CHAPTER 1

INTRODUCTION

1.1. WEIGHTED EMPIRICAL PROCESSES

A ***weighted empirical process*** (w.e.p.) corresponding to the random variables (r.v.'s) $X_{n1}, \ldots, X_{nn}$ and the non-random real weights $d_{n1}, \ldots, d_{nn}$ is defined to be

$$U_d(x) := \sum_{i=1}^{n} d_{ni}\, I(X_{ni} \le x), \qquad x \in \mathbb{R},\ n \ge 1.$$

The weights $\{d_{ni}\}$ need not be nonnegative.

The classical example of a w.e.p. is the *ordinary empirical process* that corresponds to $d_{ni} \equiv n^{-1}$. Another example is given by the two sample empirical process obtained as follows: Let m be an integer, $1 \le m \le n$, $r := n - m$; $d_{ni} = -r/n$, $1 \le i \le m$; $d_{ni} = m/n$, $m + 1 \le i \le n$. Then the corresponding U_d-process becomes

$$U_d(x) \equiv (mr/n)\, \{r^{-1} \sum_{i=m+1}^{n} I(X_{ni} \le x) - m^{-1} \sum_{i=1}^{m} I(X_{ni} \le x)\}, \qquad x \in \mathbb{R},$$

precisely the process that arises in two-sample models.

More generally, weighted empirical processes (w.e.p.'s) arise naturally in linear regression models where, for each $n \ge 1$ and each $\boldsymbol{\beta} \in \mathbb{R}^p$, the data $\{(\mathbf{x}_{ni}', Y_{ni}),\ 1 \le i \le n\}$ are related to the error variables $\{e_{ni},\ 1 \le i \le n\}$ by the linear relation

$$Y_{ni} = \mathbf{x}_{ni}'\boldsymbol{\beta} + e_{ni}, \quad 1 \le i \le n. \tag{1}$$

Here $e_{n1}, \ldots, e_{nn}$ are independent r.v.'s with respective continuous d.f.'s $F_{n1}, \ldots, F_{nn}$, $\mathbf{x}_{ni}' = (x_{ni1}, \ldots, x_{nip})$ is the i^{th} row of the known $n \times p$ design matrix $\mathbf{X}$ and $\boldsymbol{\beta}$ is the parameter vector of interest.

Consider the vector of w.e.p.'s $\mathbf{V} := (V_1, \ldots, V_p)'$ where

$$V_j(y, \mathbf{t}) := \sum_{i=1}^{n} x_{nij}\, I(Y_{ni} \le y + \mathbf{x}_{ni}'\mathbf{t}), \quad y \in \mathbb{R},\ \mathbf{t} \in \mathbb{R}^p,\ 1 \le j \le p. \tag{2}$$

Clearly, $V_j(\cdot, \mathbf{t})$ is an example of the w.e.p. $U_d(\cdot)$ with $d_{ni} \equiv x_{nij}$ and $X_{ni} \equiv Y_{ni} - \mathbf{x}_{ni}'\mathbf{t}$, $1 \le i \le n$, $1 \le j \le p$.

Observe that the data $\{(\mathbf{x}_{ni}', Y_{ni}),\ 1 \le i \le n\}$ in the model (1) are readily summarized by the vector of w.e.p.'s $\{\mathbf{V}(y, \mathbf{0}),\ y \in \mathbb{R}\}$ in the sense that the given data can be recovered from the sample paths of this vector up to a permutation. This in turn suffices for the purpose of inference about β in (1). In this sense the vector of w.e.p.'s $\{V(y, 0), y \in \mathbb{R}\}$ is at least as important to linear regression models (1) as is the ordinary empirical process to one-sample location models. One of the purposes of this monograph is to discuss the role of V-processes in inference and in proving limit theorems in models (1) in a unified fashion.

1.2. M-, R- AND SCALE ESTIMATORS

Many inferential procedures involving (1.1.1) can be viewed as functions of V. For example the least squares estimator, or more generally, the class of M-estimators corresponding to the score function ψ, (Huber: 1981), is defined as a solution $\mathbf{t}$ of the equation

$$\int \psi(y)\, V(dy, \mathbf{t}) = \text{a known constant}.$$

Similarly, rank (R) estimators of β corresponding to the score function φ are defined to be a solution $\mathbf{t}$ of the equation

$$\int \varphi(H_n(y, \mathbf{t}))V(dy, \mathbf{t}) = \text{a known constant}, \tag{1}$$

$$H_n(y, \mathbf{t}) := n^{-1} \sum_{i=1}^{n} I(Y_{ni} \le y + \mathbf{x}_{ni}'\mathbf{t}), \qquad y \in \mathbb{R}, \ \mathbf{t} \in \mathbb{R}^p.$$

A significant portion of Nonparametric Inference in models (1.1.1) deals with M- and R- estimators of β (Adichie; 1967. Huber; 1973) and linear rank tests of hypotheses about β, (Hájek–Sidák; 1967). By viewing these procedures as functions of $\{V(y, \mathbf{t}), y \in \mathbb{R}, \mathbf{t} \in \mathbb{R}^p\}$, it is possible to give a unified treatment of their asymptotic distribution theory, as is done in Chapters 3 and 4 below.

There is a vast literature in Nonparametric Inference that discusses inferential procedures based on functionals of empirical processes in the k-sample location model such as the books by Puri and Sen (1969), Serfling (1980) and Huber (1981). Yet their appropriate extensions to the linear regression model are not readily accessible. This monograph seeks to fill this void. The methodology and inference procedures studied here extend many known results in the k-sample location model to the model (1.1.1), thereby giving a unified treatment.

An important result needed for study of the asymptotic behavior of R-estimators of β is the asymptotic uniform linearity of the linear rank statistics of (1) in the regression parameter vector. Jurečková (1969, 1971) obtained this result under (1.1.1) with i.i.d. errors. A similar result was proved in Koul (1969, 1971) and Van Eeden (1972) for linear signed rank statistics under i.i.d. symmetric errors. Its extension to the case of nonidentically distributed errors is not readily available. Theorems 3.2.4 and 3.3.3 prove the asymptotic uniform linearity of linear rank and linear signed rank statistics with bounded scores under the general independent errors model (1.1.1). In the case of i.i.d. errors, the conditions in these theorems on the error d.f. are more general than requiring finite Fisher information. The results are proved uniformly over all bounded score functions and are consequences solely of the asymptotic sample continuity of V-processes and some smoothness of $\{F_{ni}\}$. The uniformity with respect to the score functions is useful when constructing adaptive rank tests that are asymptotically efficient against Pitman alternatives for a large class of error distributions.

Chapter 3 also contains a proof of the asymptotic normality of linear rank and linear signed rank statistics under independent alternatives and for indicator score functions. This proof proceeds via the weak convergence of certain basic w.e.p.'s and complements some of the results in Dupač and Hájek (1969).

Section 4.2a discusses the asymptotic distribution of M-estimators under heteroscedastic errors using the asymptotic continuity of V-processes. Section 4.2b presents some second order results on bootstrap approximations to the distributions of a class of M-estimators.

In order to make M-estimators scale invariant one often needs an appropriate robust scale estimator. One such scale estimator, as recommended by Huber (1981) and others, is

$$s_1 = \text{med}\,\{|Y_{ni} - \mathbf{x}_{ni}'\hat{\beta}|, \quad 1 \le i \le n\},$$

where $\hat{\beta}$ is an estimator of β. The asymptotic distribution of s_1 under heteroscedastic errors is given in Section 4.3. In the case of i.i.d. errors, this asymptotic distribution does not depend on $\hat{\beta}$ provided the errors are symmetric around 0. This observation naturally leads one to construct a scale estimator based on the symmetrized residuals, thereby giving another scale estimator

$$s_2 := \text{med}\,\{|Y_{ni} - \mathbf{x}_{ni}'\hat{\beta} - Y_{ni} + \mathbf{x}_{ni}'\hat{\beta}|;\ 1 \le i, j \le n\}.$$

As expected, the asymptotic distribution of s_2 is shown to be free from the estimator $\hat{\beta}$ in the case of i.i.d. errors, not necessarily symmetric. It also appears in Section 4.3.

Section 4.4 discusses the asymptotic distribution of a class of R-estimators under heteroscedastic errors using the asymptotic uniform linearity results of Chapter 3. The R-estimators considered are asymptotically equivalent to Jaeckel's estimators.

The complete rank analysis of the linear regression model (1.1.1) requires an estimate of the scale parameter

$$Q(f) := \int f \, d\varphi(F)$$

where f is density of the unknown common error d.f. F and φ is a nondecreasing function on (0, 1). This estimate is used to standardize the test statistic and estimate the standard error of the R-estimator corresponding to the score function φ. This parameter also appears in the efficiency comparisons of rank procedures and it is of interest to estimate it, after the fact, in an analysis.

Lehmann (1963), Sen (1966), Koul (1971), among others, provide estimators of Q(f) in the one- and two- sample location models and in the linear regression model. These estimators are given in terms of the lengths or Lebesgue measures of certain confidence intervals or regions. They are usually not easy to compute when the dimension p of β is larger than 1.

In Section 4.5, estimators of Q(f), based on kernel type density estimators of f and the empirical d.f. H_n, are defined and their consistency under (1.1.1) with i.i.d. errors is proved. An estimator whose window width is based on the data and is of the order of square root n, is also considered. The consistency proof presented is a sole consequence of the asymptotic continuity of certain w.e.p.'s and some smoothness of the error d.f.'s.

1.3. MINIMUM DISTANCE ESTIMATORS AND GOODNESS–OF–FIT TESTS

The practice of obtaining estimators of parameters by minimizing a certain distance between some functions of observations and parameters has been present in statistics since its beginning. The classical examples of this method are the Least Square and the minimum Chi Square estimators.

The minimum distance estimation (m.d.e.) method, where one obtains an estimator of a parameter by minimizing some distance between the empirical d. f. and the modeled d. f., was elevated to a general method of estimation by Wolfowitz (1953, 1954, 1957). In these papers he demonstrated that, compared to the maximum likelihood estimation method, the m.d.e. method yielded consistent estimators rather cheaply in several problems of varied levels of difficulty.

This methodology saw increasing research activity from the mid 1970's when many authors demonstrated various robustness properties of certain m.d. estimators. See, e.g., Beran (1977, 1978), Parr and Schucany (1979), Millar (1981, 1982, 1984), Donoho and Liu (1988 a, b), among others. All of these authors restrict their attention to the one sample setup or to the two sample location model. See Parr (1981) for additional bibliography on m.d.e. till 1980.

Inspite of many advances made in the m.d.e. methodology in one sample models, little was known till early 1980's as to how to extend this methodology to one of the most applied models, v.i.z., the multiple linear regression model (1.1.1). A ***significant*** advantage of viewing the model (1.1.1) through **V** is that one is naturally led to interesting m.d. estimators of β that are natural extensions of their one- and two- sample location model counterparts. To illustrate this, consider the m.d. estimator $\hat{\theta}$ of the one sample location parameter θ, when errors are i.i.d. symmetric around 0, defined by the relation

$$\hat{\theta} := \operatorname{argmin}\, \{T_n(t);\ t \in \mathbb{R}\},$$

with

$$T_n(t) = \int \{n^{-1/2} \sum_{i=1}^{n} I(Y_{ni} \le y + t) - I(-Y_{ni} < y - t)]\}^2 dG(y), \quad t \in \mathbb{R},$$

where $G \in \mathbb{D}\mathcal{I}(\mathbb{R})$. Since (1.1.1) is an extension of the one sample location model, it is only natural to seek an extension of $\hat{\theta}$ in this model. Assuming that $\{e_{ni}\}$ are symmetrically distributed around 0, the first thing one is tempted to consider as an extension of $\hat{\theta}$ is β_1^+ defined by the relation

$$\beta_1^+ := \operatorname{argmin}\, \{K_1^+(\mathbf{t});\ \mathbf{t} \in \mathbb{R}^p\},$$

with

$$K_1^+(\mathbf{t}) = \int \{n^{-1/2} \sum_{i=1}^{n} I(Y_{ni} \le y + \mathbf{x}_{ni}'\mathbf{t}) - I(-Y_{ni} < y - \mathbf{x}_{ni}'\mathbf{t})]\}^2 dG(y), \quad \mathbf{t} \in \mathbb{R}^p.$$

However, any extension of $\hat{\theta}$ to the linear regression model should have the property that it reduce to $\hat{\theta}$ when the model is reduced to the one sample location model and, in addition, that it reduce to an appropriate extension of $\hat{\theta}$ to the k-sample location model when the model (1.1.1) is reduced to it. In this sense β_1^+ does not provide the right extension but $\beta_{\mathbf{X}}^+$ does, where

$$\text{(1)} \qquad \beta_{\mathbf{X}}^+ := \operatorname{argmin}\, \{K_{\mathbf{X}}^+(\mathbf{t});\ \mathbf{t} \in \mathbb{R}^p\},$$

with

$$K_{\mathbf{X}}^+(\mathbf{t}) := \int \mathbf{V}^{+\prime}(y, \mathbf{t})\, (\mathbf{X}'\mathbf{X})^{-1}\, \mathbf{V}^+(y, \mathbf{t})\, dG(y), \qquad \mathbf{t} \in \mathbb{R}^p,$$

$$\mathbf{V}^{+\prime} := (V_1^+, \ldots, V_p^+),$$

$$V_j^+(y, \mathbf{t}) := V_j(y, \mathbf{t}) - \sum_{i=1}^{n} x_{nij} + V_j(-y, \mathbf{t}), \quad 1 \le j \le p, \ y \in \mathbb{R}, \mathbf{t} \in \mathbb{R}^p.$$

In the case errors are not symmetric but i.i.d. according to a known d.f. F, so that $EV_j(y, \beta) \equiv \Sigma_i\, x_{nij}\, F(y)$, a suitable class of m.d. estimators of β is defined by the relation

$$\hat{\beta}_{\mathbf{X}} := \operatorname{argmin}\, \{K_{\mathbf{X}}(\mathbf{t});\ \mathbf{t} \in \mathbb{R}^p\},$$

with

$$(2) \qquad K_X(t) := \int \|W(y, t)\|^2 \, dG(y), \qquad t \in \mathbb{R}^p,$$

$$W(y, t) := (X'X)^{-1/2}\{V(y, t) - X'1\, F(y)\}, \qquad y \in \mathbb{R},\ t \in \mathbb{R}^p,$$

$$1' := (1, \ldots, 1)_{1xn}.$$

Chapter 5 discusses the existence, the asymptotic distribution, the robustness and the asymptotic optimality of β_X^+ and $\hat{\beta}_X$ under (1.1.1) with heteroscedastic errors. For example, if $p = 1$ in (1.1.1) and the design variable is nonnegative then the asymptotic variance of β_X^+ is smaller than that of β_1^+ for a large class of symmetric error d.f.'s F and integrating measures G. A similar result holds about $\hat{\beta}_X$ and for $p \geq 1$. Chapter 5 also discusses several other m.d. estimators of β and their asymptotic theory under (1.1.1) with heteroscedastic errors. These include analogues of $\hat{\beta}_X$ when the common error d.f. is unknown and some m.d. estimators corresponding to certain supremum distances based on V.

Closely related to the problem of minimum distance estimation is the problem of testing the goodness-of-fit hypothesis H_0: $F_{ni} \equiv F_0$, F_0 a known d.f.. One test statistic for this problem is

$$\hat{D}_1 := \sup_y | n^{1/2}\{H_n(y, \hat{\beta}) - F_0(y)\} |,$$

where $\hat{\beta}$ is an estimator of β. This test statistic is suggested by looking at the estimated residuals and mimicking the one sample location model technique. In general, its large sample distribution depends on the design matrix. In addition, it does not reduce to the Kiefer (1959) tests of goodness-of-fit in the k-sample location problem when (1.1.1) is reduced to this model. Test statistics that overcome these deficiencies are

$$\hat{D}_2 := \sup_y | W^0(y, \hat{\beta}) |, \qquad \hat{D}_3 := \sup_y \|W^0(y, \hat{\beta})\|,$$

where W^0 is equal to the W of (2) with $F = F_0$. Another natural class of tests is based on $K_X^0(\hat{\beta}_X)$, where K_X^0 is equals to the K_X of (2) with W replaced by W^0 in there.

All of the above and several other goodness-of-fit tests are discussed at some length in Chapter 6. Section 6.2a discusses the asymptotic null distributions of the supremum distance statistics $\hat{D}_j$, $j = 1, 2, 3$. Also discussed in this section are asymptotically distribution free analogues of these tests, in a sense similar to that discussed by Durbin (1973, 1976) and Rao (1972) for the one-sample location model. Section 6.2b discusses smooth bootstrap approximations to the null distributions of tests based on w.e.p.'s.

Tests based on L_2-distances are discussed in Section 6.3. Some modifications of goodness-of-fit tests when F_0 has a scale parameter appear in Section 6.4 while tests of the symmetry of the errors are discussed in Section 6.5.

1.4. RANDOMLY WEIGHTED EMPIRICAL PROCESSES

A ***randomly weighted empirical process*** (r.w.e.p.) corresponding to the random variables (r.v.'s) $\zeta_{n1}, \ldots, \zeta_{nn}$, the random noise $\delta_{n1}, \ldots, \delta_{nn}$ and the random real weights $h_{n1}, \ldots, h_{nn}$ is defined to be

$$(1) \qquad V_h(x) := n^{-1} \sum_{i=1}^{n} h_{ni}\, I(\zeta_{ni} \le x + \delta_{ni}), \qquad x \in \mathbb{R},\ n \ge 1.$$

Examples of r.w.e.p.'s are provided by the w.e.p.'s $\{V_j;\ 1 \le j \le p\}$ of (1.1.2) in the case the design variables are random. More importantly, r.w.e.p.'s arise naturally in autoregression models. To illustrate this, let $\mathbf{Y}_0 = (X_0, \ldots, X_{1-p})'$ be an observable random vector, $\{\epsilon_i,\ i \ge 1\}$ be i.i.d. r.v.'s, independent of $\mathbf{Y}_0$, and $\boldsymbol{\rho}' = (\rho_1, \ldots, \rho_p)$ be a p-dimensional parameter vector. In the p^{th} order autoregression (AR(p)) model one observes $\{X_i\}$ obeying the relation

$$(2) \qquad X_i = \rho_1 X_{i-1} + \ldots + \rho_p X_{i-p} + \epsilon_i, \qquad i \ge 1,\quad \boldsymbol{\rho} \in \mathbb{R}^p.$$

Processes that play a fundamental role in the robust estimation of $\boldsymbol{\rho}$ in this model are randomly weighted residual empirical processes $\mathbf{T} = (T_1, \ldots, T_p)'$, where

$$(3) \qquad T_j(x, \mathbf{t}) := n^{-1} \sum_{i=1}^{n} g(X_{i-j})\, I(X_i \le x + \mathbf{t}'\mathbf{Y}_{i-1}), \qquad x \in \mathbb{R},\ \mathbf{t} \in \mathbb{R}^p,$$

$\mathbf{Y}'_{i-1} = (X_{i-1}, \ldots, X_{i-p})$, $i \ge 1$, and where g is a measurable function from $\mathbb{R}$ to $\mathbb{R}$. Clearly, for each $1 \le j \le p$, $T_j(x, \boldsymbol{\rho}+n^{-1/2}\mathbf{t})$ is an example of $V_h(x)$ with $\zeta_{ni} \equiv \epsilon_i$, $\delta_{ni} \equiv n^{-1/2}\,\mathbf{t}'\mathbf{Y}_{i-1}$ and $h_{ni} \equiv g(X_{i-j})$.

It is customary to expect that a method that works for linear regression models should have an analogue that will also work in autoregression models. Indeed the above inferential procedures based on w.e.p.'s in linear regression have perfect analogues in AR(p) models in terms of $\mathbf{T}$. The generalized M-estimators of $\boldsymbol{\rho}$ as proposed by Denby and Martin (1979) corresponding to the weight function g and the score functions ψ are given as a solution $\mathbf{t}$ of the p equations

$$\int \psi(x)\, T(dx, \mathbf{t}) = 0,$$

assuming that $E\psi(\epsilon) = 0$. Clearly, the classical least square estimator is obtained upon taking $g(x) \equiv x \equiv \psi(x)$ in these equations.

A generalized R-estimator $\hat{\rho}_R$ corresponding to a score function φ is defined by the relation

$$(4) \qquad \hat{\rho}_R := \text{argmin}\, \{\|S(\mathbf{t})\|;\ \mathbf{t} \in \mathbb{R}^p\})\|,$$

where

$$S(\mathbf{t}) := \int \varphi(F_n(x, \mathbf{t}))\, T(dx, \mathbf{t}),$$

$$F_n(x, \mathbf{t}) := n^{-1} \sum_{i=1}^{n} I(X_i \le x + \mathbf{t}'\mathbf{Y}_{i-1}), \qquad x \in \mathbb{R},\ \mathbf{t} \in \mathbb{R}^p.$$

An analogue of an R-estimator of (1.2.1) is obtained by taking $g(x) \equiv x$ in (4).

The m.d. estimators ρ_X^+ that are analogues of β_X^+ of (1.3.1) are defined as minimizers, w.r.t. $\mathbf{t} \in \mathbb{R}^p$, of

$$K(\mathbf{t}) := \sum_{j=1}^{p} \int [n^{-1/2} \sum_{i=1}^{n} X_{i-j}\{I(X_i \le x + \mathbf{t}'\mathbf{Y}_{i-1}) - I(-X_i \le x - \mathbf{t}'\mathbf{Y}_{i-1})\}]^2\, dG(x).$$

Observe that K involves T corresponding to $g(x) \equiv x$.

Chapter 7 discusses these and some other procedures in detail. Section 7.2 contains a result that says that the r.w.e.p.'s $\{T(x, \rho + n^{-1/2}\mathbf{t}),\ x \in \mathbb{R},\ \|\mathbf{t}\| \le B\}$ and the residual empirical processes $\{F_n(x, \rho + n^{-1/2}\mathbf{t}),\ x \in \mathbb{R},\ \|\mathbf{t}\| \le B\}$ are asymptotically uniformly linear in $\mathbf{t}$, for every $0 < B < \infty$. These results are used to investigate the asymptotic behavior of G-M and R-estimators in Sections 7.3a and 7.3b respectively. In order to carry out the rank analysis in AR(p) models, one needs a consistent estimator of $Q(f)$ where now f is the error density of $\{\epsilon_i\}$. A class of such estimators is given in Section 7.3c. A large class of m.d. estimators and their asymptotics appears in Section 7.4 whereas Section 7.5 briefly discusses some tests of goodness-of-fit hypotheses pertaining to the error d.f..

The contents of Chapter 2 are basic to those of Chapters 3, 4, and parts of Chapters 6 and 7. Sections 2.2a and 2.2b contain, respectively, proofs of the weak convergence of suitably standardized w.e.p.'s and r.w.e.p.'s to continuous Gaussian processes. Even though w.e.p.'s are a special case of r.w.e.p.'s, it is beneficial to investigate their weak convergence separately. For example, the weak convergence of U_d is obtained under a fairly general independent setup and minimal conditions on $\{d_{ni}\}$ whereas that of V_h is obtained under some hierarical dependence structure on $\{\eta_{ni}, h_{ni}, \delta_{ni}\}$ and the boundedness of the weights $\{h_{ni}\}$.

In Section 2.3, the asymptotic continuity of certain standardized w.e.p.'s is used to prove the asymptotic uniform linearity of $\mathbf{V}(., \mathbf{t})$ in $\mathbf{t}$, for $\mathbf{t}$ in certain shrinking neighborhoods of $\boldsymbol{\beta}$, under fairly general heteroscedastic errors. This result is found useful in Chapter 4 when discussing M-estimators and in Chapter 6 when discussing supremum distance test statistics for goodness-of-fit hypotheses. The asymptotic continuity is also found useful in Chapter 3 to prove various results about rank and signed rank statistics under heteroscedastic errors. The asymptotic continuity of V_{h}-processes is found useful in Chapter 7 when discussing the AR(p) model.

Chapter 2 concludes with results on functional and bounded laws of iterated logarithm pertaining to certain w.e.p.'s. It also includes an inequality due to Marcus and Zinn (1984) that gives an exponential bound on the tail probabilities of w.e.p.'s of independent r.v.'s. This inequality is an extension of the well celebrated Dvoretzky, Kiefer and Wolfowitz (1956) inequality for the ordinary empirical process. A result about the weak convergence of w.e.p.'s when r.v.'s are p-dimensional is also stated. These results are included for completeness, without proofs. They are not used in the subsequent sections. A martingale property of a properly centered $\mathrm{U_d}$ process is proved in Section 2.4.