NONPARAMETRICS: RETROSPECTIVES AND PERSPECTIVES

by

Pranab Kumar Sen

Department of Biostatistics
University of North Carolina at Chapel Hill, NC

Institute of Mimeo Series No. 1885
November 1990

N O N P A R A M E T R I C S: RETROSPECTIVES AND PERSPECTIVES

PRANAB KUMAR SEN

University of North Carolina at Chapel Hill

1. A PREAMBLE

The classical *distribution-free methods* are the precursors of the modern day *nonparametrics.* Although the use of such distribution-free methods in some areas of science and sociology sparked with the classical *Bernoulli (binomial) distribution* quite long ago [viz., Arbuthnott (1710)], a real initiation of statistical research in this area did not take place until the mid-thirities of this century. Back in those days, the *Neyman-Pearsonian theory of optimal tests of statistical hypotheses* was in an active evolutionary phase, and the *theory of statistical estimation* was neither on a very sound footing. It is in this perspective, some distribution-free methods were introduced with a sole objective: How to construct a suitable test for a statistical hypothesis such that its level of significance (size) does not depend on the form of the distribution of the underlying random variables? This explains why these tests were referred to as distribuiton-free tests. Of course, such a property of distribution-freeness holds only for a certain class of hypotheses, and hence, there were only a handful of such distribution-free tests. However, characterizations of such hypotheses (of *invariance under certain group of transformations which maps the sample space onto itself*) for which simple (exact) distribution-free (EDF) tests exist led to a big step in the late Thirties: *Randomization* or *permutation tests* were introduced by Pitman (1937), and others, and the concept of *$S(\alpha)$-structure* of EDF tests, introduced by Scheffe (1943), went through a series of generalizations and culminated in the general concept of *tests with the Neyman structure.* Randomization tests gained additional footings in the Forties. The most important aspect of the post-war developments is the fruitful incorporation of suitable *asymptotic methods* in the characterizations of *asymptotically distribution-free (ADF) and optimal nonparametric* procedures both in the hypothesis testing and estimation domains. The past forty years have witnessed a phenomenal growth of research in nonparametric methods. Domains which were once though to be outside the scope of nonparametrics were annexed in natural ways: *Multivariate nonparametrics* with adaptations in linear models and in biological assays were mostly developed in the Sixities, *sequential nonparametrics* in

---

AMS (1980) Subject Classifications: Primary 62G; Secondary 62H, J, L. M.

Keywords and phrases: Adaptive procedures; ADF tests; asymptotic relative efficiency; Bayesian nonparametrics; deficiency; empirical distributions; EDF tests; goodness of fit; linear models; multivariate nonparametrics; nonparametric estimation; permutational limit theorems; rank order statistics; resampling methods; robustness; semi-parametric models; sequential nonparametrics; statistical functionals.

the Seventies, and the *smoothing techniques* in the last two decades. *Nonparametric estimators* entered into the arena in the Sixties while the incorporation of *robust estimation* in nonparametrics did not entail any longer waiting time. The past ten years bear the witness of significant developments in *resampling methods and adaptive procedures* with special emphasis on nonparametrics. We will have a chance to discuss each of these specific areas in latter sections. All these developments would unequivocally call for a media for the cultivation of further research in this fruitful area of nonparametrics, and in this maiden issue of the Journal of Nonparametrics Statistics, I can't check the temptation of summarizing this evolution of the nonparametrics in the following theme.

Not long ago was the time when the nonparametrics
    used to reside in the behavioral and socio-metrics,
were referred to as the 'quick and dirty' methods
    and seemed unworthy of any theoretical efforts!
The sign-test was the undisputed lord of the domain
    in the courtyards of Fisher, Hodges, Bennett and Blumen.
Convincing enough was the traditional binomial distribution,
    surely, that elminated a lot of computational frustration!
The rank sums had to pay a large ransom -
    was not the exact distribution very cumbersome?
Thanks to the post-war Wilcoxon-Mann-Whitney treaty,
    the nonparametric prospects looked very much pretty.
Came the signed ranks in no disguise,
    a real savoy in the randomized paradise.
The Friedman chi square rank became a household word,
    albeit the exact tabulation could not march forward.
The randomization universe fell to compartmentations
    of permutations, matching and sign-inversions.
Came the theoreticians in colored (finite) groups,
    from Pitman to Scheffé, all joined the coups:
Just examine the Neyman-structure under invariance
    and the maximal noise elminates the ignormance.
For the sake of completeness came Blackwell and Bell,
    and the nonparametric snowball melted in the shell.
This was the new testament from Stein to Watson,
    and all of us were wondering: What's going on!
The P-P and Q-Q plots were, of course, there
    but their inter-relations were not very clear.

With the quartile tests came Dixon and Massey,
from the empirical cumulatives, what can you say?
The Russians were much ahead of US in this way,
Kolmogorov and Smirnov had already a lot to say.
Though Birnbaum and Pyke stole the Kolmogorov limelight,
with Vincze and Cśaki, did they ever have any real fight?
I don't recall when the not so mighty nonparametrics
marched into the lofty domain of order statistics,
but, from Rényi to the innumerable stochastics,
this occupation seemed to be merely logistics!
In the midst of all these diversities,
nonparametrics were about to lose identities:
The exacticity in the small sample ethics;
how to open the doors for the asymptotics?
Asymptotically, the universe of permutations
vanishes beyond the reach of manipulations.
Thus, came a genuine need for approximations
for the large sample permutation distributions!
Alas, who knew that it would be a 180° rotation
for nonparametricians, in the futre direction!
Came the permutational central limit theorems:
The initial Wald-Wolfowitz method of moments,
and the Noether-Hoeffding-Motoo refinements,
all paved the way for the final Hájek testaments!
The choice of a nonparametric rank test
remained largely a matter of personal taste.
Although, Fisher and Yates had some nice suggestions,
they were mainly intuitive guess or speculations.
Scores of rank tests were based on (generalized) U-statistics,
and the Hoeffding (1948) projection paved their logistics.
The sufficiency and completeness of sample order statistics
provided the ingredients for the optimality of U-statistics.
Is there any other single paper in the area of nonparametrics,
which has a greater impact than Hoeffding's U-statistics?
This didn't, however, resolve the issue of (asymptotic) optimality
of nonparametric tests, even for alternatives in the null locality.

As Hoeffding and Terry fortified the situation,
    a question arose regarding the allied distribution.
Hoeffding-Dwass made the way for the Chernoff-Savage theorems,
    and the journals were flooded with related developments:
Whether to assume a density with finite Fisher information
    or to impose more stringent conditions on the score function!
The ultimate goals was of course the crucification
    of the asymptotic efficiency in the Pitman definition.
Came the Hájek Theorem with the elegant LeCam contiguity,
    and, by far, this surpassed all records of academic utility!
Bahadur introduced the 'exact' and 'approximate' slopes,
    and Hodges and Lehmann led to the 'deficiency' scopes!
No matter whether you look at a fixed or local alternative,
    you would endup probably with the same parochial objective!
Upto the early Sixties, bearing this state of the art,
    who could write a superb nonparametric text fully alert!
The Fraser is of course worthy of quotation,
    though it is a collection without much unification.
The Siegel Text captured the behavioral heart
    and in despair, the theorticians fell apart.
There were traces of nonparametrics in Fisz and Wilks,
    although the treatment in Lehmann (1959) was less at risks.
The Savage Bibliography served only a part of the quest,
    rank tests were lost in the order statistics forest!
The Walsh Handbook was a matter of different taste,
    and some other cookbooks appeared in 'cut and paste'.
There was a genuine need for a standard text,
    and in right time appeared the Theory of Rank Tests.
This Hajék-Sidak book contains, as the title indicates,
    no visible trace of the popular rank based (R-) estimates;
Confidence or tolerance limits are a minimum,
    and the role of order statistics is a premimum.
Nonparametric methods in sequential and multivariate analysis
    were yet to be properly developed to register any emphasis;
Linear models and bio-assays were desparate for nonparametrics,
    however, you would not find any mention of these tricks!

But, what it contains (rightfully) is a treasure
    and browsing through the pages has been a pleasure.
Who knew about log-concavity and strong-unimodality
    of density in such a simple state of tranquality!
The unified treatment of diverse rank tests, LMPR,
    is the most general, I had known so far.
For the various rank statistics under consideration
    when it comes to the study of their null distribution:
What a beauty to fathom the basic recurrence relations;
    after all, these are the basic nonparametric connections!
The 'asymptotic bridge' on the avenue to sample-size infinity
    has been fortified with utmost care and serenity.
Have you examined the full potency of the Faddeev lemma?
    A special central limit theorem eliminates all dilemma!
For rank statistics, what a neat quadratic mean approximation,
    and a permutational central limit theorem in full generalization!
In the textbook days before Parathasarathy-Billingsley,
    weak convergence in metric spaces used to be treated sparsely.
What an elemgant weak convergence treatise on functionals,
    for nonparametericians, this is enlisted in the essentials.
The Kolmogorov-Smirnov statistics in the setup of regressions,
    who could imagine the vastness of its practical applications.
But, perhaps, the most significant development
    is the role of 'contiguity' in fulfillment.
It stores the functional Taylor expansion in the rank-basement
    and the square integrable score functions complete the assignment!
What a simple prescription for your contiguity-verification:
    LeCam's first, second and third lemmas deserve a lot of ovation!
You might argue when the finite Fisher information is in negation;
    well, accomplishment is feasible on more stringent score function.
You may as well look at the Chernoff-Savage Theorem - a different version,
    and appreciate the elegance and beauty of the contiguity led solution.
Examine the basic role of contiguity in the asymptotic efficiency
    and for ranks, in the characterization of asymptotic sufficiency.
What a superb unification of the nonparametric theory
    in the small as well as large samples,

who would not unequivocally speak of its glory
by solving all the problems and set examples!
The Theory of Rank Tests put a challenge to Erich Lehmann,
and he came up with Nonparametrics written for a layman.
Full of applications in various plausible fields of specialization,
this masterpiece written by an authority deserves a lot of ovation!
Although in terms of abstraction and mathematical sophistication,
Lehmann's Testing Statistical Hypotheses floats on a mast of citation,
in terms of adaptations by researchers in diverse fields of application,
Nonparametrics would probably surpass by any measure of appreciation.
Although, Nonparametrics did not stress on multivariate applications,
nor much in the way of unification in the field of sequentializations,
what it contains is a lucid treatment of the elements of nonparametrics,
albeit a somewhat less sophisticated Appendix dealing with asymptotics!
These two texts together have placed nonparametrics on a sound stand,
from which diversities in other areas are easy to view and understand!
We will have a chance to review some of these later,
but they may not be so prominent in this retrospective platter.
When a baby starts walking, you count every single step,
but after a while, you remember only the commencement!
In nonparametrics, the simplicity and exactness of EDF tests,
faded away under the power and robustness of ADF guests.
In Arabian Tales, one thougsand and one nights fell
which stopped the vanishing of girls in the harem!
Alas, how to curtail the flow of countless procedures,
in the nonparametrics, with some normed brochures?
There comes the Perspectives in the Nonparametrics,
and we shall comment something in constructives.

2. RETROSPECTIVES

We prefer to use the terminology 'nonparametrics' instead of 'distribution-free methods', as the former encompasses a much wider domain. From a retrospective point of view, we may put the salient features of the developments in nonparametrics in the following order:

(i) Ad hoc Distribution-free methods: The early birds.

(ii) Characterizations of EDF procedures.

(iii) Randomization procedures.

(iv) Permutation Vs. Randomization.
(v) Permutational Limit Theorems.
(vi) U-statistics and Generalizations.
(vii) Role of Order Statistics in Nonparametrics.
(viii) Rank Order Statistics
(ix) Empirical Distribution based procedures.
(x) Asymptotic Distribution Theory.
(xi) Asymptotic Efficiency Criteria.
(xii) Local and Asymptotic Optimality Criteria.
(xiii) Robustness Considerations.
(xiv) Nonparametric Estimation.
(xv) Nonparametrics in Econometrics.
(xvi) Nonparametrics in Social Sciences.
(xvii) Nonparametrics in Linear Models.
(xviii) Nonparametrics in Multivariate Analysis.
(xix) Nonparametrics in Biological Assays.
(xx) Nonparametrics in Clinical Trials and Medical Studies.
(xxi) Sequential Nonparametrics.
(xxii) Nonparametrics in Systems Analysis and Reliability Theory.
(xxiii) Nonparametrics for Directional Data Analysis.
(xxiv) Categorical Data Nonparametrics.
(xxv) Nonparametrics for Grouped Data Models.
(xxvi) Nonparametrics for Discrete Data Models.
(xxvii) Statistical Functionals and Nonparametrics.
(xxviii) Smoothing Techniques and Nonparametrics.
(xxix) Resampling Methods.
(xxx) Adaptive Procedures.
(xxxi) Bayesian Nonparametrics.
(xxxii) Nonparametrics in Stochastic Approximations.
(xxxiii) Semiparametrics Vs. Nonparametrics.
(xxxiv) Nonparametrics in Statistics Curriculum.
(xxxv) Nonparametric Textbooks and Monographs.

Nonparametrics entered the arena of Statistical Sciences, mostly, as 'quick and dirty' methods. The sign test for location (median) was undisputedly the first major step in this direction. This was based on a reduction of the orginal hypothesis testing (for location) problem to a simpler one involving a binomial parameter, and this clever idea led to adaptations of numerous combinatorial methods for some other tests which were proposed later on. For example, for the multi-sample (homogeneity) problems, the median test, quartile or quantile tests and the run test have all been formulated on similar reductions. Even the classical Kolmogorov-Smirnov test has a lot of evolutionary support from various combinatorial methods. Basically, in all the testing problems, the null hypothesis permits a reduction of the problem to a parallel one involving the simple uniform (0, 1) distribution, and this enables the use of well known properties of the uniform order statistics to formulate the relevant distribution theory (under the null hypothesis). However, these developments were rather spotty and piecemeal in the sense that they were specifically tailored for specific problems, and it seemed unclear as to how to put them into a unifying mold to cover a general class of hypothesis testing problems. Moreover, albeit their EDF nature, not much was known about any optimality properties of such tests. Computation of the exact power (under appropriate alternatives) also seemed to be difficult.

In quest of a general class of hypothesis testing problems for which EDF tests exist (and play a vital role), the next major step was to land on the so called randomization (or permutation) tests. These are the precursors of the nonparametric rank tests, and in the evolution of nonparametrics, they have played a fundamental role. These tests were proposed (mostly, in the late Thirties) for the one-sample location (symmetry) problem, ANOVA (in one-way and two-way layouts) and the bivariate independence problem. These tests are easy to interpret and apply, have a broader scope of applicability (relative to other nonparametric competitors), encompass a wider class of statistics (which may not be EDF) and they rest on less restrictive regularity conditions on the underlying probability laws as well as the sampling schemes. For the determination of critical values of such randomization tests, no statistical tables may be necessary; these are obtained directly from the sample observations (under suitable permutations). The term permutation test has its origin in this phenomenon. A more unified picture emerged in the Forties. Scheffé (1943) established a characterization of randomization tests as *similar size-α* ($0<\alpha<1$) tests having the S(α)-structure, while Lehmann and Stein (1948) enhanced the scope of such randomization tests to a wider class of problems. Lehmann and Scheffé (1950, 1955) incorporated the concept of *boundedly complete sufficient statistics* to formulate a clearer structure of randomization tests and this led to the formulation of *tests with the Neyman structure*. In this setup, various nonparametric hypotheses were characterized as *hypotheses of invariance* under certain (finite) *group of transformations* which maps the sample space onto itself and yields the *maximal invariants* which play the basic role in the construction of randomization tests. This line of attack has been extensively studied by Professor C. B. Bell (and his associates) during the past twenty five years, and a good account of these developments is given in Bell and Sen (1984). In the conventional univariate problems, randomization procedures include the conventional rank procedures as special cases. This characterization may not apply when the underlying distribuitons are not continuous (a.e.) or, even in the continuous case, when the observations are vector-valued. In the Sixties, randomization principles provided the basic characterization of multivariate nonparametric tests. Unlike their univariate counterparts, procedures based on ranks are generally not EDF in the multivariate case. However, they can be rendered conditionally (permutationally) distribution-free by appeal to the conventional randomization principle. This line of attack was mainly initiated by Chatterjee and Sen (1964) and it culminated in the monograph of Puri and Sen (1971). Another nice feature of randomization tests is a natural way of choosing a particular test having "good" power properties against a specific alternative, and the work of Hoeffding (1952) provides a clear picture of their asymptotic optimality properties.

For the classical multi-sample homogeneity problem EDF tests remain invariant under arbitrary (strictly) monotone transformation on the observations [i.e., $X \rightarrow g(X) = Y$, where g is strictly monotone], and this leads to the characterization of *ranks* as the *maximal invariants* (for the univariate models). It is not surprising to see a host of rank tests for such problems. Similarly, in the one-sample symmetry problem, the vector of signs and the ranks of the absolute values constitute the maximal invariants, and this led to the formulation of various EDF tests known as *signed-rank tests.* Similar rank tests have been constructed for the bivariate independence problem and for randomized block designs. Although these rank tests are also randomization tests, they have certain advantages:

(i) They are EDF (not conditionally DF) for the usual univariate models.

(ii) Their critical values have been extensively tabulated (for small to moderate sample sizes), so that in actual application, it is not necessary to compute them by cumbersome permutational schemes.

(iii) They have 'good' invariance properties, not shared (to that extent) by their randomization counterparts.

However, there are some other less favorable aspects too. Firstly, such EDF test statistics are usually nonlinear in the sample observations, and the evaluation of their actual distribution when the hypothesis of invariance is not true is often too cumbrous (if not prohibitively laborious). Secondly, there can be 1001 EDF tests for a specific null hypothesis, but the choice among them may depend heavily on the type of alternatives one has in mind. Characterization of (asymptotic) optimality of such EDF tests has indeed been a remarkable step, although the finite sample case may still need more careful attention. Thirdly, although these tests are EDF, the exact evaluation of their null distribution may generally become prohibitively laborious as the sample size becomes larger. The multitudes of these problems have led to the following avenues of research in nonparametrics:

(i) *Permutational limit theorems.* Sparked by the ingenuity of Wald and Wolfwitz (1944), various workers paved the way for a systematic development of permutational central limit theorems. The initial approach was to evaluate the moments of the actual permutation distribution and to incorporate them in the characterization of asymptotic normality or chi square laws. In this context, the regularity conditions were relaxed chronologically, although the minimal set was not hit. Hájek (1961) came up with an alternative approach, providing the most general result known so far. In the multivariate case, there are some additional complications, and a martingale approach may prove to be more convenient [viz. Sen (1983)].

(ii) ***U-statistics and generalized U-statistics.*** Many of the EDF test statistics (as well as others) may be expressed as functionals of sample distribution functions. Von Mises (1947) and Hoeffding (1948) considered such functionals when the ***kernel*** is of a (fixed) ***degree.*** These may be formulated by introducing ***estimable parameters*** or regular functionals as

$$\theta(F) = \int \cdots \int g(x_1, \ldots, x_m)\, dF_n(x_1) \cdots dF(x_m) \tag{1}$$

where the kernel $g(\cdot)$ is of degree m (≥ 1) and $F(\cdot)$ is the distribution function (d.f.) of a random variable X. If $F_n(\cdot)$ stands for the empirical (sample) d.f. of $X_1, \ldots X_n$, then von Mises functional is defined by

$$\begin{aligned} \theta(F_n) &= \int \cdots \int g(x_1, \ldots, x_m)\, dF_n(x_1) \cdots dF_n(x_m) \\ &= n^{-m} \textstyle\sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} g(X_{i_1}, \ldots, X_{i_m}). \end{aligned} \tag{2}$$

Hoeffding's U-statistics is defined by

$$U_n = n^{-[m]} \textstyle\sum_{\{1 \leq i_1 \neq \cdots \neq i_m \leq n\}} g(X_{i_1}, \ldots, X_{i_m}) \tag{3}$$

and is an unbiased estimator of $\theta(F)$ (when $n \geq m$). For $m=1$, (2) and (3) are the same, but for $m \geq 2$, (2) is not, in general, unbiased (but there n need not be $\geq$ m). What the two statistics have in common is that the summands are not independent, so that the classical large sample results may not apply directly. Also, in the context of multi-sample models, (1) - (3) have been extended to functionals of more than one d.f. Hoeffding (1948) succeeded in establishing the optimality of U_n and incorporating a clever projection theorem to show that under suitable regularity conditions, the asymptotic normality holds for $n^{1/2}[U_n - \theta(F)]$. The sufficiency (and completeness) of sample order statistics play a basic role in the (nonparametric) optimality of U_n. He also showed that $\theta(F_n)$ and U_n are indeed 'very close' to each other. In my estimation, Hoeffding's (1948) work is one of the most outstanding ones in the area of nonparametrics. It laid down the foundation of nonparametric estimation (which we shall discuss later on), and, besides, the projection-result has played an evolutionary role in other developments to follow [viz., Hájek (1968) and van Zwet (1984)]. The U-statistic U_n is a symmetric function of $X_1, \ldots, X_n$. What is more [viz., Berk (1966)], $\{U_n, n \geq m\}$ forms a ***reversed martingale*** sequence. This latter property provides a strong law of large numbers for U-statistics which are reported in detail in Sen (1981). These latter results are of prime importance in sequential nonparametrics (and will be discussed later on). The reverse martingale property extends to sampling from a finite population and the theory of U-statistics has a similar stand in that setup [Sen (1981)]. Lehmann (1951) extended the theory of U-statistics to the two-sample case and exhibited that many of the common rank test statistics can be expressed as such (generalized) U-statistics. The easy access to the study of the asymptotic distribution theory of U_n has been really an asset for studying the asymptotic properties of nonparametric tests. Finally, (1) extends readily to the case where m may not be finite, and we shall refer to that later on.

(iii) ***Rank order statistics.*** For the two-sample problem, the classical Wilcoxon-Mann-Whitney statistic can be expressed as a generalized U-statistics, and it also belongs to the class of ***linear rank statistics.*** Similarly, the well known Wilcoxon test statistic (in the one sample case) belongs to the class of ***signed-rank statistics.*** These rank order statistics occupy a central position in the nonparametrics. Thanks to the Hoeffding-Terry fortification and the Hájek unification, ***locally optimal*** (i.e., ***most powerful***) ***rank tests*** are generally some rank order tests. Secondly, the classical permutational limit theorems apply to such rank order statistics, so that for the limiting distributions (under the null hypothesis) one can rely on some well known results. Such rank order tests also enjoy asymptotic optimality properties for local (contiguous) alternatives when the scores are chosen appropriately. Nevertheless, the exact distribution theory when the null hypothesis does not hold is generally cumbrous, and in the asymptotic case, it demands additional regularity conditions (on the underlying distributions and the incorporated score functions). This has been a very fruitful area of research during the past 40 years. Most notable contributions are due to Chernoff and Savage (1958) and Hájek (1962, 1968). While Hájek (1962) confined himself to local (contiguous) alternatives and obtained the asymptotic normality results under minimal regularity conditions on score functions, the other two papers are geared for general alternatives. Chernoff and Savage (1958) assumed appropriate 'growth' conditions on the score function (and its first two derivatives), but allowed the two d.f.'s (F and G) to be quite arbitrary (but continuous). Hájek (1962) was able to deal with square integrable score functions in a comparatively more general regression setup, but under the protection of contiguous alternatives and for densities having finite Fisher information. Hájek (1968) finally got rid of contiguity and finite Fisher information (in a general regression setup) under a mild condition on the regression constants and a variance condition. Pyke and Shorack (1968) incorporated a novel 'weak convergence' approach for the study of asymptotic normality of two (and one) sample linear rank statistics, and this deserves a special mention too. This Pyke-Shorack approach has also been fruitfully adopted in the study of asymptotic normality of linear combinations of functions of order statistics, and this area has also been quite active in research during the past twenty years.

Like (generalized) U-statistics, the rank order statistics are generally not linear in the observations, and they may not be explicitly expressible as sums (or averages) of independnent random variables. Naturally, there was an interest to have a projections into sums of independent random variables, and the Hájek (1968) theorem provided a clear cut answer to this query. Whereas the U-statistics are reversed martingales and a similar property holds for generalized U-statistics [viz., Sen (1974)], for rank order statistics, the picture is somewhat different. For the null hypotheses situations, we have certain martingale characterizations (which we shall discuss later on), but those may not hold when the null hypothesis is not true.

(iv) *Emprical distribution(s) based procedures.* The von Mises functional in (2) is based on the empirical d.f. F_n, and the same feature holds for the several sample problems. Recall that if $X_1, \ldots, X_n$ lead to the order statistics $X_{n:1} < \ldots < X_{n:n}$ and if R_i = rank of X_i among $X_1, \ldots, X_n$, then

$$n^{-1}R_i = F_n(X_i) \text{ and } X_i = X_{n:R_i}, \text{ for } i = 1, \ldots, n. \tag{4}$$

By virtue of this relation, the rank statistics are also based on the empirical d.f.'s. There are, however, other statistics which are more explicitly based on such empirical d.f.'s. For example, the Kolmogorov-Smirnov statistics (one-sample case) are given by

$$\begin{aligned} D_n^+ &= \sup_x [F_n(x) - F(x)] \\ &= \max_{1 \le i \le n} [\tfrac{i}{n} - F(X_{n:i})], \end{aligned} \tag{5}$$

$$D_n^- = \max_{1 \le i \le n} [F(X_{n:i}) - \tfrac{i}{n}], \tag{6}$$

and

$$D_n = D_n^+ \vee D_n^- = \max_{1 \le i \le n} |F(X_{n:i}) - \tfrac{i}{n}|. \tag{7}$$

Similarly, the Cramér-von Mises statistic is given by

$$W_n = \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 \, dF(x). \tag{8}$$

All of these statistics have been extended to the multi-sample case, and these are reported in detail in Hájek and Sidăk (1967). The usual probability integral transformation [i.e., $x \to F(x)$] reduces the problem to the case where F is the uniform (0, 1) d.f., and hence, properties of the uniform order statistics $Y_{n:i} = F(X_{n:i})$, $i = 1, \ldots, n$, can be incorporated in the study of the distribution theory of the statistics in (5) - (8) [under the null hypothesis]. However, the picture becomes quite complicated when n is large. In this context, the weak convergence of $n^{1/2}(F_n - F)$ (to a tied-down Gaussian process) provides the basic key to the study of the asymptotic distribution theory of statistics of the type (5) - (8) [which are functionals of F_n]. An excellent treatise of this is available in Chapters V and

VI of Hájek and Sidák (1967). There are a number of points worthy of mentioning in this context. Firstly, the simplicity of the weak convergence approach may break down if the null hypothesis is not true. If F is the true d.f. and the null hypothesis postulates that $F \equiv F_0$, then when F holds, $n^{1/2}(F_n - F_0)$ may not have the weak convergence properties (for any fixed $F \not\equiv F_0$). Even for local alternatives (vix., $F = F_{(n)} = F_0 + n^{-1/2}w$), the weak convergence relates to a highly nonlinear 'drift function' for which the boundary crossing probabilities are not that precisely known. This creates additional problems in the study of the (asymptotic) power properties of these statistics. Secondly, if one confines himself to the conventional location/scale or other parametric alternatives, then the tests based on statistics of the type (5) - (8) generally perform less favorably than the usual rank tests, although in terms of consistency, they encompass a much wider class of alternatives than the other rank tests. Finally, in a majority of the cases, the postulated d.f. F_0 may involve some nuisance parameters which need to be eliminated through suitable estimation or transformation procedures. This generally leads to a far more complicated distribuiton theory (even under the null hypothesis), and only in some specific cases the end product is palatable [viz., Durbin (1973)].

Let me move on to categorical data, grouped data and discrete data nonparametrics. The classical Pearsonian "goodness of fit" test for categorical data models is one of the oldest nonparametric procedures, and so is the Fisher's 'exact test' for two-way contingency tables. It may be noted that Fisher's exact test for two-way contingency tables is basically a 'randomization test', so that whatever we have discussed about the latter applies to the former. The Perasonian goodness of fit test is based on the multinomial law, so that if the null hypothesis specifies all the individual probabilities, we have a genuinely distribution-free test for which the asymptotic chi square distribution theory holds under very broad conditions. The situation is somewhat more complicated when there are nuisance parameters in the specification of the hypotheses. In this context, the BAN (best asymptotically normal) estimators of the nuisance parameters pave the way for nice asymptotics with only minor adjustments for the degress of freedom for the (asymptotic) chi square test. Moreover, incorporation of BAN estimators also allows the adaptation of noncentral chi square distributions for 'local alternatives' - a feature very much comparable to other nonparametric tests. There are other points which will be discussed later on. Often, the categories represent an ordered classification, and this relates to the so called "ordinal categorical data" models. Ordinal categorical data models are close to "grouped data" models. In the later cae, often "grouping" is made on the basis of convenient "class intervals", so that the underlying variate may be assumed to be continuous. This feature makes it possible to adopt the classical nonparametrics (for the continuous case) with direct adjustments for grouping [viz., Sen (1967) and Ghosh (1973)]. Such an assumption of a continuous 'trait' underlying ordered categories has been in effect for various psychometric and educational testing problems, and the origin of *Z- (or normal-) scores* in this context rests on the underlying normality of the trait [viz., Guilford (1956)]. Discrete data models refer to random variables whose distributions admit jump discontinuities, so that for the sample observations "ties" may no longer be negligible, in probability. While adjustment for ties and adaptation of the randomization principle render conditionally (permutationally) distribution-free tests, other properties depend a lot on how the adjustments for ties are made. We may refer to Vorličkova (1970) for a nice account of these developments.

The main motivation for prescribing an EDF tests in a particular context is to make the size of the test independent of the underlying distribution. This *robustness* feature is shared by ADF tests too: The form of the underlying probability distribitions is not specifically assumed and the scope of such ADF tests is generally greater than the corresponding large sample versions of parametric tests. Nevertheless, this *validity robustness* has to be judged against the performance of such tests when the null hypothesis does not hold. Characterizations of (exact) optimality of nonparametric tests rests on rather restrictive regularity conditions, and hence, alternative avenues were explored to draw a more comprehensive picture. Characterizations of *LMPR* (locally most powerful rank) test constitute a big step in this direction; we may refer to Hájek and Sidák (1967) for an excellent account of this. There are, however, certain open problems, particularly, for the multi-parameter case, and we shall refere to that later on. Another giant step in this direction was the introduction of "*Pitman-alternatives*" leading to the *PARE* (Pitman asymptotic relative efficiency) measure. Note that if a test is *consistent*, for any fixed alternative its power will tend to one as the sample size increases. Hence, to study the asymptotic power, it may be convenient to choose alternatives approaching the null case in such a manner that the limiting power function exists and is different from 0 or 1. Pitman-alternatives pertain to this scheme. For such alternatives, for two competing tests (say, T_1 and T_2), based on two possibly different sequences of sample sizes (say $\{N_1\}$ and $\{N_2\}$), to have the same limiting power, the limit of the ratio N_1/N_2 exists (under general regularity conditions) and serves as a measure of the relative efficiency (of T_2 with respect to T_1). In the classical univariate problems, this limit (i.e., N_1/N_2) exists and is independent of the particular sequence of alternatives we choose. This feature may not be generally true in the multivariate problems [viz., Puri and Sen (1971)]. We shall comment on that later on. Bahadur (1960) considered an alternative ("non-local") approach, where in fact an exponential convergence rate of the *size* of a test (to zero) is incorporated in the definition of "*asymptotic slopes*", and in this framework the *BARE* (Bahadur asymptotic relative efficiency) is the limit of the inverse ratio of the sample sizes needed for the two tests to have the same asymptotic slope. In a similar fashion, Hoeffding (1965) showed that for multinomial populations, the likelihood ratio test is asymptotically optimal. There are some other measures of asymptotic relative efficiency (ARE) and we shall make some comments on them later onl. Defined in either way, the ARE of a test depends on the underlying probability law as well as the specific alternative one has in mind. In a nonparametric setup, this underlying distribution is generally of unspecified form, and hence, there was a genuine question: Is it possible to formulate a rank test (EDF or ADF) which is (at least, asymptotically) optimal for the specific alternative for all distributions belonging to a given class?

Some affirmative answers to such a query have been provided by the so called *adaptive procedures*, and we shall discuss them later on.

In the retrospective picture, *nonparametric estimation* occupies a special position. The quantiles and some other measures based on quantiles (such as the interquantile range) were in use for a long time. But their systematic incorporation into the development of *L-estimation theory* started in the Fifties. Initially, the emphasis was mainly on parametric models [viz., Sarhan and Greenberg (1962)], although the treatment of Huber (1981) has added a good amount of robustness considerations. Huber (1981) also contains a detailed (updated and motivating) account of so called *M-estimators* which were once proposed mainly on "local robustness" grounds but have merged into the main stream of nonparametric estimation. Nonparametric estimators based on robust rank test statistics (known as *R-estimators*) occupy the central position in this respect. The past thirty years have witnessed phenomenal developments in this sector, and we shall discuss them later on. In a sense, Hoeffding's (1948) U-statistics can be regarded as the precursors of modern nonparametric estimators (where the underlying d.f. F is allowed to be a member of a wider class $\mathfrak{F}$, and the parameter is regarded as a functional of F over the domain $\mathfrak{F}$). Such *statistical functionals* have a natural nonparametric flavor, and during the past two decades an enormous amount of research work has flowed into this stream. *Smoothing techniques* and *resampling plans* (such as *jackknifing, bootstrapping,* and *functional jackknifing*) have evolved on sound footing and applications in diverse tributaries have mushroomed. We shall discuss some of these later on. The past two decades have also witnessed significant growth of literature on two important sectors: *semi-parametric* models and *Bayesian nonparametrics*, and we intend to touch on them in the perspective section.

In any specific area of research, developments and impacts are generally judged from diverse angles. Among these, the two most notable are (i) publication of appropriate textbooks and monographs (so as to bring the research products down to the applications level), and (ii) adaptation in appropriate curriculum (so as to train people to enhance applications in other fields). Judged from either aspect, the nonparametrics have indeed made a stride. Back in the Fifties, at the graduate level (in statistics), there was hardly any textbook in nonparametrics. Mostly, EDF tests were included (in nut shells) in some contemporary texts. In applied statistics, of course, Siegel's (1956) text served very well. Fraser's (1957) advanced monograph was probably the first one in this field - although it has a lot of material collected from diverse sources, and a time-permitting unification would have served the purpose better. A number of textbooks in nonparametrics appeared in the Sixties, some aimed for mathematical readers, while others for less mathematically sophisticated ones. In this context, Hájek and Sidák (1967) deserves a special mention, and we have already elaborated that in the preamble. In the mainstream, Lehmann's (1975) text is a valuable input too - although it is geared for a somewhat lower mathematical level. The last twenty years have witnessed the appearance of a variety of nonparametric texts and monographs. Some of these are in specific area of applications [such as multivariate analysis, sequential analysis, linear models etc.], while others have introductory to intermediate levels of presentation. A variety of texts on applied nonparametric statistics has also appeared during the past fifteen years. All in all, they cater a broad domain of audience in statistics, applied statistics and in other disciplines too. In terms of adaptation of nonparametrics in various curriculum, we have a very encouraging picture too. Nonparametric statistics (or methods) have not only entered the main areana of "core" courses in mathematical statistics at the graduate as well as undergraduate levels, but also in various applied statistics programs. In the old days, anthropometry, biometry, econometry, psychometry and agricultural sciences were the ones to adore statistical methodology for their curriculum developments, and, in this context, multivariate analysis, design and planning of experiments and statistical inference were the three broad avenues for implementing statistical reasonings. However, in all of these sectors, classical parametric models used to be advocated, and usually, nonparametrics were referred to as inexact and/or inefficient tools. By now, the picture is entirely different - thanks to the robustness consciousness of statisticians and other scientists, the validity-vulnerability of parametric procedures has cast serious doubts on their adaptability (without any reservation). Robust and nonparametric procedures have gained popularity in this respect, and semi-parametric models are also being used in more and more complex situations. In life testing experiments, clinical trials and reliability theory, nonparametrics for stochastic (e.g., counting) processes are being increasingly used, and this is backed up by curriculum developments too. Would you unconditionally surrender to a normal model, or start with a nonparametric approach and show how a normal model can be reached under plausible conditions?

3. PERSPECTIVES

The roots of the 'perspectives' are, of course, in the 'retrospectives', and hence, we shall consider them in the same vein. The genesis of statistical sciences lies in a variety of other disciplines (ranging from agricultural science, anthropometry, biometry, gambling, genetics to various social sciences). Although the scenario has progessively changed over time, the basic tune has remained the same. In this interface of statistical sciences, the reflection of the mathematicians are in the sub-sigma fields of measure theory (and modern functional analysis), mathematical statisticians' in the solid foundation of methodology (mostly, pertaining to the narrow domain of statistical inference), and the statisticans' strides are in the greater realm of modelling, planning of experiments and diverse applications. With respect to the nonparametrics, the interface is quite similar. Exclusing Great Britain, most of the European researchers have chosen the abstract mathematical tracts, and their work has principally the probabilistic flavor, often dehydrated to an extent where applications may not be submerged. Of course, a most notable exception to this is the Czechoslovak School led by the late Professor Jaroslav Hájek, and during the past thirty years, their achievements have surpassed any other single group in Europe. Hungarian School also deserves a special mention. In the North American continent, India, Japan, England and in Australia, there has been a good deal of mixture of mathematical and applied works, and, to a greater extent, this interaction has enriched the modern nonparametrics to acquire a solid foundation and of being fruitfully adaptable in a variety of disciplines. At the same time, with the annexation of new disciplines in science, technology and in other diverse fields, there is a challenge for nonparametrics to cope up with the demands and to lie abreast of the needed developments.

The 'textbook' nonparametrics (up to the Sixties) were mostly confined to the classical one-sample (symmetry) problem, several sample (one-way ANOVA) problem, simple two-way ANOVA problem and the bivariate independence problem. Nonparametric tests for 'trend' made a debut in some cases. Almost thirty years ago, when I started teaching a graduate level course on "Statistical Bio-assays", the first problem I encountered: Why assume normality (or log-normality) of the "tolerance" distribution without any reservation? Specifically, in the estimation of "relative potency" of a new drug with respect to a standard one, whether be in 'direct' or 'indirect' assays, such an assumption appeared to be quite restrictive, and not very suitable in a majority of the cases. The problem was to estimate this relative potency (point as well as interval) in such a way that the end-product does not depend on the particular dosage (= dose-transformation). This invariance naturally called for nonparametric methods (which are invariant under strictly monotone transformations), and the estimator [Sen (1963)] appeared to be the same one considered by Hodges and Lehmann (1963) for the two-sample location model. This paved the way for further nonparametrics for bio-assay problems, and some of these developments are reported in Sen (1984). Interestingly enough, these problems cropped up in applications and the methodology was derived from existing as well as novel nonparametric methods. Let me iterate some other basic problems in nonparametrics which have definite genesis in the application domain. Armitage, McPherson and Rowe (1969), in the context of "interim analysis" in medical studies, confronted the problem of "repeated significance testing" on "accumulating data". For simplicity of analysis, they assumed "independent increments" and, mostly, in a parametric setup, showed how the level of significance is affected by such interim analysis? Numerical studies dominated the investigation. This study has been of paramount importance in the development of "time-sequential nonparametrics". The (U.S.) National Heart, Lung and blood Institute (NHLBI) wanted to study the effect of hypertension and high cholesterol on the risk of heart attack, and for this purpose, planned a multi-center clinical trial (1972-84). A set of 3952 healthy males (between the ages 35-50), all having cholesterol level 230 (or more), were randomly assigned to two groups: Placebo and Treament. In this double-blinded study, the flow of the events (heart attacks) wsere to be recorded (over a 12 year period) and on this "accumulating data set" appropriate statistical tests were to be performed to decide on the tenability of the null hypothesis of "no effect" due to change in the cholesterol level. There was, however, a basic point cropping up from the *medical ethics*: If the null hypothesis is not true (i.e., treatment better than placebo), then the trial should be terminated as early as (statistically) possible, and the subjects in the placebo group should be transfered to the treatment group (for better prospects). This essentially amounted to a "time-sequential problem". In a parametric setup, if one assumes that both the survival distributions (for placebo and treatment groups) are exponential, then one has "independent increments" and the Armitage-McPherson-Rowe procedure can be adopted. However, with a moderately large number of covariates (viz., diet, physical exercise, age, smoking/nonsmoking, blood-chemistry measures, etc), such an assumption is not likely to be very realistic. The next step may be use the celebrated Cox (1972) "proportional hazard model" (PHM) (which we shall discuss later on). Again, this assumption of proportionality of hazard functions may not be very realistic in many practical applications [crossing hazard functions are not at all uncommon in medical studies.] There was thus a general feeling whether alternative nonparametric time-sequential procedures (allowing the incorporation of possibly non-indpendent increments and without requiring the PHM) could be developed to suit the purpose. Chatterjee and Sen (1973) developed some martingale-characterizations for "progressively censored" rank statistics and incorporated them in the formulation of some "weak convergence" results which enabled the adaptation of the classical sequential analysis [Wald (1947)] in such time-sequential problems. Rank ANOCOVA procedures were later on extended to such time-sequential setups and "staggering entry" plans were incorporated in a natural manner. Most of these developments are reported in Sen (1981, Ch. 11) and Sen (1985, ch. 2). During the past two decardes, a number of multi-center clinical trials have been conducted, some by specific drug companies (for the promotion of their products) and some by Government Agencies, and in most of these cases, the pitfalls of the conventional parametric approaches have come to the surface. Nonparametric methods have a vast scope of appliation in this novel area of research. There is also a need to coordinate plausible nonparametric methods with appropriate semi-parametric counterparts, so that the common methodology can be developed. We shall make more comments on it later on.

Econometrics and social sciences constitute an important area where nonparametric methods sparked quite long ago. In the conventional models (for which statistical infrence is sought), it is generally assumed that the sample observations represent independent and indentically distributed random variables (i.i.d.r.v.). For linear models, the i.i.d. structure replaced by a deterministic part (involving known regression vectors) and a stochstic part for which the i.i.d. structure is imposed. In the classical parametric case, the stochastic part relates to the "errors" (or chance variables) and generally it is assumed that those "errors" are "homoscedastic" and normally distributed. In econometric time series models, departures from such basic assumptions may occur in more than one way: non-linearity of the deterministic part, heteroscedasticity, non-normality of the errors and possible dependence of these errors too. Among the types of dependence, moving average (MA), autoregressive (AR) errors and ARMA models are popular. There are other "mixing sequences" in the literature - but they often reside in pen-park of mathematicians. From statistical point of view, it is of considerable importance to judge the adequacy of a partiulcar parametric model with especial emphasis on the robustness against possible departures from the model-based assumptions. In this respect, *robust procedures* have a greater appeal than the nonparametrics. The Box-Cox and other related transformations are often advocated to achieve greater adherence to such model-based assumptions, although often without much practical interpretations. Koenker (1982) has a detailed account of such robust methods in econometrics (including some recent developments on "regression quantiles" which have come up as very handy tools in the prescription of such robust methods). In view of the fact that the "errors" are generally serially dependent, in econometrics, there is some interest in studying such "serial dependence" patterns, in addition to drawing statistical conclusions on the deterministic part. Under plausible Markov dependence structures (but still confined to normal errors), such results have been extensively studied in the literature. An excellent work on "serial statistics" in a nonparametric setup is due to late Professor M. N. Ghosh (1954). Of course, this appeared before the concept of "continguity" was introduced to nonparametricians, and it is not surprising to see that the 'method of moments' based proofs relied on comparatively stringent regularity conditions. Although in the recent past, there has been some attempt to relax some of these regularity conditions (by the use of contiguity of probability measures), these developments mostly relate to the null hypothesis of i.i.d. errors (for which the classical nonparametrics apply elegantly), and hence, much more remains to be accomplished before they are palatable in the same generality as their parametric counterparts. The functional approach (to be considered later on) has a far greater prospect in this context. In social sciences, including economics, psychometry and educational statistics, nonparametric methods have a natural appeal. Part of this impetus came from the fact that often 'ordinal categorical data' models are encountered for such research (where nonparametrics appear to be more appropriate) and, partly, the credit goes to these researchers in recognizing the inadequacy of the classical parametric methods in effectively dealing with their basic statistical analysis. The Lorenz curve (depicting the concentration of income of a community or society), Gini's mean difference, Gini's coefficient and the entire battery of income inequality measures as well as poverty and affluence indexes [see, Sen (1989)] have largely nonparametric flavors. It would be a definite mistake to overlook the role of nonparametrics in socio-economic sciences. Some of the developments in this area has been found to be very useful in other areas as well. For example, in reliability theory and system analysis, the Lorenz curve and its tributaries are finding new adaptations. Let's hope that this interaction will be even greater in the future.

In practice, very rarely, one encounters simple univariate models. Generally, on has multiple measurements, some relating to the so called *regressors* or *independent variables* and the others are the *primary* or *dependent variables*. In this setup, *linear models* in a general multivariate setup crop up naturally. The classical parametric procedures for the analysis of such multivariate models have been built in on linear structures and the multivariate normality of the underlying random vectors constitutes a basic assumption. The question is, however, how much confidence you may have in this basic assumption of (multi-) normality of error components? There are some basic difficulties in using the *variance-stabilizing* transformations in the multivariate case, and moreover, not all the variables may have continuous distributions. Thus, the adaptability of the classical parametric procedures in various practical problems remained questionable. In the Fifties, there were some scattered attempts to extend the classical univariate nonparametric procedures to the multivariate case. However, these developments were spotty and piecemeal. The basic problem was that unlike the simple univariate cases, in multivariate problems, rank based procedures are not generally distribution-free. Although ADF tests were easy to consider, a general query was to characterize suitably the EDF nature of multivariate nonparametric tests. This was systematically explored by Chatterjee and Sen (1964) and Chatterjee (1966), and their *permutationally (conditionally) distribution-free approach* gained a lot of momentum in the late Sixties; the monograph of Puri and Sen (1971) bears this testimony.

There are some distinct differences between univariate and multi-variate nonparametrics. Firstly, there are EDF rank tests for the simple univariate models, but for their multivariate counterparts, generally, one has only conditionally (permutationally) distribution-free tests. This makes the usual nonparametric tables rather unusable in the genuine multivariate cases, and one has to rely heavily on permutational algorithms or on the asymptotics. Such asymptotics may demand comparatively larger sample sizes. [Many people have argued that in a p-variate model, the sample size n should be so large that p^2/n is small, and this clearly indicates the inadequacy of asymptotic approximations in the multivariate case when p is not small and n is not so large.] Thus, we are confronted with the issue: what type of asymptotics are more appropriate when p^2/n is not so small (i.e., for moderately large n)? The resampling methods (such as bootstrapping and jackknifing) play a vital role in this context. Secondly, the usual parametric MANOVA procedures enjoy some invariance properties (under linear transformations) whereas rank based procedures are invariant under coordinatewise monotone transformations, but not necessarily under $\underset{\sim}{X} \rightarrow \underset{\sim}{Y} = \underset{\sim}{B}\underset{\sim}{X} + \underset{\sim}{c}$ where $\underset{\sim}{B}$ is nonsingular. Of course, in many problems, the coordinatewise arbitrary (but strictly) montone transformations have far more practicality than arbitrary linear transformations (on the vectors), but viewed from the study of optimality or monotonicity of power properties, this lack of invariance makes it difficult to justify the usual canonical reduction of MANOVA tests in terms of characteristic roots of the noncentrality matrix. This makes it coomparatively more difficult to characterize the optimality of nonparametric methods in multivariate cases. Since these are not generally invariant tests in the same set of their parametric counterparts, the optimal (invariant) characterizations of parametric tests may not totally hold for such nonparametric ones. However, in the asymptotic case, the parametric and nonparametric procedures do share the common properties to a larger extent. Thirdly, in the univariate models, PARE is generally defined in such a way that it does not depend on the specific alternative one chooses to derive it. On the other hand, in genuine multivariate cases, though PARE is defined on ambiguously, it may depend on the specific direction of the alternatives (viz., in the multivariate one-sample model, it depends on the direction cosines of θ). As such, the PARE may depend on the particular sequence of alternative hypotheses one has in mind. This variational picture has led to the formulation of the so called *maximum* and *minimum* PARE [c.f. Bickel (1965)], allowing the extreme fluctuations over the alternative parameter space. There may be some situation where such extrema behave very irregularly, although in the middle line, the PARE behaves steadily [viz., Puri and Sen (1971), chapters 4 and 5]. A similar picture holds for BARE, and, often, the computation of the "exact Bahadur slopes" seems to be quite cumbrous for multivariate models. Fourthly, from computational point of view too, multivariate nonparametrics are much less attractive than their univariate counterparts. This is particularly ascribable to R-estimators in multivariate (and linear) models. Finally, in multivariate problems, often, one may be interested in some "restricted type of alternatives". For examples, positive orthant alternatives in the multivariate one-sample model, generalized ordered alternatives in the multi-sample model, and, in general, some "positively homogeneous cone" in a general setup. Although, in this setup, some parametric procedures have been developed [viz., the recent monograph of Robertson et al (1988)], the theory is not so well founded [as in the case of global alternatives where the likelihood ratio tests dominate the picture]. This picture is shared by multivariate nonparametric procedures too. Roy's (1953) *union-intersection (UI-) principle* provides a workable access to such problems, although characterizations of (asymptotic) optimality of such UI-rank tests require more in depth studies.

For simple (univariate) linear models whenever the null hypothesis induces "invariance" under appropriate group of transformations on the sample sapce onto itself, EDF rank tests exist, and within this class, specific ones can be so chosen that they are (asymptotically or locally) optimal for specific families of densities. This feaure is generally not tenable when the null hypothesis is not an hypothesis of invariance. [For example, in a multiple regression model $Y = \beta X + e = \beta_1 X_1 + \beta_2 X_2 + e$, the null hypothesis H_0 relates to $\beta_1 = 0$ treating β_2 as a nuisance parameter. Under H_0, the components of Y are not i.i.d., so that permutational invariance may not hold.] ADF tests for such "subhypotheses" have been considered in the literature [viz., Chapter 7 of Puri and Sen (1985)]. In this context, some asymptotic linearity results on rank statistics [mostly, due to Jurečková (1971)] play a vital role. In randomized block designs [or other two-way layouts], permutational invariance under intra-block permutations yields EDF tests [viz., Friedman (1937)]. However, such intra-block rank tests ignore inter-block comparisons to some extent and thereby may not be very efficient. Hodges and Lehmann (1962) sparked an interesting idea of incorporating "ranking after alignment" (or aligned ranks) to improving the efficiency of such tests. Basically, aligned ranks distorts the independence of the errors, and hence, EDF tests may not generally exist. However, conditionally (permutationally) distribution-free tests exist and they generally perform better than the intra-block tests. Robustness and asymptotic efficiency aspects of such aligned rank tests have been studied in detail [see, for example, Sen (1968)], and in this context, multivariate nonparametrics have provided the basic frame.

The fruitful interaction of multivariate nonparametrics and linear model nonparametrics has actually carried on the scope of nonparametric methods to a much wider area. The so called "growth-curve" models, repeated measurement models, linear models involving stochastic predictors all belong to this domain. Some of the developments in this broad area is reported in Chapter 8 of Puri and Sen (1985) and Sen (1984). Statistical models arising in biological assays can also be fitted in this spectrum. In this context, the main feature of the nonparametrics is their "global robustness", in the sense that the underlying distributions may be quite arbitrary and may not require the usual moment conditions needed for the parametric procedures. "Local robustness" considerations have gained popularity in the recent past. Huber (1981) is a good source for this line of thinking. In many cases, through the proper use of transformations, approximate linearity and normality conditions can be justified, but there may be "outliers" or "error contaminations" which can have more serious effects on the performance of the usual normal theory (parametric) procedures. *M-procedures* have been formulated for such linear models, which without much compromise on the efficiency (under the assumed model) maintains high 'local robustness' levels. Comptationally too, such M-procedures may be somewhat less burdensome than the rank procedures. The intricate relationship of M- and R- procedures have been studied by a host of workers [viz., Jurečková (1977)], and this clearly depicts the relative merits and demerits of both of these procedures. In this context too, "asymptotics" play a dominant role, and there is ample rorom for develpment of moderate sample asymptotics (as well as small sample cases). Jurečková's linearity results rest on certain regularity conditions, and they may not be universally tenable. Recently, Heiler and Willers (1988) have considered an alternative approach which eliminates some of the stringent regularity conditions in the Jurečková approach. "Regression rank scores", germinating from "regression quantiles" also appears to be quite promising in this area [c.f. Gutenbrunner and Jurečková (1990)]. There is, however, one basic concern with these linear model nonparametrics. Most of these developments is geared towards asymptotics where local alternatives are chosen as the inhabitants. Some of these tests may behave quite well for such local alternatives, but may not be that well for non-local ones! In fact, often, for consistency of such tests (for fixed alternatives), one may require rather stringent conditions on the "scores" or the "regressors". Moreover, the basic assumption of "linearity" (sans the normality of errors) may not always appear to be that appealing. Although transformations of variables may lead to approximate symmetry (or even normality) of the chance components, they may also affect the linearity structure. This has prompted some workers to look at the "nonparametric regression" from a purely functional point of view, and we shall make some comments on it later on. But in life (and science), the foundation rests on assumptions (or projections), and so long as such assumptions can be judged tenable optimal decisions can be made with adequate margins of risks. In nonparametric linear models too, it would be a mistake, perhaps, to elminate "linear" and invite "regression" without looking for the consequences for such a substitution. The penalty may be a high price in terms of the adaptability of the related asymptotics when the size of a given data set may not be that large! Validity, robustness and adaptability for finite sample sizes should all be given due considerations in this judgement, and linear model nonparametrics would probably be a better choice than the others.

There are cetain situations where the sample observations are gathered sequentially, and in some situations, appropriate "stopping rules" are incorporated in the sampling scheme. In either case, the number of sample observations is an integer valued random variable, and the branch of statistical analysis dealing with such sequential plans is known as ***sequential analysis.*** Nonparametrics have marched into this valley too. Conceptually there were difficulties in formulating the nonparametric analogues of the classical Wald (1947) sequential probability ratio tests (SPRT), as the functional form of the density function is generally unknown in a nonparametric setup. For certain special types of nonparametric alternatives [viz., Savage and Sethuraman (1972)], such nonparametric sequential tests were considered and their termination probability was studied. However, the theory lacked the generality of the parametric SPRT and there were computational problems too. Incorporating an LMPR approach, Sen and Ghosh (1974) considered sequential rank tests for location and showed that the usual properties of the SPRT are shared by such sequential rank tests when the alternative is 'close to' the null one. the basic formulations of the these rank tests rests on (i) some martingale characterizations of rank tests (under appropriate hypotheses of invariance), (ii) weak as well as strong invariance principles for sequential rank statistics and (iii) the Hájek-LeCam formulation of "contiguity of probability measures" - all of these were studied earlier by the same authors [and others]. Repeated significance test (RST) in nonparametric setups have also been developed on the same foundation, and these have been accounted for in detail in Sen (1981). In the estimation domain, the problem of bounded-width confidence interval for location/regression parameters has also been studied in a nonparametric fashion [viz., Geertsema (1970) and Sen and Ghosh (1971)], and an upto date account of these nonparametrics is given in Chapter 10 of Sen (1981). In the context of minimum risk point estimation of location, a sequential estimation rule evolves in a natural way, and in the nonparametric case, such sequential R-estimation procedures have been developed in Sen (1980), and incorporating other related developments, an account of this branch of research is given in Chapter 4 of Sen (1985). Although on the grounds of validity-robustness and asymptotic efficiency, these sequential nonparametrics have special appeal, there remains ample room for further developments of the theory, particularly in the non-asymptotic case. How good are the Wiener process approximations for sequential rank statistics when the alternative hypothesis may not lead to "contiguity" with respect to the null one? Similarly, how close will be the minimal risk characterization when the cost per unit sample is not that small? The advent of modern computing facilities and resampling methods can provide us with valuable information in this quest.

Nonparametric methods have also been considered for various problems in reliability thoery and systems analsysis. Traditionally, exponential life distribution underlies most of the practical applications, and the "constant" (failure) hazard rate property of the exponential law yields simple solutions in a variety of cases. However, such a constant hazard rate assumption may not be very realistic in practice. "Aging" is a natural phenomenon, and change in the hazard function with the "age" is also quite common. Alternative formulations of life distributions include the following ones:

(i) Increasing failure rate (IFR) class,

(ii) Decreasing failure rate (DFR) class

(iii) IFRA and DFRA class (in averages)

(iv) New better (worse) than used (NBU or NWU),

(v) Decreasing (increasing) mean residual life class (DMRL/IMRL)

(vi) NBUE and NWUE class (in expectation).

Together they encompass a broader class of life distributions, and statistical analysis based on such general models have greater scope too. Nevertheless, in actual practice, it is not uncommon to have a change: a failure rate may be increasing (or decreasing) to start with, but then goes for a change at some point. The robustness of such a specific model based analysis to such a change-point model requires serious considerations. It may be more appropriate to allow the life distribution to be more arbitrary, so that the statisstical conclusions have far greater scope (although for such a specific model, they may not be fully efficient). Estimation of mean residual life, renewal function, availability of a system under provisions of spare and repair and many other problems in this general area have been recently studied by using alternative nonparametric and robust methods, and there is an enormous scope for further develpments in this area of vital practical interest. We may refer to an excellent review article by Hollander and Proschan (1984) where the basic concepts and methods have been outlined neatly.

Stochastic approximation has the genesis in statistical bioassays. Suppose that in a quantal assay, corresponding to a dose d, F(d) is the probability of a response (usually lethal). In an assay, a dose d may be administered to a number of subjects and the proportion p_d of their having responses is taken as an estimator of F(d). Thus, $p_d = F(d) + \text{error}$. The problem is to find a dose, say d_o, such that $F(d_o) = 1/2$; then d_o is termed the LD50 or the median lethal dose. A similar problem may crop up in a dose response regression $Y = g(d) + \text{error}$, where $g(\cdot)$ has a unique maximum (or minimum) at a point d_o (unknown), one may like to estimate d_o. In such a case, instead of choosing a set $\{d_j, 1 \leq j \leq m\}$ of doses apriori, one may proceed sequentially by letting the choice at the kth step depending on the outcome of the past. Specifically, one lets

$$d_{k+1} = d_k + \psi_k(Y_k^*), \ k \geq 0, \tag{9}$$

where Y_k^* relate to the data set upto the kth stage, and the form of $\psi_k(\cdot)$ may be chosen depending on the partiuclar problem in hand. The question arises whether such a sequential procedure converges with probability one, and if so, what other (asymptotic) properties can be ascribed to this sequential solution? Robbins and Monroe (1951) and Kiefer and Wolfowitz (1952) initiated the research in this area, and there has been a steady growth of literature during the past forty years. The relevance and importance of nonparametric techniques (particularly asymptotic methods) have been properly recognized martingale characterizations and weak as well as strong invariance principles for (sub- or super-) martingales play a vital role in this context. In a majority of the cases, the functional form of F(d) or g(d) is not assumed, so that estimation of d_o may need more detailed information on the nature of the derivative (of F or g) in a neighborhood of d_o. The stochastic approximation methods essentially provides estimators of the derivative function in a nonparametric fashion. With the advent of adaptive procedures in nonparametric estimation thoery, and the vital role of such adaptive procedures in stochastic approximations, it is highly expected to have more novel developments in nonparametrics tailored for stochastic approximations.

Watson (1983) contains a good description of directional data statistical problems. In a parametric setup, the von Mises-Fisher distribution has been widely adopted, and there are some others adopted to a much lesser extent. However, such a parametric form of distribution may not properly fit a given data set, and hence, the conclusions from a parametric analysis may not be totally valid. Thus, from validity-robustness considerations it may be wiser to seek nonparametric solutions whenever they exist. In the two-dimension case (i.e., for circular data) observations relate to the angles with respect to some arbitrary zero direction and these angles may be recorded either in a clockwise or anticlockwise manner. Thus, invariance with respect to the choice of "zero direction" and with respect to reversal of the direction may be very desirable for a nonparametric procedure. In higher dimensions (i.e., for spherical or hyperspherical data), the picture becomes more complex. For circular data, sample spacings and/or empirical distribution functions of the sample (angular) observations have been incorporated in the fomulation of suit one and multi-sample tests; their asymptotic relative efficiency properties have also been studied. There has not been matching progress for spherical data, and it seems that there is some definite scope for progress along the lines of multivariate nonparametrics.

The sample order statistics (or equivalently, the empirical distributions) occupy a central position in the nonparametrics. Instead of defining parameters as algebraic constants appearing in an assumed functional form of the underlying d.f., in a nonparametric formulation, a parameter is defined as a functional of the underlying d.f., i.e., $\theta = \theta(F)$, so that it may be quite appealing to choose an estimator $T_n = \theta(F_n)$, where F_n is the empirical d.f.. For certain types of functional $\theta(\cdot)$, the von Mises (1947) statistics belong to this class and the Hoeffding (1948) U-statistics are closely related; in both the cases, the degree of the underlying kernel is assumed to be finite. In a general setup, one may not have a kernel of finite degree, and a somewhat different approach is therefore needed. Basically, a Taylor's expansion (in a functional space) of $\theta(F_n)$ around $\theta(F)$ provides the desired solution. Assuming that F is continuous, we may use the probability integral transformation whereby the empirical d.f. F_n reduces to a uniform empirical d.f. U_n and F to the uniform (0, 1) d.f. U. As such, writing $\tau(U) = \theta(F^{-1}(U))$, $U \in (0, 1)$, we have

$$T_n = \theta(F_n) = \tau(U_n) = \tau(U) + [\tau(U_n) - \tau(U)], \tag{10}$$

where U_n belongs to the space D[0, 1], so that T_n is defined as a functional on the domain D[0, 1]. The differentiability conditions on $\tau(\cdot)$ have been formulated in an increasing order of generality (Gateaux → Fréchet → Hadamard (compact) types), and an elegant discussion of these properties is given in Fernholz (1983). Essentially, it enables one to rewrite (10) as

$$T_n = \tau(U) + \tau'_U(U_n - U) + \text{Rem}(U_n - U) \tag{11}$$

where τ'_U is the influence function (Gateaux derivative), and $\text{Rem}(U_n - U)$ is the remainder term. Actually, we may write

$$\tau'_U(U_n - U) = n^{-1}\sum_{i=1}^{n} IC(X_i; F, \tau) \tag{12}$$

where IC(·) stands for the influence function and the remainder term converges to 0 (stochastically) at a rate faster than $n^{-1/2}$. This linear approximation provides the access for using the classical central limit theorems and other asymptotic results for linear statistics for studying similar results for T_n. In some cases, a second order expansion is desired [viz., Sen (1988a)], and this provides additional information on Rem $(U_n - U)$. The main point is that in this formulation, the influence function $IC(X_i; F, \tau)$ depends on the unknown d.f. F, and hence, in statistical applications, one may need to estimate the unknown variance $E_F\{[IC(X_i; F, \tau)]^2\}$. For this purpose, the usual resampling plans can be adopted, and we shall discuss them later on. Statistical functionals are being increasingly used in more complex problems in nonparametric inference. For example, consider the usual nonparametric regression problem with X as the primary variate and Z as the regressor. Let f(x|z) be the conditional p.d.f. of X given Z=z, and define

$$m(z) = \int x f(x|z) dx, \ z \in R^p, \tag{13}$$

for some $p \geq 1$. There has been a lot of work on the nonparametric estimation of m(·) through that of f(·|z). Such procedures not only demand comparatively larger sample sizes but also more stringent smoothness conditions of f (and its derivatives). In the same setup, if we denote the conditional d.f. of X, given Z=z, by F(x|z), then we may as well formulate a nonparametric regression function as

$$m(z) = \theta(F(\cdot|z)),\ z \in R^p. \tag{14}$$

For example, $\theta(G)$ can be taken as $G^{-1}(\frac{1}{2})$, i.e., the median, so that m(z) is the conditional median function. Actually, (13) is a special case of (14). For both (13) and (14), the usual resampling plans may be needed to estimate the variance function [see, Gangopadhyay and Sen (1990)]. There is a good prospect for further developments on statistical functionals (covering multivariate distributions) which may as well be used for the usual time series problems.

In the classical (one as well as several samples) nonparametric testing and estimation problems, although the validity-robustness is guaranteed for a borad class of statistics, their (asymptotic) efficiency depends on (i) the underlying density function, (ii) the chosen score function and (iii) the specific alternative one has in mind. Thanks to the lucid treatment of Hájek and Sidǎk (1967, Ch. 6), it has become an acceptable norm to choose the score function Q(u), u $\in$ (0, 1) in such a way that it matches the Fisher information (or log-likelihood) score function [viz., $-f'(F^{-1}(u))/f(F^{-1}(u))$, for location models]. However, the p.d.f. f is generally of unknown form, and hence, we may not have any precise idea about such an optimal score function in a given context. Any chosen score function (such as the Wilcoxon, normal or median score function) can not be optimal for all p.d.f.'s. Early attempts to estimate such "optimal" scores rested on the estimation of $f(\cdot)$ and $f'(\cdot)$, and thereby needed unusually large sample sizes to clinch any desirable property of such estimators. At a later stage of development [viz., Hogg (1974)], attempts were made to choose a reasonable family ($\mathfrak{F}$) of distributions (involving only a finite number of members) and to choose an apprpriate f_o ($\in \mathfrak{F}$) based on the sample observations. Since the functional form of such an f_o is presumed, the corresponding Fisher information score function can be adapted (upto a location and/or scale factor). However, the larger is the cardinality of the set $\mathfrak{F}$, the higher is the risk in choosing a corrrect f_o ($\in \mathfrak{F}$) to match the unknown f (upto location and scale factors). Moreover, there is no guarantee that the actual f belongs to the chosen class $\mathfrak{F}$, although such an f may be closely approximated by a member of $\mathfrak{F}$. Thus, while the attainability of asymptotically optimal score functions was not guaranteed, such a restricted adaptive procedure generally yielded score functions which are at least 'nearly' optimal. The 'breakthrough' in adaptive procedures came through the nice observation that the finiteness of the Fisher information ensures the square integrability of the corresponding score function, and hence, a Fourier series representation (in terms of orthonormal functions) can be incorporated to estimate the corresponding Fourier coefficients, and the resulting estimated score function possesses the asymptotic optimality under quite general conditions. Hušková and Sen (1985) employed Legendre polynomials for this Fourier representation and also provided a review of other related work.

A completely nonparametric formulation of a problem may sometime lead to too many parameters (if not to a parametric functional), and with the increase in the number of such parameters, the precision of statistical inference procedures goes down. For this reason, often, statistical models include as unknowns both parameters and functions. For example, in an one-sample model, $f(x, \theta) = f_o(x - \theta)$, $f_o \in \mathcal{F}_o$, the class of all symmetric (about o) pdf's, θ represents the location parameter and f_o is an unknown function. This may be characterized as a semi-parametric model. A more appropriate example is the Cox (1972) regression model. Corresponding to a concomitant variant (possibly, vector) Z, the hazard rate for the primary variate X is given by

$$h(X|Z) = h_o(x) \cdot g(Z) \tag{15}$$

where $h_o(\cdot)$ is an arbitrary (unknown) function, and g(Z) has a distinct parametric form, viz., $g(Z) = \exp\{\underset{\sim}{\beta}'\underset{\sim}{Z}\}$. A completely nonparametric model would relate to h(x|z) being quite arbitrary without necessarily being factorizible into a "pure" hazard function and a multiplicitive function of the covariates. According to the same terminology, the classical nonparametric regression model:

$$Y_i = \underset{\sim}{\beta}'\underset{\sim}{x}_i + e_i,\ 1 \le i \le n;\ e_i \text{ i.i.d.r.v. with p.d.f. f} \tag{16}$$

may also be regarded as a semi-parametric model. With such merging boundaries between the nonparametrics and semi-parametrics, it is quite appropriate to embrace the latter and to unify the methodology in a common vein.

The last decade has witnessed a phenominal growth of the literature on semi-parametric models. The main impetus originated from the adaptation of the Cox (1972) proportional hazard model in a variety of *counting processes* typically arising in survial analysis; Aalen (1978) initiated the primary research in this fruitful area. *Partial likelihood methods* allow the treament of the nonparametric part as nuisance parameters and efficient estimation of the parametric part. In this context, *martingale-theory* plays a fundamental role. Such semi-parametric models have found fruitful adaptations in *time-sequential nonparametrics* [viz., Sen (1985, Ch. 2)], in neurobiology and sociology [viz., Murphy and Sen (1990)] and in various other disciplines. Adaptive procedures in nonparametric as well as semi-parametric models are likely to flourish significantly in the near future.

In nonparametric problems, the development of Bayesian approach has not matched the full generality of the parametric counterpart. This drawback mainly stems from the fact that a *prior distribution* (needed for a Bayesian analysis) should have two desirable properties: (i) the support of the prior distribution should be large, and (ii) *posterior distributions* (given sample observations from an unspecified distribution) should be manageable analytically; having both of them in full generality relates to the basic problem. Ferguson (1973) introduced the *Dirichlet process priors* and showed that they are broad enough in the sense of (i) and for them the posterior distributions are analytically manageable. These (mixtures of) Dirichlet process priors have also been successfully incorporated in Empirical Bayes estimation of distribution (and survival functions). One of the nice properties of such

Dirichlet process priors is that adaptations to censored samples can be made without much difficulty. Further, functionals of distribution functions have also been covered in Bayesian/empirical Bayesian fashion in a nonparametric setup. An excellent account of these developments is given in Ferguson, Phadia and Tiwari (1990). Although the Dirichlet processes priors have been adopted in the Bayesian analysis in nonparametrics, there remains a number of issues of both theoretical and applied interest. Exhaustive characterizations of priors for which in nonparametric Bayes analysis the posterior distributions are analytically manageable are of considerable interest. Moreover, to what extent such priors themselves have nonparametric flavors remains to be judged. From applications point of view, can a (mixture of) Dirichlet process priors always be justified in a practical context? Despite these, the Dirichlet process prior has certainly opened up a novel area in nonparametric analysis, and this is full of promise too.

Over the years, the simplicity of EDF procedures in the nonparametrics has given way to the robustness and (asymptotic) optimality of their ADF counterparts. There is, however, a fine line of demarcation between large sample paraemtric procedures and asymptotic nonparametric ones. In a parametric procedure, as the parameters appear as algebraic constants in a presummed functional form of a pdf, it is usually possible to reduce "bias" and estimate the nuisance parameters (efficiently) by some standard methodology. In a nonparametric case, however, these nuisance parameters are themselves functionals of the unknown d.f., and also the "bias" is a similar functional. Hence, there is a need to estimate such functionals in an (at least, asymptotically) efficient (and robust) manner. In this context, the usual resampling methods play a basic role. Although resampling plans are usually adapted to reduce (or estimate) the bias and to estimate the sampling variance (or mean square error (mse)) of suitable estimators, in a majority of the cases, these may as well be used to estimate the sampling distribution itself. Among these resampling plans, *jackknifing*, introduced by Quenouille (1956), is the precursor. Though it was primarily introduced for "bias reduction", the "pseudovariables" generated in this process (of jackknifing) provide a (strongly) consistent estimator of the mean square error of the estimator (and its jackknifed version). As we have discussed before, often, a smooth statistical functional T_n $(= \theta(F_n))$ can be expressed in terms of a linear functional and a residual one, where the latter component is the main source of bias (but usually adds very little to the mse). By jackknifing, this bias term is reduced [up to the order $o(n^{-1})$] while the structure of the linear term remains in tact. It may be recalled that both the *infinitesmal jackknifing* and *delta methods* are based on such a decomposition, and it is not surprising that they yield identical results on the estimated mse [see, for example, Efron (1982)]. A very smart idea in resampling plans is due to Efron (1979), and is known as the *bootstrap method.* The past decade has witness a phenomenal growth of the literature on bootstrap (and jackknife) methods, and every year some new dimensions are being added to the potential use of such resampling plans. There are certain situations where a bootstrap method may work out better than the others, although there are other situations where jackknifing may be more appealing. Albeit the emphasis on bias reduction and estimation of mse in a resampling plan, there is a greater need to have a clsoer look at them from the point of view of robustness. Efron (1982) has a very comprehensive account of the traditional aspects of some of these resampling plans where ample room has been left for further studies based on robustness considerations. As a matter of fact, both jackknifing and bootstrapping attempt to "recreate" the sampling distribution of the original statistic either by simple random sampling without replacement or with replacement from the given sample, and hence, if the original statistic is not so robust, its jackknifed/bootstrapped version may not be that either. Hence, to percolate robustness through such resampling methods, it may be desirable to start with either some robust statistics or to consider alternative resampling procedures which permit greater scope for robustness. In nonparametrics, either of these approaches sounds suitable. Functional jackknifing [viz., Sen (1988a)] dombines some advantages of both. For a general review of such functional approaches in resampling plans, we may refer to Sen (1988b), and thereby omit the details. However, in passing, we may note that such functional approaches are being increasingly adopted for nonparametric regression models. The very definition of a regression function by means of conditional expectation calls for stringent regularity conditions and raises doubts about its robustness properties. A conditional quantile function may appear to be a nice alternative - albeit it may not lead to Hadamard-differentiable functionals. There are, however, other smooth functionals (viz., trimmed/Winsorized mean or some other L-functionals) which are enough "smooth" to be adaptable in the formulation of a nonparametric regression function and are manageable from statistical analysis point of view too. In this context too, functional jackknifing may be quite useful too. We are looking forward to more fruitful research in this general area. Resampling schemes have also found their way to another area in nonparametrics: *smoothing techniques and cross validation.* In the context of nonparametric *curve estimation* [viz., density function or functionals of the density functions or its derivatives], the choice of a *bandwidth* (e.g., in a *kernel* method) hinges on some delicate points. If this bandwidth is too small, irregular fluctuations may mar the curve, while a too big width, albeit being smooth, may smoothout the curve and throw away a lot of information. An optimal choice of such a bandwidth may not only depend on the kernel chosen, but also on other quantities which depend on the unknown density function (and/or its derivatives). Adaptive procedures and resampling methods are useful tools in seeking information on these parameters so that some nearly (or asymptotically) optimal choice of bandwidth may be made. There has been a tremendous amount of research in this area during the past twelve years, and it is beyond the scope of this review to go into these details. A purposive sampling (of size 1) has led me to the reference of Müller (1984), albeit there are more developments during the past six years.

Textbooks and monographs are the coveyer belts for transportation of basic research to the domain of basic education and training, particularly at the graduate level. Impacts of the some of the books in nonparametrics in this respect have already been mentioned earlier. In all fairness to other authors, I would like to mention some of the other notable books and monographs; the list is by no means exhaustive and any omission is not based on any specific factor or bias in my own attitude towards nonparametrics. Rather, I have tried to include the following with a view to incorporating the philosophical undercurrents in research and teaching of nonparametrics. A traditional slant in teaching of statistical distributions includes only a few distinguished members: binomial, Poisson, beta, gamma and, of course, the normal distribution. Although there ae some justifications for the adaptation of normal, binomial or Poisson distribution in practice, there may not be any eminent reason to advocate their use in practice without any reservation. Until about the Fifties, the nonparametrics were the 'untouchables', specially in Europe; they were labeled 'inefficient', and often, 'inexact'. At the undergraduate level statistic courses, they were the 'refugees' in some remote corners, and even at the graduate level, they did not have any strong hold. Why would you travel through the territory of normal and other parametric distributions to come down to the simple sign test or the Wilcoxon-Mann-Whitney test? Indeed, a very bold step in this renaissance is due to Noether (1967). It is certainly enjoyable to see the direct avenue for moving into the parametrics from the nonparametrics, and appreciate the relative merits and demerits of both the approaches. Nonparametrics were indeed more adaptable in various walks of applications, and no wonder, Conover (1971) and Gibbons (1971) both collected a lot of appreciation from the applied audience. Hollander and Wolfe (1973) have a broader coverage. Hájek (1969) is a nice treatment of the methodology at the elementary level - quite different from Neother (1967). A little more towards the professional statisticians' usage, Hollander and Wolf (1973) has acquired a good reputation. At an intermediate level, Randles and Wolfe (1979) is a good text to try for the first course in nonparametrics at the graduate level, while Hettmansperger (1984) has a more detailed (albeit at an intermediate level) covarge of rank based nonparametrics. Shorack and Wellner (1986) have viewed exclusively from the galllery of (weighted) empirical processes. Compared to most of these texts and monographs, the coverage of Krishnaiah and Sen (1984) is understandably greater. However, that is in the context of a "handbook of statistics" and the purpose was somewhat different too! Nevertheless, it reflects the state of art with the developments in nonparametrics (methodology and applications), and I share a very optimistic view of the scope for future research in this fertile area.

Selected References

It is indeed a monumental task to compile a (reasonably) complete bibliography of nonparametrics. Even nearly thirty years ago, Savage (1962) encountered a tremendous difficulty in his attempt to provide an up-to-date bibliography. During the past thirty years, the volume of work has quadrupled, and the task is even hopelessly miserable! I confess that I don't have the stamina to undertake any such project, and I would rather welcome volunteers who have high energy levels. The references cited below are, mostly, the ones appearing in the text. In that way, the representativeness of this (purposive) sample may well be questioned. However, it provides a good idea of the spectrum and supplements the discussions made.

Aalen, O. O. (1978). Nonparametric inference for a family of counting processes. Ann. Statist. 6, 701-726.

Arbuthnott, J. (1710). An argument for devine providence taken from the constant regularity observed in the birth of both sexes. Philos. Trans. 27, 186-190.

Armitage, P. (1975). Sequential Medical Trials. John Wiley, New York.

Armitage, P., McPherson, C. K. and Rowe, B. C. (1969). Repeated significance tests on accumulating data. J. Roy. Statist. Soc. A132, 235-244.

Bahadur, R. R. (1960). Stochastic comparison of tests. Ann. Math. Statist. 31, 276-295.

Bell, C. B. and Sen, P. K. (1984). Randomization Procedures. In Handbook of Statistics, Vol 4: Nonparametric Methods (eds.: P. R. Krishnaiah and P. K. Sen), North Holland, Amsterdam, pp. 1-30.

Berk, R. H. (1966). Limiting behavior of posterior distributions when the model is incorrect. Ann. Math. Statist. 37, 51-58.

Bickel, P. J. (1965). On some asymptotically nonparametric competitors of Hotelling's T^2. Ann. Math. Statist. 36, 160-173.

Chatterjce, S. K. (1966). A bivariate sign-test for location. Ann. Math. Statist. 37, 1771-1782.

Chatterjee, S. K. and Sen, P. K. (1964). Nonparametric tests for the bivariate two-sample location problem. Calcutta Statist. Assoc. Bull. 13, 18-58.

Chatterjee, S. K. and Sen, P. K. (1973). Nonparametric testing under progressive censoring. Calcutta Statist. Assoc. Bull. 22, 13-50.

Chernoff, H. and Savage, I. R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. Ann. Math. Statist. 29, 972-994.

Conover, W. J. (1971). Practical Nonparametric Statistics. John Wiley, New York.

Cox, D. R. (1972). Regression models and life tables. J. Roy. Statist. Soc. B. 34, 184-220.

Durbin, J. (1973). Distribution Theory for Tests Based on the Sample Distribution Function. SIAM Publication; Philadelphia.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. Ann. Statist. 7, 1-25.

Efron, B. (1982). The Jackknife, Bootstrap and Other Resampling Methods. SIAM Publications, Philadelphia.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. Ann. Statist. 1, 209-230.

Ferguson, T. S., Phadia, E. and Tiwari, R. C. (1990). Bayesian nonparametric inference. In Issues in Statistical Inference: Essays in Honor of Debabrata Basu (eds.: M. Ghosh and P. K. Pathak).

Inst. Math. Statist. Lec. Notes Monograph. Ser. (in press).

Fernholz, L. T. (1983). Von Mises Calculus for Statistical Functionals. Springer Lec. Notes in Statistics 19, New York.

Fraser, D. A. S. (1957). Nonparametric Methods in Statistics. John Wiley, New York.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implict in the analysis of variance. J. Amer. Statist. Assoc. 32, 675-701.

Gangopadhyay, A. K. and Sen, P. K. (1990). Bootstrap confidence intervals for conditional quantile functions. Sankhya, Ser. A 52, in press.

Geertsema, J. C. (1970). Sequential confidence intervals based on rank tests. Ann. Math. Statist. 41, 1016-1026.

Ghosh, M. (1973). On a class of asymptotically optimal nonparametric tests for grouped data, I, II. Ann. Inst. Statist. Math. 25, 91-108, 109-122.

Ghosh, M. and Sen, P. K. (1971). On a class of rank order tests for regression with partially informed stochastic predictors. Ann. Math. Statist. 42, 650-661.

Ghosh, M. N. (1954). Asymptotic distribution of serial statistics and applications to problems of nonparametric tests of hypotheses. Ann. Math. Statist. 25, 218-251.

Gibbons, J. D. (1971). Nonparametric Statistical Inference. McGraw-Hill, New York.

Guilford, J. P. (1956). Fundamental Statistics in Psychology and Education. McGraw-Hill, New York.

Gutenbrunner, G. and Jurečková, J. (1990). Regression rank scores and regression quantiles. Ann. Statist. 18, inpress.

Hájek, J. (1961). Some extensions of the Wald-Wolfowitz-Noether theorem. Ann. Math. Statist. 32, 506-523.

Hájek, J. (1962). Asymptotically most powerful rank order tests. Ann. Math. Statist. 33, 1124-1147.

Hájek, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives. Ann. Math. Statist. 39, 325-346.

Hájek, J. (1969). Nonparametric Statistics. Holden-Day, San Francisco.

Hájek, J. and Šidák, Z. (1967). Theory of Rank Tests. Academic Press, New York.

Hettmansperger, T. P. (1984). Statistical Inference Based On Ranks. John Wiley, New York.

Hieler, S. and Willers, R. (1988). Asymptotic normality of R-estimates in the linear model. Statistics 19, 173-184.

Hodges, J. L. Jr. and Lehmann, E. L. (1962). Rank methods for combination of independent experiments in analysis of variance. Ann. Math. Statist. 33, 482-497.

Hodges, J. L. Jr. and Lehmann, E. L. (1963). Estimates of location based on rank tests. Ann. Math. Statist. 34, 598-611.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. Ann. Math.

Statist. 19, 293-325.

Hoeffding, W. (1952). The large sample power of tests based on permutations of observations. Ann. Math. Statist. 23, 169-192.

Hoeffding, W. (1965). Asymptotically optimal tests for multinomial distributions. Ann. Math. Statist. 36, 369-400.

Hogg, R. V. (1974). Adaptive robust procedures: partial review and some suggestions for further applications and theory. J. Amer. Statist. Assoc. 69, 909=923.

Hollander, M. and Proschan, F. (1984). Nonparametric concepts and methods in reliability. In Handbook of Statistics, Volume 4: Nonparametric Methods (eds., P. R. Krishnaiah and P. K. Sen), pp. 613-655.

Hollander, M. and Wolfe, D. A. (1973). Nonparametric Statistical Methods, John Wiley, New York.

Huber, P. J. (1981). Robust Statistics, John Wiley, New York.

Hušková, M. and Sen, P. K. (1985). On sequentially adaptive asymptotically efficient rank statistics. Sequen. Anal. 4, 125-151.

Jurečková, J. (1971). Nonparametric estimates of regression coefficients. Ann. Math. Statist. 42, 1328-1338.

Jurečková, J. (1977). Asymptotic relations of M-estimates and R-estimates in linear regression. Ann. Statist. 5, 464-472.

Kiefer, J. and Wolfowitz, J. (1952). Stochastic estimation of the maximum of a regression function. Ann. Math. Statist. 23, 462-466.

Koenker, R. (1982). Robust methods in econometrics. Economet. Rev. 1, 213-290.

Krishnaiah, P. R. and Sen, P. K. (eds.) (1984). Handbook of Statistics, Volume 4: Nonparametric Methods. North Holland, Amsterdam.

Lehmann, E. L. (1951). Consistency and unbiasedness of certain nonparametric tests. Ann. Math. Statist. 22, 165-179.

Lehmann, E. L. (1959). Testing Statistical Hypotheses. John Wiley, New York.

Lehmann, E. L. (1975). Nonparametrics: Statistical Methods Based on Ranks. Holden-Day, San Francisco.

Lehmann, E. L. and Scheffé, H. (1950). Completeness, similar regions and unbiased estiamtion, I. Sankhyā. 10, 305-340.

Lehmann, E. L. and Scheffé, H. (1955). Completeness, similar regions and unbiased estimation, II. Sankhyā. 15, 219-236.

Lehmann, E. L. and Stein, C. (1948). Most powerful tests of composite hypotheses. Ann. Math. Statist. 19, 495-516.

Müller, H. G. (1984). Smooth optimum kernel estimators of densities regression curves and modes. Ann. Statist. 12, 766-774.

Müller-Funk, U. (1984). Sequential Nonparametric Tests. In Handbook of Statistics, Vol. 4: Nonparametric Methods (eds., P. R. Krishnaiah and P. K. Sen). North Holland, Amsterdam, 657-697.

Murphy, S. A. and Sen, P. K. (1990). Time-dependent coefficients in a Cox-type regression model (preprint).

Noether, G. (1967). Elements of Nonparametrics Statistics. John Wiley, New York.

Pitman, E. J. G. (1937). Significance tests which may be applied to samples from any populations. Suppl. J. Roy. Statist. Soc. 4, 119-130, 225-232.

Puri, M. L. and Sen, P. K. (1971). Nonparametric Methods in Multivariate Analysis. John Wiley, New York.

Puri, M. L. and Sen, P. K. (1985). Nonparametric Methods in General Linear Models. John Wiley, New York.

Pyke, R. and Shorack, G. R. (1968). Weak convergence of a two-sample empirical process and a new approach to Chernoff-Savage theorems. Ann. Math. Statist. 39, 755-771.

Quenouille, M. H. (1956). Notes on bias in estimation. Biometrika 43, 353-360.

Randles, R. H. and Wolfe, D. A. (1979). Introduction to the Theory of Nonparametric Statistics. John Wiley, New York.

Robbins, H. and Monroe, S. (1951). A stochastic approximation method. Ann. Math. Statist. 22, 400-407.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). Order Restricted Statistical Inference. John Wiley, New York.

Roy, S. N. (1953). On a heuristic method of test construction and its use in multivariate analysis. Ann. Math. Statist. 24, 220-238.

Sarhan, A. E. and Greenberg, B. G., eds. (1962). Contributions to Order Statistics. John Wiley, New York.

Savage, I. R. (1962). Bibliography of Nonparametric Statistics, Harvard Univ. Press. Cambridge, Mass.

Savage, I. R. and Sethuraman, J. (1972). Asymptotic distribution of the log likelihood ratio based on rank in the two-sample problem. Proc. 6th Berkeley Symp. Math. Statist. Prob. 1, 437-458.

Scheffé, H. (1943). Statistical Inference in the nonparametric case. Ann. Math. Statist. 14, 305-332.

Sen, P. K. (1963). On the estimation of relative potency in dilution (-direct) assays by distribution-free methods. Biometrics 19, 532-552.

Sen, P. K. (1967). Asymptotically most powerful rank order tests for grouped data. Ann. Math. Statist. 38, 1229-1239.

Sen, P. K. (1968). Robustness of some nonparametric procedures in linear models. Ann. Math. Statist. 39, 1913-1922.

Sen, P. K. (1974). Weak convergence of generalized U-statistics. A
nn. Probab. 2, 90-102.

Sen, P. K. (1980). On nonparametric sequential point estimation of location based on general rank order statistics. Sankhyá, Ser. A. 42, 201-219.

Sen, P. K. (1981). Sequential Nonparametrics: Invariance Principles and Statistical Inference. John Wiley, New York.

Sen, P. K. (1983). On permutational central limit theorems for general multivariate linear rank statistics. Sankhyá, Ser A. 45, 141-149.

Sen, P. K. (1984). Nonparametric procedures for some miscellaneous problems. In Handbook of Statistics, Vol 4: Nonparametric Methods (eds., P. R. Krishnaiah and P. K. Sen). North Holland, amsterdam, pp. 699-740.

Sen, P. K. (1985). Theory and Applications of Sequential Nonparametrics. SIAM publ., Philadelphia.

Sen, P. K. (1988a). Functional jackknifing: Rationality and general asymptotics. Ann. Statist. 16, 450-469.

Sen, P. K. (1988b). Functional approaches in resampling plans: A review of some recent developments. Sankhyá, Ser. A. 50, 394-435.

Sen, P. K. (1989). Affluence and poverty indexes. In Encyclopedia of Statistical Sciences, (eds., S. Kotz and N. L. Johnson). Suppl. pp. 1-5.

Sen, P. K. and Ghosh, M. (1971). On bounded length sequential confidence intervals based on one-sample rank order statistics. Ann. Math. Statist. 42, 189-203.

Sen, P. K. and Ghosh, M. (1974). On sequential rank tests for location. Ann. Statist. 2, 540-552.

Shorack, G. R. and Wellner, J. A. (1986). Empirical Processes with Applications to Statistics. John Wiley, New York.

Siegel, S. (1956). Nonparametric Statistics for the Behavioral Sciences. McGraw-Hill, New York.

van Zwet, W. R. (1984). A Berry-Esseen bound for symmetric statistics. Zeit. Wahrsch. Verw. Geb. 66, 425-440.

von Mises, R. V. (1947). On the asymptotic distribution of differentiable statistical functions. Ann. Math. Statist. 18, 309-348.

Vorličkova, D. (1970). Asymptotic properties of rank tests under discrete distributions. Zeit. Wahrsch. Verw. Geb. 14, 275-289.

Wald, A. (1947). Sequential Analysis. John Wiley, New York.

Wald, A. and Wolfowitz, J. (1944). Statistical tests based on the permutations of observations, Ann. Math. Statist. 15, 358-372.

Watson, G. S. (1983). Statistics on the Sphere. John Wiley, New York.