Institute of Mathematical Statistics
LECTURE NOTES–MONOGRAPH SERIES

Weighted Empiricals and Linear Models

Hira L. Koul
Michigan State University

Volume 21

Institute of Mathematical Statistics
LECTURE NOTES–MONOGRAPH SERIES
Volume 21

Weighted Empiricals and Linear Models

Hira L. Koul
Michigan State University

Institute of Mathematical Statistics
Hayward, California

The production of the *IMS Lecture Notes–Monograph Series* is managed by the IMS Business Office: Jessica Utts, IMS Treasurer, and Jose L. Gonzalez, IMS Business Manager.

Library of Congress Catalog Card Number: 92-54658

International Standard Book Number 0-940600-28-5

Printed in the United States of America

To the memory of my parents

Prabhawati (Dhar) and Radhakrishen Koul (Gassi)

PREFACE

An empirical process that assigns possibly different non-random (random) weights to different observations is called a *weighted* (*randomly weighted*) *empirical process*. These processes are as basic to linear regression and autoregression models as the ordinary empirical process is to one sample models. However their usefulness in studying linear regression and autoregression models has not been fully exploited. This monograph addresses this question to a large extent.

There is a vast literature in Nonparametric Inference that discusses inferential procedures based on empirical processes in k-sample location models. However, their analogs in autoregression and linear regression models are not readily accessible. This monograph makes an attempt to fill this void. The statistical methodologies studied here extend to these models many of the known results in k-sample location models, thereby giving a unified theory.

By viewing linear regression models via certain weighted empirical processes one is naturally led to new and interesting inferential procedures. Examples include minimum distance estimators of regression parameters and goodness-of-fit tests pertaining to the errors in linear models. Similarly, by viewing autoregression models via certain randomly weighted empirical processes one is naturally led to classes of minimum distance estimators of autoregression parameters and goodness-of-fit tests pertaining to the error distribution.

The introductory Chapter 1 gives an overview of the usefulness of weighted and randomly weighted empirical processes in linear models. Chapter 2 gives general sufficient conditions for the weak convergence of suitably standardized versions of these processes to continuous Gaussian processes. This chapter also contains the proof of the asymptotic uniform linearity of weighted empirical processes based on the residuals when errors are heteroscedastic and independent. Chapter 3 discusses the asymptotic uniform linearity of linear rank and signed rank statistics when errors are heteroscedastic and independent. It also includes some results about the weak convergence of weighted empirical processes of ranks and signed ranks. Chapter 4 is devoted to the study of the asymptotic behavior of M- and R-estimators of regression parameters under heteroscedastic and independent errors, via weighted empirical processes. A brief discussion about bootstrap approximations to the distribution of a class of M-estimators appears in Section 4.2b. This chapter also contains a proof of the consistency of a class of robust estimators for certain scale parameters under heteroscedastic errors.

In carrying out the analysis of variance of linear regression models based on ranks, one often needs an estimator of the functional $\int f d\varphi(F)$, where F is the error distribution function, f its density and φ is a function from $[0, 1]$ to the real line. Some estimators of this functional and the proofs of their consistency in the linear regression setting appear in Section 4.5.

Chapters 5 and 6 deal with minimum distance estimation, via weighted empirical processes, of the regression parameters and tests of goodness-of-fit pertaining to the error distribution. One of the main themes emerging from these two chapters is that the inferential procedures based on weighted empiricals with weights proportional to the design matrix provide the right extensions of k-sample location model procedures to linear regression models.

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 many of the inferential procedures based on weighted empirical processes in linear regression that are discussed in Chapters 3-6 have precise analogs in autoregression based on certain randomly weighted empirical processes and appear in Chapter 7. In particular, the proof of the asymptotic uniform linearity of the ordinary empirical process of the residuals in autoregression appears here.

All asymptotic uniform linearity results in the monograph are shown to be consequences of the asymptotic continuity of certain basic weighted and randomly weighted empirical processes.

Chapters 2-4 are interdependent. Chapter 5 is mostly self-contained and can be read after reading the Introduction. Chapter 6 uses results from Chapters 2 and 5. Chapter 7 is almost self-contained. The basic result needed for this chapter appears in Section 2.2b.

The first version of this monograph was prepared while I was visiting the Department of Statistics, Poona University, India, on sabbatical leave from Michigan State University, during the academic year 1982–83. Several lectures on some parts of this monograph were given at the Indian Statistical Institute, New Delhi, and Universities of La Trobe, Australia, and Wisconsin, Madison. I wish to thank Professors S. R. Adke, Richard Johnson, S. K. Mitra, M. S. Prasad and B. L. S. Prakasa Rao for having some discussions pertaining to the monograph. My special thanks go to James Hannan for encouraging me to finish the project and for proof reading parts of the manuscript, to Soumendra Lahiri for helping me with sections on bootstrapping, and to Bob Serfling for taking keen interest in the monograph and for many comments that helped to improve the initial draft.

Ms. Achala Sabane and Ms. Lora Kemler had the pedestrian task of typing the manuscript. Their patient endeavors are gratefully acknowledged. Ms. Kemler's keen eye for details has been an indispensable help.

During the preparation of the monograph the author was partly supported by the University Grants Commission of India and the National Science Foundation, grant numbers NSF 82–01291, DMS–9102041.

Hira L. Koul,
East Lansing, MI. 48823
5/28/92.

NOTATION AND CONVENTIONS

The p–dimension Euclidean space is denoted by $\mathbb{R}^p$, $p \geq 1$; $\mathbb{R} = \mathbb{R}^1$; $\mathscr{B}^p$:= the σ–algebra of Borel sets in $\mathbb{R}^p$, $\mathscr{B} = \mathscr{B}^1$; λ := Lebesgue measure on $(\mathbb{R}, \mathscr{B})$. The symbol ":=" stands for "by definition".

For any set $A \subset \mathbb{R}$, $\mathbb{D}(A)$ denotes the class of real valued functions on A that are right continuous and have left limits while $\mathbb{D}\mathcal{I}(A)$ denotes the subclass in $\mathbb{D}(A)$ whose members are nondecreasing. $\mathbb{C}[0, 1]$:= the class of real valued bounded continuous functions on $[0, 1]$.

A vector or a matrix will be designated by a bold letter. A $\mathbf{t} \in \mathbb{R}^p$ is a p×1 vector, $\mathbf{t}'$ or $\mathbf{t}'$ its transpose, $\|\mathbf{t}\|^2 := \sum_{j=1}^{p} t_j^2$, $|\mathbf{t}| := \max\{|t_j|, 1 \leq j \leq p\}$. For any p-square matrix $\mathbf{C}$, $\|\mathbf{C}\|_\infty = \sup\{\|\mathbf{t}'\mathbf{C}\|;\ \|\mathbf{t}\| \leq 1\}$. For an n×p matrix $\mathbf{D}$, $\mathbf{d}_{ni}'$ denotes its i^{th} row, $1 \leq i \leq n$, and $\mathbf{D}_c$ the n×p matrix $\mathbf{D} - \overline{\mathbf{D}}$, whose i^{th} row consists of $(\mathbf{d}_{ni} - \bar{\mathbf{d}}_n)'$, with $\bar{\mathbf{d}}_n := \Sigma_i\, \mathbf{d}_{ni}/n$, $1 \leq i \leq n$.

w.e.p.('s)	:= weighted empirical process(es).
r.w.e.p.('s)	:= randomly weighted empirical process(es).
i.i.d.	:= independent identically distributed.
r.v.('s)	:= random variable(s).
d.f.('s)	:= distribution function(s).
w.r.t.	:= with respect to.
C-S	:= the Cauchy–Schwarz inequality.
D.C.T.	:= the Dominated Convergence Theorem
Fubini	:= the Fubini Theorem.
L-F CLT	:= the Lindeberg–Feller Central Limit Theorem.
$o(1)(o_p(1))$	:= a sequence of numbers (r.v.'s) converging to zero (in probability).
$O(1)(O_p(1))$	:= a sequence of numbers (r.v.'s) that is bounded (in probability).
$N(\mathbf{0}, \mathbf{C})$	:= either a r.v. with normal distribution whose mean vector is $\mathbf{0}$ and the covariance matrix $\mathbf{C}$ or the corresponding distribution.
$\|g\|_\infty$	:= the supremum norm over the domain of g, g a real valued function.
τ_a^2	:= $\sum_{i=1}^{n} a_{ni}^2$, for an arbitrary real vector $(a_{n1}, \ldots, a_{nn})'$.

Often in a discussion or in a proof the subscript n on the triangular arrays and various other quantities will not be exhibited. The index i in Σ_i or $\sum_i$ and $\max_i$ or $\max_i$ will vary from 1 to n, unless specified otherwise. All limits, unless specified otherwise, are taken as $n \longrightarrow \infty$.

For a sequence of r.v.'s $\{X, X_n, n \geq 1\}$, $X_n \underset{d}{\to} X$ means that the distribution of X_n converges weakly to that of X. For two r.v.'s X, Y, $X \underset{d}{=} Y$ means that the distribution of X is the same as that of Y.

For a sequence of stochastic processes $\{Y, Y_n, n \geq 1\}$, $Y_n \Rightarrow Y$ means that Y_n converges weakly to Y in a given topology. $Y_n \underset{fd}{\to} Y$ means that all finite dimensional distributions of Y_n converge weakly to that of Y.

Reference to an expression or a display is made by the (expression number) if referring in the same section and by the (chapter number.section number.expression number), otherwise. For example, by (3.2.1) is meant an expression (1) of Section 2 of Chapter 3. A reference to this while in Section 3.2 would appear as (1).

For convenient reference we list here some of the most often used conditions in the manuscript. For an arbitrary d.f. F on $\mathbb{R}$, conditions **(F1)**, **(F2)** and **(F3)** are as follows:

(F1) F has uniformly continuous density f w.r.t. λ.

(F2) $f > 0$, a.e. λ.

(F3) $\sup_{x \in \mathbb{R}} |xf(x)| < \infty$.

These conditions are introduced for the first time just before Corollary 3.2.1 and are used frequently subsequently.

For an $n \times p$ design matrix matrix X, the conditions (NX), (NX1) and (NX_c) are as follows:

(NX) $(X'X)^{-1}$ exists, $n \geq p$; $\max_i \mathbf{x}_{ni}'(X'X)^{-1}\mathbf{x}_{ni} = o(1)$.

(NX1)
$$(X_c'X_c)^{-1} \text{ exists, } n \geq p;$$
$$\max_i \mathbf{x}_{ni}'(X_c'X_c)^{-1}\mathbf{x}_{ni} = o(1).$$

(NX_c)
$$(X_c'X_c)^{-1} \text{ exists, } n \geq p;$$
$$\max_i (\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)'(X_c'X_c)^{-1}(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n) = o(1).$$

The condition (NX) is the most often used from Theorem 2.3.3. onwards. The letter N in these conditions stands for Noether, who was the first person to use (NX), in the case p=1, to obtain the asymptotic normality of weighted sums of r.v.'s; see Noether (1949).

TABLE OF CONTENTS