# 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 review the basic elements of this theory as needed in what follows, to make this book almost self-contained. For a more detailed treatment, see for example Billingsley (1979) and Parthasarathy (1977). However, we do assume the reader has a good knowledge of probability and statistics at an intermediate level, for example 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 $\{\Omega,\mathfrak{F},P\}$ consisting of:

— An abstract non-empty set $\Omega$, called the *sample space*. We do not impose any conditions on this set.

— A non-empty collection $\mathfrak{F}$ of subsets of $\Omega$, having the following two properties:

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

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

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

These two properties make $\mathfrak{F}$, by definition, a *Borel field* of subsets of $\Omega$. (Following Chung [1974], the term "Borel field" has the same meaning as the term "$\sigma$-Algebra" used by other authors.)

— A *probability measure* P on $\{\Omega,\mathfrak{F}\}$. This is a real-valued set function on $\mathfrak{F}$ such that:

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

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

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

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

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

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

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

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 $\Omega$ denoted by:

$$X = x(.)$$

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

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

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 shorthand 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:

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

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

$$E_n \in \mathfrak{F}, n = 1,2,\ldots \Rightarrow \cap_n E_n \in \mathfrak{F}$$

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

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

$E_n \subset E_{n+1}$, $E = \cup_n E_n \Rightarrow P(E_n) \to P(E)$ as $n \to \infty$,
$E_n \supset E_{n+1}$, $E = \cap_n E_n \Rightarrow P(E_n) \to P(E)$ as $n \to \infty$,
$P(\cup_n E_n) \leq \sum_n P(E_n)$,

where all sets involved are members of $\mathfrak{F}$. The distribution function F(t) is *right continuous*:

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

as is easily verified, and it satisfies

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

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 parts, respectively.

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

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

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

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

Let $\mathfrak{C}$ be the collection of subsets of $R^k$ of the type

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

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

$$\mathfrak{B}^k = \cap_{a\in A}\mathfrak{G}_a.$$

Since each $\mathfrak{G}_a$ is a Borel field, it follows that this collection $\mathfrak{B}^k$ is a Borel field of subsets of $R^k$ and since $\mathfrak{C}$ is contained in each $\mathfrak{G}_a$ it follows that $\mathfrak{C}$

$\subset B^k$. We shall say that the Borel field $\mathfrak{B}^k$ is the *minimal Borel field* containing the collection $\mathfrak{C}$, and for this particular collection $\mathfrak{C}$ it is called the *Euclidean Borel field*. Summarizing:

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

*Definition 1.1.2.* Let $\mathfrak{C}$ be the collection of subsets of $R^k$ of the type

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

The minimal Borel field $\mathfrak{B}^k$ containing this collection is called the *Euclidean Borel field* (also called the Borel $\sigma$-Algebra) and the members of $\mathfrak{B}^k$ 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, pp. 63-64) for an example.

We show now that for any Borel set B and any r.v. X on $\{\Omega,\mathfrak{F},P\}$, (1.1.6) is satisfied. Let $\mathfrak{D}$ be the collection of all Borel sets B such that (1.1.6) is satisfied. Then $\mathfrak{D} \subset \mathfrak{B}^k$. If $\mathfrak{D}$ is also a Borel field then $\mathfrak{B}^k \subset \mathfrak{D}$ (and hence $\mathfrak{B}^k = \mathfrak{D}$) because $\mathfrak{B}^k$ is the minimal Borel field containing the collection $\mathfrak{C}$ in definition 1.1.2, whereas obviously the collection $\mathfrak{D}$ contains $\mathfrak{C}$. So it suffices to prove that $\mathfrak{D}$ 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 $\{\Omega,\mathfrak{F},P\}$ and any Borel set B in $R^k$ we have $\{\omega \in \Omega : x(\omega) \in B\} \in \mathfrak{F}$. The collection $\mathfrak{F}\{X\}$ 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 $\mathfrak{F}$.

Consequently the definition (1.1.2) is meaningful for Borel sets. In fact, by defining a measure $\mu$ on the Euclidean Borel field $\mathfrak{B}^k$ 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 $\{R^k,\mathfrak{B}^k\}$. This probability measure $\mu$ is often referred to as the probability measure *induced* by X. Moreover, the Borel field $\mathfrak{F}\{X\}$ 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 $\{\Omega,\mathfrak{F},P\}$. 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:

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

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 $\mu$ 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 $\mu$. 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 $\mu$ on $\{R^k, \mathfrak{B}^k\}$ defining F.

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

Theorem 1.1.2 implies that there is a one-to-one correspondence between a distribution function F and its defining probability measure $\mu$. We shall employ this result later in defining mathematical expectations.

*Exercises*

1. Show that (1.1.1) and (1.1.2) imply $\cap_j E_j \in \mathfrak{F}$.
2. Consider the collection $\mathfrak{F}$ of subintervals of [0,1] with rational-valued endpoints, together with their complements and *finite* unions and intersections. Show that $\mathfrak{F}$ is not a Borel field. (Hint: Use the fact that irrational numbers can be written as limits of rational numbers.)
3. Let $\mathfrak{F}$ be a Borel field of subsets of $\Omega$. Let $A \in \mathfrak{F}$ and let $\mathfrak{G}_A$ be the collection of all subsets of the type $A\cap B$, where $B \in \mathfrak{F}$. Prove that $\mathfrak{G}_A$ is a Borel field of subsets of A.
4. Let $\Omega = \{1,2,3,4,5\}$ and let $\mathfrak{C}$ be the collection consisting of the two subsets {2}, {4}. Prove that the minimal Borel field containing $\mathfrak{C}$ consists of the following sets: $\Omega$, $\emptyset$, {2}, {4}, {2}∪{1,3,5}, {4}∪{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. Is the set Q of rational numbers a Borel set in R?
7. Let $\mathfrak{C}$ 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 $\mathfrak{C}$ 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 $\Omega$ be the set of all pairs $(x_1,x_2)$ with $x_1 \in S$, $x_2 \in S$, $x_1 < x_2$, let $\mathfrak{F}$ be the Borel field of all subsets of $\Omega$ and let P be a probability measure on $\{\Omega, \mathfrak{F}\}$. Define the random variable Y as follows:

    $y(\omega) = 1$ if $\omega = (x_1,x_2)$ with $x_1 + x_2$ odd.
    $y(\omega) = 0$ if $\omega = (x_1,x_2)$ with $x_1 + x_2$ even.

    — Determine the Borel field $\mathfrak{F}\{Y\}$ generated by Y.
    — All points in $\Omega$ 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 $\Omega$ and a Borel field $\mathfrak{F}$ of subsets of $\Omega$ such that together:
    — Y is a random vector
    — $\mathfrak{F}$ 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 $\{\Omega_1,\mathfrak{F}_1,P_1\},\{\Omega_2,\mathfrak{F}_2,P_2\},\ldots$ respectively. It is possible to construct a new probability space, $\{\Omega,\mathfrak{F},P\}$, say, such that the $X_j$'s can be regarded as *independent* random variables on $\{\Omega,\mathfrak{F},P\}$ (see Chung [1974, section 3.3]). Independence means:

> *Definition 1.2.1* Let $X_1,X_2,\ldots$ be random vectors defined on a common probability space. This sequence of random vectors is called *(totally) independent* if for any conformable sequence $(B_j)$ of Borel sets,

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

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). Then $X_j$ is a random variable defined on the probability space $\{\Omega_j,\mathfrak{F}_j,P_j\}$, where

$$\Omega_j = \{H,T\},\ \mathfrak{F}_j = \{\emptyset,\{H\},\{T\},\{H,T\}\}$$
$$P_j(\emptyset) = 0,\ P_j(\{H,T\}) = 1,\ P_j(\{H\}) = P_j(\{T\}) = \tfrac{1}{2}.$$

Now let $\Omega$ be the set of all one-sided infinite sequences of H and T, for example, let $\omega = (H,H,T,T,H,T,T,T,H,H,H,T,\ldots)$ be such an element of $\Omega$. We could take as $\mathfrak{F}$ the collection of all subsets of $\Omega$, including the empty set. Now $x_j$ can also be defined on $\{\Omega,\mathfrak{F}\}$, as follows. Let

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

Each set $E \in \mathfrak{F}$ 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

$$\delta(E_{j,i}) = ½ \text{ if } E_{j,i} \neq \emptyset,\ \delta(E_{j,i}) = 0 \text{ if } E_{j,i} = \emptyset,\ i = 1,2,$$
$$P(E) = \Pi_j[\delta(E_{j,1}) + \delta(E_{j,2})].$$

Then $(X_j)$ is a sequence of independent random variables defined on the common probability space $(\Omega,\mathfrak{F},P)$.

## 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

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

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

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

Functions satisfying the latter condition are called *Borel measurable*.

Now consider a real function $f(x_1,\dots,x_k)$ on $R^k$ and r.v.'s $X_1,\dots,X_k$ on $\{\Omega,\mathfrak{F},P\}$. If for every $t \in R$ the set

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

is a Borel set in $R^k$ then

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

for every $t \in R$ and hence $f(X_1,\dots,X_k)$ 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 $b_1,\dots,b_n$ and Borel sets $B_j$, $j = 1,2,\dots,n$ with $B_{j_1} \cap B_{j_2} = \oslash$ if $j_1 \neq j_2$, such that

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

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

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

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 $f_1,f_2,\dots$ be a sequence of Borel measurable functions on $R^k$. Then the functions $\max\{f_1,\dots,f_n\}$, $\min\{f_1,\dots,f_n\}$, $\sup_n f_n$, $\inf_n f_n$, $\limsup_{n\to\infty} f_n$ 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

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

Then

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

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

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

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 $X_1, X_2, X_3, ..$ are random variables defined on a common probability space, then so are $\max\{X_1,...,X_n\}$, $\min\{X_1,...,X_n\}$, $\sup_n X_n$, $\inf_n X_n$, $\limsup_{n\to\infty} X_n$ 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*: We prove the theorem for the univariate case $k = 1$ only. Let

$$f_n(x) = f(x) \text{ if } -n < x \le n;\ f_n(x) = 0 \text{ elsewhere},$$
$$f_{nm}(x) = \sum_{j=0}^{m-1} \sup_{x\in B_{jnm}}(x)\cdot I(x\in B_{jnm}),$$

where

$$B_{jnm} = (-n + 2n\cdot j/m,\ -n + 2n(j+1)/m].$$

Then the $f_{nm}(x)$'s are simple functions. Since by the continuity of f, $f_n(x) = \lim_{m\to\infty} f_{nm}(x)$, it follows from theorem 1.3.2 that the $f_n(x)$'s are Borel measurable. The theorem follows now from theorem 1.3.2 and the fact that

$$f(x) = \lim_{n\to\infty} f_n(x) = \limsup_{n\to\infty} f_n(x) = \liminf_{n\to\infty} f_n(x).$$ 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 $\varphi_n$ on $R^k$ such that for each $x \in R^k$,

$$0 \leq \varphi_n(x) \leq f(x), \lim_{n\to\infty}\varphi_n(x) = f(x)$$

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

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

Then the $\varphi_n$'s have all the required properties. Since by theorem 1.3.1 the simple functions $\varphi_n$ 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 to that 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 functions on R, where $g(x) \neq 0$ for $x \in R$. Prove that f/g is a simple function.
5. Consider the real function

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

   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