

consisting of modification of some of those already mentioned, are briefly stated with citation of references. Chapter 1 ends with an interesting discussion of criteria for the comparison of different algorithms.

The statement of the translation editor that this is the only book in English which deals extensively with the approximate solution of integral equations (with Cauchy kernels) may still be true. At least this reviewer could find no other. However, for the case of convolution equations there is the excellent and extensive work by Gohberg and Fel'dman [7], which forms an effective complement to the present book, since the approach and the methods in the two monographs are quite distinct for the most part.

One feature of the book which aids the reader is to set off passages which can be omitted on a first reading or are suitable "for a reader with an advanced mathematical training", by a vertical line within the margin of the text. The book is "closely written", in the sense of the theorem-proof style, sometimes proceeding without clear motivation from the readers point of view. However this style has the advantage of the presentation of a large amount of material in a short compass, and of clearly setting out what is known and not known. For a book on numerical analysis there are remarkably few numbers included. The only example noticed where any details of calculation were cited was the computer solution of the system which arose in dispersion theory (p.263)

#### REFERENCES

1. N. I. Mushelišvili, *Singular integral equations*, 2nd ed., Noordhoff, Groningen, 1953. MR 15, 434.
2. I. N. Vekua, *On the theory of singular integral equations*, Soobšč. Akad. Nauk Gruzin. SSR 3(1942), 869–876. MR 5, 268.
3. B. V. Hvedelidze, *Regularization problem in the theory of integral equations with Cauchy kernel*, Dokl. Akad. Nauk SSSR 140 (1961), 66–68 = Soviet Math. Dokl. 2 (1961), 1169. MR 26 #563.
4. A. N. Tihonov, *Solution of incorrectly formulated problems and the regularization method*, Dokl. Akad. Nauk SSSR 151 (1963), 501–504 = Soviet Math. Dokl. 4 (1963), 1035–1038. MR 28 #5576.
5. ———, *Regularization of incorrectly posed problems*, Dokl. Akad. Nauk SSSR 153 (1963), 49–52 = Soviet Math. Dokl. 4(1963), 1624–1627. MR 28 #5577.
6. J. N. Franklin, *On Tikhonov's method for ill-posed problems*, Math. Comp. 28 (1974), 889–907. MR 51 #12007.
7. I. C. Gohberg and I. A. Fel'dman, *Convolution equations and projection methods for their solution*, "Nauka", Moscow, 1971; English transl., Transl. Math. Monographs, vol. 41, Amer. Math. Soc., Providence, R. I., 1974. MR 50 #8148; #8149.

D. H. HYERS

BULLETIN OF THE  
AMERICAN MATHEMATICAL SOCIETY  
Volume 83, Number 5, September 1977

*Nonparametrics: Statistical methods based on ranks*, by E. L. Lehmann, Holden-Day, Inc., San Francisco, and McGraw-Hill International Book Company, Düsseldorf, Johannesburg, London, Mexico, New York, Panama, São Paulo, Singapore, Sydney, Toronto, 1975, xvi + 457 pp.

The methods taught in the standard basic statistics course assume that the observations come from a normal (Gaussian) distribution. Students are taught that the best estimate of the average or typical value of a population is the

average (mean) of a sample. In some applications, however, the median describes the typical member of a population more accurately than the mean. For example, in summarizing the income level of a town, a few rich families have a far greater effect on the value of the mean than they do on the median.

Practical statisticians have long noticed that a few extremely deviant observations sometimes occur even in samples from populations which look "nearly normal" and that these observations have an unduly large impact on classical statistical analyses. Professor Tukey's 1961 article [11] stimulated a great deal of research in the development of *robust* statistical estimators and tests which are relatively unaffected by deviations from the usual assumptions underlying statistical methods. Nonparametric tests are the most robust of all procedures, as they are valid for samples from any population. Rank tests, the class of nonparametric methods discussed by Professor Lehmann, are also of interest because every rank test generates a unique estimator which is often robust.

Rather than attempt to summarize the entire text, let us concentrate on the "two-sample problem" which arises in testing whether two populations differ or are essentially the same. For example, suppose one is interested in deciding whether two groups (e.g., white and Mexican Americans) have the same distribution of heights. Thus, we are testing that the height distribution of Mexicans  $G(x)$  equals that of the whites  $F(x)$ . As an *alternative* one might consider  $G(x) = F(x - \Delta)$ , i.e., the height distribution of whites is shifted to the right, i.e., for any fixed height,  $h$ , the proportion of Mexicans ( $x$ 's) who are less than  $h$  is greater than the proportion of whites ( $y$ 's) who are less than  $h$ . In addition, we are interested in estimating the size of the shift  $\Delta$  required to make the distributions identical.

To resolve the issue, we take random samples  $x_1, \dots, x_m$  from the first group and  $y_1, \dots, y_n$  from the second. Consider the combined sample of  $N = m + n$  observations. Under the null hypothesis that the two distributions are the same, i.e.,  $F(x) = G(x)$ , it can be shown that each of the original observations has probability  $1/N$  of being the  $k$ th largest in the pooled sample. Thus, the ranks in the combined sample that the  $n$   $y$ 's will have, can be considered as a random sample of  $n$  integers chosen from  $1, \dots, N$ , irrespective of the form of the distribution function  $F(x)$ . Any test which is solely a function of the ranks that one group of observations has in the combined sample is called a rank test. If we let  $R_i$  be the rank  $y_i$  has in the ordered combined sample of  $N$ , then the Wilcoxon test is defined as  $W = \sum R_i$ , and its distribution is that of the sum of  $n$  randomly selected integers from  $1, 2, \dots, N$ . Since the average of the first  $N$  integers is  $(N + 1)/2$ , under the null hypothesis the expected value of  $W$  is  $n(N + 1)/2$ . If the actual value of  $W$  is much larger than that, statisticians reject the hypothesis that the  $x$ 's and  $y$ 's have the same distribution in favor of the alternative and conclude that the distribution of the  $y$ 's is shifted to the right, i.e.,  $\Delta > 0$ . From the distribution of  $W$  we compute the probability of obtaining a value at least as distant from its expected value as the one actually obtained. Usually when this significance probability level is less than 5% we reject the null hypothesis.

At this juncture probability theory is required to calculate the exact distribution of the statistic  $W$  and show that in large samples it is approximately normally distributed. Under the null hypothesis the ranks of the  $y$ 's are a random sample from  $1, \dots, N$ . Thus, they are negatively correlated and the ordinary Central Limit Theorem does not apply. Lehmann's appendix provides an elegant treatment of asymptotic normality based on Hájek's approach to samples from finite populations. Fortunately, the asymptotic approximation to the distribution of rank statistics is quite accurate even in relatively small samples ( $m, n \geq 10$ ).

So far we have limited our attention to insuring that the probability of rejecting the hypothesis that both populations are the same when it is indeed true, is small (5% or less). The strength of nonparametric statistics is that this calculation is the same regardless of the form of the density function of the population. On the other hand, we also desire to reject the null hypothesis when the two populations truly differ. If  $c$  is the critical point of a test of size  $\alpha$  (often .05), i.e., the probability, when the null hypothesis is true, of obtaining a value  $> c$ , is  $\leq \alpha$ , written  $P_0[W > c] \leq \alpha$ , then the power of the test is the probability of  $[W > c]$  calculated under the alternative assumption (e.g.,  $\Delta = 1$ ). While the size ( $\alpha$ ) of the Wilcoxon and all rank tests *does not* depend on the form of the underlying density function, their power does. The remarkable fact about the Wilcoxon test is that it is about 95% as powerful as the usual  $t$ -test for normal data. Hence, one pays a rather small price in terms of loss of power for guaranteeing that the Type I error (size) is not affected by the form of the density.

By now the reader is probably asking if can we develop a test that is as powerful as the  $t$ -test on normal data and still has the property that its size ( $\alpha$ ) is unaffected by the shape of the density. For practical purposes the answer is yes, although the result is strictly true only for large samples. The idea is to use statistics of the form  $\sum a(R_i)$ , where  $R_i$  is the rank of  $y_i$  and  $a(R_i)$  is specified by  $a(R_i)/(N + 1)$ , where  $a(u)$  is a function on  $(0,1)$ . The fundamental result, due to Chernoff and Savage [2] is the following:

Let  $x_1, \dots, x_m; y_1, \dots, y_n$  be two independent samples from the distributions  $F(x)$  and  $G(x) = F(x - \Delta)$ , respectively, and assume that  $f = F'$  has finite Fisher information, i.e.,  $I = \int (f'/f)^2 f dx < \infty$ . The asymptotically most powerful rank test of  $H_0: \Delta = 0$  against  $\Delta \neq 0$  is based on the function

$$(1) \quad a(u) = I^{-1/2} f' [F^{-1}(u)] / f [F^{-1}(u)], \quad 0 < u < 1,$$

and is asymptotically as powerful as the best parametric (maximum likelihood) procedure. When the population sampled is the logistic,

$$f(x) = e^{-x} / (1 + e^{-x})^2,$$

then the function  $a(u) = 2\sqrt{3} (u - \frac{1}{2})$  is a straight line which corresponds to the Wilcoxon test normalized to have mean 0 and variance 1. Since the functional form of the logistic density is similar to the normal, it is not surprising that the rank test which is best for the logistic is quite powerful on normal data.

Mathematicians looking at formula (1) will notice that the functions  $a(u)$

are in  $L^2(0,1)$  and realize that there should be a connection between the inner product  $\int a_1(u)a_2(u) du$  and the closeness of the two densities  $f_1, f_2$  generating the test scores  $a_1$  and  $a_2$ .

The connection arises from the consideration of Pitman efficiency, a criterion enabling us to compare tests for large samples. A special criterion is needed here because almost any reasonable test will have power approaching 1 as the sample sizes  $(m,n)$  increase. The idea underlying Pitman efficiency is to consider testing  $H_0: \Delta = 0$  against a sequence of alternatives  $H_1: \Delta = \delta/\sqrt{N}$ . In this case the power will not approach 1 as  $N$  increases, because the difference between the null and alternative hypotheses will decrease as  $N$  increases. It can be shown that the limiting power of a test  $T_1$  (based on  $a_1$ ) is given by

$$(2) \quad \Phi(c_1\delta - u_\alpha)$$

where  $\Phi(-u_\alpha) = \alpha$  the level of the test. Notice that the larger  $c_1$  is, the more powerful the test is. The constant  $c_1$  is called the efficacy of the test and the Pitman efficiency of a test  $T_1$  relative to another test  $T_2$  is

$$(3) \quad e(T_1, T_2) = (c_1/c_2)^2.$$

The quantity  $e(T_1, T_2)$  can also be interpreted as the limiting ratio of the sample sizes required by the tests  $T_1$  and  $T_2$  to achieve the same limiting power  $\pi$  against the same sequence of alternatives. For example, if the Pitman efficiency of test  $T_1$  relative to  $T_2$  ( $e(T_1, T_2)$ ) =  $\frac{1}{2}$ , then the test  $T_2$  results require approximately half as many observations as the test  $T_1$  to achieve the same large sample power for critical regions of the same size  $\alpha$ . Besides its statistical interpretations, the Pitman efficiency can be computed readily. It can be shown that the efficiency of test  $T_1$  (based on  $a_1(u)$ ) relative to the best test,  $T_2$ , on data from the density  $f_2$  is given by

$$(4) \quad \langle a_1, a_2 \rangle = \int a_1(u)a_2(u) du,$$

where  $a_2$  is obtained from (1). By the symmetry of the inner product  $T_2$  has this same efficiency relative to the best test,  $T_1$ , for data from the density  $f_1$ .

The fact that the functions  $a(u)$  generating the most powerful rank tests for a wide family of densities are in  $L^2(0,1)$  yields insight into other problems as well. If one truly knew the form of the density, it would not make sense to use a rank test instead of the usual maximum likelihood test. Suppose one knew something about the density, e.g., it was either  $f_1$  or  $f_2$  (normal or double exponential): Is there a reasonably powerful rank test for this problem? Considering the functions  $a_1, a_2$  as vectors in  $L^2(0,1)$ , it is clear that a test corresponding to the angle bisector will maximize the minimum efficiency when data comes from either density. This problem is discussed in [1] and [6]. Again statisticians lose relatively little power against either density when they insure that the rank test is reasonably powerful for samples from a family of densities.

Another problem arises when data is censored or missing. This often occurs in reliability and clinical trials when we observe lifetimes (survivors) over time and must end the experiment early. Essentially the data allows us only to

order the  $x$ 's whose value was less than  $t$  and to know that  $m - m_1$   $x$ 's were greater than  $t$ . Similarly, we have the ordered  $y_1, \dots, y_k$  ( $k = n - n_1$ ) and the knowledge that  $n - n_1$   $y$ 's were greater than  $t$ . Thus all the  $n + m - m_1 - n_1$  observations can be considered as missing or tied. If the test based on  $a(u)$  is appropriate for complete samples we desire the closest  $L^2$  approximation to it which is constant on  $(1 - r, 1)$  where  $r = (m_1 + n_1)/(m + n)$ , the proportion of observations which are larger than  $t$ . That approximation [5], [10] is clearly the function equaling  $a(u)$  on  $(0, 1 - r)$  and the average  $(1 - r)^{-1} \int_{1-r}^1 a(u) du$  on  $(1 - r, 1)$ . Thus, we score the ranks of the true observations with the same scores we would use in the complete data situation and score the censored observations by their "average rank." Although this formula is derived from asymptotic theory, the result agrees with common sense and yields tests whose statistical properties are close to the special small sample optimal procedures that have been derived in the literature.

Mathematicians realize that Pitman efficiency is only one of several approximations that can be used to compare the power of tests in large samples. Cochran considered the rate at which the level (rejection probability under  $H_0$ ) decreases to zero when the alternative is a fixed distance from the null hypothesis and the power is also fixed ( $< 1$ ). A different formulation which often yields equivalent results is due to Bahadur. Recently Hájek [7] showed that in terms of Bahadur efficiency, rank tests are as powerful as any function of the observations against a fixed alternative. Formally, consider the null hypothesis  $H_0: F = G$ . The statistic  $\psi(x_1, \dots, x_m, y_1, \dots, y_n)$  which rejects  $H_0$  when its value is  $\geq \psi_0$  has level

$$L(\psi_0) = P_0[\psi(x_1, \dots, x_m, y_1, \dots, y_n) \geq \psi_0].$$

The Bahadur efficiency is essentially the rate at which the *observed level* tends to zero at a fixed alternative. For example, consider a fixed alternative  $F = F^*$ ,  $A = A^*$  (e.g.,  $F^* = N(0, 1)$ ,  $G = N(\Delta, 1)$ ,  $\Delta > 0$ ). Bahadur showed that for almost all sample sequences  $N^{-1} \log L(\psi(x_1, \dots, x_n; y_1, \dots, y_m))$  tends to a limit,  $-K(\psi)$ . The larger  $K(\psi)$  the faster the level (calculated under the alternative) approaches zero. Since smaller levels lead to rejection of the null hypothesis, this shows that tests with large values of  $K$  are better. Hájek showed that the Max  $K(\psi)$  over all tests  $\psi$  of  $H_0$  against any specific alternative,  $H_A(F^*, G^*)$ , is achieved by a rank test of the form  $\sum a(R_i)$ . Thus, no power is lost asymptotically by restricting attention to rank tests.

While the book briefly discusses some of the alternative notions of efficiency and refers to the recent work of Bickel, the reviewer would have appreciated both a more comprehensive analysis of when and why they give the same or different results and Professor Lehmann's expert opinion as to which notion of efficiency is most appropriate to each problem treated. In particular, Rubin and Sethuraman [9] developed the concept of Bayes risk efficiency which often agrees with Pitman's notion. In view of the "heated discussion" between Bayesians and non-Bayesians [3] in the literature on the foundations of statistics, even a short section by Professor Lehmann could aid in converting the "heat" to light.

So far our discussion has focused on a somewhat artificial problem, namely

testing the hypothesis that the two populations underlying our samples are identical. In the real world the problem of interest is how to estimate the shift (difference),  $\Delta$ . Statisticians have shown that for every rank test there is a corresponding estimator,  $\Delta$ , and its efficiency relative to the best parametric estimator is the same as the Pitman efficiency of the test to the a.m.p.r.t. Since the fundamental paper in this area is a joint work of the author [8], it is not surprising that the sections on estimating the treatment effect, illustrated by estimates derived from the Wilcoxon test, are both clear and comprehensive.

Due to space limitations we have illustrated the fundamental notions of nonparametric inference on just one problem. The text covers a much wider class of situations in which rank tests have been used, including paired comparisons (used in matching studies), the comparison of several treatments and randomized complete blocks. This last topic is quite important in biological research as it provides a method for dividing the population into homogeneous subgroups and testing to see if there is a treatment effect across all subgroups. These methods take advantage of the fact that "background" variation within each subgroup is normally much smaller than the variation in the entire population.

The mathematician reading the text will be exposed both to a wide variety of applications of nonparametric methods and to the probabilistic tools which underly the theory. Indeed, the mathematical appendix (approximately 80 pages long) is superb. Although the book discusses a variety of real problems, from a practical statistician's viewpoint it spends too much time on the problem of ties (observations with the same value) and not enough time on "heuristic" or common sense methods. For example, in the discussion of testing whether a sequence of numbers is a sample from the same distribution *or* has an increasing trend, i.e., the mean of the  $x_i$ 's increases with  $i$ , the standard test due to Mann is discussed. If one draws a picture of samples coming from both the null and alternative hypotheses, one notices that their major difference is highlighted at the ends. This suggests that a reasonably powerful test might be developed by considering the lowest third of the observations as  $x$ 's and the upper third as  $y$ 's and testing whether the  $y$ 's are larger than the  $x$ 's. In the reviewer's opinion it is unfortunate that students generally are not taught the reasoning behind this type of "quick" and relatively efficient statistical procedure [4].

In summary, Professor Lehmann's text will undoubtedly become the standard to which other mathematical introductions to nonparametric inference will be compared. It has an excellent selection of problems, ranging from routine calculations to challenging problems for the best students, and an extensive bibliography. Although students may not find the book particularly easy to read, Professor Lehmann has provided a very well organized guide to the field of rank tests. Mathematically inclined students should be encouraged to and able to pursue the study of Kolmogoroff-Smirnov and other "goodness of fit" tests which stimulated a significant amount of research in the properties of tied-down Wiener processes. The reviewer has no doubt that a student with initiative will learn a great deal about both the mathematical foundations and use of rank tests from this book.

## REFERENCES

1. A. Birnbaum and E. Laska, *Efficiency robust two-sample rank tests*, J. Amer. Statist. Assoc. **62** (1967), 1241–1251. MR **36** #6083.
2. H. Chernoff and I. R. Savage, *Asymptotic normality and efficiency of certain nonparametric test statistics*, Ann. Math. Statist. **29** (1958), 972–994. MR **20** #6755.
3. J. Cornfield, *The Bayesian outlook and its application*, Biometrics **25** (1969), 617–642. MR **41** #2833.
4. D. R. Cox and A. Stuart, *Some quick sign tests for trend in location and dispersion*, Biometrika **42** (1955), 80–95. MR **16**, 842.
5. J. L. Gastwirth, *Asymptotically most powerful rank tests for the two-sample problem with censored data*, Ann. Math. Statist. **36** (1965), 1243–1247. MR **31** #852.
6. ———, *On robust procedures*, J. Amer. Statist. Assoc. **61** (1966), 929–948. MR **34** #5226.
7. J. Hájek, *Asymptotic sufficiency of the vector of ranks in the Bahadur sense*, Ann. Statist. **2** (1974), 75–83. MR **50** #8825.
8. J. L. Hodges, Jr. and E. L. Lehmann, *Estimates of location based on rank tests*, Ann. Math. Statist. **34** (1963), 598–611. MR **27** #2050.
9. H. Rubin and J. Sethuraman, *Bayes risk efficiency*, Sankhyā Ser. A **27** (1965), 347–356. MR **34** #6928.
10. P. K. Sen, *Asymptotically most powerful rank order tests for grouped data*, Ann. Math. Statist. **38** (1967), 1229–1239. MR **35** #3820.
11. J. W. Tukey, *The future of data analysis*, Ann. Math. Statist. **33** (1962), 1–67. MR **24** #A3761.

J. L. GASTWIRTH

BULLETIN OF THE  
AMERICAN MATHEMATICAL SOCIETY  
Volume 83, Number 5, September 1977

*Representations of real numbers by infinite series*, János Galambos, Lecture Notes in Math., vol. 502, Springer-Verlag, Berlin, Hiedelberg, New York, 1976, vi + 146 pp., \$7.40.

During the last thirty years a large amount of research has been devoted to the study of various algorithms for the representation of real numbers by means of sequences of integers. In addition to the two “classical” algorithms, i.e. digit expansions and continued fractions, and to several types of additional, slightly less well-known ones like Cantor, Lüroth, Engel and Oppenheim series, more general classes of such algorithms have been defined and investigated, particularly by F. Schweiger [6] and also by the author himself [3].

The book under review introduces the reader to some of the most important features of these developments. The exposition is based on what the author calls an  $(\alpha, \gamma)$ -expansion  $y(x)$  of a real number  $x$ . For each  $j$ , two strictly decreasing sequences  $\alpha_j(n)$  and  $\gamma_j(n)$  of positive real numbers are given, satisfying the condition

$$\alpha_j(n-1) - \alpha_j(n) \leq \gamma_j(n) \quad (n = 2, 3, \dots).$$

In order to define the algorithm for a given number  $x \in (0, 1]$ , an auxiliary sequence  $d_j(x)$  of integers is defined in such a way that the infinite series

$$y(x) = \alpha_1(d_1) + \gamma_1(d_1)\alpha_2(d_2) + \gamma_1(d_1)\gamma_1(d_2)\alpha_3(d_3) + \dots$$

is always convergent and has, under fairly general assumptions, the limit  $x$ . A