## Contents

Chapter 0. Math prerequisites 7  
1. Convergence 7  
2. Set theory 7  
3. Compact set 9  
4. Riemann integral 9  
5. Product spaces 10  
6. Taylor polynomials and series expansions 10  
7. Complex numbers 10  
Additional Exercises 10

Chapter 1. Events and Probabilities 11  
1. Elementary and semi-elementary probability theory 11  
2. Sigma-fields 14  
Required Exercises 16  
Additional Exercises 17

Chapter 2. Probability measures 19  
1. Existence 19  
2. Uniqueness 22  
3. Probability measures on $\mathbb{R}$ 24  
4. Probability measures on $\mathbb{R}^k$ 28  
Required Exercises 31  
Additional Exercises 32

Chapter 3. Independence 33  
1. Independent events and sigma-fields 33  
2. Zero-one law 34  
3. Borel-Cantelli Lemmas 35  
Required Exercises 36  
Additional Exercises 36
Chapter 4. Random variables 39
  1. Measurable mappings 39
  2. Random variables with prescribed distributions 41
  3. Convergence of random variables 43
    Required Exercises 45
    Additional Exercises 46

Chapter 5. Simple random variables 43
  1. Expected value 43
  2. Inequalities 45
  3. $L_p$-norms 46
  4. The law of large numbers 47
    Required Exercises 47
    Additional Exercises 48

Chapter 6. Integration 49
  1. Approximation by simple random variables 49
  2. Expected values 50
  3. Inequalities 54
  4. Independent random variables 55
    Required Exercises 55
    Additional Exercises 55
    Moment generating functions 56

Chapter 7. Product measure and Fubini’s theorem 57
  1. Product spaces 57
  2. Product measure 57
  3. Fubini’s Theorem 58
    Required Exercises 60
    Additional Exercises 60

Chapter 8. Sums of independent random variables 61
  1. The strong law of large numbers 61
  2. Kolmogorov’s zero-one law 63
  3. Kolmogorov’s Maximal inequality and its applications 63
  4. Etemadi’s inequality and its application 66
    Required Exercises 67

Chapter 9. Weak convergence 69
  1. Convergence in distribution 69
  2. Fundamental results 71
    Required Exercises 75

Chapter 10. Characteristic functions 77
  1. Complex numbers, Taylor polynomials, etc 77
  2. Characteristic functions 79
  3. Uniqueness 81
  4. The continuity theorem 83
<table>
<thead>
<tr>
<th>Contents</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Required Exercises</td>
<td>84</td>
</tr>
<tr>
<td>Enrichments</td>
<td>85</td>
</tr>
<tr>
<td>Chapter 11. The Central Limit Theorem</td>
<td>87</td>
</tr>
<tr>
<td>1. Lindeberg and Lyapunow theorems</td>
<td>88</td>
</tr>
<tr>
<td>2. Lyapunov’s theorem</td>
<td>90</td>
</tr>
<tr>
<td>3. Strategies for proving CLT without Lindeberg condition</td>
<td>91</td>
</tr>
<tr>
<td>Required Exercises</td>
<td>91</td>
</tr>
<tr>
<td>Chapter 12. Limit Theorems in $\mathbb{R}^k$</td>
<td>93</td>
</tr>
<tr>
<td>1. The basic theorems</td>
<td>93</td>
</tr>
<tr>
<td>2. Multivariate characteristic function</td>
<td>94</td>
</tr>
<tr>
<td>3. Multivariate normal distribution</td>
<td>95</td>
</tr>
<tr>
<td>4. The CLT</td>
<td>97</td>
</tr>
<tr>
<td>Required Exercises</td>
<td>99</td>
</tr>
<tr>
<td>Appendix A. Addenda</td>
<td>101</td>
</tr>
<tr>
<td>1. Modeling an infinite number of tosses of a coin</td>
<td>101</td>
</tr>
<tr>
<td>Appendix. Bibliography</td>
<td>103</td>
</tr>
<tr>
<td>Appendix. Index</td>
<td>105</td>
</tr>
</tbody>
</table>
Math prerequisites

1. Convergence

1.1. Convergence of numbers. Recall that for a sequence of numbers, \( \lim_{n \to \infty} a_n = L \) means that ...

\[ \sum_{n=1}^{\infty} a_n = L \] means that ...

**Theorem 0.1.** If a sequence of real numbers \( \{a_n\} \) is bounded and increasing, then \( \lim_{n} a_n = \sup_{n \in \mathbb{N}} a_n \).

For unbounded increasing sequences we write \( \lim_{n} a_n = \infty \).

Recall that for a sequence of numbers \( a_n \),

\[ \limsup_{n \to \infty} a_n = \lim_{n \to \infty} \sup_{k \geq n} a_k \quad \text{and} \quad \liminf_{n \to \infty} a_n = \lim_{n \to \infty} \inf_{k \geq n} a_k. \]

**Remark 0.1.** It is clear that \( \liminf_{n \to \infty} a_n \leq \limsup_{n \to \infty} a_n \). The equality holds iff the limit \( \lim_{n \to \infty} a_n \) exists as an extended number in \( [-\infty, \infty] \).

Similarly, for a sequence of functions \( f_n : \Omega \to \mathbb{R} \), we define functions \( f_* , f^* : \Omega \to \bar{\mathbb{R}} = \mathbb{R} \cup \{ -\infty, \infty \} \) by \( f_* = \liminf_{n \to \infty} f_n \) and \( f^* = \limsup_{n \to \infty} f_n \) pointwise.

We say that the sequence of functions \( \{f_n\} \) converges pointwise, if \( f_n(\omega) \) converges for all \( \omega \in \Omega \).

We say that the sequence of functions \( \{f_n\} \) converges uniformly over \( \Omega \) to \( f \), if \( \sup_{\omega \in \Omega} |f_n(\omega) - f(\omega)| \to 0 \).

2. Set theory

(i) For a set \( \Omega \), by \( 2^\Omega \) we denote the so called power set, i.e., the set of all subsets of \( \Omega \). We use upper case letters like \( A, B, C, \ldots \) for the subsets - some (but not all) will be interpreted as "events".

(ii) The empty set is \( \emptyset \) - in handwriting this needs to be carefully distinguished from the Greek letters \( \varphi \) or \( \Phi \).
(iii) We use $A \cup B$, for the union, $A \cap B$ for the intersection, $A^c$ or $A'$ for the complement. We do not use $A + B$ and $AB$ in this course!!!

(iv) We use $A \subset B$ for what some other books denote by $A \subseteq B$. Sometimes it will be convenient to write this as $B \supseteq A$. Collections of sets will be denoted by scripted letters, like $\mathcal{A}$ or $\mathcal{F}$. We will need to consider large collections of sets, as well as collections like $\mathcal{A} = \{A_1, A_2, \ldots\}$.

(v) For a family $\mathcal{A} = \{A_t : t \in T\}$ of subsets of $\Omega$ indexed by a set $T$, the union of all sets in $\mathcal{A}$ is the set of $\omega$ with the property that there exists a set $A_t \in \mathcal{A}$ such that $\omega \in A_t$. In symbols,

$$\bigcup_{t \in T} A_t = \{\omega \in \Omega : \exists t \in T \omega \in A_t\}$$

More concisely,

$$\bigcup_{A \in \mathcal{A}} A = \{\omega \in \Omega : \omega \in A \text{ for some } A \in \mathcal{A}\} = \{\omega \in \Omega : \exists A \in \mathcal{A} \omega \in A\}$$

Similarly, we define the intersection

$$\bigcap_{t \in T} A_t = \{\omega \in \Omega : \forall t \in T \omega \in A_t\}$$

In particular, for a countable collection of sets,

$$\bigcup_{n \in \mathbb{N}} A_n = \bigcup_{n=1}^{\infty} A_n = \{\omega : \omega \in A_n \text{ for some } n \in \mathbb{N}\}$$

$$\bigcap_{n \in \mathbb{N}} A_n = \bigcap_{n=1}^{\infty} A_n = \{\omega : \omega \in A_n \text{ for all } n \in \mathbb{N}\}$$

(vi) The notation for intervals is $(a,b) = \{x \in \mathbb{R} : a < x < b\}$, $[a,b) = \{x \in \mathbb{R} : a \leq x < b\}$ and similarly $(a,b]$ and $[a,b]$.

Theorem 0.2 (DeMorgan’s law).

(0.1) \hspace{1cm} \left(\bigcup_{t \in T} A_t\right)^c = \bigcap_{t \in T} A_t^c

Since $(A^c)^c = A$, formula (0.1) is equivalent to

(0.2) \hspace{1cm} \left(\bigcap_{t \in T} A_t\right)^c = \bigcup_{t \in T} A_t^c

2.1. Indicator functions and limits of sets. This has application to the so called indicator functions:

(0.3) \hspace{1cm} I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{otherwise} \end{cases}

Since

$$I_{A_n}(\omega) = \begin{cases} 0 & 1 \end{cases},$$
it is clear that
\[ \limsup_{n \to \infty} I_{A_n}(\omega) = \begin{cases} 0 \\
1 \end{cases}. \]
This means that \( \limsup_{n \to \infty} I_{A_n}(\omega) = I_{A^*}(\omega) \) for some set \( A^* \subset \Omega \).

For the same reasons, \( \liminf_{n \to \infty} I_{A_n} = I_{A^*} \) for some set \( A^* \subset \Omega \).

**Proposition 0.3.**

\[ A_* = \bigcup_{n \in \mathbb{N}} \bigcap_{k \geq n} A_k \quad \text{and} \quad A^* = \bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} A_k \]

**Proof.** This is Exercise 0.1.

The second set has probabilistic interpretation:

\[ A^* = \{ A_n \text{ occur infinitely often } \} = \{ A_n \text{ i. o. } \} \]

It is clear that \( A_* \subset A^* \). We say that \( \lim_n A_n \) exists if \( A_* = A^* \). Exercises 0.3 and 0.4 give examples of such limits.

**2.2. Cardinality.** Sets \( A, B \) have the same *cardinality* if there exists a one-to-one and onto function \( f : A \to B \). We shall say that a set \( A \) is countable if either \( A \) is finite, or it has the same cardinality as the set \( \mathbb{N} \) of natural numbers.

It is known that the set of all rational numbers \( \mathbb{Q} \) is countable while the interval \([0, 1] \subset \mathbb{R} \) is not countable.

**3. Compact set**

Recall that if \( K \) is compact if every sequence \( x_n \in K \) has a convergent subsequence (with respect to some metric \( d \)). Equivalently, from every open cover of \( K \) one can select a finite sub-cover. If \( K \) is compact and sets \( F_n \subset K \) are closed with non-empty intersections \( \bigcap_{n=1}^\infty F_k \neq \emptyset \) for all \( n \), then the infinite intersection \( \bigcap_{k=1}^\infty F_k \) is also non-empty.

**Theorem 0.4.** *Closed bounded subsets of \( \mathbb{R}^k \) are compact.*

**4. Riemann integral**

Function \( f : [a, b] \to \mathbb{R} \) is Riemann-integrable, with integral \( S = \int_a^b f(x)dx \), if for every \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that

\[ \left| S - \sum_i f(x_i)|I_j| \right| < \varepsilon \]

for every partition of \([a, b]\) into sub-intervals \( I_j \) of length \( |I_j| < \delta \) and every choice of \( x_j \in I_j \). Every Riemann-integrable function is Lebesgue-integrable over \([a, b]\).

It is known that continuous functions are Riemann-integrable.

In calculus, the improper integral \( \int_0^\infty f(x)dx \) is defined as the limit \( \lim_{t \to \infty} \int_0^t f(x)dx \). This is not the same as the Lebesgue integral over \([0, \infty)\).
5. Product spaces

The set $\mathbb{R}^\infty$ of all infinite sequences of real numbers is a metric space with the distance

\[
d((a_n), (b_n)) = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{|a_n - b_n|}{1 + |a_n - b_n|}.
\]

In particular, a sequence of points $a_n \in \mathbb{R}^\infty$ converges to $b$ if every coordinate converges. This is the pointwise convergence of functions, with a sequence $(a_n)$ identified with function $a : \mathbb{N} \to \mathbb{R}$.

6. Taylor polynomials and series expansions

See Chapter 12 Section 1 (page 93).

7. Complex numbers

See Chapter 12 Section 1 (page 93).

---

**Exercise 0.1.** Prove Proposition 0.3.

**Solution:** Since $I_{A_n} \leq 1$, \( \limsup I_{A_n}(\omega) = 1 \) iff $\forall \varepsilon > 0 \forall n \exists k \geq n : I_{A_k}(\omega) \geq 1 - \varepsilon$. Noting that $I_{A_k}(\omega) \geq 1 - \varepsilon$ is the same as $\omega \in A_k$, the above is equivalent to $\forall n \exists k \geq n : \omega \in A_k$ which is the same as $\omega \in \bigcap_n \bigcup_{k \geq n} A_k$.

**Exercise 0.2.** Suppose $B, C$ are subsets of $\Omega$ and

\[
A_n = \begin{cases} B & \text{if } n \text{ is even} \\ C & \text{if } n \text{ is odd} \end{cases}
\]

Identify the sets $A_* = \bigcup_{n \in \mathbb{N}} \bigcap_{k \geq n} A_k$ and $A^* = \bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} A_k$.

**Solution:** The answer is $A_* = B \cap C$, $A^* = B \cup C$.

**Exercise 0.3.** Suppose $A_1 \supset A_2 \supset A_n \supset \ldots$. Show that $\lim_n A_n$ exists (and describe the limit).

**Solution:** The answer is $\lim_n A_n = \bigcap_{n=1}^{\infty} A_n$.

**Exercise 0.4.** Suppose $A_1 \subset A_2 \subset A_n \subset \ldots$. Show that $\lim_n A_n$ exists (and describe the limit).

**Solution:** The answer is $\lim_n A_n = \bigcup_{n=1}^{\infty} A_n$. 

---
Events and Probabilities


1. Elementary and semi-elementary probability theory

The standard model of probability theory is the triplet \((\Omega, \mathcal{F}, P)\), where \(\Omega\) is a set, \(\mathcal{F} \subset 2^\Omega\), and \(P\) is a function \(P : \mathcal{F} \to [0,1]\). The set \(\Omega\) is called sometimes a sample space or a probability space. Sets \(A \in \mathcal{F}\) are called events, and the number \(P(A)\) is called probability of event \(A\). We say that event \(A\) occurred, if \(\omega \in A\), and we interpret \(P(A)\) as the “likelihood” that event \(A\) occurred.

1.1. Field of events. It is natural to expect that the events form a field.

Definition 1.1. A class \(\mathcal{F}\) of subsets of \(\Omega\) is a field if:

(i) \(\Omega \in \mathcal{F}\)

(ii) if \(A \in \mathcal{F}\) then \(A^c \in \mathcal{F}\)

(iii) if \(A, B \in \mathcal{F}\) then \(A \cup B \in \mathcal{F}\)

By induction, if \(A_1, A_2, \ldots, A_n \in \mathcal{F}\) then \(A_1 \cup \cdots \cup A_n = \bigcup_{1 \leq j \leq n} A_j \in \mathcal{F}\). By DeMorgan’s law (Theorem 0.2), a field is also closed under intersections, \(\bigcap_{1 \leq j \leq n} A_j \in \mathcal{F}\). In particular, we can replace axiom (iii) by

(iii’) if \(A, B \in \mathcal{F}\) then \(A \cap B \in \mathcal{F}\)

Example 1.1. The class \(\mathcal{B}_0\) of finite unions of disjoint left-open right-closed subintervals of \((0,1]\), is a field.

Proof. \((0,1] \in \mathcal{B}_0\). If \(A = \bigcup_{j=1}^K (a_j, b_j]\) with \(a_1 < b_1 \leq a_2 < b_2 \cdots \leq a_K < b_K\) then \(A^c = (0, a_1] \cup (b_1, a_2] \cup \cdots \cup (b_{K-1}, a_K] \cup (b_K, 1]\), where some of the intervals might be empty.

If \(A = \bigcup_j I_j\) and \(B = \bigcup_k J_k\) then \(A \cup B = \bigcup_{j,k} I_j \cap J_k\) and intersections \(I_j \cap J_k\) are disjoint, possibly empty, intervals of the form \((a,b]\).

Similarly, for \(\Omega := (0,1] \times (0,1]\), the set of finite unions of rectangles \((a,b] \times (c,d]\subset\) is a field.
Remark 1.1. The field $F_0$ in Example 1.1 is the smallest field of subsets of $(0, 1]$ that contains intervals $(a, b]$.

Question 1.1. What is the smaller field of subsets of $\mathbb{R}$ that contains all half-lines $(-\infty, b]$?

1.2. Finitely additive probabilities. We want to assign the number $P(A)$, as a “measure” of the likelihood that the event $A$ occurred.

Definition 1.2. Let $F$ be a field. A function $P : F \to \mathbb{R}$ is a finitely-additive probability measure if it satisfies the following conditions

(i) $0 \leq P(A) \leq 1$ for $A \in F$
(ii) $P(\emptyset) = 0$, $P(\Omega) = 1$.
(iii) If $A, B \in F$ are disjoint, then $P(A \cup B) = P(A) + P(B)$.

A function $P : F \to \mathbb{R}$ is a probability measure on $F$ if it is finitely-additive and satisfies the following continuity condition

(iv) If $A_1 \supset A_2 \supset \ldots$ are sets in $F$ and $\bigcap_k A_k = \emptyset$, then $\lim_{n \to \infty} P(A_n) = 0$

Proposition 1.1 (Elementary properties). Suppose $P$ is a finitely additive probability measure on the field $F$ of subsets of $\Omega$. For $A, B \in F$ we have

(i) $B \subset A$ implies $P(A) \leq P(B)$
(ii) $P(A^c) = 1 - P(A)$
(iii) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Proof. (1) follows from $A = B \cup (A \setminus B)$ by finite additivity. The proof gives $P(B \setminus A) = P(B) - P(A)$

(2) follows from $\Omega = A \cup A^c$

(3) is a special case of Exercise 1.1. For example, using formula from the proof of (1): $P((A \cup B) \setminus A) = P(A \cup B) - P(A)$. Since $(A \cup B) \setminus A = (A \cup B) \cap A^c = B \cap A^c = B \setminus (A \cap B)$ using the formula again, $P((A \cup B) \setminus A) = P(B) - P(A \cap B)$. So $P(A \cup B) - P(A) = P(B) - P(A \cap B)$ and the formula follows. (There are numerous other proofs!)

Remark 1.2 (Probability measures are countably additive). Axioms (iii) and (iv) are often combined together into countable additivity,

(iii+) If $A_1, A_2, \ldots, \in F$ are pairwise disjoint and $\bigcup_k A_k \in F$, then $P(\bigcup_k A_k) = \sum_{k=1}^{\infty} P(A_k)$.

Other equivalent versions of continuity or countable additivity are:

(iv') If $A_1 \subset A_2 \subset \ldots$ are sets in $F$ and $\bigcup_k A_k \in F$, then $P(\bigcup_k A_k) = \lim_{n \to \infty} P(A_n)$
(iv") If $A_1 \supset A_2 \supset \ldots$ are sets in $F$ and $\bigcap_k A_k \in F$, then $P(\bigcap_k A_k) = \lim_{n \to \infty} P(A_n)$

(Recall Theorem 0.2.)

Example 1.2. Let $\Omega = \mathbb{N}$ and $F$ consist of all subsets $A \subset \mathbb{N}$ such that the limit $\lim_{n \to \infty} \#(A \cap \{1, \ldots, n\})/n$ exists. Then $P : F \to [0, 1]$ defined by $P(A) = \lim_{n \to \infty} \#(A \cap \{1, \ldots, n\})/n$ is a finitely additive probability measure. However, Exercise 1.8 says that $P$ is not continuous.

Constructions of finitely additive continuous measures are somewhat more involved.
1. Elementary and semi-elementary probability theory

1.2. Example: Lebesgue measure on the unit interval. In this example we consider \( \Omega = (0,1] \) and the field \( \mathcal{B}_0 \) from Example 1.1.

For \( A = \bigcup_{k=1}^{n} I_k \in \mathcal{B}_0 \) with disjoint \( I_k \), define \( \lambda(A) = \sum_{k=1}^{n} |I_k| \).

**Theorem 1.2.** \( \lambda \) is a well defined (continuous) probability measure on the field \( \mathcal{B}_0 \).

**Proof.** Since the representation \( A = \sum_{k=1}^{n} I_k \in \mathcal{B}_0 \) is not unique, we need to make sure that \( \lambda \) is well defined. Write \( A = \bigcup_{k} I_k = \bigcup_{j} J_j \) as the finite sums of disjoint intervals. Then \( I_k = I_k \cap A = \bigcup_{j} I_k \cap J_j \) so by finite additivity \( \sum_{k} |I_k| = \sum_{k} \sum_{j} |I_k \cap J_j| \) and similarly \( \sum_{j} |J_j| = \sum_{j} \sum_{k} |I_k \cap J_j| \). This shows that \( \sum_{j} |J_j| = \sum_{k} |I_k| \), so \( \lambda(A) \) is indeed well defined.

To prove continuity, we proceed by contrapositive. Suppose that \( A_n \supset A_{n+1} \) are such that \( \lambda(A_n) > \delta > 0 \). Choose \( B_n \subset K_n \subset A_n \) such that \( \lambda(A_n) - \lambda(B_n) < \delta/2^n \) and \( K_n \) is compact. Then \( P(A_n) - P(B_1 \cap \cdots \cap B_n) = P(\bigcup_{k=1}^{n} A_n \setminus B_k) \leq P(\bigcup_{k=1}^{n} A_k \setminus B_k) \leq \sum_{k=1}^{\infty} \delta/2^k = \delta/2 \). So \( P(B_1 \cap \cdots \cap B_n) \geq \delta/2 > 0 \), and \( K_1 \cap \cdots \cap K_n \neq \emptyset \). Thus, \( \bigcap_{n=1}^{\infty} A_n \supset \bigcap_{n=1}^{\infty} K_n \neq \emptyset \).

Construction of Lebesgue measure has the following generalization. Suppose \( \mathcal{B}_0 \subset 2^\mathbb{R} \) is a field consisting of finite unions of intervals \(( -\infty, b], (a, b], (b, \infty) \).

**Theorem 1.3** (Lebesgue). If \( P \) is a (continuous) probability measure on \( \mathcal{B}_0 \) then the function \( F(x) := P(( -\infty, x]) \) has the following properties:

(i) \( F \) is non-decreasing

(ii) \( \lim_{x \to -\infty} F(x) = 0 \)

(iii) \( \lim_{x \to \infty} F(x) = 1 \)

(iv) \( F \) is right-continuous, i.e., \( \lim_{y \uparrow x} F(y) = F(x) \)

Conversely, if \( F \) is a function with properties (i)-(iv) then there exists a (unique) continuous probability measure \( P \) on \( \mathcal{B}_0 \) such that \( F(x) := P(( -\infty, x]) \) for all \( x \in \mathbb{R} \).

**Sketch of the proof.** The proof of (iv) may require some care: For rational \( r \downarrow x \), we have \( (0, x] = \bigcap_{r > x} (0, r] \). Any real \( y \) lies between two rational numbers.

The proof of converse has two parts. For \( A \in \mathcal{B}_0 \) given by \( A = \bigcup_{j=1}^{n} (a_j, b_j] \), the definition \( P(A) = \sum_{j=1}^{n} (F(b_j) - F(a_j)) \) does not depend on the representation. (Here, we set \( F(-\infty) = 0 \) and \( F(\infty) = 1 \).) Uniqueness is an obvious consequence of finite additivity.

Then we need to verify continuity. This proof can proceed similarly to the proof of Theorem 1.2. (See hints for Exercise 1.13.)

A generalization of the above construction is based on the concept of a semi-algebra.

**Definition 1.3.** A collection \( S \) of subsets of \( \Omega \) is called semi-algebra, or a semi-ring, if

(i) \( \emptyset \in S \)

(ii) \( S \) is closed under intersections, i.e. if \( A, B \in S \) then \( A \cap B \in S \)

(iii) If \( A, B \in S \) then \( B \setminus A \) is a finite union of sets in \( S \).
The main (motivating) example of a semi-algebra is the family of rectangles in $\mathbb{R}^2$, and more generally, in $\mathbb{R}^d$.

The following is a version of [Durrett, Theorem 1.1.4] adapted to probability measures.

**Theorem 1.4.** Let $S$ be a semi-algebra. Suppose that $P : S \to [0,1]$ is additive, countably subadditive, i.e., if $A = \bigcup_{n=1}^{\infty} A_n$ is in $S$ for pairwise disjoint sets $A_n \in S$ then $P(A) \leq \sum_{n=1}^{\infty} P(A_n)$. If $P(\emptyset) = 0$, then $P$ has a unique extension onto the field $F$ generated by $S$, and this extension is continuous.

**Proof.**

**Question 1.2.** Why not to require uncountable continuity of probability measures?

(Compare Exercise 1.4 and Exercise 1.4.)

2. Sigma-fields

Given an infinite sequence $\{A_n\}$ of events, it is convenient to allow also more complicated events such as $A^* = \bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} A_k$ that events $A_k$ occur infinitely often. This motivates the following.

**Definition 1.4.** A class $F$ of subsets of $\Omega$ is a $\sigma$-field if it is field and if it is also closed under the formation of countable unions:

(iii+) If $A_1, A_2, \ldots \in F$ then $\bigcup_{n \in \mathbb{N}} A_n \in F$.

Note that (iii+) implies (iii) because we can take $A_1 = A$ and $A_n = B$ for other $n$.

By an application of DeMorgan’s law (Theorem 0.2), (iii+) can be replaced by

If $A_1, A_2, \ldots \in F$ then $\bigcap_{n \in \mathbb{N}} A_n \in F$.

Clearly, the power set $2^\Omega$ is the largest possible $\sigma$-field. We will often consider smallest $\sigma$-fields that contain some collections of sets of our interest.

**Proposition 1.5.** Suppose $A$ is a collection of subsets of $\Omega$. There exist a unique $\sigma$-field $F$ with the following properties:

(i) $A \in A$ implies $A \in F$. That is, $A \subset F$.

(ii) If $G$ is a $\sigma$-field such that $A \subset G$ then $F \subset G$.

We write $F = \sigma(A)$, and call $F$ the sigma-field generated by $A$.

**Proof.** Uniqueness is a consequence of (2)

To show that $F$ exists, consider a set $M$ of all sigma-fields $G$ with the property that $A \subset G$. Since $2^\Omega \in M$, this is a nonempty family of sets. Define

$$F = \bigcap_{G \in M} G.$$ 

Then $F$ is a sigma-field with the required properties, because the intersection of sigma-fields is a sigma-field (can you verify this?).

**Definition 1.5.** The Borel sigma-field is the sigma field generated by all open sets.
Proposition 1.6. For $\Omega = \mathbb{R}$, the Borel sigma-field $\mathcal{B}_1$ is generated by the intervals $\{(a,b) : a < b\}$.

For $\Omega = \mathbb{R}^d$, the Borel sigma-field $\mathcal{B}_d$ is generated by the rectangles $\prod_{k=1}^d (a_k, b_k)$.

Note that Borel-field $\mathcal{B}_0 \subset 2^\mathbb{R}$ generated by all intervals $(a, b]$ consist of finite unions of intervals $(-\infty, b], (a, b], (b, \infty)$.

2.1. Probability measures.

Definition 1.6. If $\mathcal{F}$ is a $\sigma$-field of subsets of $\Omega$ and $P$ is a probability measure on $\mathcal{F}$, then the triple $(\Omega, \mathcal{F}, P)$ is called a probability space. The sets $A \in \mathcal{F}$ are called events.

Example 1.3. Let $\mathcal{F} = 2^\Omega$ and fix $\omega_0 \in \Omega$. Then $P(A) = I_A(\omega_0)$ is a probability measure, sometimes called the point mass, denoted by $\delta_{\omega_0}$.

Since a convex combination of probability measures is a probability measure, another example of a probability measure is $P = \frac{1}{2} \delta_{\omega_0} + \frac{1}{2} \delta_{\omega_1}$.

Example 1.4 (Discrete probability space). Let $\mathcal{F}$ be the $\sigma$-field of all subsets of a countable $\Omega = \{\omega_1, \omega_2, \ldots\}$ Suppose $p_1, p_2, \ldots$ is a sequence of nonnegative numbers such that $\sum_{k=1}^\infty p_k = 1$. Define

$$P(A) = \sum_{k : \omega_k \in A} p_k$$

Then $P$ is a probability measure.

Proof. This is not entirely trivial. If $A = \sum_j A_j$ with disjoint sets then $P(A) = \sum_{\omega_i \in \bigcup A_j} p_i$ does not depend on the order of summation, and equals to the iterated series $\sum_{j=1}^\infty \sum_{i : \omega_i \in A_j} p_i$. \qed

Example 1.5 (Discrete probability measure). Let $\mathcal{F}$ be the $\sigma$-field of all subsets of an infinite set $\Omega$. Suppose $\omega_1, \omega_2, \ldots \in \Omega$ are fixed, and $p_1, p_2, \ldots$ is a sequence of nonnegative numbers such that $\sum_{k=1}^\infty p_k = 1$. Define

$$P(A) = \sum_{k : \omega_k \in A} p_k$$

Then $P$ is a probability measure.

The numbers $p_k$ are sometimes called the “probability mass function”, or the “probability density function” (as this is the density with respect to the counting measure on points $\{x_1, x_2, \ldots\}$).

More generally, if $P_1, P_2, \ldots$ is a sequence of probability measures on the common field or $\sigma$-field $\mathcal{F}$ and $p_1, p_2, \ldots$ is a sequence of nonnegative numbers such that $\sum_{k=1}^\infty p_k = 1$, then $Q(A) = \sum_{k=1}^\infty p_k P_k(A)$ is also a probability measure on $\mathcal{F}$. The probabilistic interpretation is that we have a sequence of experiments described by probability measures $P_k$ and we want to model a new experiment with additional randomization where the $k$-th experiment is performed with probability $p_k$.

2.1.1. Examples of discrete distributions on $\mathbb{R}$.

Example 1.6 (Binomial distribution). For fixed integer $n$ and $0 < p < 1$, take $x_k = k$ and

$$p_k = \binom{n}{k} p^k (1-p)^{n-k}, \; k = 0, \ldots n.$$  

The binomial formula shows that $\sum_{k=0}^n p_k = 1$. Notation: $Bin(n,p)$
Example 1.7 (Poisson distribution). For \( \lambda > 0 \) take \( x_k = k \) and

\[
p_k = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, \ldots
\]

Notation: \( \text{Poiss}(\lambda) \)

Example 1.8 (Polya’s distribution). For \( r > 0 \) and \( 0 < p < 1 \) take \( x_k = k \) and

\[
p_k = \frac{\Gamma(r + k)}{k! \Gamma(r)} (1 - p)^r p^k, \quad k = 0, 1, \ldots
\]

Notation: \( \text{NB}(p, r) \)

Required Exercises

Exercise 1.1. Prove the inclusion-exclusion formula\(^1\)

\[
(1.1) \quad P(\bigcup_{j=1}^{n} A_j) = \sum_{j=1}^{n} P(A_j) - \sum_{1 \leq j_1 < j_2 \leq n} P(A_{j_1} \cap A_{j_2}) + \sum_{1 \leq j_1 < j_2 < j_3 \leq n} P(A_{j_1} \cap A_{j_2} \cap A_{j_3}) - \cdots + (-1)^{n-1} P(A_1 \cap A_2 \cap \cdots \cap A_n)
\]

Does the proof use countable additivity?

Solution: Proceed by induction. Countable additivity is not used in this proof.

Exercise 1.2 (statistics). Prove Boole’s inequality\(^2\) \( P(\bigcup_{j=1}^{n} A_j) \leq \sum_{j=1}^{n} P(A_j) \).

Solution: Note that \( P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B) \). Now proceed by induction, with \( A = \bigcup_{j=1}^{n} A_j \) and \( B = A_{n+1} \) for the induction step.

Exercise 1.3. For a probability space \((\Omega, \mathcal{F}, P)\), if \( B_1, B_2, \ldots \) is a sequence of events such that \( \sum_{k=1}^{n} P(B_k) > n - 1 \), show that \( P(\bigcap_{k=1}^{n} B_k) > 0 \).

Solution: Proceed by contrapositive. Suppose \( 1 = P(\bigcup_{k=1}^{n} B_k^c) \leq \sum_{k=1}^{n} P(B_k^c) = n - \sum_{k=1}^{n} P(B_k) \). Thus \( \sum_{k=1}^{n} P(B_k) \leq n - 1 \).

Exercise 1.4. For a probability space \((\Omega, \mathcal{F}, P)\), suppose \( \{B_n : n \in \mathbb{N}\} \) are events with \( P(B_n) = 1 \). Show that

\[
P\left(\bigcap_{n=1}^{\infty} B_n\right) = 1
\]

\(^1\)This can also be written as

\[
\Pr\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{M=1}^{n} (-1)^{M-1} \sum_{1 \leq j_1 < j_2 < \cdots < j_M \leq n} \Pr(A_{j_1} \cap A_{j_2} \cap \cdots \cap A_{j_M})
\]

\(^2\)Named for G. Boole (1815-1864). Used for Bonferroni’s correction in multiple hypothesis testing.
Solution: 1 ≥ P(\(\bigcap_{n=1}^{\infty} B_n\)) = 1 - P(\(\bigcup_{n=1}^{\infty} B_n^c\)) ≥ 1 - \(\sum_{n=1}^{\infty} P(B_n^c)\) = 1. In other words, we prove that the complement of \(\bigcap_{n=1}^{\infty} B_n\) has zero probability.

Exercise 1.5. Suppose that \(\Omega = \mathbb{N}\) and for \(n \in \mathbb{N}\) let \(F_n\) be the \(\sigma\)-field generated by the collection of one-point sets \(A_n = \{\{1\}, \{2\}, \ldots, \{n\}\}\). Show that \(F_n \subset F_{n+1}\) and that \(F := \bigcup_n F_n\) is a field but not a \(\sigma\)-field.

Exercise 1.6. Show that measure \(P\) in Example 1.2 is additive but not continuous. (For the second statement, find \(A_n \in F\) such that \(A_n \supset A_{n+1}\) and \(\bigcap_{n=1}^{\infty} A_n = \emptyset\), but \(P(A_n) = 1\).)

Exercise 1.7. Without using Proposition 1.6, show that open intervals \((a, b)\) and closed intervals \([a, b]\) are in the sigma-field generated by the intervals \((a, b)\) in \(\mathbb{R}\). (Compare Example 1.1.)

---

**Additional Exercises**

Exercise 1.8. Without using Proposition 1.6, show that the open triangle \(T = \{(x, y) : x > 0, y > 0, x + y < 1\}\) is in the sigma-field generated by the rectangles \((a, b) \times (c, d)\) in \(\mathbb{R}^2\).

Exercise 1.9. Suppose that \(F_n\) are fields satisfying \(F_n \subset F_{n+1}\). Show that \(\bigcup_n F_n\) is a field.

Exercise 1.10. Suppose \(P\) is a finitely additive measure on a field \(F\). Show that if \(A_1, \ldots, A_n, \ldots\) are disjoint then the series \(\sum_{n=1}^{\infty} P(A_n)\) converges.

Solution: For additive measures, \(\sum_{j=1}^{n} P(A_j) = P(\bigcup_{j=1}^{n} A_j) \leq 1\). Thus the sequence of partial sums is increasing and bounded, so it converges to a number in \([0, 1]\).

Exercise 1.11. Prove that continuous finitely-additive probability measure on a field is countably additive. That is, show that property (iii+) of Remark 1.2 follows from the axioms (i)-(iv) of Definition 1.2.

Exercise 1.12. If \(P_1, P_2, \ldots\), is a sequence of continuous probability measures on the field \(F\) and \(p_1, p_2, \ldots\) is a sequence of nonnegative numbers such that \(\sum_{k=1}^{\infty} p_k = 1\), show that \(Q(A) = \sum_{k=1}^{\infty} p_k P_k(A)\) is also continuous.

Exercise 1.13. Suppose \(\Omega\) is a metric space and \(F\) is a field of subsets of \(\Omega\). Suppose that \(P\) is a finitely additive probability measure on \(F\).

Let's say that \(P\) is a tight probability measure if for every \(A \in F\) with \(P(A) > 0\) and \(\varepsilon > 0\) there exist \(B \in F\) and a compact set \(K\) such that \(B \subset K \subset A\) and \(P(A) < P(B) + \varepsilon\).

(i) In the setting of Theorem 1.2, show that the Lebesgue measure on \(B_0\) is tight.

(ii) Show that a tight finitely additive probability measure is countably additive.

Hint: Proceed by contrapositive!\(^3\)

---

\(^3\) Plan of proof: Suppose \(A_1 \supset A_2 \supset \ldots\) are sets in \(F\) such that there exists \(\delta > 0\) with \(P(A_n) > \delta\) for all \(n\). Using tightness, we can find compact sets \(K_1, K_2, \ldots\) and sets \(B_j \in F\) such that \(B_j \subset K_j\) and \(B_1 \cap B_2 \cap \cdots \cap B_n\) has positive probability. In fact, we can find such \(B_j\) with \(P(B_1 \cap B_2 \cap \cdots \cap B_n)\) of at least \(\delta(1 - \sum_{j=1}^{\infty} 1/2^j) > 0\).

Since every finite intersection \(K_1 \cap K_2 \cap \cdots \cap K_n\) contains \(B_1 \cap B_2 \cap \cdots \cap B_n\), we see that \(\bigcap_n K_n\) is nonempty. So \(\bigcap_n A_n\) cannot be empty.
Exercise 1.14 (Compare Exercise 1.5). Let \( \Omega \) be an infinite set. Consider the following classes of subsets of \( \Omega \):

\[
\mathcal{F}_n = \{ A \subset \Omega : A \text{ has at most } n \text{ elements or } A^c \text{ has at most } n \text{ elements} \}
\]

Then we have the following facts:

- \( \mathcal{F}_n \subset \mathcal{F}_{n+1} \)
- \( \mathcal{F}_0 \) is a \( \sigma \)-field
- For \( n \geq 1 \), class \( \mathcal{F}_n \) is not a field
- \( \bigcup_n \mathcal{F}_n \) is a field but it is not a \( \sigma \)-field.

Exercise 1.15 (Compare Exercise 1.4). Suppose \( \{ B_t : t \in T \} \) are events with \( P(B_t) = 1 \). Give an example where \( \bigcap_{t \in T} B_t = \emptyset \) so \( P(\bigcap_{t \in T} B_t) = 0 \). Hint: Lebesgue measure on Borel \((0,1]\)

Exercise 1.16. Let \( \Omega \) be a nonempty set and \( \mathcal{C} \) be the class of one-element sets. Show that if \( A \in \sigma(\mathcal{C}) \) then either \( A \) is countable or \( A^c \) is countable.

Exercise 1.17. Suppose \( \mathcal{A} \) and \( \mathcal{B} \) are \( \sigma \)-fields of subsets of \( \Omega \). Let \( \mathcal{F} = \mathcal{A} \land \mathcal{B} \) be the smallest \( \sigma \)-field containing both \( \mathcal{A} \) and \( \mathcal{B} \). Show that \( \mathcal{F} \) is generated by sets of the form \( A \cap B \) where \( A \in \mathcal{A} \) and \( B \in \mathcal{B} \).

Exercise 1.18. The field \( \mathcal{F}(\mathcal{A}) \) generated by a class \( \mathcal{A} \) of subsets of \( \Omega \) is defined as the intersection of all fields in \( \Omega \) containing all of the sets in \( \mathcal{A} \).

- Show that \( \mathcal{F}(\mathcal{A}) \) is indeed a field, that \( A \subset \mathcal{F}(\mathcal{A}) \) and that \( \mathcal{F}(\mathcal{A}) \) is minimal in the sense that if \( \mathcal{G} \) is a field and \( A \subset \mathcal{G} \) then \( \mathcal{F}(\mathcal{A}) \subset \mathcal{G} \).
- Show that if \( \mathcal{A} \) is nonempty then \( \mathcal{F}(\mathcal{A}) \) is the class of sets of the form \( \bigcup_{j=1}^m B_j \) where sets \( B_j \) are disjoint and are of the form \( B = \bigcap_{i=1}^n A_i \) where either \( A_i \in \mathcal{A} \) or \( A_i^c \in \mathcal{A} \).

Exercise 1.19. For \( \Omega = (0,1] \) and any \( A \subset \Omega \) define

\[
P^* = \inf \left\{ \sum_k |B_k| : B_k \in \mathcal{B}_0, \bigcup_{k=1}^{\infty} B_k \supset A \right\}
\]

where \( |B| \) is the sum of lengths of intervals forming \( B \).

(i) Show that \( 0 \leq P^*(A) \leq 1 \)
(ii) Show that \( P^*(A \cup B) \leq P^*(A) + P^*(B) \)
(iii) Show that \( P^*|_{\mathcal{B}_0} = \lambda \), the Lebesgue measure from Theorem 1.2.
(iv) Show that \( P^*(\{x\}) = 0 \).
Probability measures

Abstract. Outer measure.
Construction of a measure. \(\lambda\)-systems, \(\pi\)-systems. Dynkin’s theorem.
Probability measures on \(\mathbb{R}\) and \(\mathbb{R}^n\).

The main result

1. Existence

**Theorem 2.1** (Caratheodory). A (countably additive) probability measure on a field has an extension to the generated \(\sigma\)-field

**Proof of Theorem 2.1.** Let \(\mathcal{F}_0\) be a field of subsets of \(\Omega\) and let \(P_0\) be a probability measure on \(\mathcal{F}_0\). Put \(\mathcal{F} = \sigma(\mathcal{F}_0)\).

For each subset \(A\) of \(\Omega\), define the outer measure

\[
P^*(A) = \inf \left\{ \sum_{n=1}^{\infty} P_0(A_n) : A_n \in \mathcal{F}_0, \bigcup_{n=1}^{\infty} A_n \supset A \right\}
\]

**Question 2.1.** Can \(P^*(A) = \infty\)?

Let’s first check that \(P^*\) is a genuine extension of \(P_0\) to a set function defines on all subsets of \(\Omega\).

**Proposition 2.2.** \(P^*\) and \(P\) agree on \(\mathcal{F}_0\).

**Proof.** (Omitted in 2018)
Suppose \(A \in \mathcal{F}_0\). Clearly, \(P^*(A) \leq P(A)\) as an infimum. Given \(\varepsilon > 0\) choose \(A_n \in \mathcal{F}_0\) such that \(A \subset \bigcup_n A_n\) and \(P^*(A) + \varepsilon > \sum_n P(A_n)\). Then \(A = \bigcup_n (A_n \cap A)\) and \(A_n \cap A \in \mathcal{F}_0\), so by countable subadditivity \(P(A) \leq \sum_n P(A \cap A_n) \leq \sum_n P(A_n) < P^*(A) + \varepsilon\). Since \(\varepsilon > 0\) is arbitrary, this shows that indeed \(P(A) = P^*(A)\).

In general, \(P^*\) is not additive, at least not on \(2^\Omega\), but it still has a number of nice properties.

**Proposition 2.3.** The outer probability has the following properties:
(i) $P^∗(\emptyset) = 0$;
(ii) $P^∗(A) ≥ 0$
(iii) $A ⊂ B$ implies $P^∗(A) ≤ P^∗(B)$
(iv) $P^∗(\bigcup_n A_n) ≤ \sum_n P^∗(A_n)$

**Proof.** (Omitted in 2018)
Without loss of generality we may assume $\sum_n P^∗(A_n) < \infty$. To prove (4), choose sets $B_{nk} \in \mathcal{F}_0$ such that $A_n \subset \bigcup_k B_{nk}$ and $P^∗(A_n) ≤ \varepsilon/2^n + \sum_k P_0(B_{nk})$. Then $\bigcup_n A_n \subset \bigcup_{n,k} B_{nk}$ and $P^∗(\bigcup_n A_n) ≤ \sum_{n,k} P_0(B_{nk}) = \sum_n \sum_k P_0(B_{nk}) ≤ \varepsilon + \sum_n P^∗(A_n)$.

Next, consider the class $\mathcal{M}$ of subsets $A$ of $\Omega$ with the property that

(2.2) \[ P^∗(A \cap E) + P^∗(A^c \cap E) = P^∗(E) \] for all $E \subset \Omega$

Note that by subadditivity of $P^*$, identity (2.2) is equivalent to inequality

(2.3) \[ P^∗(A \cap E) + P^∗(A^c \cap E) ≤ P^∗(E) \] for all $E \subset \Omega$

(Omitted in 2018)

**Lemma 2.4.** $\mathcal{M}$ is a field.

**Proof.** Clearly, $\Omega \in \mathcal{M}$ and if $A \in \mathcal{M}$ then $A^c \in \mathcal{M}$. It remains to show that if $A, B \in \mathcal{M}$ then $A \cap B \in \mathcal{M}$. Choose arbitrary $E \subset \Omega$.

$P^∗(E) = P^∗(A \cap E) + P^∗(A^c \cap E)$

$= P^∗(B \cap A \cap E) + P^∗(B^c \cap A \cap E) + P^∗(B \cap A^c \cap E) + P^∗(B^c \cap A^c \cap E)$

$≥ P^∗(B \cap A \cap E) + P^∗(((B^c \cap A) \cup (B \cap A^c) \cup (B^c \cap A^c)) \cap E)$

Now notice that

$(B^c \cap A) \cup (B \cap A^c) \cup (B^c \cap A^c) = ((B^c \cap A) \cup (B^c \cap A^c)) \cup ((B^c \cap A^c) \cup (B \cap A^c))$

$= B^c \cup A^c = (B \cap A)^c$
1. Existence

Lemma 2.5. If the sets $A_n \in \mathcal{M}$ are disjoint then

$$P^*\left(E \cap \bigcup_n A_n\right) = \sum_n P^*(E \cap A_n)$$

Note that we do not yet know whether $\bigcup A_n \in \mathcal{M}$, but the formula makes sense as $P^*$ is a function on $2^\Omega$.

Proof. Consider first the case of a finite number of sets $A_1, \ldots, A_n$. WLOG, $n \geq 2$. Given disjoint $A_1, A_2$, write $E \cap (A_1 \cup A_2) = (E \cap (A_1 \cup A_2) \cap A_1) \cup (E \cap (A_1 \cup A_2) \cap A_1^c)$ and use definition (2.2) with $E$ replaced by $E \cap (A_1 \cup A_2)$. This gives

$$P^*(E \cap (A_1 \cup A_2)) = P^*(E \cap (A_1 \cup A_2) \cap A_1) + P^*(E \cap (A_1 \cup A_2) \cap A_1^c)$$

Noting that $A_1, A_2$ are disjoint, we have $E \cap (A_1 \cup A_2) \cap A_1 = E \cap A_1$ and $E \cap (A_1 \cup A_2) \cap A_1^c = E \cap A_2$, so (2.4) hold for $n = 2$ sets.

Since $\mathcal{M}$ is a field, induction now shows that (2.4) hold for $n$ sets: $P^*(E \cap \bigcup_{k=1}^n A_k) = P^*(E \cap \left(\bigcup_{k=1}^{n-1} A_k\right))$.

Now we use monotonicity:

$$P^*\left(A \cap \bigcup_{k=1}^\infty A_k\right) \geq P^*\left(A \cap \bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n P^*(E \cap E_k)$$

and we let $n \to \infty$. The reverse inequality follows by subadditivity Proposition 2.3. \hfill \Box

Lemma 2.6. $\mathcal{M}$ is a $\sigma$-field. Set function $P : \mathcal{M} \to \mathbb{R}$ defined by $P(A) = P^*(A)$ is a probability measure.

Proof. By (2.4) used with $E = \Omega$, $P^*$ restricted to $\mathcal{M}$ is countably additive. However, we do not apriori know whether $\bigcup_n A_n \in \mathcal{M}$.

Suppose $A_1, A_2, \ldots$ are disjoint with $A = \bigcup_n A_n$. Then $F_n = \bigcup_{k=1}^n A_n \in \mathcal{M}$ (field), so $P^*(E) = P^*(E \cap F_n) + P^*(E \cap F_n^c)$. Applying (2.4) to the first term and monotonicity to the second term we get $P^*(E) \geq \sum_{k=1}^n P^*(E \cap A_k) + P^*(E \cap A^c)$. Now let $n \to \infty$ and use (2.4) to see that $P^*(E) \geq P^*(E \cap A) + P^*(E \cap A^c)$. Using again subadditivity, this shows that $A \in \mathcal{M}$.

Thus $\mathcal{M}$ is closed under the countable unions of disjoint sets. It remains to prove the following lemma. \hfill \Box
Lemma 2.7. If \( \mathcal{M} \) is a field and is closed under countable unions of disjoint sets then it is a \( \sigma \)-field.

Proof. Given a collection of sets \( \{A_n\} \) in \( \mathcal{M} \) construct sets \( B_n = A_n \setminus (A_1 \cup \cdots \cup A_{n-1}) \). It is clear that \( B_n \in \mathcal{M} \) are disjoint and \( \bigcup_n A_n = \bigcup_n B_n \). □

To conclude the proof, we need to show that \( \mathcal{F}_0 \subset \mathcal{M} \) so that \( \mathcal{F} = \sigma(\mathcal{F}_0) \subset \mathcal{M} \).

Lemma 2.8. \( \mathcal{F}_0 \subset \mathcal{M} \)

Proof. Let \( A \in \mathcal{F}_0 \). In view of subadditivity, we only need to verify that (2.3) holds for every \( E \in \Omega \).

Fix \( \varepsilon > 0 \) and let \( A_n \in \mathcal{F}_0 \) be such that \( E \subset \bigcup_n A_n \) and \( \varepsilon + P^*(E) > \sum_n P(A_n) \).

Since \( A \cap E \subset \bigcup_n (A \cap A_n) \) and \( A^c \cap E \subset \bigcup_n (A^c \cap A_n) \), we have \( P^*(A \cap E) + P^*(A^c \cap E) \leq \sum_n P(A \cap A_n) + \sum_n P(A^c \cap A_n) \). By finite additivity, \( P^*(A \cap E) + P^*(A^c \cap E) \leq \sum_n P(A_n) < P(E) + \varepsilon \). □

We can now complete the proof of Theorem. Since \( P \) and \( P^* \) coincide on \( \mathcal{M} \) and \( P^* \) and \( P_0 \) coincide on \( \mathcal{F}_0 \), we already know that \( P \) and \( P_0 \) coincide on \( \mathcal{F}_0 \). Since \( \mathcal{F}_0 \subset \mathcal{M} \), therefore it is also countably additive on a smaller \( \sigma \)-field \( \mathcal{F} \) generated by the field \( \mathcal{F}_0 \).

Remark 2.1. \( P_0(A) = 1 - P^*(A) \) is called the inner measure. [Billingsley] gives other expressions for the outer and inner measures which are of importance in the theory of stochastic processes.

(Do we want anything about approximations?)

Remark 2.2. For every \( A \in \mathcal{F} \) and every \( \varepsilon > 0 \), there exists \( B \in \mathcal{F}_0 \) such that \( P((A \setminus B) \cup (B \setminus A)) < \varepsilon \).

Proof. Fix \( A \in \mathcal{F} \). We use here that by the proof of Caratheodory’s theorem, \( P(A) = P^*(A) \). In view of (2.1), for every \( \varepsilon > 0 \) there exists a countable collection of disjoint sets \( B_j \in \mathcal{F}_0 \) such that \( A \subset \bigcup_{n=1}^\infty B_n \) and \( P(A) \leq P(\bigcup_{n=1}^\infty B_n) < P(A) + \varepsilon/2 \). And then there exists \( n \) such that \( P(\bigcup_{k=1}^n B_k) < P(\bigcup_{n=1}^\infty B_n) + \varepsilon/2 \). So with \( B = \bigcup_{k=1}^n B_k \) we have

\[
P((A \setminus B) \cup (B \setminus A)) \leq P(A \setminus B) + P(B \setminus A) \leq P\left(\bigcup_{n=1}^\infty B_n \setminus B\right) + P\left(\bigcup_{n=1}^\infty B_n \setminus A\right) < \varepsilon/2 + \varepsilon/2
\]

□

2. Uniqueness

This section is based on [Billingsley, Section 3].
Theorem 2.9. A (countably additive) probability measure on a field has a unique extension to the generated σ-field.

In view of Theorem 2.1, we only need to prove uniqueness. This is accomplished using some more theory, which extracts appropriate property of the field, and combines it with “natural property” of the sets that two measures coincide. This theory yields the proof on page 24.

2.1. Dynkin’s π-λ Theorem.

Definition 2.1. A class $\mathcal{P}$ of subsets of $\Omega$ is a π-system if

(π) $A, B \in \mathcal{P}$ implies $A \cap B \in \mathcal{P}$.

Examples of π-systems are

(i) $\{\emptyset\}$, which generates sigma-field ...
(ii) Family of intervals $(-\infty, a]$ with $a \in \mathbb{R}$, which generates Borel sigma-field $\mathcal{B}_{\mathbb{R}}$
(iii) Family $(-\infty, a] \times (-\infty, b]$, which generates Borel sigma field $\mathcal{B}_{\mathbb{R}^2}$
(iv) Family of sets $B_1 \times B_2 \times \cdots \times B_d \times \mathbb{R}^\infty$ with $B_j \in \mathcal{B}_{\mathbb{R}}$ which generates the Borel sigma field $\mathcal{B}_{\mathbb{R}^\infty}$.

Definition 2.2. A class $\mathcal{L}$ of subsets of $\Omega$ is a λ-system if

(λ1) $\Omega \in \mathcal{L}$.
(λ2) $A \in \mathcal{L}$ implies $A^c \in \mathcal{L}$.
(λ3) If $A_1, A_2, \ldots, A_n, \cdots \in \mathcal{L}$ are (pairwise) disjoint then $\bigcup_n A_n \in \mathcal{L}$.

Remark 2.3. From (λ1) and (λ2) we see that $\emptyset \in \mathcal{L}$. So if $A, B \in \mathcal{L}$ are disjoint then by (λ3) we get $A \cup B = A \cup B \cup \emptyset \cup \emptyset \cup \cdots \in \mathcal{L}$.

Of course, every field is a π-system, and every σ-field is a λ-system.

Lemma 2.10. A class of sets that is both a π-system and a λ-system is a σ-field.

Proof. Clearly, if $\mathcal{F}$ is a λ-system and a π system then it is a field. Suppose $A_n \in \mathcal{F}$. Then $B_n = A_n \setminus (A_1 \cup \cdots \cup A_{n-1}) = A_n \cap A_1^c \cap \cdots \cap A_{n-1}^c \in \mathcal{F}$, too. We note that $\bigcup_n A_n = \bigcup_n B_n \in \mathcal{F}$ as a disjoint sum. \qed
Lemma 2.11. Suppose \( \mathcal{P} \) is a \( \pi \)-system and \( \mathcal{L}_0 \) is the \( \lambda \)-system generated by \( \mathcal{P} \). Then \( \mathcal{L}_0 \) is a \( \sigma \)-field.

**Sketch of proof.** Because of Lemma 2.10, to show that \( \mathcal{L}_0 \) is a \( \sigma \)-field it is enough to show that it is a \( \pi \)-system. That is, we need to show that \( A, B \in \mathcal{L}_0 \) implies \( A \cap B \in \mathcal{L}_0 \).

This is done in two steps: first fix \( A \in \mathcal{P} \) and look at the collection \( C_A \) of all sets \( B \) such that \( A \cap B \in \mathcal{L}_0 \). This collection turns out to be a \( \lambda \)-system. Since \( \mathcal{P} \subset C_A \), we have \( \mathcal{L}_0 \subset C_A \). And this holds for any \( A \in \mathcal{P} \). This shows that if \( A \in \mathcal{P} \) and \( B \in \mathcal{L}_0 \) then \( A \cap B \in \mathcal{L}_0 \).

Now fix \( B \in \mathcal{L}_0 \) and look at the collection \( C_B \) of all sets \( A \) such that \( A \cap B \in \mathcal{L}_0 \). By the previous part, \( \mathcal{P} \subset C_B \). Again, \( C_B \) turns out to be a \( \lambda \)-system, so \( \mathcal{L}_0 \subset C_B \). This proves the lemma: for every \( B \in \mathcal{L}_0 \) and every \( A \in \mathcal{L}_0 \) we have \( A \cap B \in \mathcal{L}_0 \).

It remains to prove that the collections of sets \( C_A \) and \( C_B \) are \( \lambda \)-systems. This proof is omitted.

□

Theorem 2.12 (Dynkin’s \( \pi \)-\( \lambda \) Theorem). Suppose a \( \lambda \)-system \( \mathcal{L} \) includes a \( \pi \)-system \( \mathcal{P} \). Then \( \sigma(\mathcal{P}) \subset \mathcal{L} \).

**Proof.** Let \( \mathcal{L}_0 \) be a \( \lambda \)-system generated by \( \mathcal{P} \). Then \( \mathcal{P} \subset \mathcal{L}_0 \subset \mathcal{L} \). From Lemma 2.11 we know that \( \mathcal{L}_0 \) is a \( \sigma \)-field and it contains \( \mathcal{P} \). So \( \sigma(\mathcal{P}) \subset \mathcal{L}_0 \subset \mathcal{L} \).

□

Proposition 2.13. Let \( \mathcal{P} \) be a \( \pi \)-system and denote \( \mathcal{F} = \sigma(\mathcal{P}) \). Suppose \( P_1, P_2 \) are two probability measures on \( \mathcal{F} \) that agree on \( \mathcal{P} \). Then \( P_1 = P_2 \) (on \( \mathcal{F} \)).

**Proof.** Let \( \mathcal{L} \) be the family of all sets in \( \mathcal{F} \) on which \( P_1 \) and \( P_2 \) agree. Then \( \mathcal{L} \) is a \( \lambda \)-system. By Theorem 2.12 \( \mathcal{F} \subset \mathcal{L} \).

□

Proof of Theorem 2.9. A field \( \mathcal{F}_0 \) is a \( \pi \)-system. So if \( P_1(A) = P_2(A) \) for all \( A \in \mathcal{F}_0 \), then by Proposition 2.13 the same holds for all \( A \in \mathcal{F} = \sigma(\mathcal{F}_0) \).

□

3. Probability measures on \( \mathbb{R} \)

This is based on [Billingsley, Section 12] and [Durrett, Section 1.2].

**Definition 2.3.** \( F : \mathbb{R} \to \mathbb{R} \) is a cumulative distribution function, if

(i) \( F \) is non-decreasing: \( x < y \) implies \( F(x) \leq F(y) \)

(ii) \( \lim_{x \to \infty} F(x) = 0 \) and \( \lim_{x \to -\infty} F(x) = 1 \).
(iii) $F$ is right-continuous, $\lim_{x \to x_0^+} F(x) = F(x_0)$

Suppose that $P$ is a probability measure on the Borel subsets of $\mathbb{R}$. Consider a function $F : \mathbb{R} \to \mathbb{R}$ defined by $F(x) = P((-\infty, x])$. Then $F$ is a cumulative distribution function. (You should be able to supply the proof!)

The following is a combination of Lebesgue’s Theorem 1.3, with Caratheodory’s Theorem 2.1 and uniqueness Theorem 2.9.

**Proposition 2.14.** Every cumulative distribution function $F$ corresponds to a unique probability measure $P$ on the Borel sigma-field set of $\mathbb{R}$, such that $F(x) = P((-\infty, x])$.

**Proof.** Intervals of the form $(-\infty, a]$ form a $\pi$-system, and generate the Borel $\sigma$-field. So uniqueness follows from Theorem 2.9.

Consider the field $\mathcal{B}_0$ of finite disjoint unions of intervals $(a, b]$ where $-\infty \leq a < b \leq \infty$.

For finite $a < b$, define $P((a, b]) = F(b) - F(a)$. Also define $P(-\infty, a]) = F(a)$ and $P((a, \infty)) = 1 - F(a)$.

Extend $P$ by additivity to $\mathcal{B}_0$. As in Week 1, Theorem 1.2, one needs to show that this definition is consistent, that $P$ is finitely-additive, and that $P$ is countably-additive on $\mathcal{B}_0$. Once we prove this, we invoke Theorem 2.1.

(Omitted in 2018)

Right-continuity of $F$ is used as follows: for $a < b$ are finite, given $0 < \varepsilon < P((a, b])$ there exists $0 < \delta < b - a$ such that $P((a + \delta, b]) < \varepsilon$. Therefore for every $A \in \mathcal{B}$ there exist a compact $K$ and $B \in \mathcal{B}_0$ such that $B \subset K \subset A$ and $P(B \setminus A) < \varepsilon$. (For $a = -\infty$ or $b = \infty$ the above argument needs modification, but one can still find $B \in \mathcal{B}_0$ and compact $K$ as claimed.)

This is “tightness”, so the proof is then concluded by Exercise 1.13.

(Omitted in 2018)

**Solution of Exercise 1.13.** We prove the contrapositive to the implication in Remark 1.2(3).

Suppose $A_1 \supset A_2 \supset \ldots$ are sets in $\mathcal{F}$ such that there exists $\delta > 0$ with $P(A_n) > \delta$ for all $n$. We want to show that $\bigcap_n A_n = \emptyset$ is not possible.

Using tightness, we can find compact sets $K_1, K_2, \ldots$ and sets $B_j \in \mathcal{F}$ such that $B_j \subset K_j$ and $P(B_j) > P(A_j) - \delta/2^j$. Then $P(A_n) - P(B_1 \cap B_2 \cap \cdots \cap B_n) = P(A_n \setminus B_1 \cap B_2 \cap \cdots \cap B_n) = P(\bigcup_{j=1}^n (A_j \setminus B_j)) \leq \sum_{j=1}^n P(A_j \setminus B_j) = \sum_{j=1}^n (P(A_j) - P(B_j)) < \delta/2$ Since $P(A_n) > \delta$ this shows that $P(B_1 \cap B_2 \cap \cdots \cap B_n) > \delta/2 > 0$. In particular, $K_1 \cap \cdots \cap K_n \subset B_1 \cap B_2 \cap \cdots \cap B_n \neq \emptyset$.

We now use the property of compact sets: $K_1 \cap \cdots \cap K_n \neq \emptyset$ implies that $\bigcap_{n=1}^\infty K_n \neq \emptyset$. Therefore $\bigcap_{n=1}^\infty A_n \supset \bigcap_{n=1}^\infty K_n \neq \emptyset$. 

□
3.1. Examples.

3.1.1. Uniform distributions.

Example 2.1 (Uniform I). Uniform distribution on the set of real numbers \( \{ x_1 < x_2 < \cdots < x_n \} \) is (see Examples 1.4 and 1.5) \( P = \frac{1}{n} \sum_{j=1}^{n} \delta_{x_j} \) and corresponds to \( F(x) = \# \{ j : x_j \leq x \} / n \).

Example 2.2 (Uniform II). Uniform distribution on the interval \((0, 1)\) is the probability measure \( P \) which corresponds to

\[
F(x) = \begin{cases} 
0 & x < 0 \\
\frac{x}{b-a} & \text{for } 0 \leq x \leq 1 \\
1 & x > 1 
\end{cases}
\]

Notation: \( U(0, 1) \). More generally, \( U(a, b) \) corresponds to \( F(x) = (x - a)/(b - a)1_{(a,b)} + 1_{[b,\infty)} \).

Recall the construction of the Cantor set: split \([0, 1]\) into \([0, 1/3] \cup (1/3, 2/3) \cup [2/3, 1]\) and remove the middle part. Continue recursively the same procedure with each of the closed intervals retained.

Example 2.3 (Uniform III). Uniform distribution on the Cantor set corresponds to \( F \) that is constant on all deleted intervals,

\[
F(x) = \begin{cases} 
0 & x < 0 \\
\frac{x}{3^d} & \text{for } 1/3^d \leq x < 1/3^{d-1} \\
1/2 & 1/3 \leq x < 2/3 \\
\vdots \\
7/9 & 7/9 \leq x < 8/9 \\
1 & x \leq 1 
\end{cases}
\]

The interval removed in \( d \)-th step is \( (\sum_{k=1}^{d-1} x_k/3^k, \sum_{k=1}^{d} x_k/3^k + 1/3^d) \) with \( x_d = 1 \) and \( x_1, \ldots, x_{k-1} \in \{0, 2\} \). For example, for \( d = 1 \) it is \((1/3, 1/3 + 1/3)\). For \( d = 2 \) the intervals are \((1/3^2, 1/3^2 + 1/3^2)\) and \((2/3 + 1/3^2, 2/3 + 1/3^2 + 1/3^2)\). On each removed interval, \( F(x) = \sum_{k=1}^{d-1} x_k/2^{k+1} + 1/2^d \) is constant.
3. Probability measures on $\mathbb{R}$

3.1.2. Important (absolutely) continuous distributions. Continuous distributions arise from $F(x) = \int_{-\infty}^{x} f(y) dy$, where the so called density function $f \geq 0$ and $\int_{-\infty}^{\infty} f(y) dy = 1$. Example 2.2 is absolutely continuous with $f(y) = 1_{[a,b]}$.

Example 2.4 (Exponential distribution). Take $f(x) = \lambda e^{-\lambda x} I_{(0,\infty)}(x)$, where $\lambda > 0$. This gives

$$F(x) = \begin{cases} 
0 & x < 0 \\
1 - e^{-\lambda x} & x \geq 0.
\end{cases}$$

Example 2.5 (Standard normal distribution). Take $f(x) = \exp(-x^2/2)/\sqrt{2\pi}$. Notation: $N(0,1)$.

3.1.3. Other examples.

Example 2.6 (mixed type). It is clear that

$$F(x) = \begin{cases} 
0 & x < 0 \\
x/9 & 0 \leq x < 1 \\
x/3 & 1 \leq x < 2 \\
1 & x \geq 2
\end{cases}$$

is a cumulative distribution function which cannot be written as an integral of a density.

---

Probability measures of mixed type arise in actuarial models, where the loss of an insured person might have a density but the insurance payoff may be capped, or be a fraction of the loss that changes when the loss exceeds some predefined thresholds.
4. Probability measures on $\mathbb{R}^k$

For simplicity consider only $k = 2, 3$.

4.1. Probability measures on $\mathbb{R}^2$. The $\pi$ system that generates Borel sets of $\mathbb{R}^2$ consists of sets $(-\infty, x] \times (-\infty, y]$. Thus every probability measure $P$ on Borel sets of $\mathbb{R}^2$ is determined uniquely by its values on such sets, $F(x,y) = P((-\infty, x] \times (-\infty, y])$. Function $F(x,y)$ is called a joint cumulative distribution function.

The probability measure must assign nonnegative numbers to all rectangles $A = (a_1, b_1] \times (a_2, b_2]$. It is clear (draw a picture) that

\[-\infty, b_1] \times (-\infty, b_2] = (-\infty, a_1] \times (-\infty, b_2] \cup (-\infty, b_1] \times (-\infty, a_2] \cup A\]

Thus

\[(2.5) \quad F(b_1, b_2) = P(A) + P((-\infty, a_1] \times (-\infty, b_2] \cup (-\infty, b_1])
= P(A) + F(a_1, b_2) + F(a_2, b_1) - P((-\infty, a_1] \times (-\infty, b_2) \cap (-\infty, b_1])
= P(A) + F(a_1, b_2) + F(a_2, b_1) - F(a_1, a_2)\]

Thus

\[(2.6) \quad P(A) = \Delta_A F := F(b_1, b_2) + F(a_1, a_2) - F(a_1, b_2) - F(a_2, b_1)\]

This shows that we must have $\Delta_A F \geq 0$.

It is also clear that we have the following properties:

- $F$ is "right-continuous": if $a_n, b_n > 0$ converge to 0 then $F(x + a_n, y + b_n) \to F(x, y)$.
- $\lim_{x,y \to \infty} F(x,y) = 1$
- $\lim_{y \to -\infty} F(x,y) = \lim_{x \to -\infty} F(x,y) = 0$
- $G(x) = \lim_{y \to \infty} F(x,y)$ and $H(y) = \lim_{x \to \infty} F(x,y)$ exist and define non-decreasing functions, called the marginal cumulative distribution functions

This motivates the following definition:

**Definition 2.4.** $F(x,y)$ is a bivariate cumulative distribution function, if the following conditions hold:
(i) \( \Delta_A F \geq 0 \) for all \( A = (a_1, a_2] \times (b_1, b_2] \)
(ii) \( \lim_{x,y \to \infty} F(x, y) = 1 \)
(iii) \( \lim_{y \to -\infty} F(x, y) = \lim_{x \to -\infty} F(x, y) = 0 \)
(iv) \( F \) is right-continuous.

The following is an analog of Proposition 2.14.

**Proposition 2.15.** Every cumulative distribution function \( F(x, y) \) corresponds to a unique probability measure.

**Sketch of proof.** The field \( \mathcal{B}_0 \) generated by the sets \( (-\infty, b_1] \times (-\infty, b_2] \) consists of finite unions of disjoint sets that arise as intersections of such sets or their complements, see Exercise 1.18.

This gives sets \( (-\infty, b_1] \times (-\infty, b_2] \), their complements, finite rectangles \( A \), sets of the form \( (-\infty, b_1] \times (a_2, b_2] \) and \( (a_1, b_1] \times (-\infty, b_2] \).

We define \( P((a_1, \infty) \times (a_2, \infty)) = 1 - F(a_1, a_2), P(A) = \Delta_A(F), P((-\infty, b_1] \times (-\infty, b_2]) = F(b_1, b_2) \) and \( P((-\infty, b_1] \times (a_2, b_2]) = \lim_{a_1 \to -\infty} \Delta_A F \). We extend the definition by additivity to \( \mathcal{B}_0 \).

Next we check that the assumptions of Exercise 1.13 are again satisfied, so we can conclude that \( P \) has a unique countably additive extension to the Borel \( \sigma \)-field.

It suffices to find a suitable compact set for each of the four types of the "generalized" rectangles. If \( A = (a_1, \infty) \times (a_2, \infty) \) we take \( K = [a_1 + \delta, B_1] \times [a_2 + \delta, B_2] \) and \( B = (a_1 + \delta, B_1] \times (a_2 + \delta, B_2] \).

Given \( \varepsilon > 0 \) choose \( \delta \) such that \( F(a_1 + \delta, a_2 + \delta) < F(a_1, a_2) + \varepsilon \). B_1, B_2 such that \( F(B_1, B_2) > 1 - \varepsilon \).

\[
P(B) = F(B_1, B_2) + F(a_1 + \delta, a_2 + \delta) - F(a_1 + \delta, B_2) - F(a_2 + \delta, B_1)
\]

\( \square \)

**Example 2.7.** Uniform distribution on the unit square is defined by

\[
F(x, y) = \begin{cases} 
xy & 0 \leq x \leq 1, 0 \leq y \leq 1 \\
x & 0 \leq x \leq y, y > 1 \\
y & x > 1, 0 \leq y \leq 1 \\
1 & x > 1, y > 1 \\
0 & \text{otherwise}
\end{cases}
\]

4.2. **Probability measures on \( \mathbb{R}^3 \).** The \( \pi \) system that generates Borel sets of \( \mathbb{R}^3 \) consists of sets \( (-\infty, x] \times (-\infty, y] \times (-\infty, z] \). Thus every probability measure is determined uniquely by its values on such sets, \( F(x, y, z) \).

We need to assign values of the measure to all rectangles \( A = (a_1, b_1] \times (a_2, b_2] \times (a_3, b_3] \).

It is clear that

\[
(2.7) \quad (-\infty, b_1] \times (-\infty, b_2] \times (-\infty, b_3] = A \cup (-\infty, a_1] \times (-\infty, b_2] \times (-\infty, b_3] \cup (-\infty, b_1] \times (-\infty, a_2] \times (-\infty, b_3] \cup (-\infty, b_1] \times (-\infty, b_2] \times (-\infty, a_3]
\]

Noting that \( A \) is disjoint with the remaining set, by the inclusion-exclusion formula (1.1), we get

\[
(2.8) \quad F(b_1, b_2, b_3) = P(A) + F(a_1, b_2, b_3) + F(b_1, a_2, b_3) + F(b_1, b_2, a_3)
\]

\[
- F(a_1, a_2, b_3) - F(a_1, b_2, a_3) - F(b_1, a_2, a_3) + F(a_1, a_2, a_3)
\]
2. Probability measures

So

\[ P(A) = \Delta_A(F) := F(b_1, b_2, b_3) + F(a_1, a_2, b_3) + F(a_1, b_2, a_3) - F(a_1, b_2, b_3) - F(b_1, a_2, b_3) - F(b_1, b_2, a_3) - F(a_1, a_2, a_3) \]

An analog of Definition 2.4 uses \( \Delta_A(F) \) as defined in (2.9). Proposition 2.15 has an \( R^3 \) version. Similar approach works in \( k \) dimensions, compare [Durrett, Theorem 1.1.6] or [Billingsley, Theorem 12.5], who consider general measures. (In general, \( \Delta_A(F) \) is defined using the inclusion-exclusion principle (1.1). Note that for unbounded measures \( F \) can take negative values!)

4.3. Probability measures on \( \mathbb{R}^\infty \). Recall that \( \mathbb{R}^\infty \) is the set of all infinite real sequences, with metric (0.4). Probability measures on \( \mathbb{R}^\infty \) are determined uniquely by the families of joint finite-dimensional distributions that arise from a special \( \pi \)-system of cylindrical sets, i.e. sets of the form

\[ (-\infty, a_1] \times (-\infty, a_2] \times \cdots \times (-\infty, a_n] \times \mathbb{R} \times \mathbb{R} \times \ldots \]

A special case of such a measure is constructed in Theorem 4.6. This is one place where probability theory “outperforms” the general measure theory - while there is a Lebesgue measure on \( \mathbb{R}^d \), there is no Lebesgue measure on \( \mathbb{R}^\infty \).

(Omitted in 2018)

4.4. Probability measures on \( \Omega = C[0,1] \). Constructions of probability measures on function spaces such as \( C[0,1] \) usually rely on the \( \pi \) system of sets of the form \( \{ f : f(t_1) \leq x_2, \ldots, f(t_n) \leq x_n \} \) which are indexed by \( t_1, \ldots, t_n \in [0,1] \) and \( x_1, \ldots, x_n \in \mathbb{R} \). These are sometimes referred to as cylindrical sets.

The functions

\[ F_{t_1, \ldots, t_n}(x_1, \ldots, x_n) = \Pr(f : f(t_1) \leq x_2, \ldots, f(t_n) \leq x_n) \]

are called the finite dimensional distributions. For fixed \( t_1, \ldots, t_n \), \( F_{t_1, \ldots, t_n}(x_1, \ldots, x_n) \) is a cumulative distribution function which determines a family of probability measures \( P_{t_1, t_2, \ldots, t_n} \) on Borel subsets of \( \mathbb{R}^n \). These measures determine a probability measure \( \Pr \) on \( C[0,1] \) uniquely, but it is easy to see that to do so they must be “consistent”. An example of a consistency condition is \( P_{t_1}(A) = P_{t_1, t_2}(A \times \mathbb{R}) \).

Constructions of such measures requires good understanding of compact subsets of \( C[0,1] \).
4.5. **Probability measures on** $\Omega = \mathbb{R}^{[0,1]}$. Since compact sets are easy to find in product spaces, the simplest example of a probability measure on an infinite dimensional space is the case of $\Omega = \mathbb{R}^{[0,1]}$.

**Theorem 2.16** (Kolmogorov). Suppose probability measures $P_{t_1,\ldots,t_n}$ are consistent. Then there exists a unique probability measure $\Pr$ on $\mathbb{R}^{[0,1]}$ with Borel $\sigma$-field that generates $P_{t_1,\ldots,t_n}$ as finite dimensional distributions.

**Remark 2.4.** A good description of Borel $\sigma$-field in $\mathbb{R}^{[0,1]}$ appears in [Billingsley, Section 36]. In particular, the subset $C[0,1] \subset \mathbb{R}^{[0,1]}$ is not a Borel set! However, for a given $\Pr$ one can ask what is $\Pr^*$ and $\Pr_*$ of $C[0,1]$.

Good probability measures are those for which $\Pr^*(C[0,1]) = 1$ and $\Pr^*((C[0,1])^c) = 0$.

**Proof.** The steps in the proof are:

- Introduce the field $\mathcal{F}_0$ of cylindrical sets, indexed by $t_1, \ldots, t_n$ and Borel subsets of $\mathbb{R}^n$.
- Define a probability measure $\Pr$ on $\mathcal{F}_0$ by using the finite-dimensional distributions $P_{t_1,\ldots,t_n}$.
- One then uses a variant of the compactness argument similar to Exercise 1.13 to verify that if $A_n$ is a decreasing family of sets in $\mathcal{F}_0$ with $\bigcap_n A_n = \emptyset$ then $\Pr(A_n) \to 0$. □

---

**Required Exercises**

**Exercise 2.1** (Different representations of the same measure on $\mathcal{F}$). Let $\lambda$ be the Lebesgue measure on the Borel $\sigma$-field of subsets of $\Omega = [0,1]$. Consider $\pi$-system $\mathcal{P} = \{[0,1/n) : n \in \mathbb{N}\}$ and let $\mathcal{F} = \sigma(\mathcal{P})$. Show that there exists a discrete probability measure $P = \sum_{n=1}^{\infty} p_n \delta_{\omega_n}$ on $2^\Omega$ (see Example 1.5) such that $\lambda$ restricted to $\mathcal{F}$ coincides with $P$ restricted to $\mathcal{F}$. (In formal notation, $\lambda|_{\mathcal{F}} = P|_{\mathcal{F}}$.) Is $P$ unique?

**Exercise 2.2** (Statistics). It is illustrative to produce empirical histograms at various sample sizes for the uniform distribution on the Cantor set from Example 2.3. Somewhat surprisingly, this is easy to simulate: take $2 \sum_{k=1}^{\infty} \varepsilon_k / 3^k$ where $\varepsilon_k$ represents a “toss of a fair coin” with values 0 or 1. This exercise asks you to reproduce histograms from [Proschan-Shaw].

**Exercise 2.3** (measure-preserving maps). Let $f : [0,1] \to [0,1]$ be the fractional part of $2x$. That is,

$$f(x) = \begin{cases} 2x & \text{if } x \leq 1/2 \\ 2x - 1 & \text{if } x > 1/2 \end{cases}$$
Show that for every Borel subset $A$ of $[0, 1]$ the Lebesgue measure of $f^{-1}(A)$ equals to the Lebesgue measure of $A$. (Compare Exercise 2.11.)

**Exercise 2.4.** What should be the CDF $F(x, y)$ for the distribution “uniform on the triangle” $x \geq 0, y \geq 0, x + y \leq 1$?

### Additional Exercises

**Exercise 2.5.** Let $\Omega = (0, 1] \times (0, 1]$ and let $\mathcal{F}$ be the class of sets of the form $A_1 \times (0, 1]$ with $A_1 \in \mathcal{B}$ the Borel $\sigma$-field in $(0, 1]$ and $P(A_1 \times (0, 1]) = \lambda(A_1)$ (the Lebesgue measure). Then $(\Omega, \mathcal{F}, P)$ is a probability space. For the diagonal $D = \{(x, x) : 0 < x \leq 1\}$, find $P^*(D)$ and $P^*(D^c)$.

**Exercise 2.6.** Inspect the proofs of Theorems 2.1 and 2.9. Find all places where additivity or countable additivity is used.

**Exercise 2.7** (Compare Exercise 1.19). For $\Omega = (0, 1]$ with the field $\mathcal{B}_0$ generated by intervals $I = (a, b]$, consider $\lambda_0(I) = |I|$, extended by additivity to $\mathcal{B}_0$. Let $Q$ be the set of all rational numbers in $(0, 1]$ Use the definition of $\lambda^*$ (not subadditivity) to show that $\lambda^*(Q) = 0$.

**Exercise 2.8.** The family $\mathcal{P}$ of open intervals $(-1/n, 1/n)$ with $n \in \mathbb{N}$ is a $\pi$-system in $\Omega = (-1, 1)$. Describe what sets are in the $\sigma$-field $\sigma(\mathcal{P})$. In particular, is set $\{0\}$ in $\sigma(\mathcal{P})$?

**Exercise 2.9.** Let $\mathcal{A}$ be the smallest field generated by a $\pi$-system $\mathcal{P}$ (see Exercise 1.18). Use the inclusion-exclusion formula from Exercise 1.1 to show that finitely additive probability measures that agree on $\mathcal{P}$ must also agree on $\mathcal{A}$.

**Exercise 2.10.** Suppose $\mathcal{L}$ is a $\lambda$-system. Show that $A, B \in \mathcal{L}$ and $A \subset B$ implies that $B \setminus A \in \mathcal{L}$. 

**Hint:** Show that $(B \setminus A)^c \in \mathcal{L}$.

**Exercise 2.11.** Consider $\Omega = (0, 1)$ with Lebesgue measure. Use the Dynkin’s $\pi$-$\lambda$ Theorem to prove that for all Borel sub-sets $B$ of $(0, 1/2)$ and all $x \in (0, 1/2)$, the Lebesgue measure of $B + x$ is the same as the Lebesgue measure of $B$. 

Chapter 3

Independence


1. Independent events and sigma-fields

This section follows [Billingsley, Section 4].

Definition 3.1. Events $A, B$ are independent if $P(A \cap B) = P(A)P(B)$.

Events $A_1, \ldots, A_n$ are independent, if for every $r \leq n$ and every choice of distinct $k_1, \ldots, k_r$

\begin{equation}
P(A_{k_1} \cap A_{k_2} \cap \cdots \cap A_{k_r}) = P(A_{k_1})P(A_{k_2}) \cdots P(A_{k_r})
\end{equation}

(3.1)

An infinite sequence of events $A_1, A_2, \ldots$ is independent if the events $A_1, \ldots, A_n$ are independent for every $n$.

Example 3.1. Consider $\Omega = [0, 1]^3$ with Lebesgue measure. Then events $A = \{(x, y, z) \in \Omega : x < 1/2\}$, $B = \{(x, y, z) \in \Omega : y < 1/2\}$, $C = \{(x, y, z) \in \Omega : z < 1/2\}$ are independent.

Remark 3.1. Events $A_1, \ldots, A_n$ are independent iff

\begin{equation}
P(B_1 \cap B_2 \cap \cdots \cap B_n) = P(B_1)P(B_2) \cdots P(B_n)
\end{equation}

(3.2)

for all choices of $B_j = A_j$ or $B_j = \Omega$.

Definition 3.2. Classes of sets $A_1, A_2, \ldots, A_n$ are independent, if for each choice of $A_j$ from $A_j$ the events $A_1, \ldots, A_n$ are independent.

In particular, $\sigma$-fields $A_1, A_2, \ldots, A_n$ are independent, if for every choice of $B_j$ from $A_j$, equation (3.2) holds.

Theorem 3.1. If $A_1, \ldots, A_n$ are independent $\pi$-systems then $\sigma(A_1), \ldots, \sigma(A_n)$ are independent.

Proof. Without loss of generality we may assume $\Omega \in A_j$ so the definition to use is (3.2).

Fix $B_2, \ldots, B_n$ and consider

\[ L = \{ B_1 \in \sigma(A_1) \text{ such that (3.2) holds} \} \]

It is easy to see that $L$ is a $\lambda$-system and $A_1$ is a $\pi$-system contained in $L$. So by Theorem 2.12 we see that $\sigma(A_1) \subseteq L$. This proves that

$\sigma(A_1), A_2, \ldots, A_n$

are independent $\pi$-systems.

We now repeat the same argument for $A_2$, then $A_3$, etc. \qed
Corollary 3.2. If \( A_1, \ldots, A_n, \ldots \) is an infinite set of independent \( \pi \)-systems, then \( \sigma(A_1), \ldots, \sigma(A_n), \ldots \) are independent.

Corollary 3.3. If \( A_{i,j} \) is an (possibly infinite) array of independent events then the \( \sigma \)-fields generated by each row are independent.

**Proof.** We introduce \( \pi \)-systems \( A_i \) as the collection of all finite intersections of the events in the \( i \)-th row, including \( \Omega \). If \( C_i \in A_i \) then \( C_j = \bigcap_{k=1}^{m_i} A_{i,k} \) where \( A_{i,k} \in A_i \)

\[
P(\bigcap_{i=1}^{n} C_i) = P(\bigcap_{j=1}^{n} \bigcap_{k}^{m} A_{i,j}) = \prod_{i=1}^{n} P(A_{i,j}) = \prod_{i=1}^{n} P(C_i)
\]

\[\square\]

**Example 3.2.** Suppose events \( A_1, A_2, A_3, A_4 \) are independent. Then the events \( A = A_1 \cup A_2 \) and \( B = A_3 \cup A_4 \) are also independent. This can be verified by elementary calculation, but we deduce this from Corollary 3.3:

Consider \( A = \sigma(A_1, A_2) \) and \( B = \sigma(A_3, A_4) \). By Corollary 3.3 these \( \sigma \)-fields are independent. Clearly, \( A \in \sigma(A) \) and \( B \in \sigma(B) \).

2. Zero-one law

**Definition 3.3.** The tail \( \sigma \)-field for a sequence of events \( A_1, A_2, \ldots \) is

\[
\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma(A_n, A_{n+1}, \ldots)
\]

**Example 3.3.** \( \{A_n \text{ i.o}\} = \bigcap_{n} \bigcup_{k \geq n} A_k \) and \( \{\text{all but a finite number of events } A_n \text{ holds}\} = \bigcup_{n} \bigcap_{k \geq n} A_k \) are tail events.

**Theorem 3.4** (Kolmogorov’s zero-one law). If \( A_1, A_2, \ldots \) is an independent sequence of events, then the tail \( \sigma \)-field is trivial: if \( A \in \mathcal{T} \) then \( P(A) \) is either 0 or 1.

**Proof.** By Corollary 3.3, applied to the array of independent events

\[
A_1 \\
A_2 \\
\vdots \\
A_{n-1} \\
A_n \\
A_{n+1} \\
\ldots
\]

the \( \sigma \)-fields \( \sigma(A_1), \sigma(A_2), \ldots, \sigma(A_{n-1}), \sigma(A_n, A_{n+1}, \ldots) \) are independent.

Since \( A \in \mathcal{T} \subset \sigma(A_n, A_{n+1}, \ldots) \), we see that \( A \) is independent of \( A_1, A_2, \ldots, A_{n-1} \) for every \( n \).

Corollary 3.3, applied to the array of independent events

\[
A \\
A_1 \\
A_2 \\
\ldots
\]

shows that \( \sigma(A) \) and \( \sigma(A_1, A_2, \ldots) \) are independent. But \( A \in \mathcal{T} \subset \sigma(A_1, A_2, \ldots) \), so \( P(A) = P(A \cap A) = P(A)P(A) \).

\[\square\]
Example 3.4. If $A_n$ are independent events then $P(\bigcap_{n} \bigcup_{k \geq n} A_k)$ is either 0 or 1. It is of interest to determine when each of the cases occurs.

**Proof.** $\{\bigcap_{n} \bigcup_{k \geq n} A_k = \{A_n \ i.o.\}$ is a tail event: to determine if the infinite number of events occurred, we do not need to know anything about $A_1$, $A_2$, etc. $\square$

**Corollary 3.5.** If $A_n$ are independent events and $A = \{\omega : \frac{1}{n} \sum_{n} I_{A_n}(\omega) \text{ converges} \}$ then $P(A)$ is either 0 or 1. It is of interest to determine when each of the cases occurs.

3. Borel-Cantelli Lemmas

**Theorem 3.6** (First Borel-Cantelli Lemma). If $\sum_{n=1}^{\infty} P(A_n) < \infty$ then $P(A_n \ i.o.) = 0$.

**Example 3.5.** Suppose intervals $A_n \subset (0, 1]$ have lengths $\lambda(A_n) = 1/n^2$. Then the set of $\omega \in \Omega$ for which $\sum_{n} I_{A_n}(\omega) < \infty$ has Lebesgue measure 1, because with probability one only the finite number of $I_{A_n}(\omega)$ is one, i.e. for almost every $\omega$, this is a finite sum.

**Proof.** $0 \leq P(A_n \ i.o.) = \lim_{n \to \infty} P(\bigcup_{k=n}^{\infty} A_k) \leq \lim_{n \to \infty} \sum_{k=n}^{\infty} P(A_k) = 0$ $\square$

**Theorem 3.7** (Second Borel-Cantelli Lemma). If $\sum_{n=1}^{\infty} P(A_n) = \infty$ and $A_n$ are independent events then $P(A_n \ i.o.) = 1$.

**Proof.** $P(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k) = 1 - P(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c)$ and $P(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c) = \lim_{n \to \infty} P(\bigcap_{k=n}^{\infty} A_k^c)$

Now it turns out that $P(\bigcap_{k=n}^{\infty} A_k^c) = 0$, since it is given as

$$\lim_{m \to \infty} P(\bigcap_{k=n}^{m} A_k^c) = \lim_{m \to \infty} \prod_{k=1}^{m} (1 - P(A_k)) \leq \lim_{m \to \infty} \exp \left( - \sum_{k=n}^{m} P(A_k) \right) = 0.$$
Required Exercises

Exercise 3.1. Suppose $A, B, C$ are independent events. Prove directly from the definition that $(A \cup B), C$ is a pair of independent events. Similarly, show that events $(A \setminus B), C$ are independent.

Here is a longer version of Exercise 3.1.

Exercise 3.2. Suppose $A, B, C$ are independent events. Prove directly from the definition that their complements $A^c, B^c, C^c$ are also independent events.

Exercise 3.3 (Statistics\footnote{Statistics\footnote{Statistics}}). Consider $\Omega = [0, 1]$ with Lebesgue measure. Exhibit explicitly three independent events $A, B, C \subset [0, 1]$ with $\lambda(A) = \lambda(B) = \lambda(C) = 1/2$.

Exercise 3.4. If $P(A_n) \geq \varepsilon > 0$ then $P(A_n \ i.o. > 0$

Exercise 3.5. Suppose $A_k$ are independent events such that $P(A_k) = 1/2$. Show that

$$\Pr \left( \bigcup_{n=1}^{\infty} A_n \right) = 1.$$

Exercise 3.6. Suppose $A_1, A_2, \ldots, A_n, \ldots$ are independent events with probability $P(A_n) = n^{-\theta}$. Determine all $\theta \in \mathbb{R}$ such that $P(A_n \ i.o.) = 1$.

Additional Exercises

Exercise 3.7. Suppose $P$ is a probability measure on $\mathbb{R}^2$ with the cumulative distribution function $F(x, y)$ that factors into a product of $A(x)B(y)$. Prove that the $\sigma$-fields $\mathcal{F} = \{ U \times \mathbb{R} : U \in \mathcal{B}(\mathbb{R}) \}$ and $\mathcal{G} = \{ \mathbb{R} \times U : U \in \mathcal{B}(\mathbb{R}) \}$ are independent.

Exercise 3.8. Suppose $\{A_n\}$ are independent events satisfying $P(A_n) < 1$. Show that

$$P(\bigcup_{n=1}^{\infty} A_n) = 1 \text{ iff } P(A_n \ i.o.) = 1$$

Exercise 3.9. Use the definitions to prove the claims from Example 3.3.

Exercise 3.10. Consider probability space $(\Omega, \mathcal{F}, P) = ((0, 1), \mathcal{B}, \lambda)$ (the Lebesgue measure). Let $A_n$ be "consecutive intervals" of length $p_n$, "wrapped around" if needed. Thus $A_1 = (0, p_1)$, $A_2 = (p_1, p_1 + p_2)$, and so on. (The best way to imagine this is to think of $\Omega$ as a circle, with one point removed, and $A_n$ being "rotated" into the adjacent position next to $A_{n-1}$.)

Borel-Cantelli Lemma gives a sufficient condition for $P(A_n \ i.o.) = 0$. Prove that

$$P(A_n \ i.o.) = 1 \text{ iff } \sum_{n=1}^{\infty} p_n = \infty.$$
So $P(A_n$ i.o.) is either 0 or 1 for "consecutively placed intervals", just like for independent events.

**Exercise 3.11.** Let $A_k \in \mathcal{F}$. Show that if $P(\bigcap_{n=1}^{\infty} A_k) = 1$, then for every $A \in \mathcal{F}$ of positive probability the series $\sum_n P(A \cap A_n)$ diverges.

**Exercise 3.12.** Let $A_k \in \mathcal{F}$. Show that if for every $A \in \mathcal{F}$ of positive probability the series $\sum_n P(A \cap A_n)$ diverges, then $P(\bigcap_{n=1}^{\infty} A_k) = 1$.

**Exercise 3.13.** Suppose $A_1, A_2, \ldots$ are independent events. Consider

$$C = \left\{ \omega : \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} I_{A_k}(\omega) \text{ exists} \right\}.$$

Show that either $P(C) = 0$ or $P(C) = 1$.

**Exercise 3.14.** Suppose $A_n$ is a sequence of events such that $\lim_{n \to \infty} P(A_n) = 0$ and $\sum_{n=1}^{\infty} P(A_n \cap A_{n+1}^c) < \infty$. Prove that $P(A_n$ i.o.) = 0.

**Hints:** [Resnik] gives the following hint: decompose $\bigcup_{j=n}^{m} A_j$. See also [Gut, Theorem 18.7].

**Exercise 3.15.** Suppose $\sum_{n=1}^{\infty} P(A_n) < \infty$. Let $B = \{\omega : \sum_{n=1}^{\infty} I_{A_n}(\omega) < \infty\}$. Show that $P(B) = 1$. 


Random variables


1. Measurable mappings

Suppose \( \Omega \) and \( E \) are two sets. Often \( E = \mathbb{R} \) or \( E = \mathbb{R}^d \).

Suppose \( X : \Omega \to E \) i.e. \( X \) is a function with domain \( \Omega \) and target set \( E \). Then \( X \) induces a mapping

\[
X^{-1} : 2^E \to 2^\Omega
\]
defined by \( X^{-1}(U) = \{ \omega \in \Omega : X(\omega) \in U \} \), where \( U \subseteq E \).

**Proposition 4.1.** Properties of induced mapping:

(i) \( X^{-1}(\emptyset) = \emptyset, X^{-1}(E) = \Omega \)

(ii) \( X^{-1}(U^c) = (X^{-1}(U))^c \)

(iii) \( X^{-1}\left(\bigcup_{t \in T} U_t\right) = \bigcup_{t \in T} X^{-1}(U_t) \)

**Proof.** For (iii), \( \omega \in X^{-1}\left(\bigcup_{t \in T} U_t\right) \) iff \( \exists_{t \in T} X(\omega) \in U_t \). \( \square \)

**Corollary 4.2.** If \( B \) is a \( \sigma \)-field of subsets of \( E \) then \( X^{-1}(B) \) is a \( \sigma \)-field of subsets of \( \Omega \).

**Proof.** This is based on the identities for inverse images under functions, see Proposition 4.1. \( \square \)

**Definition 4.1.** A \( \sigma \)-field generated by \( X \) is \( \sigma(X) = X^{-1}(B) \).

Exercise 4.16 says that this is the smallest \( \sigma \)-field of subsets of \( \Omega \) which makes \( X \) measurable.

1.1. Random elements and random variables. Suppose \( (\Omega, F, P) \) is a probability space and \( E \) is a set with distinguished \( \sigma \)-field \( B \). In most applications, \( E \) is a separable complete metric space and \( B \) is the *Borel* \( \sigma \)-field which is generated by the countable collection of open balls.

**Definition 4.2.** In analysis, \( X \) is called a measurable function if \( X^{-1}(B) \subseteq F \). In probability, \( X \) is then called a random element of \( E \).

If we want to indicate the \( \sigma \)-fields, we will write \( X : (\Omega, F) \to (E, B) \).

The most important special cases are \( E = \mathbb{R} \) and \( E = \mathbb{R}^d \). When \( E = \mathbb{R} \), we say that \( X \) is a random variable. When \( E = \mathbb{R}^d \), we say that \( X \) is a random vector or that \( (X_1, \ldots, X_d) \) is a multivariate random variable. In such cases, measurability can be verified somewhat easier.
Proposition 4.3. To verify whether \( X : \Omega \to \mathbb{R} \) is a random variable we only need to verify that the sets \( A_x = \{ \omega : X(\omega) \leq x \} \) are in \( \mathcal{F} \) for every real \( x \).

Similarly, to verify whether \( (X,Y) : \Omega \to \mathbb{R}^2 \) is measurable, we only need to verify whether for all \( x, y \in \mathbb{R} \) we have \( \{ \omega : X(\omega) \leq x, Y(\omega) \leq y \} \) is in \( \mathcal{F} \).

Proof. Consider the set \( \mathcal{U} \) of all sets \( U \subset \mathbb{R} \) such that \( X^{-1}(U) \in \mathcal{F} \). In view of Proposition 4.1, this is a sigma-field.

For \( x \in \mathbb{R} \), the inverse image of the set \( (-\infty, x] \) is in \( \mathcal{F} \), so \( (-\infty, x] \in \mathcal{U} \). Hence the generated sigma field \( \mathcal{B} \subset \mathcal{U} \).

\[ Q(U) = P(X^{-1}(U)) \]

Sometimes \( Q \) is called an induced measure and some authors use notation \( Q = P \circ X^{-1} \).

If \( X \) is a random variable, then its distribution is uniquely determined by the corresponding cumulative distribution function

\[ F(x) = Q((-\infty, x]) = P(\{ \omega : X(\omega) \leq x \}) \]

In probability and statistics the latter is usually abbreviated to \( F(x) = P(X \leq x) \) but this abbreviated notation is just the shorthand for the right hand side of (4.1).

Definition 4.4. We say that random variables \( X,Y \), defined perhaps on different probability spaces, are equal in distribution, if they induce the same probability measure on \( (\mathbb{R}, \mathcal{B}) \).

In view of Proposition 2.14, this is equivalent to \( X,Y \) having the same cumulative distribution function.

If \( X,Y \) are two random variables on the same probability space \( (\Omega, \mathcal{F}, P) \) then the pair \( (X,Y) \) is a measurable mapping \( \Omega \to \mathbb{R}^2 \). The joint distribution of random variables is just the induced measure on \( \mathbb{R}^2 \) and is uniquely determined by the joint cumulative distribution function

\[ F(x,y) = P(X \leq x, Y \leq y) \]

(Note the abbreviated notation for \( P(\{ \omega : X(\omega) \leq x, Y(\omega) \leq y \} \) If \( \mathbb{E} = C[0,1] \) then a measurable mapping \( X : \Omega \to C[0,1] \) is called a stochastic process with continuous trajectories. The standard notation is \( X = (X_t)_{t \in [0,1]} \). The distribution of \( X \) is uniquely determined by the family of finite dimensional distributions

\[ F_{t_1,t_2,\ldots,t_k}(x_1, x_2, \ldots, x_k) = P(X_{t_1} \leq x_1, \ldots, X_{t_k} \leq k) \]
that satisfy natural consistency conditions. The converse is not as simple here: consistent families of finite-dimensional distributions

$$\{F_{t_1,t_2,\ldots,t_k}(x_1,x_2,\ldots,x_k): 0 \leq t_1 \leq t_2 \leq \cdots \leq t_k \leq 1\}$$

define a probability measure on Borel sets of the product space $\mathbb{R}^{[0,1]}$ of all (including nonmeasurable) functions $[0,1] \to \mathbb{R}$ with pointwise convergence, see [Billingsley, Theorem 36.1] but not necessarily on Borel subsets of $C[0,1]$. (In fact, $C[0,1] \subset \mathbb{R}^{[0,1]}$ is not a Borel subset, see the discussion that follows [Billingsley, Theorem 36.3].)

2. Random variables with prescribed distributions

This section is based on [Billingsley, Section 14] or [Durrett, Theorem 1.2.2].

**Theorem 4.4.** If $F$ is a cumulative distribution function\(^1\), then there exists on some probability space \(\Omega, \mathcal{F}, P\) random variable $X$ for which $P(X \leq x) = F(x)$.

**First proof.** Proposition 2.14 gives a probability measure $P$ on $(\mathbb{R}, \mathcal{B})$ such that $F(x) = P((-\infty, x])$. Take $(\mathbb{R}, \mathcal{B}, P)$ for the probability space $(\Omega, \mathcal{F}, P)$. Define $X(\omega) = \omega$ (the identity mapping). Then $X$ has distribution $P$. \(\Box\)

**Second proof.** (This is independent of Proposition 2.14, and in fact can be used to prove it.)

Let $\Omega = (0,1)$ with Lebesgue measure $\lambda$ on Borel sigma-field. Since $F$ is non-decreasing right-continuous with limits 0, 1, for $0 < u < 1$, the set $\{x : u \leq F(x)\}$ is\(^2\) a closed\(^3\) half-line\(^4\) of the form $[\varphi(u), \infty)$ and its complement is $\{x : u > F(x)\} = (-\infty, \varphi(u))$. This shows that for every real $x$, we have $\varphi(u) \leq x$ iff $F(x) \geq u$. This also defines the quantile function

$$\varphi(u) = \inf\{x : u \leq F(x)\} = \sup\{x : F(x) < u\}$$

Define $X(\omega) = \varphi(\omega)$. Then $\lambda(\{\omega : X(\omega) \leq x\}) = \lambda(\{\omega : \varphi(\omega) \leq x\}) = \lambda((0, F(x)]) = F(x)$. \(\Box\)

The second proof of Theorem 4.4 lets us construct an infinite sequence $X_1, X_2, \ldots$ of random variables with prescribed distributions. However, this gives only very special measures on $\mathbb{R}^2$, see Exercise 4.17. In the next section we consider another special construction that gives joint distributions that are more often of interest.

**Corollary 4.5 (Proposition 2.14).** If $F$ is a CDF then there exists a unique probability measure $P$ on the Borel sets of $\mathbb{R}$ such that $P((\infty, a]) = F(a)$.

**Proof.** Existence: Take Lebesgue measure on Borel sigma-field of $(0,1)$, and $X$ as in the second proof above. Then $P$ is the induced probability measure. (Uniqueness follows from Theorem 2.9, see Proof of Proposition 2.14). from $\pi - \lambda$ theorem, see

---

1See Definition 2.3
2Can you see why isn’t it $\mathbb{R}$ or $\emptyset$?
3Why?
4Why?
Example 4.1. Write $X = X_+ - X_-$, i.e.

$$X_+(\omega) = \begin{cases} X(\omega) & \text{if } X(\omega) \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$X_- (\omega) = (-X)_+ = \begin{cases} -X(\omega) & \text{if } X(\omega) \leq 0 \\ 0 & \text{otherwise} \end{cases}$$

If $X$ has CDF $F(x)$, what are the CDFs of $X_+$ and $X_-$?

Solution.

$$P(X_+ \leq x) = \begin{cases} P(X \leq x) & x \geq 0 \\ 0 & x < 0 \end{cases}$$

So $F_+(x) = F(x)I_{[0,\infty]}(x)$.

2.1. Independent random variables. Random elements $X_1, X_2, \ldots$ are independent if $\sigma$-fields $\sigma (X_1), \sigma (X_2), \ldots$ are independent.

Example 4.2. Suppose discrete random variables $X = \sum x_jI_{A_j}, Y = \sum y_kB_k$. Then $X, Y$ are independent if $A = \{A_1, A_2, \ldots\}$ and $B = \{B_1, B_2, \ldots\}$ are independent $\pi$-systems. Thus

$$P(X = x, Y = y) = P(X = x)P(Y = y) \text{ for all } x, y \in \mathbb{R}$$

Similarly, discrete random variables $X, Y, Z$ are independent iff

$$P(X = x, Y = y, Z = 1) = P(X = x)P(Y = y)P(Z = 1) \text{ for all } x, y, z \in \mathbb{R}$$

Example 4.3. Suppose $X_1, X_2, \ldots$ take only values 0, 1 and $p_k = P(X_k = 1), q_k = 1 - p_k$. Then $X_1, X_2, \ldots$ are independent iff

$$P(X_1 = \varepsilon_1, X_2 = \varepsilon_2, \ldots, X_n = \varepsilon_n) = \prod_{k=1}^{n} p_k^{\varepsilon_k} q_k^{1-\varepsilon_k}$$

Independence is often assumed in the theorems. So it is of some interest to make sure that such random variables exist.

Theorem 4.6. If $F_1, F_2, \ldots$ are cumulative distribution functions then there exists a probability space $(\Omega, \mathcal{F}, P)$ and a sequence $X_1, X_2, \ldots$, of independent random variables such that $X_n$ has cumulative distribution function $F_n$.

Sketch of First Proof. In this proof we take $\Omega = \mathbb{R}^\infty$ with (infinite!) product measure $^5 P = P_1 \otimes P_2 \otimes \ldots$ where $P_k$ is the probability measure on $\mathbb{R}$ with cumulative distribution function $F_k$. For $\omega = (\omega_1, \omega_2, \ldots) \in \mathbb{R}^\infty$ we define $X_k(\omega) = \omega_k$.

Sketch of Second Proof. $^6$ We use $\Omega = (0, 1]$ with Lebesgue measure $\lambda$ and with binary digits function $d_n : (0, 1] \to \{0, 1\}$.

We first note that random variables $d_1, d_2, \ldots$ are independent. Indeed, as noted in the proof of Proposition A.1 we have $\lambda(d_1 = \varepsilon_1, \ldots, d_m = \varepsilon_m) = 1/2^m$. By Example 4.3 this proves independence.

$^5$We did not show how to construct infinite product measure!

$^6$This proof is from [Billingsley, Theorem 20.4]. It also answers Exercise 3.3.
Next, we arrange all of these random variables into an infinite array \( d_{i,j} \). Then random variables \( U_i(\omega) = \sum_{j=1}^{\infty} d_{i,j}(\omega)/2^j \) are independent. On the other hand, \( \lambda(\omega : U_i(\omega) \leq x) = x \); this is easiest to see for diadic rational numbers\(^7\) of the form \( x = k/2^n \).

Now take \( X_k = \varphi_k(U_k) \), where \( \varphi_k(u) \) is the quantile transform (4.3) of \( F_k \).

**Definition 4.5.** We say that \( X_1, X_2, \ldots \) are independent identically distributed (i. i. d.) random variables, if they are independent and have the same CDF.

### 2.2. Elementary examples.

**Proposition 4.7.** If \( f : \mathbb{R}^d \to \mathbb{R} \) is measurable (say, continuous) and \( X_1, \ldots, X_d : \Omega \to \mathbb{R} \) are random variables on \( \Omega, \mathcal{F}, P \), then \( Y = f(X_1, \ldots, X_d) \) is a random variable.

**Proof.** If \( B \) is a Borel subset of \( \mathbb{R} \) then \( U = f^{-1}(B) \subset \mathbb{R}^d \) is a Borel subset of \( \mathbb{R}^d \). So \( Y^{-1}(B) = (X_1, \ldots, X_d)^{-1}(U) \in \mathcal{F} \). \( \square \)

Here are some examples of such functions:

**Proposition 4.8 (Sum theorems).** Suppose \( X_1, X_2, \ldots \) are independent and \( S = X_1 + X_2 + \cdots + X_n \).

(i) If \( X_1, \ldots, X_n \) are i. i. d. Bernoulli random variables, i.e., \( P(X_j = 1) = p, P(X_j = 0) = 1 - p \), then \( S \) is Binomial \( \text{Bin}(n, p) \) (see Example 1.6)

(ii) If \( X_1, X_2, \ldots \) are Poisson random variables with parameters \( \lambda_1, \lambda_2, \ldots \) then \( S \) is Poisson with parameter \( \lambda = \lambda_1 + \cdots + \lambda_n \) (see Example 1.7)

(iii) If \( X_1, X_2, \ldots \) are i. i. d. Normal \( N(0,1) \) random variables (see Example 2.5) then \( Y = X_1 + \cdots + X_n \) is normal with mean zero and variance \( n \) (i.e., has same law as \( \sqrt{n}Z \) for some \( N(0,1) \) r.v. \( Z \)).

**Proof.** Omitted\(^8\) \( \square \)

### 3. Convergence of random variables

**Definition 4.6.** A sequence of random variables converges in probability, \( X_n \xrightarrow{P} X \), if\(^9\)

\[
\lim_{n \to \infty} P(\{X_n - X\geq \varepsilon\}) = 0
\]

for every \( \varepsilon > 0 \)

**Example 4.4.** On \( \Omega = [0, 1] \) consider \( X_n = I_{[0,n/(2n+1)]} \). Then \( X_n \xrightarrow{P} X \).

Suppose \( X_n, X \) are random variables on some probability space \( (\Omega, \mathcal{F}, P) \). Then we have

**Proposition 4.9.**

\[
\{\omega \in \Omega : \lim_{n \to \infty} X_n(\omega) = X(\omega)\} \in \mathcal{F}
\]

---

\(^7\)Observe that intervals of the form \([0, k/2^n]\) are a \( \pi \)-system that generates \( \mathcal{B} \)

\(^8\)These are “elementary” facts covered in undergraduate courses.

\(^9\)We already switched to abbreviated notation:

\[
\lim_{n \to \infty} P(\{\omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\}) = 0
\]
Proof. First we note that for a fixed \( \varepsilon > 0 \), the set \( A_n = \{ \omega : |X_n(\omega) - X(\omega)| < \varepsilon \} \in \mathcal{F} \). This is a consequence of Exercise 4.15.

Next, we note that
\[
A_\varepsilon = \{ \omega : \forall_n \exists k_n |X_k(\omega) - X(\omega)| > \varepsilon \}
\]
is in \( \mathcal{F} \). Indeed, \( A_\varepsilon = \bigcap_n \bigcup_{k \geq n} A_k^c \).

Finally, we note that
\[
\bigcap_{\varepsilon > 0} A_\varepsilon = \bigcap_{n \in \mathbb{N}} A_{1/n} \in \mathcal{F}
\]

\[\square\]

Definition 4.7. A sequence of random variables converges with probability 1 if
\[
P\left( \{ \omega \in \Omega : \lim_{n \to \infty} X_n(\omega) = X(\omega) \} \right) = 1
\]

Proposition 4.10. If \( X_n \to X \) with probability 1, then \( X_n \xrightarrow{P} X \)

Proof. The discussion of measurability shows that \( P(\forall \varepsilon > 0 \exists N \forall n > N \{ \omega : |X_n - X| < \varepsilon \}) = 1 \) iff for every rational \( \varepsilon > 0 \)
\[
P(\exists N \forall n > N |X_n - X| < \varepsilon) = P\left( \bigcup_{n > N} \bigcap |X_n - X| < \varepsilon \right) = 1
\]
This is the same as
\[
P(\bigcap_{n > N} |X_n - X| > \varepsilon) = P(|X_n - X| > \varepsilon \ i.o.) = 0
\]
Now \( P(\bigcap_{n > N} |X_n - X| > \varepsilon) = \lim_{N \to \infty} P(\bigcup_{n > N} |X_n - X| > \varepsilon) \). So convergence with probability 1 is equivalent to
\[
(4.4) \quad \forall \varepsilon > 0 \lim_{N \to \infty} P(\sup_{n > N} |X_n - X| > \varepsilon) = 0.
\]

Of course, \( P(|X_n - X| > \varepsilon) \leq P(\sup_{n > N} |X_n - X| > \varepsilon) \) \[\square\]

Proposition 4.11. Suppose \( X_n \xrightarrow{P} X \). Then there exists a subsequence \( n_k \) such that \( X_{n_k} \to X \) with probability 1

Proof. Choose positive \( \varepsilon_k \to 0 \). Given \( k \), choose \( n_k > k \) so that \( P(|X_{n_k} - X| > \varepsilon_k) < 1/2^k \). Since \( \sum_k 1/2^k < \infty \), by the first Borel-Cantelli Lemma,
\[
P(|X_{n_k} - X| > \varepsilon_k \ i.o.) = 0
\]
Therefore, for any \( \varepsilon > 0 \),
\[
P(|X_{n_k} - X| > \varepsilon \ i.o.) \leq P(|X_{n_k} - X| > \varepsilon_k \ i.o.) = 0
\]
Details: Choose \( N_0 \) such that \( \varepsilon_{n_k} < \varepsilon \) for \( k > N_0 \). Then
\[
\bigcap_{N=1}^{\infty} \bigcup_{k > N} \{|X_{n_k} - X| > \varepsilon \} \subset \bigcap_{N > N_0} \bigcup_{k > N} \{|X_{n_k} - X| > \varepsilon \} \subset \bigcap_{N > N_0} \bigcup_{k > N} \{|X_{n_k} - X| > \varepsilon_k \}
\]
\[\square\]
**Remark 4.2.** Convergence in probability is a metric convergence. Convergence with probability 1 is not a \"metric convergence\".

**Remark 4.3.** Suppose $X_n$ are random variables such that $X_n(\omega)$ converges for all $\omega \in \Omega$. Then $X(\omega) := \lim_{n \to \infty} X_n(\omega)$ is a random variable.

**Proof.**

\[
\{ \omega : X(\omega) \leq x \} = \bigcap_{j \in \mathbb{N}} \bigcup_{n \in \mathbb{N}} \bigcap_{k > n} \{ \omega : X_k(\omega) \leq x + 1/j \}
\]

\[\square\]

The third type of convergence, the so called *convergence in distribution*, is somewhat different, as it is really convergence of the induced probability measures, not random variables. This topic will appear later, but we can give a definition here:

**Definition 4.8.** We say that a sequence of $\mathbb{R}$-valued random variables $X_n$ with CDFs $F_n$ converges in distribution to a random variable $Y$ with CDF $F$, if $F_n(x) \to F(x)$ for all continuity points $x$ of $F$.

---

**Required Exercises**

**Measurability.**

**Exercise 4.1.** Suppose that $\varphi : (0, 1) \to \mathbb{R}$ is strictly increasing. Prove that $\varphi$ is measurable with respect to Borel sigma-fields.

**Exercise 4.2.** Suppose that $\varphi : (0, 1) \to \mathbb{R}$ is continuous. Prove that $\varphi$ is measurable with respect to Borel sigma-fields.

**Cumulative distribution functions.**

**Exercise 4.3.** Consider probability space $(0, 1), \mathcal{B}, \lambda$). Suppose $X : (0, 1) \to \mathbb{R}$ is given by $X(\omega) = \ln(\omega)$. Find the CDF of $X$.

**Exercise 4.4.** Suppose $X$ has CDF $F$. Let $Y = X^2$. What is the CDF of $Y$?

**Exercise 4.5.** Suppose $X$ has CDF $F$. Let $Y = X I_{|X| \leq M}$ be the truncation of r.v. $X$ at level $M$. What is the CDF of $Y$?

**Exercise 4.6.** Suppose $U$ is uniform on $(0, 1)$. Let $X = U^2$, $Y = U^3$. What is their joint CDF? (See (4.2).)

**Exercise 4.7 [Statistics].** Use the second proof of Theorem 4.4 to describe how to simulate exponential random variables (see Example 2.4) using a random number generator that produces uniform $U(0,1)$ random variables.

**Independence.**

**Exercise 4.8.** Consider $\omega = [0, 1]$ with Lebesgue measure and the measure-preserving map $f$ defined in Exercise 2.3. Show that events $A := [0, 1/2]$, $B := f^{-1}(A)$ and $C := f^{-1}(B)$ are independent. (This is one of the possible answers to Exercise 3.3.)
Convergence.

Exercise 4.9. Suppose \( F, G \) are two cumulative distribution functions. Define \( \varphi_G(u) X \) has CDF \( F \). Let \( Y = X^2 \). What is the CDF of \( Y ? \)

Exercise 4.10. Suppose random variables \( X_n = \begin{cases} n & \text{with probability } p_n \\ 0 & \text{with probability } 1 - p_n \end{cases} \)

Prove that
1. if \( p_n \to 0 \) then \( X_n \overset{P}{\to} 0 \).
2. if \( \sum_n p_n < \infty \) then \( X_n \to 0 \) with probability 1.
3. if \( X_n \) are independent then \( X_n \to 0 \) with probability 1 iff and only if \( \sum p_n < \infty \)

Exercise 4.11. Prove that if \( X_n \overset{P}{\to} X \) and \( Y_n \overset{P}{\to} Y \) then \( X_n + Y_n \overset{P}{\to} X + Y \).

Exercise 4.12. Suppose \( X_n \overset{P}{\to} X \). Show that \( \{X_n\} \) is stochastically bounded (which is the same as the sequence of laws being tight, compare Exercise 1.13), i.e. for every \( \varepsilon > 0 \) there exists \( K > 0 \) such that for all \( n \) we have \( P(|X_n| > K) < \varepsilon \).

Exercise 4.13. Use the result from Exercise 4.12 to prove that if \( X_n \overset{P}{\to} X \) and \( Y_n \overset{P}{\to} Y \) then \( X_n Y_n \overset{P}{\to} XY \).

Exercise 4.14. Suppose \( U_1, U_2, \ldots, U_n, \ldots \) are independent identically distributed \( U(0, 1) \) random variables (i.e. with cumulative distribution function \( F(x) = x \) for \( 0 < x < 1 \), see Example 2.2). Show that the sequence \( Z_n = U_1 U_2 \ldots U_n \) converges with probability 1.

Additional Exercises

Exercise 4.15. Suppose \( X : \Omega \to \mathbb{R} \) and \( Y : \Omega \to \mathbb{R} \) are two measurable functions (with respect to the Borel \( \sigma \)-field \( \mathcal{B}(\mathbb{R}) \). Prove that \((X, Y) : \Omega \to \mathbb{R}^2 \) is measurable (with respect to the Borel \( \sigma \)-field \( \mathcal{B}(\mathbb{R}^2) \). (Hint: Proposition 4.3.)

Exercise 4.16. Prove that \( \sigma(X) \) as defined in the notes (as \( X^{-1}(\mathcal{B}) \)) is in fact the smallest \( \sigma \)-field for which \( X \) is measurable. (This is the definition of \( \sigma(X) \) in [Billingsley].)

Exercise 4.17. Suppose \( X, Y \) are random variables with cumulative distribution functions \( F(x) \) and \( G(y) \), constructed as in the second proof of Theorem 4.4. Find the joint cumulative distribution function of \( X, Y \).

Exercise 4.18 (Statistics). Suppose \( X, Y \) are independent \( N(0, 1) \) random variables. Verify that \( X^2 + Y^2 \) is exponential. Hint: use polar coordinates.

Exercise 4.19. Suppose that \( X_1 \leq X_2 \leq \cdots \leq X_n \leq X_{n+1} \leq \cdots \). If \( X_n \overset{P}{\to} X \), show that \( X_n \to X \) with probability one.
Exercise 4.20. Suppose \( f : \mathbb{R} \to \mathbb{R} \) is a continuous function and \( X_n \xrightarrow{p} X \). Prove that \( Y_n = f(X_n) \) converges in probability to \( Y = f(X) \).

Exercise 4.21. Suppose \( X_n \xrightarrow{d} X \) and \( X_n \) are independent. Show that there is \( a \in \mathbb{R} \) such that the cumulative distribution of \( X \) is \( F(x) = \begin{cases} 0 & x < a \\ 1 & x \geq a \end{cases} \).
Bibliography

[Gut] A. Gut, Probability: a graduate course
[Proschan-Shaw] S M. Proschan and P. Shaw, Essential of Probability Theory for Statistitcians,
CRC Press 2016
Index

$L − p$-norm, 50
diadic interval, 105
discrete random variable, 56
discrete random variables, 42
equal in distribution, 40
events, 11, 15
expected value, 47
Exponential distribution, 57
exponential distribution, 27
Fatou’s lemma, 55
field, 11
finite dimensional distributions, 30
finitely-additive probability measure, 12
Fubini’s Theorem, 62
Geometric distribution, 57
Hölder’s inequality, 50, 58
inclusion-exclusion, 16
independent $σ$-fields, 33
independent events, 33
independent identically distributed, 42
indicator functions, 8
induced measure, 40
infinite number of tosses of a coin, 105
integrable, 54
intersection, 8
Jensen’s inequality, 49
joint cumulative distribution function, 28
joint distribution of random variables, 40
Kolmogorov’s maximal inequality, 67
Kolmogorov’s one series theorem, 68
Kolmogorov’s three series theorem, 68
Kolmogorov’s two series theorem, 68
Kolmogorov’s zero-one law, 67
Kronecker’s Lemma, 69
Lévy distance, 79
Lebesgue’s dominated convergence theorem, 55

Bernoulli random variables, 43
Binomial distribution, 15, 56
Bonferroni’s correction, 16
Boole’s inequality, 16
Borel $σ$-field, 39
Borel sigma-field, 14
Cantelli’s inequality, 52
cardinality, 9
Cauchy distribution, 87
Cauchy-Schwarz inequality, 49
centered, 51
Central Limit Theorem, 91
characteristic function, 83
characteristic function – continuity theorem, 87
Characteristic functions – uniqueness, 86
Characteristic functions – inversion formula, 86
Chebyshev’s inequality, 49
complex numbers, 82
covariance matrix, 100
cumulative distribution function, 24, 40
cylindrical sets, 30
cylindrical sets, 29
DeMorgan’s law, 8
density function, 27
distribution of a random variable, 40

$λ$-system, 23

$π$-system, 23

$σ$-field, 14

$σ$-field generated by $X$, 39

expected value, 47
Exponential distribution, 57
exponential distribution, 27
Lebesgue’s dominated convergence theorem – used, 56, 66, 75, 88
Levy’s theorem, 70
Lindeberg condition, 93
Lyapunov’s condition, 94
Lyapunov’s inequality, 49

marginal cumulative distribution functions, 28
Markov’s inequality, 49
maximal inequality, Etemadi’s, 70
maximal inequality, Kolmogorov’s, 67
measurable function, 39
measurable rectangle, 61
Minkowski’s inequality, 50
Minkowski’s inequality, 58
moment generating function, 60
Monotone Convergence Theorem, 55
multivariate normal, 99
multivariate normal distribution, 100

negative binomial distribution, 16
normal distribution, 27

Poisson distribution, 16, 57
Polya’s distribution, 16
Portmanteau Theorem, 75
power set, 7
probability, 11
probability measure, 12
probability space, 11, 15
product measure, 62

quantile function, 41, 75

random element, 39
random variable, 40
random vector, 40

sample space, 11
Scheffe’s theorem, 73
section, 61
semi-algebra, 13
semi-ring, 13
sigma-field generated by $\mathcal{A}$, 14
simple random variable, 47
Skorohod’s theorem, 75
Slutsky’s Theorem, 74
Standard normal density, 57
stochastic process with continuous trajectories, 40
stochastically bounded, 46
symmetric distribution, 71

tail $\sigma$-field, 34
Tail integration formula, 63
Taylor polynomials, 81
tight, 46
tight probability measure, 17
Tonelli’s theorem, 62
truncation of r.v., 45

uncorrelated, 51
uniform continuous, 26
Uniform density, 57
uniform discrete, 26
uniform singular, 26
uniformly integrable, 56, 77
union, 8

variance, 48
zero-one law, 34, 67