DESIGN AND ANALYSIS OF EXPERIMENTS:
NONPARAMETRIC METHODS WITH APPLICATIONS
TO CLINICAL TRIALS

by
Pranab Kumar Sen
Department of Biostatistics
University of North Carolina

DESIGN AND ANALYSIS OF EXPERIMENTS: NONPARAMETRIC METHODS WITH APPLICATIONS TO CLINICAL TRIALS[†]

Pranab Kumar Sen

1 Introduction

In a variety of experimental studies ranging from conventional agricultural, biological and industrial experiments to modern biomedical investigations and (multicenter) clinical trials, a sound statistical planning (design) of the experiment constitutes the first and foremost task.

In a conventional setup it is generally assumed that the response distribution is continuous, and further that it can be taken as normal whose mean (location parameter) depends *linearly* on the model based parameters, while the variance is a constant under varied experimental setups. The *linearity of the model* (i.e., the additivity of the effects), *homoscedasticity* and *normality* of the errors are the three basic regularity assumptions underlying the traditional parametric statistical analysis of experimental outcomes. In *biological assays*, biomedical and clinical experiments, the response variable is typically nonnegative and has a positively skewed distribution. For this reason, often, a logarithmic or square (or cubic) root transformation (belonging to the family of *Box-Cox* type transformations) is advocated to induce greater degree of symmetry of the distribution of the transformed variate (for which the normality assumption may be more likely). However, such a *nonlinear transformation* may generally affect the linearity of the regression model as well as the homoscedasticity condition. Thus, in general, there may not be any guarantee that a Box-Cox type transformation may simultaneously improve the normality of the transformed errors, their homoscedasticity and the additivity of the model. To the contrary, generally, one of the two postulations : additivity of the model effects and normality of the

error distribution, has to be compromised at the price for the other. Therefore, there is a genuine need to looking into the composite picture of linearity of regression, homoscedasticity and normality of errors from a broader perspective so as to allow plausible departures from each of these postulations. This, in turn, may dictate the need for more appropriate designs, and consequently, broader statistical analysis schemes. *Nonparametric* and *robust* statistical procedures have their genesis in this robustness and validity complex. In passing, we may remark that in biological assays, socio-economic and psychometric studies, often, a response may be *quantal* (i.e., all or nothing), *qualitative* (i.e., categorical) or *discrete* (viz., relating to count data), while in a majority of practical applications, one may also encounter *grouped data* typically arising when observations are recorded in suitable *class intervals*. Design and statistical analysis of such nonstandard experiments may call for a somewhat different statistical approach, and we may briefly mention some of the basic differences in the setups of such nonstandard statistical analyses contrasted with standard parametric formulations.

Pedagogically, developments in nonparametrics started in the hypotheses testing sector, where under suitable *hypotheses of invariance*, the sample observations have a joint distribution which remains invariant under appropriate *groups of transformations* which map the sample space onto itself. This characterization of the null hypothesis generates *exactly distribution-free (EDF)* tests for some simple hypotheses testing problems. However, in a more complex situation, such as arising in multivariate problems or in composite hypotheses testing problems, this EDF propoerty may not be generally tenable. In such a case, often, it may be possible to have some conditionally distribution-free tests, although this prescrtiption may not always work out well. Moreover, even in simple hypotheses testing problems, the distribution theory of EDF tests may become prohibitively complicated for any algebraic manipulations as the sample size increases. This unpleasant feature is shared by nonparametric (point as well as interval) estimators which are based on such EDF statistics. In the later phase of developments on nonparametrics [viz., Sen (1991) for a review], *asymptotic methods* have invaded all walks of this field. Not only such asymptotics provide simpler distribution theory for large sample sizes, but also clearer motivations for adoptations of nonparametrics instead of their parametric counterparts. Indeed, at the current state of developments on nonparametrics, such asymptotics play a vital role in all aspects.

A significant amount of research work on design and analysis of experiments relates to the so called bio-assays. Statistical methodology for bio-assays, even in a parametric setup, has a somewhat different perspective, and nonparametrics as developed in the past three decades [viz., Sen (1963, 1971b,1972)] are very relevant too. Bio-assays are the precursors of modern *clinical trials*, and during the past twenty five years, statistical perspectives in clinical trials and medical investigations have received due attention from clinical as well as statistical professionals (and *Regulatory Agencies* too). This has paved the way for development of some *clinical designs* with greater emphasis on *hazard models* and *relative risk* formulation; nonparametrics play a vital role in this context too. Perhaps, our treatise of this field would remain a bit incomplete without some discussion of such designs and their analysis schemes in the light

of such recently developed nonparametrics. Statistical planning (design) and inference (analysis) remain pertinent to two other related areas, namely, *(i) epidemiological investigations* and *(ii) environmental studies,* and more work remains to be accomplished before they can be systematically reviewed.

2 One-way layout nonparametrics

Let us consider the following linear model related to completely randomized designs :

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}; \quad \mathbf{e} = (e_1, \dots, e_n)', \tag{2.1}$$

where $\mathbf{Y}$ is the n-vector of independent observations (response variables), $\mathbf{X}$ is an $n \times p$ matrix of known constants (design variables), $\beta = (\beta_1, \dots, \beta_p)'$ is a p-vector of unknown (fixed-effects) parameters, and the error variables e_i are assumed to be independent and identically distributed random variables (i.i.d.r.v.) with a continuous univariate distribution function (d.f.) F. In a parametric setup, for testing plausible hypotheses on β (or to construct suitable confidence sets), F is assumed to be normal with 0 mean and a finite, positive variance σ^2 (unknown), but in nonparametrics this normality or even the finiteness of the error variance is not that crucial. Two basic statistical issues arising in this context are the following: (i) To choose the design matrix $\mathbf{X}$ in such a way that for a given n, the number of observations, the *information* on β acquirable from the experiment is maximized (the *optimal design* problem), and (ii) for a chosen design, to draw statistical conclusions (on β) in an optimal manner (the *optimal statistical inference* problem). For the linear model in (2.1), when F is normal, the information on β contained in the experiment is given by $\sigma^{-2}(\mathbf{X}'\mathbf{X})$, where $\mathbf{X}'\mathbf{X}$ is a positive semi-definite (p.s.d.) matrix. Therefore maximization of information on β relates to an optimal choice of the design matrix $\mathbf{X}$, such that $\mathbf{X}'\mathbf{X}$ is maximizied in a meaningful manner. Kiefer's *universal optimality of designs* is a landmark in this context. Two recent monographs by Shah and Sinha (1989) and Pukelsheim (1993) are excellent sources for detailed coverage of such optimal designs. When the normality of the error components is not taken for granted, the information criterion is taken as $I(f)\mathbf{X}'\mathbf{X}$ where $I(f)$ is the *Fisher information* for the location parameter for the density f, when the latter belongs to the location-scale family. Thus, if $I(f)$ is substituted for σ^{-2}, the situation remains comparable, so that for a given density having a finite Fisher information $I(f)$, maximization of the information (on β) through a choice of the design matrix remains isomorphic to the normal theory case.

The second problem has a somewhat different perspective. Optimal statistical inference may depend very much on the underlying density. The main point in this context is that for a possibly nonnormal d.f., having a finite variance σ^2, the dispersion matrix of the maximum likelihood estimator of β is given by $I^{-1}(f)(\mathbf{X}'\mathbf{X})^{-1}$, and the normal theory estimator has the dispersion matrix $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$, where by the classical *Cramér-Rao information-inequality,*

$$\sigma^2 \geq \{I^{-1}(f)\}, \tag{2.2}$$

so that the normal theory or linear inference procedures generally entail some loss of efficiency. This loss is measured by the difference of $I(f)$ and σ^{-2}, and for *heavy-tailed distributions,* this loss can be quite large. In the extreme case of infinite variance (viz., the Cauchy case), linear/normal theory inference procedures are inconsistent and totally inefficient too. On the top of that if in (2.1) the very additivity of the effects (i.e, linearity of the model) does not hold, the entire picture may be drastically different. Robust and nonparametric methods have mainly been considered to retain general validity as well as good efficiency properties under such plausible departures from the model based assumptions. Although such nonparametric procedures were proposed originally for hypotheses testing problems, the dual estimation problems have also received adequate attention, and we summarize them as follows.

In a completely randomized design, through a transformation on the design matrix $\mathbf{X} \to \mathbf{X}^* = \mathbf{XA}$, for a suitable p.d. $\mathbf{A}$ and a conjugate transformation on $\beta \to \beta^* = \mathbf{A}^{-1}\beta$, (2.1) can be *canonically* reduced to a multi-sample location model. In this canonical representation, the *homogeneity* of the location parameters constitute the basic null hypothesis, while estimable parameters relate to such location parameters as well as *contrasts* among them. Motivated by this canonical reduction, we partition $\mathbf{Y}$ into c (≥ 2) subvectors $\mathbf{Y}_k$, $k = 1, \ldots, c$, where $\mathbf{Y}_k$ is an n_k-vector $(Y_{k1}, \ldots, Y_{kn_k})'$ such that the Y_{ki} are i.i.d.r.v.'s with a d.f. $F_k(x) = F(x - \theta_k)$, for $k = 1, \ldots, c$. In this formulation, the d.f. F is assumed to be continuous, and it need not have a finite variance or first moment, and also the symmetry assumption is not crucial for testing the null hypothesis of equality of the θ_k. If, on the other hand, we want to estimate or test for the individual θ, then, the symmetry of the d.f. F is needed, and this will be discussed later on. Keeping this estimability and testability in mind, we may rewrite (2.1) in the following form:

$$Y_i = \theta + \beta' \mathbf{t}_i + e_i, i = 1, \ldots, n, \tag{2.3}$$

where the $\mathbf{t}_i$ are vectors of (known) regression constants, not all equal, θ is the *intercept* parameter, and $\beta = (\beta_1, \ldots, \beta_q)'$ is a vector of unknown regression parameters where q is a positive integer and $n \geq q + 1$; the e_i are i.i.d.r.v.'s having the d.f. F. In this setup, the null and alternative hypotheses of interest are

$$H_0 : \beta = \mathbf{0} \quad vs. \quad H_1 : \beta \neq \mathbf{0}. \tag{2.4}$$

Note that for the canonical multisample model, (2.4) relates to the *homogeneity* of the location parameters, and we may also want to estimate the *paired differences* $\theta_k - \theta_q, k \neq q = 1, \ldots, c$ in a robust manner. Note that under the null hypothesis in (2.4), the Y_i are i.i.d.r.v.'s with a location parameter θ, so that their joint distribution remains invariant under any permutation of the Y_i among themselves. For this reason, the null hypothesis in (2.4) is termed the *hypothesis of randomness or permutation-invariance.* In the literature, H_1 in (2.4) is termed the *regression alternative.* For such alternatives, tests for (2.4) do not require the symmetry of the d.f. F, and the same regularity assumptions pertain to the estimation of β. Further, it is possible to draw statistical inference on θ under an additional assumption that F is a symmetric d.f.

4

Let R_{ni} be the rank of Y_i among $Y_1, \ldots, Y_n$, for $i = 1, \ldots, n$ and $n \geq 1$. Also, let $a_n(1), \ldots, a_n(n)$ be a set of *scores* which depend on the sample size n and some chosen *score generating* function. Of particular interest are the following special scores: (i) *Wilcoxon scores*: $a_n(i) = i/(n+1)$, $i = 1, \ldots, n$; (ii) *Normal scores*: $a_n(i) = E(Z_{n:i})$, $i = 1, \ldots, n$, where $Z_{n:1} \leq \cdots \leq Z_{n:n}$ are the order statistics of a sample of size n drawn from a standard normal distribution; and (iii) *Log-rank scores*: $a_n(i) = \sum_{j=0}^{i-1}(n-j)^{-1} - 1$, $i = 1, \ldots, n$. With these notations, we define (a vector of) *linear rank statistics* by letting

$$\mathbf{L}_n = \sum_{i=1}^{n}(\mathbf{t}_i - \bar{\mathbf{t}}_n)a_n(R_{ni}); \tag{2.5}$$

where the $\mathbf{t}_i$ are defined as in (2.3) and $\bar{\mathbf{t}}_n = n^{-1}\sum_{i=1}^{n}\mathbf{t}_i$. Also, we define

$$\bar{a}_n = n^{-1}\sum_{i=1}^{n}a_n(i), \ A_n^2 = (n-1)^{-1}\sum_{i=1}^{n}[a_n(i) - \bar{a}_n]^2, \tag{2.6}$$

and

$$\mathbf{Q}_n = \sum_{i=1}^{n}(\mathbf{t}_i - \bar{\mathbf{t}}_n)(\mathbf{t}_i - \bar{\mathbf{t}}_n)'. \tag{2.7}$$

As has been noted earlier, under H_0 in (2.4), the Y_i are i.i.d., and hence, the vector $\mathbf{R}_n = (R_{n1}, \ldots, R_{nn})'$ takes on each permutation of $(1, \ldots, n)$ with the common probability $(n!)^{-1}$. If we denote this discrete uniform probability measure by $\mathcal{P}_n$, we obtain immediately that

$$E_{\mathcal{P}_n}(\mathbf{L}_n) = \mathbf{0} \quad \text{and} \quad E_{\mathcal{P}_n}(\mathbf{L}_n\mathbf{L}_n') = A_n^2 \cdot \mathbf{Q}_n. \tag{2.8}$$

A typical rank order test statistic is quadratic form in $\mathbf{L}_n$, and is given by

$$\mathcal{L}_n = A_n^{-2}[\mathbf{L}_n'\mathbf{Q}_n^{-}\mathbf{L}_n], \tag{2.9}$$

where $\mathbf{Q}_n^{-}$, a *generalized inververse* of $\mathbf{Q}_n$, is defined by $\mathbf{Q}_n^{-}\mathbf{Q}_n\mathbf{Q}_n^{-} = \mathbf{Q}_n^{-}$. For small values of n, the exact null hypothesis distribution of $\mathcal{L}_n$ can be obtained by direct enumeration of all possible $(n!)$ equally likely realisations of $\mathbf{R}_n$. Although this enumeration process becomes prohibitively laborious as the sample size becomes large, there are various *permutational central limit theorems* by which one may easily approximate this permutation distribution by a (central) *chi squared* distribution with q *degrees of freedom* (DF) when $\mathbf{Q}_n$ is of full rank (q). The right hand tail of the null distribution of $\mathcal{L}_n$ can therefore be used to find out a *critical level* $l_{n\alpha}$ corresponding to a given *level of significance* $\alpha : 0 \leq \alpha \leq 1$), such that if $\chi_{q,\alpha}^2$ stands for the upper α-*quantile* of the central chi squared d.f. with q DF, then

$$l_{n\alpha} \to \chi_{q,\alpha}^2, \quad \text{as} \quad n \to \infty. \tag{2.10}$$

Under quite general regularity conditions, when β is $\neq 0$, $\mathbf{Q}_n^{-1}\mathbf{L}_n$ converges to a nonnull vector, so that $n^{-1}\mathcal{L}_n$ converges to a positive number, while, by (2.10), $n^{-1}l_{n\alpha} \to 0$, as $n \to \infty$. Therefore, the test based on $\mathcal{L}_n$ is consistent against the entire class of alternatives for which the centering element of

$Q_n^{-1}L_n$ is different from 0. We may refer to Puri and Sen (1985) for some details. For this reason, *local alternatives* are chosen in such a way that under such alternatives, $\mathcal{L}_n$ has a nondegenerate asymptotic distribution which may then be incorporated in the study of various *asymptotic properties* of such rank order tests.

For a local (Pitman-type) shift alternative of the type

$$H_n : \beta = \beta_{(n)} = n^{-1/2}\lambda, \quad \text{for some fixed } \lambda \in R^q, \tag{2.11}$$

for large n, $\mathcal{L}_n$ has closely a *noncentral chi squared* d.f. with q DF and noncentrality parameter $\Delta = \gamma^2\ \lambda'Q^{-1}\lambda/A^2$, where γ is a functional of ϕ, F and is formally defined by (2.19), it is assumed that A_n^2 converges to a finite positive limit A^2 as n increases, and similarly, $n^{-1}Q_n$ converges to a p.d. Q. Also, it is assumed that the *generalized Noether-condition* holds for the t_i. For the normal scores rank test, we have asymptotic optimality when the underlying F is itself a normal d.f., and moreover, this particular test is asymptotically at least as efficient as the classical *variance ratio (ANOVA)* test for all F belonging to a general class F; the equality of their *asymptotic efficacies* holds only for a normal F which belongs to this class. For a *logistic* d.f., the test based on the Wilcoxon scores is asymptotically optimal, while for an *exponential* d.f., the log-rank test is asymptotically optimal. In general, if the d.f. F admits an absolutely continuous density f having a finite *Fisher information* $I(f)$, and if we let $a_n^*(k) = E\phi_f(U_{n:k})$, $k = 1,\ldots,n$, where $U_{n:1} \leq \cdots \leq U_{n:n}$ are the order statistics of a sample of size n from a *uniform [0,1]* d.f., and the score generating function $\phi_f(u)$ is given by

$$\phi_f(u) = -f'(F^{-1}(u))/f(F^{-1}(u)), \ u \in (0,1), \tag{2.12}$$

then the rank test based on the scores $a_n^*(k)$ is asymptotically optimal when the underlying d.f. is F; Hájek and Šidák (1967) is an excellent source for the theoretical motivation for such local optimality properties of linear rank statistics.. For the specific multisample model, as has been treated earlier, if we denote the average rank scores of the observations in the kth sample (with respect to the pooled sample observations) by $\bar{a}_{n,k}$, for $k = 1,\ldots,c$, then the rank order test styatistic in (2.9) simplifies to

$$\mathcal{L}_n = [\sum_{k=1}^{c} n_k(\bar{a}_{n,k} - \bar{a}_n)^2]/A_n^2, \tag{2.13}$$

where $\bar{a}_n$ is defined by (2.6). For this model, $q = c-1$. For the Wicoxon scores, (2.13) reduces to the well known *Kruskal-Wallis* test statistic, while the *Brown and Mood median* test statistic corresponds to the scores $a_n(k) = 0$ or 1, according as k is $\leq$ or $> (n+1)/2$. In this multisample setup, $\mathcal{L}_n$ actually tests for the homogeneity of the d.f.'s $F_1,\ldots,F_c$ against alternatives which are more general than the simple location or shift ones treated earlier. For example, if we let $\pi_{rs} = P\{Y_{r1} \geq Y_{s1}\}$, for $r \neq s = 1,\ldots,c$, then the Kruskal-Wallis test is consistent against the broader class of alternatives that the π_{rs} are not all equal to 1/2 , which relate to the so called *stochastically larger (smaller) alternatives*. Thus, the linearity of the model is not that crucial in this context. In the multisample model, when the scores are

monotone, $\mathcal{L}_n$ remains consistent against stochatically larger (smaller) class of alternatives, containing shift alternatives as a particular subclass of such a broader class of alternatives. This explains the robustness of such nonparametric tests. In this context, we may note that the ranks R_{ni} are *invariant* under any strictly monotone transformation $g(.)$ on the Y_i, so that if the $g(Y_i) = Y_i^*$ follow a linear model for some $g(\cdot)$, a rank statistic based on the Y_i and Y_i^* being the same will pertain to such a *generalized linear model* setup. This invariance eliminates the need for *Box-Cox* type transformations on the Y_i and thereby adds further to the robustness of such rank tests against plausible departures from the model based assumptions. The situation is a little less satisfactory for testing *subhypotheses* or for *multiple comparisons*, and we shall discuss them later on.

We present *R-estimators* which are based on such rank tests. Although for one and two-sample location problems, such estimates were considered by Hodges and Lehmann (1963) and Sen (1963), for general linear models, the developments took place a few years later. Adichie (1967), Sen (1968d) and Jurečková (1971), among others, considered the simple regression model. For the linear model in (2.3), let $Y_i(\mathbf{b}) = Y_i - \mathbf{b}'\mathbf{t}_i$, $i = 1, \ldots, n$, where $\mathbf{b} \in \mathcal{R}^q$, and let $R_{ni}(\mathbf{b})$ be the rank of $Y_i(\mathbf{b})$ among the $Y_r(\mathbf{b})$, $r = 1, \ldots, n, \mathbf{b} \in \mathcal{R}^q$. In (2.5), replacing the R_{ni} by $R_{ni}(\mathbf{b})$, we define the linear rank statistics $\mathbf{L}_n(\mathbf{b})$, $\mathbf{b} \in \mathcal{R}^q$. As in Jaeckel (1972), we introduce a *measure of rank dispersion* :

$$D_n(\mathbf{b}) = \sum_{i=1}^{N} \{a_n(R_{ni}(\mathbf{b})) - \bar{a}_n\} Y_i(\mathbf{b}), \ \mathbf{b} \in \mathcal{R}^q, \tag{2.14}$$

where we confine ourselves to monotone scores, so that $a_n(1) \leq \cdots \leq a_n(n)$, for every $n \geq 1$. An R-estimator of β is a solution to the minimization of $D_n(\mathbf{b})$ with respect to $\mathbf{b} \in \mathcal{R}^q$, so that we write

$$\hat{\beta}_n = \text{arg. min.} \{D_n(\mathbf{b}) : \mathbf{b} \in \mathcal{R}^q\}. \tag{2.15}$$

It can be shown [viz., Jurečková and Sen (1995, Ch. 6)] that $D_n(\mathbf{b})$ is a *nonnegative, continuous, piecewise linear* and *convex function* of $\mathbf{b} \in \mathcal{R}^q$. Note that $D_n(\mathbf{b})$ is differentiable in $\mathbf{b}$ almost everywhere and

$$(\partial/\partial \mathbf{b}) D_n(\mathbf{b})|_{\mathbf{b}^\circ} = -\mathbf{L}_n(\mathbf{b}^\circ), \tag{2.16}$$

whenever $\mathbf{b}^\circ$ is a point of differentiability of $D_n(\cdot)$. At any other point, one may work with the subgradient $\nabla D_n(\mathbf{b}^\circ)$. Thus, essentially, the task reduces to solving for the following *estimating equations* with respect to $\mathbf{b} \in \mathcal{R}^q$:

$$\mathbf{L}_n(\mathbf{b}) = 0, \tag{2.17}$$

where to eliminate multiple solutions, adopt some convention. These R-estimators are generally obtained by iterative procedures (as in the case of maximum likelihood estimators for a density not belonging to the exponential family), and, often, a one or two-step procedure starting with a *consistent* and *asymptotically normal (CAN)* initial estimator serves the purpose very well; for some theoretical developments along with an extended bibliuography, we refer to Jurečková and Sen (1995). It follows from their general

methodology that under essentially the same regularity conditions as pertaining to the hypothesis testing problem (treated before), the following *first order asymptotic distributional representation (FOADR)* result holds:

$$(\hat{\beta}_n - \beta) = \gamma^{-1} \sum_{i=1}^{n} \mathbf{d}_{ni}\phi(F(e_i)) + \mathbf{o}_p(n^{-1/2}), \tag{2.18}$$

where $\mathbf{d}_{ni} = \mathbf{Q}_n^{-1}(\mathbf{t}_i - \bar{\mathbf{t}}_n)$, for $i = 1, \ldots, n$, $\phi(\cdot)$ is the score generating function for the rank scores,

$$\gamma = \int_0^1 \phi(u)\phi_f(u)du, \tag{2.19}$$

and the Fisher information score generating function $\phi_f(\cdot)$ is defined by (2.12). Note that under a generalized *Noether Condition* on the $\mathbf{d}_{ni}$, the classical (multivariate) central limit theorem holds for the principal term on the right hand side of (2.18), so that on defining $\mathbf{Q}_n^*$ by $\mathbf{Q}_n^* \mathbf{Q}_n^{-1} \mathbf{Q}_n^* = \mathbf{I}$, we obtain from (2.18) and (2.19) that for large sample sizes,

$$\mathbf{Q}_n^*(\hat{\beta}_n - \beta) \xrightarrow{\mathcal{D}} N(0, \gamma^{-2}A^2\mathbf{I}), \tag{2.20}$$

where A^2 is the variance of the score function ϕ. For the normal theory model, the classical *maximum likelihood estimator (MLE)* agrees with the usual *least squares estimator (LSE)*, and for this (2.20) hold with $\gamma^{-2}A^2$ being replaced by σ^2, the error variance. As such, the *asymptotic relative efficiency (ARE)* of the R-estimator, based on the score function ϕ, with respect to the classical LSE is given by

$$e(R; LS) = \gamma^2\sigma^2/A^2, \tag{2.21}$$

which does not depend on the design matrix $\mathbf{Q}_n$. In particular, if we use the normal scores for the derived R-estimators, then (2.21) is bounded from below by 1, where the lower bound is attained only when the underlying distribution is normal. This explains the robustness as well as asymptotic efficiency of the normal scores R-estimators in such completely randomized designs. From robustness considerations, often, it may be better to use the Wilcoxon scores estimators. Although for this particular choice of the score generating function, (2.21) is not bounded from below by 1, it is quite close to 1 for near normal distributions and may be high for heavy tailed ones. If the error density $f(\cdot)$ is of known functional form, one may use the MLE for that pdf, and in that case, in (2.20), we need to replace $\gamma^{-2}A^2$ by $\{I(f)\}^{-1}$, where $I(f)$ is the Fisher information for location of the density f. Thus, in this case, the ARE is given by

$$e(R; ML) = \gamma^2/\{I(f)A^2\}, \tag{2.22}$$

which by the classical Cramér-Rao inequality is always bounded from above by 1. Nevertheless, it follows from the general results in Hušková and Sen (1985) that if the score generating function is chosen adaptively , then the corresponding *adaptive R*-estimator is asymptotically efficient in the sense that in (2.22) the ARE is equal to 1. The same conclusion holds for adaptive rank tests for β as well.

As has been mentioned earlier, the ranks $R_{ni}(\mathbf{b})$ are *translation-invariant* so that they provide no information on the intercept parameter θ. Thus, for testing any plausible null hypothesis on θ or to estimate the same parameter, linear rank statistics are not of much use. This problem has been eliminated to a greater extent by the use of *signed rank statistics*, which is typically defined as

$$S_n = \sum_{i=1}^{n} sign(Y_i) a_n(R_{ni}^+), \tag{2.23}$$

where the rank scores $a_n(k)$ are defined as in before and R_{ni}^+ is the rank of $|Y_i|$ among the $|Y_r|$, $r = 1, \ldots, n$. Under the null hypothesis of symmetry of the d.f. F about 0, the vector of the $|R_{ni}|$ and the vector of the $sign(Y_i)$ are stochastically independent, so that the set of 2^n equally likely *sign-inversions* generates the exact null distribution of S_n. This may also be used to derive the related R-estimator of θ. Such a test and estimator share all the properties of the corresponding test and estimator for the regression parameter. But, in the current context, there is a basic problem. The test for β based on $\mathcal{L}_n$, being translation-invariant, does not depend on the intercept parameter (which is taken as a nuisance one). On the other hand, for testing a null hypothesis on θ or estimating the same, the parameter β is treated as a nuisance one, and the signed ranks are not *regression-invariant*. Thus, the *exact distribution-freeness (EDF)* property may have to be sacrifice in favor of *asymptotically distribution-free (ADF)* ones. An exception is the case when one wants to test simulataneously for $\theta = 0$ and $\beta = 0$.

We denote a suitable R-estimator of β by $\hat{\beta}_n$, and incorporate the same to obtain the residuals:

$$\hat{Y}_{ni} = Y_i - \hat{\beta}_n' \mathbf{t}_i, \quad i = 1, \ldots, n. \tag{2.24}$$

For every real d, let $\hat{R}_{ni}^+(d)$ be the rank of $|\hat{Y}_{ni} - d|$ among the $|\hat{Y}_{nr} - d|$, $r = 1, \ldots, n$, for $i = 1, \ldots, n$. Also let

$$\hat{S}_n(d) = \sum_{i=1}^{n} sign(\hat{Y}_{ni} - d) a_n(\hat{R}_{ni}^+(d)), \quad d \in \mathcal{R}. \tag{2.25}$$

If the scores $a_n(k)$ are monotone (in k, for each n), then, it is easy to show that $\hat{S}_n(d)$ is monotone in $d \in \mathcal{R}$, and hence, we may equate $\hat{S}_n(d)$ to 0 (with respect to $d \in \mathcal{R}$) and the solution, say, $\hat{\theta}_n$, is then taken as a *tranlation-equivariant* estimator of θ. In the particular case of the sign statistic, $\hat{\theta}_n$ can be expressed as the median of the residuals $\hat{Y}_{ni}$, and for the case of the Wilcoxon signed-rank statistic, it is given by the median of the midranges of these residuals. In general, for other score functions, an iterative procedure is needed to solve for $\hat{\theta}_n$, and in such a case, one may as well start with the Wilcoxon scores estimator as the preliminary one, and in a few steps converge to the desired one. There is a basic difference between this model and the simple location model where β is null: In the latter case, the signed rank statistics based on the true value $d = \theta$ are EDF, while in this case they are only ADF. To verify that they are ADF, one convenient way is to appeal to some *asymptotic uniform linearity* results on general signed rank statistics (in the location and regression parameters), and such results have been

presented in a unified manner in Chapter 6 of Jurečková and Sen (1995), where pertinent references are also cited in detail.

Let us discuss briefly the subhypothesis testing problem for this simple disign. A prticular subhypothesis testing problem relates to the null hypothesis that $\theta = 0$ against $\theta \neq 0$, treating β as a nuisance parameter (vector). We have already observed that the basic hypothesis of sign- (or permutation) invariance does not hold when the above null hypothesis holds, and hence, EDF tests may not generally exist. However, ADF test can be considered by incorporating the residuals $\hat{Y}_{ni}$ instead of the Y_i in the formulation of suitable signed rank statistics. Such tests were termed *aligned rank tests* by Hodges and Lehmann (1962) who considered the simplest ANOVA model. Here alignment is made by substituting the estimates of the nuisance parameters as is also done in the classical normal theory linear models. A very similar picture holds for a plausible subhypothesis testing problem on the regression parameter vector. To pose such a problem in a simple manner, we partition the parameter vector β as

$$\beta' = (\beta_1', \beta_2'), \tag{2.26}$$

where β_j is a p_j-vector, $p_j \geq 1$, for $j = 1, 2$, and $p = p_1 + p_2$. Suppose now that we want to test for

$$H_0 : \beta_1 = 0 \quad \text{vs.} \quad H_1 : \beta_1 \neq 0, \quad \text{treating } \beta_2 \text{ as a nuisance parameter.} \tag{2.27}$$

Here also, under H_0 in (2.27), the hypothesis of permutational invariance may not be generally true, and hence, an EDF rank test may not generally exist. But, ADF rank tests based on aligned rank statistics can be constructed as follows. Note that if in (2.3), we partition the t_i as $(t_{i1}', t_{i2}')'$, involving p_1 and p_2 coordinates, then under H_0 in (2.27), we obtain that

$$Y_i = \theta + \beta_2' t_{i2} + e_i, \ i = 1, \ldots, n. \tag{2.28}$$

Based on the model in (2.28), we denote the R-estimator of β_2 by $\tilde{\beta}_{n2}$, and we form the residuals

$$\tilde{Y}_{ni} = Y_i - \tilde{\beta}_{n2}' t_{i2}, \quad \text{for } i = 1, \ldots, n. \tag{2.29}$$

As in after (2.4), we define the aligned ranks $\tilde{R}_{ni}$ wherein we replace the Y_i by the residuals $\tilde{Y}_{ni}$. The vector of linear rank statistics $\tilde{L}_n$ is then defined as in (2.5) with the ranks R_{ni} being replaced by $\tilde{R}_{ni}$. Also, we partition this p-vector as $(\tilde{L}_{n1}', \tilde{L}_{n2}')'$, and our test is then based on the first component of this aligned rank statistics vector. This is given by

$$\tilde{\mathcal{L}}_{n1} = A_n^{-2} \{ \tilde{L}_{n1}' Q_{n11.2}^{-1} \tilde{L}_{n1} \}, \tag{2.30}$$

where A_n^2 is defined as in (2.6), and defining Q_n as in (2.7) and partitioning it into four submatrices , we have

$$Q_{n11.2} = Q_{n11} - Q_{n12} Q_{n22}^{-1} Q_{n21}. \tag{2.31}$$

It follows from the general results in Sen and Puri (1977), further streamlined and discussed in detail in Section 7.3 of Puri and Sen (1985) that under the null hypothesis in (2.27), $\tilde{\mathcal{L}}_{n1}$ has asymptotically

chi-squared distribution with p_1 DF, so that an ADF test for H_0 in (2.27) can be based on the critical level given by $\chi^2_{p_1,\alpha}$, the upper α-percentile of this distribuion. For local alternatives, the noncentral distribution theory runs parallel to the case of the null hypothesis of $\beta = \mathbf{0}$, with the DF p being replaced by p_1 and an appropriate change in the noncentrality parameter as well. The regularity conditions governing these asymptotic distributional results have been unified and relaxed to a certain extent in Chapter 6 of Jurečková and Sen (1995).

Another important area where nonparametrics have played a vital role in such completely randomized designs is the so called *mixed effects* models. In this setup, we extend (2.3) as follows. Let $Y_1, \ldots, Y_n$ be independent random variables, such that associated with the Y_i there are (i) given design (nonstochastic) (q-)vectors $\mathbf{t}_i$ and (ii) observable stochastic *concomitant* (p-)vectors $\mathbf{Z}_i, i = 1, \ldots, n$. Then conditionally on $\mathbf{Z}_i = \mathbf{z}$, we have

$$F_i(y|\mathbf{z}) = P\{Y_i \leq y|\mathbf{Z}_i = \mathbf{z}\} = F(y - \alpha - \boldsymbol{\beta}'\mathbf{t}_i - \boldsymbol{\gamma}'\mathbf{z}), \quad i = 1, \ldots, n, \tag{2.32}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are respectively the regression parameter vector of Y on the design and concomitant variates, and α is the intercept parameter. In this linear model setup, in a nonparametric formulation, the d.f. F is allowed to be arbitrary (but, continuous), so that the finiteness of its second moment is not that crucial. In a parametric as well as nonparametric formulation a basic assumption on the concomitant variates is that they are not affected by the design variates, so that the $\mathbf{Z}_i$ are i.i.d.r.v.'s. Here also, in a nonparametric formulation, the joint distribution of $\mathbf{Z}_i$ is taken to be an arbitrary continuous one (defined on $\mathcal{R}^p$). The Chatterjee-Sen (1964) *multivariate rank permutation principle* plays a basic role in this nonparametric *analysis of covariance (ANOCOVA)* problem. Basically, for the $(p+1)$-variate observable stochastic vectors $(Y_i, \mathbf{Z}_i')'$, with respect to the q-variate design vectors $\mathbf{t}_i$, one can construct a $q \times (p+1)$ linear rank statistics vector with the elements

$$L_{njk}, \text{ for } j = 0, 1, \ldots, p, \ k = 1, \ldots, q, \tag{2.33}$$

where $\mathbf{L}_{n0} = (L_{n01}, \cdots, L_{n0q})'$ stands for the linear rank statistics vector for the primary variate (Y) and is defined as in (2.5) [with the R_{ni} being relabeled as R_{ni0}] , while for the jth coordinate of the concomitant vectors, adopting the same ranking method as in before (2.5) and denoting these ranks by $R_{nij}, i = 1, \ldots, n$), we define the linear rank statistics vector $\mathbf{L}_{nj} = (L_{nj1}, \cdots, L_{njq})'$ as in (2.5), for $j = 1, \ldots, p$. Note that ranking is done separately for each coordinate of the concomitant vector and the primary variate, so that we have a $(p+1) \times n$ *rank collection* matrix $\mathbf{R}_n$. The Chatterjee-Sen rank permutation principle applies to the $n!$ *column permutations* of $\mathbf{R}_n$ (which are conditionally equally likely), and this generates *conditionally distribution-free (CDF)* tests based on the linear rank statistics in (2.33). We may allow the scores [defined before (2.5)] to be possibly different for the primary and concomitant variates, so for the jth coordinate, these scores are taken as $a_{nj}(k), \ k = 1, \ldots, n; \ j = 0, 1, \ldots, p$, and further, without any loss of generality, we may standardize these scores in such a way that adopting the definitions in (2.6), the $\bar{a}_{nj}$ are all equal to 0 and the A^2_{nj} are all equal to one, $j = 0, 1, \ldots, p$.

11

Consider then a $(p+1) \times (p+1)$ matrix $\mathbf{V}_n$ whose diagonal elements are all equal to one, and whose elements are given by

$$v_{njl} = \sum_{i=1}^{n} a_{nj}(R_{nij})a_{nl}(R_{nil}), \quad j,l = 0,1,...,p. \tag{2.34}$$

We denote the cofactor of v_{noo} in $\mathbf{V}_n$ by $\mathbf{V}_{noo}$, $\mathbf{v}_{no} = (v_{no1},...,v_{nop})'$, and denote by

$$\mathbf{w}_n = \mathbf{V}_{noo}^{-1}\mathbf{v}_{no}, \tag{2.35}$$

$$v_{noo}^{*} = v_{noo} - \mathbf{v}_{no}'\mathbf{V}_{noo}^{-1}\mathbf{v}_{no}, \tag{2.36}$$

and

$$\mathbf{L}_{no}^{*} = \mathbf{L}_{no} - (\mathbf{L}_{n1},...,\mathbf{L}_{np})\mathbf{w}_n. \tag{2.37}$$

Let us define $\mathbf{Q}_n$ as in (2.7) and consider the quadratic form

$$\mathcal{L}_{no}^{*} = \{(\mathbf{L}_{no}^{*})'\mathbf{Q}_n^{-1}(\mathbf{L}_{no}^{*})\}/v_{noo}^{*}, \tag{2.38}$$

which may be used as a test statistic for testing the null hypothesis $H_0 : \boldsymbol{\beta} = \mathbf{0}$ against alternatives that $\boldsymbol{\beta} \neq \mathbf{0}$, treating θ and γ as nuisance parameters. Asymptotic nonnull ditribution theory, power properties and optmality of such aligned rank order tests (for local alternatives), studied first by Sen and Puri (1977), can most conveniently be unified by an appeal to the *uniform asymptotic linearity* of aligned rank statistics, and the results presented in Section 7.3 of Puri and Sen (1985) pertain to this scheme; again, the linearity results in their most general form have been presented in a unified manner in Chapter 6 of Jurečková and Sen (1995). This latter reference also contains a good account of the recent developments on *regression rank scores* procedures which may have some advantages (in terms of computational simplicity) over the aligned rank tests.

In the above formulation of a mixed-effect model, the linearity of the regression of the primary response variate on the design and concomitant variates has been taken for granted, while the normality of the errors has been waived to a certain extent by less stringent assumptions. While this can, often, be done with appropriate transformations on primary and concomitant variates, there are certain cases where it may be more reasonable to allow the regression on the concomitant variate part to be rather of some arbitrary (unknown) functional form. That is, the regression on the design variates is taken to be of a parametric (viz., linear) form, while the regression on the covariates is taken as of a *nonparametric* form. In this formulation, for the conditional d.f.'s in (2.32), we take

$$F_i(y \mid \mathbf{z}) = F(y - \boldsymbol{\beta}'\mathbf{t}_i - \theta(\mathbf{z})), \quad i = 1,...,n, \tag{2.39}$$

where the d.f. $F(.)$, $\boldsymbol{\beta}$ etc. are all defined as in before, while $\theta(\mathbf{z})$ is a *translation-equivariant* (location-regression) functional, depicting the regression of the errors $Y_i - \boldsymbol{\beta}'\mathbf{t}_i$ on the concomitant vector $\mathbf{Z}_i$. The basic difference between (2.32) and (2.39) is that in the former case, the linear regression function $\boldsymbol{\gamma}'\mathbf{z}$ involves a finite dimensional parameter $\boldsymbol{\gamma}$, while in the latter case, the nonparametric regression function

$\theta(\mathbf{z})$ may not be finite-dimensional, nor to speak of a linear one. Thus, here we need to treat $\theta(\mathbf{z})$ as a functional defined on the domain Z of the concomitant variate $\mathbf{Z}$. This formulation may generally entail extra regularity (smoothness) conditions on this nonpartametric functional, and because of that, the estimation of $\theta(\mathbf{z})$, $\mathbf{z} \in Z$, may entail a comparatively slower rate of convergence. Nevertheless, as regards the estimation of the fixed-effects parameters (i.e., $\boldsymbol{\beta}$), the conventional $\sqrt{n}$-rate of convergence still holds, although these conventional estimators may not be fully efficient, even asymptotically. A complete coverage of nonparametric methods in this type of mixed-effects models is beyond the scope of this treatise; we may refer to Sen (1995a,c) where a detailed treatment is included.

3 Two-way layouts nonparametrics

The simplest kind of designs for two-way layouts are the so called *randomized block* or *complete block* designs. 5equal number of times in each block, and the treatment combinations may Consider a randomized block design comprising $n(\geq 2)$ blocks of $p(\geq 2)$ plots each, such that p different treatments are applied to the p plots in each block. The allocation of the treatments into the plots in each block is made through randomization. Let Y_{ij} be the response of the plot in the ith block receiving the jth treatment, for $i = 1, \ldots, n, j = 1, \ldots, p$. In the normal theory model, it is assumed that

$$Y_{ij} = \mu + \beta_i + \tau_j + e_{ij}, \quad i = 1, \ldots, n; j = 1, \ldots, p, \tag{3.1}$$

where μ is the *mean effect*, β_i is the ith *block effect*, τ_j is the jth *treatment effect*, and the e_{ij} are the error components which are assumed to be independent and identically distributed according to a normal distribution with zero mean and a finite, positive variance σ^2. The block and treatment effects may either be fixed or random, rersulting in the heirarchy of fixed-, mixed- and random-effects models. As in the case of one-way layouts, a departure from such model assumptions can take place along the routes of nonlinearity of the model, possible heteroscedasticity, dependence or nonnormality of the errors. It is quite interesting to note that the method of m-ranking, one of the earliest nonparametric procedures, has a basic feature that it does not need many of these regularity assumptions, and yet works out in a very simple manner. Suppose that we desire to test the null hypothesis of no treatment effect, treating the block effects as nuisance parameters. Under this hypothesis, in (3.1), the τ_j drop out, so that the observations within a block are i.i.d.r.v. We may even allow the errors to be exchangeable (instead of i.i.d.), and this implies that under the above hypothesis, the observations within a block are *exchangeable* or *interchangeable* r.v.'s. Therefore, if we denote by r_{ij} the rank of Y_{ij} among $Y_{i1}, \ldots, Y_{ip}$, for $j = 1, \ldots, p$, then, for each $i(= 1, \ldots, n)$, under the hypothesis of no treatment effect, the ranks $r_{i1}, \ldots, r_{ip}$ are interchangeable r.v.'s. Moreover, for different blocks, such *intra-block* rank-vectors are stochastically independent of each other. Therefore, the problem of testing the null hypothesis of no treatment effect in a randomized block design can be reduced to that of testing the interchangeability of the within block rankings. On the other hand, this hypothesis can also be stated in terms of the

exchangeability of the within block response variables, and in that setup, the linearity of the block and treatment effects are not that crucial. This scenario leaves us to adopting either of the two routes for nonparametrics in two-way layouts: (i) Incorporate such intra-block rankings with the major emphasis on robustness against possible nonnormality of the errors as well as nonlinearity of the effects, and (ii) Deemphasize the normality of errors, but with due respect to the linearity of the model, incorporate inter-block comparisons in a more visible manner to develope appropriate rank procedures which are robust to possible nonnormality of errors. Aligned rank procedures are quite appropriate in this context, and we shall discuss them later on. For intra-block ranking procedures, we consider a set of scores $\{a(1), ..., a(p)\}$ which may depend on p and some underlying score generating function (but not on the number of blocks). In general these are different from the ones introduced in Section 2 (for one-way layouts). For optimal scores for specific types of local alternatives, we may refer to Sen (1968a). Then, we may define

$$T_{nj} = \sum_{i=1}^{n} a(r_{ij}), \quad j = 1, \ldots, p. \tag{3.2}$$

Moreovere let

$$\bar{a} = p^{-1} \sum_{j=1}^{p} a(j), \quad A^2 = (p-1)^{-1} \sum_{j=1}^{p} [\, a(j) - \bar{a}\,]^2. \tag{3.3}$$

Then, a suitable test statistic for testing the hypothesis of no treatment effect is the following:

$$\mathcal{L}_n = (nA^2)^{-1} \sum_{j=1}^{p} (\, T_{nj} - n\bar{a}\,)^2. \tag{3.4}$$

In particular, if we let $a(j) = j, j = 1, \ldots, p$, the T_{nj} reduce to the rank sums, $\bar{a} = (p+1)/2$ and $A^2 = p(p+1)/12$, so that (3.4) reduces to the classical Friedman (1937) χ_r^2 test statistic:

$$\chi_r^2 = \frac{12}{np(p+1)} \sum_{j=1}^{p} (\, \sum_{i=1}^{n} r_{ij} - n(p+1)/2\,)^2, \tag{3.5}$$

Similarly, letting $a(j) = 0$ or 1 according as j is $\leq (p+1)/2$ or not, we obtain the well known Brown and Mood (1951) *median* test statistic. In either case, and in general, for (3.4), the exact distribution (under the null hypothesis) can be obtained by complete enumeration of all possible equally likely $(p!)^n$ permutations of the intra-block rank vectors, each over $(1, ..., p)$. This process may become quite cumbersome as p and/or n increase. Fortunately, the central limit theorems are adoptable for the intra-block rank vectors which are independent of each other, and hence, it follows that under the null hypothesis, $\mathcal{L}_n$ has closely the central chi squared distribution with $p-1$ DF when n is large. The main advantage of using an intra-block rank test, such as $\mathcal{L}^n$ in (3.4), is that it eliminates the need for assuming additive block effects, and also, the treatment effects may not be additive too. As in the case of the Kruskal-Wallis test, introduced for one-way layouts in the last section, stochastic ordering of the treatment responses (within each block) suffices for the consistency of the test based on such intra-block ranks. Thus, such tests are very robust. The Brown-Mood median test is asymptotically optimal for

local shift alternatives when the underlying d.f. F is Laplace, while for a logistic F, the Friedman χ_r^2 is locally optimal. We may refer to Sen (1968a) for a detailed discussion of the choice of locally optimal intra-block rank tests in some specific models. The main drawback of such intra-block rank tests is that they may not adequately incorporate the *inter-block information* as is generally provided by comparisons of observations from different blocks. For example, if the block effects are additive then a contrast in the ith block observations has the same distribution as in any other block, and hence, some comparisons of such contrasts may provide additional information and may lead to more efficient tests.There are various ways of inducing such inter-block comparisons in rank tests, and among them the two popular ones are the following: (i)*Ranking after alignment*, and (ii) *weighted ranking.*

In a weighted ranking method, instead of having the sum statistics $\sum_{i=1}^n a(r_{ij})$, the intra-block ranings or rank scores are weighed to reflect possible inter-block variation, and such weights are typically inversely proportional to some measure of the within block dispersion of the observations (such as the range or standard deviation or even some rank measures of dispersion). Thus, we may use the statistics $\sum_{i=1}^n w_{ni}a(r_{ij}), j = 1, \ldots, p$, where the w_{ni} are nonnegative weights, and are typically random elements. The analysis can then be carried out in the same manner as in before. Note that such a measure of intra-block dispersion is typically independent of the ranks r_{ij}, so that given these weights, a very similar test statistic can be worked out by reference to the $(p!)^n$ permutations of the intra-block rankings. However, such a law is conditional on the given set of weights, so that we end up with conditionally distribution-free tests instead of EDF tests based on $\mathcal{L}_n$. One way of achieving the EDF property of such weighted ranking procedures is to replace the w_{ni} by their ranks and allowing these ranks to have all possible $(n!)$ realizations. Since these have been pursued in some other chapters of this volume (and also presented in detail in Chapter 10 of *HBS, Vol.4*), we shall not go into further details. The main drawback of such weighted ranking procedures is that the choice of the weights (typically stochastic) retains some arbitrariness and thereby introduces some extra variability, which in turn may generally lead to some loss of efficiency when in particular the block effects are additive. This feature is shared by the other type of weighing where the ranks of the w_{ni} are used instead of their ordinary values. However, if the block efects are not additive and the intra-block error components have the same distribution with possibly different scale parameters, weighing by some measure of dispersion alone may not be fully rational, and hence, from that perspective, such weighing procedures are also subjected to criticism.

Ranking after alignment has a natural appeal for the conventional linear model even when the errors are not normally distributed. The basic idea is due to Hodges and Lehmann (1962) who considered a very simple setup, and it has been shown by Mehra and Sarangi (1967) and in a more general setup by Sen (1968b) that such procedures are quite *robust* under plausible departures from model based assumptions (including homoscedasticity, normality and independence of the errors). As such, we may like to provide more practical aspects of this methodology. To motivate the alignment procedure, we go back to the conventional linear model in (3.1) [sans the normality of the error components]. Suppose

further that the block-effects are either random variables (which may be taken as i.i.d.) or they are fixed, and the errors in the same block are interchangeable or exchangeable random variables. In this way, we are able to include both fixed- and mixed-effects models in our formulations. Let $\tilde{Y}_i$ be a *translation-equivariant* function of $(Y_{i1}, \ldots, Y_{ip})$, such that it is symmetric in its p arguments. Typically, we choose a robust estimator of the ith block mean response, and in order to preserve robustness, instead of the block average, median, trimmed mean or other measures of central tendency can be adopted. Define then the aligned observations as

$$\tilde{Y}_{ij} = Y_{ij} - \tilde{Y}_i, \ j = 1, \ldots, p; \ i = 1, \ldots, n. \tag{3.6}$$

By (3.1) and (3.6), we may write

$$\tilde{Y}_{ij} = \tau_j - \tilde{\tau} + \tilde{e}_{ij}; \ \tilde{e}_{ij} = e_{ij} - \tilde{e}_i, \tag{3.7}$$

for $j = 1, \ldots, p; i = 1, \ldots, n$, where $\tilde{\tau}$ and $\tilde{e}_i$ are defined by the same functional form as the $\tilde{Y}_i$. Note that for each $i(= 1, \ldots, n)$, the joint distribution of $(\tilde{e}_{i1}, \ldots, \tilde{e}_{ip})$ is symmetric in its p arguments, and moreover these vectors have the same joint distribution for all blocks. Therefore, it seems very logical to adopt an overall ranking of all the $N = np$ aligned observations $(Y_{11}, \ldots, Y_{np})$ and base a rank test statistic on such aligned ranks. The only negative feature is that the overall ranking procedure distorts the independence of the rank vectors from block to block; nevertheless they retain their permutability, and this provides the access to developing conditionally distribution-free tests for testing the null hypothesis of no treatment effect.

Let $\tilde{Y}_{i:1}, \ldots, \tilde{Y}_{i:p}$ be the order statistics corresponding to the aligned observations $\tilde{Y}_{i1}, \ldots, \tilde{Y}_{ip}$ in the ith block, for $i = 1, \ldots, n$. Then under the null hypothesis of interchangeability of the $Y_{ij}, j = 1, \ldots p$, for each $i(= 1, \ldots, n)$, the $\tilde{Y}_{ij}$ has the (discrete) uniform distribution over the $p!$ possible permutations of the coordinates of $(\tilde{Y}_{i:1}, \ldots, \tilde{Y}_{i:p})$, and this permutation law is independent for different blocks. Thus, we obtain a group of $(p!)^n$ of permutations generated by the within block permutations of the aligned order statistics, and by reference to this (conditional) law, we can construct conditionally distribution-free tests for the hypothesis of interchangeability of the treatments. Under block-additivity, the vector of intra-block (aligned) order statistics are interchangeable, and hence, rakning after alignment (ignoring the blocks) remains rational. For the aligned observations, we define the ranks R_{ij} as in the preceding section, so that these R_{ij} take on the values $1, \ldots, N$, when ties among them are neglected, a case that may be done under very mild continuity assumptions on the error distributions. For the pooled sample size N, we introduce a set of scores $a_N(k), k = 1, \ldots, N$, as in Section 2, and consider the *aligned rank statistics* :

$$T_{Nj} = n^{-1} \sum_{i=1}^{n} a_N(R_{ij}), \ j = 1, \ldots, p. \tag{3.8}$$

Also, define

$$\bar{a}_{Ni} = p^{-1} \sum_{j=1}^{p} a_N(R_{ij}), \ i = 1, \ldots, n; \tag{3.9}$$

16

$$V_N = \{n(p-1)\}^{-1} \sum_{i=1}^{n} \sum_{j=1}^{p} \{ a_N(R_{ij}) - \bar{a}_{Ni} \}^2. \tag{3.10}$$

Then an aligned rank test statistic for testing the hypothesis of no treatment effect can be formulated as

$$\mathcal{L}_N^o = n\{\sum_{j=1}^{p} [T_{Nj} - \bar{a}_N]^2\}/V_N, \tag{3.11}$$

For small values of n (and p), the permutational (conditional) distribution of $\mathcal{L}_N^o$ can be incorporated to construct a conditionally distribution-free test for the above hypothesis, while, it follows from Sen (1968b) that for large sample sizes, under the null hypothesis, $\mathcal{L}_N^o$ has closely chi squared distribution with $p-1$ DF. Various robustness properties of such aligned rank tests have been studied in detail by Sen (1968c). It has been observed there that it may not be necessary that the aligned errors $\tilde{e}_{ij}$ have the common distribution for all i (i.e., blocks). In particular, for the heteroscedastic model, allowing the scale parameters to vary from block to block, it was observed that an aligned rank test may have greater ARE with respect to the classical ANOVA test than in the homoscedastic case. Some of these details are also reported in Chapter 7 of Puri and Sen (1971). Also, the alignment procedure remains applicable in the mixed-effects model too, where the block effects being stochastic or not drop out due to alignment, and hence, better robustness properties percolate. We shall discuss this aspect later on. More important is the fact that the ARE of aligned rank tests relative to the intra-block rank tests based on conjugate scores is generally greater than 1, particularly when p is not so large. For example, for the Wilcoxon score rank statistics, the ARE of the aligned rank test with respect to the Friedman χ_r^2 test is $\geq (p+1)/p$, so that for small values of p, there may be considerable gain in using an aligned rank test, albeit in terms of model robustness, the intra-block rank tests fare better.

In the above development, it has been assumed that each treatment is applied to one plot in each block. We may consider a more general case where the jth treatment is applied to $m_j(\geq 1)$ plots in each block, for $j = 1, \ldots, p$. We let $M = \sum_{j \leq p} m_j$ and $N' = nM$. Thus, the aligned ranks span over the set $\{1, \ldots, N'\}$, and the average of the nm_j rank scores for the jth treatment is denoted by $T_{N',j}, j = 1, \ldots, p$. The defdinition of V_N is modified accordingly. Then, as a direct extension of (3.11), we may consider the following aligned rank test statistic:

$$\mathcal{L}_{N'}^o = [n/V_{N'}]\{\sum_{j=1}^{p} m_j[T_{N',j} - \bar{a}_{N'}]^2. \tag{3.12}$$

This test is also conditionally distribution-free and under the null hypothesis, it has asymptotically central chi squared distribution with $p-1$ DF. It enjoys all the robustness and asymptotic efficiency properties as in the particular case of all the m_j being equal to 1.

Next, as in the case of one-way layouts treated in Section 2, we consider the problem of simultaneous testing for all paired treatment differences. We may note here that there are some variations in the formulation of such tests. Treatments vs. control tests compare simultaneously all the treatments with a control, so that there is some asymmetry in this setup. A simultaneous test for all possible

$\binom{p}{2}$ treatment differences preserves the symmetry to a greater extent. It is also possible to formulate such simultaneous tests based on intra-block rank statistics, but they may not incorporate inter-block information, and hence, may be a little less efficient than the ones based on aligned rank statistics. The basic results are due to Sen (1970a). We define the $T_{N,j}$ and V_N as in before and let

$$W_N = \max\{n^{1/2}|T_{N,j} - T_{N,l}|/V_N^{1/2} : 1 \le j \ne l \le p\} \tag{3.13}$$

and let $R_{p,\alpha}$ be the upper $100\alpha\%$ point of the distribution of the sample range of a sample of size p drawn from a standard normal distribution. Then the simultaneous (aligned) rank test is based on the decision rule that rejects all pairs (j, l) of treatments as significantly different for which

$$n^{1/2}|T_{N,j} - T_{N,l}| \ge R_{p,\alpha}V_N^{1/2}. \tag{3.14}$$

This simultaneous test is also conditionally distribution-free, and asymptotically it has the level of significance α.

In many situations the treatments represent an increasing sequence of *doses*, and it may therefore be of natural interest to test for the null hypothesis of equality of treatment effects against an *ordered alternative* $H_1 : \tau_1 \le \cdots \le \tau_p$ (with at least one strict inequality). Tests for such ordered alternatives may also be based on intra-block rank statistics, but they may be comparatively less efficient for the same reason (of not incorporating inter-block information that effectively). Hence, aligned rank tests are often advocated. We define the aligned rank statistics $T_{N,j}$, $j = 1, \ldots, p$, as in before, and let $\mathbf{a} = (a_1, \ldots, a_p)'$ be a vector of real coefficients satisfying the conditions : (i) $\sum_{j=1}^p a_j = 0$, (ii) $\sum_{j=1}^p a_j^2 = 1$ and (iii) $a_1 \le a_2 \le \ldots \le a_p$ with at least one strict inequality. Then a test statistic may be defined conveniently as

$$Q_n = n[\sum_{j=1}^p a_j\{T_{N,j} - \bar{a}_N\}]/V_N^{1/2}. \tag{3.15}$$

The choice of $\mathbf{a}$ may be made on some heuristic considerations, and in this context, Bayesian solutions have also been incorporated. Among such possibilities, the following one [viz., Sen (1968b)] provides a simple interpretation from linear trend point of view. We let

$$a_j = \sqrt{12}\{j - (p+1)/2\}/\{p(p^2 - 1)\}^{1/2}, \ j = 1, \ldots, p. \tag{3.16}$$

The resulting test statistic

$$Q_n^o = \sqrt{12n}\{\sum_{j=1}^p (j - (p+1)/2)T_{N,j}\}/\{p(p^2 - 1)V_N\}^{1/2} \tag{3.17}$$

is conditionally distribution-free (under H_0) and asymptotically normal. It provides a robust and efficient test for ordered alternatives. Some other nonparametric tests for ordered alternatives in randomized blocks are discussed in detail in Chapter 7 of Puri and Sen (1971).

As in Section 2, we will consider here R-estimators of the treatment effects $\tau_1, \ldots, \tau_p$, treating the block effects as nuisance parameters (or possibly random variables, in a mixed-effects setup). In this

context, the aligned rank statistics introduced earlier play a vital role. For each pair (j,l), consider the paired differences $Y_{i,jl} = Y_{ij} - Y_{il}$, for $i = 1, \ldots, n$. We write $\Delta_{jl} = \tau_j - \tau_l$ and $e_{i,jl} = e_{ij} - e_{il}$, for $i = 1, \ldots, n; 1 \leq j < l \leq p$. Then the $e_{i,jl}$ have a d.f. symmetric about 0, so that we may use an aligned signed rank statistic $S_{n,jl}(d)$ as in (2.25) and equating this to 0 (with respect to d), we obtain an R-estimator $(\hat{\Delta}_{n,jl})$ of Δ_{jl}. In this process, we obtain the set of $\binom{p}{2}$ estimators $\hat{\Delta}_{n,jl}, 1 \leq j < l \leq p$. We let $\Delta_{jj} = 0$ and note that by construction, $\hat{\Delta}_{n,lj} = -\hat{\Delta}_{n,jl}$, for all $j, l = 1, \ldots, p$. In the case of least squares estimation theory, the estimators are linear, so that an estimator of a contrast in the τ_j can be expressed in terms of such paired difference estimators in an arbitrary manner. But, the R-estimators may not be strictly linear, and hence, for small number of blocks at least, there remains some arbitrariness in combining such paired difference estimators to yield an estimator of an arbitrary contrast. Lehmann (1963a,b, 1964) suggested a simple modification to yield *compatible R-estimators* of contrasts in randomized block designs. For this define

$$\hat{\Delta}_{n,j\cdot} = p^{-1} \sum_{l=1}^{p} \hat{\Delta}_{n,jl}, j = 1, \ldots, p; \tag{3.18}$$

$$\hat{\Delta}_{n,jL}^{o} = \hat{\Delta}_{n,j\cdot} - \hat{\Delta}_{n,l\cdot}, j, l = 1, \ldots, p. \tag{3.19}$$

Like R-estimators in Section 2, the compatible estimators are translation-equivariant, and tey are robust, consistent and asymptotically normally distributed. However, the way (3.19) has been formulated, expressions for the asymptotic dispersion matrix of the compatible estimators are slightly different from the ones in Section 2 [cf. (2.20)]. Towards this we introduce the following parameters. Let $G(.)$ be the common marginal d.f. of the paired difference $e_{i,jl}$, and let $G^*(.)$ be the bivariate d.f. of a pair $(e_{i,jl}, e_{i,jl'})$, where $l \neq l'$. Note that the process of alignment distorts the independence of the errors even when the original e_{ij} were stochastically independent. Also, let $\phi(\cdot)$ be the score generating function for the rank statistics. Then we define $A^2 = \int_0^1 \phi^2(u)du$ and

$$\gamma(\phi, G) = \int_{\mathcal{R}} (d/dx)\phi(G(x))dG(x), \tag{3.20}$$

and

$$\lambda_\phi(G) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(G(x))\phi(G(y))dG^*(x, y). \tag{3.21}$$

The score generating function is taken to be skew-symmetric about $u = 1/2$, so that $\bar{\phi} = 0$. Moreover, by definition in (3.19), the $\hat{\Delta}_{n,jl}^o$ are expressible as paired differences of the p statistics $\hat{\Delta}_{n,j\cdot}, j = 1, \ldots, p$. Hence, to study the joint distribution of all these $\binom{p}{2}$ estimators, it suffices to consider only the $p - 1$ vector $n^{1/2}(\hat{\Delta}_{n,1p}^o - \Delta_{1p}, \ldots, \hat{\Delta}_{n,p-1p}^o - \Delta_{p-1p})$. This vector has asymptotically a $p - 1$ variate normal distribution with null mean vector and dispersion matrix $\sigma_o^2[\mathbf{I}_{p-1} + \mathbf{1}_{p-1}\mathbf{1}_{p-1}']$, where $\mathbf{1}_{p-1} = (1, \ldots, 1)'$, and

$$\sigma_o^2 = [(p-1)/p]\{A^2 + (p-2)\lambda_\phi(G)\}/\gamma^2(\phi, G). \tag{3.22}$$

If we compare the dispersion matrix of the compatible R-estimators with that of the raw R-estimators derived earlier, we obtain that the ARE of the compatible estimators relative to the raw ones is given by

$$e(\hat{\Delta}^{\circ}; \hat{\Delta}) = pA^2/\{2[A^2 + (p-2)\lambda_\phi(G)]\}, \tag{3.23}$$

and using the easily verifiable inequality that $\lambda_\phi(G) \leq (1/2)A^2$, it readily follows that (3.23) is bounded from below by 1 (albeit, this is usually quite close to 1). This shows that such compatible R-estimators are also preferable on the ground of their asymptotic efficiency properties. For details, we may refer to Puri and Sen (1971,Ch.7).

4   Nonparametric MANOVA

In a general multivarite setup, the response $\mathbf{Y}$ is a p-vector, for some $p \geq 1$, and as in Sections 2 and 3, these responses may be related to various factors (fixed-, random- or mixed-effects models). In the conventional case, one assumes that a linear model as in (2.1) or (3.1) holds where the e_i or e_{ij} are distributed according to a multinormal distribution with null mean vector and a positive definite (unknown) dispersion matrix Σ. The scope for departures from this assumed model is even more in the multivariate case, as normality is even less likely to hold in the multivariate than univariate cases. In design aspects, of course, the situation is quite similar to that in the univariate case, but in statistical analysis, the situation is more complex in multivariate models. In nonparametrics, there are additional impasses in the multivariate case. In the conventional linear model, *affine transformations* are often used to simplify the distribution theory of appropriate test stastistics which are invariant under affine transformations or of estimators which are *affine-equivariant*. In ranking procedures, usually ranking is made for each of the p coordinates separately. Thus, in one hand, such procedures are invariant for a larger group of (not necessarily linear) strictly monotone transformation for each coordinate, while, on the other hand, they are not affine-invariant. Therefore, in nonparametrics for multivariate analysis of variance (MANOVA), affine-invariance is not generally true. This does not, of course, pose a serious problem, as in many cases, the coordinate responses may not be quite conformable in a sense that an arbitrary linear compound will have a meaningful interpretation. In such a case, coordinatewise ranking with due emphasis on their dependence may serve the purpose much better. The other serious problem with nonparametrics in MANOVA is the lack of EDF property in a general setup. For example, in Section 2 we have posed the permutational invariance property which yields $n!$ equally likely permutations of $\{1, \ldots, n\}$ for the rank vector. In the multivariate case, we have the rank matrix of order $p \times n$, so that the total number of possible realizations of these matrices is equal to $(n!)^p$, and the distribution of the rank matrix over this set depends, in general, on the underlying (multivariate) distribution. Thus, a test based on the coordinatewise ranking, in general, may not be genuinely distribution-free. A very similar situation arises in two-way layouts when one uses the method of ranking separately

for each coordinate. This drawback of multivariate nonparametrics has been eliminated largely by the *rank permutation principle* due to Chatterjee and Sen (1964). This rank permutation priciple has been exploited in various directions , and up-to-date accounts of these developments are given by Puri and Sen (1971,1985). We present here only a brief synopsis of the main highlights of these developments with due emphasis on the design aspects.

First, consider the multivariate analogue of (2.3) where the $\mathbf{Y}_i$ are p-vectors, so that θ and the $\mathbf{e}_i$ are p-vectors too, while β is a $p \times q$ matrix. Then $\mathbf{e}_i$ is assumed to have a p-variate (continuous) d.f. F, which need not have independence structure. The hypotheses in (2.4) remain the same. Let us denote the rank of Y_{ij} among the $Y_{rj}, r = 1, \ldots, n$ by R_{nij}, for $i = 1, \ldots, n; j = 1, \ldots, p$. As in (2.5), for the jth coordinate, we denote the vector of linear rank statistics by $\mathbf{L}_{nj}$, for $j = 1, \ldots, p$, where, we may even take the scores $a_n(.)$ possibly different for different $j(= 1, \ldots, p)$; we may add an additional subscript j to $a_n(.)$ to do so, but for notational simplicity, this refinement is suppressed. We also define a $p \times p$ matrix $\mathbf{V}_n$ as in (2.34), and the rank collection matrix $\mathbf{R}_n$ is defined as in after (2.33). The permutational (conditional) probability law is then generated by the $n!$ (conditionally) equally likely permutations of the columns of the rank collection matrix. We define $\mathbf{Q}_n$ as in (2.7), and if we denote the $p \times q$ matrix of the L_{njl} by $\mathbf{L}_n$, as in (2.8), we will have here $E_{\mathcal{P}_n}(\mathbf{L}_n) = \mathbf{0}$ and

$$E_{\mathcal{P}_n}[\text{vec } \mathbf{L}_n \text{vec } \mathbf{L}_n'] = A^2 \cdot \mathbf{V}_n \otimes \mathbf{Q}_n, \tag{4.1}$$

where vec $\mathbf{L}_n$ is the pq-vector obtained from $\mathbf{L}_n$ by stacking the columns over each other and $\otimes$ stands for the Kronecker product of the two matrices. Then, as an extension of (2.9) to the multivariate case, we have the test statistic

$$\mathcal{L}_n = A^{-2}[\text{vec } \mathbf{L}_n'(\mathbf{V}_n \otimes \mathbf{Q}_n)^- \text{vec } \mathbf{L}_n]. \tag{4.2}$$

For small values of n, the exact (conditional) permutational distribution of $\mathcal{L}_n$ can be obtained by direct enumeration of all possible $n!$ column permutations of the rank collection matrix. This process becomes cumbersome as n increases, but, for large n, this permutation distribution aas well as the unconditional null hypothesis distribution of $\mathcal{L}_n$ can be approximated by the central chi squared distribution with pq DF. Thus, we may proceed as in (2.10) with q replaced by pq, and also, we may consider, in the same vein, local Pitman-type alternatives as in (2.11) and consider asymptotic power properties of such multivariate rank tests. There are some difficulties concerning the characterization of asymptotically optimal rank tests (compare with (2.12)), and further regularity conditions are needed to establish such properties in a general multivariate setup. We may refer to Puri and Sen (1985, Chs.5-7). For the specific multisample multivariate case, treated in the univariate case in Section 2, the expression for $\mathcal{L}_n$ in (4.2) simplifies to the following:

$$\mathcal{L}_n = \sum_{k=1}^{c} n_k(\bar{\mathbf{a}}_{n,k} - \bar{\mathbf{a}}_n)'\mathbf{V}_n^-(\bar{\mathbf{a}}_{n,k} - \bar{\mathbf{a}}_n)/A_n^2, \tag{4.3}$$

where the $\bar{\mathbf{a}}_{n,k}$ and $\bar{\mathbf{a}}_n$ are defined as in (2.13), but for the vector case. The discussion following (2.13) pertaining to the specific choice of scores also applies here.

Let us consider the R-estimation problem for the parameter matrix β. As a natural generalization of (2.14), we may consider here a measure of rank dispersion $D_n(\mathbf{B})$ defined for $\mathbf{B} \in \mathbf{R}^{pq}$, wherein we replace the $a_N(R_{ni}(\mathbf{b}))$ by $\mathbf{a}_n(\mathbf{R}_{ni}(\mathbf{B}))$, $\bar{a}_n$ by $\bar{\mathbf{a}}_n$ and $Y_i(\mathbf{b})$ by $\mathbf{Y}_i(\mathbf{B}) = \mathbf{Y}_i - \mathbf{B}(\mathbf{t}_i - \bar{\mathbf{t}}_n)$; also we need to introduce a matrix to depict the scale factors and possible dependence of the coordinates of the vector $\mathbf{Y}$. Because the rankings are made separately for each coordinate and affine invariance may not generally hold for such R-estimators, we find it convenient to adopt a coordinatewise R-estimation procedure. Based on the aligned scores and aligned observations on the jth coordinate, we define a measure of rank dispersion as in (2.14), for $j = 1, \ldots, p$. Then proceeding virtually as in (2.14) through (2.17), we arrive at the following:

$$\mathbf{L}_{nj}(\mathbf{b}_j) = \mathbf{0}, \; j = 1, \ldots, p, \tag{4.4}$$

where each $\mathbf{b}_j$ is a q-vector. If we express $\beta' = (\beta'_1, \ldots, \beta'_p)$, then the jth estimating equation in (4.4) yields the R-estimator $\hat{\beta}_{nj}$ of β_j, for $j = 1, \ldots, p$. Each of these estimators (vectors) satisfies a FOADR result given in (2.18) where we need to attach a subscript j to each $\hat{\beta}_n, \beta, \gamma, \phi, F$ and e_i to indicate their dependence on the jth coordinate, for $j = 1, \ldots, p$. We denote the dispersion matrix of the p-vector $(\gamma_1^{-1}\phi_1(F_{[1]}(e_{i1})), \ldots, \gamma_p^{-1}\phi(F_{[p]}(e_{ip})))$ by $\boldsymbol{\Psi}$, and then as a direct extension of (2.20) in the multivariate case, we arrive at the following: As $n \to \infty$,

$$(\hat{\beta}_n - \beta)\mathbf{Q}_n^* \xrightarrow{\mathcal{D}} \mathcal{N}(0, \boldsymbol{\Psi} \otimes \mathbf{I}), \tag{4.5}$$

where $\mathbf{Q}_n^*$ is defined as in (2.20). In this matrix case, the definition of the ARE in (2.21) needs some modification, and the usual A-, D- and E-optimality (efficiency) criteria can be incorporated to suit the purpose. Again, we may refer to Puri and Sen (1985, Ch.6) for some of these details.

Let us consider next the MANOVA nonparametrics for the two-way layouts. As in Section 3, here also, we may consider the intra-block ranking and ranking after alignment cases separately, and compare their merits and demerits. The intra-block ranking method was considered by Gerig (1969) who developed a multivariate extension of the Friedman χ_r^2 test statistic and used the same intra-block permutation groups to develop a permutationally (conditionally) distribution-free test for the null hypothesis of no treatment differences accross the q variates. Let us go back to (3.5) and denote by $r_{ij}^{(k)}$ the rank on the kth variate for the jth treatment in the ith block, when ranking is made separately for each variate and within each block. Thus, we will have n rank matrices $\mathbf{R}_i, i = 1, \ldots, n$, where each $\mathbf{R}_i$ is a $q \times p$ matrix with the elements $r_{ij}^{(k)}$. We define a $q \times p$ matrix of rank statistics $\bar{r}_{nj}^{(k)} = n^{-1} \sum_{i=1}^n r_{ij}^{(k)}, j = 1, \ldots, p; k = 1, \ldots, q$, express this into a pq-vector (as in earlier this section), and also define a $q \times q$ matrix $\mathbf{V}_n = ((v_{nkl}))$ by letting

$$v_{n,kl} = [n(p-1)]^{-1} \sum_{i=1}^n \sum_{j=1}^p \{r_{ij}^{(k)} - (p+1)/2\}\{r_{ij}^{(l)} - (p+1)/2\}, \; k,l = 1, \ldots, q. \tag{4.6}$$

Then, the multivariate analogue of (3.5) is given by

$$\mathcal{L}_n = n[\sum_{j=1}^p \sum_{k=1}^q \sum_{l=1}^q v_n^{kl} \{\bar{r}_{nj}^{(k)} - (p+1)/2\}\{\bar{r}_{nj}^{(l)} - (p+1)/2\}], \tag{4.7}$$

22

where the v_n^{kl} stand for the elements of the matrix $\mathbf{V}_n^{-1}$. The permutational (conditional) distribution of $\mathcal{L}_n$ over the set of $(p!)^n$ intra-block column permutations of the rank matrics can be approximated well by the chi squared distribution with $q(p-1)$ DF when n is large, and asymptotic power properties have been studied by Gerig (1969) and others. Gerig (1975) has also extended this test for the multivariate analysis of covariance (MANOCOVA) problem, and has studied its robustness properties too. The modifications follow along the lines of (2.35) through (2.39), and hence, we omit the details. As in the univariate case, such intra-block rank tests are generally not fully informative (as they may not recover the inter-block information to a satisfactory extent), and for this reason, aligned rank procedures in MANOVA and MANOCOVA are often preferred. The prospect for weighted rankings is somewhat less apparent here as the weight would depend on the intra-block dispersion matrices, and hence, more delicate considerations are needed in a rational formulation. Most of the developments on aligned rank procedures in MANOVA/MANOCOVA are due to Sen (1969b, 1984a), and presented succinctly here.

We start with the conventional linear model in (3.1), as extended to the multivariate case, and we drop the assumption of normality of errors. Thus, the response vector for the jth treatment in the ith block is a q-vector $\mathbf{Y}_{ij}$, and as such, in (3.1), we change the elemnents on the right hand side by appropriate q-vectors. For each coordinate $(k = 1\ldots,q)$, we cosider a suitable translation-equivariant intra-block measure of central tendency, and denote these by $\tilde{\mathbf{Y}}_i, i = 1,\ldots,n$. Then, the aligned response vectors are defined as in (3.7) as $\tilde{\mathbf{Y}}_{ij} = \mathbf{Y}_{ij} - \tilde{\mathbf{Y}}_i, j = 1,\ldots,p; i = 1,\ldots,n$. For each $k(=1,\ldots,q)$, we introduce the aligned ranks $R_{ij}^{(k)}, j = 1,\ldots,p; i = 1,\ldots,n$, as in before (3.8), and we define the set of aligned rank statistics $T_{nj}^{(k)}, j = 1,\ldots,p; k = 1,\ldots,q$, as in (3.8) [where we may even allow the score function $a_n(.)$ to vary from one coordinate to another, but for notational simplicity this is dropped]. As a natural extension of (3.10) to the multivariate case, we define then $\mathbf{V}_N = ((v_{N,kl}))$, by letting for each $k, q = 1,\ldots,q$,

$$v_{N,kl} = [n(p-1)]^{-1} \sum_{i=1}^{n} \sum_{j=1}^{p} \{a_N(R_{ij}^{(k)}) - \overline{a}_{N,i}^{(k)}\}\{a_N(R_{ij}^{(l)}) - \overline{a}_{N,i}^{(l)}\}, \tag{4.8}$$

where the intrablock average rank score vectors have the elements denoted by $\overline{a}_{N,i}^{(k)}$, for $k = 1,\ldots,q; i = 1,\ldots,n$. Then, a multivariate version of the general aligned rank statistic in (3.11) is given by

$$\mathcal{L}_N^o = n \sum_{j=1}^{p} \sum_{k=1}^{q} \sum_{l=1}^{q} v_N^{kl} \{T_{N,j}^{(k)} - \overline{a}_N^{(k)}\}\{T_{N,j}^{(l)} - \overline{a}_N^{(l)}\}, \tag{4.9}$$

where the v_N^{kl} are the elements of $\mathbf{V}_N^{-1}$. Here also, a conditionally (permutationally) distribution-free test based on $\mathcal{L}_N^o$ can be obtained by reference to the set of $(p!)^N$ intrablock column permutations of the aligned observation matrices, and for large n, this conditional distribution as well as the unconditional null hypothesis distribution can be well approximated by the central chi squared distribution with $q(p-1)$ DF. Various asymptotic (power and efficiency) properties of this nonparametric MANOVA procedure are studied by Sen (1969b). MANOCOVA nonparametrics also follow the same line of attack as in (2.35) through (2.38). Multivariate extensions of simultaneous (aligned) rank tests for all possible pairs

23

of treatments and all possible coordinates follow by using the Roy (1953) *largest root* criterion.

5 Nonparametrics for Incomplete Block Designs

In this section, we deal with a general subclass of two-way layouts where possibly due to a large number of treatments, blocks of smaller size are used, so that not all treatments are applied to all blocks. In the literature, these are referred to as *incomplete block designs (IBD)*. Consider n replications of an IBD consisting of b blocks of constant size $k(\geq 2)$ to which $r(\geq k)$ treatments are applied in such a way that (i) no treatment occurs more than once in any block, (ii) the jth treatment occurs in $r_j(\leq b)$ blocks, and (iii) the (j, j')th treatments occur together in $r_{jj'}(> 0)$ blocks, for $j \neq j' = 1, \ldots, v$. Let then $\mathcal{S}_j$ stand for the set of treatments occurring in the ith block, $i = 1, \ldots, b$. In the sth replicate $(s = 1, \ldots, n)$, the response of the plot in the ith block and receiving the jth treatment is a stochastic p-vector $\mathbf{X}_{sij}$, for $j \in \mathcal{S}_i; i = 1, \ldots, b$. In the univariate case (i.e., $p = 1$), intra-block rank tests for IBD's are due to Durbin (1951), Benard and Elteren (1953) and Bhapkar (1961a), among others. For some special IBD's, the studies made by Elteren and Noether (1959) and Bhapkar (1961a) reveal the low (Pitman-) efficiency of such tests, particularly when k is small. For this reason and motivated by the results in the preceding two sections, we shall mainly consider here suitable aligned rank tests for IBD's. We shall only summarize the results and for details, we may refer to Sen (1971a).

We consider the model

$$\mathbf{X}_{sij} = \boldsymbol{\mu}_s + \boldsymbol{\beta}_{si} + \boldsymbol{\tau}_j + \boldsymbol{\varepsilon}_{sij}, \ j \in \mathcal{S}_i, \tag{5.1}$$

for $i = 1, \ldots, b; s = 1, \ldots, n$, where the $\boldsymbol{\mu}_s$ stand for the *replicate effects*, $\boldsymbol{\beta}_{si}$ for the block effects (nuisance parameters in fixed-effects models or spurious random vectors in mixed-effects models), $\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_v$ are the treatment effects (parameters of interest) and the $\boldsymbol{\varepsilon}_{sij}$ are the error vectors. We may set without any loss of generality $\sum_{j=1}^v \boldsymbol{\tau}_j = \mathbf{0}$. Instead of the specific multinormality assumption on the errors, it is assumed that for each (s, i), $\{\boldsymbol{\varepsilon}_{sij}, j \in \mathcal{S}_i\}$ have jointly a continuous cumulative d.f. $G(\mathbf{x}_1, \ldots, \mathbf{x}_k)$ which is symmetric in its k argument vectors. This includes the conventional assumption of independence and identity of distributions of all the $N(= nbk)$ error vectors as a special case. As in Section 3 or 4, we define the aligned observations by

$$\mathbf{Y}_{sij} = \mathbf{X}_{sij} - k^{-1} \sum_{l \in \mathcal{S}_i} \mathbf{X}_{sil}, \ j \in \mathcal{S}_i, \ i = 1, \ldots, b, \ s = 1, \ldots, n, \tag{5.2}$$

and let $R_{sij}^{(k)}$ be the rank of $Y_{sij}^{(k)}$ among the N aligned observations on the kth response variate, for $j \in \mathcal{S}_i, i = 1, \ldots, b, s = 1, \ldots, n; k = 1, \ldots, q$. Also, for each $i(= 1, \ldots, b)$, let

$$\tau_{j,i} = \tau_j - k^{-1} \sum_{j \in \mathcal{S}_i} \tau_l, \quad \mathbf{e}_{i,j} = \varepsilon_{ij} - k^{-1} \sum_{l \in \mathcal{S}_i} \varepsilon_{il}, \tag{5.3}$$

for $j \in \mathcal{S}_i, i = 1, \ldots, b; s = 1, \ldots, n$. Then, we have

$$\mathbf{Y}_{sij} = \tau_{j,i} + \mathbf{e}_{i,j}, \ j \in \mathcal{S}_i, \ 1 \leq i \leq b; \ 1 \leq s \leq n. \tag{5.4}$$

We want to test the null hypothesis of no treatment effect, i.e.,

$$H_o : \tau_1 = \cdots = \tau_v = 0, \tag{5.5}$$

against the set of alternatives that at least one of the τ_j is different from 0. Note that for each (s, i), the $e_{sij}, j \in \mathcal{S}_i$ are exchangeable random vectors, so that under the null hypothesis, by (5.4), the aligned vectors within each block (in each replicate) are also exchangeable. This provides the same permutational invariance structure as in the case of complete blocks, and hence, similar conditionally (permutationally) distribution-free aligned rank tests can be constructed. Let us denote by

$$\mathcal{P}_j = \{i (\in [1, b]) : j \in \mathcal{S}_i\}, \quad j = 1, \ldots, v. \tag{5.6}$$

Then, for each $k(= 1, \ldots, q)$, we introduce scrores $a_{Nk}(\alpha), \alpha = 1, \ldots, N$ as in Section 3 (or 4), and denote the block averages of these scores by $\overline{a}_{Nsi.}^{(k)}$, replicate averages by $\overline{a}_{Ns..}^{(k)}$ and the grand average by $\overline{a}_N^{(k)}$. Let then

$$T_{N,j}^{(k)} = n^{-1} \sum_{s=1}^{n} \sum_{i \in \mathcal{P}_j} a_{Nk}(R_{ij}^{(k)}), \; j = 1, \ldots, v; \; k = 1, \ldots, q. \tag{5.7}$$

Recall that the r_j may not be all equal, and so may not be the $r_{jj'}$. This calls for some adjustments for the permutational covariance matrix. We define two matrices $\mathbf{V}_N^{(l)} = ((v_{N,kk'}^{(l)})), l = 1, 2,$. where for $k, k' = 1, \ldots, q,$

$$v_{N,kk'}^{(1)} = (nbk)^{-1} \sum_{s=1}^{n} \sum_{i=1}^{b} \sum_{j \in \mathcal{S}_i} \{a_{Nk}(R_{ij}^{(k)}) - \overline{a}_{Nsi.}^{(k)}\} \{a_{Nk}(R_{ij}^{(k')}) - \overline{a}_{Ns.}^{(k')}\}; \tag{5.8}$$

and

$$v_{N,kk'}^{(2)} = (nb)^{-1} \sum_{s=1}^{n} \sum_{i=1}^{b} [\overline{a}_{Nsi.}^{(k)} - \overline{a}_{Ns..}^{(k)}][\overline{a}_{Nsi.}^{(k')} - \overline{a}_{Ns..}^{(k')}]. \tag{5.9}$$

Also, define two (design) matrices $\mathbf{A}^{(l)} = ((a_{jj'}^{(l)})), l = 1, 2,$ by letting

$$a_{jj'}^{(1)} = [kr_j \delta_{jj'} - r_{jj'}]/(k - 1), a_{jj'}^{(2)} = [br_{jj'} - r_j R_{j'}]/(b - 1), \tag{5.10}$$

for $j, j' = 1, \ldots, v,$ where $\delta_{jj'}$ is the usual Kronecker delta, and $r_{jj} = r_j$. Further, let

$$\mathbf{W}_N = \mathbf{A}^{(1)} \otimes \mathbf{V}_N^{(1)} + \mathbf{A}^{(2)} \otimes \mathbf{V}_N^{(2)}. \tag{5.11}$$

Finally, let

$$\mathbf{T}_N = (T_{N,1}^{(1)}, \ldots, T_{N,v}^{(1)}, \ldots, T_{N,1}^{(q)}, \ldots, T_{N,v}^{(q)}), \tag{5.12}$$

and

$$\boldsymbol{\eta} = (\overline{a}_N^{(1)}(r_1, \ldots, r_v), \ldots, \overline{a}_N^{(q)}(r_1, \ldots, r_v)). \tag{5.13}$$

Then, as in Sen (1971a), we may consider the following aligned rank order test statistic:

$$\mathcal{L}_N^* = (\mathbf{T}_N - \boldsymbol{\eta}) \mathbf{W}_N^- (\mathbf{T}_N - \boldsymbol{\eta})', \tag{5.14}$$

where $\mathbf{W}_N^-$ stands for a generalized inverse of $\mathbf{W}_N$. Keeping in mind the class of balanced, partially balanced and group divisible IBD's, as in Sen (1971), we may assume that

$$\text{Rank of} \quad \mathbf{A}^{(1)} \; = \; v - 1, \tag{5.15}$$

and

$$\mathbf{A}^{(2)} \quad \text{and} \quad b\mathbf{A}^{(1)} - (b-1)\mathbf{A}^{(2)} \quad \text{are positive semi-definite.} \tag{5.16}$$

For small n, the exact permutation distribution of $\mathcal{L}_N^*$ can be obtained by direct enumeration, but the task becomes prohibitively laborious as n increases. However, as in Sen (1971a), we claim that under the null hypothesis, the permutational (conditional) distribution of $\mathcal{L}_N^*$ as well as the unconditional distribution can be approximated by the central chi squared distribution with $q(v-1)$ DF. For studies of asymptotic power and relative efficiency properties of such aligned rank tests in IBD's, we may refer to Sen (1971a). In the univariate case (i.e., for $q = 1$), for balanced incomplete block designs, the ARE results for intra-block rank tests were studied by Elteren and Noether (1959). In general, aligned rank tests fare better for IBD's, particularly when k is small. It has been observed in this context that aligned rank tests are also robust to possible heteroscedasticity of the joint error distributions from replicate to replicate, a case that may often arise in practice when the replicates are not so homogeneous in a statistical sense. In the nonparametric case, so long as the linearity of the model can be assumed (but the errors need not be normally distributed), the same block totals can be put into an alignment scheme for generating aligned rank tests which provides additional information on the hypothesis testing problem. These details can be sketched as in Sen (1971a), although it would be more advantageous for us to report on this nonparametric recovery of interblock information in a comparatively more general setup in clinical trials in a later section. As such, we omit the details here. We conclude this section with a note that robust R-estimation of treatment effects (contrasts) in IBD's can be formulated very much in the sameway as in Section 3 [viz., (3.18) through (3.23)]. In an univariate setup, this was done by Greenberg (1966) and Puri and Sen (1967), while by virtue of the comments on R-estimation in the multivariate case made in Section 4, these findings extend readily to the general model treated in this section. Therefore, we omit these details.

6 Nonparametrics in Factorial Designs

Nonparametric ANOVA, MANOVA and (M)ANOCOVA models presented in the preceding sections relate mostly to the case where treatmentwise there is a one-way layout, although incorporating blockwise variations, it may be a two-factor model. There are many situations where the treatments represent the combinations (at two or more levels) of two or more factors, so that we may not only be interested in their *main effects* but also in their possible *interactions*. In a normal theory model, such interaction effects and main effects all can be handled by suitable (linear) transformations on the original response

26

variables, and similar test statistics can be used to test for plausible null hypothesis of no interaction or no main effects. The situation is more complex in the nonparametric case. The primary impasse stems out of the fact that whereas the least squares methodology addresses well the invariance under affine transformations on the response vectors, their (coordinatewise) ranks are not affine-invariant. Moreover, for testing the null hypothesis of no interaction, it may be more reasonable to assume that the main effects of the various factors may not have insignificant differences, so that they should be treated as nuisance parameters. Although some people have tried the *rank transformation* approach, referred to in Section 3, to mimic the usual ANOVA tests based on such transformed vectors, there are some serious theoretical deficiencies of such procedures in a general multi-factor experiment. The formulation of null and alternative hypotheses requires a much more restricted setup for such rank transformed data sets (hinging on strictly monotone but arbitrary nonlinear transformations), and, often, there is a big compromise on the underlying robustness aspects (which were the original motivations for favoring a nonparametric approach). For these reasons, we shall not emphasize on such rank transformations in factorial designs, and we continue exploring aligned rank procedures in such designs. We shall mainly follow the approach of Mehra and Sen (1969).

We consider the case of replicated two-factor experiments with one observations per cell. Let Y_{ijk} be the response variate for the cell (j, k) in the ith replicate, and assume that the following fixed-effects factorial model holds:

$$Y_{ijk} = \mu_i + \nu_j + \tau_k + \gamma_{jk} + \omega_{ijk}, \tag{6.1}$$

for $i = 1, \ldots, n; j = 1, \ldots, p; k = 1, \ldots, q,$, with $n \geq 2, p \geq 2, q \geq 2$. Here the μ_i relate to the replicate effects, ν_j and τ_k to the main effects for the two factors, γ_{jk} for the interaction effects of the two factors, and ω_{ijk} are the residual error components. We may set without any loss of generality

$$\sum_{i=1}^{n} \mu_i = 0, \sum_{j=1}^{p} \nu_j = 0, \sum_{k=1}^{q} \tau_k = 0, \tag{6.2}$$

and

$$\gamma_{j\cdot} = q^{-1} \sum_{k=1}^{q} \gamma_{jk} = 0, j = 1, \ldots, p; \tag{6.3}$$

$$\gamma_{\cdot k} = p{-}1 \sum_{j=1}^{p} \gamma_{jk} = 0, k = 1, \ldots, q. \tag{6.4}$$

It is further assumed that for each i, $(\omega_{i11}, \ldots, \omega_{ipq})$ have a joint d.f. G which is a symmetric function of its pq arguments, and these n $(pq\text{-})$vectors are independent. This includes the conventional assumption of i.i.d. structure of the ω_{ijk} as a particular case, and more generally, it allows each replicate error vector to have interchangeable components which may still be dependent, a case that may arise if we allow the replicate effects to be possibly stochastic, so that we would have then a mixed effects factorial model. The null hypothesis of interest is

$$H_0 : \boldsymbol{\Gamma} = ((\gamma_{jk})) = \mathbf{0}, \tag{6.5}$$

27

against alternatives that $\boldsymbol{\Gamma}$ is non-null. We would like to formulate suitable aligned rank tests for this hypothesis testing problem.

For an $m(\geq 1)$, let $\mathbf{1}_m = (1, \ldots, 1)'$, and consider the following intra-block transformations which eliminates the replicate and main effects. Let $\mathbf{Y}_i = ((Y_{ijk}))_{p \times q}$, $\boldsymbol{\Omega}_i$ be the corresponding matrix of the error components, and let

$$\mathbf{Z}_i = (\mathbf{I}_p - p^{-1}\mathbf{1}_p\mathbf{1}_p')\mathbf{Y}_i(\mathbf{I}_q - q^{-1}\mathbf{1}_q\mathbf{1}_q'), i = 1, \ldots, n; \tag{6.6}$$

$$\mathbf{E}_i = (\mathbf{I}_p - p^{-1}\mathbf{1}_p\mathbf{1}_p')\boldsymbol{\Omega}_i(\mathbf{I}_q - q^{-1}\mathbf{1}_q\mathbf{1}_q'), i = 1, \ldots, n. \tag{6.7}$$

Then from (6.4), (6.6) and (6.7), we have

$$\mathbf{Z}_i = \boldsymbol{\Gamma} + \mathbf{E}_i, \ i = 1, \ldots, n. \tag{6.8}$$

So that on this transformed model, the nuisance parameters are all eliminated. Note that the assumed interchangeability condition on the intra-block error components implies that for each $i(= 1, \ldots, n)$, the components of $\mathbf{E}_i$ remain interchangeable too. This provides the access to using permutationally distribution-free procedures based on the stochastic matrices $\mathbf{Z}_i, i = 1, \ldots, n$. On the other hand, the $\mathbf{E}_i$ satisfy the same restrains as in (6.3) and (6.4), so that there are effectively only $(p-1)(q-1)$ linearly independent components among the pq ones (for each i).

It follows from (6.7) and the assumed interchangeability of the elements of $\boldsymbol{\Omega}_i$ that the joint distribution of $\mathbf{E}_i$ remains invariant under any of the possible $p!$ permutations of its columns , and also under any of the possible $q!$ permutations of its rows. Thus, there is a finite group G of $(p!q!)$ permutations which maps the sample space of $\mathbf{E}_i$ onto itself and leaves the joint distribution invariant, so that working with the n independent aligned error matrices, we arrive at a group $\mathcal{G}_n$ of transformations having $(p!q!)^n$ elements, and this provides the access to the exact permutation distribution of sutable test statistics based on these aligned observations. We may proceed as in Section 3 with intra-block rankings of these aligned observations and get a robust test, although it may not generally compare favorably in terms of power with aligned rank tests based on overall rankings, justifiable on the ground that the $\mathbf{Z}_i$ do not contain any block effect.

Let R_{ijk} be the rank of Z_{ijk} among the $N = npq$ aligned observations $Z_{suv}, s = 1, \ldots, n; u = 1, \ldots, p; v = 1, \ldots, q$, and define the scores $a_N(r), r = 1, \ldots, N$ as in Section 3. For notational simplicity, we let $\eta_{ijk} = a_N(R_{ijk}), i = 1, \ldots, n; j = 1, \ldots, p; k = 1, \ldots, q$ and let

$$\eta_{ij.} = q^{-1}\sum_{k=1}^{q}\eta_{ijk}, \ j = 1, \ldots, p; \tag{6.9}$$

$$\eta_{i.k} = p^{-1}\sum_{j=1}^{p}\eta_{ijk}, \ k = 1, \ldots, q; \tag{6.10}$$

$$\eta_{i..} = (pq)^{-1}\sum_{j=1}^{p}\sum_{k=1}^{q}\eta_{ijk}, \tag{6.11}$$

for $i = 1, \ldots, n$, and let $\eta_{\cdots} = n^{-1} \sum_{i=1}^{n} \eta_{i \cdots}$. Define the aligned rank statistics as

$$\mathbf{L}_{N,jk} = n^{-1} \sum_{i=1}^{n} \eta_{ijk}; \quad \mathbf{L}_N = ((L_{N,jk})). \tag{6.12}$$

Then the rank-adjusted statistics are defined by

$$\mathbf{L}_N^* = (\mathbf{I}_p - p^{-1} \mathbf{1}_p \mathbf{1}_p') \mathbf{L}_N (\mathbf{I}_q - q^{-1} \mathbf{1}_q \mathbf{1}_q') = ((L_{Njk}^*)), say. \tag{6.13}$$

Let us also define the rank measure of dispersion:

$$V_n = [n(p-1)(q-1)]^{-1} \sum_{i=1}^{n} \sum_{j=1}^{p} \sum_{k=1}^{q} (\eta_{ijk} - \eta_{ij\cdot} - \eta_{i\cdot k} + \eta_{i\cdots})^2. \tag{6.14}$$

Then, as in Mehra and Sen (1969), we consider the following test statistic:

$$\mathcal{L}_N^* = [n/V_N] \sum_{j=1}^{p} \sum_{k=1}^{q} \{L_{Njk}^*\}^2, \tag{6.15}$$

which is analogous to the classical parametric test statistic based on the variance ratio criterion. It may be appropriate here to mention that as in the case of two-way layouts, if we have a mixed-effects model, where the treatment effects and their interactions are fixed effects, while the block effects are stochastic, the alignment process eliminates the block-effects (fixed or not), and hence, aligned rank tests are usable for such mixed-effects models too. At this stage, it may be appropriate to point out the basic difference between the current alignment procedure and an alternative one, the rank transformation procedure. In the latter case, one simply replaces the original Y_{ijk} by their ranks (within the overall set) and performs the usual ANOVA test for interactions based on such rank matrices. Basically, rank transformations relate to the sample counterpart of the classical *probability integral* transformation. For a contnuous d.f. F, the latter is continuous, but still the former is a step function. Moreover, the latter is a bounded and typically nonlinear (monotone) function, so that the original linear model fitted to the Y_{ijk} may not fit to their transformed counterparts $F(Y_{ijk})$, when F is highly nonlinear. Thus, even if the block effects are eliminated by intra-block transformations, such nonlinearity effects are present in the foundation of rank transformations, and this makes them generally much less adoptable in factorial designs. In particular, if the main effects are not null, their latent effects in the rank transformation procedure may cause serious problems with respect to the validity and efficiency criteria. The aligned ranking procedure sketched here is free from this drawback as long as the basic linearity of the model in (6.1) is tenable.

For small values of n, p and q, the exact (conditional) permutational distribution of $\mathcal{L}_N^*$ can be obtained by considering the $(p!q!)^n$ (conditionally) equally likely row and column permutations of the matrices $\mathbf{H}_1 = ((\eta_{ijk})), i = 1, \ldots, n$, and as this process becomes unpracticable for large n, we appeal to the following large sample result: As n increases, the permutational (conditional) as well as the unconditional null distribution of $\mathcal{L}_N^*$ can be approximated by the central chi squared distribution with $(p-1)(q-1)$ DF. For various asymptotic properties we may refer to Mehra and Sen (1969). The procedure

extends readily to more than two-factor designs, and all we have to do is to define the aligned observations first to eliminate the nuisance parameters, and on such aligned observations we need to incorporate appropriate groups of transformations preserving invariance of their joint distributions, and with respect to such a group, we can obtain the permutational (rank) measure of dispersion. This provides the access to constructing variance-ratio type statistics based on such aligned rank statistics. This prescription is of sufficient general form so as to include the general class of IBD's treated in Section 5. Moreover, the results discussed here for univariate response variates percolate through general *incomplete multiresponse designs (IMD)* pertaining to clinical trials and medical studies [viz., Sen (1994a)]. At this stage we may refer to rank transformations as have been advocated by a host of researchers. However, one has to keep in mind that the scope of such procedures for blocked designs may be considerably less than the aligned ones presented here.

In practice, replicated $m(\geq 2)$-factor experimental designs crop up in a variety of ways, and in this setup, often, each of these factors is adapted at two levels, say 1 and 2. This way, we are led to a class of *n replcated 2^m factorial experiments*. For such designs, a similar ranking after alignment procedure, due to Sen (1970b), works out well. Let $\mathbf{j} = (j_1, \ldots, j_m)$ represent the combination of the *levels* $j_1, \ldots, j_m$ of m factors $(A_1, \ldots, A_m)$, where $j_k = 1, 2$, for $k = 1, \ldots, m$. We denote by $\mathbf{J}$ the set of all (2^m) realizations of $\mathbf{j}$. For the ith replicate, the response of the plot receiving the treatment combination $\mathbf{j}$ is denoted by $X_{i\mathbf{j}}$, and we consider the usual linear model (sans the normality of the errors):

$$X_{i\mathbf{j}} = \beta_i + [\sum_{\mathbf{r} \in \mathbf{R}} (-1)^{<\mathbf{j}, \mathbf{r}>} \tau_{\mathbf{r}}]/2 + e_{i\mathbf{j}}, \; \mathbf{j} \in \mathbf{J}, \; i = 1, \ldots, n, \tag{6.16}$$

where $< \mathbf{a}, \mathbf{b} > = \mathbf{a}'\mathbf{b}$, the β_i represent the block effects, the $e_{i\mathbf{j}}$ are the error components, $\mathbf{r} = (r_1, \ldots, r_m)'$ with each r_j either 0 or 1, $\mathbf{R}$ is the set of all possible (2^m) realizations of $\mathbf{r}$, and the treatment effects $\tau_{\mathbf{r}}$ are defined as follows.

$$\tau_{\mathbf{r}} = \tau_{A_1^{r_1} \ldots A_m^{r_m}}, \; \mathbf{r} \neq \mathbf{0}; \; \tau_{\mathbf{0}} = 0, \tag{6.17}$$

where for each $j(= 1, \ldots, m)$, $A_j^0 = 0$. Thus, $\tau_{A_1} = \tau_{1,0,\ldots,0}, \ldots, \tau_{A_m} = \tau_{0,\ldots,0,1}$ represent the main effects, $\tau_{A_1, A_2} = \tau_{1,1,0}$ etc. represent a two-factor interaction, and so on; $\tau_{\mathbf{r}}$ is a k-factor interaction effect if $< \mathbf{r}, \mathbf{1} > = k$, for $k = 1, \ldots, m$. As in earlier sections, we assume here that for each $i(= 1, \ldots, n)$, the set $\{e_{i\mathbf{j}}, \mathbf{j} \in \mathbf{J}\}$ consists of interchangeable r.v.'s, and the block-effects need not be fixed; they may as well be stochasic. Let now $\mathbf{P}$ be a subset of $\mathbf{R}$, and suppose that we want to test the null hypothesis

$$H_{0,\mathbf{P}} : \{\tau_{\mathbf{r}}, \mathbf{r} \in \mathbf{P}\} = \mathbf{0}, \tag{6.18}$$

against the set of alternatives that these effects are not all equal to 0.

Since (6.16) involves the block effects as nuisance parameters (or spurious r.v.'s), by means of the following intra-block transformations, we obtain the aligned observations. These aligned observations

provide both the least squares and R-estimators of the $\tau_{\mathbf{r}}$. Let

$$t_{i,\mathbf{r}} = 2^{-(m-1)} \sum_{\mathbf{j} \in b f J} (-1)^{<\mathbf{j}, \mathbf{r}>} X_{i\mathbf{j}}, \quad \mathbf{r} \in \mathbf{R}, \quad i = 1, \ldots, n. \tag{6.19}$$

Then we may write $t_{i,\mathbf{r}} = \tau_{\mathbf{r}} + g_{i,\mathbf{r}}$, for every $\mathbf{r} \in \mathbf{R}$, where the $g_{i,\mathbf{r}}$ are the corresponding aligned error components. It is easy to verify that these $g_{i,\mathbf{r}}$ remain exchangeable r.v.'s too, within each block. Moreover, it has been shown by Sen (1970b) by simple arguments that univariate d.f.'s for these aligned errors are all symmetric about zero, and all their bivariate d.f.'s are diagonally symmetric about 0. Actually, the joint distribution of these aligned errors (within each block) is also diagonally symmetric about 0. Thus, for the R-estimation of the $\tau_{\mathbf{r}}$, we may use the (marigal) set $t_{i,\mathbf{r}}, i = 1, \ldots, n$, and as in Section 2 [see (2.25)], incorporate a general signed rank statistic to yield the desired estimator. These are based on i.i.d.r.v.'s, and hence, no residuals are needed to reconstruct the estimators. As regards rank tests for the null hypothesis in (6.18), we may consider the n i.i.d.r. vectors

$$(t_{i,\mathbf{r}}, \mathbf{r} \in \mathbf{P}), \quad i = 1, \ldots, n, \tag{6.20}$$

and use multivariate signed-rank test statistics, displayed in detail in Chapter 4 of Puri and Sen (1971). Asymptotic properties of such tests, studied in detail there, remain in tact for such aligned rank tests in 2^m factorial experiments. Extensions to *confounded* or *partially confounded* designs have also been covered in Sen (1970b).

7 Paired Comparisons Designs: Nonparametrics

In order to compare a number (say, $t(\geq 2)$) of objects which are presented in pairs to a set of (say, $n(\geq 2)$) judges who verdict (independently) a relative preference of one over the other within each pair, the method of *paired comparisons (PC)*, developed mostly by the psychologists, allows one to draw statistical conclusions on the relative positions of all the objects. *Paired comparisons designs (PCD)* are thus incomplete block designs with blocks of size two and a dichotomous response on the ordering of the intra-block plot yields. There are several detours from this simple description of PCD. For example, it may be possible to have observable responses (continuous variates) for each pair of objects: This will relate to the classical IMD with two plots in each block, so that, the results developed in earlier sections would be appliocable here. Hence, we skip these details. Another route relates to *paired characteristics* so that ordering of the two objects within each pair may have four possible outcomes (instead of the two in the case of a single characteristic). Nonparametrics for such paired comparisons for paired characteristics were developed by Sen and David (1968) and Davidson and Bradley (1969,1970),among others. A general account of such PCD methodology is given in David (1988) where other references are also cited. A general characteristics of such paired comparisons procedures is that *circular triads* may arise in a natural way, and this may lead to *intransitiveness* of statistical inference tools when viewed from

a *decision theoretics* point; the problem becomes even more complex in a multivariate setup. However, following David (1988) we may say that it is a valuable feature of the method of paired comparisons that it allows such contradictions to show themselves..., and hence, the methodology developed addresses this isuue in a sound statistical manner. As in earlier sections, it is also possible to work out the (M)ANOVA and (M)ANOCOVA models side by side, and following Sen (1995b), we summarize the main results along the same vein.

Paired comparisons procedures in a multivariate setup rest on suitable representations of probability laws for multiple dichotomous attributes. Let us consider $p(\geq 1)$ dichotomous attributes, and let $\mathbf{i} = (i_1, \ldots, i_p)'$, where each i_j can take only two values 0 and 1, for $j = 1, \ldots, p$. The totality of all such 2^p realizations of $\mathbf{i}$ is denoted by the set I, and consider a stochastic p-vector $\mathbf{X} = (X_1, \ldots, X_p)'$, such that

$$P\{\mathbf{X} = \mathbf{i}\} = \pi(\mathbf{i}), \quad \mathbf{i} \in \mathcal{I}. \tag{7.1}$$

This probability law is defined on a 2^p-simplex $\Pi = \{\pi(\mathbf{i}) \geq 0, \forall \mathbf{i} \in \mathcal{I}; \sum_{\mathbf{i} \in \mathcal{I}} \pi(\mathbf{i}) = 1\}$, so that there are $2^p - 1$ linearly independent elements in Π. Since there ate t objects (forming $\binom{t}{2}$ pairs), the total number of linearly independent parameters is equal to $\{2^p - 1\} \binom{t}{2}$, and this is generally large when t and /or p is not small. We consider the following modification of the Bahadur (1961) representation for multiple dichotomous attributes. Let

$$\pi_{i*}^{(j)} = P\{X_j = i\}, \quad i = 0, 1; 1 \leq j \leq p. \tag{7.2}$$

We denote by $\theta_j = \pi_{0*}^{(j)}, j = 1, \ldots, p$. Also, for every $l : 2 \leq l \leq p; 1 \leq i_1 < \cdots < i_l \leq p$, define an lth order *association parameter*

$$\theta_{i_1 \cdots i_l} = \theta_{i_1} \cdots i_l(\Pi), \tag{7.3}$$

where there are $\binom{p}{l}$ such parameters, for $l = 2, \ldots, p$. Taking into account the set of $\theta's$, marginal and association parameters, we exhaust the totality of $2^p - 1$ linearly independent parameters. We denote this set by $\Theta = \{\theta_{i_1 \cdots i_l}, 1 \leq i_1 < \ldots < i_l \leq p; 1 \leq l \leq p\}$, and arrive at the following.

$$
\begin{aligned}
\pi(\mathbf{i}) \;=\; & \prod_{j=1}^{p} \pi_{i_j*}^{(j)} + \sum_{1 \leq j_1 < j_2 \leq p} (-1)^{i_{j_1}+i_{j_2}} \theta_{j_1 j_2} \prod_{r=1}^{2} \theta_{j_r} \prod_{s=1, \neq j_1, j_2}^{p} \pi_{i_s*}^{s} \\
& + \sum_{1 \leq j_1 < j_2 < j_3 \leq p} (-1)^{i_{j_1}+i_{j_2}+i_{j_3}} \theta_{j_1 j_2 j_3} \prod_{r=1}^{3} \theta_{j_r} \prod_{s=1, \neq j_1, j_2, j_3}^{p} \pi_{i_s*}^{s} \\
& + \cdots + (-1)^{i_1 + \cdots + i_p} \theta_{1 \cdots p} \prod_{r=1}^{p} \theta_r.
\end{aligned}
\tag{7.4}
$$

The PCD models are atuned to such representations. By reference to the PCD model under consideration, for the pair (i, j), we denote the response vector by $\mathbf{X}'_{ij} = (X_{ij}^{(1)}, \cdots, X_{ij}^{(p)})$ (each coordinate being a dichotomous variable), and the probability law of $\mathbf{X}_{ij}$ over the 2^p-simplex is denoted by Π_{ij}, and its transformation to Θ is denoted by Θ_{ij}, for $1 \leq i < j \leq t$. Basically, PC methodology relates to

comparing these $\binom{t}{2}$ probability laws inducing a reduction of the parameter space to a subset of the Θ_{ij} by an appeal to the Bahadur representation in (7.4). Thus, the basic null hypothesis of interest is

$$H_o : \Theta_{ij} = \Theta^o(\text{unknown}), \forall 1 \leq i < j \leq t, \tag{7.5}$$

against the set of alternatives that the Θ_{ij} are not all the same. To motivate suitable nonparametric testing procedures, we make an appeal to the classical Bradley-Terry (1952) model in the univariate case, which has also been extended to the multivariate case by Davidson and Bradley (1969,1970) and others. In the univariate case, dropping the superscript (j) in (7.2), we denote the corresponding probabilities for the $\binom{t}{2}$ pairs by $\pi_{ij}, 1 \leq i < j \leq t$. Then, conceive of a set $\{\alpha_1, \cdots, \alpha_t\}$ of positive numbers, such that $\sum_{j=1}^{t} \alpha_j = 1$, and write

$$\pi_{ij} = \frac{\alpha_i}{\alpha_i + \alpha_j}, 1 \leq i < j \leq t. \tag{7.6}$$

Thus, the set of $\binom{t}{2}$ unknown (probabilities) parameters is expressed in terms of $t - 1$ unknown α's, and the null hypothesis can equivalently be expressed as $H^0 : \alpha_1 = \cdots = \alpha_t = t^{-1}$. With this formulation, the (nonparametric) MLE of the α can be obtained from a given data set and can then be incorporated in the usual (likelihood ratio-)type tests for the null hypothesis of homogeneity of the α. In the multivariate case, for each of the p coordinates, we would have such α-parameters (in all $p(t - 1)$ in number), and homogeneity of these $(p$-)vectors $\alpha_i, i = 1, \ldots, t$ constitutes the null hypothesis of interest. In this formulation of the hypotheses, the association parameters (i.e., the Θ_{ij}) are to be treated as nuisance parameters, and usually they are assumed to be homogeneous. Davidson and Bradley (1969) in their formulation assumed that any thrid or higher order association parameter is null, and incorporated the classical likelihood ratio principle to formulate some large sample PC tests in a multivariate setup. Sen (1995b) has shown that part of this assumption may not be that crucial, and also the results extend directly to the MANOCOVA model. As in the case of multivariate nonparametric procedures, treated earlier, such tests may no longer be genuinely distribution-free, and hence, suitable permutational invariance structures are to be exploited to render them as permutationally (conditionally) distribution-free; such procedures have been considered by Sen and David (1968) and Sen (1995b), among others. We therefore summarize here these results in a general setup, and indicate the simplifications for simpler models.

In a MANOCOVA setup, we partition the p responses into two subsets: *primary* responses, $p_1(\geq 1)$ in number, and *concomitant* responses, $p_2(\geq 1)$ in number; $p = p_1 + p_2$. Consider then the component null hypotheses: $H_{or} : \alpha_{r1} = \cdots = \alpha_{rp} = t^{-1}$, for $r = 1, \ldots, p$. The intersection of the p_1 null hypotheses $H_{or}, r = 1, \ldots, p_1$ is denoted by H_{o1}^*, and similarly, H_{o2}^* denotes the intersection of the p_2 component null hypotheses for the concomitant responses. Then the MANOCOVAPC null hypothesis can be stated as

$$H_0^* = H_{o1}^* \mid H_{o2}^*. \tag{7.7}$$

For every pair $(j, l) : 1 \leq j < l \leq t$, we denote the observed cell frequencies by $n_{jl}(\mathbf{i})$, where $\mathbf{i}$ ranges

over the set I, and the total number of observations for this pair is n_{jl}. The marginal frequencies are denoted by $n_{jl,r}(i_r)$, for $i_r = 0,1; r = 1,\ldots,p$, and the bivariate marginals by $n_{jl,rs}(i_r, i_s)$, for $i_r = 0,1, i_s = 0,1; r \neq s = 1,\ldots,p$. Then as in Sen and David (1968), we may obtain some partial maximum likelihood estimator (PMLE) of the association parameters as follows.

$$\hat{\theta}_{n,rs} = n^{-1} \sum_{1 \leq j < l \leq t} (n_{jl,rs}(00) + n_{jl,rs}(11) - n_{jl,rs}(01) - n_{jl,rs}(10)), \tag{7.8}$$

for $r \neq s = 1,\ldots,p$. Conventionally, we let $\hat{\theta}_{rr} = 1$, for $r = 1,\ldots,p$, and consider the following matrix, partitioned appropriately:

$$\hat{\boldsymbol{\Theta}}_n = ((\hat{\theta}_{n,rs})) = \left(\left(\hat{\boldsymbol{\Theta}}_{n,jl} \right) \right)_{j,l=1,2}, \tag{7.9}$$

where the partitioned matrices are of the order $p_j \times p_l$, for $j,l = 1,2$. Let us also introduce the statistics

$$T_{n,jr} = \sum_{l=1,\neq j}^{t} n_{jl}^{-1/2} [n_{jl,r}(0) - n_{jl,r}(1)], \quad j = 1,\ldots,t; \; r = 1,\ldots,p, \tag{7.10}$$

and write

$$\mathbf{T}_{n,j} = (T_{n,j1},\ldots,T_{n,jp})', \quad j = 1,\ldots,t. \tag{7.11}$$

For the MANOVA test, as in Sen and David (1968) and Sen (1995), we consider the following test statistic:

$$\mathcal{L}_n = t^{-1} \sum_{j=1}^{t} \mathbf{T}'_{n,j} \hat{\boldsymbol{\Theta}}_n^{-1} \mathbf{T}_{n,j}. \tag{7.12}$$

In the univariate case (i.e., for $p = 1$), $\mathcal{L}_n$ is exactly distribution-free under the null hypothesis; for $p > 1$, this EDF property may not generally hold, but permutational (conditional) distribution-freeness holds. Thus, for small sample sizes, a finite group of 2^n possible sign-inversions can be incorporated to generate the permutational distribution of $\mathcal{L}_n$. This procedure has been elaborated in Sen and David (1968) in the bivariate case, and a very similar picture holds for general multivariate paired comparisons models. For large sample sizes, the null hypothesis distribution of $\mathcal{L}_n$ can be well approximated by the central chi squared distribution with $t-1$ DF. For local alternatives, noncentral chi-square approximations also hold.

Let us consider the MANOCOVAPC model in the same setup as explained before. We confine ourselves only to the set of (p_2) concomitant traits, and based on these responses, we construct a MANOVAPC test statistic in the same manner as in (7.12). Let us denote this test statistic by $\mathcal{L}_{n2}$. Under H_{02}, $\mathcal{L}_{n2}$ is permutationally (conditionally) distribution-free, and asymptotically, it has the chi squared distribution with $(t-1)p_2$ DF. Then for testing the null hypothesis H_{01} assuming that H_{02} holds, we consider the test statistic

$$\mathcal{L}_{n1}^* = \mathcal{L}_n - \mathcal{L}_{n2}. \tag{7.13}$$

It may be noted that $\mathcal{L}_{n1}^*$ may also be expressed in terms of concomitant variate adjusted $\mathbf{T}_{nj}$ and a similarly adjusted covariance matrix; for some details, we may refer to Sen (1995). It follows from

the above discussion that under the MANOCOVA model null hypothesis, $\mathcal{L}_{n1}^*$ is also permutationally (conditionally) distribution-free. Moreover, under the same null hypothesis, it has asymptotically the central chi squared distribution with $(t-1)p_1$ DF. Asymptotic nonnull distribution theory (for local alternatives) for such MANOCOVAPC tests has been studied in detail by Sen (1995b), and it has been shown that the ARE of this MANOCOVA test with respect to the corresponding MANOVA test (ignoring the p_2 concomitant traits), in a general multivariate setup, is bounded from below by 1. Thus, at least asymptotically, the MANOCOVAPC tests are better alternatives than the corresponding MANOVA PC tests. A similar picture holds for the ARE of the MANOCOVAPC test with respect to the MANOVAPC test based on all the p traits (when, in fact, there are p_2 concomitant traits). The intuitive reason for this better ARE picture is that the concomitant traits do not contribute to the growth of the noncentrality of the MANOVAPC test (based on the entire set), so that a larger DF with a common noncentrality parameter leads to a decrease in the (asymptotic) power function.

8 Nonparametrics for Crossover Designs

In the context of clinical trials and/or biomedical studies, each experimental unit (or subject) receives several treatments at different time-periods, so that there is a *repeated measurement design (RMD)* flavor in the statistical modeling and analysis schemes pertaining to such experimental plans. The simplest situation relates to a two-period design where for some subjects two treatments are administered exclusively in these two periods in a specific order, while for others, it is done in the reverse order. For this reason, it is also called a *changeover* or *crossover design*. A basic feature of such RMD's is that *residual effects* or *carryover effects* are likely to be a vital part of the response pattern, and hence, in the modeling, such effects are to be incorporated in an appropriate manner. The more generality one may want to achieve in this formulation, the more complex may be the actual statistical modeling and analysis schemes, and hence, often, it is assumed that such residual effects are additive and have some structural form. While most of such technicalities are discussed in some other chapters of this volume, we may like to introduce only some specific crossover designs, and examine in that context how far nonparamterics can be accepted as an alternative to standard normal theory parametrics.

Let us consider a $p(\geq 2)$ period model wherein n experimental units are used, in such a way that for the jth unit, in the ith period, treatment $d(i,j)$ is used, where the $d(i,j)$ belong to an index set relating to the treatments administered in the experiment. We denote the response of the jth unit at the ith period by Y_{ij}, and consider the conventional linear model incorporating the *first order carryover* effects $\rho(d(i-1,j))$:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \tau_{d(i,j)} + \rho_{d(i-1,j)} + \epsilon_{ij}, \tag{8.1}$$

for $i = 1, \ldots, p; j = 1, \ldots, n$, where μ is the mean effect, α_i are the period effects, β_j are the unit effects, $\tau_{d(i,j)}$ are the treatment effects, $\rho(d(0,j)) = 0$, and the errors ϵ_{ij} are assumed to be i.i.d.r.v.'s with zero

mean and a finite positive variance σ^2. In the normal theory model, again their error distribution is assumed to be normal. This is the so called *fixed effects model with first order residual effects* and it corresponds to the familiar completely randomized design in the conventional case. Motivated by the two-way layouts discussed earlier, we may also extend the model to randomized block designs where the experimental units may be blocked into relatively homogeneous groups, and for each group, we have a model as in above. Thus, there is a need to introduce block effects as well as interaction parameters with respect to block vs. carryover parameters. In this context, it may be remarked that in clinical trials or biological assays and other biomedical experiments, the very mechanism by which the blocks are formulated, these block effects may be random, and, in turn, the *carryover×block* interactions may also be random. Therefore, one may encounter a so called *mixed effects* model, which may be presented as:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \xi_{jk} + \tau_{d(i,j,k)} + \rho_{d(i-1,j,k)} + \epsilon_{ijk}, \tag{8.2}$$

where i stand for the experimental unit, j for the period, and k for the block, with the parameters defined accordingly. A simplified version of this mixed-effects model was considered by Grizzle (1965). In the above formualtion when we treat the ξ_{jk} as random, it is quite likely that the residual or carryover effects $\rho_{d(i-1,j,k)}$ are also to be treated as stochastic. In the normal theory model, all these stochastic elements are assumed to be independent and normally distributed with zero means and appropriate (unknown, positive) variances, so that the model can be interpreted in the light of conventional *variance components models*. Although such an independence assumption may not be that unrealistic in a practical application, the assumption of normality may, however, not be tenable in a variety of situations. Sans this normality assumption, the classical parametric procedures may lose their appeal on theoretical (viz., *optimality*) as well as practical *robustness*) grounds, and hence, there is a general feeling that nonparametric and robust statistical procedures are to be advocated in this context. The basic idea is to incorporate the *alignment* principle as far as possible, so as to reduce the number of estimable parameters and error components, so that classical multivariate nonparametric and robust methods discussed in earlier sections can be implemented successfully. This alignment principle is isomorphic to the one in the classical normal theory models. From more elementary practical considerations, Koch (1972) initiated the use of some nonparametric methods in the statistical analysis of two-period-change-over design with emphasis on applications, and Tudor and Koch (1994) have a recent review of applied works in this field, which cast additional light on related applicational developments. As such, instead of providing a general but abstract formulaion, we shall try to motivate the basic ideas with specific simplifications.

Let there be a complete block design with p periods (indexed as $i = 1, \ldots, p$), b blocks (indexed as $j = 1, \ldots, b$), and m units in each block (indexed as $k = 1, \ldots, m$). Consider a contrast in the p measurements on the kth unit in the jth block :

$$Y_{jk}^* = \sum_{i=1}^{p} l_i Y_{ijk}, \text{ for } j = 1, \ldots, b, \ k = 1, \ldots, m, \tag{8.3}$$

where $\sum_{i=1}^{p} l_i = 0$. Then, by (8.2) and (8.3), we obtain that

$$Y_{jk}^* = \sum_{i=1}^{p} l_i \{\alpha_i + \tau_{d(i,j,k)} + \rho_{d(i-1,j,k)}\} + \sum_{i=1}^{p} l_i \epsilon_{ijk} = \theta_{jk} + \epsilon_{jk}^*, \text{ say.} \tag{8.4}$$

In this formulation, sans normality of the errors ϵ_{ijk}, the aligned ϵ_{jk}^* may have distributions dependent on the chosen l_i, and this may cause some problems. To eliminate this drawback, we proceed as in the case of aligned rank procedures treated earlier, and define the set of aligned observations as

$$Y_{ijk}^o = Y_{ijk} - p^{-1} \sum_{i=1}^{p} Y_{ijk}, \ i = 1, \ldots, p, \tag{8.5}$$

for every $j = 1, \ldots, b; \ k = 1, \ldots, m$. In the case of complete *balanced* (B)RMD's, $\sum_{i=1}^{p} \{\alpha_i + \tau_{d(i,j,k)} + \rho_{d(i-1,j,k)}\} = 0$, for every $j = 1, \ldots, b; \ k = 1, \ldots, m$, so that if we consider the set of intra-block intra-unit aligned observations:

$$Y_{ijk}^o = Y_{ijk} - p^{-1} \sum_{i=1}^{p} Y_{ijk}, \ k = 1, \ldots, m; \ j = 1, \ldots, b; \ i = 1, \ldots, p, \tag{8.6}$$

then for the jth block, we have a set of m independent p-vectors $\mathbf{Y}_{jk}^o = (Y_{1jk}^o, \ldots, Y_{pjk}^o)'$, for $k = 1, \ldots, m$, and for different blocks too, these aligned vectors are stochastically independent. We denote the collection of m vectors $\mathbf{Y}_{jk}^o, k = 1, \ldots, m$ in the jth block by $\mathbf{Y}_j^o$ and the combined collection by $\mathbf{Y}_{..}^o$. In a similar manner, we define the aligned error vectors $\mathbf{e}_{jk}^o$, for $k = 1, \ldots, m; j = 1, \ldots, b$. Further, for each j, k, we let $\boldsymbol{\Theta}_{jk} = (\theta_{1jk}, \ldots, \theta_{pjk})'$, where $\theta_{ijk} = \alpha_i + \tau_{d(i,j,k)} + \rho_{d(i-1,j,k)}$, for $i = 1, \ldots, p; \ j = 1, \ldots, b; \ k = 1, \ldots, m$. Let $n = km$ and $N = kpm$. Then we consider the aligned observation matrix $\mathbf{Y}^o$ of order $p \times n$, expressed as $(\mathbf{Y}_{11}^o, \ldots, \mathbf{Y}_{bm}^o)$, and a similar representation is made for $\mathbf{e}^o$ and $\boldsymbol{\Theta}^o$. As such, we write

$$\mathbf{Y}^o = \boldsymbol{\Theta}^o + \mathbf{e}^o. \tag{8.7}$$

This representation enables us to incorporate the general theory and methodology of multivariate nonparametrics, discussed in Section 4. In this context, we may remark that for each j, k, whenever the $\epsilon_{ijk}, 1 \leq i \leq p$ are interchangeable r.v.'s (a condition implied by the usual assumption that they are i.i.d.), the $e_{ijk}, 1 \leq i \leq p$ are also interchangeable, and for different j, k, these stochastic vectors are independent. Thus, for these aligned error vectors, we have the same exchangeability assumption as in the classical nonparametric (M-)ANOVA model. This intuitively suggests that aligned rank procedures for the MANOVA model, considered in Section 4, can be incorporated in the current context too. There is an additional simplification in this setup. The marginal d.f. of each e_{ijk}^o is the same, so that while ranking these aligned observations, we do not have to rank the elements separately for each of the p rows in the matrix $\mathbf{Y}^o$; rather, we consider the overall ranking of all the N aligned observations. In order to achieve this simplification, we need, however, to check a basic condition that estimable parameters among the sets of period-effects, treatment-effects and carryover-effects can be expressed in terms of $\boldsymbol{\Theta}^o$-contrasts. This can easily be verified in the case of balanced RMD's, while for general unbalanced RMD's, we may set appropriate design restraints which would ensure the same.

The aligned observation matrix $\mathbf{Y}^o$ plays also a vital role in robust estimation and testing procedures for RMD's. In the normal theory model, if we confine ourselves to the fixed-effects case in (8.1), then the classical weighted least squares estimation (WLSE) methodology can be incorporated to characterize the optimality of estimators and tests based on this aligned observation matrix. For the mixed-effects model in (8.2), the situation is somewhat more complex due to the random block and interaction effects, and the assumption of independence as well as normality of these components seems to be even more vulnerable in actual practical applications. On the other hand, a characterization of the optimality of the WLSE is, of course, limited to the basic assumption that all the stochastic elements in model (8.2) are normally distributed and their independence-homoscedasticity condition holds. Any departure from such model-assumptions can not only take away the optimality properties but also may signal lack of validitty. As such, *robust methods* for drawing statistical conclusions are quite appealing in this context. While the aligned rank based estimates and tests are generally globally robust, they may not be (even asymptotically) fully efficient when the form of the underlying error distribution is not known. In a local robustness perspective, if only small departures from the assumed model are contemplated, it may be more appropriate to use *M-estimators* and related *M-tests* based on appropriate score functions. Generally, such score functions are smooth but non-linear, and based on consideration of robustness, they are usually bounded and monotone. For some detail discussion of such robust procedures for general linear models, we may refer to Hampel et al. (1986) and Jurečková and Sen (1995), among others. o

9 Clinical Trials and Survival Analysis

In clinical trials and life testing problems, although the setups are related to classical statistical designs, general objectives and operational constraints, often, call for different types of designs and appropriate statistical analysis schemes. Semi-parametrics and nonpametrics are more appealing than standard parametrics in such designs. We illustrate this with a very simple life-testing model, which may as well be adapted to a clinical trial setup. Suppose that we want to study the impact of *smoking* on longivity of human beings. Usually, before a clinical trial is initiated on a specific human sector, it is planned to have an *animal study* to study safe dosage, side effects and other causal effects, so that the actual clinical trial can be administered with less restraints from medical ethics and other humanatarian points. Suppose that two groups of monkeys are chosen, one for the *placebo* (no smoking) and the other for the *treatment* (smoking) group. After a period of study, say eighteen months, these monkeys are sacrificed, and their arterial cross-section at some specified location is examined for the constriction of the arterial channel. Thus, for this *arteriosclerosis* problem, the response variable is the ratio of the open space to the entire cross-section. It is hypothesized that smoking tends to make this ratio stochastically smaller, so that essentially we have a two sample model for testing homogeneity against a one-sided alternative. But, there are some basic differences between this life testing and the classical two-sample model. First, some

animals may die before the study period is over, resulting in *censoring*. Such censoring may also arise due to dropouts or withdrawals of the subjects due to causes other than early failures. Secondly, the duration of the study period has to be decided on the basis of extraneous information on the associated inhalation toxicity problem, and the outcome response variable may depend in a rather complex manner on this study period, in the sense that a linear or log-linear regression may not be reasonable. Thirdly, in an agricultural experiment, the treatment can be applied in a reasonably controlled manner, whereas, in such life testing models, such a controlled experimental setup is not that expected. Finally, it may not be very reasonable to assume that the response variable is exponentially or normally or lognormally distributed, so that standard parametric procedures may not be very appealing in such a context. We may refer to Sen (1984a,b) for some discussion of these aspects of multivariate nonparametrics relating to medical studies. If the study has to cover human subjects, occupational factors, sex, diet, physical exercise, age and many other concomitant variates may make the model far more complex for adoption of parametric statistical analysis schemes. *Robustness* and *validity* are therefore of important considerations in this respect.

In designing such clinical trials and/or life testing models, it may therefore be necessary to incorporate various auxiliary and concomitant variates in the model, plan judiciously on the duration of the study, identify the follow-up nature of the study, and check the appropriateness of *repeated significance tests* and/or *interim analysis* for such experimental data sets. These factors have a great bearing on the formulation of appropriate statistical analysis schemes.

Consider a simple (say, placebo vs. treatment) clinical trial or life testing model which we put in a slightly more general regression model setup. Note that in survival analysis or life testing problems, the failure times are nonnegative random variables typically with skewed distribution, and hence, log-transformations are used to induce more symmetry (if not normality) on the distribution of the transformed response variable (termed, the *response metameter*. Similarly, on the dosage, often such a transformation is used to induce more linearity of the response-dosage regression, and such a transormed dose is referred to as a *dose metameter*. With such transformations, we may have greater confidence on the linearity of regression as well as symmetry of the response distribution, although it may not be entirely satisfactory to assume normality or logistic form for such a distribution, as is typically done in parametric analysis. Thus, from robustness and validity points of view, we shall allow the response distribution to be arbitrary to a greater extent. Let there be n sujects under study, and let their (transformed) responses be denoted by $X_1, \ldots, X_n$ respectively. We assume that the X_i are independent with continuous distribution functions $F_1(x), \ldots, F_n(x), x \in \mathbf{R}$ respectively. These d.f.'s may depend on the dose levels to which the subjects are subjected, and we conceive of (nonstochastic) constants (regressors) $t_1, \ldots, t_n$ (not all equal) which can be used to formulate a semi-parametric model :

$$F_i(x) = F_o(x - \beta t_i), x \in \mathbf{R}, i = 1, \ldots, n, \tag{9.1}$$

where β stands for the regression parameter (unknown) and the d.f. F_o is assumed to be continuous but

otherwise of arbitrary form. As in Section 2, we may also consider a more general linear model [viz., (2.3)] where the t_i are known vectors of regression constants and β is a vector of regression parameters. The classical placebo vs. treatment setup is thus a special case of (9.1) where the t_i are binary. While this model is isomorphic to the classical linear regression model treated in Section 2, there is a basic difference between the two setups. There, all the X_i were assumed to be observable at the same time, while here we have follow-up scheme. Let us denote the order statistics corresponding to $X_1, \ldots, X_n$ by $X_{n:1} \leq \cdots \leq X_{n:n}$, where by virtue of the assumed continuity of F_o, ties among these observations may be neglected with probability 1. Let us also denote the *anti-ranks* by $S_1, \ldots, S_n$, so that

$$X_{S_i} = X_{n:i}, \quad \text{for} \quad i = 1, \ldots, n, \tag{9.2}$$

and $(S_1, \ldots, S_n)$ is a (random) permutation of $(1, \ldots, n)$. At a time point t within the study period, the observable random elements are the failures occurring before that time and the corresponding c_{S_i} values; although the entire set of $c_1, \ldots, c_n$ is known at the beginning of the study. Thus, as we move along the study, we gather an accumulating data set

$$\{X_{n:j}, \ c_{S_j}; j \leq i\}, \quad \text{for } i = 1, \ldots, n, \tag{9.3}$$

where it may be possible only to have a subset of these elements due to censoring as well as with-drawals/dropouts. For example, during the study period a (random) subset of failures may occur, while the other subjects have failure times larger than the set endpoint of the study; this is referred to as *right truncation*. Alternatively, the study may be so planned that it would be conducted until a prespecified number ($m \leq n$) of failures take place, resulting in a stochastic duration of the study; this is referred to as *right censoring*. In either case, there is some incompleteness in the observed data set due to possible immature termination of the study. Actually, in biomedical studies, including clinical trials and epidemiological or environmental investigations, censoring may be much more complex than the ones referred to above; we may refer to Sen (1995d) for some details of the interface of statistical censoring in practice and the controversies arising in such a context. In the current context, we stick to some simple censoring patterns as have been described before. In the above setup, from medical ethics point of view, it may be reasonable to set an underlying condition that if there is a significant difference in the placebo vs. treatment group responses (in our setup, $\beta \neq 0$), then we should terminate the study as early as possible and switch all the subjects to the better group for better health prospects. For this reason, instead of waiting until the study period is over, it may be desirable to look into the accumulating data set at regular time intervals, resulting in the so called interim analysis schemes, and at each such prechosen time-point, to perform a test of significance (on β) for possible stopping of the trial, resulting in the so called repeated significance testing (RST) schemes. Such statistical analysis schemes in a broader setup are also referred to as *time-sequential procedures* (Sen (1981b)), and monitoring on a continual basis (i.e., at every failure point) is referred to as *progressively censored schemes (PCS)* (Chatterjee and Sen (1973)).

We start with the setup of (2.3)-(2.5). Let us examine the picture at the kth failure point $X_{n:k}$. We have the knowledge of the previous failure points as well as the corresponding anti-ranks $S_1, \ldots, S_k$; also, we know that the remaining observations are right-cenrored. Thus, it seems very natural to project the linear rank statistic $\mathbf{L}_n$, defined in (2.5), onto the subspace generated by the anti-ranks $S_1, \ldots, S_k$. For this let us define

$$a_{nk}^* = (n-k)^{-1} \sum_{j=k+1}^{n} a_n(j), \text{ for } k = 0, \ldots, n-1, \qquad (9.4)$$

and conventionally, we let $a_{nn}^* = 0$. Note that the ranks for the k observed failures are $1, \ldots, k$ respectively, while each of the remaining $(n-k)$ censored observation is given the average rank score a_{nk}^*, so that at the kth failure point, the censored linear rank statistic is given by

$$\begin{aligned} \mathbf{L}_{nk} &= \sum_{i \leq k} (\mathbf{t}_{S_i} - \bar{\mathbf{t}}_n) a_n(i) + a_{nk}^* \sum_{j > k} (\mathbf{t}_{S_j} - \bar{\mathbf{t}}_n) \qquad (9.5) \\ &= \sum_{i \leq k} (\mathbf{t}_{S_i} - \bar{\mathbf{t}}_n)[a_n(i) - a_{nk}^*], \text{ for } k = 1, \ldots, n, \end{aligned}$$

and conventionally, we let $\mathbf{L}_{n0} = \mathbf{0}$. We define $\mathbf{Q}_n$ as in (2.7), and in the case of scalar t_i, we denote this by Q_n^2. Let us also define $\bar{a}_n$ and A_n^2 as in (2.6), and for every $k : 0 \leq k \leq n$, we let

$$A_{nk}^2 = A_n^2 - (n-1)^{-1}(n-k)[a_{nk}^* - \bar{a}_n]^2, \qquad (9.6)$$

so that it follows by some simple arguments that

$$0 = A_{n0}^2 \leq A_{n1}^2 \leq \cdots \leq A_{nn}^2 = A_n^2. \qquad (9.7)$$

Consider now an experimental scheme wherein the study is planned to be curtailed at the rth failure point, for some prefixed positive integer $r(\leq n)$ (so that we have a Type II censoring scheme). In the case of scalar t_i, we consider the test statistic

$$Z_{nr} = \{L_{nr}\}/\{Q_n A_{nr}\}. \qquad (9.8)$$

If the null hypothesis (H_0) relates to the homogeneity of the F_i (or, equivalelently, $\beta = 0$), following the line of attack of Chatterjee and Sen (1973), we can claim that under H_0 Z_{nr} is EDF, and for large n, Z_{nr} has normal distribution with 0 mean and unit variance. Thus, an appropriate (one or two-sided) test can be based Z_{nr}. In the case of vector $\mathbf{t}_i$, parallel to (2.9), we define

$$\mathcal{L}_{nr} = A_{nr}^{-2}\{\mathbf{L}_{nr}' \mathbf{Q}_n^- \mathbf{L}_{nr}\}. \qquad (9.9)$$

Here also, under the null hypothesis of homogeneity of the d.f.'s F_i, $\mathcal{L}_{nr}$ is EDF, and proceeding as in after (2.9), we may argue that under the null hypothesis, $\mathcal{L}_{nr}$ has asymptotically central chi squared distribution with q DF when $\mathbf{Q}$ is of rank q. Therefore an appropriate test for the null hypothesis $H_0 : \beta = \mathbf{0}$ can be based on the test statistic $\mathcal{L}_{nr}$. Let us now consider the Type I censoring or truncation case. Here, for some prefixed time-point T_o, the experiment is planned for the prefixed

41

duration $(0, T_o]$. Let $r(T_o)$ be the number of failures occurring in the study period $(0, T_o]$. In this setup, $r(T_0)$ is a nonnegative integer valued random variable. It may be tempting to use the statistics $Z_{nr(T_o)}$ or $\mathcal{L}_{nr(T_o)}$ as appropriate test statistics. This is indeed possible. But, it may be kept in mind that such statistics are not generally EDF, even under the hypothesis of homogeneity of the underlying d.f.'s. However, as in Chatterjee and Sen (1973), we may argue that conditionally on $r(T_o) = r$, $\mathcal{L}_{nr(T_o)}$ or $Z_{nr(T_o)}$ is distribution-free under H_0, and they enjoy the same properties as in the case of Type II censoring. In clinical trials, often, an interim analysis scheme is adopted wherein one plans to review the accumulating dataset either at regular time intervals or after regular failure intervals. In that way, one has an extension of Type I or II censoring schemes. For such schemes, RST procedures are generally adopted to guard against an inflation of Type I error for the overall significance testing procedures. In an extreme case, one may also like to monitor the study more or less on a continual basis, and in that setup, a progressive censoring scheme (PCS) is more appropriate. With this motivation, we first consider some PCS schemes, mainly adapted from Chatterjee and Sen (1973) and Sen (1981b).

Consider the (double-)sequence of PCS linear rank statistics $\{L_{nk}; 0 \leq k \leq n\}$, defined by (9.9), and let $\mathcal{B}_{nk}$ be the sigma-field generated by $\{S_j; j \leq k\}$, for $k = 0, \ldots, n$. Then, it follows from Chatterjee and Sen (1973) and Majumdar and Sen (1978) that under the null hypothesis of homogeneity of the d.f.'s $F_1, \ldots, F_n$, for every n, $\{\mathbf{L}_{nk}, \mathcal{B}_{nk}; 0 \leq k \leq n\}$ is a null mean (vector) martingale (array). Moreover, under the null hypothesis, these $\mathbf{L}_{nk}$ are (jointly) distribution-free, so that a test based on this collection is also distribution-free under the null hypothesis. With the possibility of early termination in mind, we may consider the following Kolmogorov-Smirnov type test statistics. First, consider the case of scalar t_i, so that the L_{nk} are scalar too. Let then

$$K_n^+ = \max\{Q_n^{-1}A_n^{-1}L_{nk} : 0 \leq k \leq n\}; \tag{9.10}$$

$$K_n = \max\{Q_n^{-1}A_n^{-1}|L_{nk}| : 0 \leq k \leq n\}. \tag{9.11}$$

It may be remarked that (9.10) is designed for testing agaianst one-sided alternatives $\beta > 0$, while (9.11) is for the two-sided ones $\beta \neq 0$. The exact null hypothesis distribution of either of these statistics can be obtained by enumeration of all possible $n!$ realizations of $S_1, \ldots, S_n$ over the permutations of $1, \ldots, n$. This task becomes prohibitively laborious as n increases, and hence, for large values of n, suitable distributional approximations are generally used. Towards these limit laws, we may construct suitable stochastic processes $W_n = \{W_n(t), 0 \leq t \leq 1\}, n \geq 1$, by letting

$$W_n(t) = Q_n^{-1}A_n^{-1}L_{nk_n(t)}, \ t \in [0, 1], \tag{9.12}$$

where

$$k_n(t) = \max\{k : A_{nk}^2 \leq tA_n^2\}, \ t \in [0, 1], \tag{9.13}$$

and the A_{nk}^2 are all defined as in (9.6). Note that by definition,

$$K_n^+ = \sup\{W_n(t) : t \in [0, 1]\}; \tag{9.14}$$

$$K_n = \sup\{|W_n(t)| : t \in [0,1]\}. \tag{9.15}$$

Incorporating the martingale property described earlier, it was shown by Chatterjee and Sen (1973) that under the null hypothesis,

$$W_n \text{ converges in law to } W, \text{ as } n \to \infty, \tag{9.16}$$

where $W = \{W(t), t \in [0,1]\}$ is a standard Brownian motion on the unit interval [0,1]. This weak convergence result in turn leads us to the following: Under the null hypothesis of homogeneity of the d.f.'s $F_1, \ldots, F_n$, for every $\lambda \geq 0$, as n increases,

$$P\{K_n^+ \geq \lambda | H_0\} \to 2[1 - \Phi(\lambda)], \tag{9.17}$$

$$P\{K_n \geq \lambda | H_0\} \to 2[1 - \Phi(\lambda)] - 2\sum_{k \geq 1}(-1)^k[\Phi((2k+1)\lambda) - \Phi((2k-1)\lambda)], \tag{9.18}$$

where $\Phi(\cdot)$ is the standard normal d.f. Therefore, the asymptotic critical levels for K_n^+ and K_n can be obtained from the above two expressions. Let us denote the actual α-level critical values for K_n^+ and K_n by $K_{n,\alpha}^+$ and $K_{n,\alpha}$ respectively. Then, we have the following *time-sequential* testing procedures: (i) To test the null hypothesis H_0 against the one-sided alternatives H_1, at the kth failure point $X_{n:k}$, compute the statistic $W_{nk} = Q_n^{-1}A_n^{-1}L_{nk}$, for $k \geq 0$. As long as these W_{nk} lie below the level $K_{n,\alpha}$, continue in having more accumulating data; if for the first time, at some $k = M$, say, W_{nM} is $\geq K_{n,\alpha}^+$, stop at this Mth failure point and reject the null hypothesis. If no such M exists, stop at the last failure point and accept the null hypothesis. (ii) For the two-sided alternatives, work with the $|W_{nk}|$ and have a similar procedure where the critical level is taken as $K_{n,\alpha}$.

Several modifications of this procedure are quite easy to workout. First, consider the case where the experiment is preplanned to a maximum of r out of n failures, so that a time-sequential procedure has a maximal duration $X_{n:r}$. To this end, one may simply define the statistics

$$W_{nk;r} = Q_n^{-1}A_{nr}^{-1}L_{nk}, \quad k = 0, 1, \ldots, r, \tag{9.19}$$

and define the sup-norm statistics as in before (with $k \leq r$). Thus, effectively, we shrink the range of k and rescale by the truncated variance function. In that way, the EDF character of the tests (under the null hypothesis) remains in tact, and the same limiting distributions hold. Secondly, often, this maximum duration of a time-sequential procedure is set in terms of a given time point. Borrowing the analogy with the Type I censoring scheme, we may again derive some conditionally distribution-free time sequential tests, for which in the above setup, we are to restrict ourselves to values of $k \leq R$, the number of failures within that set time period. Thirdly, instead of the unweighted sup-norm statistics, we could have used some weighted version wherein the $W_n(t)$ are to multiplied by a nongegative scalar factor $q(t)$, such that $\int_0^1 q^2(t)dt < \infty$. The choice of this weight function $q(\cdot) = \{q(t); t \in (0,1)\}$ may be made on the basis of the importance of early stopping (from clinical point of view). This will relate to the weak convergence to a general Gaussian process and would therefore rely on the boundary crossing

probabilities for such a process. In some simple cases, such probabilites are known [viz., Chapter 2 of Sen (1981b)], and these can then be incorporated in the simplification of asymptotic critical levels for the test statistics. But, in general, for an arbitrary square integrable $q(\cdot)$, such algebraic expressions may not be available, and hence, numerical or simulation methodology may have to be used. We may refer to Sinha and Sen (1982) for some related results. Fourthly, instead of a more or less continuously monitoring, as is the case with PCS, it may be desirable to have a prefixed number of looks into the accumulating data set either on a calender time basis or on the response outcome one; indeed this is usual for conventional interim analyses schemes. In this finite dimensional version, even for the asymptotic case, one needs to look into multivariate normal distributional probability (multiple) integrals, and exact evaluation seems to be rather impracticable. There are some simplifications when the number of looks is as small as 2 or 3, although they are to be obtained by numerical quadrature formulae, and we may refer to Flemming and Harrington (1991) for some related studies. Lan and DeMets (1983) introduced a novel concept of *spending function* in this context, and that can be used with some advantages, particularly when the interim time points are based on the response outcome. In passing, we may remark that if the number of looks exceeds 10 and these points are not too concentrated in any particular patch of the unit interval, then a solution obtained from the contonuous process provides a close (upper) bound to the one from the finite discrete version. In long range, multi-center clinical trials, generally such an interim analysis involves a moderately large number of looks into the accumulating data set on a fairly regular basis, and hence, such weak convergence based approximations work out fairly well.

We consider next the general case of vector $\mathbf{t}_i$, and as in (9.9) define the statistics $\mathcal{L}_{nr}, r = 0, 1, \ldots, n$. Let then

$$\mathcal{L}_{nr}^* = \{A_{nr}^2 / A_n^2\}\mathcal{L}_{nr}, \ r \leq n. \tag{9.20}$$

In this case, as a test statistic, we consider the following:

$$K_n^* = \max\{\ \mathcal{L}_{nk}^* \ : \ k \leq n\}. \tag{9.21}$$

In order to express this statistic in terms of a suitable stochastic process, we adopt the same definition of $k_n(t), t \in [0, 1]$, as in (9.13). Also, let $B_n^2 = \{B_n^2(t); t \in [0, 1]\}$ be defined by

$$B_n^2(t) = \mathcal{L}_{nk_n(t)}^*, \ t \in [0, 1]. \tag{9.22}$$

Then, we may write equivalently

$$K_n^* = \sup\{B_n^2(t) : t \in [0, 1]\} \tag{9.23}$$

Let $W_1, \ldots, W_q$ be independent copies of a standardized Brownian motion W. Let us define then $B^2 = \{B^2(t), t \in [0, 1]\}$ by letting

$$B^2(t) = \sum_{j=1}^{q} W_j^2(t), \ t \in [0, 1]. \tag{9.24}$$

In the literature these are known as the *Bessel (squared) processes*, and boundary crossing probabilities for them have been extensively studied by DeLong (1981) and others. It follows from Majumdar and Sen (1978) that under the null hypothesis of homogeneity of the d.f.'s $F_1, \ldots, F_n$, as n increases, B_n^2 converges weakly to B^2, so that K_n^* converges in law to a Bessel squared process functional $K^* = \sup\{B^2(t) : t \in [0, 1]\}$. This enables us to use the extensive tables provided by DeLong (1981) for asymptotic approximations for the exact null hypothesis distribution of K_n^*. In this respect we may note that this test is essentially against a multi-sided alternative, and hence, one-sided versions are not that suitable in this setup. Secondly, as in the case of K_n^+ or K_n, here also, we may work with suiably weighted versions of K_n^*. Thirdly, modifications for a Type II censoring scheme in connection with PCS can be made by replacing A_n^2 by the correponding value at the target number r^*, and a conditionally distribution-free version may similarly be considered for the trucation PCS scheme.

In the discussion made above, we have mainly confined ourselves to identification of EDF structures and simplifications of the null hypothesis distributions of suitable rank based procedures. The study of their non-null distributions entails even more mathematical complexities where often the regularity assumptions may not match the reality of practical applications. However, under the usual local alternatives, such asymptotic distributions, albeit being more complex than in the conventional situations, have been studied in a unified manner, and a general account of these asymptotics is given in Sen (1981b,Ch.11). In clinical trials or survival analysis, it is not uncommon to encounter *noncompliance* due to dropouts or withdrawals. One common approach to accomodating noncompliance in an objective perspective is to introduce the concept of *random censoring* wherein it is assumed that a *censoring variable* C_i is associated with the primary variate (Y_i), such that the observable r.v.'s are

$$T_i = min\{Y_i, C_i\} \text{ and } I_i = I(T_i = Y_i), i \geq 1, \tag{9.25}$$

and there may be, in general, other concomitant variates too. In this setup, it is assumed that the C_i are stochastically independent of the Y_i (a condition that may not generally hold in practice), and further the distribution of C_i is not affected by the treatments to which the subjects are subjected in the study; this latter condition is known as *noninformative censoring*, and again, this may not meet the light of reality in all applications. In a nonparametric formulation, such a random censoring scheme may introduce complications beyond a simple amendment range, and may therefore require more structural assumptions on the survival functions of the associated variables. In this respect, a fundamental contribution is due to D.R. Cox (1972,1975) who incorporated the novel idea of *partial likelihood functions* along with the basic assumption of *proportional hazards (PH)* for the survival functions for various treatment groups. In the simplest case of a palcebo vs. treatment study, if F_0 and F_1 stand for the respective d.f.'s, we denote the corresponding survival functions by $\bar{F}_0$ and $\bar{F}_1$, and let f_j and h_j be the density and hazard functions corresponding to F_j, for $j = 0, 1$. Thus, $h_j(x) = f_j(x)/\bar{F}_j(x)$, for $x \in (0, \infty)$ and $j = 0, 1$. It is then assumed that

$$h_1(x) = c \cdot h_0(x), \text{ for all } x \in (0, \infty), \tag{9.26}$$

45

where c is a positive constant. In terms of the survival functions, this PH assumption leads to the following formulation:

$$\bar{F}_1(x) = [\bar{F}_0(x)]^c, \text{ for all } x \in (0, \infty). \tag{9.27}$$

In this formulation, one can allow the d.f. F_0 to be arbitrary (but absolutely continuous), so that the hazard function $h_0(x)$ is treated as an arbitrary function on $(0, \infty)$. If both F_0 and F_1 are exponential d.f.'s, then this PH condition is automatically true, and the above formulation extends this characterization to a wider class. Motivated by this simple formulation, let us consider a set of n d.f.'s $F_i, i = 1, \ldots, n$, denote the corresponding survival and hazard functions by $\bar{F}_i, i = 1, \ldots, n$ and $h_i(x), i = 1, \ldots, n$ respectively, and consider a PH formulation incorporating a regression function $\beta' t_i$ as follows:

$$h_i(x) = h_i(x|t_i) = h_o(x) \cdot exp\{\beta' t_i\}, \ x \in (0, \infty), i = 1, \ldots, n, \tag{9.28}$$

where the t_i are given design variates, and $h_o(x)$, the base line hazard function, is treated as arbitrary. In this formulation, we may even include (stochastic) concomitant variates in the t_i, so that we may term the h_i as conditional hazard function, and in terms of the log-hazard functions, we have then a linear regression model. Since the regression part of the model is of parametric form while the base line hazard function is nonparametric in character, this model is also referred to as a *semi-parametric model*. The ingenuity of the Cox formulation lies in incorporation of the partial likelihood formulation for drawing valid and efficient statistical conclusions on the *regression* parameter (vector) β, treating the baseline hazard as a nusance parameter (functional). A complete treatment of this novel methodology is beyond the scope of this article, although some other chapters in this Volume are likely to deal with this in some details. In passing, we may refer to the recent monograph of Andersen et al. (1993) where this semi-parametric approach has been dealt with in a much more general and sound theoretical basis. Use of rank based procedures for this PH model is naturally appealing on the ground of *invariance* under strictly monotone transformations on the primary variate, and it is not surprising to see that the classical log-rank procedure described in Section 2 has a close affinity to this model as well. PCS modeling for such PH models has also been worked out along the same line as in before, and we may refer to Sen (1981b) for some details. There are a few points to ponder in this respect. First, the PH assumption itself may not hold in general in all applications ; we may refer to Sen (1994 b) for some exposition of the nonrobustness aspects of the classical PH models in survival analysis and clinical trials. Secondly, *staggered entry* or batch arrival models are usually encountered in practice, where the entry pattern may be quite arbirary and stochastic in nature. In such a case, the partial likelihood approach may not lead to the usual log-rank procedures, and may call for more complex statistical analysis schemes; we refer to Sen (1985) for some details. Thirdly, if the study involves some concomitant variates (and usually they abound in practical applications), often, we have some of them as *time-dependent.* In such a case, the Cox formulation of the PH model may require more sophisticated statistical solutions, and much of the simplicities may be lost in this quest; we refer to Murphy and Sen (1991) for some accounts of such developments. Finally,

we may remark that censoring in statistical theory and practice may not be very complementary to each other or coherent in a natural sense. Part of this difficulty [viz., Sen (1995d)] stems from the fact that the basic regularity assumptions, as are generally needed in biomedical studies (including clinical trials, and epidemiological and environmental investigations), may not provide an easy access to incorporating the simple censoring schemes referred to before; the more complex is the nature of such censoring the greater is the price one has to pay for implementation of suitable statistical designs and for developing suitable statiastical parametric or nonparametric analysis schemes to accomodate such complexities. We complete this section with some discussion on the *staggered entry* plans which are common in medical studies and clinical trials. Recruitment of subjects for such investigations often require extensive search for valid subjects, and as these are generally people having some medical problems requiring some treatments, they may arrive at the clinics either in batches or sometimes sporadically over time. This may result in a differential exposure time if the study is planned for a fixed duration of time, and hence, the basic formulation of Type I, II or random censoring schemes may not be very appealing in such a case. To handle such a relatively more complex censoring pattern, one may need to look into the composite picture from a multiple time-parameter point of view. Sinha and Sen (1982) considered such a scheme based on the usual empirical distributions adjusted for staggered entries and formulated suitable nonparametric testing procedures. A more general treatment with rank statistics in a staggering entry plan is due to Sen (1985). Weak convergence to multi-dimensional Gaussian processes provides the desired statistical tools to implement such methodology in practical applications.

10 Nonparametrics in Incomplete Multiresponse Clinical Designs

In clinical trials, from epidemiological perspectives, often, information is gathered on more than one response variable, and, in addition, on relevant concomitant variables too. Nevertheless, from clinical or medical perspectives, it is not uncommon to single out one of the response variates as the primary endpoint and the others as auxiliary ones. In some cases, characterizing a primary endpoint in this conventional manner may entail stringent cost constraints primarily related to its precise measurement, and some *surrogate* endpoints are advocated to cast vital information at relatively lower levels of cost or time consumptions. The multitude of such response variates through their mutual statistical dependence can cast light on the primary endpoint and hence their simultaneous measurements generally lead to comparatively more precise statistical conclusions. On the other hand, based on other practical considerations, such as relative cost, ease and/or precision of measurement, it may not be very convenient to include the entire battery of simultaneous measurements on all the response as well as concomitant variates for all the subjects (units). For this reason, often, an *incomplete multiresponse design (IMD)* or a *hierarchical design (HD)* is adopted. In the latter design, there is a hierarchy on the response variates

in the sense that there is a (partial) ordering with respect to the number of experimental units on which their measurements are recorded; for this reason, they are also termed *nested designs (ND)*. To illustrate this point, suppose that there are $p(> 1)$ response variates, denoted by $Y_1, \ldots, Y_p$ respectively. On a smallest set, sat $\mathcal{S}_1$, of experimental units, all these p responses are measured simultaneously; for a larger set $\mathcal{S}_2$, containing $\mathcal{S}_1$ as a subset, $Y_2, \ldots, Y_p$ (but not Y_1) are recorded on the subset $\mathcal{S}_2 \setminus \mathcal{S}_1$, and so on. For the largest set $\mathcal{S}_p$, containing $\mathcal{S}_{p-1}$ as a subset, Y_p alone is recorded on the subset $\mathcal{S}_p \setminus \mathcal{S}_{p-1}$. Such a multiresponse design, determined by the inherent nesting $\mathcal{S}_1 \subset \mathcal{S}_2 \subset \cdots \subset \mathcal{S}_p$ is termed a hierarchical design [viz., Roy et al. (1971, Ch.8)]. It may not always be desirable or even practicable to impose this basic hierarchy condition. For example, the (random pattern) *missing observations* in multiresponse designs may distort this hierarchy condition to a certain extent. Nevertheless, it may be feasible to incorporate some IMD's wherein the set $\mathbf{Y} = \{Y_1, \ldots, Y_p\}$ can be partitioned into various subsets $\{Y_{i_1}, \ldots, Y_{i_r}\}, 1 \leq r \leq p, 1 \leq i_1 < \ldots, < i_r \leq p$, such that these subsets are not necessarily nested and they are adoptable for possibly different number of experimental units. For example, for $p = 2$, we have three possible subsets $\{Y_1\}, \{Y_2\}$ and $\{Y_1, Y_2\}$, and possibly different designs (say, $\mathcal{D}_1, \mathcal{D}_2$ and $\mathcal{D}_{12}$) may be chosen for these subsets. In this context it may be recalled that in clinical trials, often, the primary emphasis is on a comparative study of a placebo and one or more treatments, so that these designs are to be chosen in a conventional sense with due emphasis on these treatments.

In clinical trials, a primary endpoint, in spite of being the most relevant one, may encounter some basic problems regarding its precise measurement (due to possibly excessive cost or some other practical limitations); therefore, it is not uncommon to make use of a very closely related but presumably, relatively less expensive variate, termed a surroage endpoint. Generally, surrogate endpoints may not contain as much information as contained in the primary endpoint, and such a substitution may have serious effects on valid and efficient statistical modeling and analysis, unless the surrogate variate has some statistical concordance with the primary one. The situation may particularly be very bleak when this statistical interface of surrogate and primary endpoints is not that clearly known, and this case arises typically when no specific data are available on simultaneous measurement of both these variables. Nevertheless, the use of such surrogate endpoints in clinical trials and medical investigations has generally been accepted by the allied medical community and has caught the attention of statisticians as well. A nice statistical account of such uses, and abuses too, is given in a set of articles published in *Statistics in Medicine*, Volume 8, No.2 (1989). More technical exposition of this field are due to Pepe (1992) and Sen (1994a), among others. Not all auxiliary variables qualify for surrogates, and for qualified ones, it seems very reasonable (if not essential) to design a study in such a way that for a majority of the experimental units, termed the *surrogate sample*, valid surrogate endpoints and concomitant variates are recorded, while for a smaller subset of experimental units, termed the *validation sample*, simultaneous recording of the primay and surrogate endpoints throws light on their statistical relation which enables us to combine the evidence from both the subsets of data and draw better statistical conclusions. If statistical conclusions are to be

drawn only from the surrogate sample observations, some stronger regularity assumptions are generally needed to justify the conclusions, while the use of a validation sample may enhance the scope of the study considerably. We may refer to Prentice (1989) and Pepe (1992) for some useful accounts of these pros and cons of surrogate endpoints in clinical trials. The Prentice-Pepe setups can be characterized as both a hierarchical and incomplete multiresponse model with $p = 2$. Many clinical trials encounter a more complex setup involving multiple endpoints resulting in multiresponse primary variates. We may refer to Wei et al. (1989) and Prentice and Cai (1992) for some statistical treatments for such designs. A more comprehensive IMD/HD approach in a nonparametric setup is due to Sen (1994a), and we summarize these results here.

There may be in general more than one primary endpoints, and we denote this by $\mathbf{Y} = (Y_1, \ldots, Y_p)'$, where $p(\geq 1)$ and there may be a partial ordering of the importance of these primary endpoints, which may also be taken into account in the design and statistical analysis of the study. Similarly, the surrogate endpoint may also be represented by a q-vector $\mathbf{Y}_o$, where q is a positive integer. Thus, in general, we have a set of $p + q$ responses, some of which may be costly to record. In order to extract information on the statistical relations between $\mathbf{Y}$ and $\mathbf{Y}_o$, and to incorporate the same in drawing statistical conclusions, it may be desirable to use IMD's or HD's. In this respect in a conventional approach, one adopts *multivariate general linear models (MGLM)* for statistical modeling and analysis; however, the basic regularity assumptions are even more unlikely to be tenable in this multivariate situation. Thus, the appropriateness of MGLM's in clinical trials is questionable. Use of *generalized linear models (GLM)* is also subject to similar limitations, and on model robustness grounds they are even more vulnerable to plausible departures from the assumed regularity assumptions. The Cox (1972) PHM based *partial likelihood* approach is also subject to serious nonrobustness constraints, and hence, sans sufficient confidence on such a parametric or semi-parametric model, such procedures should not be advocated in real applications. Some aspects of nonrobustness of the Cox PHM approach are discussed in Sen (1994a), and these remarks pertain to general IMD's as well. Nonparametrics, on the other hand, possess good robustness properties, and are better competitors to these alternative ones. This has been the main motivation of Sen (1994a) in pursuing general nonparametrics for such IMD's with adequate emphasis on the related asymptotics.

The basic rank procedures described in the earlier sections, particularly, for randomized blocks, incomplete block designs, factorial experiments and multivariate models, provide the necessary access for this development, and we shall unify these in a convenient mold. Aligned rank procedures are particularly useful in this context. It may be recalled that a surrogate endpoint is a qualified substitute for the primary endpoint only if it reflects a picture with reference to the treatment difference concordant with the primary endpoint; in the literature this condition is also referred to as the *validity* criterion for a surrogate. Such a condition can be tested if one has a validation sample where both the primary and surrogate endpoints are recorded. But, generally, such a validation sample has a smaller size compared

49

to the surrogate sample. Hence, the general nonparametric approach is based on the following scheme: (i)For testing a plausible hypothesis relating to treatment differences construct a suitable nonparametric test statistic based on the surrogate sample observations and adjusted for covariates, if any. (ii) For the validation sample, construct a similar nonparametric statistic for both the primary and surrogate endpoints (using the multivariate approach treated in Section 4), also adjusted for concomitant variates, if any. (iii) Test for the concordance of the primary and surrogate endpoints with respect to treatment differences based on the statistics in Step (ii). Again, nonparametric tests can be used here. (iv) Regress the primary endpoint statistics on the surrogate endpoint ones (in the validation sample in Step (ii)), and obtain the aligned statistics for the primary endpoint as residuals from this fitted regression. (v) Combine the statistics in Step (i) and (iv) by the usual weighted least squares principle and use the same in the formulation of the actual test statistic to be used for testing a hypothesis on treatment differences.

In this context the joint (asymptotic) normality of multivariate rank statistics provides the theoretical justifications for the various steps sketched above, and also provides the foundation of general asymptotics relevant to this topics. These details are provided in Sen (1994a).

For linear models in the parametric case, IMD's entail a secondary task: *Recovery of interblock information* from the block totals. In the nonparametric case, although the basic linearity of the model may not be fully appreciated, such recovery of interblock information is possible. The basic motivation is the same, although a somewhat different alignment process is needed to incorporate this recovery in nonparametric analysis. This alignment procedure is very similar to the classical parametric case: Block averages or some other measure of central tendency are used for construction of interblock aligned rank statistics, while the residuals within each block are pooled together for all blocks and replicates to construct aligned rank statistics for intra-block analysis. These two sets of rank statistics are then combined [as in Step (v) above] in a convenient way to construct suitable test statistics which have greater power than the one based solely on the intra-block residuals. For clinical trials involving (a treatment-wise) IMD and a surrogate endpoint, recovery of interblock information in nonparametric analysis of covariance models has recently been treated in a unified manner by El-Moalem and Sen (1995). The idea is quite simple. In addition to using the aligned rank statistics (for the primary and surrogate endpoints as well as the concomitant variates), in a replicated IMD, it is also possible to use the within replicate block totals, align them with a view to eminiating the replicate effects, and then to use aligned rank statistics on such aligned block totals to extract further information on the treatment effects. Use of such aligned rank statistics eliminates the cruciality of the linearity of the model to a certain extent and makes it possible to use the usual weighted least squares methodology to construct suitable pooled aligned rank statistics which may be incorporated in the construction of a plausible test statistic having, at least asymptotically, better power properties. In this context it is not necessary to assume that the treatments are replicated equal number of times within each replicate or any pair of them are done so, and the treatise covers a general class of IMD's.

11 Concluding Remarks

The current state of art with the developments on nonparametrics in design and analysis of various types of experiments really calls for a far more thorough treatise of the subject matter than presented in this writeup. For lack of space, it has not been possible to include the entire battery of topics in design and analysis of experiments where nonparametrics are relevant.

As regards the basic nonparametrics presented in the first eight sections of this writeup, the treatment here is fairly thorough. However, the last two sections are presented with more motivations from applications point of view but from methodological point of view somewhat less technical details are provided than they deserve. The nonparametric task remains as much more challenging, although some work has already been in progress in this direction. On the top of this there is another important consideration underlying practical adoption of statistical designs and analysis packages in clinical trials and medical studies in general. *Missing observations* may be a part of such experimental data, and statistical analysis should address this issue adequately with due considerations to practical adoptions. In clinical trials, such a missing pattern may be due to censoring of various types discussed in earlier sections, while in a general setup, it may be due to other factors as well. In epidemiological studies, it is not uncommon to encounter multiple causes of failures, and hence, a *competing risk* setup is often judged as appropriate. Again design and statistical analysis (parametrics as well as nonparametrics) for such studies follow somewhat different tracks, and it may be desirable to pay due attention to the developments of nonparametrics for competing risks models in as much generality as possible. *Random missing patterns* are often introduced as a part of the basic assumptions to deal with messy data sets arising in such studies. In a nonparametric MANOVA setup, for some developments on such random missing patterns, we may refer to Servy and Sen (1987), where other pertinent references are also cited. There remains much more to be accomplished in this direction. Competing risks models in a general multiresponse (or multiple endpoint) clinical trial poses even more complex statistical designing and analysis tasks. Only in some simplest situations, some relevant nonparametrics have been developed; we may refer to DeMasi (1994) where other references are cited in detail. Since more complex censoring patterns may arise in this context, statistical modeling (underlying either parametric or robust procedures) needs to address the infrastructure in an adequate manner; this not only increases the number of parameters associated with the model, but also may raise some *identifiability* issues which call for more delicate treatments. These need to be addressed in a more general and integrated manner than done here.

Throughout this presentation the major emphasis has been on nonparametrics based on rank statistics and allied estimators. Although such rank procedures can mostly be justified from a *global robustness* point of view (with very little emphasis on the form of the underlying error distributions), there are some other situations where it may be wiser to take recourse to *local robustness* properties wherein only small departures from an assumed model are contemplated, so that high efficiency mingled with

low sensitiveness to such local departures dominate the scenario. In this setup, as viable competitors to such nonparametrics, robust procedures based on suitable L- and M-statistics are often advocated. *Regression quantiles* have their genesis in this complex, but in the recent past, they have paved the way for the related *regression rank scores* estimators and test statistics which compare very favorably with nonparametric procedures based on R-estimators. Recently, Jurečková and Sen (1993) have established certain asymptotic equivalence results on the classical R-estimators and regression rank scores estimators in a linear model based on a common score generating function, and as such, taking into account the relative computational complexities of these two approaches, in some cases, we may advocate the use of such regression rank scores procedures as well. For some general accounts of these findings we may refer to Jurečková and Sen (1995, Ch.6), where mostly fixed-effects models are considered, and to Sen (1995a) where some mixed-effects models have also been treated.

In the clinical and epidemiological sectors, due to medical ethics standards and current policies of some of the regulatory agencies in USA or other industrized nations, a *multi-phase design* approach for human usage is generally adopted. In Phase I, primary emphasis is on exploration of biochemical/biomedical effects, toxicity etc, while in Phase II, some *therapeutic* factors are taken into account. In this setup, it is quite common to have first some *animal studies*, and the conclusions as may be gathered from such studies are then to be incorporated in the design and general formulation of the main study: Phase III clinical trials. The emerging sub-discipline: *clinical epidemiology* has been geared to address more complex issues arising in this interdisciplinary field. Because of apparently conflicting attitudes of statisticians and epidemiologists to some of these clinical problems, in clinical epidemiology, there is, often, a blending of *ecology* and *etiology* for which the design as well as analysis aspects may differ drastically even in some simple parametric setups [see for example, Sen (1994c)]. Nonparametrics play a fundamental role in this setup too. For example, extrapolating the statistical findings from experiments conducted on subhuman primates to human beings raises the question of their validity and scope. In the statistical lieterature, such methodologies are categorized under the topic: *Accelerated Life Testing (ATL)* procedures. In parametric setups, the basic regularity assumptions appear to be quite stringent, and hence, nonparametrics are generally advocated for greater scope and reliability. However, in this context too, validity and reliability of statistical regularity assumptions need to be assessed properly. *Biological assays* are the main statistical assessment tools in this venture. Designs for such bio-assays may often be somewhat different, and we may refer to the classical text of Finney (1964) for a detailed account of such developments. His treatise has mainly been on a conventional parametric walk, wherein due emphasis has been laid down on *transformations* on the response and dose variables (termed the *response metameter* and *dosage* or *dose metameter* respectively) under which suitable parametric models can be justified. Nevertheless, in practice, such transformations may not simultaneously achieve the basic linearity of the model and normality (or logistic or some other simple form) of the tolerance distribution. As such, there is good scope for nonparametrics. Some developments in this sector took

place during the sixties [viz., Sen (1963)] and early seventies, and a systematic review of this work is reported in Sen (1984a), where pertinent references are also cited.

A new area of statistical awareness relates to our endangered environment and the statistical endeavors to cope with such problems; these have led to the developments of another frontier of statistical sciences: the *Envirnometrics*. The tasks are truely challenging and statistical considerations are overwhelming in this venture. Unlike the case of conventional agricultural experiments, animal studies or even the clinical trials, environmental problems are generally characterized by the lack of control in the conduct of a scientific study to a much greater extent. Also, a large number of factors contributes to unaccountable variations in the response patterns. Moreover, the response variables are often imprecisely defined and may also encounter serious measurement problems. For example, to assess the air pollution standard of various urban, suburban and rural areas in USA, the basic task may be to define precisely the response variables, identify their probable causes or factor variables (viz., auto-exhaution, environonmental smoking, industrial emissions, etc.,), variation with the whether conditions. day-to-night variation, and many other factors which may not be properly defined and may hardly be controllable to a satisfactory extent. Some of these variables may even be binary or polychotomous in nature. On the top of that even when a variable is quantitative, it may usually be recorded in class intervals leading to the so called *interval censoring* schemes. Thus, *measurement errors* and to a certain extent *misclassifications* are usually encountered as a vital part of such response as well as dose variables. Even in the simplest case of two or several sample problems, for such *grouped data* ties among the observations may not be negligible, and there may not be a unique way to handle such ties; we may refer to Hájek and Šidak (1967, Ch. 3) for some treatment of ties for the exat null hypothesis ditributions of rank statistics, and to Sen (1967) for asymptotic optimality of rank tests for grouped data in a simple regression model. These results extend directly to general linear models. However, linear models are hardly appropriate in this complex setup when the response-dose regression, subject to possible measurement errors/misclassifications, may be quite nonlinear in nature, and suitable transformations on the factor as well as response variables are generally used to induce more linearity in the models; their impacts on the distributional assumptions are needed to be assessed carefully. Assumptions of independence, homoscedasticity and even the symmetry of the error components are to be examined critically in the particular contexts, and for these reasons, statistical designs and analysis schemes are to be developed in more practically adoptable settings. Such environmental problems are not totally out of the reach of clinical trials and biomedical studies. The emerging field of *environmental health sciences* deals with the impact of the environment on human health and prospects for long-range healthy living. Environmental health effects have been identified to be far more outreaching than in a simple chemical or biochemical setting, and *Genotoxicity* has also been identified as an important ingredient in this phenomenon. In this quest, biological assays involving *biological markers* are vital tools for assessments on subhuman primates, and suitable design of (mutagenetic) experiments are generally advocated for extrapolation of

the findings from animals to human beings. Because of the fundamental roles of *Molecular Biology* and *Human Genetics* in these complex experimental schemes, such designs are generally quite different from the conventional ones considered in this volume. The appropriateness of an interdisciplinary approach is crucial in this context. *Inhalation toxicology, water contamination, air pollution* and scores of other serious environmental threats are affecting the *Quality of Life* (QOL) and in many ways, endangering our lives too, and, in this respect, an *interdisciplinary approach* is very much needed to provide scientifically sound and operationally manageable solutions. We may refer to Sen and Margolin (1995) for some of the basic statistical issues in some environmental studies with major emphasis on inhalation toxicology, and conclude that statistical planning and analysis schemes are most vital in this venture. Parametrics or semi-parametrics are less likely to be appropriate in this emerging research field, and nonparametrics are indispensible in this context to a far greater extent.

12   Acknowledgments

References

[1] Adichie, J.N. (1967). Estimates of regression parameters based on rank tests. *Ann. Math. Statist. 38,* 894-904.

[2] Andersen, P.K., Borgan, O., Gill, R.D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* Springer-Verlag, New York.

[3] Armitage, P. (1975). *Sequential Medical Trials,* 2nd ed. Blackwell, Oxford.

[4] Bahadur, R.R. (1961). A representation of the joint distribution of responses to n dichotomous items. In *Studies in Item Analysis and Prediction* (H. Solomon,Ed.) Stanford Univ. Press, Calif.

[5] Benard, A. and van Elteren, P. (1953). A generalization of the method of m-rankings. *Indag. Math. 15,* 358-369.

[6] Bhapkar, V.P. (1961a). Some nonparametric median procedures. *Ann. Math. Statist.32,* 846-863.

[7] Bhapkar, V.P. (1961b). A nonparametric test for the problem of several samples. *Ann. Math. Statist. 32,* 1108-1117.

[8] Bradley, R.A. and Terry, M.E. (1952). Rank analysis of incomplete block designs, I. The method of paired comparison. *Biometrika 39,* 324-345.

[9] Brown, G.W. and Mood, A.M. (1951).On median tests for linear hypotheses. *Proc. 2nd Berkeley Symp. Math. Statist. Probability,* 159-166.

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

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

[12] Cox, D.R. (1972). Regression models and life tables (with discussion).*J. Roy. Statist. Soc. Ser.B, 34*,187-220.

[13] Cox, D.R. (1975). Partial Likelihood. *Biometrika 62*,369-375.

[14] David, H.A. (1988). *The Method of Paired Comparisons*, 2nd Ed. Oxford Univ. Press, New York.

[15] Davidson, R.R. and Bradley, R.A. (1969). Multivariate paired comparisons: The extension of a univariate model and associated estimation and test procedures. *Biometrika 56*, 81-94.

[16] Davidson, R.R. and Bradley, R.A. (1970). Multivariate paired comparisons: Some large sample results on estimation and test of equality of preference. In *Nonpartametric Techniques in Statistical Inference* (M.L. Puri, Ed.), Cambridge Univ. Press, New York, pp.111-125.

[17] DeLong, D.M. (1981). Crossing probabilities for a square root boundary by a Bessel process. *Commun. Statist. Theor. Meth. A10*, 2197-2213.

[18] DeMasi, R.A. (1994). *Proportional Hazards Models for Multivariate Failure Time Data with Generalized Competing Risks.* Unpublished Doctoral Dissertation, Univ. N. Carloina, Chapel Hill.

[19] Durbin, J. (1951). Incomplete blocks in ranking experiments. *Brit. J. Statist. Psychol. 4*, 85-90.

[20] El-Moalem, H. and Sen, P.K. (1995). Nonparametric recovery of interblock information in clinical trials with a surrogate endpoint.(submitted for publication).

[21] van Elteren, P. and Noether, G.E. (1959). The asymptotic efficiency of the χ_r^2-test for a balanced incomplete block design. *Biometrika 46*, 475-477.

[22] Finney, D.J. (1964).*Statistical Methods in Biological Assay,*2nd ed. Griffin, London.

[23] Fleming, T.R. and Harrington, D.P. (1991). *Counting Processes and Survival Analysis.* John Wiley, New York.

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

[25] Gerig, T.M. (1969). A multivariate extension of Friedman's χ^2-test. *J. Amer. Statist. Assoc. 64*, 1595-1608.

[26] Gerig, T.M. (1975). A multivariate extension of Friedman's χ^2-test with random covariates. *J. Amer. Statist. Assoc. 70*, 443-447.

[27] Greenberg, V.L. (1966). Robust estimation in incomplete block designs. *Ann. Math. Statist. 37*, 1331-1337.

[28] Grizzle, J.E. (1965). The two-period change-over design and its use in clinical trials. *Biometrics 21*, 467-480.

[29] Hájek, J. and Šidak, Z. (1967). *Theory of Rank Tests.* Academic Press, New York.

[30] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986). *Robust Statistics: The Approach Based on Influence Function*, John Wiley, New York.

[31] Hodges, J.L.,Jr. and Lehmann, E.L. (1962). Rank methods for combination of independent experiments in analysis of variance. *Ann. Math. Ststist. 33*, 487-497.

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

[33] Huber, P.J. (1981). *Robust Statistics*, John Wiley, New York.

[34] Jaeckel, L.A. (1972). Estimating regression coefficients by minimizing dispersion of the residuals. *Ann. Math. Statist. 43*, 1449-1458.

[35] Jurečková, J. (1971). Nonparametric estimate of regression coefficients. *Ann. Math. Statist. 42*, 1328-1338.

[36] Jurečková, J. (1977). Asymptotic relations of M-estimates and R-estimates in linear models. *Ann. Statist.* *5*, 464-472.

[37] Jurečková, J. and Sen, P.K. (1993). Asymptotic equivalence of regression rank scores estimators and R-estimators in linear models. In *Statistics and Probability: A Raghu Raj Bahadur Festschrift* (Eds. J.K. Ghosh et al.), Wiley Eastern, New Delhi, pp. 279-292.

[38] Jurečková, J. and Sen, P.K. (1995). *Robust Statistical Procedures: Asymptotics and Interrelations.* John Wiley, New York.

[39] Koch, G.G. (1972). The use of nonparametric methods in the statistical analysis of two-period change-over design. *Biometrics 28*, 577-584.

[40] Krishnaiah, P.R.(ed.) (1981). *Handbook of Statistics, Volume 1: Analysis of Variance.*North Holland, Amsterdam.

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

[42] Lan, K.K.B. and DeMets, D.L. (1983). Discrete sequential boundaries for clinical trials. *Biometrika 70*, 659-663.

[43] Lehmann, E.L. (1963a). Robust estimation in analysis of variance. *Ann. Math. Statist. 34*, 957-966.

[44] Lehmann, E.L. (1963b). Asymptotically nonparametric inference: An alternative approach to linear models. *Ann. Math. Statist. 34*, 1494-1506.

[45] Lehmann, E.L. (1964). Asymptotically nonparametric inference in some linear models with one observations per cell. *Ann. Math. Statist. 35*, 726-734.

[46] Majumdar, H. and Sen, P.K. (1978). Nonparametric tests for multiple regression under progressive censoring. *J. Multivariate Anal.8*, 73-95.

[47] Mantel, N. and Haenszel, W. (1959). Statistical aspects of analysis of data from retrospective studies of disease. *J. Nat. Cancer Inst. 22*, 719-748.

[48] Mehra, K.L. and Sarangi, J. (1967). Asymptotic efficiency of some rank tests for comparative experiments. *Ann. Math. Statist. 38*, 90-107.

[49] Mehra, K.L. and Sen, P.K. (1969). On a class of conditionally distribution-free tests for interactions in factorial experiments.*Ann. Math. Statist. 40*, 658-666.

[50] Murphy, S.A. and Sen, P.K. (1991). Time-dependfent coefficients in a Cox-type regression model. *stochast. Proc. Appl. 39*, 153-180.

[51] Pepe, M.S. (1992). Inference using surrogate outcome data and a validation sample. *Biometrika 79*, 495-512.

[52] Prentice, R.L. (1989). Surrogate endpoints in clinical trials: Definition and operational criteria. *Statist. Med. 8*, 431-440.

[53] Prentice, R.L. and Cai, J. (1992). Covariance and survival function estimation using censored multivariate failure time data.*Biometrika 79*, 495-512.

[54] Pukelsheim, F. (1993). *Optimal Experimental Design*, John Wiley, New York.

[55] Puri, M.L. and Sen, P.K. (1967). On robust estimation in incomplete block designs. *Ann. Math. Statist. 38*, 1587-1591.

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

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

[58] Quade, D. (1984). Nonparametric methods in two-way layouts. In *Handbook of Statistics, Vol.4: Nonparametric Methods*(Eds: P.R. Krisnaiah, P.R. and Sen, P.K.), North Holland, Ameterdam,pp.185-228.

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

[60] Roy, S.N., Gnanadesikan, R. and Srivastava, J.N. (1970).*Analysis and Design of Certain Quantitative Multiresponse Experiments.*Pergamon Press, New York.

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

[62] Sen, P.K. (1967a). Asymptotically mostpowerful rank order tests for grouped data. *Ann. Math. Statist. 38*,1229-1239.

[63] Sen, P.K. (1967b). A note on the asymptotic efficiency of Friedman's χ_r^2-test. *Biometrika 54*, 677-679.

[64] Sen, P.K. (1968a). Asymptotically efficient test by the method of n-ramking. *J. Roy. Statist. Soc. Ser.B 30*, 312-317.

[65] Sen, P.K. (1968b). On a class of aligned rank order tests in two-way layouts. *Ann. Math. Statist. 39*, 1115-1124.

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

[67] Sen, P.K. (1968d). Estimates of the regression coefficient based on Kendall's tau. *J. Amer. Statist. Assoc. 63*, 1379-1389.

[68] Sen, P.K. (1969a). On nonparametric T-method of multiple comparisons in randomized blocks. *Ann. Inst. Statist. Math. 21*, 329-333.

[69] Sen, P.K. (1969b). Nonparametric tests for multivariate interchangeability. Part Two: The problem of MANOVA in two-way layouts. *Sankhya, Ser.A 31*, 145-156.

[70] Sen, P.K. (1970a). On the robust efficiency of the combination of independent nonparametric tests. *Ann. Inst. Statist. Math. 22*, 277-280.

[71] Sen, P.K. (1970b). Nonparametric inference in n replicated 2^m factorial experiments. *Ann. Inst. Statist. Math. 22*, 281-294.

[72] Sen, P.K. (1971a). Asymptotic efficiency of a class of aligned rank order tests for multiresponse experiments in some incomplete block designs. *Ann. Math. Statist.42*, 1104-1112.

[73] Sen, P.K. (1971b). Robust statistical procedures in problems of linear regression with special reference to quantitative bio-assays, I. *Internat. Statist. Rev. 39*, 21-38.

[74] Sen, P.K. (1972). Robust statistical procedures in problems of linear regression with special reference to quantitative bio-assays, II. *Internat. Statist. Rev. 40*, 161-172.

[75] Sen, P.K. (1979). Rank analysis of covariance under progressive censoring. *Sankhya, Ser.A 41*, 147-169.

[76] Sen, P.K. (1980). Nonparametric simultaneous inference for some MANOVA models. In *Handbook of Statistics, Vol. 1: Analysis of Variance* (ed. P.R. Krishnaiah), North Holland, Amsterdam, pp. 673-702.

[77] Sen, P.K. (1981a). The Cox regression model, invariance principles for some induced quantile processes and some repeated significance tests. *Ann. Statist. 9*, 109-121.

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

[79] Sen, P.K. (1984a). Some miscellaneous problems in nonparametric inference. In *Handbook of Statistics, Vol. 4: Nonparametric Methods* (Eds. P.R. Krishnaiah and P.K. Sen), North Holland, Amsterdam, pp. 699-739.

[80] Sen, P.K. (1984b). Multivariate nonparametric procedures for certain arteriosclerosis problems. In *Multivariate Analysis VI* (Ed. P.R. Krishnaiah), North Holland, Amsterdam, pp. 563-581.

[81] Sen, P.K. (1985). *Theory and Applications of Sequential Nonparametrics¿* SIAM, Philadelphia.

[82] Sen, P.K. (1988). Combination of statistical tests for multivariate hypotheses against restricted alternatives. In *Statistics: Applications and New Directions.* (Eds. S. Dasgupta and J.K. Ghosh), Ind. Statist. Inst., Calcutta, pp. 377-402.

[83] Sen, P.K. (1991a). Nonparametrics: Retrospectives and perspectives (with discussion). *J. Nonpara. Statist. 1*, 3-53.

[84] Sen, P.K. (1991b). Repeated significance tests in frequency and time domains. In *Handbook of Sequential Analysis* (Eds: B.K. Ghosh and P.K. Sen), Marcel Dekker, New York, pp.169-198.

[85] Sen, P.K. (1993a). Statistical perspectives in clinical and health sciences: The broadway to modern applied statistics. *J. Appl. Statist. Sc. 1*, 1-50.

[86] Sen, P.K. (1993b). Perspectives in multivariate nonparametrics: Conditional functionals and ANOCOVA models. *Sankhya, Ser.A 55*, 516-532.

[87] Sen, P.K. (1994a). Incomplete multiresponse designs and surrogate endpoints in clinical trials. *J. Statist. Plan. Infer. 42*, 161-186.

[88] Sen, P.K. (1994b). Some change-point problems in survival analysis: Relevance of nonparametrics in applications. *J. Appl. Statist. Sc. 1*, 425-444.

[89] Sen, P.K. (1994c). Bridging the biostatistics-epidemiology gap: The Bangladesh task. *J. Statist. Res. 28,* 21-39.

[90] Sen, P.K. (1995a). Nonparametric and robust methods in linear models with mixed effects. *Tetra Mount. Math. J.* , in press.

[91] Sen, P.K. (1995b). Paired comparisons for multiple characteristics: an ANOCOVA approach. In *H.A. David Festschrift* (Eds: H.N. Nagaraja, D.F. Morrison and P.K. Sen), Springer-Verlag, New York, in press.

[92] Sen, P.K. (1995c). Regression rank scores estimation in ANOCOVA. Submitted for publication.

[93] Sen, P.K. (1995d). Censoring in Theory and Practice: Statistical perspectives and controversies. (to appear).

[94] Sen, P.K. and David, H.A. (1968). Paired comparisons for paired characteristics. *Ann. Math. Statist. 39,* 200-208.

[95] Sen, P.K. and Margolin, B.H. (1995). Inhalation toxicology: Awareness, identifiability and statistical perspectives. *Sankhya, Ser.B 57,* in press.

[96] Sen, P.K. and Puri, M.L. (1977). Asymptotically distribution-free aligned rank order tests for composite hypotheses for general linear models. *Zeit. Wahrsch. verw. Geb. 39,* 175-186.

[97] Senn, S. (1993). *Cross-over Trials in Clinical Research.* John Wiley, New York.

[98] Servy, E.C. and Sen, P.K. (1987). Missing values in multisample rank permutation tests for MANOVA and MANOCOVA. *Sankhya, Ser.A 49,* 78-95.

[99] Shah, K.R. and Sinha, B.K. (1989). *Theory of Optimal Designs.* Lect. Notes Statist. No. 54, Springer-Verlag, New York.

[100] Sinha, A.N. and Sen, P.K. (1982). Tests based on empirical processes for progressive censoring schemes with staggering entry and random withdrawal. *Sankhya, Ser.B 44,* 1-18.

[101] Tudor, G. and Koch, G.G. (1994). Review of nonparametric methods for the analysis of cross over studies. *Statist. Meth. Med. Res. 3,* 345-381.

[102] Wei, L.J., Lin, D.Y. and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Amer. Statist. Assoc. 84,* 1065-1073.