Schrödinger equation, the neutron transport equation, Maxwell's equations, and the Dirac equation. A notable feature of the book is the treatment of second-order elliptic and parabolic problems in L^2 and L^p spaces. Fattorini does a nice job of explaining the Agmon-Douglis-Nirenberg elliptic machinery (in the second-order case), making it accessible to a wide audience. An important feature of the book is its extensive and useful bibliography occupying more than a hundred pages.

JEROME A. GOLDSTEIN

BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 12, Number 2, April 1985

0273-0979/85 $1.00 + $.25 per page

Bayes theory, by J. A. Hartigan, Springer Series in Statistics, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1983, xii + 145 pp., $16.80. ISBN 0-387-90883-8

A basic problem of statistics is to infer something about a parameter or state of nature θ after observing a random variable x whose distribution p_θ depends on θ. A neat, but controversial, solution to this problem of inference is provided by the Bayesian approach. Assume that θ is a random variable with distribution π *prior* to observing x. The inference is made by calculating q_x, the conditional or posterior distribution of θ, given x. If p_θ and π have probability density functions $f(x|\theta)$ and $g(\theta)$, respectively, then q_x has density $h(\theta|x)$ given by Bayes's formula

$$h(\theta|x) = \frac{f(x|\theta)g(\theta)}{\int f(x|\varphi)g(\varphi)\,d\varphi} \tag{1}$$

or, briefly,

$$h(\theta|x) \propto f(x|\theta)g(\theta). \tag{2}$$

(For simplicity, assume the densities are with respect to Lebesgue measure. However, any σ-finite dominating measure will do.) There is no disagreement about Bayes's formula. The controversy is about its application and its interpretation.

The two major interpretations of the probability of an event E, both of which can be traced back to the seventeenth-century origins of the subject, are as the limiting relative frequency of E in a sequence of trials, or as a measure of the degree of belief in the occurrence of E. For the past half century the majority of probabilists and statisticians have accepted the frequency interpretation, even though it is of limited application and seems somewhat circular in its "dependence" on the law of large numbers. The frequency view is disastrous for Bayesian inference because it rarely happens that prior probabilities make sense as frequencies. They do make sense when viewed as degrees of belief, and this explains why Bayesians are often identified with subjective probability (de Finetti (1974), Savage (1954)). However, there have been, and are, prominent Bayesians who advocate the use of logical or canonical prior distributions (Jeffreys (1939), Jaynes (1968)). The present author, Hartigan, points out that an overview "may be taken from the point of view of a subjective probabilist, since subjective probability includes all other theories".

In addition to their unorthodox interpretation of probability, many Bayesians are unwilling to accept the conventional Kolmogorov (1933) axioms. Some, like de Finetti, reject the requirement of countable additivity, and others, including Hartigan, retain the assumption of countable additivity, but allow *improper* probability distributions that have an infinite total mass. There are foundational reasons for these technical heresies, but a simple example will illustrate their utility.

Think of θ as a physical constant, like the speed of light or the weight of some object, and suppose x is a measurement of θ subject to a normal error distribution with mean 0 and variance 1. Thus x is $N(\theta, 1)$. If there is little prior knowledge of θ, it is tempting to say that, posterior to x, θ is $N(x, 1)$. The frequentists make essentially this claim using different language. R. A. Fisher (1973) would say that the "fiducial distribution" of θ is $N(x, 1)$, while many others would use the same distribution to construct confidence intervals. Now it is easy to see that there is no proper, countably additive prior π having the posterior above. However, if π is a finitely additive, translation-invariant probability, it does have the desired posterior (Heath and Sudderth (1978)). The same posterior can also be obtained from (2) if π is Lebesgue measure or, equivalently, if $g(\theta) = 1$ for all θ. In the past, improper priors have often been used formally in (2) and justified by the results. A major virtue of Hartigan's book is that a theory of improper probability is developed and Bayes's formula is proved.

Many statisticians now follow Wald (1950) and regard the classical statistical problems, such as estimation and testing, as *statistical decision problems*. In addition to θ and x, there are a set of possible actions and a loss function l that assesses a penalty $l(\theta, a)$ when θ is the state of nature and action a is taken. The decision maker gets to see x before choosing an action $d(x)$, and the *decision function* d is evaluated through the expected loss

$$R_d(\theta) = \int l(\theta, d(x))\, p_\theta(dx).$$

It hardly ever happens that one decision function is better than all the others, and the usual approach is to study the class of admissible decision functions: i.e., those d such that there is no d' for which $R_{d'}(\theta) \leqslant R_d(\theta)$ for all θ, with strict inequality holding for some θ. Thus the non-Bayesian decision maker must employ some other principle, such as maximum likelihood for estimation, and still faces what may be the difficult problem of determining whether the d selected is admissible.

Again the Bayesian, equipped with a prior π, has a relatively straightforward approach. The loss corresponding to a decision function d can now be evaluated as

$$B(d) = \int R_d(\theta)\, \pi(d\theta),$$

and one seeks to minimize the real number $B(d)$. Furthermore, this can often be done in an algorithmic fashion by choosing $d(x)$ to minimize the posterior expected loss

$$\int l(\theta, d(x))\, q_x(d\theta). \tag{3}$$

It is somewhat of a surprise and a strong theoretical argument for the Bayesians, that, for many problems, all of the admissible decision functions are Bayes for some prior π or the limit in an appropriate sense of Bayes decision functions. Thus the choice of a decision function is often tantamount to the choice of a prior.

Return to the example of the physical constant θ and suppose the task is to estimate θ by a function $d(x)$ subject to a loss equal to the square of the error. The usual estimator $d(x) = x$ is admissible for this one-dimensional problem, but is not admissible for the analogous problem in dimensions $n \geqslant 3$ (Stein (1956), discussed in Chapter 9 of Hartigan). This estimator is not Bayes for any proper, countably additive prior, nor is it Bayes in Hartigan's improper theory (cf. p. 65). However, the estimator is Bayes (even in higher dimensions) with respect to a finitely additive, invariant prior. This illustrates Hartigan's point on p. 58 that "optimality by a finitely additive probability is rather too easy" and suggests that optimality in his theory is a bit too hard. The estimator is "formal Bayes" with respect to Lebesgue measure in that it minimizes the posterior loss (3). However, it is ruled out by Hartigan's theory because $B(d)$ is infinite.

In addition to the topics already mentioned, Hartigan treats a number of others. He surveys some of the techniques for getting a prior and proposes a new one based on perceived similarities between observables. Classical and some recent work on Bayesian asymptotic theory are presented. Recent work of de Robertis and Hartigan (1981) on the extent to which Bayesian procedures depend on the prior π is explained.

A topic conspicuous by its absence is Bayesian prediction theory. It can be argued that prediction problems are more fundamental than the more commonly discussed estimation and testing problems (Aitchison and Dunsmore (1975), Geisser (1980)).

The book has the appearance of a textbook, and there are problem sets for each chapter. However, the pace is brisk and sometimes hard to follow. (A variation of the Daniell integral is developed in about three pages. Properties of the integral that have not been discussed are used later.) Terminology is not standard. (A "probability space" is not what you think.) The reader who overcomes these difficulties will be rewarded by a stimulating and original work on current Bayesian theory written by a major contributor to the theory. Another selling point for the book is its uncommonly reasonable price.

REFERENCES

1. J. Aitchison and I. R. Dunsmore, *Statistical prediction analysis*, Cambridge Univ. Press, 1975.
2. B. de Finetti, *Theory of probability*, Wiley, New York, 1974.
3. R. A. Fisher, *Statistical methods and scientific inference*, 3rd ed., Macmillan, New York, 1973.

4. S. Geisser, *A predictivistic primer*, Bayesian Analysis in Econometrics and Statistics, North-Holland, Amsterdam, 1980.

5. D. Heath and W. Sudderth, *On finitely additive priors, coherence, and extended admissibility*, Ann. Statist. **6** (1978), 333–345.

6. E. T. Jaynes, *Prior probabilities*, IEEE Trans. System Sci. Cybernet. **SSC-4**, (1968), 227–241.

7. H. Jeffreys, *Theory of probability*, Oxford Univ. Press, London, 1939.

8. A. N. Kolmogorov, *Foundations of the theory of probability*, Chelsea, New York, 1950. (German ed., 1933).

9. L. de Robertis and J. A. Hartigan, *Bayesian inference using intervals of measures*, Ann. Statist. **9** (1981), 235–244.

10. L. J. Savage, *The foundations of statistics*, Wiley, New York, 1954.

11. C. Stein, *Inadmissibility of the usual estimator for the mean of a multivariate normal population*, Proc. Third Berkeley Sympos., Vol. 1, 1956, pp. 197–206.

12. A. Wald, *Statistical decision functions*, Wiley, New York, 1950.

WILLIAM D. SUDDERTH

BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 12, Number 2, April 1985

0273-0979/85 $1.00 + $.25 per page

Combinatorial enumeration, by Ian P. Goulden and David M. Jackson, John Wiley & Sons, Inc., Somerset, New Jersey, 1983, xxiv + 569 pp., $47.50. ISBN 0-4718-6654-7

The most important idea in enumerative combinatorics is that of a generating function. According to the classical viewpoint, if the function $f(x)$ has a power series expansion $\sum_{n=0}^{\infty} a_n x^n$, then $f(x)$ is called the generating function for the sequence a_n. Sometimes the coefficients b_n, defined by

$$f(x) = \sum_{n=0}^{\infty} b_n \frac{x^n}{n!},$$

are more useful; here $f(x)$ is called the exponential (or factorial) generating function for the sequence b_n. Generating functions are often easier to work with than explicit formulas for their coefficients, and they are useful in deriving recurrences, congruences, and asymptotics.

Although generating functions have been used in enumeration since Euler, only in the past twenty years have theoretical explanations been developed for their use. Some of these, such as those of Foata and Schützenberger [**6**, **7**] and Bender and Goldman [**2**] use decompositions of objects to explain generating function relations. Other approaches, such as those of Rota [**13**], Doubilet, Rota, and Stanley [**4**], and Stanley [**14**], use partially ordered sets. Goulden and Jackson's book is a comprehensive account of the decomposition-based approach to enumeration.

In the classical approach to generating functions, one has a set A of configurations (for example, finite sequences of 0's and 1's) satisfying certain conditions. Each configuration has a nonnegative integer "length". The problem is to find the number a_i of configurations of length i. One first finds a recurrence for the a_i by combinatorial reasoning; the recurrence then leads to