Chapter 2

Probability measures

The main results about probability measures are the following two facts:

**Theorem 2.1** (extension). If $P$ is a (continuous) probability measure on a field $\mathcal{F}_0$ then it has a unique extension to $\mathcal{F} = \sigma(\mathcal{F}_0)$.

**Theorem 2.2** (uniqueness). Suppose $\mathcal{F} = \sigma(\mathcal{A})$ and $\mathcal{A}$ is closed under finite intersections (the so called $\pi$-system). If two probability measures on $\mathcal{F}$ agree on $\mathcal{A}$ then they are equal also on $\mathcal{F}$.

The details are somewhat technical, and not very illuminating.

1. Existence

**Theorem 2.3** (Caratheodory). A (countably additive) probability measure on a field has an extension to the generated $\sigma$-field.

**Proof of Theorem 2.3.** 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 2.4.** $P^*$ and $P$ agree on $\mathcal{F}_0$. 
2. Probability measures

Proof. (Omitted in 2019)
Suppose $A \in F_0$. Clearly, $P^*(A) \leq P(A)$ as an infimum. Given $\varepsilon > 0$ choose $A_n \in F_0$ 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 $A_n \cap A \in F_0$, 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 2.5.** The outer probability has the following properties:

(i) $P^*(\emptyset) = 0$;

(ii) $P^*(A) \geq 0$

(iii) $A \subset B$ implies $P^*(A) \leq P^*(B)$

(iv) $P^*(\bigcup_n A_n) \leq \sum_n P^*(A_n)$

Proof. (Omitted in 2019)
Without loss of generality we may assume $\sum_n P^*(A_n) < \infty$. To prove (4), choose sets $B_{nk} \in F_0$ such that $A_n \subset \bigcup_k B_{nk}$ and $P^*(A_n) \leq \varepsilon/2^n + \sum_k P_0(B_{nk})$. Then $\bigcup_n A_n \subset \bigcup_{n,k} B_{nk}$ and $P^*(\bigcup_n A_n) \leq \sum_{n,k} P_0(B_{nk}) = \sum_n \sum_k P_0(B_{nk}) \leq \varepsilon + \sum_n P^*(A_n)$.

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

$$ P^*(A \cap E) + P^*(A^c \cap E) = P^*(E) \quad \text{for all} \quad E \subset \Omega $$

(2.2)

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

$$ P^*(A \cap E) + P^*(A^c \cap E) \leq P^*(E) \quad \text{for all} \quad E \subset \Omega $$

(2.3)
Lemma 2.6. \( M \) is a field.

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

\[
P^*(E) = P^*(A \cap E) + P^*(A^c \cap E)
= P^*(B \cap A \cap E) + P^*(B^c \cap A \cap E) + P^*(B \cap A^c \cap E) + P^*(B^c \cap A^c \cap E)
\geq P^*(B \cap A \cap E) + P^*((B^c \cap A) \cup (B \cap A^c) \cup (B^c \cap A^c)) \cap E)
\]

Now notice that

\[
(B^c \cap A) \cup (B \cap A^c) \cup (B^c \cap A^c) = ((B^c \cap A) \cup (B^c \cap A^c)) \cup ((B^c \cap A^c) \cup (B \cap A^c))
= B^c \cup A^c = (B \cap A)^c \quad \square
\]

Lemma 2.7. If the sets \( A_n \in M \) are disjoint then

\[
(2.4) \quad P^* \left( E \cap \bigcup_n A_n \right) = \sum_n P^*(E \cap A_n)
\]

Note that we do not yet know whether \( \bigcup A_n \in M \), 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 \( E \cap (A_1 \cup A_2) = (E \cap (A_1 \cup A_2) \cap A_1) \cup (E \cap (A_1 \cup A_2) \cap A_2) \) and use definition (2.2) with \( E \) replaced by \( E \cap (A_1 \cup A_2) \). This gives

\[
P^* (E \cap (A_1 \cup A_2)) = P^* (E \cap (A_1 \cup A_2) \cap A_1) + P^* (E \cap (A_1 \cup A_2) \cap A_2)
\]

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 \( E \cap (A_1 \cup A_2) \cap A_2 = E \cap A_2 \), so (2.4) hold for \( n = 2 \) sets.

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

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 2.5. \( \square \)
Lemma 2.8. $\mathcal{M}$ 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 $P^*(E) \geq \sum_{n=1}^{\infty} P^*(E \cap A_n) + P^*(E \cap A_n^c)$. 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. □

Lemma 2.9. 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 2.10. $\mathcal{F}_0 \subset \mathcal{M}$

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

Fix $\epsilon > 0$ and let $A_n \in \mathcal{F}_0$ be such that $E \subset \bigcup A_n$ and $\epsilon + P^*(E) > \sum P(A_n)$.

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 $P^*(A \cap E) + P^*(A^c \cap E) \leq \sum P(A \cap A_n) + \sum P(A^c \cap A_n)$. By finite additivity, $P^*(A \cap E) + P^*(A^c \cap E) \leq \sum P(A_n) < P(E) + \epsilon$. □

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 $\mathcal{F}_0$. □
2. Uniqueness

Remark 2.11. $P_s(A) = 1 - P^s(A)$ 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 2.12. 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 $P(A) \leq \sum_{n=1}^{\infty} P(B_n) < P(A) + \varepsilon/2$. 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

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

2. Uniqueness

This section is based on [Billingsley, Section 3].

Theorem 2.13. A (countably additive) probability measure on a field has a unique extension to the generated $\sigma$-field.

In view of Theorem 2.3, 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 16.

2.1. Dynkin’s $\pi$-$\lambda$ Theorem.

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

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

Examples of $\pi$-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 $\lambda$-system if

(\(\lambda_1\)) $\Omega \in \mathcal{L}$.
(\(\lambda_2\)) $A \in \mathcal{L}$ implies $A^c \in \mathcal{L}$.
(\(\lambda_3\)) If $A_1, A_2, \ldots, A_n, \cdots \in \mathcal{L}$ are (pairwise) disjoint then $\bigcup_n A_n \in \mathcal{L}$.

Remark 2.14. From (\(\lambda_1\)) and (\(\lambda_2\)) we see that $\emptyset \in \mathcal{L}$. So if $A, B \in \mathcal{L}$ are disjoint then by (\(\lambda_3\)) 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 2.15.** 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.

(Omitted in 2019)

**Lemma 2.16.** Suppose \( \mathcal{P} \) is a π-system and \( \mathcal{L}_0 \) is the λ-system generated by \( \mathcal{P} \). Then \( \mathcal{L}_0 \) is a σ-field.

**Sketch of proof.** Because of Lemma 2.15, to show that \( \mathcal{L}_0 \) is a σ-field it is enough to show that it is a π-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 \( \mathcal{C}_A \) of all sets \( B \) such that \( A \cap B \in \mathcal{L}_0 \). This collection turns out to be a λ-system. Since \( \mathcal{P} \subset \mathcal{C}_A \), we have \( \mathcal{L}_0 \subset \mathcal{C}_A \). 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 \( \mathcal{C}_B \) of all sets \( A \) such that \( A \cap B \in \mathcal{L}_0 \). By the previous part, \( \mathcal{P} \subset \mathcal{C}_B \). Again, \( \mathcal{C}_B \) turns out to be a λ-system, so \( \mathcal{L}_0 \subset \mathcal{C}_B \). 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 \( \mathcal{C}_A \) and \( \mathcal{C}_B \) are λ-systems. This proof is omitted.

(Omitted in 2019)

**Theorem 2.17** (Dynkin’s π-λ Theorem). Suppose a λ-system \( \mathcal{L} \) includes a π-system \( \mathcal{P} \). Then \( \sigma(\mathcal{P}) \subset \mathcal{L} \).

**Proof.** Let \( \mathcal{L}_0 \) be a λ-system generated by \( \mathcal{P} \). Then \( \mathcal{P} \subset \mathcal{L}_0 \subset \mathcal{L} \). From Lemma 2.16 we know that \( \mathcal{L}_0 \) is a σ-field and it contains \( \mathcal{P} \). So \( \sigma(\mathcal{P}) \subset \mathcal{L}_0 \subset \mathcal{L} \).

(Omitted in 2019)

**Proposition 2.18.** Let \( \mathcal{P} \) be a π-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 λ-system. By Theorem 2.17 \( \mathcal{F} \subset \mathcal{L} \).

**Proof of Theorem 2.13.** A field \( \mathcal{F}_0 \) is a π-system. So if \( P_1(A) = P_2(A) \) for all \( A \in \mathcal{F}_0 \), then by Proposition 2.18 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 1.6, with Caratheodory’s Theorem 2.3 and uniqueness Theorem 2.13.

**Proposition 2.19.** 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 2.13.

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 Theorem 1.5, one needs to show that this definition is consistent, that \( P \) is finitely-additive, and that \( P \) is continuous (countably-additive) on \( \mathcal{B}_0 \). Once we prove this, we invoke Theorem 2.3.

(Omitted in 2019)

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 1.11.

\( \square \)
Solution of Exercise 1.11. We prove the contrapositive to the implication in Remark 1.4(3).

Suppose \( A_1 \supseteq A_2 \supseteq \ldots \) 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 \( P(A_n) - P(B_1 \cap B_2 \cap \cdots \cap B_n) = P(A_n \setminus B_1 \cap B_2 \cap \cdots \cap B_n) \leq P(\bigcup_{j=1}^n (A_j \setminus B_j)) \leq \sum_{j=1}^n P(A_j \setminus B_j) = \sum_{j=1}^n (P(A_j) - P(B_j)) < \delta/2 \). (As \( (A_n \setminus B_1 \cap B_2 \cap \cdots \cap B_n) = \bigcup_{j=1}^n A_n \cap B_j' \subset \bigcup_{j=1}^n A_k \cap B_j' = \bigcup_{j=1}^n (A_j \setminus B_j) \).) Since \( P(A_n) > \delta \) this shows that \( P(B_1 \cap B_2 \cap \cdots \cap B_n) > \delta/2 > 0 \). In particular, \( K_1 \cap \cdots \cap K_n \cap B_1 \cap B_2 \cap \cdots \cap B_n \neq \emptyset \).

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 \( \bigcap_{n=1}^\infty A_n \supset \bigcap_{n=1}^\infty K_n \neq \emptyset \). \( \square \)

### 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 1.3 and 1.4) \( P = \frac{1}{n} \sum_{j=1}^n \delta_{x_j} \) 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

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

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,

$$F(x) = \begin{cases} 0 & x < 0 \\ \vdots & \vdots \\ 1/4 & 1/9 \leq x < 2/9 \\ \vdots & \vdots \\ 1/2 & 1/3 \leq x < 2/3 \\ \vdots & \vdots \\ 3/4 & 7/9 \leq x < 8/9 \\ \vdots & \vdots \\ 1 & x \leq 1 \end{cases}$$

The interval removed in $d$-th step is $(\sum_{k=1}^{d} x_k/3^k, \sum_{k=1}^{d} x_k/3^k+1/3^d)$ 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.

![Graph of $F(x)$](image)

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\(^2\).

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

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

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

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

Thus

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

\(^2\)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 of loss that changes when the loss exceeds some predefined thresholds.
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 2.19.

**Proposition 2.20.** 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 1.16.

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 1.11 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 $F(B_1, B_2) > 1 - \varepsilon$.

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

**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

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

Noting that $A$ is disjoint with the remaining set, by the inclusion-exclusion formula (1.2), we get

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

So

\[(2.9) \quad P(A) = \Delta_A(F) :=
F(b_1, b_2, b_3) + F(a_1, a_2, b_3) + F(b_1, a_2, a_3) - F(a_1, b_2, b_3) - F(b_1, a_2, a_3) - F(b_1, b_2, a_3) - F(a_1, a_2, a_3)
\]

An analog of Definition 2.4 uses $\Delta_A(F)$ as defined in (2.9). Proposition 2.20 has an $\mathbb{R}^3$ 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 (1.2). 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 (A.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 4.9. 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 $\mathbb{R}^\infty$. 
4. Probability measures on \( \mathbb{R}^k \)

(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 \( P_{t_1,\ldots,t_n} \) 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 \( \Omega = [0,1] \). 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 \( \Omega = [0,1] \).

Theorem 2.21 (Kolmogorov). Suppose probability measures \( P_{t_1,\ldots,t_n} \) are consistent. Then there exists a unique probability measure \( \Pr \) on \( [0,1] \) with Borel \( \sigma \)-field that generates \( P_{t_1,\ldots,t_n} \) as finite dimensional distributions.

Remark 2.22. A good description of Borel \( \sigma \)-field in \( [0,1] \) appears in [Billingsley, Section 36]. In particular, the subset \( C[0,1] \subset [0,1] \) 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 1.11 to verify that if \( A_n \) is a decreasing family of sets in \( \mathcal{F}_0 \) with \( \bigcap_n A_n = \emptyset \) then \( \Pr(A_n) \to 0 \).
2. Probability measures

5. Lebesgue measure

Occasionally we need to consider positive measures on a σ-field that are not probability measures. The main example is the Lebesgue measure \( \lambda \) on \((\mathbb{R}, \mathcal{B})\), which corresponds to strictly increasing and continuous but unbounded function \( F(x) = x \) and the Lebesgue measure \( \lambda \) on \((\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})\).

6. Transport of measure

Suppose we have two measure spaces \( \Omega_1, \mathcal{F}_1 \) and \( \Omega_2, \mathcal{F}_2 \).

**Definition 2.5.** We say that a function \( \varphi : \Omega_1 \to \Omega_2 \) is measurable if for every \( B \in \mathcal{F}_2 \) its inverse image \( \varphi^{-1}(B) \in \mathcal{F}_2 \).

Recall that \( \varphi^{-1}(B) = \{ \omega \in \Omega_1 : \varphi(\omega) \in B \} \).

A transport of measure is a construction of a probability measure on \((\Omega_2, \mathcal{F}_2)\) from measure \( P_1 \) and function \( \varphi \). This topic will return in Definition 4.3, where \( P_2 \) is called the induced measure. The definition of \( P_2 \) is as follows. For \( B \in \mathcal{F}_2 \), we define

\[
P_2(B) := P_1(\varphi^{-1}(B))
\]

It is known that a continuous function \((0, 1) \to \mathbb{R}\) is measurable. We will show this in Proposition 4.5 on page 40, and for now we use it without proof to give examples with \( \Omega_1 = (0, 1) \) with Lebesgue measure \( \lambda \).

**Example 2.8.** If \( f : (0, 1) \to \mathbb{R} \) is \( \varphi(t) = \ln t \) then the CDF of \( P_2 \) is

\[
P_2((\infty, x]) = \lambda(\{t \in (01) : \varphi(t) \leq x\}) = \lambda(\{t \in (01) : e^x\}) = \lambda((0, 1) \cap (\infty, e^x)) = \begin{cases} 1 & x \geq 0 \\ e^x & x < 0 \end{cases}
\]

Measure \( P_2 \) has density

\[
f(x) = \begin{cases} 0 & x > 0 \\ e^x & x < 0 \end{cases}
\]

**Example 2.9.** If \( \varphi : (0, 1) \to \mathbb{R} \) is \( \varphi(t) = \tan(\pi t - \frac{\pi}{2}) \) then a similar calculation shows that the CDF of the induced measure \( P_2 \) on \((\mathbb{R}, \mathcal{B})\) is \( F_2(x) = \frac{1}{2} + \frac{1}{\pi} \arctan x \). The corresponding density

\[
f(x) = \frac{1}{\pi} \frac{1}{1 + x^2}
\]

is known as the **Cauchy density**.
**Additional Exercises**

**Required Exercises**

**Exercise 2.1** (Different representations of the same probability measure on a $\sigma$-field $\mathcal{F}$). 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 1.4) such that $\lambda$ restricted to $\mathcal{F}$ coincides with $P$ restricted to $\mathcal{F}$. (In formal notation, $\lambda|_{\mathcal{F}} = P|_{\mathcal{F}}$.)

**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 distribution is easy to simulate by taking $2\sum_{k=1}^{\infty} \varepsilon_k/3^k$ where $\varepsilon_k$ represents a “toss of a fair coin” with values 0 or 1. This exercise asks you to reproduce some of the histograms that appear in [Proschan-Shaw].

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

$$
\varphi(x) = \begin{cases} 
2x & \text{if } x \leq 1/2 \\
2x - 1 & \text{if } x > 1/2
\end{cases}
$$

Show that for every Borel subset $A$ of $[0, 1]$ the Lebesgue measure of $\varphi^{-1}(A)$ equals to the Lebesgue measure of $A$. (Compare Exercise 2.12.)

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

**Exercise 2.5**. Consider $(\Omega_1, \mathcal{F}, P_1) = ((0, 1), \mathcal{B}, \lambda)$ and measurable function $\varphi : (0, 1) \to \mathbb{R}$ given by $\varphi(t) = 4t(1-t)$. Find the CDF for the measure induced by $\varphi$ on $(\mathbb{R}, \mathcal{B})$.

**Additional Exercises**

**Exercise 2.6**. 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 2.7**. Inspect the proofs of Theorems 2.3 and 2.13. Find all places where additivity or countable additivity is used.

**Exercise 2.8** (Compare Exercise 1.17). 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 2.9**. 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 2.10. Let $\mathcal{A}$ be the smallest field generated by a $\pi$-system $\mathcal{P}$ (see Exercise 1.16). 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 2.11. 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 2.12. 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 $B$. 


Bibliography

[Gut] A. Gut, Probability: a graduate course
# Index

$L_1$ metric, 11
$L_2$ metric, 11
$L_p$-norm, 59, 76
$\lambda$-system, 25
$\pi$-system, 25
$\sigma$-field, 16
$\sigma$-field generated by $X$, 41
distribution of a random variable, 42
Bernoulli random variables, 46
Binomial distribution, 17, 73
bivariate cumulative distribution function, 30
Bonferroni's correction, 18
Boole's inequality, 18
Borel $\sigma$-field, 41
Borel sigma-field, 16
Cantelli's inequality, 65
cardinality, 9
Cauchy distribution, 114
Cauchy-Schwarz inequality, 58
centered, 63
Central Limit Theorem, 117
characteristic function, 111
characteristic function – continuity theorem, 114
Characteristic functions – uniqueness, 113
Characteristic functions – inversion formula, 113
Chebyshev's inequality, 57
complex numbers, 108
conjugate exponents, 60
continuity condition, 14
Continuous mapping theorem, 103
converge in $L_p$, 61
converge in mean square, 61
convergence in distribution, 49, 99
Convergence of types, 119
converges in distribution, 127
converges in probability, 46
converges pointwise, 7
converges uniformly, 7
converges with probability one, 47
correlation coefficient, 58
countable additivity, 14
covariance matrix, 130
cumulative distribution function, 26, 43
cylindrical sets, 32, 33
cylindrical sets, 32
de Moivre formula, 108
DeMorgan’s law, 8
density function, 29, 74
diadic interval, 135
discrete random variable, 73
discrete random variables, 45
equal in distribution, 43
events, 13, 17
expected value, 54, 69
Exponential distribution, 74
exponential distribution, 29
Fatou's lemma, 71
field, 13
finite dimensional distributions, 32
finitely-additively probability measure, 14
Fubini’s Theorem, 82
Geometric distribution, 73
Hölder’s inequality, 60, 76
inclusion-exclusion, 18
independent $\sigma$-fields, 35
independent events, 35
independent identically distributed, 46
independent random variables, 44
indicator functions, 9
induced measure, 42
infinite number of tosses of a coin, 135
integrable, 69
intersection, 8
Jensen’s inequality, 58
joint cumulative distribution function, 30
joint distribution of random variables, 43
<table>
<thead>
<tr>
<th>Term</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kolmogorov’s maximal inequality</td>
<td>91</td>
</tr>
<tr>
<td>Kolmogorov’s one series theorem</td>
<td>91</td>
</tr>
<tr>
<td>Kolmogorov’s three series theorem</td>
<td>92</td>
</tr>
<tr>
<td>Kolmogorov’s two series theorem</td>
<td>92</td>
</tr>
<tr>
<td>Kolmogorov’s zero-one law</td>
<td>90</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov metric</td>
<td>10, 11</td>
</tr>
<tr>
<td>Kronecker’s Lemma</td>
<td>94</td>
</tr>
<tr>
<td>Lévy distance</td>
<td>106</td>
</tr>
<tr>
<td>law of $X$, 42</td>
<td></td>
</tr>
<tr>
<td>Lebesgue’s dominated convergence theorem</td>
<td>71, 72</td>
</tr>
<tr>
<td>Lebesgue’s dominated convergence theorem – used</td>
<td>73, 89, 102, 115</td>
</tr>
<tr>
<td>Levy’s metric</td>
<td>11</td>
</tr>
<tr>
<td>Levy’s theorem</td>
<td>93</td>
</tr>
<tr>
<td>Lindeberg condition</td>
<td>120</td>
</tr>
<tr>
<td>Lyapunov’s condition</td>
<td>121</td>
</tr>
<tr>
<td>Lyapunov’s inequality</td>
<td>58</td>
</tr>
<tr>
<td>marginal cumulative distribution functions</td>
<td>30</td>
</tr>
<tr>
<td>Markov’s inequality</td>
<td>57</td>
</tr>
<tr>
<td>maximal inequality, Etemadi’s</td>
<td>95</td>
</tr>
<tr>
<td>maximal inequality, Kolmogorov’s</td>
<td>91</td>
</tr>
<tr>
<td>mean square convergence</td>
<td>76</td>
</tr>
<tr>
<td>measurable function</td>
<td>41</td>
</tr>
<tr>
<td>measurable rectangle</td>
<td>81</td>
</tr>
<tr>
<td>metric, 10</td>
<td></td>
</tr>
<tr>
<td>metric space</td>
<td>10</td>
</tr>
<tr>
<td>Minkowski’s inequality</td>
<td>59</td>
</tr>
<tr>
<td>Minkowski’s inequality, 76</td>
<td></td>
</tr>
<tr>
<td>moment generating function</td>
<td>56, 77</td>
</tr>
<tr>
<td>moments, 55</td>
<td></td>
</tr>
<tr>
<td>Monotone Convergence Theorem, 69</td>
<td></td>
</tr>
<tr>
<td>multivariate normal</td>
<td>129</td>
</tr>
<tr>
<td>multivariate normal distribution, 130</td>
<td></td>
</tr>
<tr>
<td>multivariate random variable, 41</td>
<td></td>
</tr>
<tr>
<td>negative binomial distribution, 18</td>
<td></td>
</tr>
<tr>
<td>normal distribution, 29</td>
<td></td>
</tr>
<tr>
<td>Poisson distribution, 18, 73</td>
<td></td>
</tr>
<tr>
<td>Polya’s distribution</td>
<td>18</td>
</tr>
<tr>
<td>Portmanteau Theorem, 102</td>
<td></td>
</tr>
<tr>
<td>power set, 7</td>
<td></td>
</tr>
<tr>
<td>probability, 13</td>
<td></td>
</tr>
<tr>
<td>probability measure</td>
<td>14</td>
</tr>
<tr>
<td>probability space, 13, 17</td>
<td></td>
</tr>
<tr>
<td>product measure</td>
<td>82</td>
</tr>
<tr>
<td>quantile function, 44, 101</td>
<td></td>
</tr>
<tr>
<td>random element</td>
<td>41</td>
</tr>
<tr>
<td>random variable</td>
<td>41</td>
</tr>
<tr>
<td>random vector</td>
<td>41</td>
</tr>
<tr>
<td>sample space, 13</td>
<td></td>
</tr>
<tr>
<td>Scheffe’s theorem, 100</td>
<td></td>
</tr>
<tr>
<td>section, 81</td>
<td></td>
</tr>
<tr>
<td>semi-algebra, 15</td>
<td></td>
</tr>
<tr>
<td>semi-ring, 15</td>
<td></td>
</tr>
<tr>
<td>sigma-field generated by $\mathcal{A}$, 16</td>
<td></td>
</tr>
<tr>
<td>simple random variable</td>
<td>53</td>
</tr>
<tr>
<td>simple random variables</td>
<td>45</td>
</tr>
<tr>
<td>Skorohod’s theorem</td>
<td>101</td>
</tr>
<tr>
<td>Slutsky’s Theorem</td>
<td>101</td>
</tr>
<tr>
<td>Standard normal density</td>
<td>74</td>
</tr>
<tr>
<td>stochastic process with continuous trajectories, 43</td>
<td></td>
</tr>
<tr>
<td>stochastic processes</td>
<td>42</td>
</tr>
<tr>
<td>stochastically bounded</td>
<td>51</td>
</tr>
<tr>
<td>symmetric distribution</td>
<td>97</td>
</tr>
<tr>
<td>tail $\sigma$-field</td>
<td>36</td>
</tr>
<tr>
<td>Tail integration formula</td>
<td>84</td>
</tr>
<tr>
<td>Taylor polynomials</td>
<td>107</td>
</tr>
<tr>
<td>tight, 51</td>
<td></td>
</tr>
<tr>
<td>tight probability measure</td>
<td>19</td>
</tr>
<tr>
<td>Tonelli’s theorem</td>
<td>82</td>
</tr>
<tr>
<td>total variation metric</td>
<td>11</td>
</tr>
<tr>
<td>truncation of r.v., 50</td>
<td></td>
</tr>
<tr>
<td>uncorrelated</td>
<td>63</td>
</tr>
<tr>
<td>uniform continuous</td>
<td>28</td>
</tr>
<tr>
<td>Uniform density</td>
<td>74</td>
</tr>
<tr>
<td>uniform discrete</td>
<td>28</td>
</tr>
<tr>
<td>uniform singular</td>
<td>28</td>
</tr>
<tr>
<td>uniformly integrable</td>
<td>72, 105</td>
</tr>
<tr>
<td>union, 8</td>
<td></td>
</tr>
<tr>
<td>variance, 56</td>
<td></td>
</tr>
<tr>
<td>Waserstein distance</td>
<td>11</td>
</tr>
<tr>
<td>weak convergence</td>
<td>49</td>
</tr>
<tr>
<td>weak law of large numbers</td>
<td>63</td>
</tr>
<tr>
<td>zero-one law, 36, 90</td>
<td></td>
</tr>
</tbody>
</table>