Chapter 4

Random variables

Printed: February 8, 2019

1. Measurable mappings

Suppose Ω and E are two sets. Often E = ℝ or E = ℝ^d.

Suppose X : Ω → E i.e. X is a function with domain Ω 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 \( U \subset E \).

**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) \( X^{-1}(\bigcup_{t \in T} U_t) = \bigcup_{t \in T} X^{-1}(U_t) \)

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

**Corollary 4.2.** If \( \mathcal{B} \) is a σ-field of subsets of \( E \) then \( X^{-1}(\mathcal{B}) \) is a σ-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 σ-field generated by \( X \) is \( \sigma(X) = X^{-1}(\mathcal{B}) \).

Exercise 4.21 says that this is the smallest σ-field of subsets of \( \Omega \) which makes \( X \) measurable.

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

**Definition 4.2.** In analysis, \( X \) is called a measurable function if \( X^{-1}(\mathcal{B}) \subset \mathcal{F} \). In probability, \( X \) is then called a random element of \( E \).
If we want to indicate the \( \sigma \)-fields, we will write \( X : (\Omega, \mathcal{F}) \to (\mathbb{E}, \mathcal{B}) \).

The most important special cases for us are \( \mathbb{E} = \mathbb{R} \) and \( \mathbb{E} = \mathbb{R}^d \). When \( \mathbb{E} = \mathbb{R} \), we say that \( X \) is a random variable. When \( \mathbb{E} = \mathbb{R}^d \), 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. 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 \( \mathcal{F} \).

Somewhat more generally, we have the following.

**Proposition 4.3.** If \( \sigma(A) = \mathcal{B}_{\mathbb{E}} \) and \( X^{-1}(A) \subset \mathcal{F} \) then \( X : (\Omega, \mathcal{F}) \to (\mathbb{E}, \mathcal{B}) \) is measurable with respect to \((\mathcal{F}, \mathcal{B}_{\mathbb{E}})\).

**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 \( A \in \mathcal{A} \), the inverse image of the set \( A \) is in \( \mathcal{F} \), so \( A \in \mathcal{U} \). Thus \( \mathcal{A} \subset \mathcal{U} \), and the generated sigma field \( \sigma(A) = \mathcal{B}_{\mathbb{E}} \) is in \( \mathcal{U} \).

\[ \square \]

**Remark 4.4.** The collection \( X_1, \ldots, X_d \) of random variables (on the same probability space) defines random vector \((X_1, \ldots, X_d)\). (For \( d = 2 \), this is Exercise 4.20.)

We also remark that random elements of spaces of functions, such as \( \mathbb{E} = C[0,1] \), the space of all continuous functions on \([0,1]\), or \( \mathbb{E} = D[0,\infty) \), the space of right-continuous functions with left limits, are called stochastic processes rather than random functions. So we say ”Wiener process” or ”Poisson process”, rather than random continuous function, or random piecewise-linear function.

The following properties are often useful.

**Proposition 4.5.** Consider \( \mathbb{R} \) or \( \mathbb{R}^d \) with Borel sigma field.

- If \( A \in \mathcal{F} \) then \( I_A : \Omega \to \mathbb{R} \) is measurable.
- A continuous function \( \mathbb{R} \to \mathbb{R} \) is measurable.
- A continuous function \( \mathbb{R}^m \to \mathbb{R}^n \) is measurable.
- Composition of measurable transformations is measurable.
- Sum of two measurable functions is measurable.
- Product of two measurable functions is measurable.
- A pointwise limit of a sequence of measurable functions \( \mathbb{R} \to \mathbb{R} \) is a measurable function.

For example, as a product of two measurable function \( x \mapsto e^{it}I_{(a,b)}(x) \) is a measurable function \((\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})\). For measurable \( f, g \) the sum \( f + g \) is measurable as a composition of \((f, g) : \Omega \to \mathbb{R}^2 \) with continuous function \((x,y) \mapsto x + y \).

**1.2. Induced probability measures.**

**Definition 4.3.** The distribution of a random variable \( X : (\Omega, \mathcal{F}) \to (\mathbb{E}, \mathcal{B}) \) is a probability measure \( \mu \) on \((\mathbb{E}, \mathcal{B})\) defined by

\[
\mu(U) = P(X^{-1}(U))
\]
Sometimes $\mu$ is called an induced measure and some authors use notation $Q = P \circ X^{-1}$. We will sometimes write $L(X) = \mu$ and say that $\mu$ is the law of $X$.

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

$$F(x) = \mu((\infty, x]) = P(\{\omega : X(\omega) \leq x\})$$

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 2.19, this is equivalent to $X, Y$ having the same cumulative distribution function.

**Example 4.1.** Consider $\Omega = [0, 1]$ with Lebesgue measure on Borel subsets. Let $X(\omega) = \omega$ and $Y(\omega) = 1 - \omega$. Then $X, Y$ have the same distribution, which is easily checked by computing both cumulative distribution functions.

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 $\Omega \rightarrow \mathbb{R}^2$. The joint distribution of random variables is just the induced measure on $(\mathbb{R}^2, \mathcal{B})$ 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 $P(\{\omega : X(\omega) \leq x, Y(\omega) \leq y\})$.)

Taken together, pair $(X, Y)$ from Example 4.1 defines a probability measure on $(\mathbb{R}^2, \mathcal{B})$, and Exercise 4.9 asks for its joint cumulative distribution function. Cumulative distribution function of pairs of uniformly distributed random variables are called copulas and are useful in applications.

Our second proof of Theorem 4.10 below shows that one can find an infinite sequence of such functions on $[0, 1]$, and furthermore, unlike in Example 4.1, the random variables can be independent.

---

**Omitted in 2019** If $\mathbb{E} = C[0, 1]$ then a measurable mapping $X : \Omega \rightarrow C[0, 1]$ 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 $[0, 1] \rightarrow \mathbb{R}$ 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].) One way to ensure properties of trajectories that we need, is to construct the Wiener process and the Poisson process directly on some probability space $(\Omega, \mathcal{F}, P)$.

---

2. Random variables with prescribed distributions

This section is based on [Billingsley, Section 14] or [Durrett, Theorem 1.2.2].
Theorem 4.6. If \( F \) is a cumulative distribution function\(^1\), then there exists on some probability space a random variable \( X \) for which \( P(X \leq x) = F(x) \).

\[ \text{(Omitted in 2019)} \]

First proof. Proposition 2.19 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 \( P \). \(\square\)

The second proof is independent of Proposition 2.19, and can be used to prove it.

Second proof. 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 \( \{ x : u \leq F(x) \} \) \(\text{\(\square\) is a closed}^3 \text{ of the form } [\varphi(u), \infty) \text{ and its complement is } \{ x : u > F(x) \} = (-\infty, \varphi(u)). \text{ This shows that for every real } x, \text{ we have } \varphi(u) \leq x \text{ iff } F(x) \geq u. \text{ 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 \( \lambda(\{ \omega : X(\omega) \leq x \}) = \lambda(\{ \omega : \omega \leq F(x) \}) = \lambda([0, F(x)]) = F(x). \)

\[ \square \]

Example 4.2 (Cauchy distribution). Cauchy distribution with density \( \frac{1}{\pi(1+x^2)} \) has cumulative distribution function \( F(x) = \frac{1}{2} + \frac{\arctan(x)}{\pi} \). For Cauchy random variable, the proof above gives formula \( X(\omega) = \tan(\pi \omega - \pi/2) \).

Corollary 4.7 (Proposition 2.19). 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 from \( \pi - \lambda \) Theorem 2.13, see Proof of Proposition 2.19).

Example 4.3. 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 \( F_+(x) = F(x)I_{[0,\infty)}(x) \).

\[ \text{\(\square\)} \]

\(^1\)See Definition 2.3
\(^2\)Can you see why isn’t it \( \mathbb{R} \) or \( \emptyset? \)
\(^3\)Why?
\(^4\)Why?
2. Random variables with prescribed distributions

2.1. Independent random variables. The second proof of Theorem 4.6 lets us construct a finite or an infinite sequence $X_1 = \varphi_1(\omega), X_2 = \varphi_2(\omega), \ldots$ of random variables with prescribed distributions. However, this gives only very special measures on $\mathbb{R}^\infty$, see Exercise 4.22. We now consider another special construction that gives joint distributions that are of more interest.

Definition 4.5. Random variables $X_1, X_2, \ldots$ are independent if the generated $\sigma$-fields $\sigma(X_1), \sigma(X_2), \ldots$ are independent.

In other words, $X_1, X_2, \ldots$ are independent if the events $X_1 \in U_2, X_2 \in U_2, \ldots$ are independent for any Borel sets $U_1, U_2, \ldots$.

Proposition 4.8. Random variables are independent, iff for every $n$ and every $x_1, \ldots, x_n \in \mathbb{R}$ the joint cumulative distribution function factors:

$$F_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = F_1(x_1) \ldots F_n(x_n) \quad (4.4)$$

Proof. Clearly, independence implies (4.4) - take $U_k = (-\infty, x_k]$. (Omitted in 2019) To prove the converse, consider the families of $\pi$-systems $A_k = \{X_1^{-1}((-\infty, x_k)], k = 1, 2, \ldots$. Then (4.4) implies that $A_1, A_2, \ldots$ are independent, so by Corollary 3.3 the generated sigma fields are also independent.

Example 4.4. Consider discrete random variables $X = \sum_{j=1}^{\infty} x_j I_{A_j}, Y = \sum_{k=1}^{\infty} y_k B_k$ with disjoint sets. Then $X, Y$ are independent iff $A = \{\emptyset, A_1, A_2, \ldots\}$ and $B = \{\emptyset, B_1, B_2, \ldots\}$ are independent $\pi$-systems. Thus $X, Y$ are independent iff

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

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

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

Remark 4.9. An important special case of discrete random variables are the simple random variables, which take only a finite number of values.

Example 4.5. 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

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

for all $n$ and all choices of $\varepsilon_1, \ldots, \varepsilon_n \in \{0, 1\}$.

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

Theorem 4.10. If $F_1, F_2, \ldots$ are cumulative distribution functions then there exists a probability space $(\Omega, F, P)$ and a sequence $X_1, X_2, \ldots$, of independent random variables such that $X_n$ has cumulative distribution function $F_n$. 
4. Random variables

(Omitted in 2019)

**Sketch of First Proof.** In this proof we take $\Omega = \mathbb{R}^\infty$ with (infinite!) product measure $P = P_1 \otimes P_2 \otimes \ldots$ 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.** The second proof is based on the idea that digits of $\omega \in (0, 1)$ under Lebesque measure are independent and can be arranged into infinite number of sequences of (still independent) digits. This can be done in many ways, for example if $\omega = .d_1 d_2 \ldots$ then we can rearrange its digits into

\[
    \begin{array}{cccccc}
        d_1 & d_2 & d_6 & d_7 & \ldots \\
        d_3 & d_5 & d_8 & \phantom{d} & \phantom{d} \\
        d_4 & d_9 & \phantom{d} & \phantom{d} & \phantom{d} \\
        d_{10} & \phantom{d} & \phantom{d} & \phantom{d} & \phantom{d} \\
        \vdots & \phantom{d} & \phantom{d} & \phantom{d} & \phantom{d}
    \end{array}
\]

splits $\omega \in (0, 1)$ into the infinite sequence of numbers $\omega_1 = .d_1 d_2 d_6 d_7 \ldots$, $\omega_2 = .d_3 d_5 d_8 \ldots$, $\omega_3 = .d_4 d_9 \ldots$, and so on.

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 B.1 we have $\lambda(d_1 = \varepsilon_1, \ldots, d_m = \varepsilon_m) = 1/2^m$. By Example 4.5 this proves independence.

Next, we arrange all of these random variables into an infinite array $d_{i,j}$. Then by Corollary 3.5 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 4.6.** 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.

Data collected from repeated runs of an experiment in statistics are modeled by i. i. d. random variables.

2.2. Elementary examples. The following is a repeat of one of the points in Proposition 4.5.

**Proposition 4.11.** 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 the probability space $(\Omega, \mathcal{F}, P)$, then $Y = f(X_1, \ldots, X_d)$ is a random vector.

**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 $Y^{-1}(B) = (X_1, \ldots, X_d)^{-1}(U) \in \mathcal{F}$. □

Here are some examples of such functions:

**Proposition 4.12 (Sum theorems).** Suppose $X_1, X_2, \ldots$ are independent and $S = X_1 + X_2 + \cdots + X_n$.

\(^6\)For more details see [Billingsley, Theorem 20.4]. (This proof also answers Exercise 3.2.)

\(^7\)Observe that the diadic intervals $[0, k/2^n]$ with $k, n \in \mathbb{N}$ form a $\pi$-system that generates $\mathcal{B}$
(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 Bin($n, p$) (see Example 1.5)

(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 1.6)

(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$^8$

3. Convergence of random variables

The following definition requires us to know that $X_n - X$ is measurable. This is a consequence of Exercise 4.20.

**Definition 4.7.** A sequence of random variables converges in probability to a random variable $X$, abbreviated as $X_n \xrightarrow{P} X$, if for every $\varepsilon > 0$, we have

$$\lim_{n \to \infty} P(\{\omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\}) = 0$$

We will use the abbreviated notation

$$\lim_{n \to \infty} P(|X_n - X| \geq \varepsilon) = 0.$$

It is also clear that it is enough to consider only rational $\varepsilon > 0$.

**Example 4.6.** On $\Omega = [0,1]$ consider $X_n = I_{[0,n/(2n+1)]}$. Then $X_n \xrightarrow{P} X$. In act, here convergence holds for every $\omega$. Exercise 4.12 shows that this does not have to be so. See also next example.

**Example 4.7.** Suppose $\Omega$ is a unit circle with (with probability measure from arclength, i.e. measure induced by $\theta \mapsto (\cos \theta, \sin \theta)$). Suppose $X_n = I_{\Theta_n}$ where $\Theta_n$ are consecutive arcs of length $1/n$ on the unit circle. Then $X_n \xrightarrow{P} 0$, but for every $\omega$ the sequence $X_n(\omega)$ does not converge.

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

**Proposition 4.13.**

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

$^8$These are “elementary” facts covered in undergraduate courses.
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 4.20.

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

Finally, we note that $\bigcap_{\varepsilon > 0} A_\varepsilon = \bigcap_n A_{1/n} \in \mathcal{F}$. □

In view of the above proposition, the probability that a sequence of random variables converges to a random variable is well-defined.

**Definition 4.8.** A sequence of random variables converges with probability one if

$$P\left( \{ \omega \in \Omega : \lim_{n \to \infty} X_n(\omega) = X(\omega) \} \right) = 1$$

**Proposition 4.14.** If $X_n \to X$ with probability one, then $X_n \xrightarrow{P} X$.

**Proof.** We need to observe that $P(\forall \varepsilon > 0 \exists N \forall n > N \{ \omega : |X_n(\omega) - X(\omega)| < \varepsilon \}) = 1$ iff for every rational $\varepsilon > 0$

$$P(\exists N \forall n > N \ |X_n - X| < \varepsilon) = P\left( \bigcup_N \bigcap_{n > N} |X_n - X| < \varepsilon \right) = 1$$

This is the same as

$$P\left( \bigcap_N \bigcup_{n > N} |X_n - X| > \varepsilon \right) = P(\sup_{n > N} |X_n - X| > \varepsilon \text{ i.o.}) = 0$$

Now $P\left( \bigcap_N \bigcup_{n > N} |X_n - X| > \varepsilon \right) = \lim_{N \to \infty} P\left( \bigcup_{n > N} |X_n - X| > \varepsilon \right)$. So convergence with probability one is equivalent to

$$\forall \varepsilon > 0 \lim_{N \to \infty} P(\sup_{n > N} |X_n - X| > \varepsilon) = 0.$$ (4.5)

Of course, $P(|X_N - X| > \varepsilon) \leq P(\sup_{n > N} |X_n - X| > \varepsilon)$. □

**Proposition 4.15.** Suppose $X_n \xrightarrow{P} X$. Then there exists a subsequence $n_k$ such that $X_{n_k} \to X$ with probability one.

**Proof.** Choose positive $\varepsilon_k \to 0$, say $\varepsilon_k = 1/k$. Fix $k \in \mathbb{N}$. By convergence in probability, $\lim_{n \to \infty} P(|X_n - X| > \varepsilon_k) = 0$ so there is an integer $N$ such that for all $n \geq N$ we have $P(|X_n - X| > \varepsilon_k) < 1/2^k$. Choose one such $n$, and make sure that it is also large enough to satisfy $n > k$, too. Since this $n$ depends on $k$, call it $n(k)$ or $n_k$. This is the sub-sequence that we want. We now verify that $\lim_{k \to \infty} X_{n_k} = X$ with probability one. Since $\sum_k 1/2^k < \infty$, by the first Borel-Cantelli Lemma,

$$P(|X_{n_k} - X| > \varepsilon_k \text{ i.o.}) = 0$$

Therefore, for any $\varepsilon > 0$,

$$P(|X_{n_k} - X| > \varepsilon \text{ i.o.}) \leq P(|X_{n_k} - X| > \varepsilon_k \text{ i.o.}) = 0$$
3. Convergence of random variables

Details: Choose $N_0$ such that $\varepsilon_{n_k} < \varepsilon$ for $k > N_0$. Then

\[
\bigcap_{N=1}^{\infty} \bigcup_{k>N} \{|X_{n_k} - X| > \varepsilon\} \subset \bigcap_{N>N_0} \bigcup_{k>N} \{|X_{n_k} - X| > \varepsilon\} \subset \bigcap_{N>N_0} \bigcup_{k>N} \{|X_{n_k} - X| > \varepsilon_k\}
\]

Since $P(\bigcap_{N>N_0} \bigcup_{k>N} \{|X_{n_k} - X| > \varepsilon_k\}) = 0$, we have $P(\bigcap_{N=1}^{\infty} \bigcup_{k>N} \{|X_{n_k} - X| > \varepsilon\}) = 0$. □

Remark 4.16. Convergence in probability is a metric convergence, although it is too early to write down the metric, see Exercise 6.4s. Convergence with probability one is not a "metric convergence".

Remark 4.17. Suppose $X_n$ are random variables such that $X_n(\omega)$ converges for all $\omega \in \Omega$. Then the limit $X(\omega) := \lim_{n \to \infty} X_n(\omega)$ is a random variable. Somewhat more generally, if $\lim_{n \to \infty} X_n(\omega)$ converges with probability one, then $\Omega = \Omega_0 \cup \Omega_1$ decomposes into the two measurable sets where convergence holds and where convergence fails. Then with $X(\omega)$ defined arbitrarily (say as 0) for the $\omega \in \Omega_0$, we can still claim that $X_n \to X$ with probability one. the limit can be used to define the random variable $X$.

Proof.

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

Convergence in probability is a “metric convergence” (the metric cannot yet be written, but appears on page 141), so it has the following property.

Proposition 4.18. If every subsequence of $\{X_n\}$ has a $P$-convergent subsequence, then all the limits must be equal (with probability one) and $X_n$ converges in probability.

Proof. If different subsequences converge to say $X'$ and $X''$, then by choosing a subsequence that alternates between the two subsequences we can check that $|Pr(|X' - X''| > \varepsilon)| = 0$ for every $\varepsilon > 0$, so $X' = X''$ with probability one. Lets denote the common limit by $X$.

To prove convergence, to the above $X$, we proceed by contradiction. Suppose that $X_n$ does not converge in probability. Then there exists a subsequence $\{X_{n_k}\}_{k \in \mathbb{N}}$ such that $P(|X_{n_k} - X| > \varepsilon) > \delta > 0$. This subsequence cannot have a further subsequence that would converge to $X$.

To see that this is quite useful, try solving Exercise 4.25 without using Proposition 4.18. (Yes, it can be done!)

Every convergent sequence of numbers is bounded. An analog of this involves a separate concept which is introduced in Exercise 4.14.

3.1. Convergence in distribution. The third type of convergence, the so called convergence in distribution, or sometimes weak convergence, is somewhat different, as it is really convergence of the induced probability measures, not of the random variables. This topic will appear in Chapter 9, but we can give the definition here:

Definition 4.9. We say that a sequence of $\mathbb{R}$-valued random variables $X_1, X_2, \ldots, X_n, \ldots$ with cumulative distribution functions $F_1, \ldots, F_n, \ldots$ converges in distribution to a random variable $Y$ with cumulative distribution function $F$, if $F_n(x) \to F(x)$ for all continuity points $x$ of $F$. 
Notation $X_n \xrightarrow{D} X$ is often used, but one has to keep in mind that this is the convergence of induced measures $\mu_n$, and that random variables themselves could come from different probability spaces.

Example 4.8 (Normal approximation to Binomial). Suppose $X_n$ is $\text{Bin}(n,p)$ and $Z$ is normal $N(0,1)$, see Example 2.5. In Chapter 11, we will show a theorem (Theorem 11.2) that will imply that
$$\frac{X_n - np}{\sqrt{np(1-p)}} \xrightarrow{D} Z.$$

Example 4.9 (Poisson approximation to Binomial). Suppose $X_n$ is binomial $\text{Bin}(n,p) = \lambda/n$. In Example 9.4 we will show that $X_n \xrightarrow{D} Y$ where $Y$ is Poiss($\lambda$).

Here is a more elementary example that can be worked out directly from the definition.

Example 4.10 (extrema). Suppose $U_n$ are i.i.d. $\text{U}(0,1)$ and $X$ is exponential with parameter $\lambda = 1$, see Example 2.4. Then $n \min\{U_1, U_2, \ldots, U_n\} \xrightarrow{D} X$.

Proof. Let $x > 0$ and take $n$ large enough so that $x/n < 1$. Then
$$F_n(x) := P(n \min\{U_1, U_2, \ldots, U_n\} \leq x) = 1 - P(n \min\{U_1, U_2, \ldots, U_n\} \geq x/n) = 1 - P(U_1 \geq x/n) \cap U_2 \geq x/n \cap \cdots \cap U_n \geq x/n) = 1 - (1 - x/n)^n = 1 - e^{-x}.$$ This proves convergence for all $x > 0$. If $x \leq 0$ then $F_n(x) = 0$ converges to $F_X(x)$, too. \qed

Convergence in distribution is a metric convergence (of measures on $(\mathbb{R}, \mathcal{B})$) with Levy’s metric defined on page 141 and in Exercise 9.10.). It is a good exercise to check that if $X_n \xrightarrow{P} X$ then $X_n \xrightarrow{D} X$. (This is Theorem 9.2 in the notes.)

Example 4.11. Suppose $\Omega = (0,1)$ with Lebesgue measure. Define
$$X_n(\omega) = \begin{cases} \omega & \text{if } n \text{ is even} \\ 1 - \omega & \text{if } n \text{ is odd} \end{cases}$$
Then $X_n \xrightarrow{D} X_1$ because $F_n(x) = F_1(x)$. But $X_n$ cannot converge in probability, as it has two subsequence with two different limits $X_1$ and $X_2$.

An analog of Exercise 4.13 does not hold, as $X_n + Y_n$ is undefined, or ill-defined. (However there is a ”substitute” in Theorem 9.4 later on.)

The relations between the three modes of convergence are complicated:

(i) Convergence with probability one implies convergence in probability. (Proposition 4.14)
(ii) Convergence in probability implies convergence in distribution. (Theorem 9.2, which could be proven here.)
(iii) Sequences convergent in probability have subsequences that converge with probability one. This is Proposition 4.15.
(iv) Sequences convergent in distribution can be redefined onto a common probability space on which they converge pointwise (hence almost surely). This is Theorem 9.6, which could have been proven here.

Later on, we will also need a fact from Exercise 4.25, so this exercise will be solved in class.
Measurability.

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

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

Exercise 4.3. Prove one/some/all of the statements in Proposition 4.5.

Cumulative distribution functions.

Exercise 4.4. 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 4.5. Suppose

\[
F(x) = \begin{cases} 
0 & x < 1 \\
1/2 & 1 \leq x < 2 \\
3/4 & 2 \leq x < 3 \\
1 & x \geq 3 
\end{cases}
\]

Compute \( \varphi(u) \) from the proof of Theorem 4.10.

Exercise 4.6. Suppose \( X : \Omega \to \mathbb{R} \) has CDF \( F \). Let \( Y = X^2 \). What is the CDF of \( Y \)?

Exercise 4.7. Suppose \( X : \Omega \to \mathbb{R} \) 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 4.8. Suppose \( U \) is uniform on \( (0, 1) \). Let \( X = U^2, Y = U^3 \). What is their joint CDF? (See (4.2).)

Exercise 4.9 (an example of copula). Find the joint cumulative distribution function for a random vector \( (X, Y) \) defined in Example 4.1.

Exercise 4.10 (Statistics). Use the second proof of Theorem 4.6 to describe how to simulate exponential random variables (see Example 2.4) using a random number generator that produces an infinite sequence of uniform \( U(0, 1) \) random variables.

Independence.

Exercise 4.11. 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 3.2.)
Convergence.

**Exercise 4.12.** Suppose random variables

\[ X_n = \begin{cases} 
  n & \text{with probability } p_n \\
  0 & \text{with probability } 1 - p_n 
\end{cases} \]

Prove that

(i) if \( p_n \to 0 \) then \( X_n \xrightarrow{P} 0 \).

(ii) if \( \sum_n p_n < \infty \) then \( X_n \to 0 \) with probability one.

(iii) if \( X_n \) are independent then \( X_n \to 0 \) with probability one iff and only if \( \sum p_n < \infty \)

**Exercise 4.13.** Prove that if \( X_n \xrightarrow{P} X \) and \( Y_n \xrightarrow{P} Y \) then \( X_n + Y_n \xrightarrow{P} X + Y \).

**Exercise 4.14.** Suppose \( X_n \xrightarrow{P} X \). Show that \( \{X_n\} \) is stochastically bounded (which is the same as the sequence of laws being tight, compare Exercise 1.11), i.e. for every \( \varepsilon > 0 \) there exists \( K > 0 \) such that for all \( n \) we have \( P(|X_n| > K) < \varepsilon \).

**Exercise 4.15.** Use the result from Exercise 4.14 to prove that if \( X_n \xrightarrow{P} X \) and \( Y_n \xrightarrow{P} Y \) then \( X_nY_n \xrightarrow{P} XY \).

**Exercise 4.16.** 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 on page 23). Show that the sequence \( Z_n = U_1U_2 \cdots U_n \) converges with probability one.

**Exercise 4.17** (Hw 4). If \( X_n \leq Y_n \leq Z_n \) for all \( n \) and \( X_n \to X \), \( Z_n \to X \) in probability then \( Y_n \to X \) in probability.

**Exercise 4.18** (Hw 4). If \( X_n \to X \) in probability and \( X_n \geq 1 \) for all \( n \), then \( X \geq 1 \) (that is, \( P(X \geq 1) = 1 \) and \( \sqrt{X_n} \to \sqrt{X} \) in probability.

**Exercise 4.19** (Hw 6). Give a formula for a subsequence \( n_k \) such that \( X_{n_k} \to X \) with probability one, if it is known that

(i) for all \( t > 0 \) we have \( P(|X_n - X| > t) \leq \frac{17}{t^2\sqrt{n}} \)

(ii) for all \( u > 0 \) we have \( P(|X_n - X| > u) \leq \frac{17}{u \log(n+1)} \)

---

**Additional Exercises**

Some of these exercises appeared on past prelims!

**Exercise 4.20.** 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 4.21.** 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 4.22.** Suppose $X, Y$ are random variables with cumulative distribution functions $F(x)$ and $G(y)$, constructed as in the second proof of Theorem 4.6. Find the joint cumulative distribution function of $X, Y$.

**Exercise 4.23** (Statistics). Suppose $X, Y$ are independent $N(0, 1)$ random variables. Verify that $X^2 + Y^2$ is exponential. Hint: compute CDF using polar coordinates.

**Exercise 4.24.** Suppose that $X_1 \leq X_2 \leq \cdots \leq X_n \leq X_{n+1} \leq \cdots$. If $X_n \xrightarrow{p} X$, show that $X_n \rightarrow X$ with probability one.

**Exercise 4.25.** Suppose $f : \mathbb{R} \rightarrow \mathbb{R}$ is a continuous function and $X_n \xrightarrow{p} X$. Prove that $Y_n = f(X_n)$ converges in probability to $Y = f(X)$. *Hint:* an elementary proof relies on Proposition 4.18.

The following generalization of Exercise 4.25 can be proved by slight elaboration on the same techniques.

**Exercise 4.26.** Suppose $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ is a continuous function and $X_n \xrightarrow{p} X$, $Y_n \xrightarrow{p} Y$. Prove that $Z_n = f(X_n, Y_n)$ converges in probability to $Z = f(X, Y)$.

**Exercise 4.27.** Suppose $X_n \xrightarrow{p} X$ and $X_n$ are independent. Show that there is $a \in \mathbb{R}$ such that the cumulative distribution of $X$ is $F(x) = \begin{cases} 0 & x < a \\ 1 & x \geq a \end{cases}$.
Review of 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 ... (you should be able to write a formal definition!)

\[ \sum_{n=1}^{\infty} a_n = L \] means that ... (you should be able to write a formal definition!)

**Theorem A.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 \( a_n \),

\[ \limsup_{n \to \infty} a_n = \lim \sup_{n \to \infty} a_k \text{ and } \liminf_{n \to \infty} a_n = \lim \inf_{n \to \infty} a_k. \]

**Example A.1.** Here are a couple of examples of that illustrate several different possibilities, and also indicates standard conventions about the use of \( \infty \).

(i) Consider \( a_k = \frac{k}{k+1} \). Then \( \inf_{k \geq n} a_k = \min_{k \geq n} a_k = \frac{n}{n+1} \). So \( \liminf_{n \to \infty} a_n = 1 \).

Similarly, \( \sup_{k \geq n} a_k = 1 \). So \( \limsup_{n \to \infty} a_n = 1 \).

(ii) Suppose \( a_k = (-1)^k \). Then \( \inf_{k \geq n} a_k = \min_{k \geq n} a_k = -1 \) and \( \sup_{k \geq n} a_k = \max_{k \geq n} a_k = 1 \). So \( \liminf_{n \to \infty} a_n = -1 \) and \( \limsup_{n \to \infty} a_n = 1 \).

(iii) Suppose \( a_k = (1 + (-1)^k) k \). Then \( \inf_{k \geq n} a_k = 0 \) and \( \sup_{k \geq n} a_k = \infty \). So \( \liminf_{n \to \infty} a_n = 0 \) and \( \limsup_{n \to \infty} a_n = \infty \).

**Remark A.2.** 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 $\sup_{\omega \in \Omega} |f_n(\omega) - f(\omega)| \to 0$.

Example A.2. Consider $f_n(x) = x^n$. If $\Omega = [0, 1/2]$ then $f_n$ converges uniformly over $\Omega$. If $\Omega = [0, 1]$ then $f_n$ converges pointwise.

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 $\varnothing$ 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 $B \supset A$. 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,

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

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 A.3** (DeMorgan’s law).

\[
(A.1) \quad \left( \bigcup_{t \in T} A_t \right)^c = \bigcap_{t \in T} A_t^c
\]

Since \((A^c)^c = A\), formula (A.1) is equivalent to

\[
(A.2) \quad \left( \bigcap_{t \in T} A_t \right)^c = \bigcup_{t \in T} A_t^c
\]

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

\[
(A.3) \quad I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{otherwise} \end{cases}
\]

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 A.4.**

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

**Proof.** This is Exercise A.1. \(\square\)

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 A.3 and A.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.
A. Review of math prerequisites

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 $\bigcap_{k=1}^{n} F_k \neq \emptyset$ for all $n$, then the infinite intersection $\bigcap_{k=1}^{\infty} F_k$ is also non-empty.

**Theorem A.5.** Closed bounded subsets of $\mathbb{R}^k$ are compact.

**Proof.** To see what this involves, let's indicate why closed interval $[0,1]$ is 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

$$\left| S - \sum_{i} f(x_j) |I_j| \right| < \varepsilon$$

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 Cartesian product $U \times V$ of sets $U,V$ is the set of ordered pairs $\{(u,v) : u \in U, v \in V\}$. In particular, $\mathbb{R}^2 : \mathbb{R} \times \mathbb{R}$.

The cartesian product $\times_{k=1}^{\infty} U_k$ is the set of all sequences $\{u_k\}$ of elements of the sets $U_1, U_2, \ldots$. In particular, $\mathbb{R}^\infty := \times_{k=1}^{\infty} \mathbb{R}$

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

$$d((a_n), (b_n)) = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{|a_n - b_n|}{1 + |a_n - b_n|}. \tag{A.4}$$

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 129).

7. Complex numbers

See Chapter 12 Section 1 (page 129).
8. Metric spaces

**Definition A.1.** A function \(d : E \times E \to \mathbb{R}\) is called a metric if

(i) \(d(x, y) \geq 0\) and if \(d(x, y) = 0\) then \(x = y\)

(ii) \(d(x, y) = d(y, x)\)

(iii) \(d(x, z) \leq d(x, y) + d(y, z)\)

We then call the pair \((E, d)\) a metric space.

Here are some “elementary” examples of metric spaces.

(i) \(\mathbb{R}\) with \(d(x, y) = |x - y|\)

(ii) \(\mathbb{R}^d\) with \(d(x, y) = \|x - y\|\).

(iii) \(C[0, 1]\) with \(d(f, g) = \sup\{|f(x) - g(x) : 0 \leq x \leq 1\}\).

(iv) The set of all CDFs on \(\mathbb{R}\) with Kolmogorov-Smirnov metric

\[d(F, G) = \sup\{|F(x) - G(x)| : 0 < x < \infty\}\]

(v) The set of all probability measures on \((\mathbb{R}, B)\) with total variation metric

\[\delta(P, Q) = \sup\{|P(A) - Q(A)| : A \in B\}\]

It is known that \(\delta(P, Q) = \inf\{P(X \neq Y) : \mathcal{L}(X) = P, \mathcal{L}(Y) = Q\}\)

(vi) The set of all CDFs on \(\mathbb{R}\) with Levy’s metric:

\[L(F, G) = \inf\{\varepsilon : G(x) \in [F(x - \varepsilon) - \varepsilon, F(x + \varepsilon) + \varepsilon] \text{ for all } x\}\]

This is a metric for weak convergence: \(F_n \to F\) iff \(L(F_n, F) \to 0\)

Here are some of the metric spaces encountered in probability:

(i) The set of all CDFs (see Definition 2.3) on \(\mathbb{R}\) with Kolmogorov-Smirnov metric

\[d(F, G) = \sup\{|F(x) - G(x)| : 0 < x < \infty\}\]

This distance is used in statistics.

(ii) The set of all probability measures on \((\mathbb{R}, B)\) with total variation metric

\[\delta(P, Q) = \sup\{|P(A) - Q(A)| : A \in B\}\]

It is known that \(\delta(P, Q) = \inf\{P(X \neq Y) : \mathcal{L}(X) = P, \mathcal{L}(Y) = Q\}\)

(iii) The set of all CDFs on \(\mathbb{R}\) with Levy’s metric:

\[L(F, G) = \inf\{\varepsilon : G(x) \in [F(x - \varepsilon) - \varepsilon, F(x + \varepsilon) + \varepsilon] \text{ for all } x\}\]

This is a metric for weak convergence: \(F_n \to F\) iff \(L(F_n, F) \to 0\), see Exercise Exercise 9.10.

(iv) The set of (classes of equivalence of) all random variables on \((\Omega, \mathcal{F}, P)\) with the distance

\[d(X, Y) = E\left(\frac{|X - Y|}{1 + |X - Y|}\right)\]

This is a metric for convergence in probability: \(X_n \xrightarrow{P} X\) iff \(d(X_n, X) \to 0\).
There are numerous other distances of interest. The following are frequently encountered and useful.

(i) The set of all (classes of equivalence of) integrable random variables $L_1(\Omega, \mathcal{F}, P)$ with the $L_1$ metric $\|X - Y\|_1 = E(|X - Y|)$

(ii) The set of all (classes of equivalence of) square integrable random variables $L_2(\Omega, \mathcal{F}, P)$ with the $L_2$ metric $\|X - Y\|_2 = \sqrt{E((X - Y)^2)}$.

(iii) The set of probability measures (CDFs) on $\mathbb{R}$ with the Waserstein distance

$$d(P, Q) = \inf \{E|X - Y| : \mathcal{L}(X) = P, \mathcal{L}(Y) = Q\}$$

It is known that $d(P, Q) = \sup \int f(x)dP - \int f(x)dQ : f$ Lipschitz with constant 1

Additional Exercises


Exercise A.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 $A^* = \bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} A_k$.

Exercise A.3. Suppose $A_1 \supset A_2 \supset A_n \supset \ldots$. Show that $\lim_n A_n$ exists (and describe the limit).

Exercise A.4. Suppose $A_1 \subset A_2 \subset A_n \subset \ldots$. Show that $\lim_n A_n$ exists (and describe the limit).

Exercise A.5. If $a_n \to L$ for a finite $L$, show that $\{a_n\}$ is bounded.
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

$L_1$ 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
deoivre’s 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-additive 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
Kolmogorov’s maximal inequality, 91
Kolmogorov’s one series theorem, 91
Kolmogorov’s three series theorem, 92
Kolmogorov’s two series theorem, 92
Kolmogorov’s zero-one law, 90
Kolmogorov-Smirnov metric, 10, 11
Kronecker’s Lemma, 94
Lévy’s distance, 106
law of $X$, 42
Lebesgue’s dominated convergence theorem, 71, 72
Lebesgue’s dominated convergence theorem – used, 73, 89, 102, 115
Levy’s metric, 11
Levy’s theorem, 93
Lindeberg condition, 120
Lyapunov’s condition, 121
Lyapunov’s inequality, 58
marginal cumulative distribution functions, 30
Markov’s inequality, 57
maximal inequality, Etemadi’s, 95
maximal inequality, Kolmogorov’s, 91
mean square convergence, 76
measurable function, 41
measurable rectangle, 81
metric, 10
metric space, 10
Minkowski’s inequality, 59
Minkowski’s inequality, 76
moment generating function, 56, 77
moments, 55
Monotone Convergence Theorem, 69
multivariate normal, 129
multivariate normal distribution, 130
multivariate random variable, 41
negative binomial distribution, 18
normal distribution, 29
Poisson distribution, 18, 73
Polya’s distribution, 18
Portmanteau Theorem, 102
power set, 7
probability, 13
probability measure, 14
probability space, 13, 17
product measure, 82
quantile function, 44, 101
random element, 41
random variable, 41
random vector, 41
sample space, 13
Scheffe’s theorem, 100
section, 81
semi-algebra, 15
semi-ring, 15
sigma-field generated by $\mathcal{A}$, 16
simple random variable, 53
simple random variables, 45
Skorohod’s theorem, 101
Slutsky’s Theorem, 101
Standard normal density, 74
stochastic process with continuous trajectories, 43
stochastic processes, 42
stochastically bounded, 51
symmetric distribution, 97
tail $\sigma$-field, 36
Tail integration formula, 84
Taylor polynomials, 107
tight, 51
tight probability measure, 19
Tonelli’s theorem, 82
total variation metric, 11
truncation of r.v., 50
uncorrelated, 63
uniform continuous, 28
Uniform density, 74
uniform discrete, 28
uniform singular, 28
uniformly integrable, 72, 105
union, 8
variance, 56
Waserstein distance, 11
weak convergence, 49
weak law of large numbers, 63
zero-one law, 36, 90