SERIE RESEARCH MEMORANDA

BASIC PROBABILITY THEORY

Herman J. Bierens

Researchmemorandum 1987-18 April '87

VRIJE UNIVERSITEIT
Faculteit der Economische Wetenschappen en Econometrie
A M S T E R D A M

BASIC PROBABILITY THEORY *

Herman J.Bierens
Depa:tment of Econometrics
Free University, Amsterdam

December 1988

---

* Chapter 1 of:
PRINCIPLES OF NONLINEAR AND NONPARAMETRIC REGRESSION ANALYSIS
To be published by Marcel Dekker Inc., New York.

PRINCIPLES OF NONLINEAR AND NONPARAMETRIC REGRESSION ANALYSIS

TABLE OF CONTENTS:

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 parametric 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.

1. BASIC PROBABILITY THEORY

The asymptotic theory of nonlinear regression models, in particular consistency results, heavily depends on uniform laws of large numbers. Understanding these laws requires knowledge of abstract probability theory. In this chapter we shall therefore review the basic elements of this theory so far as needed in the sequel, in order to make this book almost self-contained. However, we do assume that the reader has already a good knowledge of probability and statistics on intermediate level, say on the level of Hogg and Craig (1978). The material in this chapter is a revision and extension of section 2.1 in Bierens (1981).

1.1 *Measure-theoretical foundation of probability theory*

The basic concept of probability theory is the *probability space*. This is a triple (Ω,F,P) consisting of:

- An abstract non-empty set Ω, called the *sample space*. We do not impose any conditions on this set.
- A non-empty collection F of subsets of Ω, having the following two properties:

$$\text{If } E \in F, \text{ then } E^c \in F, \tag{1.1.1}$$

where E^c denotes the complement of the subset E with respect to Ω: $E^c = \Omega\backslash E$.

$$\text{If } E_j \in F \text{ for } j=1,2,\ldots, \text{ then } \cup_j E_j \in F. \tag{1.1.2}$$

These two properties make F, by definition, a *Borel field** of subsets of Ω.

- A *probability measure* P on (Ω,F). This is a real-valued set function on F such that:

$$P(\Omega) = 1, \tag{1.1.3}$$

$$P(E) \geq 0 \text{ for all } E \in F, \tag{1.1.4}$$

---

*) Following Chung (1974), the term "Borel field" has the same meaning as the term "σ-algebra" used by other authors.

$E_j \in F$ for $j=1,2,\ldots$ and $E_{j_1} \cap E_{j_2} = \emptyset$ if $j_1 \neq j_2$ imply $P(\cup_j E_j) = \Sigma_j P(E_j)$. (1.1.5)

Example: Toss a fair coin. The possible outcomes are head (H) or tail (T). Thus $\Omega = \{H,T\}$. The collection F of all subsets of Ω, i.e.,

$$F = \{\Omega,\emptyset,\{H\},\{T\}\}$$

is a Borel field. Finally, the appropriate probability measure in this case is

$$P(\{H\}) = P(\{T\}) = 1/2,\ P(\Omega) = 1,\ P(\emptyset) = 0.$$

Now let X be a *random variable* (r.v.) and let F be its distribution function. In the measure-theoretical approach of probability theory a random variable is considered as a real valued *function* on the set Ω denoted by:

$$X = x(.)$$

with value $x(\omega)$ at $\omega \in \Omega$, such that for every real number t:

$$\{\omega \in \Omega : x(\omega) \leq t\} \in F.$$

The *distribution function* F with value F(t) at $t \in R$ is then defined by:

$$F(t) = P(\{\omega \in \Omega : x(\omega) \leq t\}),$$

which will often be denoted by the short-hand notation:

$$F(t) = P(X \leq t).$$

Example: In the coin tossing case the function

$$x(H) = 1,\ x(T) = 0$$

determines a random variable X. The corresponding distribution

function is:

$$F(t) = 1 \text{ if } t \geq 1,$$
$$F(t) = 1/2 \text{ if } 0 \leq t < 1,$$
$$F(t) = 0 \text{ if } t < 0.$$

It is not hard to see that the axioms (1.1.1) and (1.1.2) imply:

$$E_n \in F\ ,\quad n=1,2,\ldots \ \Rightarrow\ \cap_n E_n \in F$$

and that from (1.1.3), (1.1.4) and (1.1.5):

$$P(\emptyset) = 0,\ P(E^c) = 1 - P(E),\ P(E\cup D) + P(E\cap D) = P(E) + P(D),$$

$$E \subset D \Rightarrow P(E) \leq P(D\backslash E) + P(E) = P(D),$$

$$E_n \subset E_{n+1},\quad E = \cup_n E_n \ \Rightarrow\ P(E_n) \to P(E) \text{ as } n \to \infty,$$

$$E_n \supset E_{n+1},\quad E = \cap_n E_n \ \Rightarrow\ P(E_n) \to P(E) \text{ as } n \to \infty,$$

$$P(\cup_n E_n) \leq \Sigma_n P(E_n),$$

where all sets involved are members of F. Moreover, the distribution function $F(t)$ is *right continuous*:

$$F(t) = \lim_{\varepsilon \downarrow 0} F(t+\varepsilon),$$

as is easily verified, and it satisfies

$$F(\infty) = \lim_{t\to\infty} F(t) = 1\ ,\quad F(-\infty) = \lim_{t\to-\infty} F(t) = 0.$$

Furthermore, by $F(t-)$ we denote:

$$F(t-) = \lim_{\varepsilon \downarrow 0} F(t-\varepsilon),$$

which clearly satisfies $F(t-) \leq F(t)$.

A finite dimensional *random vector* can now be defined as a vector with random variables as components, where these random components are assumed to be defined on a *common* probability space. Moreover, a *complex random variable* Z can be defined by $Z = X + i\cdot Y$ with real valued random variables X and

Y defined on a common probability space as real and imaginary part, respectively.

Next we shall construct a Borel field B^k of subsets of R^k such that for every set $B \in B^k$ and any k-dimensional random vector X on a probability space $\{\Omega,F,P\}$ we have

$$\{\omega \in \Omega : x(\omega) \in B\} \in F, \tag{1.1.6}$$

because only for such subsets B of R^k we can define the probability

$$P(X \in B) = P(\{\omega \in \Omega : x(\omega) \in B\}), \tag{1.1.7}$$

Let C be the collection of subsets of R^k of the type

$$\times_{m=1}^{k}(-\infty,t_m], \quad t_m \in R,$$

and let G be the Borel field of all subsets of R^k. Clearly we have $C \subset G$, that means that $E \in C \Rightarrow E \in G$. But next to G there may be other Borel fields of subsets of R^k with this property, say G_a, $a \in A$, where A is an index set. Assuming that all Borel fields containing C are represented this way, we then have a non-empty collection of Borel fields G_a, $a \in A$, of subsets of R^k such that $C \subset G_a$ for each $a \in A$. Now consider the collection

$$B^k = \cap_{a\in A}G_a .$$

Since each G_a is a Borel field, it follows that this collection B^k is a Borel field of subsets of R^k and since C is contained in each G_a it follows that $C \subset B^k$. We shall say that the Borel field B^k is the *minimal Borel field* containing the collection C, and for this particular collection C it is called the *Euclidean Borel field*. Summarizing:

Definition 1.1.1. Let $\mathcal{C}$ be any collection of subsets of a set Γ and let the Borel fields of subsets of Γ containing C be G_a, $a \in A$. Then $G = \cap_{a\in A}G_a$ is called the *minimal Borel field* containing C.

Definition 1.1.2. Let C be the collection of subsets of R^k of the type

$$\times_{m=1}^{k}(-\infty, t_m], \quad t_m \in R.$$

The minimal Borel field B^k containing this collection is called the *Euclidean Borel field* (also called the Borel σ-Algebra) and the members of B^k are called *Borel sets*.

The concept of Borel sets is very general. Roughly speaking, any subset of R^k you can imagine is a Borel set. For example, each singleton in R^k is a Borel set, the area in a circle is a Borel set in R^2, the circle itself and the straight line are Borel sets in R^2, the (hyper)cube is a Borel set, etc. Thus "almost" every subset of R^k is a Borel set. However, there are exceptions, but the shape of a set in R^k that is not a Borel set is complicated beyond our imagination. See Royden (1968, p.63-64) for an example.

We show now that for any Borel set B and any r.v. X on (Ω,F,P), (1.1.6) is satisfied. Let D be the collection of all Borel sets B such that (1.1.6) is satisfied. Then $D \subset B^k$. If D is also a Borel field then $B^k \subset D$ (and hence $B^k = D$) because B^k is the minimal Borel field containing the collection C in definition 1.1.2, whereas obviously the collection D contains C. So it suffices to prove that D is a Borel field. However, this is not too hard and therefore left to the reader. This proves the first part of theorem 1.1.1. below. The proof of the second part is left as an easy exercise.

Theorem 1.1.1. For any random vector X in R^k defined on (Ω,F,P) and any Borel set B in R^k we have $\{\omega \in \Omega : x(\omega) \in B\} \in F$. The collection $F(X)$ of sets $\{\omega \in \Omega : x(\omega) \in B\}$ with B an arbitrary Borel set in R^k is a Borel field itself, contained in F.

Consequently the definition (1.1.2) is meaningful for Borel sets. In fact, by defining a measure μ on the Euclidean Borel field B^k as

$$\mu(B) = P(X \in B) = P(\{\omega \in \Omega : x(\omega) \in B\})$$

for any Borel set B in R^k we have created a probability measure on (R^k,B^k). This probability measure μ is often referred to as

the probability measure *induced* by (the random variable or -vector) X. Moreover, the Borel field $F(X)$ is called the *Borel field generated by* X. This concept plays an important role in defining conditional expectations.

We are now able to define a (joint) distribution function on R^k. Let X be a random vector in R^k defined on a probability space (Ω,F,P). The product sets

$$\times_{j=1}^{k}(-\infty,t_j]$$

are Borel sets in R^k, where the t_j's are the components of a vector $t \in R^k$. Thus:

$$\{\omega \in \Omega : x(\omega) \in \times_{j=1}^{k}(-\infty,t_j]\} \in F.$$

The (joint) *distribution function* F, say, of X is now defined for all $t \in R^k$ by:

$$F(t) = P(\{\omega \in \Omega : x(\omega) \in \times_{j=1}^{k}(-\infty,t_j]\}) = \mu(\times_{j=1}^{k}(-\infty,t_j]),$$

where μ is the probability measure induced by X. However, F(t) will often also be denoted by the shorthand notation:

$$F(t) = P(X \leq t).$$

Clearly, F is uniquely determined by μ. The reverse is also true, i.e.:

Theorem 1.1.2. Given a distribution function F on R^k there exists a *unique* probability measure μ on (R^k,B^k) defining F.

Proof: Similarly to Royden (1968, Proposition 12, p. 262).

Thus there is a one-to-one correspondence between a distribution function F and its defining probability measure μ. We shall employ this result later in defining mathematical expectations.

Exercises:

1. Show that (1.1.1) and (1.1.2) imply $\cap_j E_j \in F$.

2. Consider the collection F of subintervals of [0,1] with rational-valued endpoints, together with their complements and *finite* unions and intersections. Show that F is not a Borel field. (Hint: Use the fact that irrational numbers can be written as limits of rational numbers.)

3. Let F be a Borel field of subsets of Ω. Let $A \in F$ and let G_A be the collection of all subsets of the type $A \cap B$, where $B \in F$. Prove that G_A is a Borel field of subsets of A.

4. Let $\Omega = \{1,2,3,4,5\}$ and let C be the collection consisting of the two subsets $\{2\}$, $\{4\}$. Prove that the minimal Borel field containing C consists of the following sets: Ω, $\emptyset$, $\{2\}$, $\{4\}$, $\{2\} \cup \{1,3,5\}$, $\{4\} \cup \{1,3,5\}$, $\{2,4\}$, $\{1,3,5\}$.

5. Prove that the following subsets of R^2 are Borel sets:
- a rectangle,
- the area in a circle,
- a straight line.

6. Prove that the set Q of rational numbers is a Borel set in R. (Hint: Use the countability of Q)

7. Let C be the collection of all intervals of the type [a,b], where a and b are finite and $a < b$. Prove that the minimal Borel field containing C is just the Euclidean Borel field.

8. Complete the proof of theorem 1.1.1.

9. Prove that a distribution function is always right continuous.

10. Let $S = \{1,2,3,4\}$, let Ω be the set of all pairs (x_1, x_2) with $x_1 \in S$, $x_2 \in S$, $x_1 < x_2$, let F be the Borel field of all subsets of Ω and let P be a probability measure on (Ω, F). Define the random variable Y as follows:

$$y(\omega) = 1 \quad \text{if} \quad \omega = (x_1, x_2) \quad \text{with} \quad x_1 + x_2 \quad \text{is odd.}$$

$$y(\omega) = 0 \quad \text{if} \quad \omega = (x_1, x_2) \quad \text{with} \quad x_1 + x_2 \quad \text{is even.}$$

- Determine the Borel field $F(Y)$ generated by Y.
- All points in Ω have equal probability. Derive the distribution function F(y) of Y.

11. A bowl contains 9 white balls and 1 red ball of equal size. Draw randomly one ball and assign to it the value $X_1 = 1$ if the ball is red and the value $X_1 = 0$ if the ball is white. Draw randomly a second ball without replacing the first ball, and

assign the value $X_2 = 1$ to it if it is red and the value $X_2 = 0$ if it is white. Consider the vector $Y = (X_1, X_2)'$. Define a sample space Ω and a Borel field F of subsets of Ω such that together:
- Y is a random vector,
- F is equal to the Borel field generated by Y.
Also, define an appropriate probability measure P.

1.2 *Independence*

Let $X_1, X_2, \ldots$ be a sequence of random variables with corresponding probability spaces $\{\Omega_1, F_1, P_1\}, \{\Omega_2, F_2, P_2\}, \ldots$ respectively. It is possible to construct a new probability space, $\{\Omega, F, P\}$, say, such that the X_j's can be regarded as *independent* random variables on $\{\Omega, F, P\}$ [see Chung (1974, section 3.3)]. Independence means:

Definition 1.2.1. Let $X_1, X_2, \ldots$ be random vectors in $X_1, X_2, \ldots$, respectively, defined on a common probability space, where the X_j are Euclidean spaces (possibly with different dimensions). This sequence of random vectors is called *(totally) independent* if for any sequence (B_j) of Borel sets (with B_j a Borel set in X_j),

$$P(\cap_j \{\omega \in \Omega: x_j(\omega) \in B_j\}) = \prod_j P(\{\omega \in \Omega: x_j(\omega) \in B_j\})$$

and it is called *mutually (or pairwise) independent* if for $j_1 \neq j_2$, X_{j_1} and X_{j_2} are independent.

As an example, consider the tossing of a fair coin. Assign to X_j the value 1 if the outcome of the j-th tossing is head (H) and assign the value 0 if the outcome is tail (T). The X_j is a random variable defined on the probability space $\{\Omega_j, F_j, P_j\}$, where

$$\Omega_j = \{H,T\}, \quad F_j = \{\emptyset, \{H\}, \{T\}, \{H,T\}\}$$

$$P_j(\emptyset) = 0, \quad P_j(\{H,T\}) = 1, \quad P_j(\{H\}) = P_j(\{T\}) = 1/2.$$

Now let Ω be the set of all one-sided infinite sequence of H and T, for example, let $\omega = (H,H,T,T,H,T,T,T,H,H,H,T,\ldots\ldots)$ be

such an element of Ω. We could take as F the collection of all subsets of Ω, including the empty set. Now x_j can also be defined on (Ω,F), as follows. Let

$$x_j(\omega) = 1 \quad \text{if the j-th element of } \omega \text{ is H},$$
$$x_j(\omega) = 0 \quad \text{if the j-th element of } \omega \text{ is T}.$$

Each set $E \in F$ can be written as $E = E_{j,1} \cup E_{j,2}$, where $E_{j,1}$ and $E_{j,2}$ are disjoint sets defined by:

$$E_{j,1} = \{\omega \in E : \text{j-th element of } \omega \text{ is H}\},$$
$$E_{j,2} = \{\omega \in E : \text{j-th element of } \omega \text{ is T}\}.$$

Of course, $E_{j,1}$ or $E_{j,2}$ may be empty. Now define

$$\delta(E_{j,i}) = \tfrac{1}{2} \quad \text{if} \quad E_{j,i} \neq \emptyset,$$
$$\delta(E_{j,i}) = 0 \quad \text{if} \quad E_{j,i} = \emptyset \ , \ i = 1,2,$$

$$P(E) = \prod_j [\delta(E_{j,1}) + \delta(E_{j,2})].$$

Then (X_j) is a sequence of independent random variables defined on the common probability space (Ω,F,P).

1.3 *Borel measurable functions*

If X is a r.v. and f(x) is a real function on R, is then f(X) a r.v.? The answer is: not always. There are functions [see for example Royden (1968, problem 3.28)] for which this is not the case. The condition for f(X) being an r.v. is that for all $t \in R$ we have

$$\{\omega \in \Omega : f(x(\omega)) \leq t\} \in F,$$

where (Ω,F,P) is the probability space involved, and referring to theorem 1.1.1 we see that this will be the case if for every $t \in R$,

$$\{x \in R : f(x) \leq t\} \quad \text{is a Borel set in R.}$$

Functions satisfying the latter condition are called *Borel measurable*. Now consider a real function $f(x_1,\ldots,x_k)$ on R^k and r.v.'s $X_1,\ldots,X_k$ on (Ω,F,P). If for every $t \in R$ the set

$$B_t = \{(x_1,\ldots,x_k) \in R^k : f(x_1,\ldots,x_k) \leq t\}$$

is a Borel set in R^k then

$$\{\omega \in \Omega : f(x_1(\omega),\ldots,x_k(\omega)) \leq t\} \in F$$

for every $t \in R$ and hence $f(X_1,\ldots,X_k)$ is an r.v. Also such functions are called Borel measurable.

Definition 1.3.1. A real function $f(x)$ on R^k is called *Borel measurable* if for every $t \in R$ the set $\{x \in R^k : f(x) \leq t\}$ is a Borel set in R^k.

A first example of a Borel measurable function is the so-called *simple function*:

Definition 1.3.2. A real function $f(x)$ on R^k is called a simple function if there are finite real numbers $b_1,\ldots,b_n$ and Borel sets B_j, $j=1,2,\ldots,n$ with $B_{j_1} \cap B_{j_2} = \emptyset$ if $j_1 \neq j_2$, such that

$$f(x) = \Sigma_{j=1}^{n} b_j I(x \in B_j),$$

where $I(.)$ is the indicator function, i.e.,

$$I(x \in B_j) = 1 \quad \text{if } x \in B_j;\ I(x \in B_j) = 0 \quad \text{if } x \notin B_j.$$

Simple functions differ from the well-known step functions in that for step functions the disjoint sets B_j are restricted to intervals. Since intervals are Borel sets, step functions are simple functions.

Realizing that for a simple function f the set $\{x \in R^k : f(x) \leq t\}$ is always a finite union of Borel sets, we have:

Theorem 1.3.1. Simple functions are Borel measurable.

From this result we can derive other Borel measurable functions using the following theorem.

Theorem 1.3.2. Let $f_1, f_2, \ldots$ be a sequence of Borel measurable functions on R^k. Then the functions $\max\{f_1, \ldots, f_n\}$, $\min\{f_1, \ldots, f_n\}$, $\sup_n f_n$, $\inf_n f_n$, $\limsup_{n\to\infty} f_n$ and $\liminf_{n\to\infty} f_n$ are also Borel measurable.

Proof: We only consider the case $k = 1$. Moreover, it is not hard to see that if f is Borel measurable then so is $-f$, hence it suffices to prove the theorem for the "max" and "sup"-cases. Let

$$h_n(x) = \max\{f_1(x), \ldots, f_n(x)\}.$$

Then

$$\{x \in R : h_n(x) \leq t\} = \cap_j \{x \in R : f_j(x) \leq t\},$$

which is a Borel set since the f_j's are Borel measurable. Moreover, replacing n by ∞ we see that $\sup_n f_n(x)$ is Borel measurable. Since

$$\limsup_{n\to\infty} f_n(x) = \inf_n \sup_{k\geq n} f_k(x)$$

and since $\inf_n g_n(x)$ is Borel measurable if the g_n are Borel measurable, it follows directly that $\limsup_{n\to\infty} f_n(x)$ is Borel measurable. Q.E.D.

Along the same lines it can be shown:

Corollary 1.3.1. If $X_1, X_2, X_3, \ldots$ are random variables defined on a common probability space, then so are $\max\{X_1, \ldots, X_n\}$, $\min\{X_1, \ldots, X_n\}$, $\sup_n X_n$, $\inf_n X_n$, $\limsup_{n\to\infty} X_n$ and $\liminf_{n\to\infty} X_n$.

From theorems 1.3.1 and 1.3.2 it follows now:

Theorem 1.3.3. Continuous real functions on R^k are Borel measurable.

Proof: Let $f(x)$ be a continuous real function on R^k. It is not hard to show that for fixed n the functions $f_n(x)$ defined by

$$f_n(x) = f(x) \quad \text{if } |x| \leq n; \quad f_n(x) = 0 \quad \text{if } |x| > n; \quad n = 1, 2, \ldots$$

can be written as limits of simple functions, so that by theorem 1.3.2 the $f_n(x)$ s are Borel measurable. Since

$$f(x) = \lim_{n\to\infty} f_n(x) = \limsup_{n\to\infty} f_n(x) = \liminf_{n\to\infty} f_n(x)$$

it follows from theorem 1.3.2 that f is Borel measurable.

Q.E.D.

Let f be any Borel measurable function on R^k. Since the functions $\max\{0,x\}$ and $\max\{0,-x\}$, $x \in R$, are continuous and hence Borel measurable, it follows that

$$f^+(.) = \max\{0,f(.)\},\ f^-(.) = \max\{0,-f(.)\}$$

are non-negative Borel measurable functions. Moreover, we obviously have

$$f = f^+ - f^-. \tag{1.3.1}$$

This representation is important because it means that without loss of generality we can limit our attention to non-negative Borel measurable functions. Thus the following theorem gives a full characterization of Borel measurable functions:

Theorem 1.3.4. A non-negative real function f on R^k is Borel measurable if and only if there is a non-decreasing sequence of simple functions ψ_n on R^k such that for each $x \in R^k$,

$$0 \le \psi_n(x) \le f(x),\ \lim_{n\to\infty} \psi_n(x) = f(x)$$

Proof: Take for given non-negative Borel measurable f and integers m with $1 \le m \le n2^n$,

$$\psi_n(x) = (m-1)/2^n \text{ if } (m-1)/2^n \le f(x) < m/2^n;$$
$$\psi_n(x) = n \text{ otherwise.}$$

Then the ψ_n's have all the required properties. Since by theorem 1.3.1 the simple functions ψ_n are Borel measurable, the limit is Borel measurable by theorem 1.3.2. Q.E.D.

Combining (1.3.1) and theorem 1.3.4 now yields:

Theorem 1.3.5. A real function on R^k is Borel measurable if and only if it is a (pointwise) limit of a sequence of simple functions on R^k.

Exercises:

1. Let f be a Borel measurable real function on R^k and let B be an arbitrary Borel set in R. Prove that the set $\{x \in R^k : f(x) \in B\}$ is a Borel set in R^k. (Hint: Use a similar argument as in the proof of theorem 1.1.1)
2. Let f and g be Borel measurable real functions on R. Prove that f+g and f-g are Borel measurable. (Hint: Use theorem 1.3.5)
3. Prove that the product of two simple functions on R^k is a simple function itself.
4. Let f and g be simple function on R , where $g(x) \neq 0$ for $x \in R$. Prove that f/g is a simple function.
5. Consider the real function

$f(x) = x$ if x is rational, $f(x) = -x$ if x is irrational.

Prove that f is Borel measurable.

1.4 *Mathematical expectation*

The theorems 1.3.4 and 1.3.5 can be used for defining the mathematical expectation of f(X), where f is a Borel measurable function on R^k and X is a random vector in R^k defined on a probability space (Ω,F,P), as a limit of mathematical expectations of simple functions. The latter expectations are defined as follows:

Definition 1.4.1. Let f(x) be the simple function on R^k as defined in definition 1.3.2. and let X a random vector in R^k defined on a probability space (Ω,F,P). Then the mathematical expectation of f(X) is defined by:

$$E\ f(X) = \Sigma_{j=1}^{n} b_j P(\{\omega \in \Omega : x(\omega) \in B_j\}) = \Sigma_{j=1}^{n} b_j \mu(B_j),$$

where μ is the probability measure on (R^k, B^k) induced by X.

For any non-negative Borel measurable function we define:

Definition 1.4.2. Let $f(x)$ be a non-negative Borel measurable function on R^k and let X be a random vector in R^k. Then

$$E\ f(X) = \sup E\ \psi(X),$$

where the supremum is taken over all the simple functions satisfying $0 \leq \psi(x) \leq f(x)$.

Using the representation (1.3.1) we now have:

Definition 1.4.3. Let $f(x)$ be a Borel measurable function on R^k and let X be a random vector in R^k. If

$$E\ f^+(X) < \infty \text{ and/or } E\ f^-(X) < \infty,$$

then $E\ f(X) = E\ f^+(X) - E\ f^-(X)$.

This is also denoted by the general integral with respect to the measure P of the probability space on which X is defined:

$$E\ f(X) = \int f(x(\omega))P(d\omega) = \int f(x)dP$$

(the latter integral is only a short-hand notation of the former) or by the integral with respect to the measure μ induced by X:

$$E\ f(X) = \int f(x)\mu(dx).$$

If both $E\ f^+(X) = \infty$ and $E\ f^-(X) = \infty$, the mathematical expectation and the corresponding integral are undefined. Moreover, for any set $\Lambda \in F$ we define:

$$\int_\Lambda f(x(\omega))P(d\omega) = \int y(\omega)P(d\omega),$$

where

$$y(\omega) = f(x(\omega)) \text{ if } \omega \in \Lambda,\ y(\omega) = 0 \text{ if } \omega \in \Omega\backslash\Lambda,$$

provided that the latter integral is defined. Similarly, we

define for arbitrary Borel sets B in R^k:

$$\int_B f(x)\mu(dx) = \int g(x)\mu(dx),$$

where:

$$g(x) = f(x) \text{ if } x \in B, \quad g(x) = 0 \text{ if } x \notin B$$

The mathematical expectation is often denoted by the classical Riemann-Stieltjes integral [see Rudin (1976)]:

$$E\ f(X) = \int f(x)dF(x), \tag{1.4.1}$$

where F is the joint distribution function of the random vector X. Strictly speaking this definition requires that the function f can be written as a pointwise limit of step functions [compare Rudin (1976, chapter 6)]. If so, then (1.4.1) is also a mathematical expectation according to the general definition, for step functions are obviously simple functions. For notational convenience we shall often use the notation (1.4.1) even if f is not a pointwise limit of step functions. Thus:

Definition 1.4.4. $\int f(x)dF(x) = \int f(x)\mu(dx)$, where μ is the probability measure defining F. Moreover, $\int_B f(x)dF(x) = \int_B f(x)\mu(dx)$ for Borel sets B in R^k.

Most of the properties of the general integral with respect to a probability measure go through for the clasical Riemann-Stieltjes integral with respect to the distribution function F, and in the following we shall assume that the reader is familiar with them. [Otherwise, see for example Chung (1974) and Royden (1968)]. Some of these properties are listed below since they are frequently used in this book. Assuming that the integrals and mathematical expectations involved are well defined, these properties are:

- Let X and Y be random variables defined on the probability space (Ω,F,P). Let Λ and Λ_n be sets in F. Let α and β be real numbers. We have:

$$\int_\Lambda (\alpha x + \beta y)dP = \alpha\int_\Lambda x dP + \beta\int_\Lambda y dP.$$

If the Λ_n's are disjoint, then $\int_{\cup\Lambda_n} x dP = \Sigma_n \int_{\Lambda_n} x dP$.

If $x(\omega) \geq 0$ for every $\omega \in \Lambda$, then $\int_\Lambda x dP \geq 0$.

If $x(\omega) \leq y(\omega)$ for every $\omega \in \Lambda$, then $\int_\Lambda x dP \leq \int_\Lambda y dP$.

$|\int_\Lambda x dP| \leq \int_\Lambda |x| dP$.

These properties are not hard to verify from definitions 1.4.2 and 1.4.3. See also Kolmogorov and Fomin (1961, chapter VII), for related properties of the Lebesgue integral.

- If $X_1, X_2, \ldots, X_n, \ldots$ are independent random variables or -vectors and if $f_1(x), f_2(x), \ldots, f_n(x), \ldots$ are conformable Borel measurable functions, then:

$$E \prod_j f_j(X_j) = \prod_j [E\, f_j(X_j)].$$

- *Chebishev's inequality*. If φ is a Borel measurable function on R such that $\varphi(x)$ is positive and monotonic increasing on $(0,\infty)$ and $\varphi(x)=\varphi(-x)$, then for every r.v. X and every $\delta > 0$ we have $P(|X| > \delta) \leq E\,\varphi(X)/\varphi(\delta)$.

- *Holder's inequality*. Let X and Y be r.v. Then for p>1 and 1/p+1/q=1,

$$|E(X \cdot Y)| \leq E|X \cdot Y| \leq (E|X|^p)^{1/p} (E|Y|^q)^{1/q}.$$

(For p = 2 we have the well-known *Cauchy-Schwarz-inequality*).

- *Minkowski's inequality*. Let X and Y be random variables. If for some $p \in [1,\infty]$, $E|X|^p < \infty$ and $E|Y|^p < \infty$, then

$$(E|X + Y|^p)^{1/p} \leq (E|X|^p)^{1/p} + (E|Y|^p)^{1/p}.$$

- *Liapounov's inequality*. Let X be a r.v. Then for $1 \leq p \leq q \leq \infty$,

$$(E|X|^p)^{1/p} \leq (E|X|^q)^{1/q}.$$

(This follows straightforwardly from Holder's inequality by putting Y = 1 and replacing X by $|X|^p$ and p by q/p.)

- *Jensen's inequality*. Let φ be a convex real function on R and let X be a random variable such that $E|X| < \infty$, $E|\varphi(X)| < \infty$. Then

$$\varphi(E(X)) \leq E[\varphi(X)].$$

(N.B.: Convex real functions are Borel measurable).

Since the mean of a finite number of non-random variables in R may be considered as a mathematical expectation, it follows from Holder's inequality that for real numbers x_j, y_j, $p>1$, $1/p+1/q=1$:

$$|(1/n)\Sigma_{j=1}^{n} x_j y_j| \leq \{(1/n)\Sigma_{j=1}^{n}|x_j|^p\}^{1/p}\{(1/n)\Sigma_{j=1}^{n}|y_j|^q\}^{1/q}, \tag{1.4.2}$$

and consequently, taking $y_j = 1$,

$$|\Sigma_{j=1}^{n} x_j|^p \leq n^{p-1}\Sigma_{j=1}^{n}|x_j|^p \ , \quad p \geq 1. \tag{1.4.3}$$

The latter inequality is a sharpening of the following trivial but useful inequality:

$$|\Sigma_{j=1}^{n} x_j|^p \leq n^p(\max|x_j|)^p \leq n^p\Sigma_{j=1}^{n}|x_j|^p, \quad p > 0. \tag{1.4.4}$$

Moreover, by Minkowski's inequality we have for $1 \leq p < \infty$,

$$(1/n)\{\Sigma_{j=1}^{n}|x_j+y_j|^p\}^{1/p} \leq \{(1/n)\Sigma_{j=1}^{n}|x_j|^{1/p} + \{(1/n)\Sigma_{j=1}^{n}|y_j|^p\}^{1/p} \tag{1.4.5}$$

Finally we shall also use (in chapter 2) the following result.

Theorem 1.4.1. Let X be a r.v. on (Ω,F,P) such that $E|X| < \infty$. Let (Λ_n) be a sequence of sets in F such that $\lim_{n\to\infty}P(\Lambda_n) = 0$. Then:

$$\lim_{n\to\infty}\int_{\Lambda_n}|x(\omega)|P(d\omega) = 0.$$

Proof: Let for $n = 1,2,\ldots$

$$\Delta_n = \{\omega \in \Omega : |x(\omega)| \geq n\}; \quad \Gamma_n = \{\omega \in \Omega : n \leq |x(\omega)| < n+1\}.$$

From the fact that

$$\int |x(\omega)| P(d\omega) = \int_{\cup_{n=0}^{\infty}} |x(\omega)| dP(d\omega) = \Sigma_{n=0}^{\infty} \int_{\Gamma_n} |x(\omega)| P(d\omega) < \infty$$

we conclude that

$$\int_{\Delta_n} |x(\omega)| P(d\omega) = \Sigma_{k=n}^{\infty} \int_{\Gamma_k} |x(\omega)| P(d\omega) \to 0 \text{ as } n \to \infty.$$

The theorem now follows from:

$$\int_{\Lambda_n} |x(\omega)| P(d\omega) = \int_{\Lambda_n \cap \Delta_k} |x(\omega)| P(d\omega) + \int_{\Lambda_n \cap \Delta_k^c} |x(\omega)| P(d\omega)$$

$$\leq \int_{\Delta_k} |x(\omega)| P(d\omega) + \int_{\Lambda_n \cap \Delta_k^c} kP(d\omega)$$

$$\leq \int_{\Delta_k} |x(\omega)| P(d\omega) + kP(\Lambda_n)$$

by letting first $n \to \infty$ and then $k \to \infty$. Q.E.D.

Exercises:

1. Prove the first set of properties for finite-valued random variables.
2. Prove the second property for simple functions.
3. Prove the Cauchy-Schwarz inequality.
4. Prove Jensen's inequality for the case that X is discretely distributed, i.e.,

$$P(X = x_j) = p_j > 0 \; ; \; j=1,\ldots,n \; ; \; \Sigma_{j=1}^{n} p_j = 1.$$

1.5 *Characteristic functions*

A special and very useful mathematical expectation is the so-called characteristic function:

Definition 1.5.1. Let X be a random vector in R^k with distribution function F. The characteristic function of this distribution is the following complex valued function $\varphi(t)$ on R^k:

$$\varphi(t) = E \exp(i \cdot t'X) = E \cos(t'X) + i \cdot E \sin(t'X),$$

where $t \in R^k$ is a non-random vector.

Distributions are fully determined by their characteristic functions: *distributions are equal if and only if their characteristic functions are equal*. For absolutely continuous distributions this one-to-one correspondence can even be stated explicitly in an inversion formula. We recall that a distribution function F on R^k is absolutely continuous if there exists a non-negative function f on R^k, called the density of F, satisfying:

$$F(t) = \int_{(u \leq t)} f(u)du$$

(where u and t are vectors in R^k). For such distributions we have the following results.

Theorem 1.5.1. (Inversion formula for characteristic functions) Let F be a distribution function on R^k and let φ be its characteristic function. If φ is absolutely integrable then F is absolutely continuous and its density F satisfies:

$$f(x) = (1/2\pi)^k \int \exp(-i \cdot t'x)\varphi(t)dt$$

$$= (1/2\pi)^k \int \cos(t'x)\mathrm{Re}[\varphi(t)]dt + (1/2\pi)^k \int \sin(t'x)\mathrm{Im}[\varphi(t)]dt.$$

Proof: Cf. Wilks(1963).

Moreover, we have:

Theorem 1.5.2. A distribution function F is continuously differentiable up to the m-th order if its characteristic function $\varphi(t)$ satisfies:

$$\int |t|^m |\varphi(t)| dt < \infty. \qquad (1.5.1)$$

This can be proved by generalizing theorem 1.5.1 to derivatives, realizing that by condition (1.5.1) we may differen-

tiate m-times under the integral.

Exercises:

1. The distribution F on R has characteristic function $\exp(i \cdot t)$. Determine F.
2. The distribution F on R has characteristic function

$$\exp(-|t|).$$

Determine the density of F.

3. Let X be a r.v. with characteristic function $\varphi(t)$. Show that for m=1,2,..,

$$E\, X^m = [(d/dt)^m \varphi(t)|_{t=0}]/i^m,$$

provided $E\,|X|^m < \infty$.

1.6 *Random functions*

A random function is a function which is a r.v. for each value of its argument. Usually random functions occur as a function of both random variables and parameters, for example the sum of squares of a regression model. Their definition is similar to that of random variables:

Definition 1.6.1. Let (Ω,F,P) be a probability space and let Θ be a subset of R^k. The real function $f(.) = f(.,\omega)$ on $\Theta\times\Omega$ is called a (real) random function on Θ if for every $t \in R$ and every $\theta_0 \in \Theta$, $\{\omega \in \Omega : f(\theta_0,\omega) \leq t\} \in F$.

However, dealing with random functions one must be aware of some pitfalls. First, if f(.) is a random function on an uncountable subset Θ of a Euclidean space, then $\sup_{\theta\in\Theta} f(\theta)$ and $\inf_{\theta\in\Theta} f(\theta)$ are not automatically random variables, because

$$\{\omega \in \Omega : \inf_{\theta\in\Theta} f(\theta,\omega) \leq t\} = \cup_{\theta\in\Theta}\{\omega \in \Omega : f(\theta,\omega) \leq t\}$$

and

$$\{\omega \in \Omega : \sup_{\theta\in\Theta} f(\theta,\omega) \leq t\} = \cap_{\theta\in\Theta}\{\omega \in \Omega : f(\theta,\omega) \leq t\}$$

are then *uncountable* unions and intersections, respectively, of members of the Borel field F and therefore not necessarily members of F themselves. Another pitfall is that if θ is a random vector in an uncountable subset Θ of a Euclidean space and if $f(.)$ is a random function on Θ , then $f(\theta)$ is not necessarily a random variable, because:

$$\{\omega \in \Omega: f(\theta(\omega),\omega) \leq t\}$$

$$= \cup_{\theta_1 \in \Theta}[\{\omega \in \Omega: f(\theta_1,\omega) \leq t\} \cap \{\omega \in \Omega: \theta(\omega) = \theta_1\}]$$

is an uncountable union of members of F. These problems can be ovecome if we assume that the random function $f(.)$ is *separable* [see Gihman and Skorohod (1974), chapter III, section 2]. However, in this study we shall only deal with random functions of the type $f(.) = \varphi(.,X)$, where φ is a Borel measurable real function on $\Theta \times R^m$ with Θ a compact Borel set in R^k and X is a random vector in R^m. If in addition $\varphi(.,x)$ is for each $x \in R^m$ a continuous function on Θ we do not need the separability concept, due to the following theorem:

Theorem 1.6.1. Let Θ be a compact set in R^k and let $\varphi(\theta,x)$ be a Borel measurable real function on $\Theta \times R^m$ which is continuous in θ for each $x \in R^m$. There exists a mapping $\theta(x)$ from R^k into Θ with Borel measurable components such that

$$\varphi(\theta(x),x) = \sup_{\theta \in \Theta} \varphi(\theta,x).$$

Consequently the supremum involved is a Borel measurable real function on R^m. Moreover, if $\varphi(\theta,x)$ is continuous on $\Theta \times R^m$ then $\sup_{\theta \in \Theta} \varphi(\theta,x)$ is a continuous function on R^m.

Proof: Jennrich (1969).

(N.B.: Compact sets in R^k are Borel sets).

Of course, a similar result holds for the 'inf' case. The condition in theorem 1.6.1 that the set Θ is compact (hence bounded) is not strictly necessary for the Borel measurability of $\sup_{\theta \in \Theta} \varphi(\theta,x)$ and $\inf_{\theta \in \Theta} \varphi(\theta,x)$, provided that

$$\Theta_n = \Theta \cap \{\theta \in R^k : |\theta| \leq n\}$$

is compact for each n. Then we have:

$$\sup_{\theta\in\Theta}\varphi(\theta,x) = \lim_{n\to\infty}\sup_{\theta\in\Theta_n}\varphi(\theta,x),$$

$$\inf_{\theta\in\Theta}\varphi(\theta,x) = \lim_{n\to\infty}\inf_{\theta\in\Theta_n}\varphi(\theta,x),$$

which by theorems 1.3.2 and 1.6.1 are Borel measurable functions.

Let us return to more general random functions. The properties of a random function f(.) may differ for different ω in Ω. For two points ω_1 and ω_2 in Ω it is for example possible that $f(\tau,\omega_1)$ is continuous and $f(\tau,\omega_2)$ is discontinuous at the same τ.

In this study we shall always consider properties of random functions holding almost surely, which means that a property of $f(.) = f(.,\omega)$ holds for all ω in a set $E \in F$ with $P(E) = 1$. Thus for example the statement: "f(.) is a.s. continuous on Θ" means that there is a null set N such that $f(.,\omega)$ is continuous on Θ for all $\omega \in \Omega\backslash N$.

Exercises:

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

$$f(\theta) = 1 \text{ if } X \geq \theta, \quad f(\theta) = 0 \text{ if } X < \theta.$$

Is $f(\theta)$ a random function on [0,1]?

2. Let X be N(0,1) distributed and let for $\theta \in R$,

$$f(\theta) = X \text{ if } \theta \text{ is rational}, \quad f(\theta) = -X \text{ if } \theta \text{ is irrational}.$$

Prove that $f(\theta)$ is a.s. continuous.

3. Let for $x \in R$ and $\theta \in R$,

$$\varphi(x,\theta) = 1-(\theta-2)^2 \text{ if } 1 \leq \theta \leq 3 \text{ and } x > 0,$$

$$\varphi(x,\theta) = [1-(\theta-2)^2]/[1+x^2] \text{ if } 1 \leq \theta \leq 3 \text{ and } x \leq 0,$$

$\varphi(x,\theta) = 1-(\theta+2)^2$ for $-3 \leq \theta \leq -1$ and $x \leq 0$,

$\varphi(x,\theta) = [1-(\theta+2)^2]/[1+x^2]$ if $-3 \leq \theta \leq -1$ and $x > 0$,

$\varphi(x,\theta) = 0$ elsewhere,

and let $\hat{\theta}(x)$ be such that

$$\varphi(\hat{\theta}(x),x) = \sup_{\theta\in[-3,3]}\varphi(\theta,x).$$

- Is φ continuous?
- Determine $\hat{\theta}(x)$.
- What is your conclusion regarding the second part of theorem 1.6.1 ?

References:

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

Gihman,I.I. and A.V.Skorohod (1974), *The Theory of Stochastic Processes I*, Berlin: Springer Verlag

Hogg,R.V. and A.T.Craig (1978), *Introduction to Mathematical Statistics*, London: MacMillan

Jennrich,R.I. (1969), "Asymptotic Properties of Nonlinear Least Squares Estimators", *Annals of Mathematical Statistics* 40, 633-643

Kolmogorov,A.N. and S.V.Fomin (1957), *Elements of the Theory of Functions and Functional Analysis, Vol.1: Metric and Normed Spaces*, Rochester: Graylock Press

Royden,H.L. (1968), *Real Analysis*, London: Macmillan

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

Wilks,S.S. (1963), *Mathematical Statistics*, Princeton: Princeton University Press