

ET

19

05348

1987

SERIE RESEARCH MEMORANDA

CONVERGENCE

Herman J. Bierens

Researchmemorandum 1987-19 April '87



VRIJE UNIVERSITEIT

Faculteit der Economische Wetenschappen en Econometrie

A M S T E R D A M

CONVERGENCE *

Herman J. Bierens
Department of Econometrics
Free University, Amsterdam

December 1988

* Chapter 2 of:

PRINCIPLES OF NONLINEAR AND NONPARAMETRIC REGRESSION ANALYSIS

To be published by Marcel Dekker Inc., New York.



TABLE OF CONTENTS:

PREFACE

PART 1: PRELIMINARY MATHEMATICS

- 1. BASIC PROBABILITY THEORY
 - 1.1 Measure-theoretical foundation of probability theory
 - 1.2 Independence
 - 1.3 Borel measurable functions
 - 1.4 Mathematical expectation
 - 1.5 Characteristic functions
 - 1.6 Random functions
- 2. CONVERGENCE
 - 2.1 Weak and strong convergence of random variables
 - 2.2 Convergence of mathematical expectations
 - 2.3 Convergence of distributions
 - 2.4 Central limit theorems
 - 2.5 Further results on convergence of distributions and mathematical expectations, and laws of large numbers
 - 2.6 Convergence of random functions
 - 2.7 Uniform strong and weak laws of large numbers
- 3. INTRODUCTION TO CONDITIONING
 - 3.1 Definition of conditional expectation
 - 3.2. Basic properties of conditional expectations
 - 3.3 Identification of conditional expectations

PART 2: REGRESSION ANALYSIS UNDER INDEPENDENCE

- 4. NONLINEAR PARAMETRIC REGRESSION ANALYSIS
 - 4.1 Nonlinear regression models and the nonlinear least squares estimator
 - 4.2 Consistency and asymptotic normality: General theory
 - 4.2.1 Consistency
 - 4.2.2 Asymptotic normality
 - 4.3 Consistency and asymptotic normality of nonlinear least squares estimators in the i.i.d. case
 - 4.3.1 Consistency

- 4.3.2 Asymptotic normality
- 4.3.3 Consistent estimation of the asymptotic variance matrix
- 4.4 Consistency and asymptotic normality of the nonlinear least squares estimator under data heterogeneity
 - 4.4.1 Data heterogeneity
 - 4.4.2 Strong and weak consistency
 - 4.4.3 Asymptotic normality
- 4.5 Testing parameter restrictions: The Wald test
- 5. TESTS FOR MODEL MISSPECIFICATION
 - 5.1 White's version of Hausman's test
 - 5.2 Newey's M-test
 - 5.2.1 Introduction
 - 5.2.2 The conditional M-test
 - 5.3 A consistent conditional M-test
 - 5.4 The integrated M-test
- 6. THE NADARAYA-WATSON KERNEL REGRESSION FUNCTION ESTIMATOR
 - 6.1 Introduction
 - 6.2 Asymptotic normality in the continuous case
 - 6.3 Uniform consistency in the continuous case
 - 6.4 Discrete and mixed continuous-discrete regressors
 - 6.4.1 The discrete case
 - 6.4.2 The mixed continuous-discrete case
 - 6.5 The choice of the kernel
 - 6.6 The choice of the window width
 - 6.7 An empirical application to specification of household expenditure systems and equivalence scales
 - 6.7.1 Introduction
 - 6.7.2 Model and data
 - 6.7.3 The results
 - 6.7.4 Sample selection
- 7. SAMPLE MOMENTS INTEGRATING NORMAL KERNEL ESTIMATORS
 - 7.1 Kernel density estimators
 - 7.1.1 Integral conditions
 - 7.1.2 Sample moments integrating kernel density estimators
 - 7.1.3 The Dirac-catastrophe and the modified SMINK density estimator

- 7.1.4 How to choose the window width of the modified SMINK density estimator
- 7.2 Regression
 - 7.2.1 SMINK estimators of a regression function
 - 7.2.2 How to choose the window width parameters
- 7.3 A numerical example
- 7.4 Proofs
- 8. NONLINEAR REGRESSION WITH DISCRETE EXPLANATORY VARIABLES
 - 8.1. The earnings function
 - 8.2. The functional form of a regression model with discrete explanatory variables
 - 8.3. The choice of the linear separator
 - 8.4. Estimating and testing the regression function
 - 8.4.1 Estimation
 - 8.4.2 Model specification testing
 - 8.4.3 The selection of the polynomial order
 - 8.5 Proofs
 - 8.5.1 Proof of theorem 8.4.2
 - 8.5.2 Proof of theorem 8.4.3

PART 3: TIME SERIES

- 9. CONDITIONING AND DEPENDENCE
 - 9.1 Conditional expectations relative to a Borel field
 - 9.1.1 Definition and basic properties
 - 9.1.2 Martingales
 - 9.1.3 Martingale convergence theorems
 - 9.1.4 A martingale difference central limit theorem
 - 9.2 Measures of dependence
 - 9.2.1 Mixingales
 - 9.2.2 Uniform and strong mixing
 - 9.2.3 ν -Stability
 - 9.3 Weak laws of large numbers for dependent random variables
 - 9.4 Proper heterogeneity and uniform laws for functions of infinitely many random variables
- 10. FUNCTIONAL SPECIFICATION OF TIME SERIES MODELS
 - 10.1 Linear times series regression models
 - 10.1.1 Introduction

- 10.1.2 The Wold decomposition
- 10.1.3 Linear vector time series models
- 10.2 ARMA memory index models
 - 10.2.1 Introduction
 - 10.2.2 Finite conditioning of univariate rational-valued time series
 - 10.2.3 Infinite conditioning of univariate rational-valued time series
 - 10.2.3 The multivariate case
 - 10.2.4 The nature of the ARMA memory index parameters and the response functions
 - 10.2.5 Discussion
- 10.3 Nonlinear ARMAX models
- 11. ARMAX MODELS: ESTIMATION AND TESTING
 - 11.1 Estimation of linear ARMAX models
 - 11.1.1 Introduction
 - 11.1.2 Consistency and asymptotic normality
 - 11.2 Estimation of nonlinear ARMAX models
 - 11.3 A consistent $N(0,1)$ model specification test
 - 11.4 A consistent Hausman-type model specification test
 - 11.5 An autocorrelation test
- 12. NONPARAMETRIC TIME SERIES REGRESSION
 - 12.1 Assumptions and preliminary lemmas
 - 12.2 Consistency and asymptotic normality of time series kernel regression function estimators



PREFACE

This book deals with statistical inference of nonlinear regression models from two opposite points of view, namely the case where the functional form of the model is completely specified as a known function of regressors and unknown parameters, and the opposite case where the functional form of the model is completely unknown. First it is assumed that the response function of the regression model under review belongs to a certain well-specified parametric family of functional forms, by which estimation of the model merely amounts to estimation of the unknown parameters. For this class of models we review the asymptotic properties of the nonlinear least squares estimator for independent data as well as for time series.

In practice assumptions on the functional form are often made on the basis of computational convenience rather than on the basis of precise a priori knowledge of the empirical phenomenon under review. Therefore the linear regression model is still the most popular model specification in applied research. However, even if the specification of the functional form is based on sound theoretical considerations there is quite often a large range of functional forms that are theoretically admissible, so that there is no guarantee that the actually chosen functional form is true. Functional specification of a parametric nonlinear regression model should therefore always be verified by conducting model misspecification tests. Various model misspecification tests will therefore be discussed, in particular consistent tests which have asymptotic power 1 against all deviations from the null hypothesis that the model is correct.

The opposite case of parametric regression is nonparametric regression. Nonparametric regression analysis is concerned with estimation of a regression model without specifying in advance its functional form. Thus the only source of information about the functional form of the model is the data set itself. In this book we shall review various nonparametric regression approaches, with special emphasis on the kernel method, under various distributional assumptions.

This book is divided into three parts. In the first part we review the elements of abstract probability theory we need in part 2. Part 2 is devoted to the asymptotic theory of para-

metric and nonparametric regression analysis in the case of independent data generating processes. In part 3 we extend the analysis involved to time series.

The selection of the topics mainly reflexes my own interest in the subject. Instead of providing an encyclopedic survey of the literature, I have chosen for a setup which aims to fill the gap between intermediate statistics (including linear time series analysis) and the level necessary to get access to the recent literature on nonlinear and nonparametric regression analysis, with emphasis on my own contributions. The ultimate goal is to provide the student with the tools for his own independent research in this area, by showing what tools I and others have used and what they have been used for. Thus, this book may be viewed as an account of my own struggle with the material involved. I think this book is particularly suitable for self-tuition (at least it aims to be), and may prove useful in a graduate course in mathematical statistics and advanced econometrics.

Acknowledgements:

The first five chapters of this book have been disseminated in draft form as working papers. I am grateful to Anil Bera, Alexander Georgiev and Jan Magnus for suggesting additional references, and in particular to Lourens Broersma, Johan Smits and Ton Steerneman who suggested various improvements.

A large body of the material in chapter 6 has been published earlier in Truman F. Bewley (ed.), *Advances in Econometrics, Fifth World Congress*, Cambridge University Press. I am indebted to Cambridge University Press for granting permission to reprint it.

2. CONVERGENCE

In this chapter we consider various modes of convergence, i.e., weak and strong convergence of random variables, weak and strong laws of large numbers, convergence in distribution and central limit theorems, weak and strong uniform convergence of random functions and uniform weak and strong laws. The material in this chapter is a revision and extension of sections 2.2-2.4 in Bierens (1981).

2.1 Weak and strong convergence of random variables

In this section we shall deal with the concepts of convergence in probability and almost sure convergence, and various laws of large numbers. Throughout we assume that the random variables involved are defined on a common probability space (Ω, \mathcal{F}, P) . The first concept is well-known:

Definition 2.1.1. Let (X_n) be a sequence of r.v.'s. We say that X_n converges in probability to a r.v. X if for every $\epsilon > 0$, $\lim_{n \rightarrow \infty} P(|X_n - X| \leq \epsilon) = 1$, and we write: $X_n \rightarrow X$ in pr. or $\text{plim}_{n \rightarrow \infty} X_n = X$.

However, almost sure convergence is much stronger a convergence concept:

Definition 2.1.2. Let (X_n) be a sequence of r.v.'s. We say that X_n converges almost surely (a.s.) to a r.v. X if there is a null set $N \in \mathcal{F}$ (that is a set in \mathcal{F} satisfying $P(N) = 0$) such that for every $\omega \in \Omega \setminus N$, $\lim_{n \rightarrow \infty} X_n(\omega) = x(\omega)$, and we write: $X_n \rightarrow X$ a.s. or $\lim_{n \rightarrow \infty} X_n = X$ a.s.

Note that this definition is equivalent with:

$$X_n \rightarrow X \text{ a.s. if } P(\{\omega \in \Omega: \lim_{n \rightarrow \infty} X_n(\omega) = x(\omega)\}) = 1.$$

A useful criterion for almost sure convergence of random variables is given by the following theorem.

Theorem 2.1.1. Let X and X_1, X_2, \dots , be random variables. Then $X_n \rightarrow X$ a.s. if and only if

$$\lim_{n \rightarrow \infty} P(\cap_{m=n}^{\infty} \{|X_m - X| \leq \varepsilon\}) = 1 \quad \text{for every } \varepsilon > 0. \quad (2.1.1)$$

Proof: First we prove that $x(\omega) = \lim_{n \rightarrow \infty} x_n(\omega)$ pointwise on $\Omega \setminus N$ implies (2.1.1). Let $\omega_0 \in \Omega \setminus N$. Then for every $\varepsilon > 0$ there is a number $n_0(\omega_0, \varepsilon)$ such that

$$|x_n(\omega_0) - x(\omega_0)| \leq \varepsilon \quad \text{for all } n \geq n_0(\omega_0, \varepsilon).$$

Now consider the following set in F .

$$A_n(\varepsilon) = \cap_{m=n}^{\infty} \{\omega \in \Omega : |x_m(\omega) - x(\omega)| \leq \varepsilon\}.$$

Then $\omega_0 \in A_{n_0}(\omega_0, \varepsilon)(\varepsilon)$ and hence $\omega_0 \in \cup_n A_n(\varepsilon)$. Thus we have:

$$\Omega \setminus N \subset \cup_n A_n(\varepsilon)$$

and consequently

$$P(\Omega \setminus N) \leq P(\cup_n A_n(\varepsilon)).$$

But $P(\Omega \setminus N) = P(\Omega) - P(N) = 1$ since N is a null set, hence

$$P(\cup_n A_n(\varepsilon)) = 1.$$

Since $A_n(\varepsilon) \subset A_{n+1}(\varepsilon)$, we have $A_k(\varepsilon) = \cup_{n=1}^k A_n(\varepsilon)$ and thus

$$\lim_{n \rightarrow \infty} P(A_k(\varepsilon)) = \lim_{n \rightarrow \infty} P(\cup_{n=1}^k A_n(\varepsilon)) = P(\cup_n A_n(\varepsilon)) = 1,$$

which proves the first part of the theorem.

Next we prove that if $\lim_{n \rightarrow \infty} P(A_n(\varepsilon)) = 1$ for every $\varepsilon > 0$ then there exists a null set N such that

$$x(\omega) = \lim_{n \rightarrow \infty} x_n(\omega) \quad \text{pointwise on } \Omega \setminus N.$$

For $\delta > 0$ put $N_\delta = \Omega \setminus \cup_n A_n(\delta)$. Then $N_\delta \in F$ and

$$P(N_\delta) = P(\Omega) - P(\cup_n A_n(\delta)),$$

hence N_δ is a null set. Define

$$N = \cup_k N_{1/k}.$$

Then N is a countable union of null sets in F and therefore a null set itself. Let $\omega_0 \in \Omega \setminus N$, then $\omega_0 \in \cap_k (\cup_n A_n(1/k))$. Suppose that k is arbitrarily chosen but fixed. Then $\omega_0 \in A_n(1/k)$. Therefore, there exists an $n_0 = n_0(1/k)$ such that

$$|x_n(\omega_0) - x(\omega_0)| \leq 1/k \text{ for } n \geq n_0.$$

It is obvious now that $x_n(\omega_0)$ converges to $x(\omega_0)$ pointwise on $\Omega \setminus N$. This proves the 'only if' part of the theorem. Q.E.D.

From this theorem we see that

Corollary 2.1.1. $X_n \rightarrow X$ a.s. implies $X_n \rightarrow X$ in pr.,

Proof: Note that

$$\cap_{m=n}^{\infty} \{|X_m - X| \leq \varepsilon\} \subset \{|X_n - X| \leq \varepsilon\}$$

and consequently

$$P(\cap_{m=n}^{\infty} \{|X_m - X| \leq \varepsilon\}) \leq P(\{|X_n - X| \leq \varepsilon\}).$$

Q.E.D.

The following simple but important theorem provides another useful criterion for almost sure convergence.

Theorem 2.1.2. (Borel-Cantelli lemma) If for every $\varepsilon > 0$,

$$\sum_n P(|X_n - X| > \varepsilon) < \infty,$$

then $X_n \rightarrow X$ a.s.

Proof: Consider the set

$$\begin{aligned} A_n(\varepsilon) &= \cap_{m=n}^{\infty} \{\omega \in \Omega: |x_m(\omega) - x(\omega)| \leq \varepsilon\} \\ &= \cap_{m=n}^{\infty} \{|X_m - X| \leq \varepsilon\}. \end{aligned}$$

From theorem 2.1.1 it follows that it suffices to show

$$P(A_n(\varepsilon)) \rightarrow 1 \text{ or equivalently, } P(A_n(\varepsilon))^c \rightarrow 0 .$$

But

$$A_n(\varepsilon)^c = \bigcup_{m=n}^{\infty} \{ |X_m - X| > \varepsilon \},$$

and hence

$$P(A_n(\varepsilon))^c \leq \sum_{m=n}^{\infty} P(|X_m - X| > \varepsilon).$$

Since the latter sum is a tail sum of the convergent series

$$\sum_n P(|X_n - X| > \varepsilon)$$

we must have

$$\sum_{m=n}^{\infty} P(|X_m - X| > \varepsilon) \rightarrow 0 \text{ as } n \rightarrow \infty ,$$

which proves the theorem.

Q.E.D.

The a.s. convergence concept arises in a natural way from the strong laws of large numbers. Here we give three versions of these laws.

Theorem 2.1.3. Let (X_j) be a sequence of uncorrelated random variables satisfying $E(X_j - EX_j)^2 = O(j^\mu)$ for some $\mu < 1$. Then

$$(1/n) \sum_{j=1}^n (X_j - EX_j) \rightarrow 0 \text{ a.s.}$$

Proof: This theorem is a further elaboration of the strong law of large numbers of Rademacher-Menchov [see Révész (1968), theorem 3.2.1. or Stout (1974), theorem 2.3.2.], which states:

Let (Y_j) , $j > 0$, be a sequence of orthogonal random variables (orthogonality means that $EY_{j_1}Y_{j_2} = 0$ if $j_1 \neq j_2$). If

$$\sum_j (\log j) E Y_j^2 < \infty$$

then $\sum_j Y_j$ converges a.s. (which means that $Z = \sum_j Y_j$ is a.s. a

finite valued random variable).

Now let $Y_j = (X_j - EX_j)/j$. Then

$$\sum_j (\log j) E Y_j^2 = \sum_j (\log j) O(j^{\mu-2}) < \infty \text{ for } \mu < 1,$$

hence $\sum_j (X_j - EX_j)/j$ converges a.s. From the Kronecker lemma [see Révész (1968), theorem 1.2.2. or Chung (1974), p.123] it follows that this result implies the theorem under review.

Q.E.D.

Theorem 2.1.4. Let (X_j) be a sequence of uncorrelated random variables satisfying:

$$\sup_n (1/n) \sum_{j=1}^n E |X_j - EX_j|^{2+\delta} < \infty \text{ for some } \delta > 0.$$

Then $(1/n) \sum_{j=1}^n (X_j - EX_j) \rightarrow 0$ a.s.

Proof: The moment condition in this theorem implies

$$E |X_j - EX_j|^{2+\delta} = O(j),$$

so that by Liapounov's inequality

$$\begin{aligned} E(X_j - EX_j)^2 &\leq (E |X_j - EX_j|^{2+\delta})^{2/(2+\delta)} \\ &= O(j^{2/(2+\delta)}). \end{aligned}$$

The theorem now follows from theorem 2.1.3.

Q.E.D.

If the X_j 's are independent identically distributed (i.i.d.) the condition on the second moment is not needed:

Theorem 2.1.5: (Strong law of large numbers of Kolmogorov) Let (X_j) be a sequence of i.i.d. random variables satisfying $E |X_1| < \infty$. Then

$$(1/n) \sum_{j=1}^n X_j \rightarrow EX_1 \text{ a.s.}$$

Proof: Chung (1974, theorem 5.4.2).

We already have mentioned that almost sure convergence implies convergence in probability. There is also a converse connection, given by the following theorem.

Theorem 2.1.6. Let X and X_1, X_2, \dots be r.v.'s. Then $X_n \rightarrow X$ in pr. if and only if every subsequence (n_k) of the sequence (n) contains a further subsequence (n_{k_j}) such that

$$X_{n_{k_j}} \rightarrow X \text{ a.s. as } j \rightarrow \infty \quad (2.1.2)$$

Proof: Suppose that every subsequence (n_k) contains a further subsequence (n_{k_j}) such that (2.1.2) holds, but that not $X_n \rightarrow X$ in pr. Then there exist numbers $\epsilon > 0$, $\delta > 0$, and a subsequence (n_k) such that

$$P(|X_{n_k} - X| \leq \epsilon) \leq 1 - \delta,$$

hence for every further subsequence we have the same, which contradicts our assumption. Thus the 'if' part is now proved. Next, suppose that $X_n \rightarrow X$ in pr. Then for every positive integer k ,

$$\lim_{n \rightarrow \infty} P(|X_n - X| > 1/2^k) = 0.$$

For each k we can find an n_k such that

$$P(|X_{n_k} - X| > 1/2^k) \leq 1/2^k,$$

hence

$$\sum_k P(|X_{n_k} - X| > 1/2^k) \leq \sum_k 1/2^k < \infty$$

and consequently:

$$\sum_k P(|X_{n_k} - X| > \epsilon) < \infty \text{ for every } \epsilon > 0.$$

By the Borel-Cantelli lemma it follows now that $X_{n_k} \rightarrow X$ a.s., which proves the 'only if' part. Q.E.D.

Theorem 2.1.7. Let X and X_1, X_2, \dots be r.v.'s such that (a) $X_n \rightarrow X$ a.s. or (b) $X_n \rightarrow X$ in pr., respectively. Let $f(x)$ be a Borel measurable real function on R . If f is continuous on a Borel set B such that $P(X \in B) = 1$, then (a) $f(X_n) \rightarrow f(X)$ a.s. or (b) $f(X_n) \rightarrow f(X)$ in pr., respectively.

Proof:

(a) There is a null set N_1 such that for every $\epsilon > 0$ and every $\omega \in \Omega \setminus N_1$,

$$|x_n(\omega) - x(\omega)| \leq \epsilon \text{ if } n \geq n_0(\omega, \epsilon).$$

Let $N_2 = \Omega \setminus \{\omega \in \Omega : x(\omega) \in B\}$. Then N_2 is a null set in F and $x(\omega)$ is for every $\omega \in \Omega \setminus N_2$ a continuity point of f . Thus for every $\epsilon > 0$ and every $\omega \in \Omega \setminus N_2$ there is a number $\delta(\epsilon, \omega) > 0$ such that:

$$|f(x_n(\omega)) - f(x(\omega))| \leq \epsilon \text{ if } |x_n(\omega) - x(\omega)| \leq \delta(\epsilon, \omega),$$

hence for every $\epsilon > 0$ and every $\omega \in \Omega \setminus (N_1 \cup N_2)$ we have:

$$|f(x_n(\omega)) - f(x(\omega))| \leq \epsilon \text{ if } n \geq n_0(\omega, \delta(\epsilon, \omega)).$$

Since $N_1 \cup N_2$ is a null set in F , this proves part (a) of the theorem.

(b) For an arbitrary subsequence (n_k) we have a further subsequence (n_{k_j}) such that (2.1.2) holds and consequently by part (a) of the theorem:

$$f(X_{n_{k_j}}) \rightarrow f(X) \text{ a.s.}$$

By theorem 2.1.6 this implies $f(X_n) \rightarrow f(X)$ in pr. Q.E.D.

Remark: Until so far in this section we only have dealt with random variables in R . However, generalisation of the definitions and the theorems in this section to finite dimensional random vectors is straightforward, simply by changing random variable to random vector. The only notable extension is:

Theorem 2.1.8. Let $X_n = (X_{1n}, \dots, X_{kn})'$ and $X = (X_1, \dots, X_k)'$ be random vectors. Then $X_n \rightarrow X$ a.s. (in prob.) if and only if for each j , $X_{jn} \rightarrow X_j$ a.s. (in prob.).

Exercises:

1. Let (X_j) be a sequence of random variables such that for each j

$$E X_j = \mu, E (X_j - \mu)^2 = \sigma^2, \sup_j E |X_j - \mu|^3 < \infty,$$

$$\text{cov}(X_j, X_{j-m}) = 0 \text{ if } m > 1, \text{cov}(X_j, X_{j-m}) \neq 0 \text{ if } m = 1.$$

Prove that $(1/n) \sum_{j=1}^n X_j \rightarrow \mu$ a.s.

Hint: Let $Y_{1j} = X_{2j}$, $Y_{2j} = X_{2j+1}$. Prove first that for $i=1,2$,

$$(1/n) \sum_{j=1}^n Y_{ij} \rightarrow \mu \text{ a.s.}$$

2. Let (Y_j) , $j > 0$, be a sequence of random variables satisfying $E Y_j = 0$, $E Y_j^2 = 1/j^2$ and let $X_n = (Y_n)^n$, $n=1,2,\dots$. Prove that $X_n \rightarrow 0$ a.s. Hint: Combine Chebishev's inequality and the Borel-Cantelli lemma.

3. Let $X_{n,m}$ and X_m be random variables, where $n=1,2,\dots$ and $m=1,2,\dots$ such that for $m=1,2,\dots,k < \infty$, $X_{n,m} \rightarrow X_m$ a.s. as $n \rightarrow \infty$. Prove that

$$\max_{m=1,2,\dots,k} |X_{n,m} - X_m| \rightarrow 0 \text{ a.s.}$$

4. Let (X_j) be a sequence of i.i.d. random variables satisfying $E X_j = 0$, $0 < E X_j^2 < \infty$. Prove that

$$[(1/n) \sum_{j=1}^n X_j] / [(1/n) \sum_{j=1}^n X_j^2] \rightarrow 0 \text{ a.s.}$$

5. Let X be a random variable satisfying

$$P(X = 1) = P(X = .1) = \frac{1}{2}.$$

Let $Y_n = X^n$, $n=1,2,\dots$. Does Y_n converges a.s. or in pr.?

6. Prove theorem 2.1.8.

2.2 Convergence of mathematical expectations

If X and X_1, X_2, \dots are random variables or vectors such that for some $p > 0$, $E|X_n - X|^p \rightarrow 0$ as $n \rightarrow \infty$, then it follows from Chebishev's inequality that $X_n \rightarrow X$ in pr. The converse is not always true. A partial converse is given by the following theorem.

Theorem 2.2.1. If $X_n \rightarrow X$ in pr. and if there is a r.v. Y satisfying $|X_n| \leq Y$ a.s. for $n = 1, 2, \dots$ and $E Y^p < \infty$ for some $p > 0$, then $E|X_n - X|^p \rightarrow 0$.

Proof: If $P(|X| > Y) > 0$ then $X_n \rightarrow X$ in pr. is not possible, hence $|X| \leq Y$ a.s. Since now $|X_n - X| \leq 2Y$ there is no loss of generality in assuming $X = 0$ a.s. We then have:

$$\begin{aligned} \int |x_n(\omega)|^p P(d\omega) &= \int_{\{|x_n(\omega)| > \varepsilon\}} |x_n(\omega)|^p P(d\omega) \\ &\quad + \int_{\{|x_n(\omega)| \leq \varepsilon\}} |x_n(\omega)|^p P(d\omega) \\ &\leq \varepsilon^p + \int_{\{|x_n(\omega)| > \varepsilon\}} y(\omega)^p P(d\omega). \end{aligned}$$

The theorem follows now from theorem 1.4.1.

Q.E.D.

Putting $p = 1$ in theorem 2.2.1 we have:

Theorem 2.2.2. (Dominated convergence theorem) If $X_n \rightarrow X$ in pr. and if $|X_n| \leq Y$ a.s., where $EY < \infty$, then $EX_n \rightarrow EX$.

We shall use this theorem for proving first Fatou's lemma which in its turn will be used for proving the monotone convergence theorem.

Theorem 2.2.3. (Fatou's lemma) If $X_n \geq 0$ a.s., then $E \liminf_{n \rightarrow \infty} X_n \leq \liminf_{n \rightarrow \infty} E X_n$.

Proof: Put $X = \liminf_{n \rightarrow \infty} X_n$ and let $\varphi(x)$ be any simple function satisfying $0 \leq \varphi(x) \leq x$. Put $Y_n = \min(\varphi(X), X_n)$. Then $Y_n \rightarrow \varphi(X)$ in pr. because

$$\begin{aligned} P(|\min(\varphi(X), X_n) - \varphi(X)| \geq \varepsilon) &= P(X_n \leq \varphi(X) - \varepsilon) \\ &\leq P(X_n \leq X - \varepsilon) \rightarrow 0. \end{aligned}$$

Moreover, since $\varphi(x)$ is a simple function we must have $E \varphi(X) < \infty$. From the dominated convergence theorem and from $Y_n \leq X_n$ a.s. it follows now:

$$E \varphi(X) = \lim_{n \rightarrow \infty} E Y_n = \liminf_{n \rightarrow \infty} E Y_n \leq \liminf_{n \rightarrow \infty} E X_n.$$

Taking the supremum over all such simple functions φ it follows now from definition 1.4.2 that the theorem holds. Q.E.D.

Theorem 2.2.4. (Monotone convergence theorem) Let (X_n) be a nondecreasing sequence of r.v.'s. Then

$$E \lim_{n \rightarrow \infty} X_n = \lim_{n \rightarrow \infty} E X_n \leq \infty.$$

Proof: Since our sequence (X_n) is nondecreasing, we have:

$$\lim_{n \rightarrow \infty} X_n = \liminf_{n \rightarrow \infty} X_n, \quad \lim_{n \rightarrow \infty} E X_n = \liminf_{n \rightarrow \infty} E X_n,$$

so that by Fatou's lemma, $E \lim_{n \rightarrow \infty} X_n \leq \liminf_{n \rightarrow \infty} E X_n$. However, for any n we have $X_n \leq \lim_{n \rightarrow \infty} X_n$ a.s. because X_n is nondecreasing, hence $E X_n \leq E \lim_{n \rightarrow \infty} X_n$ and consequently

$$\lim_{n \rightarrow \infty} E X_n \leq E \lim_{n \rightarrow \infty} X_n.$$

This proves the theorem. Q.E.D.

Exercises:

1. Let Y_n be defined in exercise 5 of section 2.1. Does $E Y_n$ converge?
2. Let (f_n) be a sequence of Borel measurable real functions on R^k and let μ be a probability measure on (R^k, B^k) . Suppose there exists a nonnegative Borel measurable real function g on R^k such that for $n=1, 2, \dots$,

$$\int |f_n(x)| \mu(dx) \leq \int g(x) \mu(dx) < \infty.$$

Moreover, assume that $f(x) = \lim_{n \rightarrow \infty} f_n(x)$ exists for each x in a set $S \subset R^k$ with $\mu(S) = 1$. Prove that

$$\lim_{n \rightarrow \infty} \int f_n(x) \mu(dx) = \int f(x) \mu(dx).$$

(This is another version of the dominated convergence theorem.)

2.3 Convergence of distributions

If X and X_1, X_2, \dots are r.v.'s with distribution functions F, F_1, F_2, \dots , respectively, then one would like to say that X_n converges in distribution to X if for every $t \in \mathbb{R}$, $F_n(t) \rightarrow F(t)$. However, if X and F are given and if we define $X_n = X + 1/n$, then:

$$F_n(t) = P(X_n \leq t) = P(X \leq t - 1/n) = F(t - 1/n),$$

so that for every discontinuity point t_0 of F we have:

$$\lim_{n \rightarrow \infty} F_n(t_0) = \lim_{n \rightarrow \infty} F(t_0 - 1/n) = F(t_0^-) < F(t_0),$$

while intuitively we would expect that in this case we also have convergence in distribution. Furthermore, if $X_n = X + n$ we have:

$$F_n(t) = P(X_n \leq t) = F(t-n) \rightarrow F(-\infty) = 0 \text{ for every } t.$$

Thus not every sequence of distribution functions converges to another distribution function. In the latter case we say that the convergence is improper.

Definition 2.3.1. A sequence $(F_n(t))$ of distribution functions converges properly pointwise if $F_n(t) \rightarrow F(t)$ pointwise for all continuity points of F , where F is a distribution function. We then write: $F_n \rightarrow F$ properly, pointwise.

The exclusion of discontinuity points avoids the complication that otherwise the function $F(t) = \lim_{n \rightarrow \infty} F_n(t)$ may not be right continuous. In view of the above example we now define:

Definition 2.3.2. A sequence (X_n) of random variables (or random vectors) converges in distribution to X , if their underlying distribution functions (F_n) , F , respectively, satisfy $F_n \rightarrow F$ properly pointwise. We then write: $X_n \rightarrow X$ in distr..

Remark: If this 'limit' distribution F is the distribution function of (for example) the normal distribution $N(\mu, \sigma^2)$, we shall also write: $X \rightarrow N(\mu, \sigma^2)$ in distr..

There is a close connection between proper pointwise convergence of distribution functions and convergence of mathematical expectations, as is shown by the following theorem. This theorem is very fundamental as it allows for a variety of applications.

Theorem 2.3.1. Let F and F_n , $n = 1, 2, \dots$ be distribution functions on R^k . Then $F_n \rightarrow F$ properly pointwise if and only if for every bounded continuous real function φ on R^k ,

$$\int \varphi(t) dF_n(t) \rightarrow \int \varphi(t) dF(t).$$

Proof: Since the proof for the general case $k > 1$ is a straightforward extension of that for $k = 1$, we assume $k = 1$. Suppose $F_n \rightarrow F$ properly pointwise. For given $\epsilon > 0$ we can always find continuity points a and b of F such that $F(b) - F(a) > 1 - \epsilon$. Let φ be any bounded continuous real function on R with uniform bound 1 (which is no restriction). By the uniform continuity of φ on $[a, b]$ we can find continuity points t_2, t_3, \dots, t_{m-1} of F satisfying $a = t_1 < t_2 < \dots < t_{m-1} < t_m = b$ and

$$\sup_{t \in (t_i, t_{i+1}]} \varphi(t) - \inf_{t \in (t_i, t_{i+1}]} \varphi(t) \leq \epsilon$$

for $i = 1, 2, \dots, m-1$. Now define:

$$\psi(t) = \inf_{t \in (t_i, t_{i+1}]} \varphi(t) \text{ for } t \in (t_i, t_{i+1}], i = 1, 2, \dots, m-1,$$

$$\psi(t) = 0 \text{ elsewhere.}$$

Then

$$0 \leq \varphi(t) - \psi(t) \leq \epsilon \text{ for } t \in (a, b],$$

$$0 \leq \varphi(t) - \psi(t) \leq 1 \text{ for } t \notin (a, b],$$

hence

$$\begin{aligned} & \left| \int \psi(t) dF_n(t) - \int \varphi(t) dF_n(t) \right| \\ & \leq \int_{\{t \in (a, b]\}} \epsilon dF_n(t) + \int_{\{t \notin (a, b]\}} dF_n(t) \\ & = \epsilon(F_n(b) - F_n(a)) + 1 - F_n(b) + F_n(a) \\ & \rightarrow \epsilon(F(b) - F(a)) + 1 - F(b) + F(a) \leq 2\epsilon. \end{aligned}$$

Moreover,

$$\begin{aligned} & \int \psi(t) dF_n(t) = \\ & \sum_{i=1}^{m-1} (\inf_{t \in (t_i, t_{i+1}]} \varphi(t)) (F_n(t_{i+1}) - F_n(t_i)) \\ & \rightarrow \sum_{i=1}^{m-1} (\inf_{t \in (t_i, t_{i+1}]} \varphi(t)) (F(t_{i+1}) - F(t_i)) = \int \psi(t) dF(t) \end{aligned}$$

and

$$\left| \int \varphi(t) dF(t) - \int \psi(t) dF(t) \right| \leq 2\epsilon.$$

So we have:

$$\begin{aligned} & \left| \int \varphi(t) dF_n(t) - \int \varphi(t) dF(t) \right| \\ & \leq 4\epsilon + \left| \int \psi(t) dF_n(t) - \int \psi(t) dF(t) \right| \leq 5\epsilon \end{aligned}$$

for sufficiently large n , which proves the 'only if' part of the theorem.

Now let u be a continuity point of F and define

$$\varphi(t) = 1 \text{ if } t \leq u, \quad \varphi(t) = 0 \text{ if } t > u,$$

$$\varphi_{1,m}(t) = 1 \text{ if } t \leq u - 1/m,$$

$$\varphi_{1,m}(t) = -m \cdot t + m \cdot u \text{ if } t \in (u - 1/m, u],$$

$$\varphi_{1,m}(t) = 0 \text{ if } t > u,$$

$$\varphi_{2,m}(t) = 1 \text{ if } t \leq u,$$

$$\varphi_{2,m}(t) = -m \cdot t + 1 + m \cdot u \text{ if } t \in (u, u + 1/m],$$

$$\varphi_{2,m}(t) = 0 \text{ if } t > u + 1/m.$$

Then $\varphi_{1,m}$ and $\varphi_{2,m}$ are bounded continuous functions on \mathbb{R} satisfying $\varphi_{1,m}(t) \leq \varphi(t) \leq \varphi_{2,m}(t)$, hence for $m = 1, 2, \dots$ and $n \rightarrow \infty$:

$$\int \varphi_{1,m}(t) dF_n(t) \leq F_n(u) = \int \varphi(t) dF_n(t) \leq \int \varphi_{2,m}(t) dF_n(t),$$

$$\int \varphi_{1,m}(t) dF(t) \leq F(u) = \int \varphi(t) dF(t) \leq \int \varphi_{2,m}(t) dF(t).$$

Moreover:

$$\begin{aligned} 0 &\leq \int (\varphi_{2,m}(t) - \varphi_{1,m}(t)) dF(t) \leq \int_{\{t \in (u-1/m, u+1/m)\}} dF(t) \\ &= F(u+1/m) - F(u-1/m). \end{aligned}$$

Since u is a continuity point of F , $F(u + 1/m) - F(u - 1/m)$ can be made arbitrarily small by increasing m , hence $F_n(u) \rightarrow F(u)$, which proves the 'if' part. Q.E.D.

A direct consequence of this theorem is that:

Theorem 2.3.2. $X_n \rightarrow X$ in pr. implies $X_n \rightarrow X$ in distr.,

Proof: By theorem 2.1 7 it follows that for any continuous function φ we have $X_n \rightarrow X$ in pr. implies $\varphi(X_n) \rightarrow \varphi(X)$ in pr., whereas by theorem 2.2 2, $\varphi(X_n) \rightarrow \varphi(X)$ in pr. implies $E\varphi(X_n) \rightarrow E\varphi(X)$ if φ is a bounded continuous function. Q.E.D.

The converse of this theorem is not generally true, but it is if X is constant a.s., that is: $P(X = c) = 1$ for some constant c . In that case the proper limit F involved is:

$$F(t) = 1 \text{ if } t \geq c, \quad F(t) = 0 \text{ if } t < c.$$

The proof of this proposition is very simple:

$$P(|X_n - c| \leq \varepsilon) = P(c - \varepsilon \leq X_n \leq c + \varepsilon) = F_n(c + \varepsilon) - F_n((c - \varepsilon)-) \\ + F(c + \varepsilon) - F(c - \varepsilon) = 1$$

for every $\varepsilon > 0$, since $c + \varepsilon$ and $c - \varepsilon$ are continuity points of F . Thus:

Theorem 2.3.3. Convergence in distribution to a constant implies convergence in probability to that constant.

Let X_n and X be random vectors in \mathbb{R}^k such that $X_n \rightarrow X$ in distr. and let f be any continuous real function on \mathbb{R}^k . For any bounded continuous real function φ on \mathbb{R} it follows that $\varphi(f)$ is a bounded continuous real function on \mathbb{R}^k , so that by theorem 2.3.1,

$$E\varphi(f(X_n)) \rightarrow E\varphi(f(X))$$

and consequently $f(X_n) \rightarrow f(X)$ in distr. Thus we have:

Theorem 2.3.4. Let X_n and X be random vectors in \mathbb{R}^k and let f be a continuous real function on \mathbb{R}^k . Then $X_n \rightarrow X$ in distr. implies $f(X_n) \rightarrow f(X)$ in distr.

Remark: The continuity of f in theorem 2.3.4 is crucially a condition. For example, let X be a continuously distributed random variable, let X_n the value of X rounded off to n decimal digits and let

$$f(x) = 1 \text{ if } x \text{ is rational, } f(x) = 0 \text{ if } x \text{ is irrational.}$$

Then $X_n \rightarrow X$ in distr. and f is Borel measurable. However, $f(X_n) = 1$ a.s., $f(X) = 0$ a.s., which renders $f(X_n) \rightarrow f(X)$ in distr. impossible.

A more general result is given by the following theorem.

Theorem 2.3.5. Let X_n and X be random vectors in R^k , Y_n a random vector in R^m and c a nonrandom vector in R^m . If $X_n \rightarrow X$ in distr. and $Y_n \rightarrow c$ in distr., then $f(X_n, Y_n) \rightarrow f(X, c)$ in distr. for any continuous real function f on $R^k \times C$, where C is some subset of R^m with interior point c .

Proof: Again we prove the theorem for the case $k = m = 1$ since the proof of the general case is similar. It suffices to prove that for any bounded continuous real function φ on R^2 we have $E\varphi(X_n, Y_n) \rightarrow E\varphi(X, c)$, because then $E\psi(f(X_n, Y_n)) \rightarrow E\psi(f(X, c))$ for any bounded continuous real function ψ on R , which by theorem 2.3.1 implies $f(X_n, Y_n) \rightarrow f(X, c)$ in distr.

Let M be the uniform bound of φ and let F_n and F be the distribution functions of X_n and X , respectively. For every ϵ we can choose continuity points a and b of F such that:

$$P(X \in (a, b]) = F(b) - F(a) > 1 - \epsilon/(2M).$$

Moreover, for any $\delta > 0$ we have:

$$\begin{aligned} |E\varphi(X_n, Y_n) - E\varphi(X_n, c)| &\leq \int_{\{x_n \in (a, b]\}} |\varphi(x_n, y_n) - \varphi(x_n, c)| dP \\ &\quad + \int_{\{x_n \notin (a, b]\}} |\varphi(x_n, y_n) - \varphi(x_n, c)| dP \\ &\leq \int_{\{x_n \in (a, b]\} \cap \{|y_n - c| \leq \delta\}} |\varphi(x_n, y_n) - \varphi(x_n, c)| dP \\ &\quad + 2M \cdot P(\{x_n \in (a, b]\} \cap \{|y_n - c| > \delta\}) + 2M \cdot P(X_n \notin (a, b]) \\ &\leq \int_{\{x_n \in (a, b]\} \cap \{|y_n - c| \leq \delta\}} |\varphi(x_n, y_n) - \varphi(x_n, c)| dP \\ &\quad + 2M \cdot P\{|y_n - c| > \delta\} + 2M(1 - F_n(b) + F_n(a)). \end{aligned}$$

Since by theorem 2.3.3, $Y_n \rightarrow c$ in pr., we have

$$P\{|Y_n - c| > \delta\} \rightarrow 0 \text{ for any } \delta > 0,$$

whereas

$$\lim_{n \rightarrow \infty} 2M(1 - F_n(b) + F_n(a)) = 2M(1 - F(b) + F(a)) < \epsilon.$$

Furthermore, since $\varphi(t_1, t_2)$ is uniformly continuous on the bounded set

$$\{(t_1, t_2) \in \mathbb{R}^2: a < t_1 \leq b, |t_2 - c| \leq \delta\},$$

provided that δ is so small that this set is contained in $\mathbb{R}^k \times C$, we can choose δ such that the last integral is smaller than ε . So we conclude:

$$E\varphi(X_n, Y_n) - E\varphi(X_n, c) \rightarrow 0.$$

Since obviously $E\varphi(X_n, c) \rightarrow E\varphi(X, c)$ because $X_n \rightarrow X$ in distr., the theorem follows. Q.E.D.

Remark: It should be stressed that the constancy of c is crucial in theorem 2.3.5. Thus, e.g., $X_n \rightarrow X$ in distr. and $Y_n \rightarrow Y$ in distr., where X and Y are nonconstant random variables, does not generally imply $X_n + Y_n \rightarrow X + Y$ in distr. or $X_n Y_n \rightarrow X Y$ in distr. Moreover, theorem 2.1.8 does not carry over to convergence in distribution.

Finally we note that convergence in distribution is closely related to convergence of characteristic functions:

Theorem 2.3.6. Let (F_n) be a sequence of distribution functions on \mathbb{R}^k and let (φ_n) be the sequence of corresponding characteristic functions. If $F_n \rightarrow F$ properly pointwise, then $\varphi_n(t) \rightarrow \int \exp(i \cdot t'x) dF(x)$ pointwise on \mathbb{R}^k . If $\varphi_n(t) \rightarrow \varphi(t)$ pointwise on \mathbb{R}^k and $\varphi(t)$ is continuous at $t = 0$ then there exists a unique distribution function F such that $\varphi(t) = \int \exp(i \cdot t'x) dF(x)$ and $F_n \rightarrow F$ properly pointwise.

Proof: Cf. Feller (1966).

This theorem is basic for proving central limit theorems. Moreover, the following corollary of theorem 2.3.6 is very useful in proving *multivariate* asymptotic normality results.

Theorem 2.3.7. Let (X_n) be a sequence of random vectors in \mathbb{R}^k . If for all vectors $\xi \in \mathbb{R}^k$, $\xi'X_n$ converges in distribution to a normal distribution $N(\mu'\xi, \xi'\Lambda\xi)$, where Λ is a positive (semi) definite matrix, then X_n converges in distribution to the k -variate normal distribution $N_k(\mu, \Lambda)$.

Proof: We recall that the characteristic function of the k-variate normal distribution $N_k(\mu, \Lambda)$ is given by

$$\exp(i \cdot \mu' t - \frac{1}{2} t' \Lambda t)$$

(see, e.g., Anderson (1958)). Now consider a random vector $X = (X_1, \dots, X_k)'$ in R^k with mean vector

$$E X = (E X_1, \dots, E X_k)' = (\mu_1, \dots, \mu_k)' = \mu$$

and variance matrix

$$E[(X - \mu)(X - \mu)'] = (E[(X_1 - \mu_1)(X_j - \mu_j)]) = \Lambda.$$

(Recall that if Λ is a singular matrix then the normal distribution involved is said to be singular). Suppose that for every vector ξ in R^k , $\xi'X$ is normally distributed. Then $\xi'X \sim N(\xi'\mu, \xi'\Lambda\xi)$, hence for every $t \in R$ and every $\xi \in R^k$,

$$E \exp(i \cdot t \xi' X) = \exp[i \cdot t \mu' \xi - \frac{1}{2} t^2 \xi' \Lambda \xi].$$

Substituting $t = 1$ now yields

$$E \exp(i \cdot \xi' X) = \exp[i \cdot \mu' \xi - \frac{1}{2} \xi' \Lambda \xi] \quad \text{for all } \xi \in R^k,$$

which is just the characteristic function of the k-variate normal distribution $N_k[\mu, \Lambda]$. Q.E.D.

Exercises:

1. Let Y_n be defined in exercise 5 of section 2.1. Does Y_n converge in distribution?
2. Let (X_n) be a sequence of independent $N_k(\mu, \Lambda)$ distributed random vectors in R^k , where Λ is nonsingular. Let

$$\hat{\Lambda}_n = (1/n) \sum_{j=1}^n (X_j - \mu)(X_j - \mu)'$$

Prove that

$$(X_n - \mu)' \hat{\Lambda}_n^{-1} (X_n - \mu) \rightarrow \chi_k^2 \quad \text{in distr.}$$

Hint: use the fact that the elements of an inverse matrix are

continuous functions of the elements of the inverted matrix, provided the latter is nonsingular.

3. The χ_n^2 distribution has characteristic function

$$\varphi_n(t) = (1 - 2it)^{-n/2}.$$

Let $X_n = Y_n/n$, where Y_n is distributed χ_n^2 . Prove $X_n \rightarrow 1$ in pr., using theorem 2.3.6 and the fact that $(1 + z/n)^n \rightarrow e^z$ for real or complex valued z .

2.4 Central limit theorems

In this section we consider a number of central limit theorems (CLT). These CLT's are well known, but stated here for convenience. For the proofs we refer to textbooks like Feller (1966) or Chung (1974).

Theorem 2.4.1. Let X_1, X_2, X_3, \dots be i.i.d. random variables with $E X_j = \mu$, $\text{var}(X_j) = \sigma^2 < \infty$. Then

$$(1/\sqrt{n})\sum_{j=1}^n (X_j - \mu) \rightarrow N(0, \sigma^2) \text{ in distr.}$$

Proof: E.g., Chung (1974, theorem 6.4.4).

The next central limit theorem is due to *Liapounov*.

Theorem 2.4.2. Let

$$S_n = \sum_{j=1}^{k_n} X_{n,j},$$

where for each n the r.v.'s $X_{n,1}, \dots, X_{n,k_n}$ are independent and $k_n \rightarrow \infty$. Put

$$E X_{n,j} = \alpha_{n,j}, \quad \alpha_n = \sum_{j=1}^{k_n} \alpha_{n,j}$$

$$\sigma^2(X_{n,j}) = E(X_{n,j} - \alpha_{n,j})^2 = \sigma_{n,j}^2, \quad \sigma_n^2 = \sum_{j=1}^{k_n} \sigma_{n,j}^2,$$

assuming $\sigma_n^2 < \infty$ (but not necessarily $\limsup_{n \rightarrow \infty} \sigma_n^2 < \infty$). If for some $\delta > 0$,

$$\lim_{n \rightarrow \infty} \sum_{j=1}^{k_n} E |(X_{n,j} - \alpha_{n,j})/\sigma_n|^{2+\delta} = 0$$

then $(S_n - \alpha_n)/\sigma_n \rightarrow N(0,1)$ in distr.

Proof: Chung (1974, p.209).

This theorem is less general than the Lindeberg-Feller central limit theorem, but its conditions are easier to verify.

A special case of theorem 2.4.2 is:

Theorem 2.4.3. For each $n \geq 1$ let $X_{n,j}$, $j=1, \dots, n$, be independent random variables with $E X_{n,j} = 0$. If

$$\lim_{n \rightarrow \infty} (1/n) \sum_{j=1}^n E X_{n,j}^2 \rightarrow \sigma^2$$

and

$$\lim_{n \rightarrow \infty} \sum_{j=1}^n E |X_{n,j}/\sqrt{n}|^{2+\delta} = 0 \text{ for some } \delta > 0 \quad (2.4.1)$$

then $(1/\sqrt{n}) \sum_{j=1}^n X_{n,j} \rightarrow N(0, \sigma^2)$ in distr.

Remark: Note that condition (2.4.1) holds if

$$\sup_n (1/n) \sum_{j=1}^n |X_{n,j}|^{2+\delta} < \infty$$

Exercises:

1. Let (X_j) be a sequence of i.i.d. random vectors in R^k with

$$E X_j = \mu, E (X_j - \mu)(X_j - \mu)' = \Lambda,$$

where Λ is nonsingular. Let

$$\bar{X} = (1/n) \sum_{j=1}^n X_j,$$

$$\hat{\Lambda} = (1/n) \sum_{j=1}^n (X_j - \bar{X})(X_j - \bar{X})',$$

$$Y_n = n(\bar{X} - \mu)' \hat{\Lambda}^{-1} (\bar{X} - \mu).$$

Prove that $Y_n \rightarrow \chi_k^2$ in distr. (Cf. exercise 2 of section 2.3).

2. Let (U_j) be a sequence of i.i.d. random variables satisfying $E U_j = 0$, $E U_j^2 = 1$. Let $X_j = U_j + U_{j-1}$. Prove that

$$(1/\sqrt{n})\sum_{j=1}^n X_j \rightarrow N(0,4) \text{ in distr.}$$

2.5 *Further results on convergence of distributions and mathematical expectations, and laws of large numbers.*

The condition in theorem 2.3.1 that the function φ is bounded is only necessary for the 'if' part. If we assume proper pointwise convergence then convergence of expectations will occur under more general conditions, as is shown in the following extension of the 'only if' part of theorem 2.3.1.

Theorem 2.5.1 Let (F_n) be a sequence of distribution functions on R^k satisfying $F_n \rightarrow F$ properly pointwise. Let $\varphi(x)$ be a continuous real function on R^k such that

$$\sup_n \int |\varphi(x)|^{1+\delta} dF_n(x) < \infty \quad \text{for some } \delta > 0.$$

Then $\int \varphi(x) dF_n(x) \rightarrow \int \varphi(x) dF(x)$.

Proof: Define for $a > 0$

$$\begin{aligned} \varphi_a(x) &= \varphi(x) \text{ if } |\varphi(x)| \leq a, \quad \varphi_a(x) = a \text{ if } \varphi(x) > a, \\ \varphi_a(x) &= -a \text{ if } \varphi(x) < -a. \end{aligned} \tag{2.5.1}$$

Obviously $\varphi_a(x)$ is a bounded continuous real function on R^k , hence by theorem 2.3.1:

$$\int \varphi_a(x) dF_n(x) \rightarrow \int \varphi_a(x) dF(x) \tag{2.5.2}$$

Moreover,

$$\begin{aligned} \left| \int \varphi(x) dF_n(x) - \int \varphi_a(x) dF_n(x) \right| &\leq 2 \int_{|\varphi(x)| > a} |\varphi(x)| dF_n(x) \\ &\leq 2 \cdot a^{-\delta} \int |\varphi(x)|^{1+\delta} dF_n(x) = O(a^{-\delta}) \end{aligned} \tag{2.5.3}$$

uniformly in n , and similarly we have:

$$\begin{aligned}
|\int \varphi(x) dF(x) - \int \varphi_a(x) dF(x)| &\leq 2a^{-\delta} \int |\varphi(x)|^{1+\delta} dF(x) \\
&= o(a^{-\delta}) \quad (2.5.4)
\end{aligned}$$

provided

$$\int |\varphi(x)|^{1+\delta} dF(x) < \infty \quad (2.5.5)$$

The theorem follows easily from (2.5.2) through (2.5.4) by letting first $n \rightarrow \infty$ and then $a \rightarrow \infty$. Thus it suffices to show that (2.5.5) is true. Now observe that $|\varphi_a(x)|^{1+\delta}$ is monotonic nondecreasing in a and that $|\varphi_a(x)|^{1+\delta} \rightarrow |\varphi(x)|^{1+\delta}$ as $a \rightarrow \infty$. It follows therefore from the monotone convergence theorem (theorem 2.2.4) and theorem 2.3.1 that:

$$\begin{aligned}
\int |\varphi(x)|^{1+\delta} dF(x) &= \lim_{a \rightarrow \infty} \int |\varphi_a(x)|^{1+\delta} dF(x) \\
&= \lim_{a \rightarrow \infty} \lim_{n \rightarrow \infty} \int |\varphi_a(x)|^{1+\delta} dF_n(x) \\
&\leq \lim_{n \rightarrow \infty} \int |\varphi(x)|^{1+\delta} dF_n(x) \leq \sup_n \int |\varphi(x)|^{1+\delta} dF_n(x) < \infty
\end{aligned} \quad (2.5.6)$$

This result completes the proof of theorem 2.5.1. Q.E.D.

Along similar lines we can prove the following version of the weak law of large numbers.

Theorem 2.5.2. Let X_1, X_2, \dots be a sequence of independent random vectors in R^k , and let $(F_j(x))$ be the sequence of corresponding distribution functions. Let $\varphi(x)$ be a continuous function on R^k . If

$$(1/n) \sum_{j=1}^n F_j \rightarrow G \text{ properly, pointwise}$$

and

$$\sup_n (1/n) \sum_{j=1}^n E |\varphi(X_j)|^{1+\delta} < \infty \text{ for some } \delta > 0$$

then $\text{plim}_{n \rightarrow \infty} (1/n) \sum_{j=1}^n \varphi(X_j) = \int \varphi(x) dG(x)$.

Proof: Consider the function $\varphi_a(x)$ defined in (2.5.1). Then obviously by the independence of the X_j 's, the boundedness of

$\varphi_a(x)$ and Chebishev's inequality

$$\text{plim}_{n \rightarrow \infty} (1/n) \sum_{j=1}^n (\varphi_a(X_j) - E \varphi_a(X_j)) = 0, \quad (2.5.7)$$

while from theorem 2.3.1 it follows

$$\lim_{n \rightarrow \infty} E(1/n) \sum_{j=1}^n \varphi_a(X_j) = \int \varphi_a(x) dG(x). \quad (2.5.8)$$

Hence

$$\text{plim}_{n \rightarrow \infty} (1/n) \sum_{j=1}^n \varphi_a(X_j) = \int \varphi_a(x) dG(x). \quad (2.5.9)$$

Moreover, since $\varphi_a(x)$ is bounded, it follows from (2.5.9) and theorem 2.2.1. that

$$E \left| (1/n) \sum_{j=1}^n \varphi_a(X_j) - \int \varphi_a(x) dG(x) \right| \rightarrow 0 \text{ as } n \rightarrow \infty. \quad (2.5.10)$$

Furthermore, similarly to (2.5.3) it follows that

$$\limsup_{n \rightarrow \infty} E \left| (1/n) \sum_{j=1}^n \varphi(X_j) - (1/n) \sum_{j=1}^n \varphi_a(X_j) \right| \rightarrow 0 \text{ as } a \rightarrow \infty \quad (2.5.11)$$

and similarly to (2.5.4) that

$$\left| \int \varphi(x) dG(x) - \int \varphi_a(x) dG(x) \right| \rightarrow 0 \text{ as } a \rightarrow \infty. \quad (2.5.12)$$

Combining (2.5.10), (2.5.11) and (2.5.12) we see that

$$\lim_{n \rightarrow \infty} E \left| (1/n) \sum_{j=1}^n \varphi(X_j) - \int \varphi(x) dG(x) \right| \rightarrow 0. \quad (2.5.13)$$

The theorem follows now from (2.5.13) and Chebishev's inequality. Q.E.D.

Remark: The difference of this theorem with the classical weak law of large numbers is that the finiteness of second moments is not necessary.

If we combine the theorems 2.1.4 and 2.5.1. we easily obtain the following strong version of theorem 2.5.2.

Theorem 2.5.3. Let the conditions of theorem 2.5.2 be satisfied, and assume in addition that

$$\sup_n (1/n) \sum_{j=1}^n E |\varphi(X_j)|^{2+\delta} < \infty \quad \text{for some } \delta > 0. \quad (2.5.14)$$

Then $(1/n) \sum_{j=1}^n \varphi(X_j) \rightarrow \int \varphi(x) dG(x)$ a.s.

The continuity condition on the function φ in theorems 2.5.2 and 2.5.3 is mainly due to theorem 2.5.1, i.e., without this condition theorem 2.5.1 is not generally true. Suppose for example

$$\varphi(x) = x^2 \text{ if } x \text{ is rational, } \varphi(x) = -x^2 \text{ if } x \text{ is irrational.}$$

Then $\varphi(x)$ is a Borel measurable real function on \mathbb{R} . The proof of this proposition is left as an exercise. Now let X be a random drawing from an absolutely continuous distribution, say the standard normal distribution, and let X_n be the value of X rounded off to n decimal digits. Then $X_n \rightarrow X$ in distr., hence the distribution function F_n of X_n converges properly to the distribution function F of X . However, since X is a.s. irrational we have:

$$E \varphi(X) = \int -x^2 dF(x) = -1$$

whereas X_n is a.s. rational and thus

$$E \varphi(X_n) = \int x^2 dF_n(x) \rightarrow \int x^2 dF(x) = 1.$$

This counter-example shows that theorem 2.5.1 does not carry over for general Borel measurable functions φ . In order that a similar result as in theorem 2.5.1 does hold for Borel measurable functions we need a stronger convergence in distribution concept, namely setwise proper convergence:

Definition 2.5.1. Let (F_n) be a sequence of distribution functions on \mathbb{R}^k with corresponding sequence (μ_n) of probability measures on $(\mathbb{R}^k, \mathcal{B}^k)$ (cf. section 1.1). This sequence (F_n) converges properly setwise if for each Borel set B in \mathcal{B}^k , $\lim_{n \rightarrow \infty} \mu_n(B) = \mu(B)$, where μ is a probability measure on $(\mathbb{R}^k, \mathcal{B}^k)$. We then write $F_n \rightarrow F$ properly setwise, where F is the distribution function induced by μ .

Using this concept and definition 1.4.4 we can now state:

Theorem 2.5.4. Let F_n and F be distribution functions on \mathbb{R}^k . Then $F_n \rightarrow F$ properly setwise if and only if for every bounded Borel measurable real function φ on \mathbb{R}^k ,

$$\lim_{n \rightarrow \infty} \int \varphi(x) dF_n(x) = \int \varphi(x) dF(x).$$

Proof: The 'if' part is easy and therefore left to the reader. Moreover, the 'only if' part follows easily from definition 1.3.2 if φ is a simple function. Thus, assume that φ is not simple. From the proof of theorem 1.3.4 it easily follows that for arbitrary $\varepsilon > 0$ and each bounded Borel set B we can construct a simple function ψ such that

$$|\psi(x) - \varphi(x)| < \varepsilon \quad \text{if } x \in B, \quad \psi(x) = 0 \quad \text{for } x \notin B.$$

Moreover, we may choose B such that

$$\mu(B) = \int_B dF(x) > 1 - \varepsilon.$$

Furthermore, since

$$\mu_n(B) = \int_B dF_n(x) \rightarrow \mu(B)$$

there exists an n_ε such that

$$\mu_n(B) > 1 - 2\varepsilon \quad \text{if } n \geq n_\varepsilon.$$

Thus we have for $n \geq n_\varepsilon$

$$\begin{aligned} \left| \int \varphi(x) dF_n(x) - \int \varphi(x) dF(x) \right| &\leq \left| \int_B \varphi(x) dF_n(x) - \int_B \varphi(x) dF(x) \right| \\ &\quad + \left| \int_{\mathbb{R}^k \setminus B} \varphi(x) dF_n(x) - \int_{\mathbb{R}^k \setminus B} \varphi(x) dF(x) \right| \\ &\leq \left| \int_B (\varphi(x) - \psi(x)) dF_n(x) \right| + \left| \int_B (\varphi(x) - \psi(x)) dF(x) \right| \\ &\quad + \left| \int_B \psi(x) dF_n(x) - \int_B \psi(x) dF(x) \right| + M\mu_n(\mathbb{R}^k \setminus B) + M\mu(\mathbb{R}^k \setminus B) \end{aligned}$$

$$\begin{aligned} &\leq \varepsilon \mu_n(B) + \varepsilon \mu(B) + 3\varepsilon M + \left| \int_B \psi(x) dF_n(x) - \int_B \psi(x) dF(x) \right| \\ &\leq (2+3M)\varepsilon + \left| \int \psi(x) dF_n(x) - \int \psi(x) dF(x) \right|, \end{aligned}$$

where M is the bound of $\varphi(x)$. Since the second term converges to zero (for ψ is simple) and the first term can be made arbitrarily small, the theorem follows. Q.E.D.

Now observe from the proof of theorem 2.5.1 that the continuity of φ and φ_a is only necessary for (2.5.2). However, if $F_n \rightarrow F$ properly setwise then (2.5.2) carries over for Borel measurable φ_a , as we just have shown in theorem 2.5.4. Consequently we have:

Theorem 2.5.5. Let (F_n) be a sequence of distribution functions on R^k satisfying $F_n \rightarrow F$ properly setwise. Let $\varphi(x)$ be a Borel measurable real function such that

$$\sup_n \int |\varphi(x)|^{1+\delta} dF_n(x) < \infty \quad \text{for some } \delta > 0.$$

Then $\int \varphi(x) dF_n(x) \rightarrow \int \varphi(x) dF(x)$.

Replacing theorem 2.5.1 by theorem 2.5.5 the laws of large numbers (theorems 2.5.2 and 2.5.3) can now be generalized to:

Theorem 2.5.6. Let (X_j) be a sequence of independent random vectors in R^k , and let $(F_j(x))$ be the sequence of corresponding distribution functions. Let $\varphi(x)$ be a Borel measurable real function on R^k . If

$$(1/n) \sum_{j=1}^n F_j \rightarrow G \text{ properly setwise}$$

and

$$\sup_n (1/n) \sum_{j=1}^n E |\varphi(X_j)|^{1+\delta} < \infty \quad \text{for some } \delta < 0$$

then $\text{plim}_{n \rightarrow \infty} (1/n) \sum_{j=1}^n \varphi(X_j) = \int \varphi(x) dG(x)$.

Theorem 2.5.7. Let the conditions of theorem 2.5.6 be satisfied. If

$$\sup_n (1/n) \sum_{j=1}^n E |\varphi(X_j)|^{2+\delta} < \infty \text{ for some } \delta > 0,$$

then $(1/n) \sum_{j=1}^n \varphi(X_j) \rightarrow \int \varphi(x) dG(x)$ a.s.

Finally we consider convergence of random variables of the type

$$(1/n) \sum_{j=1}^n \varphi_j(X_j),$$

where the φ_j 's are Borel measurable (respectively continuous) functions.

Theorem 2.5.8. Let X_j be a sequence of independent random vectors in R^k and let (φ_j) be a sequence of Borel measurable (continuous) real functions on R^k . Denote $Y_j = \varphi_j(X_j)$ and let F_j be the distribution function of Y_j . If

$$(1/n) \sum_{j=1}^n F_j \rightarrow G \text{ properly setwise (pointwise)}$$

and

$$\sup_n (1/n) \sum_{j=1}^n E |\varphi_j(X_j)|^{1+\delta} < \infty \text{ for some } \delta > 1 [\delta > 0]$$

then $(1/n) \sum_{j=1}^n \varphi_j(X_j) \rightarrow \int y dG(y)$ a.s. [in pr.].

We shall not use this theorem in the sequel, but it is stated because it covers theorems 2.5.2, 2.5.3, 2.5.6 and 2.5.7. Moreover, its proof is an easy but useful exercise.

Exercises:

1. Prove theorem 2.5.8.
2. Restate theorems 2.5.2 and 2.5.6 for double arrays $(X_{n,j})$, $j=1, \dots, n$, $n=1, 2, \dots$ of random vectors in R^k and prove the modified theorems involved.

2.6 Convergence of random functions

Dealing with convergence of random functions one should be aware of some pitfalls. The first one concerns pointwise a.s. convergence. Let $f(\theta)$ and $f_n(\theta)$ be random functions on a subset Θ on \mathbb{R}^k such that for each $\theta \in \Theta$, $f_n(\theta) \rightarrow f(\theta)$ a.s. as $n \rightarrow \infty$. At first sight we would expect from definition 2.1.2 that there is a null set N and an integer function $n_0(\omega, \theta, \varepsilon)$ such that for every $\varepsilon > 0$ and every $\omega \in \Omega \setminus N$,

$$|f_n(\theta, \omega) - f(\theta, \omega)| \leq \varepsilon \text{ if } n \geq n_0(\omega, \theta, \varepsilon).$$

However, reading definition 2.1.2 carefully we see that this is not correct, because *the null set N may depend on θ* : $N = N_\theta$. Then again at first sight we might reply that this does not matter because we could choose $N = \bigcup_{\theta \in \Theta} N_\theta$ as a null set. But the problem now is that we are not sure whether $N \in \mathcal{F}$, for only countable unions of members of \mathcal{F} are surely members of \mathcal{F} themselves. Thus although $N_\theta \in \mathcal{F}$ for each $\theta \in \Theta$, this is not necessarily the case for $\bigcup_{\theta \in \Theta} N_\theta$ if Θ is uncountable. Moreover, even if $\bigcup_{\theta \in \Theta} N_\theta \in \mathcal{F}$, it may fail to be a null set itself if Θ is uncountable. For example, let $\Theta = \Omega = [0, 1]$, let P be the Lebesgue measure on $[0, 1]$ and let $N_\theta = \{\theta\}$ for $\theta \in [0, 1]$. Then $P(\bigcup_{\theta \in \Theta} N_\theta) = P(\Omega) = 1$, while obviously the N_θ 's are null sets.

The second pitfall concerns uniform convergence of random functions. As is well known, uniform convergence of (real) nonrandom functions, for example $\varphi_n(\theta) \rightarrow \varphi(\theta)$ uniformly on Θ as $n \rightarrow \infty$, can be defined by

$$\sup_{\theta \in \Theta} |\varphi_n(\theta) - \varphi(\theta)| \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Dealing with uniform a.s. convergence of random functions, i.e.,

$$f_n(\theta) \rightarrow f(\theta) \text{ a.s. uniformly on } \Theta,$$

a suitable definition is therefore:

$$\sup_{\theta \in \Theta} |f_n(\theta) - f(\theta)| \rightarrow 0 \text{ a.s. as } n \rightarrow \infty,$$

However, this has only a probabilistic meaning if the supremum involved is a random variable. If so, then uniform a.s. convergence is equivalent with the following:

There is a null set N and an integer function $n_0(\omega, \epsilon)$ both independent of θ , such that for every $\epsilon > 0$, every $\omega \in \Omega \setminus N$ and every $\theta \in \Theta$,

$$|f_n(\theta, \omega) - f(\theta, \omega)| \leq \epsilon \quad \text{if } n \geq n_0(\omega, \epsilon).$$

Thus going from pointwise a.s. convergence to uniform a.s. convergence we have to check three things, namely that the null set N is independent of θ , that the integer function $n_0(\omega, \epsilon)$ is independent of θ , and that

$$\sup_{\theta \in \Theta} |f_n(\theta) - f(\theta)|$$

is a random variable for each n . Only if so, we shall say that $f_n(\theta) \rightarrow f(\theta)$ a.s. uniformly on Θ . Nevertheless, if this is not the case but $f_n(\theta, \omega) \rightarrow f(\theta, \omega)$ uniformly on Θ for every ω in Ω except in a null set not depending on θ , then we still have a useful property, as will turn out in chapter 4. In this case we shall say that $f_n(\theta) \rightarrow f(\theta)$ a.s. pseudo-uniformly on Θ . Summarizing:

Definition 2.6.1. Let $f(\theta)$ and $f_n(\theta)$ be random functions on a subset Θ of a Euclidean space, and let (Ω, F, P) be the probability space involved. Then:

(a) $f_n(\theta) \rightarrow f(\theta)$ a.s. pointwise on Θ if for every $\theta \in \Theta$ there is a null set N_θ in F and for every $\epsilon > 0$ and every $\omega \in \Omega \setminus N_\theta$ a number $n_0(\omega, \theta, \epsilon)$ such that $|f_n(\theta, \omega) - f(\theta, \omega)| \leq \epsilon$ if $n \geq n_0(\omega, \theta, \epsilon)$;

(b) $f_n(\theta) \rightarrow f(\theta)$ a.s. uniformly on Θ if

(I) $\sup_{\theta \in \Theta} |f_n(\theta) - f(\theta)|$ is a random variable for $n = 1, 2, \dots$, and if

(II) there is a null set N and an integer function $n_0(\omega, \epsilon)$, both independent of θ , such that for every $\epsilon > 0$ and every $\omega \in \Omega \setminus N$, $|f_n(\theta, \omega) - f(\theta, \omega)| \leq \epsilon$ if $n \geq n_0(\omega, \epsilon)$.

(c) $f_n(\theta) \rightarrow f(\theta)$ a.s. pseudo-uniformly on Θ if condition (II) in (b) holds, but not necessarily condition (I).

Similarly to the case of a.s. uniform convergence of random functions the uniform convergence in probability of $f_n(\theta)$ to $f(\theta)$ can be defined by $\text{plim}_{n \rightarrow \infty} \sup_{\theta \in \Theta} |f_n(\theta) - f(\theta)| = 0$, provided that $\sup_{\theta \in \Theta} |f_n(\theta) - f(\theta)|$ is a random variable

for $n = 1, 2, \dots$. In that case it follows from theorem 2.1.6 that $f_n(\theta) \rightarrow f(\theta)$ in pr. uniformly on Θ if and only if every subsequence (n_k) of (n) contains a further subsequence (n_{k_j}) such that

$$f_{n_{k_j}}(\theta) \rightarrow f(\theta) \text{ a.s. uniformly on } \Theta.$$

This suggests how to define pseudo-uniform convergence in pr.:

Definition 2.6.2. Let $f_n(\theta)$ and $f(\theta)$ be random functions on a subset Θ of a Euclidean space. Then:

(a) $f_n(\theta) \rightarrow f(\theta)$ in pr. uniformly on Θ if $\sup_{\theta \in \Theta} |f_n(\theta) - f(\theta)|$ is a random variable for $n = 1, 2, \dots$ satisfying

$$p\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} |f_n(\theta) - f(\theta)| = 0;$$

(b) $f_n(\theta) \rightarrow f(\theta)$ in pr. pseudo-uniformly on Θ if every subsequence (n_k) of (n) contains a further subsequence (n_{k_j}) such that $f_{n_{k_j}}(\theta) \rightarrow f(\theta)$ a.s. pseudo-uniformly on Θ .

Remark: In this study we shall often conclude:

$$\sup_{\theta \in \Theta} |f_n(\theta) - f(\theta)| \rightarrow 0 \text{ a.s. or}$$

$$p\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} |f_n(\theta) - f(\theta)| = 0$$

instead of $f_n(\theta) \rightarrow f(\theta)$ a.s. Θ or $f_n(\theta) \rightarrow f(\theta)$ in pr., uniformly on Θ , respectively. In these cases it will be clear from the context that $\sup_{\theta \in \Theta} |f_n(\theta) - f(\theta)|$ is a random variable for $n = 1, 2, \dots$

We are now able to generalize theorem 2.1.7 to random functions.

Theorem 2.6.1. Let $(f_n(\theta))$ be a sequence of random functions on a Borel subset Θ of a Euclidean space. Let $f(\theta)$ be an a.s. continuous random function on Θ . Let X_n and X be random vectors in Θ such that $P(X \in \Theta) = 1$ and $P(X_n \in \Theta) = 1$ for $n = 1, 2, \dots$. Moreover, suppose that $f(X)$ is a random variable and that $f_n(X_n)$ is a random variable for $n = 1, 2, \dots$. If

- (a) $X_n \rightarrow X$ a.s. and $f_n(\theta) \rightarrow f(\theta)$ a.s. pseudo-uniformly on Θ , or
 (b) $X_n \rightarrow X$ in pr. and $f_n(\theta) \rightarrow f(\theta)$ in pr. pseudo-uniformly on Θ ,

then (a) $f_n(X_n) \rightarrow f(X)$ a.s. or (b) $f_n(X_n) \rightarrow f(X)$ in pr., respectively.

Proof:

(a) Let (Ω, F, P) be the probability space. Let N_1 be the null set on which $x_n(\omega) \rightarrow x(\omega)$ fails to hold, let N_2 be the null set on which $f(\theta, \omega)$ fails to be continuous, let N_3 and $N_{3,n}$ be null sets on which $x(\omega) \in \Theta$ and $x_n(\omega) \in \Theta$, respectively, fail to hold and finally let N_4 be the null set on which

$$\sup_{\theta \in \Theta} |f_n(\theta, \omega) - f(\theta, \omega)| \rightarrow 0$$

fails to hold. Put $N = N_1 \cup N_2 \cup N_3 \cup (\cup_{n=1}^{\infty} N_{3,n}) \cup N_4$. Then $N \in F$, $P(N) = 0$ and for $\omega \in \Omega \setminus N$ we have:

$$\begin{aligned} & |f_n(x_n(\omega), \omega) - f(x(\omega), \omega)| \\ & \leq \sup_{\theta \in \Theta} |f_n(\theta, \omega) - f(\theta, \omega)| + |f(x_n(\omega), \omega) - f(x(\omega), \omega)| \rightarrow 0. \end{aligned}$$

This proves part (a). Part (b) follows from (a) by using theorem 2.1.6. Q.E.D.

Exercise:

1. Let X be uniformly distributed on $[0, 1]$. Define for $\theta \in [0, 1]$,

$$f_n(\theta) = n^{-1} |X - \theta|.$$

Show that $f_n(\theta) \rightarrow 0$ a.s. pointwise on $\theta = [0,1]$ while $\sup_{\theta \in \Theta} |f_n(\theta)| = 1$. (This is a counter-example that pointwise a.s. convergence on compact spaces does not imply uniform a.s. convergence.)

2.7 Uniform strong and weak laws of large numbers

Next we shall extend the theorems 1 and 2 of Jennrich (1969). We shall closely follow Jennrich's proof, but instead of the Helly-Bray theorem (theorem 2.3.1) we shall now use theorems 2.5.1 and 2.5.5. The extension involved is:

Theorem 2.7.1. Let X_1, X_2, \dots be a sequence of independent random vectors in R^k with distribution functions F_1, F_2, \dots , respectively. Let $f(x, \theta)$ be a continuous real function on $R^k \times \Theta$, where Θ is a compact Borel set in R^m . If

$$(1/n) \sum_{j=1}^n F_j \rightarrow G \text{ properly, pointwise} \quad (2.7.1)$$

and

$$\sup_n (1/n) \sum_{j=1}^n E \sup_{\theta \in \Theta} |f(x, \theta)|^{2+\delta} < \infty \quad (2.7.2)$$

then

$$(1/n) \sum_{j=1}^n f(X_j, \theta) \rightarrow \int f(x, \theta) dG(x) \text{ a.s. uniformly on } \Theta,$$

where the limit function involved is continuous on Θ .

Proof: First we note that by theorem 1.6.1 the supremum in (2.7.2) is a random variable.

For the sake of convenience and clarity we shall label the main steps of the proof.

Step 1: Choose θ_0 arbitrarily in Θ and put for $\delta \geq 0$

$$\Gamma_\delta = \{\theta \in R^m : |\theta - \theta_0| \leq \delta\} \cap \Theta,$$

Then for any $\delta \geq 0$,

$$\sup_{\theta \in \Gamma_\delta} f(x, \theta) \text{ and } \inf_{\theta \in \Gamma_\delta} f(x, \theta)$$

are continuous functions on R^k , because Γ_δ is a closed subset of a compact set and therefore compact itself. See Rudin (1976, theorem 2.35) and compare theorem 1.6.1. Moreover,

$$|\sup_{\theta \in \Gamma_\delta} f(x, \theta)| \leq \sup_{\theta \in \Theta} |f(x, \theta)|, \quad (2.7.3)$$

$$|\inf_{\theta \in \Gamma_\delta} f(x, \theta)| \leq \sup_{\theta \in \Theta} |f(x, \theta)|. \quad (2.7.4)$$

Thus it follows from theorem 2.5.3 and the conditions (2.7.1) and (2.7.2) that

$$(1/n) \sum_{j=1}^n \sup_{\theta \in \Gamma_\delta} f(X_j, \theta) \rightarrow \int \sup_{\theta \in \Gamma_\delta} f(x, \theta) dG(x) \rightarrow \text{a.s.}, \quad (2.7.5)$$

and

$$(1/n) \sum_{j=1}^n \inf_{\theta \in \Gamma_\delta} f(X_j, \theta) \rightarrow \int \inf_{\theta \in \Gamma_\delta} f(x, \theta) dG(x) \text{ a.s.} \quad (2.7.6)$$

Step 2: By continuity,

$$\sup_{\theta \in \Gamma_\delta} f(x, \theta) - \inf_{\theta \in \Gamma_\delta} f(x, \theta) \rightarrow 0 \text{ as } \delta \downarrow 0,$$

pointwise in x . It follows now from the dominated convergence theorem that

$$\lim_{\delta \downarrow 0} \left| \int \sup_{\theta \in \Gamma_\delta} f(x, \theta) dG(x) - \int \inf_{\theta \in \Gamma_\delta} f(x, \theta) dG(x) \right| = 0. \quad (2.7.7)$$

Step 3: Choose $\varepsilon > 0$ arbitrarily. From (2.7.7) it follows that $\delta > 0$ can be chosen so small, say $\delta = \delta(\varepsilon)$, that

$$0 \leq \int \sup_{\theta \in \Gamma_{\delta(\varepsilon)}} f(x, \theta) dG(x) - \int \inf_{\theta \in \Gamma_{\delta(\varepsilon)}} f(x, \theta) dG(x) \leq \frac{1}{2}\varepsilon. \quad (2.7.8)$$

Let (Ω, F, P) be the probability space involved. From (2.7.5) and (2.7.6) it follows that there is a null set N and for each $\omega \in \Omega \setminus N$ a number $n_0(\omega, \varepsilon)$ such that:

$$\left| (1/n) \sum_{j=1}^n \sup_{\theta \in \Gamma_{\delta}(\varepsilon)} f(x_j(\omega), \theta) - \int \sup_{\theta \in \Gamma_{\delta}(\varepsilon)} f(x, \theta) dG(x) \right| \leq \frac{1}{2} \varepsilon, \quad (2.7.9)$$

$$\left| (1/n) \sum_{j=1}^n \inf_{\theta \in \Gamma_{\delta}(\varepsilon)} f(x_j(\omega), \theta) - \int \inf_{\theta \in \Gamma_{\delta}(\varepsilon)} f(x, \theta) dG(x) \right| \leq \frac{1}{2} \varepsilon, \quad (2.7.10)$$

if $n \geq n_0(\omega, \varepsilon)$. From (2.7.8), (2.7.9) and (2.7.10) it follows now that for every $\omega \in \Omega \setminus N$, every $n \geq n_0(\omega, \varepsilon)$ and every $\theta \in \Gamma_{\delta}(\varepsilon)$:

$$\begin{aligned} & (1/n) \sum_{j=1}^n f(x_j(\omega), \theta) - \int f(x, \theta) dG(x) \\ & \leq (1/n) \sum_{j=1}^n \sup_{\theta \in \Gamma_{\delta}(\varepsilon)} f(x_j(\omega), \theta) - \int \inf_{\theta \in \Gamma_{\delta}(\varepsilon)} f(x, \theta) dG(x) \\ & \leq \left| (1/n) \sum_{j=1}^n \sup_{\theta \in \Gamma_{\delta}(\varepsilon)} f(x_j(\omega), \theta) - \int \sup_{\theta \in \Gamma_{\delta}(\varepsilon)} f(x, \theta) dG(x) \right| \\ & \quad + \left| \int \sup_{\theta \in \Gamma_{\delta}(\varepsilon)} f(x, \theta) dG(x) - \int \inf_{\theta \in \Gamma_{\delta}(\varepsilon)} f(x, \theta) dG(x) \right| \leq \varepsilon \end{aligned}$$

and similarly:

$$(1/n) \sum_{j=1}^n f(x_j(\omega), \theta) - \int f(x, \theta) dG(x) \geq -\varepsilon.$$

Thus for $\omega \in \Omega \setminus N$ and $n \geq n_0(\omega, \varepsilon)$ we have:

$$\sup_{\theta \in \Gamma_{\delta}(\varepsilon)} \left| (1/n) \sum_{j=1}^n f(x_j(\omega), \theta) - \int f(x, \theta) dG(x) \right| \leq \varepsilon.$$

We note that the null set N and the number $n_0(\omega, \varepsilon)$ depend on the set $\Gamma_{\delta}(\varepsilon)$, which in its turn depends on θ_0 and ε . Thus the above result should be restated as follows. For every θ_0 in Θ and every $\varepsilon > 0$ there is a null set $N(\theta_0, \varepsilon)$ and an integer function $n_0(\cdot, \varepsilon, \theta_0)$ on $\Omega \setminus N(\theta_0, \varepsilon)$ such that for $\omega \in \Omega \setminus N(\theta_0, \varepsilon)$ and $n \geq n_0(\omega, \varepsilon, \theta_0)$:

$$\sup_{\theta \in \Gamma_{\delta}(\varepsilon)(\theta_0)} \left| (1/n) \sum_{j=1}^n f(x_j(\omega), \theta) - \int f(x, \theta) dG(x) \right| \leq \varepsilon, \quad (2.7.11)$$

where

$$\Gamma_{\delta}(\theta_0) = \{ \theta \in R^k : |\theta - \theta_0| \leq \delta \} \cap \Theta. \quad (2.7.12)$$

Step 4: The collection of sets $\{ \theta \in R^k : |\theta - \theta_0| < \delta \}$ with $\theta_0 \in \Theta$ is an open covering of Θ . Since Θ is compact, there exists by

definition of compactness a finite covering. Thus there are a finite number of points in Θ , say

$$\theta_{\delta,1}, \dots, \theta_{\delta,r_{\delta}} \text{ with } r_{\delta} < \infty$$

such that

$$\Theta \subset \bigcup_{i=1}^{r_{\delta}} \{\theta \in \mathbb{R}^k : |\theta - \theta_{\delta,i}| < \delta\}.$$

Using (2.7.12) we therefore have:

$$\Theta = \bigcup_{i=1}^{r_{\delta}} \Gamma_{\delta}(\theta_{\delta,i}). \quad (2.7.13)$$

Now put:

$$N_{\varepsilon} = \bigcup_{i=1}^{r_{\delta}(\varepsilon)} N(\theta_{\delta}(\varepsilon), i, \varepsilon),$$

$$n_{*}(\omega, \varepsilon) = \max_{1 \leq i \leq r_{\delta}(\varepsilon)} n_0(\omega, \varepsilon, \theta_{\delta}(\varepsilon), i).$$

Then by (2.7.12) and (2.7.13) we have for $\omega \in \Omega \setminus N_{\varepsilon}$ and $n \geq n_{*}(\omega, \varepsilon)$,

$$\begin{aligned} & \sup_{\theta \in \Theta} \left| (1/n) \sum_{j=1}^n f(x_j(\omega), \theta) - \int f(x, \theta) dG(x) \right| \\ & \leq \max_{1 \leq i \leq r_{\delta}(\varepsilon)} \sup_{\theta \in \Gamma_{\delta}(\varepsilon)(\theta_{\delta}(\varepsilon), i)} \left| (1/n) \sum_{j=1}^n f(x_j(\omega), \theta) \right. \\ & \quad \left. - \int f(x, \theta) dG(x) \right| \leq \varepsilon. \end{aligned}$$

Since it can be shown, similarly to the proof of theorem 2.2.1, that the null set N_{ε} can be chosen independently of ε , it follows now that

$$(1/n) \sum_{j=1}^n f(X_j, \theta) \rightarrow \int f(x, \theta) dG(x) \text{ a.s. pseudo-uniformly on } \Theta. \quad (2.7.14)$$

Step 5: From (2.7.7) it follows that $\int f(x, \theta) dG(x)$ is a continuous function on Θ . Using theorem 2.3.1, it is now easy to verify that

$$\sup_{\theta \in \Theta} \left| (1/n) \sum_{j=1}^n \hat{f}(X_j, \theta) - \int f(x, \theta) dG(x) \right|$$

is a random variable, so that (2.7.14) becomes

$$(1/n)\sum_{j=1}^n f(X_j, \theta) \rightarrow \int f(x, \theta) dG(x) \text{ a.s. uniformly on } \Theta.$$

This completes the proof.

Q.E.D.

If condition (2.7.2) is only satisfied with $1+\delta$ instead of $2+\delta$ then we can no longer apply theorem 2.5.3 for proving (2.7.5) and (2.7.6). However, applying theorem 2.5.2 we see that (2.7.5) and (2.7.6) still hold in probability. From theorem 2.1.6 it then follows that any subsequence (n_k) of (n) contains further subsequences

$$(n_{k_m}^{(1)}) \text{ and } (n_{k_m}^{(2)}),$$

say such that for $m \rightarrow \infty$,

$$(1/n_{k_m}^{(1)}) \sum_{j=1}^{n_{k_m}^{(1)}} \sup_{\theta \in \Gamma_\delta} f(X_j, \theta) \rightarrow \int \sup_{\theta \in \Gamma_\delta} f(x, \theta) dG(x) \text{ a.s.}$$

$$(1/n_{k_m}^{(2)}) \sum_{j=1}^{n_{k_m}^{(2)}} \inf_{\theta \in \Gamma_\delta} f(X_j, \theta) \rightarrow \int \inf_{\theta \in \Gamma_\delta} f(x, \theta) dG(x) \text{ a.s.}$$

Note that we may assume without loss of generality that these further subsequences are equal:

$$n_{k_m} = n_{k_m}^{(1)} = n_{k_m}^{(2)}.$$

We now conclude from the argument in the proof of theorem 2.7.1 that

$$\sup_{\theta \in \Theta} |(1/n_{k_m}) \sum_{j=1}^{n_{k_m}} f(X_j, \theta) - \int f(x, \theta) dG(x)| \rightarrow 0 \text{ a.s.}$$

as $m \rightarrow \infty$. Again using theorem 2.1.6 we then conclude:

Theorem 2.7.2. Let the conditions of theorem 2.7.1 be satisfied, except (2.7.2). If

$$\sup_n (1/n) \sum_{j=1}^n E \sup_{\theta \in \Theta} |f(X_j, \theta)|^{1+\delta} < \infty \text{ for some } \delta > 0,$$

then

$$(1/n) \sum_{j=1}^n f(X_j, \theta) \rightarrow \int f(x, \theta) dG(x) \text{ in pr. uniformly on } \Theta,$$

where the limit function involved is continuous on Θ .

Next, let $f(x, \theta)$ be Borel measurable in both arguments and for each $x \in R^k$ continuous on Θ . Referring to theorems 2.5.6 and 2.5.7 instead of theorems 2.5.2 and 2.5.3, respectively, we have:

Theorem 2.7.3. Let X_1, X_2, \dots be a sequence of independent random vectors in R^k with distribution functions F_1, F_2, \dots , respectively. Let $f(x, \theta)$ be a Borel measurable function on $R^k \times \Theta$, where Θ is a compact Borel set in R^m , which is continuous in θ for each $x \in R^k$. If

$$(1/n) \sum_{j=1}^n F_j \rightarrow G \text{ properly setwise} \quad (2.7.15)$$

and

$$\sup_n (1/n) \sum_{j=1}^n E \sup_{\theta \in \Theta} |f(X_j, \theta)|^{2+\delta} < \infty \text{ for some } \delta > 0 \quad (2.7.16)$$

then

$$(1/n) \sum_{j=1}^n f(X_j, \theta) \rightarrow \int f(x, \theta) dG(x) \text{ a.s. uniformly on } \Theta,$$

where the limit function involved is continuous on Θ .

Theorem 2.7.4. Let the conditions of theorem 2.7.3 be satisfied, except condition (2.7.16). If

$$\sup_n (1/n) \sum_{j=1}^n E \sup_{\theta \in \Theta} |f(X_j, \theta)|^{1+\delta} < \infty \text{ for some } \delta > 0 \quad (2.7.17)$$

then

$$(1/n) \sum_{j=1}^n f(X_j, \theta) \rightarrow \int f(x, \theta) dG(x) \text{ in pr. uniformly on } \Theta,$$

where the limit function involved is continuous on Θ .

Finally, if the sequence (X_j) is i.i.d. we can relax the moment conditions (2.7.16) and (2.7.17) further, due to Kolmogorov's strong law (cf. theorem 2.1.5):

Theorem 2.7.5. Let the conditions of theorem 2.7.3 be satisfied, except condition (2.7.16). If (X_j) is i.i.d. with

$$F_j = G \text{ and } E \sup_{\theta \in \Theta} |f(X_j, \theta)| < \infty$$

then the conclusion of theorem 2.7.3 carries over.

Exercise:

1. Restate theorems 2.7.2 and 2.7.4 for double arrays $(X_{n,j})$, $j=1,2,\dots,n$, $n=1,2,\dots$ of random vectors in R^k (cf. exercise 2 in section 2.5).

References:

Anderson, T.W. (1958), *An Introduction to Multivariate Statistical Analysis*, New York: John Wiley

Bierens, H.J. (1981), *Robust Methods and Asymptotic Theory in Nonlinear Econometrics*, Lecture Notes in Economics and Mathematical Systems, Vol. 192, Berlin: Springer Verlag

Chung, K.L. (1974), *A Course in Probability Theory*, New York: Academic Press

Feller, W. (1966), *An Introduction to Probability Theory and Its Applications, Vol. II*, New York: John Wiley.

Jennrich, R.I. (1969), "Asymptotic Properties of Nonlinear Least Squares Estimators", *Annals of Mathematical Statistics* 40, 633-634

Revesz, P. (1968), *The Laws of Large Numbers*, New York: Academic Press

Rudin, W. (1976), *Principles of Mathematical Analysis*, Tokyo: Mac Graw Hill Kogakusha

Stout, W.F. (1974), *Almost Sure Convergence*, New York: Academic Press