, set up a mathematical model for the transmission of information. The model insisted upon the statistical nature of information, a point which Norbert Wiener had been making for years. It gave a measure for information, definition to channel capacity, and in the coding theorem had a substantial mathematical theorem with surprising practical implications. The paper was written for an engineering audience, but, while mathematically sophisticated for that audience, was short on rigor by the strict mathematical canon. The work was taken up eagerly in electrical engineering circles, giving rise to the IRE Professional Group in Information Theory, its Transactions, cited earlier, and the journal *Information and control* as well as several international symposia; but it was regarded somewhat prissily by mathematical and statistical circles in this country (an attitude somewhat strengthened by poor early textbooks and some periodical contributions from the lunatic fringe). This could not be said about the English statisticians, who perhaps have a stronger tradition for serious consideration of applications and for recorded public discussions. The centers of activity in this country have been Bell Telephone Laboratories (Shannon, McMillan, Lloyd, Gilbert, Slepian, and Hamming with Mandelbrot and Feinstein temporarily) where work has included extending the theory in depth and generality, presenting it in a more rigorous form, as well as implementing it with concrete codes; and Massachusetts Institute of Technology

(Elias, Fano, Feinstein (no longer there), Mandelbrot and Schutzenberger (temporarily), and now Shannon) where work has been heavily concentrated in a series of doctoral dissertations attempting to realize the practical implications of the coding theorem.

While the Russian mathematicians cannot be said to have invented information theory they can justly claim to be the first substantial mathematical group to give it appreciation, critical examination, clear exposition, and serious extensions. In fact the *Uspekhi* articles by Khinchin, now available in English by way of a paper-bound translation *Mathematical foundations of information theory*, Dover Publications, 1957, seemed to raise the subject to mathematical respectability in this country. Its clear statement of the open issues, in particular the relation between ergodic and stationary capacity as well as the converse to the coding theorem, led to a series of recent papers by Wolfowitz, Feinstein, and Blackwell, Breiman, and Thomasian in the United States. Among the Russian contributors to this subject besides Khinchin are Kolmogorov, Gelfand, and Yaglom. The Russian journal *Teoriia Veroiatnostei i ee Primeneniia* (Moscow) publishes contributions to information theory at the rate of half a dozen to a dozen articles per year. Excerpts are frequently available in English in *Automation express*.

While information theory is a subject which can be treated either from a mathematical or applied point of view, the aim of the author is to give a "concise but rigorous exposition of the fundamentals of the mathematical theory of information" more or less limited to "the general properties of channels of various types proving various theorems concerning their suitability for transmitting information" but, in general, not going into the implementation of these theorems. The spirit of his approach is given by "Our subject matter is mainly deductive in nature, i.e., it is possible to start with a small number of definitions and derive everything else from them, and indeed this is the path we shall essentially take. At the same time, the theory deals with terms such as information content, information source, rate of transmission of information, i.e., terms which carry a certain amount of intuitive meaning. It will therefore be of interest, at each step of the mathematical development, to compare our results with the dictates of intuition. We shall see that the basic concepts of the theory are readily interpreted in terms of intuitive notions. Of course we cannot expect the same of the more advanced results of the theory. These are basically limit theorems of an involved nature, which may serve as a guide to the intuition of anyone wishing to delve further into the field."

Chapter I, *Introductory concepts*, shows that the real function $H(p_1, p_2, \dots, p_n)$ giving the amount of information in a probability space or, if you like, the information content of a source, is uniquely determined up to a multiplicative constant by the three requirements: (1) $H(p, 1-p)$ is a continuous function of p for $0 \leq p \leq 1$. (2) $H(p_1, p_2, \dots, p_n)$ is a symmetric function of all its variables. (3) If $p_n = q_1 + q_2 > 0$, then $H(p_1, p_2, \dots, p_{n-1}, q_1, q_2) = H(p_1, \dots, p_n) + p_n H(q_1/p_n, q_2/p_n)$.

Chapter II, *Basic properties of $H(x)$* , contains a number of useful inequalities and the noiseless coding theorem which relates H of the source to the average length required to represent elements of the source by sequences of letters from an alphabet of D different letters, where it is required that no sequence of letters representing an element of the source be obtainable from a shorter sequence of letters representing another element of the source by adding letters to the shorter sequence.

Chapter III, *The discrete channel without memory*, sets up the notion of channel as well as transmission rate and capacity of channels.

Chapter IV, *The coding theorem for discrete channels without memory*, using a lemma proved by the author in his thesis, shows that for a discrete channel without memory and a source whose information rate is less than the channel capacity, if $1 > e > 0$ then there is a code so that the uniform error bound on the probability of incorrect reception is less than e . It is understood that this more or less error free reception is at the expense of a delay in reception. This theorem, whose converse was first proved by Wolfowitz, together with the version in Chapter VI, is the heart of the theory. While one can crudely combat noise by simply repeating a message often enough and relying on the law of large numbers, such a procedure cuts down the information rate much below the channel capacity. The coding theorem asserts that there is a cleverer way of combatting noise, a way which ensures asymptotic error free reception while maintaining any information rate less than the channel capacity, at the possible expense of complicated computing for encoding and decoding as well as a time lag for the arrival of blocks of symbols. The coding theorem indicates that one can fully capitalize on a noisy channel if one has sufficient computing equipment.

Chapter V, *The semi-continuous channel without memory*, gives a somewhat syncopated presentation of measure theory and integration which will undoubtedly generate business from engineers for less rapid approaches by way of mathematics courses.

Chapter VI, *The discrete channel with memory* develops the coding theory for this case.

Chapter VII, *The binary symmetric channel*, deals with the case of two message symbols, two received signals such that $p(y_1|x_1) = p(y_2|x_2) = q > (1/2)$. This is the simplest nontrivial case and the idea is to find how the probability of incorrect reception goes to zero with the length of the block.

A comparison of Feinstein's book with the English translation of Khinchin's papers is in order. Both of these books represent an enormous advance in clean mathematical presentation of this material over previous books. Khinchin's book is excellent and readable. It is leisurely, while assuming more mathematical sophistication of the reader, including martingale convergence theorems, it is relatively self complete and does not require as much specialized material as does some of the recent periodical literature. The work is more motivated than is typical of presentations from this side of the Atlantic. Khinchin's work, as stated before, is historically important insofar as it provided the first complete critical survey of the subject, sorted out the open problems, and broke the barrier to serious treatment of the subject. It may be the more attractive to a mathematical reader.

Feinstein's book is addressed to an engineering audience. As such it is well motivated and unusually careful and mathematically clean. For example, while Khinchin slipped in his formulation of a discrete channel with memory, making his work wrong at one point (see K. Takano, *On the basic theorems of information theory*, Ann. Inst. Statist. Math. vol. 9 no. 2 (1957) pp. 53-77), Feinstein does not make this slip. Feinstein's work is more recent and the Remarks at the ends of chapters are devoted in part to valuable discussions of open questions, although publication of recent papers including one by Feinstein himself, make these comments outdated.

One minor negative comment. The author's habit, which follows some engineering practice, of using the same letter for each probability distribution and distinguishing among distributions by using different letters for the sample points, e.g. using $p(x)$ and $p(y)$ to denote input probability distribution and output probability distribution, can be confusing.

S. SHERMAN

Funkcje Rzeczywiste. By Roman Sikorski. (Monografie Matematyczne, vol. 35). Warsaw, Państwowe Wydawnictwo Naukowe, 1958. 534 pp.

The introduction to this Monograph contains an account of the ideas and theorems of set theory and topology, necessary for the understanding of the subject treated in the book, which is divided