Weak convergence

This is based on [Billingsley, Section 25]

1. Convergence in distribution

A cumulative distribution function $F$ can have at most a countable number of discontinuity points. In fact, the set $\{x : F(x) - F(x^-) \geq 1/n\}$ can have at most $n$ points. (This observation will be used in many proofs.)

**Definition 9.1.** Let $F_n, F$ be cumulative distribution functions. We say that $F_n \xrightarrow{D} F$ if $F_n(x) \rightarrow F(x)$ for every point $x$ of continuity of $F$.

We say that $X_n \xrightarrow{D} X$ if $F_n \xrightarrow{D} F$.

We first show how to use the definition to prove weak convergence.

**Example 9.1.** Suppose $\{X_k\}$ are independent exponential. Then $\max_{1 \leq k \leq n} X_k - \ln n \xrightarrow{D} Y$ where $Y$ has the Gumbel distribution: $P(Y \leq x) = \exp(-e^{-x})$.

Indeed, $P(\max_{1 \leq k \leq n} X_k - \ln n \leq x) = P(\max_{1 \leq k \leq n} X_k \leq x + \ln n) = P(X_1 \leq x + \ln n)^n = (1 - e^{-x \ln n})^n = (1 - \frac{e^{-x}}{n})^n \rightarrow e^{-e^{-x}}$

The following example illustrates that we cannot require convergence for all $x \in \mathbb{R}$.

**Example 9.2.** Suppose $X_n$ are uniform $U(0, 1/n)$ with

$$F_n(x) = \begin{cases} 1 & x > 1/n \\ x/n & x \in [0, 1/n] \\ 0 & x < 0 \end{cases}$$

It is clear that $X_n \rightarrow 0$ with probability one, so we expect (and can prove, see Theorem 9.2 below) that $X_n \xrightarrow{D} 0$, i.e. that $F_n(x) \rightarrow F(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$. And indeed, $F_n(x) = 0$ for $x < 0$ and $F_n(x) \rightarrow 1$ for $x > 0$. But note that $F_n(0) = 0$ does not converge to $F(0)$. 
The following example illustrates that a popular interpretation of weak convergence as “approximating all probabilities” for $X_n$ by the asymptotic probabilities for $X$ has significant restrictions.

**Example 9.3.** Suppose $P(X_n = k) = 1/n$. Then $\frac{1}{n}X_n \xrightarrow{D} U(0, 1)$. Indeed, $F_n(x) = [nx + 1]/n \to x$. Note however that $P(\frac{1}{n}X_n \in A)$ does not converge to $\lambda(A)$ for all Borel sets $A$.

In view of Example 9.3, it is interesting to have a criterion where a stronger form of weak convergence holds.

**Theorem 9.1** (Schéffe’s theorem). Suppose $X_n$ has a density $f_n(x)$ with respect to a (possibly infinite, possibly discrete) measure $\nu(dx)$ on $\mathbb{R}$. If $f_n(x) \to f(x)$ pointwise and $f$ is a density of a random variable $X$, then

$$\sup_A |P(X_n \in A) - P(X \in A)| \to 0$$

**Proof.** Consider $g_n = f - f_n$. Then $g_n^+ \to 0$ and $0 \leq g_n^+ \leq f$ so by the dominated convergence theorem $\int g_n^+ \nu(dx) \to 0$. Now $\int |g_n| \nu(dx) = \int_{g_n > 0} g_n \nu(dx) - \int_{g_n \leq 0} g_n \nu(dx)$. Since $\int g_n \nu(dx) = 0$ we have $\int_{g_n \leq 0} g_n \nu(dx) = -\int_{g_n > 0} g_n \nu(dx)$ so

$$\int |g_n| \nu(dx) = 2 \int g_n^+ \nu(dx) \to 0$$

Thus $|P(X_n \in A) - P(X \in A)| \leq \int_A |g_n(x)| \nu(dx) \to 0$ for any $A$. □

**Example 9.4.** Suppose $X_n$ is binomial $\text{Bin}(n, p = \lambda/n)$. Then $X_n \xrightarrow{D} Y$ where $Y$ is Poiss($\lambda$). Indeed, the density with respect to the counting measure $\nu$ converges pointwise

$$P(X_n = k) = \binom{n}{k} p^k (1-p)^{n-k} \to \frac{n(n-1) \ldots (n-k+1) \lambda^k}{k!} (1-\lambda/n)^n \to e^{-\lambda} \lambda^k / k!$$

Thus in this case $P(X_n \in A) \to P(Y \in A)$ for all $A$. For example,

$$P(X_n \text{ is even}) \to e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{2k!}$$

Similarly, as the number of degrees of freedom $d \to \infty$, the density of student $T_d$ distribution converges to the standard normal density.

**Theorem 9.2.** If $X_n \xrightarrow{P} X$ then $X_n \xrightarrow{D} X$.

**Proof.** Let $x$ be a point of continuity of $F(x)$. Then

$$P(X_n \leq x) = P(X_n \leq x, |X_n - X| \leq \varepsilon) + P(X_n \leq x, |X_n - X| > \varepsilon)$$

So

$$P(X_n \leq x) \leq P(X \leq x + \varepsilon) + P(|X_n - X| > \varepsilon)$$

Similarly,

$$P(X \leq x - \varepsilon) \leq P(X_n \leq x) + P(|X_n - X| > \varepsilon)$$

So

$$F(x - \varepsilon) \leq \lim \inf P(X_n \leq x) \leq \lim \sup P(X_n \leq x) \leq F(x + \varepsilon)$$

Taking the limit $\varepsilon \to 0$,

$$F(x^-) \leq \lim \inf P(X_n \leq x) \leq \lim \sup P(X_n \leq x) \leq F(x)$$
Remark 9.3. If $X_n \xrightarrow{P} a$ for a deterministic random variable $a$ then $X_n \xrightarrow{\mathbb{P}} a$. Indeed, $P(X_n \leq a - \varepsilon) \rightarrow 0$ and $P(X_n \leq a + \varepsilon) \rightarrow 1$ so

$$P(|X_n - a| > \varepsilon) \leq P(X_n \leq a - \varepsilon) + 1 - P(X_n \leq a + \varepsilon) \rightarrow 0$$

Theorem 9.4 (Slutsky’s Theorem). Suppose $(X_n, Y_n)$ are defined on the same probability space. If $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{P} 0$ then $X_n + Y_n \xrightarrow{D} X$

Proof. Take $y' < y''$ two continuity points of the law of $X$ and $y' < x - \varepsilon < x < x + \varepsilon < y''$. Then

$$P(X_n \leq y') - P(|Y_n| > \varepsilon) \leq P(X_n + Y_n \leq x) \leq P(X_n \leq y'') + P(|Y_n| > \varepsilon)$$

So

$$F(y') \leq \lim \inf P(X_n + Y_n \leq x) \leq \lim \sup P(X_n + Y_n \leq x) \leq F(y'')$$

We now note that the set $\{ x : F(x-)-F(x) \geq 1/n \}$ has at most $n$ points, so the set of all discontinuities of $F$ is at most countable. Therefore, if $x$ is a continuity point of $F$ we can find continuity points $y' < x < y''$ that are arbitrarily close to $x$. Thus taking a sequence $y' \rightarrow x$ and $y'' \rightarrow x$ of such points we get

$$F(x) \leq \lim \inf P(X_n + Y_n \leq x) \leq \lim \sup P(X_n + Y_n \leq x) \leq F(x)$$

The following corollary is often useful. (Its proof requires tightness!).

Corollary 9.5. Suppose $(X_n, Y_n)$ are defined on the same probability space. If $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{P} c$ then $X_n Y_n \xrightarrow{D} cX$.

Proof. See Exercise 9.2

2. Fundamental results

Theorem 9.6 (Skorohod’s theorem). Suppose $X_n \xrightarrow{D} X$ i.e. $F_n \xrightarrow{D} F$. Then there exist a probability space $(\Omega, F, P)$ and random variables $Y_n, Y$ with CDF $F_n, F$ on $(\Omega, F, P)$ such that $Y_n \rightarrow Y$ for all $\omega \in \Omega$.

Proof. We choose $\Omega = (0, 1)$ with Lebesgue measure. Recalling the quantile function (4.3) we define

$$Y_n(\omega) = \inf \{ x : F_n(x) \geq \omega \}$$

$$Y(\omega) = \inf \{ x : F(x) \geq \omega \}$$

Recall that $Y(\omega) \leq x$ iff $F(x) \geq \omega$ so $Y(\omega) > x$ implies $F(x) < \omega$.

Given $\varepsilon > 0$ choose $Y(\omega) - \varepsilon < x < Y(\omega)$ such that $F(x^-) = F(x)$.

Since $F_n(x) \rightarrow F(x) < \omega$, this implies that $F_n(x) < \omega$ for large $n$. Thus $Y_n(\omega) > x > Y(\omega) - \varepsilon$. Since $\varepsilon > 0$ this shows that $\lim \inf Y_n \geq Y$.

Now choose $\omega < \omega'$ and a continuity point $y$ of $F$ such that $Y(\omega') < y < Y(\omega') + \varepsilon$. The first inequality then implies that $\omega < \omega' \leq F(y)$, so for large $n$ we have $F_n(y) > \omega$. Thus $Y_n(\omega) \leq y < Y(\omega') + \varepsilon$. This shows that $\lim \sup Y_n(\omega) \leq Y(\omega)$ for all points $\omega$ of continuity of $Y$. Note that $Y$ is an increasing function so it can have at most countable number of discontinuities.
At such points we re-define \( Y_n(\omega) \) to be \( Y(\omega) \). This changes \( Y_n \) on the set of measure zero, so does not affect the result. \( \square \)

**Theorem 9.7** (Portmanteau Theorem). The following conditions are equivalent:

(i) \( X_n \xrightarrow{D} X \)

(ii) \( E(f(X_n)) \to E(f(X)) \) for every bounded continuous function \( f \)

(iii) \( E(f(X_n)) \to E(f(X)) \) for every bounded Lipschitz (uniformly continuous) function \( f \)

(iv) \( P(X_n \in U) \to P(X \in U) \) for every Borel set \( U \) such that \( P(X \in \delta U) = 0 \)

**Proof.** (Omitted in 2018)

(1)\(\Rightarrow\)(2) Using Theorem 9.6 we have \( f(Y_n) \to f(Y) \) so by Lebesgue’s dominated convergence theorem (Theorem 6.10), the integrals converge. Note that this proof works for \( \mathbb{R} \) but not for \( \mathbb{R}^2 \), so it is of interest to have a direct proof that will not use Theorem 9.6. See [Billingsley, Theorem 29.1]

(2)\(\Rightarrow\)(3) is obvious

(3)\(\Rightarrow\)(1) Fix a point of continuity \( x_0 \) of \( F \) and let

\[
f(x) = \begin{cases} 
1 & x \leq x_0 \\
\text{linear} & x_0 < x < x_0 + \varepsilon \\
0 & x > x_0 + \varepsilon
\end{cases}
\]

Then \( F_n(x_0) \leq E(f(X_n)) \to E(f(X) \leq F(x_0 + \varepsilon) \) so \( \limsup_n F_n(x_0) \leq F(x_0) \).

Next, take

\[
f(x) = \begin{cases} 
1 & x \leq x_0 - \varepsilon \\
\text{linear} & x_0 - \varepsilon < x < x_0 \\
0 & x \geq x_0
\end{cases}
\]

Then \( F_n(x_0) \geq E(f(X_n)) \to E(f(X) \geq F(x_0 - \varepsilon) \) Thus \( \liminf_n F_n(x_0) \geq F(x_0) \).

(Omitted in 2018)

(3)\(\Rightarrow\)(4) The plan here is to prove that for any Borel set \( U \) with interior \( U^\circ \) and closure \( \bar{U} \) we have

\[\mu(U^\circ) \leq \liminf \mu_n(U^\circ) \leq \liminf \mu_n(U) \leq \limsup \mu_n(U) \leq \limsup \mu_n(\bar{U}) \leq \mu(\bar{U})\]

To prove the last inequality, take continuous \( f \) with values between 0, 1 and such that \( f = 1 \) on \( \bar{U} \) and \( f = 0 \) outside of an \( \varepsilon \)-closure of \( U \). Then \( \mu_n(U) \leq \int f(x)\mu_n(dx) \to \int f(x)\mu(dx) \leq \mu(U^{[\varepsilon]})\)

Noting that \( \bar{U} = \bigcap_{\varepsilon > 0} U^{[\varepsilon]} \) we get the last inequality. The first inequality then follows by taking complements.

(4)\(\Rightarrow\)(1) is obvious
(1)⇒(2) [Second proof] Suppose $f$ is continuously differentiable and $f' = 0$ outside of a finite interval $[-K, K]$. Then from Fubini’s theorem\(^1\) we get
\[
Ef(X_n) = f(0) + \int_0^K f'(t)P(X_n > t)dt - \int_{-K}^0 f'(t)P(X_n \leq t)dt
\]
Since $P(X_n \leq t) \to P(X \leq t)$ except for a countable (Lebesgue-measure zero) set of $t$, by Lebesgue dominated convergence theorem we get $Ef(X_n) \to Ef(X)$.

(Omitted in 2018)
To extend this result to all bounded continuous functions, let’s say bounded by $M$, we use tightness (see Definition 9.2) to choose $K$ such that $\int_{|X_n| \geq K} |f(X)|dP < \varepsilon$ for all $n$ and all functions $f$ bounded by $2M$. That is, choose $K$ such that $P(|X_n| > K) \leq \frac{\varepsilon}{2M}$ and $P(|X| > K) \leq \frac{\varepsilon}{2M}$.

Given continuous $f$ bounded by $M$, use Weierstrass theorem to find a smooth function $g : [-K, K] \to \mathbb{R}$ such that $|g(x) - f(x)| < \varepsilon$ for $x \in [-K, K]$ and $g$ is bounded by $2M$. Finally, extend $g$ to $\mathbb{R}$ in such a way that $g(x) = 0$ outside of $[-K - 1, K + 1]$ and without increasing the bound $|g(x)| \leq 2M$. Then
\[
|Ef(X_n) - Ef(X)| \leq 2\varepsilon + \int_{|X_n| \leq K, |X| \leq K} (f(X_n) - f(X))dP
\]
\[
\leq 4\varepsilon + \int_{|X_n| \leq K, |X| \leq K} (g(X_n) - g(X))dP
\]
\[
\leq 6\varepsilon + |Eg(X_n) - Eg(X)|
\]
Since $|Eg(X_n) - Eg(X)| \to 0$ by the previous part of the proof, and $\varepsilon > 0$ is arbitrary, we see that $Ef(X_n) - Ef(X) \to 0$

\[\square\]

As an immediate corollary, we get an important result.

**Theorem 9.8** (Continuous Mapping Theorem). If $X_n \xrightarrow{D} X$ and $f$ is a continuous (but perhaps unbounded) function then $f(X_n) \xrightarrow{D} f(X)$.

**Example 9.5.** In the setting of Example 9.3, we have $E(f(X_n/n)) = \frac{1}{n} \sum_{k=1}^{n} f(k/n) \to \int_0^1 f(x)dx$

**Definition 9.2.** A sequence of probability measures $\mu_n$ on $\mathbb{R}$ is tight if for every $\varepsilon > 0$ there exists $K$ such that $\mu_n([K - \varepsilon, K + \varepsilon]) > 1 - \varepsilon$.

It is clear that if $X_n \xrightarrow{D} X$ then $X_n$ is tight.

**Theorem 9.9** (Helly, Prokhorov). If $\mu_n$ is a tight family of probability measures then there is a probability measure $\mu$ and a subsequence $n_k \to \infty$ such that $\mu_{n_k} \xrightarrow{D} \mu$.

\[1f(x) = f(0) + \int_{x>0} f'(t)dt - \int_{x<0} f'(t)dt\]
Proof. Since $F_n(r)$ is a bounded sequence of numbers, there is a subsequence that converges. In fact, by using a diagonal method, there is a subsequence $n_k$ such that $F_{n_k}(r) \to G(r)$ for all $r \in \mathbb{Q}$.

To see this, enumerate all rational numbers $q_1, q_2, \ldots$. Since $[0, 1]$ is compact, we can choose a sequence $n(k) = n_1(k)$ such that $F_{n(k)}(q_1)$ converges to, say, $G(q_1)$. Choose a subsequence $n_2(k)$ of $n_1(k)$ such that $F_{n_2(k)}(q_1)$ converges to, say, $G(q_1)$ and so on.

Then the diagonal subsequence $m_k := n_k(k)$ has the property that $F_{m_k}(q) \to G(q)$ for every $q \in \mathbb{Q}$.

Define

$$F(x) = \inf \{G(r) : r > x\}$$

Note that $F(x) = \lim_{r_k \uparrow x} G(r)$, so $F$ is non-decreasing and right-continuous. By tightness, $F(x) < \varepsilon$ if $x < K$ and $F(x) > 1 - \varepsilon$ if $x > K$. Next we check that $F$ is right-continuous:

Now we verify the weak convergence. Let $x$ be a point of continuity of $F$. Choose $r_k \uparrow x$ and $r_k' \downarrow x$.

Then

$$F_n(r_k) \leq F_n(x) \leq F_n(r_k')$$

so for every $k$ we have

$$G(r_k) \leq \lim \inf F_n(x) \leq \lim \sup F_n(x) \leq G(r_k')$$

But $G(r_k') \to F(x)$ as $k \to \infty$. And for any $\varepsilon_k > 0$ converging to 0 we have $G(r_k) \geq F(r_k - \varepsilon_k) \to F(x)$ by continuity.

\[\square\]

Example 9.6. Suppose $X_n$ are uniform on $(0, n)$. Then

$$F_n(x) = \begin{cases} 0 & x < 0 \\ x/n & 0 \leq x \leq n \\ 1 & x > n \end{cases}$$

So $F_n(x) \to F_\infty(x)$ for all $x$. Clearly $F_\infty(x)$ is not a cumulative distribution function, and $X_n$ is not a tight sequence.

We will need the following corollary.

Theorem 9.10. If $\mu_n$ is a tight family of probability measures and if each subsequence converges to the same probability measure $\mu$ then $\mu_n \xrightarrow{D} \mu$.

Proof. Suppose $\mu_n$ fails to converge to $\mu$ with CDF $F$. Then there is a point of continuity $x$ of $F$ and an infinite sequence $n_k$ such that $|F_{n_k}(x) - F(x)| > \delta$ for all $k$. Since subsequence
\( \mu_{n_k} \) is tight, choose a convergent subsequence. By assumption, this sequence converges to \( \mu_u \), so \( F_{n_k}(x) \to F(x) \), which contradicts that \( |F_{n_k}(x) - F(x)| > \delta \) for all \( k \). □

Recall Definition 6.3.

**Proposition 9.11.** If \( \{X_n\} \) is uniformly integrable, then \( \sup_n E|X_n| < \infty \)

**Proof.** (This should have been an assigned exercise!) \( E|X_n| = \int_{|X_n| \leq K} |X_n| dP + \int_{|X_n| > K} |X_n| dP < K\varepsilon + \varepsilon \).

**Theorem 9.12.** Suppose \( X_n \xrightarrow{D} X \) and \( \{X_n\} \) is uniformly integrable. Then \( X \) is integrable and \( E(X_n) \to E(X) \).

**First proof.** From Theorem 9.6 there exists a sequence \( Y_n \to Y \) such that \( E(Y_n) = E(X_n) \). By Lebesgue’s dominated convergence theorem (See Remark 6.13), \( E(Y_n) \to E(Y) \). □

**Second proof.** The first step is to prove that \( X \) is integrable, which we will omit\(^2\).

Given \( \varepsilon > 0 \) choose \( K \) such that \( \int_{|X_n| > K} |X_n| dP < \varepsilon \). Since \( X \) is integrable, we can increase \( K \) to ensure that we also have \( \int_{|X| > K} |X| dP < \varepsilon \).

Let \( f : \mathbb{R} \to \mathbb{R} \) be a piecewise-linear bounded continuous function such that \( f(x) = x \) for \( x \in [-K, K] \) and \( f(x) = 0 \) for \( x \notin [-K - 1, K + 1] \). (Note that \( f = f_K \) depends on \( K \).) By Theorem 9.7, \( E(f(X_n)) = E(f(X)) \). On the other hand, \( X_n = f(X_n) \) for \( |X_n| \leq K \) and \( |f(x)| \leq |x| \) for all \( x \), so

\[
|E(X_n) - E(X)| \leq |E(X_n) - Ef(X_n)| + |E(X) - E(f(X))| + |Ef(X_n) - Ef(X)|
\]

\[
\leq 2 \int_{|X_n| \geq K} |X_n| dP + 2 \int_{|X| \geq K} |X| dP + |Ef(X_n) - Ef(X)|
\]

\[
\leq 4\varepsilon + |Ef(X_n) - Ef(X)|
\]

Since \( |Ef(X_n) - Ef(X)| \to 0 \) this implies convergence. □

The proof of the next result contains solution of (a generalization of) Exercise 6.5.

**Corollary 9.13.** Suppose \( \sup_n E|X_n|^{r+\delta} < \infty \) for a natural \( r \) and \( \delta > 0 \). If \( X_n \xrightarrow{D} X \) then \( E(|X|^r) < \infty \) and \( E(X_n^r) \to E(X^r) \).

**Proof.** We verify that \( |X_n|^r \) is uniformly integrable, compare Exercise 6.5.

\[
\int_{|X_n| > t} |X_n|^r dP = \int_{|X_n| > t} |X_n|^r 1 dP \leq \int_{|X_n| > t} |X_n|^r \delta \|X_n\|^\delta dP \leq \frac{1}{\delta^\delta} \sup_n E|X_n|^r + \delta
\]

Choose bounded continuous \( f_K \) as in the main part of the proof but apply it to \( |X_n| \) so that \( 0 \leq f_K(|X_n|) \leq |X_n| \). Then \( E(|X||X_n| \leq K) \leq E f_K(|X|) \). But \( E f_K(|X|) = \lim_{K \to \infty} E f_K(|X_n|) \). And \( E f_K(|X_n|) \leq E|X_n| \leq M \) by Proposition 9.11. So \( E|X| = \lim_{K \to \infty} E(|X||X_n| \leq K) \leq M < \infty \). □
9. Weak convergence

Required Exercises

**Exercise 9.1.** Suppose \( \{X_k\} \) are independent uniform \( U(0,1) \) random variables. Show that 
\[
\min_{1 \leq k \leq n} X_k \xrightarrow{D} Y
\]
and determine the law of \( Y \).

**Exercise 9.2.** Prove Corollary 9.5: if \( X_n \xrightarrow{P} c \) for a constant \( c \) and \( Y_n \xrightarrow{D} Y \), show that \( X_n Y_n \xrightarrow{D} cY \).

**Exercise 9.3.** Suppose \( X_n \xrightarrow{D} X \). Show that the laws of \( X_n \) are tight.

**Exercise 9.4.** Suppose \( X_n \in \mathbb{Z} \) and \( X_n \xrightarrow{D} X \). Show that \( \mathbb{P}(X \in \mathbb{Z}) = 1 \) and that \( \mathbb{P}(X_n = k) \rightarrow \mathbb{P}(X = k) \) for every \( k \in \mathbb{Z} \).

**Exercise 9.5.** Suppose \( X_n \) has density \( f_n(x) = 1 + \cos(2\pi nx) \) on \([0,1]\). Prove that \( X_n \xrightarrow{D} X \) (and determine the law of \( X \)).

**Exercise 9.6.** Suppose \( E(X_n^2) = 1 \). Show that \( F_n \) is tight.

**Exercise 9.7.** Suppose \( E(X_n^2) = 1 \). Show that \( \{X_n\} \) is uniformly integrable.

**Exercise 9.8.** Show that \( X \) is integrable if and only if for every \( \varepsilon > 0 \) there exists \( K \) such that 
\[
\int_{|X| > K} |X|dP < \varepsilon.
\]
(This is Corollary 6.6 on page 74)

**Exercise 9.9.** Suppose that \( \sup_n E(|X_n|f(|X_n|)) < \infty \) for some non-decreasing function \( f \) such that \( \lim_{x \to \infty} f(x) = \infty \). Show that \( \{X_n\} \) is uniformly integrable.

**Exercise 9.10.** The Lévy distance between two probability measures on \( \mathbb{R} \) is defined as
\[
d(F,G) = \inf\{\varepsilon > 0 : G(x - \varepsilon) - \varepsilon \leq F(x) \leq G(x + \varepsilon) + \varepsilon \text{ for all } x\}
\]
(i) Verify that this is a metric.

(ii) Verify that \( F_n \xrightarrow{D} F \) iff \( d(F_n, F) \to 0 \)

(iii) Verify that for every probability measure \( \mu \) on Borel sets of \( \mathbb{R} \) there exist probability measures \( \mu_n \) with finite support such that \( \mu_n \xrightarrow{D} \mu \). Show further that the support can be taken from \( \mathbb{Q} \), so that the space of distribution functions is separable in the Lévy metric.

**Definition 9.3.** We say that \( (X_n,Y_n) \xrightarrow{D} (X,Y) \) if for every bounded continuous function \( f : \mathbb{R}^2 \to \mathbb{R} \) we have \( E(f(X_n,Y_n)) \to E(f(X,Y)) \).

**Exercise 9.11.** Suppose \( (X_n,Y_n) \) are independent and \( X_n \xrightarrow{D} X \ Y_n \xrightarrow{D} Y \). Prove that \( (X_n,Y_n) \xrightarrow{D} \mu \) where \( \mu = F_X \otimes F_Y \) is the product measure.

**Exercise 9.12.** Suppose \( (X_n,Y_n) \xrightarrow{D} (X,Y) \). Prove that \( X_n^2 + Y_n^2 \) converges in distribution.

---

3Non-negative, triangle inequality, and \( d(F,G) = 0 \) only for \( F = G \)
Bibliography

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

$L_1$ metric, 5
$L_2$ metric, 5
$L_p$-norm, 63, 80
$\lambda$-system, 21
$\pi$-system, 21
$\sigma$-field, 10
$\sigma$-field generated by $X$, 41
distribution of a random variable, 42
Bernoulli random variables, 46
Binomial distribution, 12, 77
bivariate cumulative distribution function, 26
Bonferroni's correction, 12
Boole's inequality, 12
Borel $\sigma$-field, 41
Borel sigma-field, 11
Cantelli’s inequality, 70
cardinality, 3
Cauchy distribution, 117
Cauchy-Schwarz inequality, 62
centered, 68
Central Limit Theorem, 121
characteristic function, 113
characteristic function – continuity theorem, 117
Characteristic functions – uniqueness, 116
Characteristic functions – inversion formula, 116
Chebyshev’s inequality, 61
complex numbers, 112
conjugate exponents, 64
certainty condition, 8
Continuous mapping theorem, 107
converge in $L_p$, 65
converge in mean square, 65
convergence in distribution, 49, 103
converges in distribution, 127
converges in probability, 46
converges pointwise, 1
converges uniformly, 1
correlates with probability one, 47
correlation coefficient, 63
countable additivity, 8
covariance matrix, 130
cumulative distribution function, 22, 43
cylindrical sets, 28, 29
cylindrical sets, 28
DeMorgan’s law, 2
density function, 25, 78
diadic interval, 135
dependent random variables, 45
dependent random variables, 45
equal in distribution, 43
events, 7, 11
expected value, 58, 73
Exponential distribution, 78
exponential distribution, 25
Fatou’s lemma, 75
field, 7
finite dimensional distributions, 28
finitely-additive probability measure, 8
Fubini’s Theorem, 86
Geometric distribution, 77
Hölder’s inequality, 64, 80
inclusion-exclusion, 12
independent $\sigma$-fields, 33
independent events, 33
independent identically distributed, 46
independent random variables, 44
indicator functions, 3
induced measure, 43
infinite number of tosses of a coin, 135
integrable, 73
intersection, 2
Jensen’s inequality, 62
joint cumulative distribution function, 26
joint distribution of random variables, 43
Kolmogorov’s maximal inequality, 95
Kolmogorov’s one series theorem, 95
Kolmogorov’s three series theorem, 96
Kolmogorov’s two series theorem, 96
Kolmogorov’s zero-one law, 94
Kolmogorov–Smirnov metric, 4, 5
Kronecker’s Lemma, 98

Lévy distance, 110
law of $X$, 43
Lebesgue’s dominated convergence theorem, 75, 76
Lebesgue’s dominated convergence theorem – used, 77, 93, 106, 118
Levy’s metric, 5
Levy’s theorem, 97
Lindeberg condition, 123
Lyapunov’s condition, 124
Lyapunov’s inequality, 62

marginal cumulative distribution functions, 26
Markov’s inequality, 61
maximal inequality, Etemadi’s, 99
maximal inequality, Kolmogorov’s, 95
mean square convergence, 80
measurable function, 41
measurable rectangle, 85
metric, 4
metric space, 4
Minkowski’s inequality, 63
Minkowski’s inequality, 80
moment generating function, 60, 81
moments, 59
Monotone Convergence Theorem, 73
multivariate normal, 129
multivariate normal distribution, 130
multivariate random variable, 42

negative binomial distribution, 12
normal distribution, 25

Poisson distribution, 12, 77
Polya’s distribution, 12
Portmanteau Theorem, 106
power set, 1
probability, 7
probability measure, 8
probability space, 7, 11
product measure, 86

quantile function, 44, 105

random element, 41
random variable, 42
random vector, 42

sample space, 7
Schtefe’s theorem, 104
section, 85
semi-algebra, 9
semi-ring, 9
sigma-field generated by $\mathcal{A}$, 10
simple random variable, 57

simple random variables, 45
Skorohod’s theorem, 105
Slutsky’s Theorem, 105
Standard normal density, 78
stochastic process with continuous trajectories, 43
stochastic processes, 42
stochastically bounded, 53
symmetric distribution, 101
tail $\sigma$-field, 34
Tail integration formula, 88
Taylor polynomials, 111
tight, 53
tight probability measure, 13
Tonelli’s theorem, 86
total variation metric, 5
truncation of r.v., 52

uncorrelated, 68
uniform continuous, 24
Uniform density, 78
uniform discrete, 24
uniform singular, 24
uniformly integrable, 76, 109
union, 2

variance, 60

Wasserstein distance, 5
weak convergence, 49
weak law of large numbers, 68
zero-one law, 34, 94