Chapter 2

Measurable Functions and Convergence

1 Mappings and $\sigma$-Fields

Notation 1.1 (Inverse images) Suppose $X$ denotes a function mapping some set $\Omega$ into the extended real line $\bar{R} \equiv R \cup \{\pm \infty\}$; we denote this by $X : \Omega \to \bar{R}$. Let $X^+$ and $X^-$ denote the positive part and the negative part of $X$, respectively:

\begin{align*}
(1) \quad X^+(\omega) &\equiv \begin{cases} X(\omega) & \text{if } X(\omega) \geq 0, \\
0 & \text{else}, \end{cases} \\
(2) \quad X^-(\omega) &\equiv \begin{cases} -X(\omega) & \text{if } X(\omega) \leq 0, \\
0 & \text{else}. \end{cases}
\end{align*}

Note that

\begin{align*}
(3) \quad X = X^+ - X^- \quad \text{and} \quad |X| = X^+ + X^- = X + 2X^- = 2X^+ - X.
\end{align*}

We also use the following notation:

\begin{align*}
(4) \quad [X = r] &\equiv X^{-1}(r) \equiv \{\omega : X(\omega) = r\} \quad \text{for all real } r, \\
(5) \quad [X \in B] &\equiv X^{-1}(B) \equiv \{\omega : X(\omega) \in B\} \quad \text{for all Borel sets } B, \\
(6) \quad X^{-1}(\mathcal{B}) &\equiv \{X^{-1}(B) : B \in \mathcal{B}\}. 
\end{align*}

We call these the inverse images of $r, B$, and $\mathcal{B}$, respectively. We let

\begin{align*}
(7) \quad \bar{\mathcal{B}} &\equiv \sigma[\mathcal{B}, \{+\infty\}, \{-\infty\}].
\end{align*}

Inverse images are also well-defined when $X : \Omega \to \Omega'$ for arbitrary sets $\Omega$ and $\Omega'$. \hfill \square

For $A, B \in \Omega$ we define $A \Delta B \equiv AB^c \cup A^cB$ and $A \setminus B \equiv AB^c$. There is use for the notation
CHAPTER 2. MEASURABLE FUNCTIONS AND CONVERGENCE

(8) \[\|X\| \equiv \sup_{\omega \in \Omega} |X(\omega)|,\]
and we will also reintroduce this sup norm in other contexts below.

**Proposition 1.1 (Basics of inverse images)** Let \( X : \Omega \to \Omega' \) and \( Y : \Omega' \to \Omega'' \). Let \( T \) denote an arbitrary index set. Then for all \( A, B, A_t \subset \Omega' \) we have

\[
\begin{align*}
(9) \quad & X^{-1}(B^c) = [X^{-1}(B)]^c, \quad X^{-1}(A \setminus B) = X^{-1}(A) \setminus X^{-1}(B), \\
(10) \quad & X^{-1}\left(\bigcup_{t \in T} A_t\right) = \bigcup_{t \in T} X^{-1}(A_t), \quad X^{-1}\left(\bigcap_{t \in T} A_t\right) = \bigcap_{t \in T} X^{-1}(A_t).
\end{align*}
\]

For all sets \( A \subset \Omega'' \), the composition \( Y \circ X \) satisfies

\[
(11) \quad (Y \circ X)^{-1}(A) = X^{-1}(Y^{-1}(A)) = X^{-1} \circ Y^{-1}(A).
\]

**Proof.** Trivial. \( \square \)

**Proposition 1.2 (Preservation of \( \sigma \)-fields)** Let \( X : \Omega \to \Omega' \). Then:

\[
\begin{align*}
(12) \quad & \mathcal{A} \equiv X^{-1}(a \text{-field of subsets of } \Omega') = (a \text{-field of subsets of } \Omega), \\
(13) \quad & X^{-1}(\sigma(C')) = \sigma[X^{-1}(C')] \quad \text{for any collection } C' \text{ of subsets of } \Omega', \\
(14) \quad & \mathcal{A}' \equiv \{A' : X^{-1}(A') \in (a \text{-field of subsets of } \Omega)\} \\
& \quad = (a \text{-field of subsets of } \Omega').
\end{align*}
\]

**Proof.** Now, (12) is trivial from proposition 1.1. Consider (14). Now:

\[
\begin{align*}
(a) \quad & A' \in \mathcal{A}' \implies X^{-1}(A') \in \mathcal{A} \implies X^{-1}(A') = [X^{-1}(A')]^c \in \mathcal{A} \implies A'^c \in \mathcal{A}', \\
(b) \quad & A_n's \in \mathcal{A}' \implies X^{-1}(\bigcup_n A'_n) = \bigcup_n X^{-1}(A'_n) \in \mathcal{A} \implies \bigcup_n A'_n \in \mathcal{A}'.
\end{align*}
\]

This gives (14). Consider (13). Using (12) gives

\[
\begin{align*}
(c) \quad & X^{-1}(\sigma[C']) = (a \text{-field containing } X^{-1}(C')) \supset \sigma[X^{-1}(C')].
\end{align*}
\]

Then (14) shows that

\[
\begin{align*}
(d) \quad & \mathcal{A}' \equiv \{A' : X^{-1}(A') \in \sigma[X^{-1}(C')]\} = (a \text{-field containing } C') \supset \sigma[C'], \text{ so that (using first } \sigma[C'] \subset \mathcal{A}' \text{ from (d), and then the definition of } \mathcal{A}' \text{ in (d)}) \\
(e) \quad & X^{-1}(\sigma[C']) \subset X^{-1}(\mathcal{A}') \subset \sigma[X^{-1}(C')].
\end{align*}
\]

Combining (c) and (e) gives (13). [We will apply (13) below to obtain (2.2.6).] \( \square \)

Roughly, using (12) we will restrict \( X \) so that \( \mathcal{F}(X) \equiv X^{-1}(\mathcal{B}) \subset \mathcal{A} \) for our original \((\Omega, \mathcal{A}, \mu)\), so that we can then “induce” a measure on \((\mathcal{R, B})\). Or, (14) tells us that the collection \( \mathcal{A}' \) is such that we can always induce a measure on \((\Omega', \mathcal{A}')\). We do this in the next section. First, we generalize our definition of Borel sets to \( n \) dimensions.
Example 1.1 (Euclidean space) Let

\[ R_n \equiv R \times \cdots \times R \equiv \{(r_1, \ldots, r_n) : \text{each } r_i \text{ is in } R\} \]

Let \( U_n \) denote all open subsets of \( R_n \), in the usual Euclidean metric. Then

(15) \[ B_n \equiv \sigma[U_n] \text{ is called the class of Borel sets of } R_n. \]

Following the usual notation, \( B_1 \times \cdots \times B_n \equiv \{(b_1, \ldots, b_n) : b_1 \in B_1, \ldots, b_n \in B_n\} \)

Now let

(16) \[ \prod_{i=1}^n B \equiv B \times \cdots \times B \equiv \sigma[\{B_1 \times \cdots \times B_n : \text{all } B_i \text{ are in } B\}]. \]

Now consider

(17) \[ \sigma\{(-\infty, r_1] \times \cdots \times (-\infty, r_n] : \text{all } r_i \text{ are in } R\}. \]

Note that the three \( \sigma \)-fields of (15), (16), and (17) are equal. Just observe that each of these three classes generates the generators of the other two classes, and apply exercise 1.1.1. (Surely, we can define a generalization of area \( \lambda_2 \) on \( (R_2, \mathcal{B}_2) \) by beginning with \( \lambda_2(B_1 \times B_2) = \lambda(B_1) \times \lambda(B_2) \) for all \( B_1 \) and \( B_2 \) in \( \mathcal{B} \), and then extending to all sets in \( \mathcal{B} \). We will do this in theorem 5.1.1, and we will call it Lebesgue measure on two-dimensional Euclidean space. This clearly extends to \( \lambda_n \) on \( (R_n, \mathcal{B}_n) \).) \[ \square \]
2 Measurable Functions

We seek a large usable class of functions that is closed under passage to the limit. This is the fundamental property of the class of measurable functions. Propositions 2.2 and 2.3 below will show that the class of measurable functions is also closed under all of the standard mathematical operations. Thus, this class is sufficient for our needs.

Definition 2.1 (Simple functions, etc.) Let the measure space \((\Omega, \mathcal{A}, \mu)\) be given and fixed throughout our discussion. Consider the following classes of functions. The indicator function \(1_A(\cdot)\) of the set \(A \subset \Omega\) is defined by

\[
1_A(\omega) \equiv \begin{cases} 
1 & \text{if } \omega \in A, \\
0 & \text{else}.
\end{cases}
\]

A simple function is of the form

\[
X(\omega) \equiv \sum_{i=1}^{n} x_i 1_{A_i}(\omega) \quad \text{for} \quad \sum_{i=1}^{n} A_i = \Omega \quad \text{with all} \quad A_i \in \mathcal{A}, \quad \text{and} \quad x_i \in \mathbb{R}.
\]

An elementary function is of the form

\[
X(\omega) \equiv \sum_{i=1}^{\infty} x_i 1_{A_i}(\omega) \quad \text{for} \quad \sum_{i=1}^{\infty} A_i = \Omega \quad \text{with all} \quad A_i \in \mathcal{A}, \quad \text{and} \quad x_i \in \mathbb{R}.
\]

Definition 2.2 (Measurability) Suppose that \(X : \Omega \to \Omega'\), where \((\Omega, \mathcal{A})\) and \((\Omega', \mathcal{A}')\) are both measurable spaces. We then say that \(X\) is \(\mathcal{A}'-\mathcal{A}\)-measurable if \(X^{-1}(\mathcal{A}') \subset \mathcal{A}\). We also denote this by writing either

\[
X : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}') \quad \text{or} \quad X : (\Omega, \mathcal{A}, \mu) \to (\Omega', \mathcal{A}')
\]

(or even \(X : (\Omega, \mathcal{A}, \mu) \to (\Omega', \mathcal{A}', \mu')\) for the measure \(\mu'\) “induced” on \((\Omega', \mathcal{A}')\) by the mapping \(X\), as will soon be defined). In the special case \(X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})\), we simply call \(X\) measurable; and in this special case we let \(\mathcal{F}(X) \equiv X^{-1}(\mathcal{B})\) denote the sub \(\sigma\)-field of \(\mathcal{A}\) generated by \(X\).

Proposition 2.1 (Measurability criteria) Let \(X : \Omega \to \mathbb{R}\). Suppose \(\sigma[\mathcal{C}] = \mathcal{B}\). Then measurability can be characterized by either of the following:

\[
\begin{align*}
(5) \quad & X \text{ is measurable if and only if } X^{-1}(\mathcal{C}) \subset \mathcal{A}. \\
(6) \quad & X \text{ is measurable if and only if } X^{-1}([-\infty, x]) \in \mathcal{A} \text{ for all } x \in \mathbb{R}.
\end{align*}
\]

Note that we could replace \([-\infty, x]\) by any one of \([-\infty, x), [x, +\infty], \) or \((x, +\infty]\).

Proof. Consider (5). Let \(X^{-1}(\mathcal{C}) \subset \mathcal{A}\). Then

\[
(a) \quad X^{-1}(\mathcal{B}) = X^{-1}(\sigma[\mathcal{C}]) = \sigma[X^{-1}(\mathcal{C})] \quad \text{by proposition 2.1.2}
\]

(b) \(\subset \mathcal{A}\) since \(X^{-1}(\mathcal{C}) \subset \mathcal{A}\), and \(\mathcal{A}\) is a \(\sigma\)-field.

The other direction is trivial. Thus (5) holds. To demonstrate (6), we need to show that \(\mathcal{B}\) satisfies...
2. MEASURABLE FUNCTIONS

(c) \( \sigma[\{[-\infty, x] : x \in R\}] = \bar{B} \equiv \sigma[\mathcal{B}, \{-\infty\}, \{+\infty\}] \).

Since \( \mathcal{B} = \sigma[C_I] \) for \( C_I \) as in (1.3.2) and since

\[
\begin{align*}
(a, b) &= [-\infty, b] \cap [-\infty, a]^c, \\
[-\infty, b) &= \bigcup_{1}^{\infty} [-\infty, b - 1/n],
\end{align*}
\]

(e) \{−∞\} = \bigcap_{n}[−∞, −n], \{+∞\} = \bigcap_{n}[−∞, n]^c, etc.,

the equality (c) is obvious. The rest is trivial. □

Proposition 2.2 (Measurability of common functions) Let \( X, Y, \) and \( X_n \)'s be measurable functions. Consider \( cX \) with \( c > 0 \), \( −X \), \( \inf X_n \), \( \sup X_n \), \( \liminf X_n \), \( \limsup X_n \), \( \lim X_n \) if it exists, \( X^2, X \pm Y \) if it is well-defined, \( XY \) where \( 0 \cdot \infty \equiv 0, X/Y \) if it is well-defined, \( X^+, X^−, |X| \), and the composite \( g(X) \) for a continuous \( g \) and for any measurable function \( g \). All of these are measurable functions.

Proposition 2.3 (Measurability via simple functions)

(7) Simple and elementary functions are measurable.

(8) \( X : \Omega \to \bar{R} \) is measurable if and only if

\( X \) is the limit of a sequence of simple functions.

Moreover:

If \( X \geq 0 \) is measurable, then \( X \) is

(9) the limit of a sequence of simple functions that are \( \geq 0 \) and \( \nearrow \).

The \( X_n \)'s and \( Z_n \)'s that are defined in (10) and (12) below are important.

Proof. The functions in proposition 2.2 are measurable, as is now shown.

(a) \( [cX < x] = [X < x/c], \quad [−X < x] = [X > −x] \).

(b) \( \inf X_n < x] = \bigcup [X_n < x], \quad \sup X_n = −\inf(−X_n) \).

(c) \( \liminf X_n = \sup (\inf X_k), \quad \limsup X_n = −\liminf(−X_n) \).

(d) \( \lim X_n = \liminf X_n \), provided that \( \lim X_n(\omega) \) exists for all \( \omega \).

(e) \( [X^2 < x] = [−\sqrt{x} < x < \sqrt{x}] = [X < \sqrt{x}] \cap [X \leq −\sqrt{x}] \).

Each of the sets where \( X \) or \( Y \) equals 0, \( \infty \), or \( −\infty \) is measurable; use this below.

(f) \( [X > Y] = \bigcup_r \{ X > r > Y : r \text{ is rational} \} \), so \( [X > Y] \) is a measurable set.

So, \( [X + Y > z] = [X > z − Y] \in \mathcal{A} \) since \( z − Y \) is trivially measurable.

(Here \( [X = \infty] \cap [Y = −\infty] = \emptyset \) is implied, as \( X + Y \) is well defined. Etc., below.)
(g) \(X - Y = X + (-Y)\) and \(XY = [(X + Y)^2 - (X - Y)^2]/4\).

(h) \(X/Y = X \times (1/Y)\),

since \([1/Y < x] = [Y > 1/x]\) for \(x > 0\) in case \(Y > 0\), and for general \(Y\) one can write \(1/Y = 1[Y>0] - 1[Y<0]\) with the two indicator functions measurable.

(i) \(X^+ = X \vee 0\) and \(X^- = (-X) \vee 0\).

For \(g\) measurable, \((g \circ X)^{-1}(\bar{B}) = X^{-1}(g^{-1}(\bar{B})) \subset X^{-1}(\bar{B}) \subset \mathcal{A}\). Then continuous \(g\) are measurable, since

(j) \(g^{-1}(\mathcal{B}) = g^{-1}(\sigma[\text{open sets}]) = \sigma[g^{-1}(\text{open sets})] \subset \sigma[\text{open sets}] \subset \bar{B}\), and both \(g^{-1}(\{+\infty\})\) and \(g^{-1}(-\infty)\) are a (possibly void) subset of \((-\infty, +\infty)\). Now apply the result for measurable \(g\).

We now prove proposition 2.3. Claim (7) is trivial. Consider (8). Define simple functions \(X_n\) by

\[X_n \equiv \sum_{k=1}^{n2^n} k - \frac{1}{2^n} \times \left\{1_{\left[k-\frac{1}{2n}\leq X < k\frac{1}{2n}\right]} - 1_{\left[k-\frac{1}{2n}\leq -X < k\frac{1}{2n}\right]}\right\} + n \times \left\{1_{\{X \geq n\}} - 1_{\{-X \geq n\}}\right\}.\]

Since \(|X_n(\omega) - X(\omega)| \leq 2^{-n}\) for \(|X(\omega)| < n\), we have

(k) \(X_n(\omega) \to X(\omega)\) as \(n \to \infty\) for each \(\omega \in \Omega\).

Also, the nested subdivisions \(k/2^n\) cause \(X_n\) to satisfy

(l) \(X_n \nearrow\) when \(X \geq 0\).

We extend proposition 2.3 slightly by further observing that

\[\|X_n - X\| \to 0\] as \(n \to \infty\), if \(X\) is bounded.

Also, the elementary functions

\[Z_n \equiv \sum_{k=1}^{\infty} k - \frac{1}{2^n} \times \left\{1_{\left[k-\frac{1}{2n}\leq X < k\frac{1}{2n}\right]} - 1_{\left[k-\frac{1}{2n}\leq -X < k\frac{1}{2n}\right]}\right\} + \infty \times \left\{1_{\{X = \infty\}} - 1_{\{X = -\infty\}}\right\}\]

are always such that

\[\|(Z_n - X) \times 1_{[-\infty<X<\infty]}\| \leq 1/2^n \to 0\] as \(n \to \infty\).

Proposition 2.4 (The discontinuity set is measurable; Billingsley) If \((M, d)\) and \((M', d')\) are metric spaces and \(\psi : M \to M'\) is any function (not necessarily a measurable function), then the discontinuity set of \(\psi\) defined by
2. MEASURABLE FUNCTIONS

(14) \[ D_\psi \equiv \{ x \in M : \psi \text{ is not continuous at } x \} \]

is necessarily in the Borel \( \sigma \)-field \( B_d \) (that is, the \( \sigma \)-field generated by the \( d \)-open subsets of \( M \)).

Proof. Let

\[ A_{\epsilon, \delta} \equiv \{ x \in M : d(x, y) < \delta, d(x, z) < \delta \text{ and } d'(\psi(y), \psi(z)) \geq \epsilon \text{ for distinct } y, z \in M \}. \]

Note that \( A_{\epsilon, \delta} \) is an open set, since \( \{ u \in M : d(x, u) < \delta_0 \} \subset A_{\epsilon, \delta} \) will necessarily occur if \( \delta_0 \equiv \{ \delta - [d(x, y) \lor d(x, z)] \}/2 \); that is, the \( y \) and \( z \) that work for \( x \) also work for all \( u \) in \( M \) that are sufficiently close to \( x \). (Note: The \( y \) that worked for \( x \) may have been \( x \) itself.) Then

\[ D_\psi = \bigcup_{i=1}^{\infty} \bigcap_{j=1}^{\infty} A_{\epsilon_i, \delta_j} \in B_d, \]

where \( \epsilon_1, \epsilon_2, \ldots \) and \( \delta_1, \delta_2, \ldots \) both denote the positive rationals, since each \( A_{\epsilon, \delta} \) is an open set.

\[ \square \]

Induced Measures

Example 2.1 (Induced measures) We now turn to the “induced measure” previewed above. Suppose \( X : (\Omega, \mathcal{A}, \mu) \to (\Omega', \mathcal{A}') \), so that \( X \) is \( \mathcal{A}' \)-\( \mathcal{A} \)-measurable. We define \( \mu_X \equiv \mu' \) by

\[ (15) \quad \mu_X(A') \equiv \mu'(A') \equiv \mu(X^{-1}(A')) \quad \text{for each } A' \in \mathcal{A}'. \]

Then \( \mu_X \equiv \mu' \) is a measure on \( (\Omega', \mathcal{A}') \) called the induced measure. This is true, since we verify that

(a) \[ \mu'(\emptyset) = \mu(X^{-1}(\emptyset)) = \mu(\emptyset) = 0, \]

and

\[ \mu'(\sum_{n=1}^{\infty} A'_n) = \mu(X^{-1}(\sum_{n=1}^{\infty} A'_n)) = \mu(\sum_{n=1}^{\infty} X^{-1}(A'_n)) = \sum_{n=1}^{\infty} \mu(X^{-1}(A'_n)) = \sum_{n=1}^{\infty} \mu'(A'_n). \]

Note also that

(c) \[ \mu'(\Omega') = \mu(X^{-1}(\Omega')) = \mu(\Omega). \]

Thus if \( \mu \) is a probability measure, then so is \( \mu_X \equiv \mu' \). Note also that we could regard \( X \) as an \( \mathcal{A}' \)-\( \mathcal{F}(X) \)-measurable transformation from the measure space \( (\Omega, \mathcal{F}(X), \mu) \) to \( (\Omega', \mathcal{A}', \mu_X) \).

Suppose further that \( F \) is a generalized df on the real line \( R \), and that \( \mu_F(\cdot) \) is the associated measure on \( (R, \mathcal{B}) \) satisfying \( \mu_F([a, b]) = F(b) - F(a) \) for all \( a \) and \( b \) (as was guaranteed by the correspondence theorem (theorem 1.3.1)). Thus \( (R, \mathcal{B}, \mu_F) \) is a measure space. Define

\[ (16) \quad X(\omega) = \omega \quad \text{for all } \omega \in R. \]

Then \( X \) is a measurable transformation from \( (R, \mathcal{B}, \mu_F) \) to \( (R, \mathcal{B}) \) whose induced measure \( \mu_X \) is equal to \( \mu_F \). Thus for any given df \( F \) we can always construct a measurable function \( X \) whose df is \( F \). \[ \square \]
Exercise 2.1 Suppose $(\Omega, \mathcal{A}) = (R_2, B_2)$, where $B_2$ denotes the $\sigma$-field generated by all open subsets of the plane. Recall that this $\sigma$-field contains all sets $B \times R$ and $R \times B$ for all $B \in B$; here $B_1 \times B_2 = \{(r_1, r_2) : r_1 \in B_1, r_2 \in B_2\}$. Now define measurable transformations $X_1((r_1, r_2)) = r_1$ and $X_2(r_1, r_2) = r_2$. Then define $Z_1 \equiv (X_1^2 + X_2^2)^{1/2}$ and $Z_2 \equiv \text{sign}(X_1 - X_2)$, where $\text{sign}(r)$ equals $1, 0, -1$ according as $r$ is $>0, = 0, <0$. The exercise is to give geometric descriptions of the $\sigma$-fields $\mathcal{F}(Z_1)$, $\mathcal{F}(Z_2)$, and $\mathcal{F}(Z_1, Z_2)$.

Proposition 2.5 (The form of an $\mathcal{F}(Z)$-measurable function) Suppose that $Z$ is a measurable function on $(\Omega, \mathcal{A})$ and that $Y$ is $\mathcal{F}(Z)$-measurable. Then there must exist a measurable function $g$ on $(\bar{R}, \bar{B})$ such that $Y = g(Z)$.

Proof. (The approach of this proof is to consider indicator functions, simple functions, nonnegative functions, general functions. This approach will be used again and again. Learn it!) Suppose that $Y = 1_D$ for some set $D \in \mathcal{F}(Z)$, so that $Y$ is an indicator function that is $\mathcal{F}(Z)$-measurable. Then we can rewrite $Y$ as $Y = 1_D = 1_{\mathcal{F}^{-1}(B)} = 1_B(Z) \equiv g(Z)$, for some $B \in B$ that depends on $D$, where $g(r) \equiv 1_B(r)$. Thus the proposition holds for indicator functions. It holds for simple functions, since when all $B_i \in B$,

$$Y = \sum_{i=1}^m c_i 1_{D_i} = \sum_{i=1}^m c_i 1_{Z^{-1}(B_i)} = \sum_{i=1}^m c_i 1_{B_i}(Z) \equiv g(Z).$$

Let $Y \geq 0$ be $\mathcal{F}(Z)$-measurable. Then there do exist simple $\mathcal{F}(Z)$-measurable functions $Y_n$ such that $Y \equiv \lim_n Y_n = \lim_n g_n(Z)$ for the simple $\bar{B}$-measurable functions $g_n$. Now let $g = \lim g_n$, which is $\bar{B}$-measurable, and note that $Y = g(Z)$. For general $Y = Y^+ - Y^-$, use $g = g^+ - g^-$. \qed

Exercise 2.2 (Measurability criterion) Let $\mathcal{C}$ denote a $\pi$-system of subsets of $\Omega$. Let $\mathcal{V}$ denote a vector space of functions; that is, $X + Y \in \mathcal{V}$ and $\alpha X \in \mathcal{V}$ for all $X, Y \in \mathcal{V}$ and all $\alpha \in R$—and, all the usual elementary facts hold.

(a) Suppose that:

17) $1_C \in \mathcal{V}$ for all $C \in \mathcal{C}$.

18) If $A_n \not\overset{\mathcal{C}}{\sim} A$ with $1_{A_n} \in \mathcal{V}$, then $1_A \in \mathcal{V}$.

Show that $1_A \in \mathcal{V}$ for every $A \in \sigma[\mathcal{C}]$.

(b) It then follows trivially that every simple function

$$X_n \equiv \sum_{i=1}^m \alpha_i 1_{A_i} \in \mathcal{V};$$

here $m \geq 1$, all $\alpha_i \in R$, and $\sum_{i=1}^m A_i = \Omega$ with all $A_i \in \sigma[\mathcal{C}]$.

(c) Now suppose further that $X_n \not\overset{\mathcal{C}}{\sim} X$ for $X_n$’s as in (19) implies that $X \in \mathcal{V}$. Show that $\mathcal{V}$ contains all $\sigma[\mathcal{C}]$-measurable functions.
3 Convergence

Convergence Almost Everywhere

**Definition 3.1** (→.e.) Let $X_1, X_2, \ldots$ denote measurable functions on $(\Omega, \mathcal{A}, \mu)$ to $(\bar{\mathbb{R}}, \bar{\mathcal{B}})$. Say that the sequence $X_n$ converges almost everywhere to $X$ (denoted by $X_n \to_{.e.} X$) if for some $N \in \mathcal{A}$ for which $\mu(N) = 0$ we have $X_n(\omega) \to X(\omega)$ as $n \to \infty$ for all $\omega \notin N$. If for all $\omega \notin N$ the sequence $X_n(\omega)$ is a Cauchy sequence, then we say that the sequence $X_n$ mutually converges a.e. and denote this by writing $X_n - X_m \to_{.e.} 0$ as $m \wedge n \to \infty$. (Here, $m \wedge n \equiv \min(m, n).$)

**Exercise 3.1** Let $X_1, X_2, \ldots$ be measurable functions from $(\Omega, \mathcal{A}, \mu)$ to $(\bar{\mathbb{R}}, \bar{\mathcal{B}})$.

(a) If $X_n \to_{.e.} X$, then $X = \tilde{X}$ a.e. for some measurable $\tilde{X}$.

(b) If $X_n \to_{.e.} X$ and $\mu$ is complete, then $X$ itself is measurable.

**Proposition 3.1** A sequence of measurable functions $X_n$ that are a.e. finite converges a.e. to a measurable function $X$ that is a.e. finite if and only if these functions $X_n$ converges mutually a.e. (Thus we can redefine such functions on null sets and make them everywhere finite and everywhere convergent and/or follow the convention of corollary 2 to the Carethéodory theorem 1.2.1 and automatically complete every measure.)

**Proof.** The union of the countable number of null sets on which finiteness or convergence fails is again a null set $N$. On $N^c$, the claim is just a property of the real numbers. □

**Proposition 3.2 (The convergence and divergence sets are measurable)** Consider the finite measurable functions $X, X_1, X_2, \ldots$ (perhaps redefined on null sets to achieve this); thus, they are $\mathcal{B} \cdot \mathcal{A}$-measurable. Then the convergence and mutual convergence sets are measurable. In fact, the **convergence set** is given by

$$(1) \quad [X_n \to X] \equiv \bigcap_{k=1}^\infty \bigcup_{n=1}^\infty \bigcap_{m=n}^\infty \left[|X_m - X| < \frac{1}{k}\right] \in \mathcal{A},$$

and the **mutual convergence set** is given by

$$(2) \quad [X_n - X_m \to 0] \equiv \bigcap_{k=1}^\infty \bigcup_{n=1}^\infty \bigcap_{m=n}^\infty \left[|X_m - X_n| < \frac{1}{k}\right] \in \mathcal{A}.$$ 

**Proof.** Just read the right-hand side of (1) as, for all $\epsilon \equiv 1/k > 0$ there exists an $n$ such that for all $m \geq n$ we have $|X_m(\omega) - X(\omega)| < 1/k$. (Practice saying this until it makes sense.) □

Taking complements in (1) allows the **divergence set** to be expressed via

$$(3) \quad [X_n \to X]^c = \bigcup_{k=1}^\infty \bigcap_{n=1}^\infty \bigcup_{m=n}^\infty \left[|X_m - X| \geq \frac{1}{k}\right] \equiv \bigcup_{k=1}^\infty A_k \quad \text{with} \quad A_k \nearrow \in k,$$

where

$$(4) \quad A_k = \bigcap_{n=1}^\infty D_{kn}, \quad \text{and} \quad D_{kn} = \bigcup_{m=n}^\infty [\|X_m - X| \geq 1/k] \quad \text{are} \quad \searrow \in n.$$
Proposition 3.3  Consider finite measurable $X_n$’s and a finite measurable $X$ on any $(\Omega, \mathcal{A}, \mu)$. (i) We have

\[ X_n \xrightarrow{\text{a.e.}} (\text{such an } X) \quad \text{iff} \quad X_n - X_m \xrightarrow{\text{a.e.}} 0 \quad \text{iff} \quad \mu\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \{|X_m - X_n| > \epsilon\}\right) = 0, \quad \text{for all } \epsilon > 0. \]

(A finite limit $X(\omega)$ exists if and only if the Cauchy criterion holds; and we want to be able to check for the existence of a finite limit $X(\omega)$ without knowing its value.)

(ii) (Most useful criterion for $\xrightarrow{\text{a.e.}}$. On any $(\Omega, \mathcal{A}, \mu)$, we have

\[ X_n \rightarrow \text{a.e.} \quad \text{(some finite measurable } X) \quad \text{provided} \]

\[ \mu\left(\bigcup_{m=n}^{\infty} \{|X_m - X_n| > \epsilon\}\right) \rightarrow 0, \quad \text{for all } \epsilon > 0, \]

\[ \mu\left(\bigcup_{n \leq m \leq N} \{|X_m - X_n| > \epsilon\}\right) \leq \epsilon \quad \text{for all } N \geq n \geq (\text{some } n_\epsilon), \quad \text{for all } \epsilon > 0. \]

Proof. Use proposition 1.1.2 on the $\uparrow$ sets in the mutual convergence analog of the sets $A_k$ in (3) to obtain (5). Then the intersection of sets in (5) is a subset of each set in the intersection; thus (6) yields (5). Finally, the sets in (7) increase to the set in (6); so use proposition 1.1.2 yet again. (Replace $X_n$ by $X$ in (5), (6), and (7) and require $\mu(\Omega) < \infty$. Then the converse that (5) implies (6) holds, as the events in (6) are then $\downarrow$.) □

Remark 3.1 (Additional measurability for convergence and divergence) Suppose we still assume that $X_1, X_2, \ldots$ are finite measurable functions. Then the following sets are seen to be measurable:

\[ [\omega : X_n(\omega) \to X(\omega) \in \bar{R}]^c = [\lim \inf X_n < \lim \sup X_n] \]

\[ = \bigcup_{\text{rational } r} [\lim \inf X_n < r < \lim \sup X_n] \in \mathcal{A}, \]

\[ [\lim \sup X_n = +\infty] = \bigcap_{m=1}^{\infty} [\lim \sup X_n > m] \in \mathcal{A}. \]

These comments reflect the following fact: If $X_n(\omega)$ does not converge to a finite number, then there are several different possibilities; but these interesting events are all measurable. □

Convergence in Measure

Definition 3.2 ($\rightarrow_{\mu}$). A given sequence of measurable and a.e. finite functions $X_1, X_2, \ldots$ is said to converge in measure to the measurable function $X$ taking values in $\bar{R}$ (to be denoted by $X_n \rightarrow_{\mu} X$ as $n \to \infty$) if

\[ \mu([|X_n - X| \geq \epsilon]) \to 0 \quad \text{as } n \to \infty, \quad \text{for all } \epsilon > 0. \]

(Such convergence implies that $X$ must be finite a.s., as

\[ [|X| = \infty] \subset \left\{\bigcup_{k=1}^{\infty} [|X_k| = \infty]\right\} \cup [|X_n - X| \geq \epsilon] \]

shows.) We say that these $X_n$ converge mutually in measure, which we denote by writing

\[ X_m - X_n \rightarrow_{\mu} 0 \quad \text{as } m \wedge n \to \infty, \quad \text{if } \mu([|X_m - X_n| \geq \epsilon]) \to 0 \quad \text{as } m \wedge n \to \infty, \quad \text{for each } \epsilon > 0. \]
Proposition 3.4  (a) If \( X_n \to_{\mu} X \) and \( X_n \to_{\mu} \tilde{X} \), then \( X = \tilde{X} \) a.e.

(b) On a complete measure space, \( X = \tilde{X} \) on \( N^c \), for a null set \( N \).

Proof.  For all \( \epsilon > 0 \)

(a) \( \mu( \{|X - \tilde{X}| \geq 2\epsilon\} ) \leq \mu( \{|X_n - X| \geq \epsilon\} ) + \mu( \{|X_n - \tilde{X}| \geq \epsilon\} ) \to 0 \),

giving \( \mu( \{|X - \tilde{X}| \geq \epsilon\} ) = 0 \) for all \( \epsilon > 0 \). Thus

(b) \( \mu( \{|X \neq \tilde{X}| \} ) = \mu( \bigcup_k \{|X - \tilde{X}| \geq 1/k\} ) \leq \sum_1^\infty \mu( \{|X - \tilde{X}| \geq 1/k\} ) = \sum_1^\infty 0 \),
as claimed. \( \square \)

Exercise 3.2  (a) Show that in general \( \to_{\mu} \) does not imply \( \to_{a.e.} \).

(b) Give an example with \( \mu(\Omega) = \infty \) where \( \to_{a.e.} \) does not imply \( \to_{\mu} \).

Theorem 3.1 (Relating \( \to_{\mu} \) to \( \to_{a.e.} \).) Let \( X \) and \( X_1, X_2, \ldots \) be measurable and finite a.e. functions. The following are true.

(11) \( X_n \to_{a.e.} \) (such an \( X \)) if and only if \( X_n - X_m \to_{a.e.} 0 \).

(12) \( X_n \to_{\mu} \) (such an \( X \)) if and only if \( X_n - X_m \to_{\mu} 0 \).

(13) Let \( \mu(\Omega) < \infty \). Then \( X_n \to_{a.e.} \) (such an \( X \)) implies \( X_n \to_{\mu} X \).

(14) (Riesz) If \( X_n \to_{\mu} X \), then for some \( n_k \) we have \( X_{n_k} \to_{a.e.} X \). (See (16)).

(Reducing \( \to_{\mu} \) to \( \to_{a.e.} \) by going to subsequences) Suppose \( \mu(\Omega) < \infty \). Then

\begin{equation}
(15) \quad X_n \to_{\mu} X \quad \text{if and only if}
\end{equation}

each subsequence \( n' \) has a further \( n'' \) on which \( X_{n''} \to_{a.e.} \) (such an \( X \)).

Proof.  Now, (11) is proposition 3.1, and (12) is exercise 3.3 below. Result (13) comes from the elementary observation that

(a) \( \mu( \{|X_n - X| \geq \epsilon\} ) \leq \mu( \bigcup_{m=n}^\infty \{|X_m - X| \geq \epsilon\} ) \to 0 \), by (6).

To prove (14), choose \( n_k \uparrow \) such that

(b) \( \mu(A_k) \equiv \mu( \{|X_{n_k} - X| > 1/2^k\} ) < 1/2^k \),

with \( \mu( \{|X_n - X| > 1/2^k\} ) < 1/2^k \) for all \( n \geq n_k \). Now let

(c) \( B_m \equiv \bigcup_{k=m}^\infty A_k \), so that \( \mu(B_m) \leq \sum_{k=m}^\infty 2^{-k} \leq 1/2^{m-1} \)

On \( B_m^c = \bigcap_{k=m}^\infty A_k^c \) we have \( |X_{n_k} - X| \leq 1/2^k \) for all \( k \geq m \), so that

(d) \( |X_{n_k}(\omega) - X(\omega)| \leq 1/2^k \to 0 \) as \( k \to \infty \), for each \( \omega \in B_m^c \),

with \( \mu(B_m) \leq 1/2^{m-1} \). Since convergence occurs on each \( B_m^c \), we have
CHAPTER 2. MEASURABLE FUNCTIONS AND CONVERGENCE

(e) \( X_{n_k}(\omega) \to X(\omega) \) as \( k \to \infty \) for each \( \omega \in C \equiv \bigcup_{m=1}^{\infty} B_m^c \),
where \( B_m = \bigcup_{k=m}^{\infty} A_k \) is \( \subset \) with \( (\bigcap_{m=1}^{\infty} B_m) \subset (\text{every } B_m) \). So

(f) \( \mu(C^c) = \mu(\bigcap_{m=1}^{\infty} B_m) \leq \limsup \mu(B_m) \leq \lim \frac{1}{2^{m-1}} = 0 \),
completing the proof of (14).

(Comment on exercise 3.3: When \( X_n \to \mu X \), analogy with (a) gives
(16) \( \mu(\{|X_m - X_n| \geq 1/2^k\}) \leq 1/2^k \) for all \( m, n \geq (\text{some } n_k) \).
Thus \( A_k \equiv \{ |X_{n_k} - X_{n_k+1}| \geq 1/2^k \} \) has \( P(A_k) \leq 1/2^k \) for all \( k \). In analogy with the first paragraph, prove the a.s. convergence of the \( X_{n_k} \) to some \( X \) on this subsequence by considering

(17) \( |X_{n_k} - X_{n_\ell}| \leq |(X_{n_k} - X_{n_{k+1}}) + \cdots + (X_{n_{\ell-1}} - X_{n_\ell})| \).

Then show that the whole sequence converges in measure to this \( X \).

Consider the unproven half of (15). Suppose that every \( n' \) contains a further \( n'' \) as claimed
(with a particular \( X \)). Assume that \( X_n \to \mu X \) fails. Then for some \( \epsilon_o > 0 \) and some \( n' \)

(g) \( \lim_{n'} \mu(\{|X_n' - X| > \epsilon_o\}) = (\text{some } a_o) > 0 \).

But we are given that some further subsequence \( n'' \) has \( X_{n''} \to_{\text{a.e.}} X \), and thus \( X_{n''} \to \mu X \) by (13), using \( \mu(\Omega) < \infty \). Thus

(h) \( \lim_{n''} \mu(\{|X_{n''} - X| > \epsilon_o\}) = 0 \);
but this is a contradiction of (g). \( \square \)

**Exercise 3.3** As in (12), show that \( X_n \to \mu X \) if and only if \( X_m - X_n \to \mu 0 \). (Hint. Adapt the proof of (16).)

**Exercise 3.4** (a) Suppose that \( \mu(\Omega) < \infty \) and \( g \) is continuous a.e. \( \mu_X \) (that is, \( g \) is continuous except perhaps on a set of \( \mu_X \) measure 0). Then \( X_n \to \mu X \) implies that \( g(X_n) \to \mu g(X) \).

(b) Let \( g \) be uniformly continuous on the real line. Then \( X_n \to \mu X \) implies that \( g(X_n) \to \mu g(X) \). (Here, \( \mu(\Omega) = \infty \) is allowed.)

**Exercise 3.5** (a) (Dini) Consider continuous transformations \( X_n \) from a compact space \( \Omega \) to \( R \) for which \( X_n(\omega) \not\to X(\omega) \) for each \( \omega \in \Omega \), where \( X \) is continuous. Then \( X_n \) converges uniformly to \( X \) on \( \Omega \). (Likewise, if \( X_n(\omega) \not\to X(\omega) \) for all \( \omega \).
(b) In general, a uniform limit of bounded and continuous functions \( X_n \) is also bounded and continuous.
4 Probability, RVs, and Convergence in Law

Definition 4.1 (Random variable and df) (a) A probability space \((\Omega, \mathcal{A}, P)\) is just a measure space for which \(P(\Omega) = 1\). Now, \(X : (\Omega, \mathcal{A}, P) \to (R, \mathcal{B})\) will be called a random variable (to be abbreviated rv); thus it is a \(\mathcal{B}\)-\(\mathcal{A}\)-measurable function. If \(X : (\Omega, \mathcal{A}, P) \to (\bar{R}, \mathcal{B})\), then we will call \(X\) an extended rv.

(b) The distribution function (to be abbreviated \(df\)) of a rv is defined by

\[
F_X(x) \equiv P(X \leq x) \quad \text{for all } -\infty < x < \infty.
\]

We recall that \(F \equiv F_X\) satisfies

\[
F \text{ is increasing and right continuous, with } F(-\infty) = 0 \text{ and } F(+\infty) = 1.
\]

We let \(C_F\) denote the continuity set of \(F\) that contains all points at which \(F\) is continuous. (That \(F \not\equiv \) is trivial, and the other three properties all follow from the monotone property of measure, since \((\infty, x] = \bigcap_{n=1}^{\infty} (-\infty, n] = 0\), and \(\bigcup_{n=1}^{\infty} (-\infty, n] = R\).)

(c) If \(F\) is increasing and right continuous with \(F(\infty) \geq 0\) and \(F(-\infty) \leq 1\), then \(F\) will be called a sub \(df\).

(d) The induced measure on \((R, \mathcal{B})\) (or \((\bar{R}, \mathcal{B})\)) will be denoted by \(P_X\). It satisfies

\[
P_X(B) = P(X^{-1}(B)) = P(X \in B) \quad \text{for all } B \in \mathcal{B}
\]

(for all \(B \in \mathcal{B}\) if \(X\) is an extended rv). We call this the induced distribution of \(X\). We use the notation \(X \equiv F\) to denote that the induced distribution \(P_X(\cdot)\) of the rv \(X\) has \(df\) \(F\).

(e) We say that rv \(X_n\) (with dfs \(F_n\)) converge in distribution or converge in law to a rv \(X_0\) (with df \(F_0\)) if

\[
F_n(x) = P(X_n \leq x) \to F_0(x) = P(X_0 \leq x) \quad \text{at each } x \in C_{F_0}.
\]

We abbreviate this by writing either \(X_n \to_d X_0, F_n \to_d F_0\), or \(\mathcal{L}(X_n) \to \mathcal{L}(X_0)\).

Notation 4.1 Suppose now that \(\{X_n : n \geq 0\}\) are rv's on \((\Omega, \mathcal{A}, P)\). Then it is customary to write \(X_n \to_p X_0\) (in place of \(X_n \to_{\mu} X_0\)) and \(X_n \to_{a.s.} X_0\) (as well as \(X_n \to_{a.e.} X_0\)). The “\(p\)” is an abbreviation for in probability, and the “a.s.” is an abbreviation for almost surely. Anticipating the next chapter, we let \(Eg(X)\) denote \(\int g(X) d\mu\) or \(\int g(X) dP\) when \(\mu\) is a probability measure \(P\). We say that \(X_n\) converges to \(X_0\) in \(r\)th mean if \(E|X_n - X_0|^r \to 0\). We denote this by writing \(X_n \to_r X_0\) or \(X_n \to_{\mathcal{L}_r} X_0\).

Proposition 4.1 Suppose that the rv's \(X \equiv F\) and \(X_n \equiv F_n\) satisfy \(X_n \to_p X\). Then \(X_n \to_d X\). (Thus, \(X_n \to_{a.s.} X\) implies that \(X_n \to_d X\).)

Proof. (This result has limited importance. But the technique introduced here is useful; see exercise 4.1 below.) Now,

(a) \(F_n(t) = P(X_n \leq t) \leq P(X \leq t + \epsilon) + P(|X_n - X| \geq \epsilon)\)

(b) \(\leq F(t + \epsilon) + \epsilon\) for all \(n \geq \text{some } n_{\epsilon}\).

Also,

\[
F_n(t) = P(X_n \leq t) \geq P(X \leq t - \epsilon \text{ and } |X_n - X| \leq \epsilon) \equiv P(AB) \\
\geq P(A) - P(B^c) = F(t - \epsilon) - P(|X_n - X| > \epsilon) \\
\geq F(t - \epsilon) - \epsilon \quad \text{for } n \geq \text{(some } n'_{\epsilon})\)
\]
Thus for \( n \geq (n_\epsilon \lor n'_\epsilon) \) we have

\[
(c) \quad F(t - \epsilon) - \epsilon \leq \liminf F_n(t) \leq \lim F_n(t) \leq F(t + \epsilon) + \epsilon.
\]

If \( t \) is a continuity point of \( F \), then letting \( \epsilon \to 0 \) in \( (c) \) gives \( F_n(t) \to F(t) \). Thus \( F_n \to_d F \).

The following elementary result is extremely useful. Often, one knows that \( X_n \to_d X \), but what one is really interested in is a slight variant of \( X_n \), rather than \( X_n \) itself. The next result was designed for just such situations.

**Definition 4.2** (Type) Two rvs \( X \) and \( Y \) are of the same type if \( Y \sim = aX + b \).

**Theorem 4.1** (Slutsky) Suppose that \( X_n \to_d X \), while the rvs \( Y_n \to_p a \) and \( Z_n \to_p b \) as \( n \to \infty \) (here \( X_n, Y_n \), and \( Z_n \) are defined on a common probability space, but \( X \) need not be). Then

\[
(5) \quad U_n \equiv Y_n \times X_n + Z_n \to_d aX + b \quad \text{as } n \to \infty.
\]

**Exercise 4.1** Prove Slutsky’s theorem. (Hint. Recall the proof of proposition 4.1. Then write \( U_n = (Y_n - a)X_n + (Z_n - b) + aX_n + b \) where \( Y_n - a \to_p 0 \) and \( Z_n - b \to_p 0 \). Note also that \( P(|X_n| > (\text{some large } M_\epsilon)) < \epsilon \) for all \( n \geq (\text{some } n_\epsilon) \).

**Exercise 4.2** Let \( c \) be a constant. Show that \( X_n \to_d c \) if and only if \( X_n \to_p c \).

**Remark 4.1** Suppose \( X_1, X_2, \ldots \) are independent rvs with a common df \( F \). Then \( X_n \to_d X_0 \) for any rv \( X_0 \) having df \( F \). However, there is no rv \( X \) for which \( X_n \) converges to \( X \) in the sense of \( \to_{a.s.}, \to_p, \) or \( \to_r \). (Of course, we are assuming that \( X \) is not a degenerate rv (that is, that \( \mu_F \) is not a unit point mass).)

\( \square \)
5 Discussion of Sub σ-Fields*

Consider again a sequence of rvs $X_1, X_2, \ldots$, where each quantity $X_n$ is a measurable transformation $X_n : (\Omega, \mathcal{A}, P) \rightarrow (R, \mathcal{B}, P_{X_n})$, and where $P_{X_n}$ denotes the induced measure. Each rv $X_n$ is $\mathcal{B}\cdot \mathcal{F}(X_n)$-measurable, with $\mathcal{F}(X_n)$ a sub $\sigma$-field of $\mathcal{A}$. Even though the intersection of any number of $\sigma$-fields is a $\sigma$-field, the union of even two $\sigma$-fields need not be a $\sigma$-field. We thus define the **sub $\sigma$-field generated by** $X_1, \ldots, X_n$ as

\[ \mathcal{F}(X_1, \ldots, X_n) \equiv \sigma[\bigcup_{k=1}^{n} \mathcal{F}(X_k)] = \mathcal{X}^{-1}(B_n) \quad \text{for} \quad X_n \equiv (X_1, \ldots, X_n)', \]

where the equality will be shown in the elementary proposition 5.2.1 below.

Note that $\mathcal{F}(X_1, \ldots, X_n) \subset \mathcal{F}(X_1, \ldots, X_n, X_{n+1})$, so that these necessarily form an increasing sequence of $\sigma$-fields of $\mathcal{A}$. Also, define

\[ \mathcal{F}(X_1, X_2, \ldots) \equiv \sigma[\bigcup_{k=1}^{\infty} \mathcal{F}(X_k)]. \]

It is natural to say that such $X_n = (X_1, \ldots, X_n)'$ are adapted to the $\mathcal{F}(X_1, \ldots, X_n)$. In fact, if $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots$ is any sequence of $\sigma$-fields for which $\mathcal{F}(X_1, \ldots, X_n) \subset \mathcal{F}_n$ for all $n$, then we say that the $X_n$’s are adapted to the $\mathcal{F}_n$’s.

Think of $\mathcal{F}(X_1, \ldots, X_n)$ as the **amount of information** available at time $n$ from $X_1, \ldots, X_n$; that is, you have available for inspection all of the probabilities

\[ P((X_1, \ldots, X_n) \in B_n) = P((X_1, \ldots, X_n)^{-1}(B_n)) = P(X_1, \ldots, X_n)(B_n), \]

for all Borel sets $B_n \in \mathcal{B}_n$. Rephrasing, you have available for inspection all of the probabilities $P(A)$, for all $A \in \mathcal{F}(X_1, \ldots, X_n)$.

At stage $n+1$ you have available $P(A)$ for all $A \in \mathcal{F}(X_1, \ldots, X_n, X_{n+1})$; that is, you have more information available. (Think of $\mathcal{F}_n \setminus \mathcal{F}(X_1, \ldots, X_n)$ as the amount of information available to you at time $n$ that goes beyond the information available from $X_1, \ldots, X_n$; perhaps some of it comes from other rvs not yet mentioned, but it is available nonetheless.)

Suppose we are not given rvs, but rather (speaking informally now, based on your general feel for probability) we are given joint dfs $F_n(x_1, \ldots, x_n)$ that we think ought to suffice to construct probability measures on $(R_n, \mathcal{B}_n)$. In (2.2.16) we saw that for $n = 1$ we could just let $(\Omega, \mathcal{A}, \mu) = (R, \mathcal{B}, \mu\mu)$ and use $X(\omega) = \omega$ to define a rv that carried the information in the df $F$. How do we define probability measures $P_n$ on $(R_n, \mathcal{B}_n)$ so that the **coordinate rvs**

\[ X_k(\omega_1, \ldots, \omega_n) = \omega_k \quad \text{for all} \quad (\omega_1, \ldots, \omega_n) \in R_n \]

satisfy

\[ P_n(X_1 \leq x_1, \ldots, X_n \leq x_n) = F_n(x_1, \ldots, x_n) \quad \text{for all} \quad (x_1, \ldots, x_n) \in R_n, \]

and thus carry all the information in $F_n$? Chapter 5 will deal with this construction. But even now it is clear that for this to be possible, the $F_n$’s will have to satisfy some kind of consistency condition as we go from step $n$ to $n+1$. Moreover, the consistency problem should disappear if the resulting $X_n$’s are “independent.”

But we need more. We will let $R_\infty$ denote all infinite sequences $\omega_1, \omega_2, \ldots$ for which each $\omega_i \in R$. Now, the construction of (5) and (6) will determine probabilities on the collection $\mathcal{B}_n \times \prod_{k=n+1}^{\infty} R$ of all subsets of $R_\infty$ of the form

\[ B_n \times \prod_{k=n+1}^{\infty} R \equiv \{(\omega_1, \ldots, \omega_n, \omega_{n+1}, \ldots) : (\omega_1, \ldots, \omega_n) \in B_n, \omega_k \in R \quad \text{for} \quad k \geq n+1\}, \]
with \(B_n \in B_n\). Each of these collections is a \(\sigma\)-field (which within this special probability space can be denoted by \(\mathcal{F}(X_1, \ldots, X_n)\)) in this overall probability space \((R_\infty, B_\infty, P_\infty)\), for some appropriate \(B_\infty\). But what is an appropriate \(\sigma\)-field \(B_\infty\) for such a probability measure \(P_\infty\)? At a minimum, \(B_\infty\) must contain
\[
\sigma \left[ \bigcup_{n=1}^{\infty} \left\{ B_n \times \prod_{k=n+1}^{\infty} R \right\} \right] = \sigma \left[ \bigcup_{n=1}^{\infty} \mathcal{F}(X_1, \ldots, X_n) \right],
\]
and indeed, this is what we will use for \(B_\infty\). Of course, we also want to construct the measure \(P_\infty\) on \((R_\infty, B_\infty)\) in such a way that
\[
P_\infty \left( \prod_{k=1}^{n} (-\infty, x_k] \times \prod_{k=n+1}^{\infty} R \right) = F_n(x_1, \ldots, x_n) \quad \text{for all } n \geq 1
\]
and for all \(x_1, \ldots, x_n\) in \(R\). The details are given in chapter 5.

Until chapter 5 we will assume that we are given the rvs \(X_1, X_2, \ldots\) on some \((\Omega, \mathcal{A}, P)\), and we will need to deal only with the known quantities \(\mathcal{F}(X_1, \ldots, X_n)\) and \(\mathcal{F}(X_1, X_2, \ldots)\) defined in (1) and (2). This is probability theory: Given \((\Omega, \mathcal{A}, P)\), we study the behavior of rvs \(X_1, X_2, \ldots\) that are defined on this space. Now contrast this with statistics: Given a physical situation producing measurements \(X_1, X_2, \ldots\), we construct models \(\{(R_\infty, B_\infty, P_\infty^\theta) : \theta \in \Theta\}\) based on various plausible models for \(F_n^\theta(x_1, \ldots, x_n)\), \(\theta \in \Theta\), and we then use the data \(X_1, X_2, \ldots\) and the laws of probability theory to decide which model \(\theta_0 \in \Theta\) was most likely to have been correct and what action to take. In particular, the statistician must know that the models to be used are well-defined.

We also need to extend all this to uncountably many rvs \(\{X_t : t \in T\}\), for some interval \(T\) such as \([a, b]\), or \([a, \infty)\), or \([a, \infty)\), or \((\infty, \infty)\), \ldots. We say that rvs \(X_t : (\Omega, \mathcal{A}, P) \to (R, B)\) for \(t \in T\) are adapted to an \(\mathcal{F}_t\) sequence of \(\sigma\)-fields \(\mathcal{F}_t\) if \(\mathcal{F}_s \subset \mathcal{F}_t\) for all \(s \leq t\) with both \(s, t \in T\) and if each \(X_t\) is \(\mathcal{F}_t\)-measurable. In this situation we typically let \(R_T \equiv \prod_{t \in T} R_t\) and then let
\[
\mathcal{F}_t \equiv \mathcal{F}(X_s : s \leq t) \equiv \sigma \left[ \bigcup_{s} X_s^{-1}(\mathcal{B}) : s \leq t \text{ and } s \in T \right] \quad \text{for all } t \in T.
\]
This is also done in chapter 5 (where more general sets \(T\) are, in fact, considered).

The purpose in presenting this section here is to let the reader start now to become familiar and comfortable with these ideas before we meet them again in chapter 5 in a more substantial and rigorous presentation. (The author assigns this as reading at this point and presents only a very limited amount of chapter 5 in his lectures.)

**Exercise 5.1** (a) Show that the class \(\mathcal{C} \equiv \{X_1^{-1}(B_1) \cap \{X_2^{-1}(B_2) : B_1, B_2 \in \mathcal{B}\}\) is a \(\pi\)-system that generates the \(\sigma\)-field \(\mathcal{F}(X_1, X_2)\).

(b) Recall the Dynkin \(\pi\)-\(\lambda\) theorem, and state its implications in this context.

(c) State an extension of this part (a) to \(\mathcal{F}(X_1, \ldots, X_n)\) and to \(\mathcal{F}(X_1, X_2, \ldots)\).
Probability for Statisticians
Shorack, G.R.
2017, XXII, 510 p. 19 illus., 15 illus. in color., Softcover
ISBN: 978-3-319-52206-7